

**Análisis y predicción de enfermedades pandémicas
mundiales**

Universidad del Valle de Guatemala

Facultad de Ingeniería



Análisis y predicción de enfermedades pandémicas mundiales

Trabajo de graduación en modalidad de Tesis presentado por

Ana Lucia Diaz Leppe


para optar el grado de Licenciada en Ingeniería en Ciencia de la
Computación y Tecnologías de la Información


Guatemala
2020


Vo.Bo:

(f) 
Ing. Tomás Gálvez

Tribunal Examinador:

(f) 
Ing. Douglas Barrios

(f) 
Ing. Tomás Gálvez

(f) 
Ing. Franz Josef Haidacher

Fecha de Aprobación: Guatemala, 14 de diciembre del 2020

Prefacio

Alrededor del mundo han existido enfermedades que ocurren dentro de poblaciones, lo que provoca que pasen cierto tiempo en circunstancias anormales. El objetivo de este proyecto es utilizar los sistemas de información geográficos GIS los cuales permiten almacenar, analizar y manipular datos, y a la vez desplegar su información geográficamente referenciada, así como herramientas de análisis de datos, con el fin de resolver problemas de planificación y gestión sobre el manejo de las pandemias, así poder asociar las enfermedades con distintos elementos geográficos. Las enfermedades seleccionadas para el presente trabajo son COVID-19, Ébola y VIH Sida.

El desarrollo de este proyecto está compuesto por análisis espacial y análisis de datos, utilizando la ubicación como un hilo conector para determinar la cantidad de personas infectadas y elementos de localización geográfica que pueden influenciar a la propagación de cada enfermedad. Se desarrolla el análisis exploratorio de los datos y predicciones sobre la propagación de las enfermedades, estos resultados se obtienen de un API para mantener las gráficas de predicciones actualizadas y con ello popular un Tablero Web para proveer una visualización objetiva de los resultados obtenidos en la investigación.

Agradezco a mis docentes y casa de estudio, por su alta calidad de enseñanza y distintos aprendizajes que he aplicado durante toda mi carrera. Por último, agradezco el haber convertido cada reto, en un triunfo más

ÍNDICE

Prefacio	v
LISTA DE FIGURAS	ix
LISTA DE CUADROS	xi
RESUMEN	xiii
ABSTRACT	xv
I. INTRODUCCIÓN	1
II. JUSTIFICACIÓN	3
III. OBJETIVOS	5
A. Objetivo general	5
B. Objetivos específicos	5
IV. MARCO TEÓRICO	7
A. Sistemas y tecnologías Web	7
1. Página Web	7
2. Framework	7
3. Angular	7
4. Arquitectura de angular	9
5. Angular CLI	10
6. iFrame	10
B. Interfaz de Programacion de Aplicaciones (API)	10
1. Flask	11
C. Servicios en la nube	11
1. Cloud Computing	12
2. Amazon Web Services (AWS)	12
3. Calendarización con CRON en Linux	12
4. Unicorn	12
5. Reverse Proxy NGINX	12
D. Data Science	12
1. ¿Qué es R?	13
2. ¿Qué es Python?	13
3. Clustering	13

4. Series de tiempo	14
5. Tablero de datos	16
E. Machine Learning	16
1. Redes Neuronales	16
2. Regresión de Cresta Bayesiana	17
3. Métricas para validar un modelo predictivo	19
F. Servicios de Información Geográfica	19
1. Bases de datos geográficas	20
2. ¿Qué es una capa en GIS?	20
3. ArcGis	20
G. Sigmoide	21
F. Enfermedades	22
1. Pandemia	22
2. Coronavirus	22
3. Ébola	24
4. VIH Sida	25
V. MARCO METODOLÓGICO	29
A. Análisis del estado actual	29
B. Limpieza de datos	30
C. Análisis Exploratorio y Clustering	30
D. Construcción de Sistema GIS	30
E. Predicciones	34
F. Diseño de la Arquitectura del Proyecto	35
G. API	36
H. Sistemas en la nube	36
I. FRONT-END	34
J. Pruebas de Usuario	37
VI. RESULTADOS	41
VII. DISCUSIÓN	81
VIII. CONCLUSIONES	87
IX. RECOMENDACIONES	89
X. BIBLIOGRAFÍA	91

LISTA DE FIGURAS

1. Características de angular.	8
2. Arquitectura de angular.	10
3. Arquitectura de una red neuronal.	17
4. Ejemplo de una sigmoide.	21
5. Diagrama de infraestructura del proyecto.	35
6. Mapa de COVID-19 con capa de mapa de calor por contagios confirmados.	41
7. Mapa de COVID-19 con capa de puntos por contagios confirmados.	42
8. Mapa de e COVID-19 con capa de densidad poblacional.	42
9. Mapa de COVID-19 este de Estados Unidos densidad poblacional versus contagios.	43
10. Mapa de COVID-19 oeste de Estados Unidos densidad poblacional versus contagios.	43
11. Mapa de COVID-19 con capa de áreas turísticas.	43
12. Mapa de COVID-19 con capa de hospitales y clínicas con la función de iniciar una ruta.	44
13. Mapa de COVID-19 con capa de vuelos.	44
14. Mapa de COVID-19 con capa de personas mayores por país.	45
15. Gráficas por país para el caso de COVID-19.	45
16. Mapa de Ébola con capa de mapa de calor por contagios confirmados.	46
17. Mapa de Ébola con capa de puntos por contagios confirmados.	46
18. Mapa de Ébola con capa de hospitales y clínicas con la función de iniciar una ruta.	46
19. Gráficas por país para el caso de Ébola.	47
20. Mapa de VIH SIDA con capa de mapa de calor para personas que reciben tratamiento.	47
21. Mapa de VIH SIDA con capa de puntos por personas que reciben tratamiento.	48
22. Gráficas por país para el caso de VIH SIDA.	48
23. Mapa de VIH SIDA con capa de índice de progreso social global.	48
24. Mapa de VIH SIDA con capa de hoteles y moteles alrededor del mundo.	49
25. Gráficos comparativos por países del caso de COVID-19.	49
26. Gráficos comparativos por provincias del caso de COVID-19.	50
27. Gráfico circular sobre comparación entre casos.	50
28. Gráficos de barras que muestra el aumento de cada caso respecto al día anterior.	51
29. Serie de tiempo para casos de contagio del caso de Ébola.	51
30. Serie de tiempo para casos de mortalidad del caso de Ébola.	51
31. Gráfico de caja y bigotes para el caso de Ébola respecto a años y meses.	52
32. Gráfico comparativo sobre los diferentes tipos de contagios del caso de VIH SIDA.	53
33. Gráfico sobre el comportamiento a través del tiempo por contagios homosexuales.	54

34. Gráfico sobre el comportamiento a través del tiempo por contagios de maternidad.	54
35. Gráfico circular sobre la proporción de contagios de VIH SIDA según el género.	55
36. Gráfico de serie de tiempo sobre contagios por medio de una relación heterosexual.	55
37. Gráfico sobre serie de tiempo por contagios por medio de inyección.	56
38. Gráfico de codo para el caso de casos confirmados de COVID-19.	56
39. Clúster para el caso de casos confirmados de COVID-19.	57
40. Query que muestra la agrupación de países de casos confirmados por COVID-19.	58
41. Gráfico de codo para el caso de casos de mortalidad sobre el caso de COVID-19.	59
42. Cluster para el caso de casos de mortalidad sobre el caso de COVID-19.	59
43. Query que muestra la agrupación de casos de mortalidad de COVID-19.	60
44. Gráfico de codo para el caso de VIH SIDA.	61
45. Cluster para el caso de VIH SIDA.	61
46. Query que muestra la agrupación de VIH SIDA.	62
47. Predicciones con Redes Neuronales a nivel general para casos de mortalidad.	63
48. Predicciones con Redes Neuronales a nivel general para casos de recuperación.	64
49. Predicciones con Redes Neuronales a nivel general para casos confirmados.	64
50. Predicciones con Redes Neuronales sobre casos confirmados por país.	65
51. Predicciones con Lasso sobre el caso de VIH SIDA.	65
52. Predicciones con ARIMA sobre el caso de VIH SIDA.	66
53. Codificación y decodificación de una gráfica en base 64.	62
54. Máquina EC2 en AWS.	68
55. Máquina EC2 en AWS ejecutándose.	69
56. Menú y leyenda para el área del tablero de COVID-19.	70
57. Mapa y menú para el área del tablero de COVID-19 sobre el estado actual.	70
58. Mapa y menú para el área del tablero de COVID-19 mostrando el área de clusters.	71
59. Mapa y menú para el área del tablero de COVID-19 mostrando el área de predicciones.	71
60. Menú y leyenda para el área del tablero de VIH SIDA.	72
61. Mapa y menú para el área del tablero de VIH SIDA mostrando el estado actual.	72
62. Mapa y menú para el área del tablero de VIH SIDA mostrando el área de clusters.	73
63. Mapa y menú para el área del tablero de VIH SIDA mostrando el área de predicciones.	73
64. Mapa y menú para el área del tablero de VIH SIDA sobre su análisis de estado actual.	74
65. Gráfico circular que busca la funcionalidad del botón de menú.	75
66. Gráfico circular que busca la funcionalidad de encontrar un país en específico.	76
67. Gráfico circular que busca la funcionalidad de ver la curva de crecimiento por país.	76
68. Resultados de la encuesta sobre el porcentaje funcionalidades encontradas.	76

69. Resultados de la encuesta sobre comentarios en donde se podría mejorar el tablero.	77
70. Resultados en la encuesta sobre sugerencias de funcionalidades.	77
71. Resultados en la encuesta sobre sugerencias de arreglos dentro del tablero.	77
72. Pregunta sobre la profesión de los encuestados.	78
73. Resultados en la encuesta en un gráfico circular del cumplimiento del objetivo general.	78
74. Resultados en la encuesta sobre las funcionalidades finales encontradas.	79
75. Resultados en la encuesta sobre la utilidad del tablero en sus diferentes profesiones.	79
76. Resultados en la encuesta sobre las necesidades a cubrir con sus profesiones.	79

LISTA DE CUADROS

1. Casos de uso.	38
2. Rutas y funciones del API.	67

RESUMEN

El objetivo general del proyecto consistió en apoyar en la preparación, respuesta y control de las enfermedades de COVID-19, Ébola y VIH Sida mediante una plataforma de información que muestra el estado actual y la predicción sobre el crecimiento de los contagios de cada enfermedad por país y por su situación general. Al mismo tiempo, que realiza un análisis de las estaciones del año en donde existe una alta posibilidad de contagio, y la edad y género en que predominan los casos de contagios por estas enfermedades. Para cumplir con esto se realizó una plataforma de información que se actualiza diariamente y que unifica los resultados de las distintas enfermedades ,por medio de mostrar los datos en un mapa geográfico interactivo que muestra la proporción entre los casos confirmados, recuperados y de mortalidad en cada país, su comportamiento en el tiempo y las áreas en el mundo en donde hay una alta cantidad de contagios, al igual que se estudian las distintas variables geográficas que pueden llegar a afectar en la propagación o control de la enfermedad. Por otro lado, se integró un tablero adicional que integra diferentes gráficas y resultados que permiten entender la situación de cada enfermedad y su comportamiento futuro el cual fue determinado por diferentes algoritmos de aprendizaje. Para cumplir con la validación de la utilidad y funcionamiento del sistema, se realizaron varias iteraciones del prototipo existente, en donde cada una de estas, se realizaron cambios según los comentarios recibidos por los diferentes usuarios.

Dentro de los resultados más significativos se pudo determinar en base a las encuestas que la aplicación si puede ser útil para brindar información necesaria para la preparación, respuesta y control de diferentes enfermedades dado a su grado de aceptación en visualizaciones de datos y diferentes comentarios, al igual que su alta exploración en los tableros GIS y sus comentarios positivos al respecto. Por otro lado, según los resultados se recomienda poner especial atención en los países que se encuentran con un alto nivel de pobreza y un alto nivel de vida salvaje ya que estos países no solo son de los más grandemente afectados, sino que podrían ser el inicio de una futura enfermedad.

ABSTRACT

The overall objective of the project was to support the preparation, response and control of the diseases of COVID-19, Ebola and HIV AIDS through an information platform that shows the current state and prediction of the growth of the contagion of each disease by country and its overall situation. At the same time, it conducts an analysis of the seasons of the year where there is a high possibility of contagion, and the age and gender in which the cases of infection by these diseases predominate. In order to achieve this, an information platform was created that is updated daily and that unifies the results of different diseases by showing the data in an interactive geographical map that shows the ratio of confirmed, recovered and mortality cases in each country, their behavior in time and the areas in the world where there are a high number of contagions, as well as the different geographical variables that may affect the spread or control of the disease the disease. On the other hand, an additional board was integrated that integrates different graphs and results that allow understanding the situation of each disease and its future behavior, which was determined by different learning algorithms. In order to comply with the validation of the utility and operation of the system, several iterations of the existing prototype were performed, in which each of these changes were made according to the comments received by the different users.

Among the most significant results, it was possible to determine from surveys that the application can be useful to provide information necessary for the preparation, response and control of different diseases given its degree of acceptance in data visualizations and different comments, as well as its high exploration in the GIS boards and its positive comments. On the other hand, it is recommended that special attention be paid to countries with high poverty and a high standard of wildlife, since these countries are not only the most severely affected but could also be the beginning of a future disease.

I. INTRODUCCIÓN

Los datos cuantitativos sobre las enfermedades seleccionadas para el presente trabajo pueden brindar información que ayuda a identificar la propagación, prevenir y controlar las enfermedades por medio del análisis de datos correspondiente. Esto se realiza por medio de relacionar los datos de la enfermedad tales como su localización, fallecidos, contagiados, cantidad de pruebas realizadas, hospitales, población y épocas del año.

El análisis de la propagación de enfermedades puede ser un factor muy importante en las consideraciones de acción y riesgo de una enfermedad pandémica. El presente proyecto consiste en el análisis de los datos cuantitativos y geográficamente referenciados de las pandemias y predicción de la propagación de estas enfermedades que se han presentado alrededor del mundo mostrando un alto impacto. Las enfermedades que se incluyen dentro del presente trabajo incluyen COVID-19, Ébola y VIH Sida. Con el objetivo de determinar posibles áreas de aparición y propagación de los contagios a través del tiempo.

II. JUSTIFICACIÓN

“Una pandemia es la propagación de una nueva enfermedad” (Vargas Dominguez and Palacios Alvarez, 2018) a lo largo de la historia estas enfermedades han disminuido el número de seres humanos y aumentan las calamidades dentro de las cuales poblaciones enteras quedan arruinadas, los servicios de salud se llenan hasta colapsar, se crea inseguridad y desconfianza entre la población por lo que dan flor a rumores de todo tipo que solo crean confusión y caos. Por otro lado, los estratos pobres de la población quedan más empobrecidos y sin medios para defenderse. Mientras las personas de clase media y alta van desmejorando su calidad de vida. Debido a todas estas razones dar una respuesta inmediata a este tipo de eventualidades es vital para la sociedad humana (Policlínica metropolitana, 2020).

Según la Organización Mundial de la Salud la pandemia de COVID-19 ha presentado un desafío sin precedentes para la salud pública debido a que ha presentado una dramática pérdida de vidas humanas. Provocando repercusiones en los sistemas alimentarios y exponiendo su fragilidad, los cierres de las fronteras, las restricciones comerciales y las medidas de confinamiento han impedido que los agricultores accedan a los mercados ya sea para comprar insumos y vender sus productos y que los trabajadores agrícolas cosechen sus cultivos interrumpiendo de esta manera cadenas de suministro de alimentos nacionales e internacionales. El trastorno social y económico causado por la pandemia ha provocado que decenas de millones de personas corran el riesgo de caer en pobreza extrema, el número de personas desnutridas que se estima que son de 690 millones pueda llegar a aumentar hasta 132 millones para el final del dos mil veinte. Sin medios para obtener ingresos durante los encierros, muchos no pueden alimentarse a sí mismos ni a sus familias. Dentro del área laboral, muchas personas han perdido su trabajo o han obtenido ingresos bajos e irregulares por lo que se ven motivados a seguir trabajando en condiciones inseguras exponiéndose a sí mismos y a su familia a riesgos adicionales. Los países que enfrentan crisis humanitarias están particularmente expuestos a los efectos del COVID-19. Por lo tanto, es fundamental responder rápidamente a la pandemia y al mismo tiempo garantizar que la asistencia humanitaria y de recuperación llegue a los más necesitados, para evitar que este se convierta en una catástrofe humanitaria y de una disminución en seguridad alimentaria que posiblemente provoque un retraso para los desarrollos que ya han sido logrados. Siendo así, es esencial desarrollar estrategias sostenibles a largo plazo que ayuden a los diferentes países a tomar decisiones sobre su población y estado actual.

Según la CDC (Centers for Disease Control and Prevention), la enfermedad del Ébola provocó un gran impacto socioeconómico especialmente en Guinea, Liberia y Sierra Leona.

Según las proyecciones de 2014 del Banco Mundial se estima que en el 2015 se perdieron 2,200 millones de dólares en el producto interno bruto (PIB) dentro de estos 3 países. Esto resultó en una menor inversión dentro del país y una pérdida sustancial en el crecimiento del sector privado. Por otro lado, tuvo un gran impacto dentro de la salud pública especialmente en los trabajadores de salud, desde el inicio del brote hasta noviembre del 2015 se notificaron un total de 881 contagios confirmados por trabajadores de salud, provocando 513 muertes en Guinea, Liberia y Sierra Leona. Dentro de Liberia se perdió 8% de los doctores, enfermeras y parteras a causa de esta enfermedad. Al mismo tiempo, el caso del ébola afectó severamente los servicios de salud y causó retrocesos en el tratamiento y control de enfermedades graves que incluyen a enfermedades como el VIH SIDA, Tuberculosis y Malaria provocando un descenso de 10,600 personas.

El VIH SIDA sigue siendo uno de los mayores desafíos de la salud pública mundial en especial en países de ingresos medios y bajos. Desde el inicio de la enfermedad, 75 millones de personas han sido contagiadas con el virus de VIH y alrededor de 32 millones de personas han muerto por el mismo. Esta enfermedad puede crear un impacto económico tanto para el individuo como para su familia debido a la pérdida de ingresos y un aumento de gastos como resultado al cuidado de los contagiados. Esto provoca que las ganancias, ahorros y las inversiones de las empresas puedan reducirse debido al aumento de los gastos relacionados con el sida y provoquen una menor productividad laboral. (Whiteside, 2008)

El proyecto tiene el propósito de informar a los usuarios y ayudarlos a desarrollar diferentes estrategias para cada respectivo país, continente o provincia, ya sea para observar el comportamiento de la enfermedad en el futuro y tomar decisiones tanto personales como gubernamentales. Sino también, para ayudarlos a definir los puntos geográficos más vulnerables y determinar así los puntos geográficos que necesitan un mayor abastecimiento de medicinas o atención médica, o se encuentren más susceptibles a una futura enfermedad. Esto se realiza por medio de un tablero de datos actualizado que es accesible para toda persona que contenga internet y quiera informarse de diferentes enfermedades pandémicas.

III. OBJETIVOS

A. Objetivo general

Apoyar en la preparación, respuesta y control de pandemias mediante una plataforma de información y predicción sobre el crecimiento de los contagios por país y género en que predominan los casos de personas afectadas.

B. Objetivos específicos

- Presentar la información sobre casos de personas confirmadas de contagio, casos de personas recuperadas, casos de mortalidad y distintos gráficos sobre los datos de cada enfermedad pandémica que incluyen curva de crecimiento, áreas de contagiado, gráficos de contraste en un tablero actualizado.
- Como parte del tablero realizar un análisis descriptivo de la data para proporcionar al usuario un conjunto de gráficos que describan la información de manera objetiva.
- Realizar una predicción sobre el crecimiento de los contagios de la enfermedad pandémica por país en base a la data histórica y su curva de crecimiento.

IV. MARCO TEÓRICO

Dado que el proyecto se centra en ciencias de los datos, ArcGis, sistemas web, servicios de la nube y análisis de datos de enfermedades, es necesario plantear algunos conceptos y parámetros que funcionen como base de apoyo para la comprensión de la investigación y entendimiento del proceso realizado.

A. Sistemas y tecnologías WEB

1. Página Web

Es un documento digital de carácter multimediático ya que es capaz de incluir audio, videos, imágenes y texto, se adapta a los estándares de WWW (World Wide Web) y se puede acceder a ella misma a través de un navegador web.

2. Framework

Es una plantilla, esquema o estructura conceptual de base tecnológica que nos permite desarrollar y organizar software, evitando así posibles errores de programación. Se trata de un conjunto de herramientas y módulos que se pueden reutilizar para varios proyectos.

Un framework se utiliza cuando se desea trabajar de una manera más organizada y con el objetivo de minimizar el riesgo de errores, al mismo tiempo que se desea aumentar el nivel de productividad. Esto nos permite ahorrar tiempo creando un espacio de trabajo propio, facilita la colaboración entre diferentes programadores ya que se encarga de estandarizar el código, por otro lado, dispone herramientas para atacar problemas específicos de la arquitectura de código, evitando escritura de código repetitivo y permitiendo concentrarse en la funcionalidad de alto nivel de la aplicación.

3. Angular

Angular obtiene su nombre del hecho de que las etiquetas de HTML están encerradas entre corchetes angulares. Angular se encuentra basado en el lenguaje de programación llamado TypeScript y está creado para crear aplicaciones web dinámicas. Angular se utiliza para desarrollar aplicaciones del lado del cliente de última generación, especialmente aplicaciones de una sola página "single page application". Cuenta con una serie de funcionalidades y herramientas que simplifican el desarrollo de las aplicaciones al mismo tiempo garantiza excelentes resultados de desempeño. Angular permite desarrollar para varias plataformas, como nativas móviles, web y de escritorio. Por otro lado, se pueden aplicar patrones de diseño para construir las asignaciones de una manera mantenible. Algunas funcionalidades de Angular son las siguientes:

1. Se requiere una cantidad mínima de código HTML al utilizar Angular.

2. **Modelo Vista Controlador (MVC):** es una arquitectura de software utilizada para separar el código por sus distintas responsabilidades, manteniendo distintas capas que se encargan de hacer una tarea concreta. Dentro de este mismo se realiza la separación de código en tres capas diferentes separados por su responsabilidad en Modelos, Vistas y Controladores. Los Modelos es la capa donde se trabaja con los datos, por lo tanto, tiene mecanismos para acceder a la información y para actualizar su estado. Las Vistas contienen el código que va a producir visualización de las interfaces del usuario, por lo que el código nos permite renderizar los estados de nuestra aplicación en HTML y JavaScript. Por último, los controladores contienen el código necesario para responder a las acciones que solicitan en la aplicación como visualizar un elemento, etc. Esta es una capa que sirve de enlace entre las vistas y los modelos, respondiendo a los mecanismos que puedan requerirse para implementar las necesidades de nuestra aplicación.
3. **Pruebas unitarias (Unit Testing):** Son un excelente marco de prueba conocido como Karma el cual fue diseñado por Google. Estas son muy útiles para verificar programáticamente que las secciones de código realizan los comportamientos esperados.
4. **Data Model Binding:** Se requiere menor cantidad de código para vincular la data de la aplicación a los identificadores de HTML que es el código que puede procesar el navegador.
5. **Desplazamiento virtual (Virtual Scrolling):** Utilizado para cargar dinámicamente elementos visuales en el desplazamiento virtual en Angular.

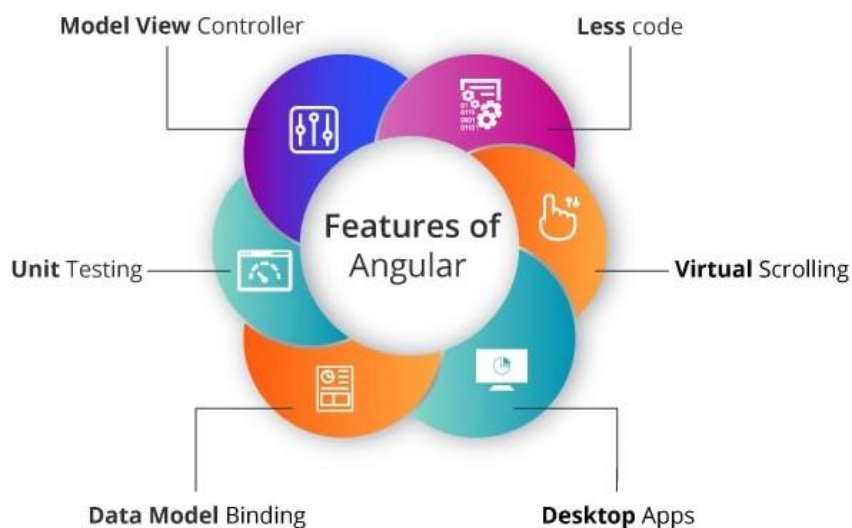


Figura 1 : Características de Angular.

4. Arquitectura de Angular

SPA (Single Page Application) es una arquitectura que nació de la evolución del navegador y la aparición de HTML5 como una tecnología de desarrollo de frontend que reemplaza a RIA (Rich Internet Application), utilizando una sola página actualizable para mostrar todo el contenido de la aplicación. Los bloques fundamentales que construyen una aplicación dentro de Angular se denominan NgModules. Los módulos son contenedores para almacenar los componentes y servicios de una aplicación. En Angular cada programa puede verse como un árbol de módulos jerárquico, declarados como clases de typescript. Los módulos son contenedores, por lo que lo primero que se debe guardar en ellos son componentes. Estos contienen una parte visual en código HTML y una parte funcional en Typescript, siendo este mismo el controlador.

Los bloques de construcción más importantes de la aplicación en Angular, y utilizados en el presente trabajo, se enumeran a continuación:

1. Módulos: las aplicaciones construidas con angular son modulares, esto significa que la aplicación se subdivide en bloques con funcionalidades distribuidas. El sistema de modularidad unificado con Angular se denomina "NgModules".
2. Componentes: el componente es útil para definir uno o más elementos mostrados en pantalla. Los componentes se crean, actualizan y se destruyen mientras los usuarios operan a través de la aplicación.
3. Directivas: Hay dos tipos de directivas en Angular, las cuales son directivas de atributo y de estructura. Al reemplazar, agregar o eliminar los componentes, las directivas estructurales pueden modificar los diseños. Para modificar el comportamiento o la apariencia de un componente, se utilizan las directivas de atributo.
4. Servicios: Un Servicio en Angular es una Clase con un objetivo o función específica a realizar.
5. Inyección de dependencias: El framework de Angular provee la facilidad de utilizar servicios como dependencia, la cual se puede llegar a utilizar en cualquier lugar del código sin la necesidad de volver a implementar la definición del servicio dentro de la clase.

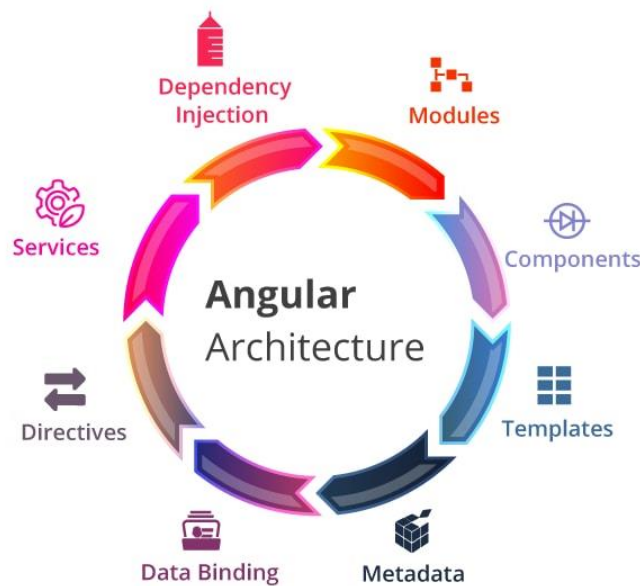


Figura 2: Arquitectura de Angular.

5. Angular CLI

El intérprete de línea de comandos de Angular que permite iniciar proyectos de desarrollo con Angular. Ayuda a facilitar la creación, ejecución, pruebas y puesta en producción de las aplicaciones; ya que nos arma el esqueleto de archivos y carpetas que se van a necesitar, junto con las herramientas de desarrollo del Framework.

Angular CLI es una herramienta de NodeJS por lo que se instala via npm. Esto es debido a que Angular CLI y sus componentes utilizados por angular se empaquetan como paquetes npm y se distribuyen del registro npm.

6. iFrame

Un iFrame es un componente que permite la inclusión de código en HTML o JavaScript externo dentro de una aplicación Web.

B. Interfaz de Programación de Aplicaciones (API)

“Un API provee una abstracción a un problema y específica como los clientes deben interactuar con los componentes del software que implementan una solución a un problema. Los componentes están distribuidos por la librería del software permitiendo tener múltiples aplicaciones. El API permite reutilizar construcciones de bloques que permiten modular piezas de la funcionalidad que sean incorporadas en una aplicación de usuario final. Un API es una interfaz bien definida que provee un servicio a otras piezas del software. Su propósito es proveer

una interfaz lógica a la funcionalidad de un componente mientras esconde detalles de su implementación." (API design for C++, 2011)

Un API permite que un servicio se comunice con esta aplicación sin necesidad de conocer su implementación a detalle, solamente con conocimiento de los métodos de uso del API. Los API son una forma simplificada de proporcionar acceso a datos o secciones de infraestructura de una entidad sin el riesgo de proporcionar la lógica de negocio al que consume los servicios del API. Una API es útil cuando los usuarios necesitan acceder a datos en tiempo real o sus datos cambian o se actualizan con frecuencia

1. Flask

Flask nos permite crear Web APIs con python, este mismo es un framework para python, el cual proporciona funcionalidad para crear aplicaciones web, incluida la gestión de soluciones HTTP y plantillas de representación. Flask se encarga de asignar solicitudes HTTP a funciones de Python como asignar rutas a una función. Cuando nos conectamos a la url flask comprueba si hay una coincidencia entre la ruta proporcionada y una función definida. Flask ejecuta el código en la función y muestra el resultado devuelto en el navegador. El proceso de asignación de URL a funciones se denomina como enrutamiento. Por último, utiliza una librería para utilizar el uso compartido de recursos entre orígenes (CORS) el cual es un mecanismo que utiliza encabezados HTTP adicionales para indicar a los navegadores que den a una aplicación web que se ejecuta en un origen, acceso a recursos seleccionados de un origen diferente.

Python tiene varios frameworks que se pueden utilizar para desarrollo de aplicaciones web y APIs. El más conocido es Django, el cual incluye muchas herramientas integradas. Aunque esto mismo puede ahorrar tiempo y esfuerzo, también puede resultar abrumador. Las aplicaciones en Flask resultan más sencillas ya que es posible aplicarlas sobre código ya existente con poco cambio a las mismas, por esta razón, Flask genera un valor agregado mayor al uso en el presente trabajo

C. Servicios en la nube

"Los servicios de la nube son servicios de software, datos o infraestructura que los proveedores externos alojan y ponen a disposición de los usuarios a través de internet. Estos facilitan el flujo de datos de los usuarios a través de internet" (RedHat, 2020).

1. Cloud Computing

Cloud computing consiste en ejecutar cargas de trabajo dentro de un entorno computacional en la nube.

2. Amazon Web Services (AWS)

EC2, por sus siglas en inglés, “Elastic Compute Cloud” es decir nube de computación elástica, es un servicio web que proporciona capacidad informática segura y redimensionable en la nube. Está diseñado para facilitar la computación en la nube a escala web para los desarrolladores. Proporciona un control completo de recursos informáticos y permite ejecutar el entorno informático proveniente de Amazon. Por otro lado, ofrece una plataforma informática amplia y profunda con opciones de procesamiento, almacenamiento, redes, sistema operativo, entre otras.

3. Calendarización con CRON en Linux

CRON se encarga de programar tareas, permitiendo a los administradores de sistemas programar tareas para que se ejecuten en un momento específico, ya sea de forma recurrente, diaria, semanal o mensual.

4. Unicorn

CRON se encarga de programar tareas, permitiendo a los administradores de sistemas programar tareas para que se ejecuten en un momento específico, ya sea de forma recurrente, diaria, semanal o mensual.

5. Reverse Proxy NGINX

Un servidor proxy inverso (NGINX) es un tipo de servidor proxy que normalmente se encuentra detrás del firewall en una red privada y dirige las resoluciones de los clientes al servicio en un puerto específico. Un proxy inverso proporciona a un nivel adicional de abstracción y control para garantizar el flujo fluido del tráfico de red entre clientes y servidores. Esto mismo ayuda en el equilibrio de carga ya que distribuye las solicitudes de los clientes en un grupo de servidores que maximiza la velocidad y la capacidad. Garantizando que ningún servidor se encuentre sobrecargado. Si un servidor deja de funcionar, el equilibrador de carga redirige el tráfico a los servidores en las líneas restantes. Por otro lado, NGINX puede comprimir datos entrantes y salientes, así como almacenar en caché el contenido solicitado, lo cual termina acelerando el flujo de tráfico entre clientes y servidores.

D. Data Science

Data Science o las ciencias de los datos, es el estudio de la extracción generalizable de conocimientos a partir de datos (Dhar, 2013). La información se refiere a los datos organizados

de diversas fuentes, como periódicos, internet, televisión, discusión etc. El conocimiento se refiere a la conciencia o comprensión sobre el tema adquirido de la educación o la experiencia de una persona. La información es la forma en la que llegamos a entender los datos. Por último, el conocimiento es la información relevante y objetiva que ayuda a sacar conclusiones. Los datos que han sido recopilados proporcionan información y cuando esta se combina con experiencia se traduce como conocimiento. ("Information vs data vs knowledge", 2021). Siendo así, Data Science es un área emergente de trabajo relacionado con la recopilación, preparación, análisis, visualización, administración y preservación de grandes colecciones de información (Stanton, 2012). Sus pilares fundamentales están formados por datos, tecnología y personas. Su proceso comienza desde realizar una pregunta sobre la problemática, obtener la información necesaria, explorar la data, modelar y exponer resultados por medio de visualizaciones. Los lenguajes más populares para trabajar Data Science son R, Python, Sql, Java, Scala, Julia y MATLAB, por su fácil manipulación de datos y amplias librerías de análisis, proyección y visualización de datos y resultados.

1. ¿Qué es R?

R fue lanzado en 1995 siendo un descendiente de la programación S. Fue escrito en C y Fortran. Entre sus ventajas está que R tiene un paquete para casi todas las aplicaciones cuantitativas y estadísticas, desde regresiones no lineales, redes neuronales, mapas etc. Utiliza ggplot2 para visualizar los datos y es de uso gratuito. Su lado negativo es que R no es un lenguaje rápido y no está creado para otro tipo de programación externa al análisis de datos. ("Information vs data vs knowledge", 2021)

2. ¿Qué es Python?

Python es un lenguaje de programación de uso general creado a finales de los 80, cuenta con una amplia gama de módulos específicos y alto soporte dentro de la comunidad. Cuenta con una curva de aprendizaje baja lo cual hace un lenguaje ideal para personas que están aprendiendo programación. Cuenta paquetes como pandas, scikit-learn y tensorflow que son utilizados para análisis y predicción de datos. ("Information vs data vs knowledge", 2021)

3. Clustering

Sirve para agrupar datos que tengan características similares. Es decir, sirve para segmentar datos en grupos iguales o similares. Para realizar clustering se deben seguir los siguientes pasos("Qué es y para qué sirve el clustering - un ejemplo de aplicación práctico - Conceptos Claros", 2021):

1. Decidir el número de clusters

2. Formar grupos con similitudes y asignar un centroide.
3. Medir el error de la máquina en su capacidad para encontrar el centroide de los grupos
4. Volver a formar grupos similares partiendo del centroide anterior. ("Qué es y para qué sirve el clustering - un ejemplo de aplicación práctico - Conceptos Claros", 2021)

4. Series de Tiempo

Es una secuencia de observaciones medidas en determinados momentos del tiempo, ordenados cronológicamente y generalmente espaciados entre sí de manera uniforme. (Instituto Superior Tecnológico de Panuco, 2021)

Se puede denotar como (modelo aditivo):

$$X_t = T_t + E_t + I_t$$

en donde:

T_t es la tendencia

E_t es la componente estacional

I_t es la componente aleatoria

Los tipos de series de tiempo son univariantes cuando solo analizan una serie temporal en función a su propio pasado. Mientras que las multivariantes analizan varias series temporales a la vez. Cuando se llega a construir ese modelo, generan cierta dependencia o relación entre pasado de las diversas series. (Instituto Superior Tecnológico de Pánuco, 2021).

Los componentes de tendencia se definen como un cambio a largo plazo que se produce el cambio a largo plazo de la media. Esta tendencia se identifica por un movimiento suave de la serie a largo plazo. Siendo así, los componentes estacionales se dan cuando las series temporales presentan periodicidad o variación cada cierto tiempo. Por otro lado, los componentes aleatorios se dan cuando no responde a ningún patrón de comportamiento, dando en sí el resultado de factores aleatorios que inciden de forma aislada en una serie de tiempo.

Una serie de tiempo estacionaria es estable a lo largo del tiempo, por lo que tiene una media y varianza constante en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos dos periodos de tiempo no de tiempo en el cual se encuentra calculando la covarianza:

Donde X_t es una serie de tiempo entonces cumple con las siguientes propiedades en donde Y_k , la covarianza al rezago k , es la covarianza entre los valores de X_t y X_{t+k} , los cuales están separados en k periodos:

$$\text{Media: } E(X_t) = E(X_{t+k}) = \mu$$

$$\text{Varianza: } V(X_t) = V(X_{t+k}) = \sigma^2$$

$$\text{Covarianza: } Y_k = E[(X_t - \mu)(X_{t+k} - \mu)]$$

Para volver una serie de tiempo estacionaria es difícil diferenciarlo. Es decir, restar el valor anterior del valor actual. A veces, dependiendo de la complejidad de la serie, es posible que se necesite más de una diferenciación. Por ejemplo, el paro laboral aumenta en general en invierno y disminuye en verano.

Las series de tiempo no estacionarias son series en las cuales la tendencia y variabilidad cambian con el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

Para realizar series de tiempo con Arima se debe entender cómo funciona el modelo ARIMA. El modelo de ARIMA es una "media móvil integrada auto regresiva". Es decir, es una clase de modelos que explica una serie tiempo determinada en función a sus propios valores pasados, es decir, sus propios retrasos y los errores de pronóstico retrasados, por lo que se puede utilizar su ecuación para pronosticar valores futuros. El término "Autorregresivo en ARIMA significa que es un modelo de regresión lineal que usa sus propios retrasos como predictores.

Los modelos AR significa que es un modelo auto regresivo puro es en donde Y_t depende solo de sus propios retrasos. Es decir, Y_t es una función de los rezados de Y_t .

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t$$

Por otro lado, MA es en el que Y_t depende solo de los errores de pronóstico retrasados.

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

El modelo de arima está compuesto por tres términos siendo: p el que marca el orden del término AR, q el orden del término MA y d es el número de diferenciados necesarios para que la serie temporal sea estacionaria. Por lo tanto, el modelo de ARIMA es uno en el que la serie de tiempo se diferencia al menos una vez para hacerlo estacionario y se combinan los términos de AR y MA. Obteniendo la siguiente ecuación:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

Siento la predicción Y_t igual a la constante en el cual se suma la combinación lineal de retrasos de Y más la combinación lineal de errores de pronóstico retrasados (Prabhakaran, 2020).

5. Tablero de datos

Un tablero de datos es una representación gráfica de métricas que nos permite visualizar los resultados del proceso de Data Science como vistas de tablas, gráficos de Análisis Exploratorio o visualización de gráficos de predicciones. Su último propósito es transformar los datos en información útil para orientar nuestra estrategia hacia la consecución de los objetos planteados. Un buen tablero de datos debe tener un gráfico limpio y ordenado que nos permita entender los datos que se plantean, los cuales nos permitan comparar datos de forma que nos permita proporcionarle al usuario soluciones visuales y comprensibles.

E. Machine Learning

Es una disciplina científica en el ámbito de inteligencia artificial que crea sistemas que aprenden a través del tiempo, esto significa que tomas los datos históricos y mediante un modelo predefinido intentan predecir valores o comportamientos futuros. Por lo que identifica patrones complejos entre millones de datos. La máquina que realiza el aprendizaje es realmente un algoritmo que revisa datos y predecir comportamientos futuros. ("¿Qué es Machine Learning? - Cleverdata", 2021) A Continuación se mencionan algunos algoritmos trabajados dentro de este proyecto:

1. Redes neuronales

Su nombre y estructura están inspirados en el cerebro humano, imitando la forma en que las neuronas biológicas se transmiten entre sí ya que el conocimiento es adquirido a través de un proceso de aprendizaje, al mismo tiempo el conocimiento se almacena en la fuerza de las conexiones sinápticas es decir los pesos de la red y la topología de la red define la forma de interconexión de las neuronas. Por otro lado, el proceso de aprendizaje o su algoritmo de entrenamiento, modifica los pesos de la red de la forma adecuada para obtener la función deseada para el sistema. Las redes neuronales son sistemas que no requieren que la tarea a ejecutar se programe, pues estas mismas generan y aprenden de la experiencia. (What are Neural Networks?, 2020). Estas se componen de capas de un nodo, que contiene una capa de entrada, capas ocultas y una capa de salida. Cada nodo, o neurona artificial, se conecta a otro y tiene un peso y un umbral asociados. Si la salida de cualquier nodo individual está por encima del valor del umbral que fue especificado, ese nodo se activa y envía datos a la siguiente capa de la red. De lo contrario, no se transmiten datos a la siguiente capa de la red. El nodo sumatorio acumula todas las señales de entrada multiplicadas por los pesos o ponderadas y las pasa a la salida a través de una función umbral o función de transferencia. En una red con varias capas

cada capa tiene su propia matriz de peso W , su propio vector de ganancias b , un vector de entradas netas n y un vector de salida a . Es importante mencionar que una red multicapa son más poderosas que las redes que cuentan con una sola capa.

La entrada neta a cada unidad puede describirse de la siguiente manera (Ejemplos didácticos de redes neuronales - Diinf, 2020):

$$neta_i = \sum_{i=1}^n W_i X_i = \overline{XW}$$

Las redes neuronales no requieren linealidad, tienen adaptabilidad a través del reentrenamiento y generan robustez por medio de la carga computacional distribuida y el paralelismo de la entrada y la salida. (Ejemplos didácticos de redes neuronales - Diinf, 2020)

Entre sus desventajas se encuentra que puede tener un tiempo de aprendizaje elevado además de que este cuenta con la necesidad de tener muchos datos de entrenamiento para realizar un correcto aprendizaje. (Ejemplos didácticos de redes neuronales - Diinf, 2020)

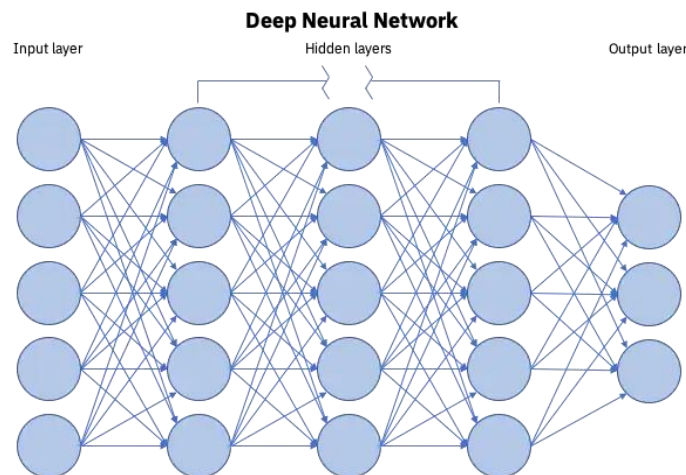


Figura 3: Arquitectura de una red neuronal.

2. Regresión de Cresta Bayesiana

Se utilizan técnicas de regresión bayesiana para incluir parámetros de regularización dentro del procedimiento de estimación, el cual se ajusta a los datos disponibles. La regularización utilizada en la regresión y clasificación de Ridge es equivalente a encontrar una estimación máxima posterior bajo un resultado anterior sobre los coeficientes con precisión. Mientras tanto λ se trata como una variable aleatoria para estimar los datos. Para obtener un modelo probabilístico la salida que tiene un resultado Gaussiano y es distribuida. (Soriano Flores, 2019)

$$p(y|\chi, \omega, \alpha) = \eta(y|\chi\omega, \alpha)$$

Las ventajas de utilizar este modelo es que se adapta a los datos disponibles y puede utilizarse para incluir parámetros de regularización en el procedimiento de estimación. Es muy útil cuando el tamaño del conjunto de datos es limitado. El enfoque bayesiano es un enfoque probado y comprobado y es muy sólido matemáticamente. Por lo que puede ser bastante útil cuando los datos van cambiando día a día ya que, en comparación con el aprendizaje por lotes, donde tenemos todo el conjunto de datos completo antes de entrenar el modelo, la regresión bayesiana arregla este problema al no necesitar almacenar datos y ajustarse a ellos. (Soriano Flores, 2019).

Dentro de la Regresión de Cresta Bayesiana los parámetros se estiman conjuntamente durante el ajuste del modelo, los parámetros de regularización se estiman minimizando la probabilidad marginal logarítmica. (Soriano Flores, 2019). El algoritmo puede ser bastante útil cuando tenemos datos insuficientes. Por otro lado, el resultado de un modelo de regresión bayesiana se obtiene a partir de una distribución de probabilidad, en comparación con las técnicas de regresión regulares. En comparación con otros algoritmos los pesos de los coeficientes se desplazan hacia los ceros, lo que provoca que se vuelva más estable.

3. Lasso

Su nombre deriva de operador de selección y contracción mínima absoluta. Es una técnica de regresión lineal que se creó con el objetivo de encontrar una técnica que fuera estable y que realizará una selección de variables. Por lo tanto, es una técnica de regresión lineal regularizada como regresión de cresta bayesiana que utiliza contracción. La contracción es donde los valores de los datos se reducen hacia un punto central, como una media. (Aprendizaje Automático y aplicaciones IMERL / FING / UdelaR, 2013). Lasso utiliza la regularización L1, el cual agrega una penalización igual al valor absoluto de la magnitud de los coeficientes. Este tipo de regularización puede resultar en modelos dispersos con pocos coeficientes. Este tipo de regularización puede resultar en modelos dispersos con pocos coeficientes. Algunos de los coeficientes pueden volverse cero y se eliminan del modelo. Por otro lado, las penalizaciones mayores dan como resultado valores de coeficiente más cercanos a ceros lo cual es ideal para producir modelos más simples. (Aprendizaje Automático y aplicaciones IMERL / FING / UdelaR, 2013).

Lasso proporciona una buena precisión dentro de la predicción, debido al reducir y la eliminar de los coeficientes pueden reducir la varianza. Por otro lado, ayuda a aumentar la interpretación de los datos al eliminar las variables irrelevantes que no

están asociadas con la variable de respuesta, de esta manera también reduce el riesgo de "overfitting" (VU Amsterdam, 2020).

4. Métricas para validar un modelo predictivo

Una vez se ha creado el modelo con el conjunto de datos de entrenamiento es necesario evaluar si el modelo generado tiene buena precisión con respecto a la variable de estudio, es decir es saber si los valores predichos a través del modelo que fue definido están lejos de los valores de los conjuntos de datos de validación ("How to validate a predictive model ? - Aspexit", 2020)

A.) MAE

Es el promedio de la diferencia absoluta entre el valor observado y los valores predichos. El error absoluto medio o MAE es un puntaje lineal, lo que significa que todas las diferencias individuales se ponderan por igual en el promedio. ("Evaluando el error en los modelos de regresión", 2021).

B.) MSE

El error medio cuadrático MSE mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. ("Evaluando el error en los modelos de regresión", 2021).

C.) Cross Validation

Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo que se llevará a cabo a la práctica. (LONDOÑO, 2016)

F. Servicios de Información Geográfica

Por sus siglas en inglés, GIS, o Sistema de Información Geográfica, es un conjunto de herramientas utilizadas para organizar, almacenar, comunicar y analizar grandes cantidades de datos geográficamente referenciados. Por medio de las herramientas que proporcionan los Sistemas de Información Geográfica se pueden realizar análisis de datos para encontrar patrones sociales, ambientales, de salud, económicos, entre otros. Dentro de un Sistema de Información Geográfica los datos se dividen en series de capas lógicas de información (Crosier, 1999).

1. Bases de datos geográficas

Una base de datos geográfica hace posible que datos geográficamente referenciados se almacenen en un formato estructurado que simplifica la administración, actualización, reutilización y uso compartido de los mismos. Las bases de datos son el lugar de

almacenamiento y administración de capas básicas en GIS. Dentro de cada capa se crean símbolos los cuales representan un punto objetivo dentro del mapa. (Crosier, 1999).

1. ¿Qué es una capa en GIS?

Son colecciones lógicas de elementos y tipos de datos como puntos, líneas, polígonos, matrices de información (Rasters), entre otros, los cuales proveen información descriptiva sobre la representación del mapa proporcionado. (Crosier, 1999).

Representaciones fundamentales de capa de información geográfica:

a. Entidades

Están conformadas por puntos, líneas y polígonos. Las entidades son representaciones de elementos ubicados sobre la tierra. Los puntos definen ubicaciones discretas de entidades geográficas que simbolizan una coordenada en específico. Las líneas representan la forma y ubicación de objetos geográficos demasiado estrechos para mostrarse como áreas. Las líneas se utilizan para representar una entidad como la distancia entre dos coordenadas. Por último, los polígonos son áreas cerradas que representan la forma y ubicación de entidades homogéneas como estados, parcelas, 5 tipos de suelo y zonas de uso.

b. Atributos

Los mapas transmiten información descriptiva a través de símbolos de mapa, códigos de colores, etiquetas, entre otros. En GIS los atributos descriptivos se administran en tablas que se basan en una serie de conceptos de base de datos relacionales. Las tablas de atributos proporcionan un modelo de datos para almacenar y trabajar con la información de los atributos mencionados anteriormente y como ligarlos estos (símbolos, colores, etiquetas, marcas de tiempo, operadores lógicos, etc) a una referencia geográfica. (Crosier, 1999)

c. Imágenes o Rasters

Para Sistema de Información Geográfica un Ráster se refiere a una o un conjunto de matrices de datos basados en celdas o píxeles, donde cada celda contiene data referenciable. (Crosier, 1999)

3. ArcGis

“ArcGis es un framework de desarrollo que implementa los conceptos de un Sistema de Información Geográfica y facilita el recopilar, organizar, administrar, analizar, compartir y distribuir soluciones sobre información geográfica. ArcGIS es utilizada por personas y entidades de todo el mundo para poner el conocimiento geográfico al servicio de los sectores del gobierno, empresa privada, investigación científica, educación y medios. “Este mismo se encarga de publicar información que sea accesible para cualquier usuario. ArcGIS se

encuentra disponible en aplicaciones de Escritorio, Web y Teléfonos Inteligentes, en los cuales incluye herramientas profesionales, mapas, recursos configurables y contenido compartido por la comunidad de usuarios, todo bajo una arquitectura online basada en la nube. ArcGIS también contiene librerías que amplían las capacidades de un Sistema de Información Geográfica, como los 'Living Atlas of the World', que son la principal colección de información geográfica de ArcGIS. En la cual incluye mapas, aplicaciones y capas de datos que se pueden integrar al ambiente de trabajo. (Esri, 2020)

Dentro de ArcGIS se puede crear, compartir y utilizar mapas inteligentes que se encargan de organizar, comprender y proporcionar grandes cantidades de información. Un mapa dentro de ArcGIS muestra, integra y sintetiza completas capas de información geográfica y descriptiva de cualquier fuente de datos georreferenciados, entre los cuales se pueden incluir datos topográficos, imágenes satelitales de La Tierra, mapas, entre otros. Dentro de su escalabilidad se permite configurar herramientas visuales como representaciones de datos (atributos) en un mapa geográfico (Living Atlas), acceder a modelos de análisis de datos, distintos gráficos de Barras, gráficos de tiempo, etc.(Crosier, 1999)

ArcGIS API Permite la comunicación con servicios de ArcGIS para retornar las soluciones de Información Geográfica programadas en el tablero en código de JavaScript. (Esri, 2020)

G. Sigmoide

Muchos procesos naturales y curvas de aprendizaje de sistemas complejos muestran una progresión temporal desde unos niveles bajos que se van acercando a un clímax que transcurre después de cierto tiempo. La función de sigmoide permite describir la evolución teniendo una forma de S. (Mitchell, 1997)

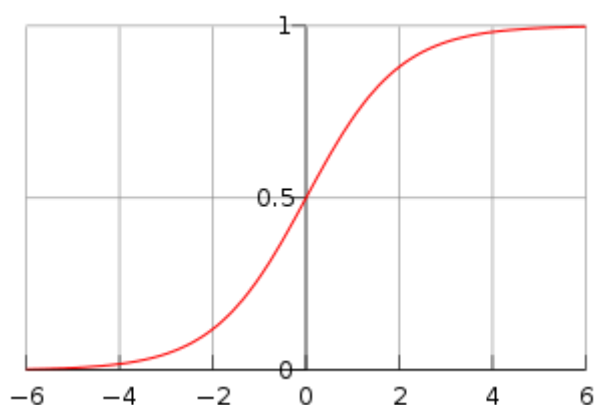


Figura 4: Ejemplo de una sigmoide.

H. Enfermedades

1. Pandemia

Una base de datos geográfica hace posible que datos geográficamente referenciados se almacenen en un formato estructurado que simplifica la administración, actualización, reutilización y uso compartido de los mismos. Las bases de datos son el lugar de almacenamiento y administración de capas básicas en GIS. Dentro de cada capa se crean símbolos los cuales representan un punto objetivo dentro del mapa. (Crosier, 1999).

2. Coronavirus

“Los coronavirus son una extensa familia de virus que puede causar enfermedades tanto en animales como humanos. En los humanos el coronavirus causa infecciones respiratorias que puede ir desde un resfriado a un síndrome respiratorio de Oriente Medio (MERS) y el síndrome respiratorio agudo severo (SRAS)” (Organización Mundial de la Salud, 2020).

a. COVID-19

“El COVID-19 es la enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente. Tanto este nuevo virus como la enfermedad que provoca eran desconocidos antes de que estallara el brote en Wuhan (China) en diciembre de 2019. Actualmente el COVID-19 es una pandemia que afecta a muchos países del mundo.” (Organización Mundial de la Salud, 2020).

Los síntomas de la enfermedad del COVID-19 pueden variar entre cada persona mientras algunas personas no presentan algún síntoma otros pueden presentar tos seca, cansancio y fiebre. Algunos síntomas poco frecuentes son la congestión nasal, dolor de cabeza, conjuntivitis, dolor de garganta, diarrea, pérdida de gusto u olfato, erupciones cutáneas o cambio de color en manos y pies.

Mientras el 80% de las personas se recuperan sin la necesidad de asistencia médica. Alrededor de 1 de cada 5 personas que contraen el COVID-19 presentan complicaciones y dificultad para respirar.

Entre los mayores afectados se encuentran las personas de edad avanzada, esto se debe a que el riesgo de presentar síntomas aumenta con la edad, las personas que se encuentran entre 85 años y mayores se encuentran entre el mayor riesgo.” En los Estados Unidos, aproximadamente un 80% de las muertes por esta enfermedad ha sido entre las personas de 65 años y mayores. Los riesgos son aún más altos para personas mayores cuando tienen afecciones de salud subyacentes.” (Mayo Clinic, 2020). Los residentes de hospitales y clínicas están a alto riesgo debido a la frecuencia que tienen con múltiples problemas subyacentes de salud los cuales pueden estar combinados con la edad avanzada. Dentro de estos mismos, los microbios se llegan a propagar con mayor facilidad entre las personas. Por otro lado, las

personas que presentan problemas pulmonares pueden llegar a presentar síntomas graves entre las enfermedades pulmonares que pueden llegar a producir el mayor daño se encuentran las personas que sufren de cáncer de pulmón y la enfermedad pulmonar obstructiva crónica. Sin embargo, si se cuenta con fibrosis quística, fibrosis pulmonar y asma, puede llegar a aumentar los síntomas del COVID-19. En casos aparte, las enfermedades cardíacas, diabetes, obesidad y problemas en la sangre pueden aumentar síntomas del COVID-19. En el caso de la diabetes y la obesidad es debido a que reducen la eficiencia del sistema inmunitario, por lo que aumentan el riesgo a cualquier infección en general. Las enfermedades cardíacas con mayor riesgo se encuentran la miocardiopatía, hipertensión pulmonar, enfermedades cardíacas congénitas, insuficiencias cardíacas, enfermedades de las arterias coronarias y presión alta. Las personas que sufren cáncer tienen un alto riesgo de sufrir grandes complicaciones con el contagio de COVID-19, sin embargo, el riesgo varía dependiendo del tipo de tratamiento que se está recibiendo. En los casos de anemia se puede dar un aumento de sufrir síntomas graves ya que causa que los glóbulos rojos se vuelvan duros, pegajosos y en forma de C, por lo que los glóbulos rojos mueren temprano y el oxígeno no puede ser transportado en todo el cuerpo. Si se presenta talasemia con anemia aumenta los síntomas debido a que el cuerpo no es capaz de producir suficiente hemoglobina afectando como los glóbulos rojos transportan el oxígeno. Las personas que presentan un sistema inmunitario debilitado como estar contagiado con VIH/sida, trasplante de médula ósea, uso de prednisona o algún trasplante de órganos se encuentran en alto riesgo de sufrir síntomas graves.

Una persona puede contraer COVID-19 al estar en contacto con otra persona infectada. “La enfermedad se propaga principalmente de persona a persona a través de las gotículas que salen despedidas de la nariz o de la boca de una persona infectada al toser, estornudar o hablar. Estas gotículas son pesadas por lo que no llegan muy lejos y caen rápidamente al suelo. Una persona puede contraer COVID-19 si inhala las gotículas procedentes de una persona infectada por el virus. Por lo que se debe mantener a un metro de distancia de los demás. Estas gotículas pueden caer sobre los objetos y superficies que rodean a la persona. De modo que una persona se puede infectar al tocar estos objetos y luego tocarse los ojos, nariz o boca.” (Organización Mundial de la Salud, 2020).

“Diversos estudios han demostrado que el virus de COVID-19, puede sobrevivir hasta 72 horas en las superficies de plástico y acero inoxidable, y al menos 4 horas en las superficies de cobre y menos de 24 horas en superficies de cartón.” (Organización Mundial de la Salud, 2020). Sin embargo, con desinfectantes domésticos se puede matar al covid en segundos.

3. Ébola

La enfermedad por virus del Ébola (EVE), antes conocida como fiebre hemorrágica del Ébola. Es una enfermedad grave que puede llegar a ser mortal para el ser humano. El virus es transmitido por animales salvajes y se propaga en las poblaciones humanas por transmisión de contacto con otras personas. “Los brotes de enfermedad por el virus del Ébola (EVE) tienen una tasa de letalidad que es aproximadamente del 50%. En brotes anteriores fue de 25% a 90%.” (Organización Mundial de la Salud, 2020).

El virus fue por primera vez detectado en 1976 en brotes simultáneos ocurridos en Nzara (Sudán del Sur) y Yambuku (República Democrática del Congo). La aldea en la que se produjo está situada cerca del río Ébola por lo que este mismo le da nombre al virus.

El brote del 2014 al 2016 fue mucho más extenso y complejo que el que se produjo en 1976. Además se extendió a diferentes países: empezó en Guinea, luego se propagó a través de Sierra Leona y Liberia

Mientras el Ébola se puede transmitir por medio de animales salvajes y tener contacto con la sangre de estos mismos. También se pueden transmitir vía sexual y pueden permanecer en el semen por un largo tiempo. Es debido a esto que se realizan los siguientes sistemas de control: todos los supervivientes de ébola y sus parejas sexuales reciben asesoramiento con el fin de adoptar prácticas sexuales seguras hasta que el semen arroje resultados negativos, los hombres deben ser sometidos a una prueba de detección del virus del semen 3 meses después del inicio de los síntomas y si en caso ya tengan caso positivo todos los meses hasta que se presenten dos veces negativo en RT-PCR en un intervalo de una semana. Antes de que el semen de negativo en 2 pruebas de detección debe de mantener una buena higiene personal y mantener lavándose las manos exhaustivamente con agua y jabón después de cualquier contacto físico con el semen. Por otro lado, se debe evitar el contacto con las personas que han muerto por esta enfermedad. Puede por otro lado, haber persistencia del virus en personas que se han recuperado de la enfermedad ya que el virus persiste en zonas menos accesibles por sistema inmunitario como los testículos, ojos y sistema nervioso central. Para las mujeres infectadas durante el embarazo el virus persiste en la placenta, líquido amniótico y el feto, al igual que en la lactancia.

El periodo de incubación es de 2 a 21 días. Las personas no son contagiosas hasta que aparecen los síntomas. Se caracterizan hasta la aparición de fiebre, dolor muscular, dolor de cabeza y garganta, lo cual va seguido de vómitos, diarrea, erupciones cutáneas, disfunción renal y hepática, y en algunos casos hemorragias internas y externas. Los laboratorios han mostrado una disminución del número de leucocitos y plaquetas, así como la elevación de las enzimas hepáticas. (Organización Mundial de la Salud, 2020).

Para poder llevar un buen control de la enfermedad es necesario tomar diferentes intervenciones; como la atención a los casos, vigilancia, rastreo de los casos, entierros en condiciones seguridad o movilización social. La participación de la comunidad es fundamental para el éxito y control de brotes. (Organización Mundial de la Salud, 2020).

Se mencionan a continuación diferentes medidas de control:

- Reducir el riesgo de transmisión por medio de animales salvajes: se deben utilizar guantes y otras prendas protectoras para manipular carne cruda de algún animal siendo principalmente el mono o murciélagos de fruta. Se debe cocinar hasta que se encuentren bien cocidos antes de consumirlo.
- Reducir el riesgo de transmisión de persona a persona: se debe reducir el contacto con pacientes infectados especialmente con sus líquidos corporales. Por lo que se debe utilizar guantes y equipo de protección para atender enfermos.
- Reducir el riesgo de posible transmisión sexual: los varones supervivientes a la enfermedad deben tener relaciones sexuales seguras e higiénicas durante los 12 meses siguientes después de haber presentado síntomas o hasta que sus muestras de semen den dos veces negativo para el virus del Ebola. “La OMS no recomienda el aislamiento de los pacientes convalecientes de ambos sexos una vez que sus muestras de sangre hayan dado negativo para el virus del Ébola”.
- Medidas de contención de los brotes: inhumación rápida y segura de los cadáveres infectados al igual que se deben identificar rápidamente las personas que hayan tenido contacto con algún infectado y someterlas en observación durante 21 días. Por otro lado, se deben separar las personas sanas de las personas enfermas para evitar la propagación, mantener una buena higiene y limpieza en los centros de salud.

4. VIH Sida

El VIH continúa siendo uno de los mayores problemas para la salud pública mundial tomando ya 33 millones de vidas. Debido a esto se ha mantenido una mejor prevención, diagnóstico, tratamiento y atención en los siguientes años. Se estima que a finales del 2019 había 38 millones de personas con VIH. Siendo la proporción de 68% de adultos y 53% menores se encontraban en un tratamiento antirretrovírico (TAR) de por vida, al igual que el 85% de las mujeres embarazadas previniendo que el virus fuese transmitido a sus hijos. La OMS define como poblaciones clave a las siguientes personas: hombres que tienen relaciones sexuales con hombres, personas que se inyectan drogas, trabajadores sexuales y clientes y personas transgénero. El aumento de la vulnerabilidad de VIH suele estar asociado a factores legales y sociales, lo que aumenta la exposición a situaciones de riesgo y crea obstáculos para acceder a servicios de prevención, detección y tratamiento de la respectiva

enfermedad. Más de dos tercios de todas las personas con VIH viven en la región de África. El VIH se puede diagnosticar mediante pruebas de diagnóstico rápido los cuales brindan resultados el mismo día. Actualmente, aunque no existe cura para la infección por el VIH existe el tratamiento antirretroviral que permiten mantener controlado el virus y prevenir la transmisión hacia otras personas.

El VIH sida fue identificada en la década de los 80 en estados unidos, la cual sigue siendo uno de los mayores desafíos de la salud pública mundial en especial en países de ingresos medios y bajos. Se estima que desde que surgió 32 millones de personas han muerto por causas relacionadas con VIH. (Organización Mundial de la Salud, 2020).

El virus de la inmunodeficiencia humana (VIH) ataca el sistema inmunitario y debilita la defensa contra muchas infecciones y determinados tipos de cáncer. Mientras destruye células inmunitarias, impide el funcionamiento normal de la inmunidad terminando en inmunodeficiencia. La fase más avanzada de la infección por el VIH es el síndrome de inmunodeficiencia adquirida (sida) que puede tomar años en manifestarse.

El VIH se transmite a través del intercambio de determinados líquidos corporales de la persona infectada, como la sangre, leche materna, semen y secreciones vaginales. También se puede transmitir de madre a hijo durante el embarazo y el parto. Es importante reconocer que las personas con VIH que reciben tratamiento antirretrovírico (TAR) y han suprimido la carga vírica ya no pueden transmitir VIH a sus parejas. ("HIV/AIDS", 2020)

Los comportamientos que pueden aumentar el riesgo de que una persona contraiga VIH son los siguientes ("HIV/AIDS", 2020):

- Tener relaciones sexuales de penetración sin preservativo.
- Padecer otra enfermedad de transmisión sexual como sífilis, herpes, clamidiasis, gonorrea etc.
- Compartir agujas, jeringuillas, soluciones de droga o materiales de drogas inyectables.
- Recibir inyecciones, transfusiones sanguíneas sin garantías de seguridad.
- Pincharse accidentalmente con una aguja.

El VIH se puede prevenir utilizando preservativos femeninos y masculinos durante la penetración puede proteger contra la propagación de las enfermedades. Los preservativos masculinos de látex tienen un efecto protector de 85% contra la infección de VIH. (Organización Mundial de la Salud, 2020). Por otro lado, toda persona que se encuentre entre las poblaciones clave mencionadas por la OMS debe someterse a pruebas de detección de VIH al igual que su pareja. La circuncisión médica masculina voluntaria es una buena práctica ya que este reduce el 50% de riesgo de que un hombre resulte infectado por el VIH, esta

práctica es clave ya en lugares donde hay una gran cantidad de infectados de VIH como África oriental y meridional. Por otro lado, los consumidores de droga inyectables se pueden proteger utilizando una inyección material estéril siendo la aguja y jeringa la cual no deben compartir bajo ninguna circunstancia. El tratamiento de la dependencia, como el tratamiento sustitutivo con opioides para los dependientes de estas sustancias ayuda a reducir el riesgo de transmisión de VIH. Por otro lado, en los casos de transmisión vertical (madre a hijo durante el parto o lactancia) se puede reducir utilizando antirretrovíricos para la madre e hijo durante la gestión y periodo de lactancia. Un número cada vez mayor de países están alcanzando tasas muy bajas de transmisión maternoinfantil de VIH por medio de la utilización de antirretrovírico de por vida. Por último, aunque la tuberculosis no está relacionada directamente con el VIH es una de las causas de muertes más comunes por esta enfermedad ya que provoca 1 de cada 3 muertes de los infectados por VIH. Por lo que para prevenir se recomienda que se realicen un tratamiento antituberculoso y antirretrovírico. Por lo que, si una persona presenta tuberculosis o VIH deben hacerse pruebas de ambas enfermedades. Por último, si una persona tuvo algún encuentro de riesgo con esta enfermedad debe tomar antirretrovirales en las 72 horas siguientes a la exposición.

V. MARCO METODOLÓGICO

A. Análisis del estado actual

Se inicia el presente trabajo con la investigación sobre las distintas enfermedades que se han propagado en varias partes del mundo y de las cuales las personas no presentan inmunidad. Dentro de estos, se selecciona COVID-19, VIH Sida y Ébola. Se escogió COVID-19 ya que es una enfermedad que ha presentado un desafío durante el año 2020, siendo nuestro tiempo actual, este ha provocado una "dramática pérdida de vidas humanas a nivel mundial, presentando un desafío para la salud pública, sistemas alimentarios y el área laboral, provocando una caída en la economía y social, provocando que millones de personas caigan en el riesgo de caer en pobreza extrema mientras aumenta la cantidad de personas desnutridas." (World Health Organization , 2020). Por otro lado, aunque el VIH SIDA fue una enfermedad que apareció en los ochentas, es una enfermedad que sigue teniendo una alta presencia hoy en día el cual sigue siendo una enfermedad mortal que afecta a personas en todo el mundo afectando no solo la economía de las personas contagiadas, sino que sigue siendo un estigma o tabú en el entorno. Por último, el caso del Ébola, aunque ya no es una enfermedad de gran impacto hoy en día si provocó una gran impacto no solo es descensos y contagios sino que también a nivel socioeconómico y salud pública dentro de África y su propagación en diferentes áreas del mundo. Este último nos permite estudiar una enfermedad que ya ha pasado y su propagación geográfica.

Los datos que se escogieron fueron "Novel Corona Virus Datasets" este repositorio es operado por el centro de ciencia e ingeniería de sistemas de la Universidad de Hopkins al igual que es apoyado del equipo ESRI living Atlas. Se escogieron estos datos debido a que estos se actualizan diariamente y recopilan información de diversas fuentes tales como World Health Organization, DXY.cn pneumonia, BNO News, Comisión Nacional de salud de la república popular china , China DCD, departamento de salud de Hong Kong, Gobierno de Macau, Taiwan CDC, US CDC, Gobierno de Canadá, Departamento gubernamental de salud de Australia, centro europeo de prevención y control de enfermedades (ECDC), ministerio de salud de Singapur y ministerio de salud de Italia.

Dentro del caso de VIH SIDA se manejaron dos diferentes datasets; el primero fue obtenido de World Health Organization (WHO) el cual tiene data actualizada desde el 2019 y nos muestra la cantidad de personas que se encuentran en tratamiento de VIH SIDA alrededor del mundo. Sin embargo, para poder cubrir las diferentes poblaciones clave definida por la OMS se trabajó con el dataset proporcionado por el centro europeo para prevención y control de enfermedades. Dentro de este se obtiene información de los siguientes perfiles; contagios por

relación heterosexual, contagios por relaciones homosexuales, contagios por transmisión vertical y contagios por inyecciones u drogas.

Para el caso de Ébola se utilizaron los datos de “Ébola 2014-2016 outbreak” se escogió este dataset debido a que este es un dataset organizado que recopila información de who y de OCHA Services. Permitiendo unificar las diferentes fuentes de información y realizar un análisis sobre el mismo.

B. Limpieza de datos

Se prosiguió con realizar una limpieza de datos, con el objetivo de asegurar la calidad de los datos y minimizar el riesgo que supondría tomar decisiones con datos poco precisos, erróneos o incompletos. Es decir, no tomar datos erróneos o ignorar datos de valor.

Para el caso de COVID-19 se realiza una limpieza dentro de Python, reemplazando por ceros las variables que se encuentran como dato nulo, dado que algunos datos se encontraban como nulos debido a que no había comenzado la enfermedad. Es necesario manejar los valores nulos ya que los valores nulos afectan nuestro desempeño y capacidad predictiva, ya que estos tienen el potencial de afectar nuestras estadísticas (Sayyed, 2020). Por lo tanto, puede producir predicciones engañosas. Por consiguiente, se realiza una comparación entre la fecha actual y la primera fecha de inicio de los contagios provocados por la enfermedad, esto se realizó con el objetivo de poder realizar una comparación de fechas y obtener como incrementan o decrecientan los contagios diariamente, conocido como diferenciación.

Para el caso de Ébola y VIH SIDA, se tuvo que tratar de manera diferente debido a que estos no se encontraban en un formato que no solo no se podían analizar ya que contenía columnas informativas o no se encontraban en un formato que permitiera analizar dentro de un tablero GIS el cual necesita estar establecido por columnas de país, latitud, longitud y diferentes columnas para cada fecha de datos. De la misma manera se determinaron que todos los valores nulos dentro de las fechas fueran marcados como cero, y se eliminaron las columnas y filas informativas. Esto se realizó para los diferentes casos como dentro de VIH SIDA para la información obtenida de WHO y para las diferentes poblaciones clave obtenidas, y para los casos de mortalidad y confirmados de Ébola.

C. Análisis exploratorio y clustering

Se llevó a cabo un análisis exploratorio para cada enfermedad. Dentro del caso de COVID-19 debido a que los datos cambian diariamente se decidió realizar un diagrama que compare el aumento de casos respecto al día anterior. Del mismo modo, se decidió trabajar un

“Pie Chart” o gráfico circular en donde podemos observar casos de personas confirmadas, casos de recuperación y casos de mortalidad, esto se realizó para mostrar proporciones con el propósito de comunicar la magnitud de estos casos y poder comparar cual es el caso predominante. Por otro lado, se realizaron gráficos donde se compara los países y provincias más y menos afectadas. Sin embargo, gracias al tablero GIS se descartaron estos gráficos dentro de la versión final del tablero. Para el caso de VIH Sida se realizó un gráfico de barras que muestra la comparación de contagios entre las diferentes poblaciones clave con el propósito de determinar cuál es la población clave que se encuentra más afectada. Al mismo tiempo, se realizó diferentes gráficos para cada población en base a su evolución en los diferentes años. Por último, se realizó un gráfico “Pie Chart” o gráficos circulares con el propósito de ver los contagios respecto al género. Dentro del caso de Ébola se realizaron gráficos de series de tiempo para visualizar la actividad de la enfermedad en base a contagios y casos de mortalidad, siempre visualizando los diferentes países por separado. Por otro lado, se realizó un diagrama de caja y bigotes para visualizar los grupos de años y meses a través de cuartiles. Al mismo tiempo que nos permite identificar los valores atípicos que se encontraron en los años y meses. Al mencionar que es un valor atípico se hace referencia a que su valor es numéricamente distante al resto de los datos. La ventaja que se obtiene al utilizar estos datos es que ocupan menos espacio lo cual es bastante conveniente para no saturar el tablero de información.

Se utilizó un algoritmo de clustering para los casos de COVID-19 y VIH SIDA, con el objetivo de clasificar los países en grupos de transmisión, es decir grupos de países que fueron infectados por el virus en periodos de tiempo similares. Sin embargo, “40% de las infecciones en los estados unidos, están dadas por personas que están contagiadas por coronavirus pero que nunca, nunca se sienten enfermos ni muestran los síntomas de la enfermedad” (National Geographic, 2020). Por lo que se realizó un cluster para casos de mortalidad y casos confirmados, mostrando dentro del tablero el caso del cluster del caso de mortalidad. Para definir la cantidad de clusters a utilizar se hizo un gráfico de codo. Por otro lado, para el caso de VIH SIDA se utilizó para los datos de WHO en donde se determina la cantidad de personas que se encuentran en tratamiento de VIH SIDA. Del mismo modo, el gráfico de codo ayudó a determinar la cantidad de clusters que se deban utilizar.

D. Construcción de Sistema GIS

Se escogió trabajar con un tablero que muestre un mapa geográfico debido a que estamos hablando de enfermedades que se propagan alrededor del mundo. Por otro lado, se escogió trabajar con un tablero GIS de Esri ya que este permite crear, compartir y utilizar mapas inteligentes en los cuales no solo se puede crear y administrar una base de datos geográfica sino

que también permite utilizar capas de Esri sino de otras instituciones lo cual nos permite enriquecer el mapa aumentando la comprensión y conocimientos para llegar a tomar mejores decisiones y de esta manera ayudar al usuario a resolver problemas con análisis espacial.

Dentro de ArcGIS Online se crea un nuevo Map Viewer que nos permite agregar un tablero de mapa geográfico para cada enfermedad a analizar, con el objetivo de luego ser utilizadas dentro de código en JavaScript por medio del ArcGis API que se agrega con una sola invocación del Map View solicitado. Se fueron agregando las respectivas capas para cada enfermedad de la manera siguiente:

1. COVID-19

Se agregó “Novel Corona Virus Datasets” como nuevo Living Atlas de ArcGIS Online, aunque inicialmente se había utilizado capa web para cargar los datos este se cambió por el Living Atlas proporcionado por Esri ya que realizaba las mismas tareas de manera más optimizada que el la capa Web. Se creó un “HeatMap” o mapa de calor para los datos de “Novel Corona Virus Datasets” para conocer en qué lugares se encuentra una mayor concentración de personas infectadas con la enfermedad en base a una gama de colores, funcionando como un raster, el cual fue determinado por el atributo de confirmados el cual es una variable cuantitativa. Al mismo tiempo y con el mismo atributo, se seleccionó la misma data pero para “Counts and Amounts (Size)” o Recuentos y cantidades, que funciona como un punto dentro del mapa. Estos mapas de símbolos utilizan una lógica intuitiva según la cual los símbolos más grandes equivalen a un número de casos mayor, determinando de esa manera en que los países con mayor cantidad de contagios tengan un tamaño mucho mayor que los que tienen menor cantidad de contagios. Dentro de esta última capa se programó una Gráfica de Pie o gráfico circular que nos brinda la proporción de fallecidos a razón del tiempo, recuperados y oficialmente confirmados para cada país, con el objetivo de ir realizando una proporción de los diferentes casos por país. Permittiéndonos entender a mayor complejidad la enfermedad.

Una vez terminado el análisis exploratorio sobre el dataset se prosiguió a tomar en cuenta todas las variables que pueden afectar a la propagación de la enfermedad dados por los resultados de clustering. Se creó una capa de adultos mayores, alrededor del mundo, datos obtenidos por las Naciones Unidas, sobre la cantidad de personas mayores a 65 años por país el cual es un dataset que se tiene actualizado hasta el año 2019. Se incluyó lo anterior dentro del mapa como “Counts and Amounts (Size)” determinando de esta manera los países que cuentan con una mayor proporción de adultos mayores. Se incluyó un símbolo que simboliza al adulto mayor para diferenciar esta capa de la anterior. Por otro lado, se incluyó una capa de densidad poblacional en cada país, datos obtenidos de “simplemaps” validada por la NASA, United States Census y USGS para ver la población en países, pueblos y ciudades, con el objetivo de determinar

zonas de riesgo debido a una alta densidad poblacional dentro de un sector. Por otro lado, se incluyó un “Living Atlas” de “OpenStreetMap” con el objetivo de mostrar hospitales y clínicas en todos lados del mundo para que las personas puedan localizar hospitales o clínicas y poder iniciar una ruta a algún hospital, esta capa fue solicitada por químicos farmacéuticos debido a que para entender y decidir los pasos a seguir siempre es bueno conocer acerca tanto de la situación actualizada de las enfermedades y sus hospitales cercanos ya que los hospitales pueden llegar a actuar como una variable externa. Por último, fue incluido un “Living Atlas” sobre rutas aéreas con el objetivo de que el usuario pueda tomar decisiones en base a futuros viajes a partir del contagio de COVID a nivel mundial o un análisis de si un país se está viendo contagiado por la cantidad de vuelos que lo transitan.

2. Ébola

Se agregó “Ebola 2014-2016 outbreak” como nueva capa Web dentro de ArcGis Online. Se creó un nuevo “HeatMap” o mapa de calor para determinar las regiones mayor concentración de Ébola. Similar al caso anterior, se incluyeron un hospitales, clínicas y población. Sin embargo, al este ser un conjunto de datos estático (es decir datos que ya no se actualizan) Este fue el mismo caso para Recuentos y cantidades, los cuales muestran la concentración de Ébola por país, debido a que el Ébola es un caso que se ve afectado por la cantidad de animales que existen dentro de la región, debido a partir de ellos se pueden transmitir el contagio, se incluyó un “Living Atlas” que marca los lugares en el mundo en donde se encuentran puntos críticos de biodiversidad los cuales son regiones caracterizadas por niveles excepcionales de endemismo vegetal. Dentro de esta capa se marcan los 36 puntos críticos de biodiversidad en donde se contiene 1,500 especies de plantas vasculares y existe una alta actividad humana que ha provocado una pérdida de 70% de su superficie original.

3. VIH/SIDA

Se agregó la capa de World Health Organisation (WHO) en donde podemos observar a personas que se encuentran en tratamiento de VIH SIDA alrededor del mundo como nueva capa Web dentro de ArcGis Online tanto como “Counts and Amounts (Size)” y como mapa de calor, con el objetivo de analizar si esta enfermedad por región y por país, al mismo tiempo que se agregó el dataset para ver los lugares con mayor población, hospitales y clínicas. En este caso, se agregaron los diferentes moteles registrados que existen a lo largo del mundo debido a que se está observando diferentes posibles lugares con actividad sexual. Por otro lado, se agregaron los países que tienen una alta comunidad de LGBTQ+ con el objetivo de determinar si los índices de contagio de la enfermedad son influenciados por el tamaño o la presencia de dicha comunidad.

Para poder observar los diferentes casos planteados por las poblaciones clave que fue definida por la OMS se agregó una capa como “Counts and Amounts (Size)” en donde se colocó

un símbolo para cada caso planteado observando de esta manera: para el caso de transmisión por inyecciones se agregó un símbolo de jeringa, para transmisión por medio de relaciones heterosexuales se agregó una bandera heterosexual, para transmisión homosexual se incluyó una bandera LGBTQ+ con un símbolo de VIH SIDA y, por último se agregó un símbolo para mostrar maternidad.

E. Predicciones

Para COVID-19 se utilizaron dos algoritmos clave: el primero fue regresión de la cresta bayesiana y el segundo fue redes neuronales estos fueron principalmente escogidos debido a que es una data que se actualiza y cambia diariamente. Se utilizó regresión de cresta bayesiana bayesiana, ya que este es un algoritmo utilizado cuando los datos cambian o se actualizan diariamente debido a que no necesita almacenar datos y se ajusta a los datos que recibe en tiempo real, en el tipo de regresión se trabajó con regresión de cresta bayesiana ya que se encarga de minimizar los riesgos en el aprendizaje del algoritmo por medio de reducir en gran medida la varianza y de esta manera mejorar la eficacia de los estimadores para la predicción. Por otro lado, se utilizó el algoritmo de redes neuronales debido a que estas tienen la capacidad de aprender y modelar relaciones complejas y no lineales, lo cual ayuda ya que significa que las relaciones de los datos entre las entradas y las salidas pueden ser no complejas sin tener que ser lineales. Este también cuenta con la ventaja de que una vez tenga aprendido las entradas iniciales y sus relaciones puede llegar a inferir relaciones entre los datos actuales y los datos futuros, lo que hace que el modelo generalice y prediga los datos futuros. Por otro lado, las redes neuronales no imponen ninguna restricción sobre las variables de entrada y como estas deben distribuirse. Por último, las redes neuronales tienen la capacidad de manejar datos que pueden ir variando en el tiempo y tienen una varianza no constante, gracias a su capacidad para aprender relaciones ocultas en los datos sin imponer relaciones fijas entre los datos. (Mahanta, 2017). Todo esto resulta beneficioso debido a que el dataset de COVID-19 es un dataset que tiene nuevos resultados y cambia diariamente. Este algoritmo se ejecuta para los casos sobre confirmados, recuperados y fallecidos, al mismo tiempo que se creó una función con el objetivo que se ejecute este algoritmo sobre cada país al ser solicitado, siendo una predicción dinámica ya que necesita que el usuario ingrese por país para generar la predicción.

En el caso de VIH SIDA se utilizaron dos algoritmos clave: siendo el algoritmo de series de tiempo con ARIMA y el algoritmo de Lasso estos fueron principalmente escogidos debido a que se contaba una data limitada. Utilizando el algoritmo de ARIMA se determinó si era una serie de tiempo estacionaria o no estacionaria, se determinó que esta misma era no estacionaria ya que se realizó un proceso de diferenciación en donde se restó la fecha inicial a la fecha actual. Se

escogió utilizar el modelo de ARIMA ya que nos ayudan a pronosticar series de tiempo que son relativamente cortas y no se cuenta con suficiente data para realizar una predicción, tal y como se encuentra la data actual. Al mismo tiempo el modelo de ARIMA tiene una estructura fija y está diseñada específicamente para datos de series de tiempo, por lo que se ajusta bien es decir puede manejar bien los altibajos(Davydenko,2016). Por otro lado, se escogió trabajar con el algoritmo de Lasso, ya que este puede proporcionar una buena precisión dentro de la predicción, porque la reducción y la eliminación de los coeficientes pueden reducir la varianza sin un aumento sustancial de la tendencia de los datos, lo cual nos ayuda a la limitada cantidad de datos que se tiene del caso de VIH SIDA que se presenta por años. Por otro lado, Lasso ayuda a aumentar la interpretación de los datos al eliminar las variables irrelevantes que no están asociadas con la variable de respuesta, de esta manera también reduce el riesgo de "overfitting"(VU Amsterdam, 2020).

Para el caso de Ébola ya que tiene un "dataset" limitado y la data obtenida hace referencia al brote que tuvo el caso de Ébola por lo que no cuenta con los datos de cuando este mismo comenzó a bajar ni los datos actuales. Se abstuvo de realizar predicciones sobre el mismo.

F. Diseño de la arquitectura del proyecto

Se muestra un diagrama de la estructura del proyecto, a continuación, se explicará en detalle cada uno de los componentes.

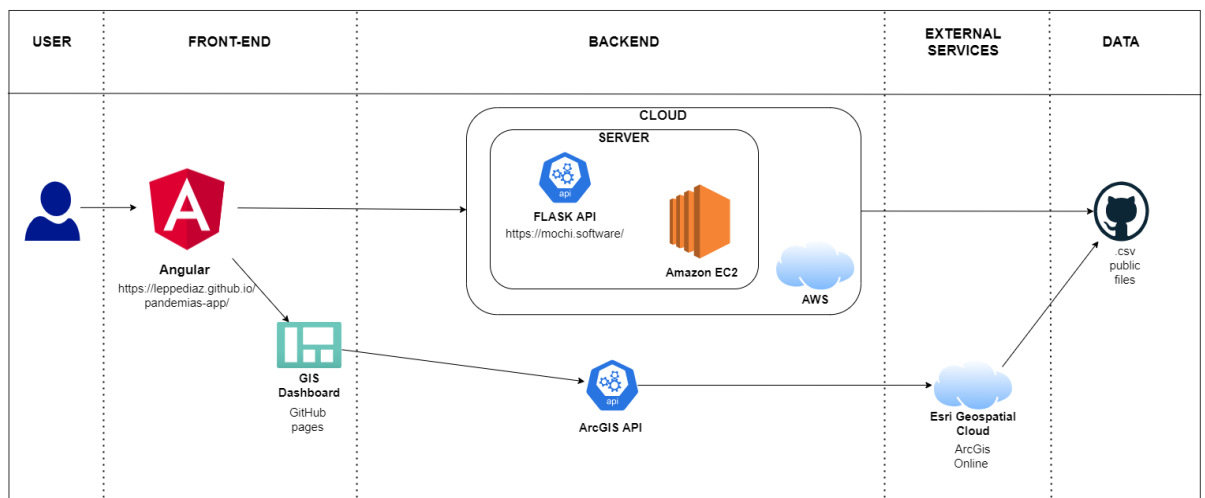


Figura 5: Diagrama de infraestructura del proyecto.

G. API

Se realizó un API utilizando Flask debido a que este permite ser agregado a aplicaciones ya existentes de manera rápida y sencilla, por lo que es más adecuadas para una aplicación contenida al mismo tiempo que nos brinda herramientas, librerías, tecnologías que nos permiten construir un API web en una corta cantidad de tiempo. Dentro de flask se otorgan rutas las cuales corren un proceso que ejecuta la predicción o gráfica y retorna una imagen en formato de base 64, permitiendo obtener resultados actualizados a la fecha por medio de una solicitud HTTP "GET" de COVID-19. Por último, se emplea una librería que facilita el uso compartido de recursos entre orígenes (CORS) para indicar a los navegadores que den a una aplicación web que se ejecuta en un origen, acceso a recursos seleccionados de un origen diferente con el objetivo de poder realizar el dashboard con los resultados del API. Se realizó una limpieza dentro del API de manera en que se diferenciara las rutas en donde se desea volver a realizar una predicción a cuando solo se desea consultar una predicción. Al igual que se eliminaron todas las impresiones y dejaron las funciones lo óptimas posibles. Esto se llevó a cabo con el objetivo de no saturar al servidor al mismo tiempo que aumentar la rapidez del API.

H. Sistemas en la nube

Para publicar el API y poder hacerla accesible en cualquier parte del internet, se utilizó el servicio de Máquinas Virtuales en la Nube de AWS, denominado EC2 (Elastic Cloud Compute). Se empezó creando la cuenta en el portal de Amazon Web Services, luego se procedió a seguir los pasos por defecto para crear la Máquina Virtual en EC2. Creada la máquina virtual, se le colocó una IP Pública para poder acceder a la máquina virtual desde cualquier lugar, se le apertura el puerto de http para que el API tenga un medio de salida del servidor y se accedió a la máquina por medio del protocolo SSH, protocolo de comunicación segura por medio de una consola. Ya dentro de la máquina se procedió a copiar el repositorio dentro de los archivos de la máquina y por medio del Servicio de Calendarización CRON de Linux se colocó que diariamente se ejecuten las funciones del API para generar las predicciones, así poder proveer las predicciones actualizadas y prevenir que las mismas se generen cada vez que se requiera el contenido del API.

Para poder colocar el API como un servicio de la máquina virtual se instaló Gunicorn, el cual ayuda a que el API se reinicie cuando encuentre cualquier tipo de error, así como poder correr el API en múltiples instancias denominadas 'threads' o hilos, así mejorar el rendimiento y disponibilidad del servicio del API. El siguiente requerimiento para colocar el API accesible reside en un servicio 'Reverse Proxy' para realizar un redireccionamiento interno del programa

que corre dentro de la máquina virtual hacia a la IP Pública, externa y accesible desde cualquier punto en el internet, en el puerto 80, el cual es el puerto http estándar y no requiere ningún agregado al dominio al momento de ingresar la dirección del API en el código o en el navegador web.

El objetivo de publicar el API consiste en poder realizar consultas que retorna las imágenes de las predicciones diarias como 'base64' y colocar las mismas en la página Web para formar un Tablero actualizado. Para poder realizar esto dentro del código de Angular se obtuvieron las dificultades de que las consultas al API por medio de este código se deben de realizar por medio de consultas seguras y con headers de CORS, por lo que se agregó el código lo necesario para habilitar flask-cors y para volver la consulta al API de forma segura se realizó una reserva del dominio mochi.software con ayuda de el plan de estudiante de GitHub en 'name.com', se ingresó a la consola de la página y se redireccionó el dominio a la IP del servidor EC2. Luego, para la seguridad SSL se procedió a crear una cuenta de CloudFlare que ofrece servicio de SSL gratis, se generaron DNS y se cambiaron los DNS Records del dominio para que apunten a los de CloudFlare.

I. FRONT-END

Se decidió utilizar Angular debido a su facilidad de vincular datos a los controles HTML sin necesidad de código especial al igual que su cantidad mínima de código, nivel de escalabilidad y el manejo de sus componentes los cuales nos permiten desarrollar porciones de código HTML con código TypeScript con funcionalidades específicas.

El proyecto se manejó con Angular CLI el cual nos permitió iniciar el proyecto y crear el esqueleto para el desarrollo de la aplicación. Dentro de Angular se cargó los mapas GIS de las diferentes enfermedades por medio de un iframe utilizando la conexión de ArcGis API. De igual manera, se cargó el dashboard que tiene las predicciones y el análisis de COVID-19 actualizadas dentro de un iframe. Cabe resaltar que este caso solo se produjo con los datos de COVID-19 debido a que para Ébola y VIH Sida solamente se cuenta con data histórica, por lo que los resultados se muestran directamente desde angular sin necesidad de realizar ninguna conexión.

J. Pruebas de Usuario

Durante este proyecto se realizaron varias encuestas para ir validando la experiencia de usuario que tiene el proyecto, validación de objetivos, exploración de las funcionalidades y posibles alcances que puede llegar a adoptar. Para la experiencia del usuario se brindó una lista de funcionalidades para los primeros prototipos en donde se fue validando que los usuarios encontrarán las diferentes funcionalidades y brindarán sus comentarios. Para el caso de

validación de objetivos, se llevó a cabo en la última fase del prototipo en donde se realizaron preguntas de una selección, al mismo tiempo que se dejó preguntas abiertas en donde el usuario pueda escribir el uso que le puede dar a la aplicación y los diferentes cambios que le haría.

Se definieron los diferentes casos de uso que puede llegar a adoptar el proyecto siendo los siguientes:

Nombre de caso de uso	Actor	Función
Análisis Gubernamental	Cualquier entidad gubernamental	Análisis y visualización de la información con el objetivo de tomar una decisión dentro de su país.
Análisis y entendimiento de distintas enfermedades pandémicas para uso personal	Cualquier persona de cualquier edad, género y profesión.	Ayuda a enriquecer el conocimiento de cualquier usuario de una manera segura y completa, al igual que a tomar decisiones personales.
Administración de productos por lugares geográficos	administradores, ingenieros industriales, químicos farmacéuticos, ingenieros bioquímicos	ayuda a planificar, proyectar, instalar, poner en marcha, dirigir y controlar productos al mismo tiempo que se toman medidas sanitarias.
Análisis y recolección de data para orientar al diagnóstico de un paciente	Personal médico	Ayuda al personal médico a tener data de la enfermedad que le permita orientar su diagnóstico.

Tabla No 1: Casos de uso

Según estos casos de uso, se realizaron encuestas en donde se compartió el link: <https://leppediaz.github.io/pandemias-app/> a diferentes personas siendo a sistemas de personal médico o practicantes de medicina de la Universidad Francisco Marroquín, estudiantes de la Universidad del Valle u otras universidades y profesionales de distintas carreras. Al mismo tiempo que se compartió encuestas en 'Google Forms'.

Las preguntas, aunque varían dependiendo de la fase del prototipo fueron primordialmente las siguientes:

A.) Primera fase:

1. ¿Has encontrado el botón de menú?
2. ¿Has cambiado las diferentes capas de información para observar mejor una enfermedad?
3. ¿Has buscado un país específico?

4. ¿Qué cambiarías o modificarías en este tablero?
 5. ¿Has visto la curva de crecimiento?
- B.) Segunda fase:
1. ¿Qué funciones viste en el Dashboard? (se mostró una lista de checkbox con las funcionalidades actuales en las cuales iban presionando check)
 2. ¿Qué te perdió en el dashboard?
 3. ¿Qué te gustaría ver en el dashboard?
 4. ¿Qué errores se deben arreglar dentro del dashboard?
- C.) Tercera fase:
1. ¿Cuál es tu profesión?
 2. ¿Te serviría este tablero para apoyar en la preparación, respuesta y control de enfermedades? (SÍ/NO/PARCIALMENTE)
 3. ¿Qué funciones lograste encontrar? (listado)
 4. ¿Para qué te serviría este tablero?
 5. ¿Qué más te gustaría ver en este tablero?

VI. RESULTADOS

Durante la realización del presente trabajo se realizaron diferentes fases que son de suma importancia para validar el cumplimiento de los objetivos del proyecto. Por lo que se irá presentando los resultados en base a las diferentes fases del proyecto:

A. Construcción de Sistema GIS

1. Tablero COVID-19 en GIS

Se creó una capa de "HeatMap" dentro de GIS en donde se observó una alta concentración de COVID en las áreas de Norteamérica expandiéndose hacia centro américa, Europa con excepción a Europa del norte, países árabes e india, es de recalcar que estos mapas se actualizan de manera diaria por lo que los resultados pueden ir variando en base a cómo se comporta esta enfermedad con el tiempo. La capa de HeatMap o Mapa de Calor nos deja observar la propagación de la enfermedad a nivel continental.

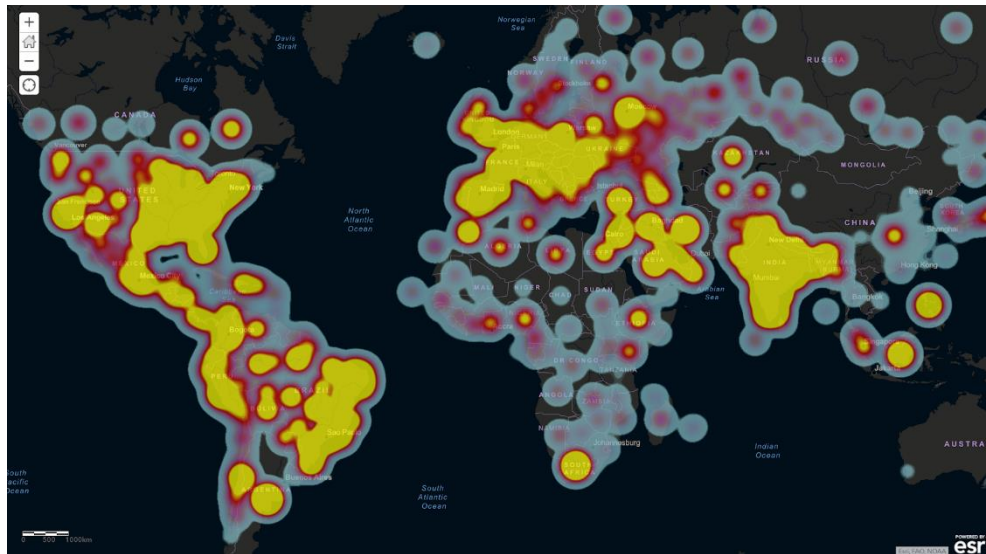


Figura 6: Mapa de COVID-19 con capa de mapa de calor por contagios confirmados.

Por otro, la capa "Counts and Amounts (Size)" o Recuentos y cantidades, nos permite ver la expansión por países y realizar comparaciones entre los diferentes países alrededor del mundo.



Figura 7: Mapa de COVID-19 con capa de puntos por contagios confirmados.

Se realizó una programación para que en base a cuando se acerque el usuario en el mapa pueda ir observando diferentes características de cada país. Dentro de este mismo, cada punto morado significa un punto en donde existe alta población y por lo tanto el riesgo a contagio puede ser mayor. Del mismo modo, se agregó una capa para observar los lugares turísticos de cada país en donde puede haber aglomeraciones de personas.

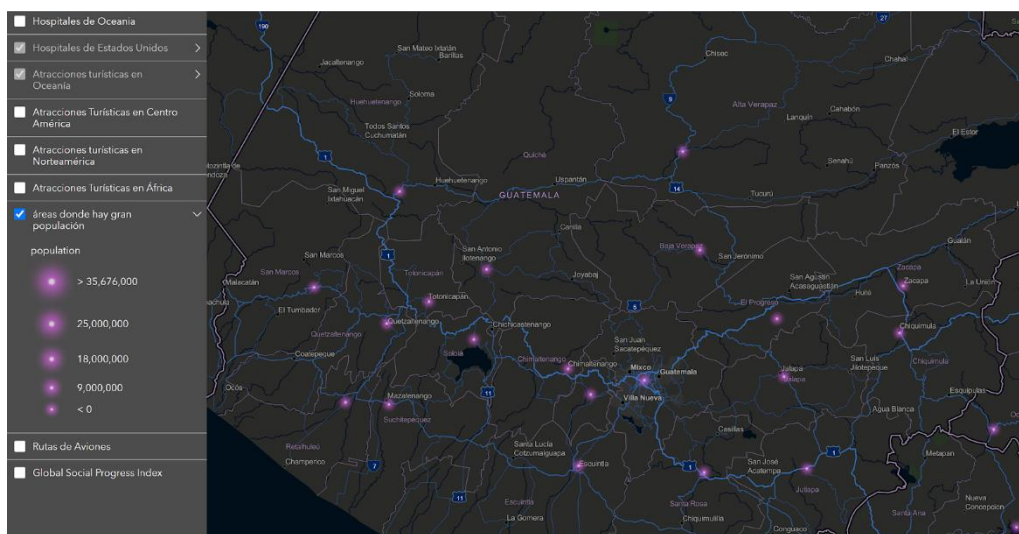


Figura 8: Mapa de e COVID-19 con capa de densidad poblacional.

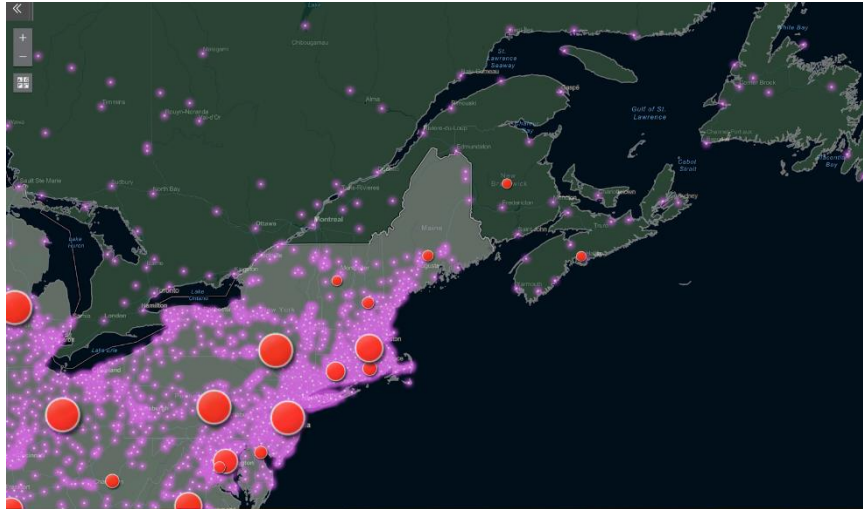


Figura 9: Mapa de COVID-19 este de Estados Unidos densidad poblacional versus contagios.

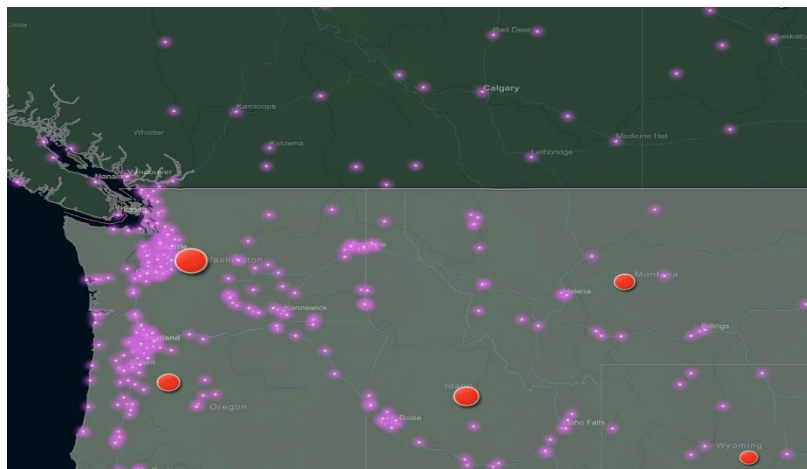


Figura 10: Mapa de COVID-19 oeste de Estados Unidos densidad poblacional versus contagios.

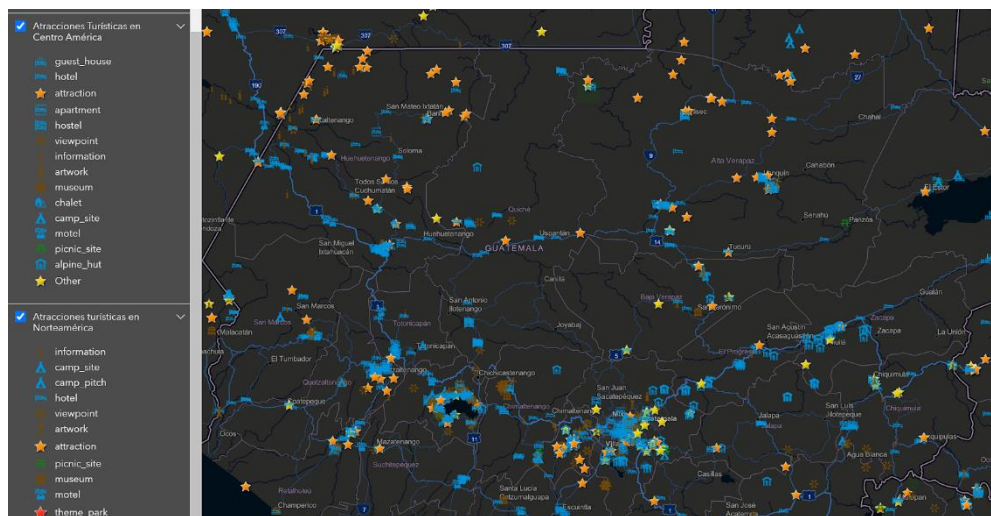


Figura 11: Mapa de COVID-19 con capa de áreas turísticas.

Se agregó una capa de hospitales y clínicas en donde al realizar click sobre ella el usuario es capaz de obtener la ruta al hospital.

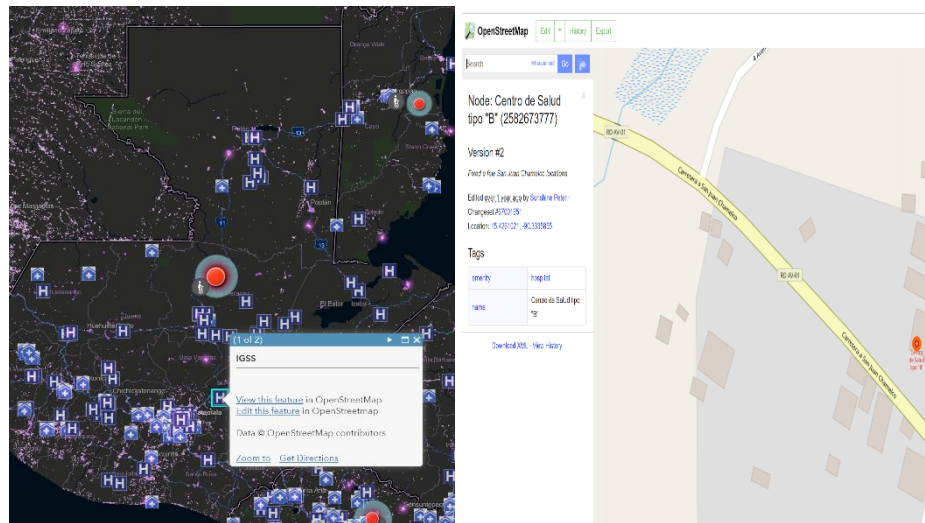


Figura 12: Mapa de COVID-19 con capa de hospitales y clínicas con la función de iniciar una ruta.

Por otro lado, se agregó una capa de vuelos de aviones con el objetivo de que el usuario pueda ver la situación de la enfermedad y en base a esta misma tomar la decisión de un vuelo.

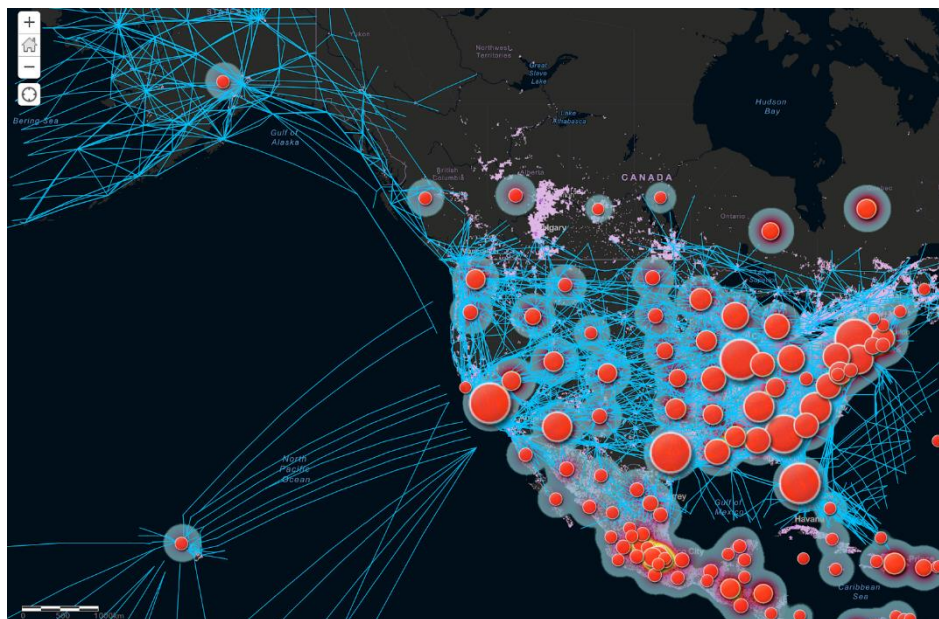


Figura 13: Mapa de COVID-19 con capa de vuelos.

Se utilizaron los datos de Naciones Unidas en “World Population Prospects” sobre personas de mayor edad en cada país. En donde se puede observar que los países europeos tienen una alta cantidad de personas mayores, lo cual contrasta con la baja cantidad de personas

mayores en África. La primera imagen nos muestra la situación de personas mayores en Europa, mientras la segunda nos muestra la poca cantidad de personas mayores en África.

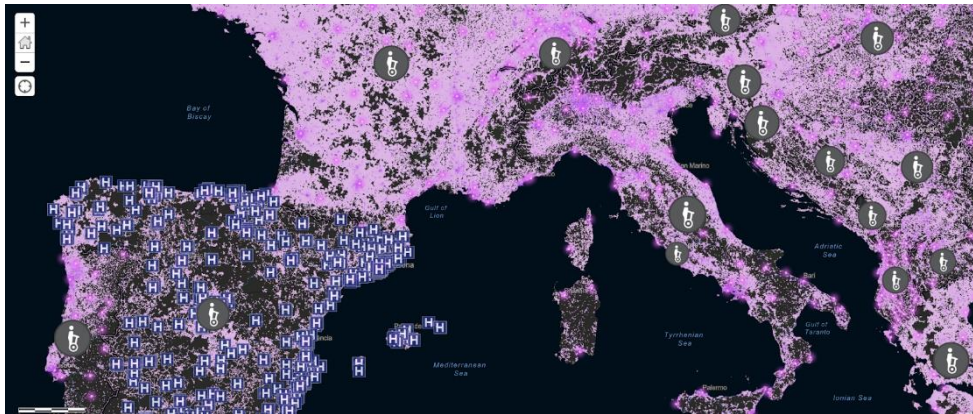


Figura 14: Mapa de COVID-19 con capa de personas mayores por país.

Dentro de cada país se realizaron diferentes tablas que muestran la proporción de sus casos confirmados, recuperados y fallecidos. Se puede validar el nombre y la cantidad de cada componente realizando click sobre ellos o posicionando el puntero encima. Por otro lado, se utilizaron los datos de Naciones Unidas sobre personas de mayor edad alrededor del mundo para mostrar la curva de crecimiento de personas de mayor edad en cada país.

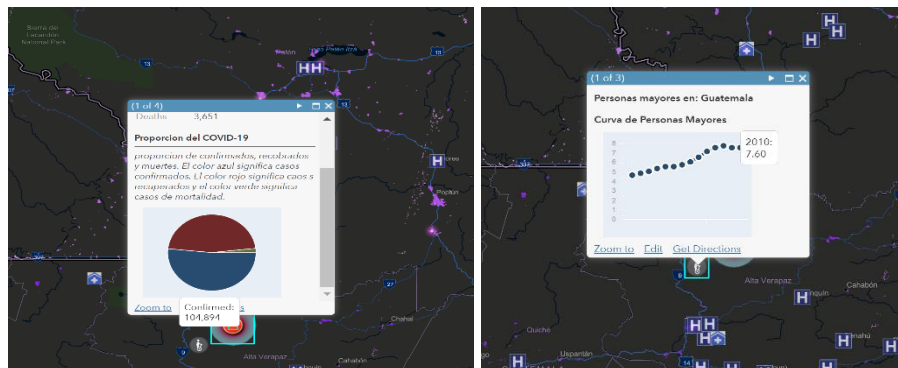


Figura 15: Gráficas por país para el caso de COVID-19.

2. Tablero Ébola en GIS

El caso de la enfermedad de Ébola es muy diferente a la enfermedad de COVID-19, por lo cual tuvo un alcance distinto. Se puede observar que la concentración de la enfermedad del Ébola se encuentra en África Occidental. Siendo la mayor Afectada Sierra Leona seguida de Liberia.

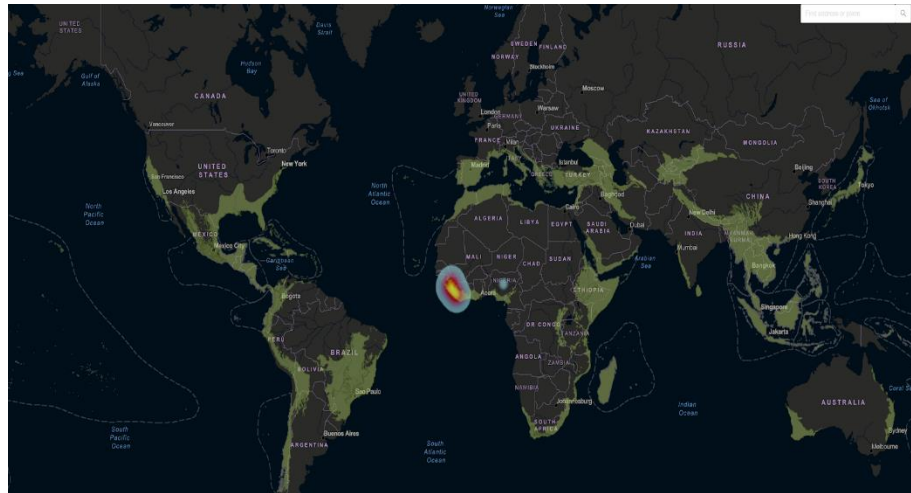


Figura 16: Mapa de Ébola con capa de mapa de calor por contagios confirmados.

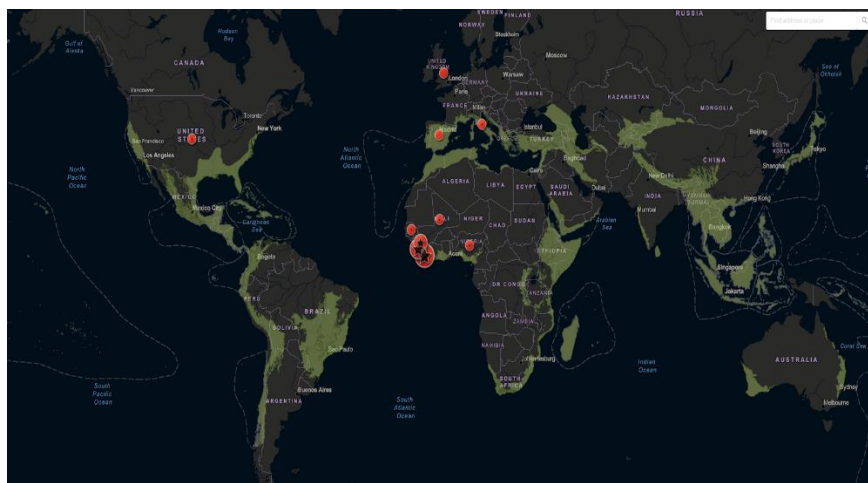


Figura 17: Mapa de Ébola con capa de puntos por contagios confirmados.

Es de notar, que ambos países se encuentran en áreas con una alta cantidad de vida salvaje, flora y fauna. Por otro lado, se puede observar que fue incluida la capa de clínicas y hospitales al igual que en COVID-19.

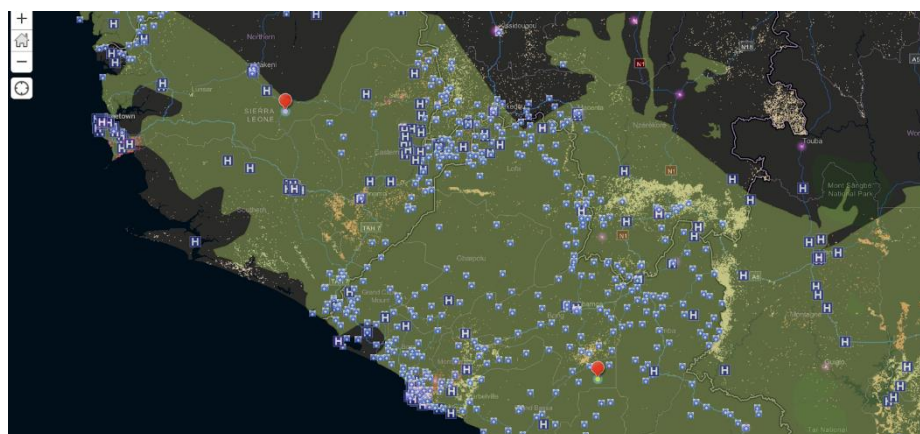


Figura 18: Mapa de Ébola con capa de hospitales y clínicas con la función de iniciar una ruta.

Dentro de los gráficos se incluyeron gráficos de curva de crecimiento para cada país sobre la cantidad de personas que fueron confirmadas enfermas de Ébola. Al mismo tiempo que

se puede obtener un resumen simple del estado de la enfermedad o la variable seleccionada al realizar un click o dejar el mouse encima de la gráfica. Es de notar que en caso contrario en donde Sierra Leona y Liberia fueron los países con una mayor cantidad de casos, los países que tuvieron una mayor tasa de mortalidad fueron Guinea y Sierra Leona.

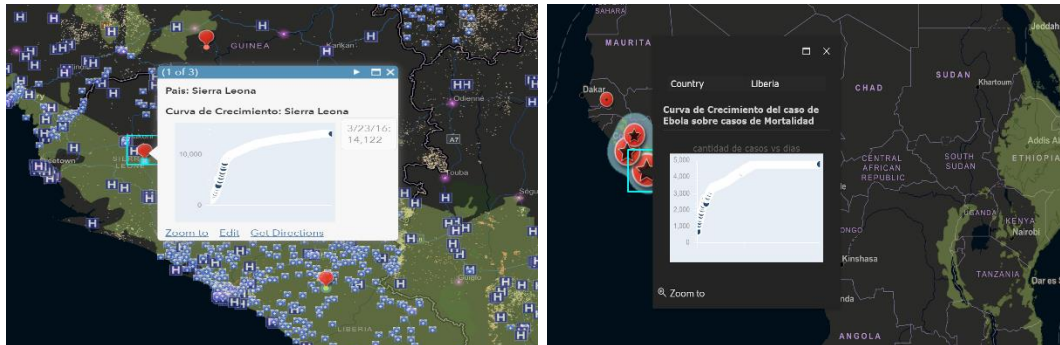


Figura 19: Gráficas por país para el caso de Ébola.

3. Tablero VIH SIDA en GIS

Para el tablero de VIH SIDA se puede observar que existe una mayor concentración en África siendo el país más afectado Sudáfrica. Se buscaron diferentes capas para las poblaciones clave definidas por los diferentes casos de contagio los cuales son por inyecciones, transmisión vertical, relaciones heterosexuales o relaciones homosexuales. Los puntos rojos mostraron la cantidad de personas que se encuentran recibiendo tratamiento de retrovirales según los diferentes países con el objetivo de observar las personas que ya se encuentran oficialmente confirmadas por VIH SIDA. Por otro lado, se manejó una capa de índice de progreso social global el cual es una medición del desempeño social de un país es decir su capacidad de satisfacer las necesidades humanas fundamentales de los humanos.

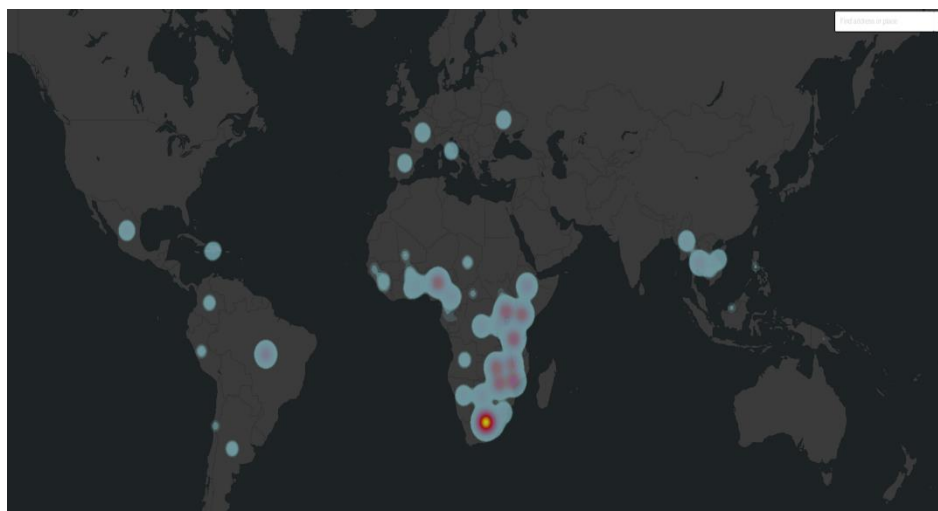


Figura 20: Mapa de VIH SIDA con capa de mapa de calor para personas que reciben tratamiento.

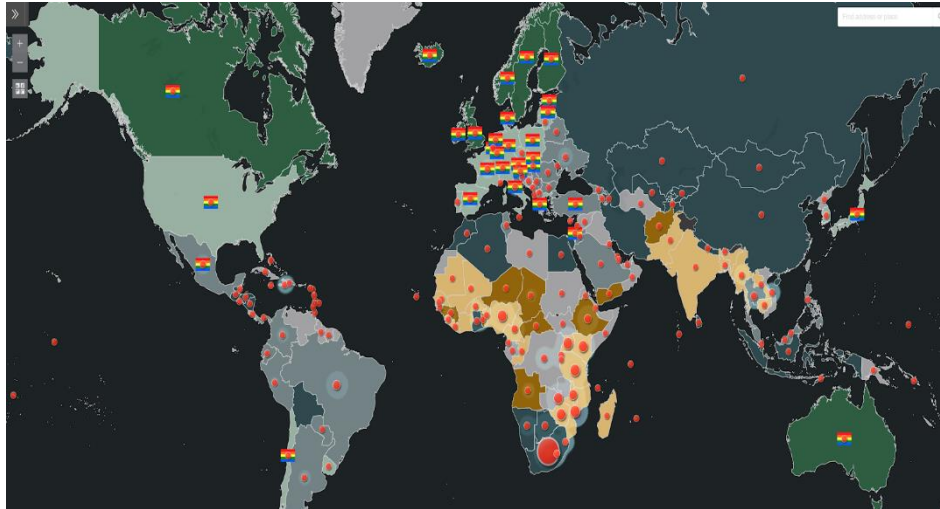


Figura 21: Mapa de VIH SIDA con capa de puntos por personas que reciben tratamiento.

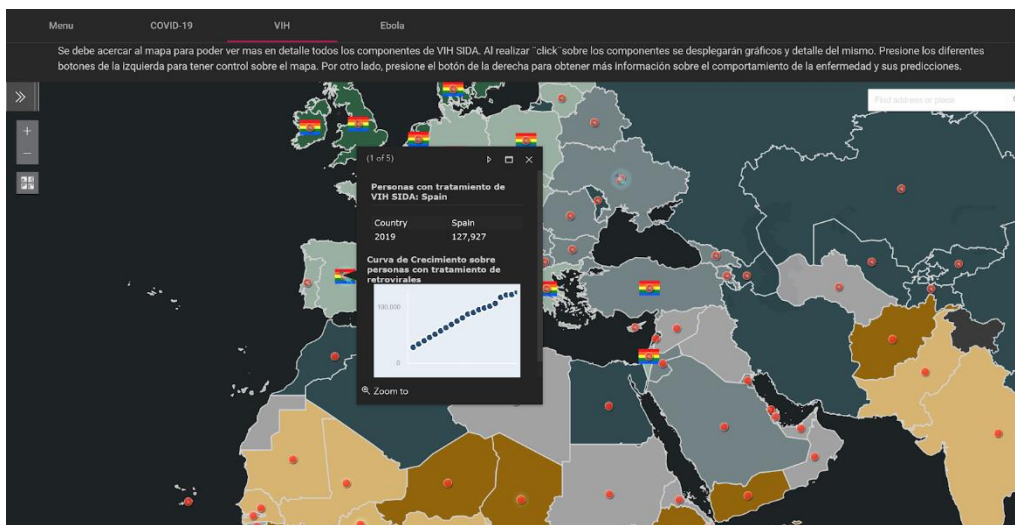


Figura 22: Gráficas por país para el caso de VIH SIDA.

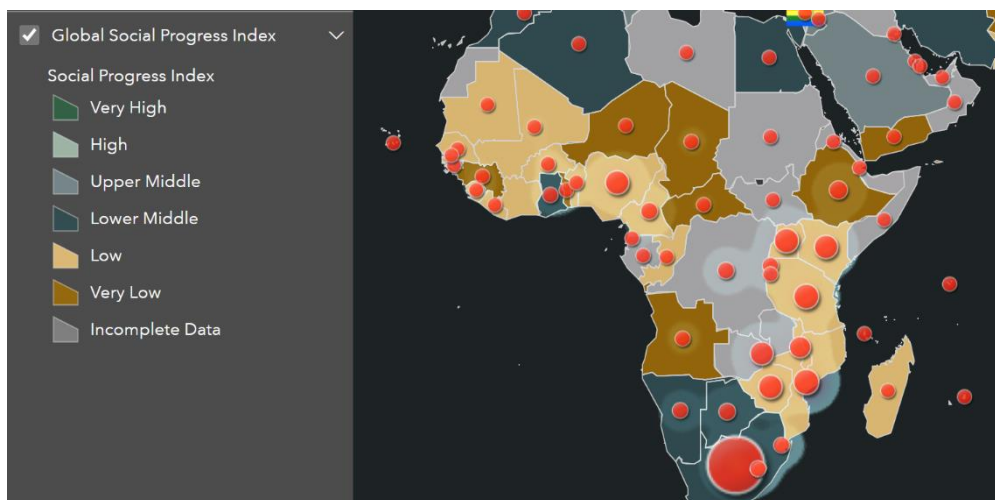


Figura 23: Mapa de VIH SIDA con capa de índice de progreso social global.

Se programó para que al acercarse a un país pueda observar los diferentes moteles o lugares donde las personas pueden mantener relaciones sexuales. Buscando crear una relación

en que mientras más moteles se tengan en el país más se propaga la enfermedad. Esto surgió a partir de la noticia de cbc en donde se busca eliminar la propagación de la enfermedad por medio de brindar condones gratis en los moteles y hoteles ("Free condoms in motel rooms may prevent AIDS: study | CBC News", 2021).

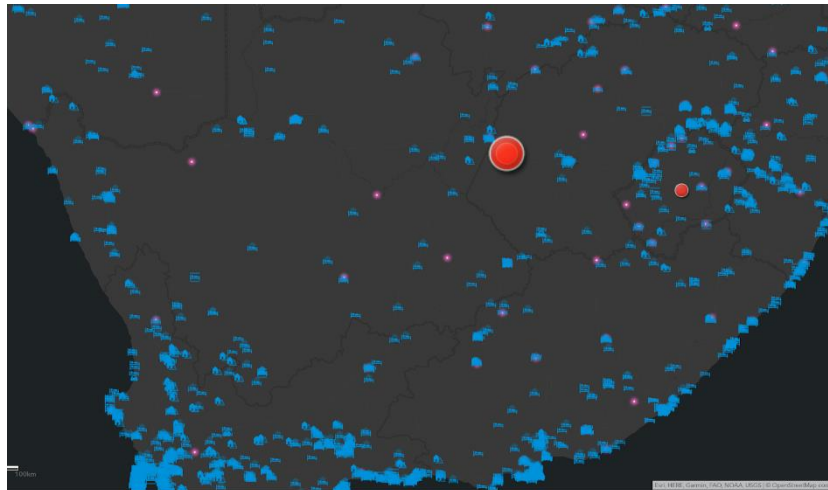


Figura 24: Mapa de VIH SIDA con capa de hoteles y moteles alrededor del mundo.

B. Análisis exploratorio

1. Resultados COVID-19

Tal y como se mencionó anteriormente, los datos del COVID se actualizan diariamente por lo que los resultados pueden ir variando, dependiendo del día en que se realiza la consulta.

Se puede observar que en la fecha 10/22/20 el país más afectado son los Estados Unidos, seguido de India y Brasil (gráfico de la izquierda). En el caso contrario el país menos afectado fueron las islas Solomon, MS Zaandam, Western Sahara y Saint Kitts y Nevis (gráfico de la derecha).

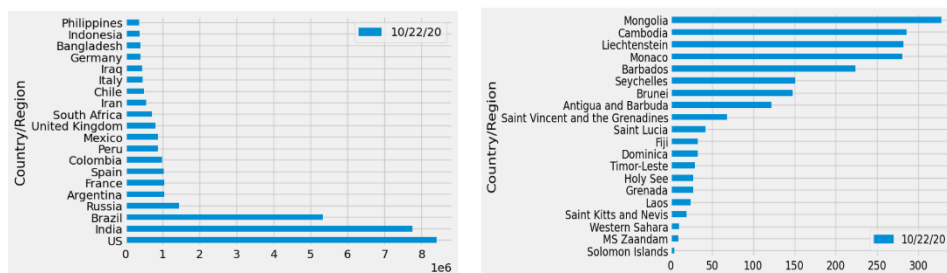


Figura 25: Gráficos comparativos por países del caso de COVID-19.

Entre las diferentes provincias registradas las provincias más afectadas fueron Quebec y Ontario el cual se encuentran en Canadá y Hubei en China. En el caso contrario las provincias menos afectadas en la fecha 10/22/20 fueron Tibet que se encuentra en Asia y Anguilla que se

encuentran en el caribe. Se adjunta el caso de Diamond Princess el cual fue un crucero que quedó varado en Tokio por la situación del COVID-19.

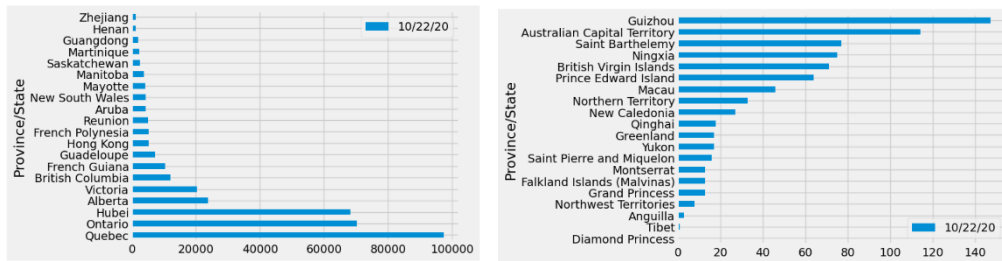


Figura 26: Gráficos comparativos por provincias del caso de COVID-19.

Se mostró un gráfico dentro del tablero que se va actualizando diariamente y muestra la proporción entre los casos recuperados, confirmados y los casos de mortalidad.

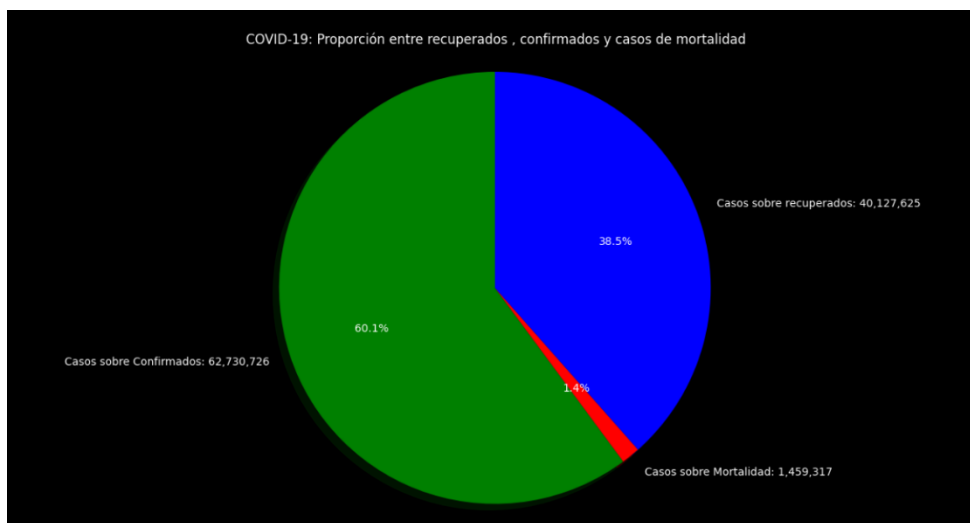


Figura 27: Gráficos circular sobre comparación entre casos confirmados, casos recuperados y casos de mortalidad.

Se puede observar el crecimiento en casos confirmados, casos de mortalidad y casos de recuperación respecto al día anterior.

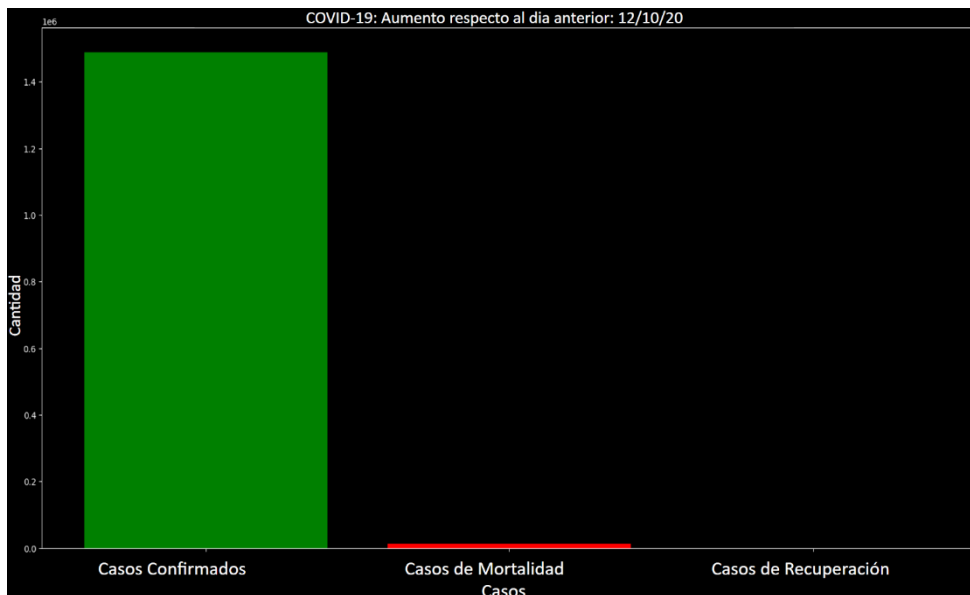


Figura 28: Gráficos de barras que muestra el aumento de cada caso respecto al día anterior entre casos confirmados, casos recuperados y casos de mortalidad.

2. Resultados Ébola

Se realizaron series de tiempo para los casos confirmados y de mortalidad en donde se puede ir observando el comportamiento de la enfermedad a lo largo de diferentes fechas.

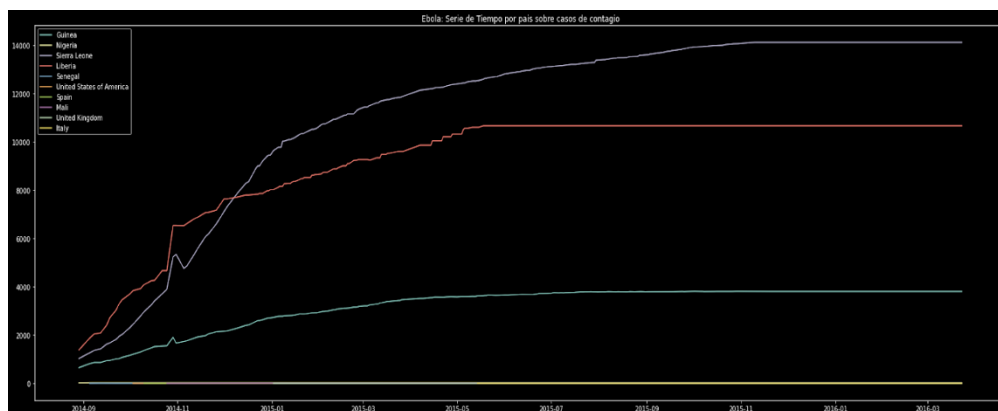


Figura 29: Serie de tiempo para casos de contagio del caso de Ébola.

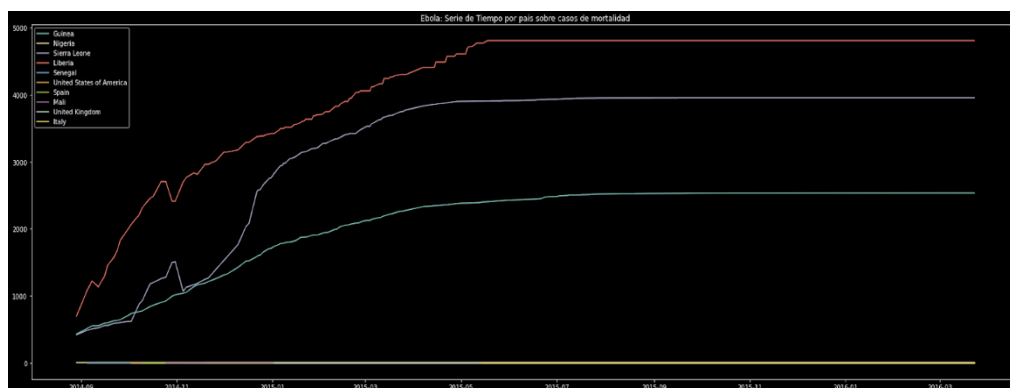


Figura 30: Serie de tiempo para casos de mortalidad del caso de Ébola.

Por otro lado, se agregaron dos diagramas de caja y bigotes para observar el comportamiento del ébola en base a las diferentes temporadas. En donde se puede observar que el año que generó mayor repercusión fue el 2015 y el mes más afectado fue el mes de Enero. Observando que en su mayoría en la temporada de invierno se obtuvo una mayor cantidad de contagios.

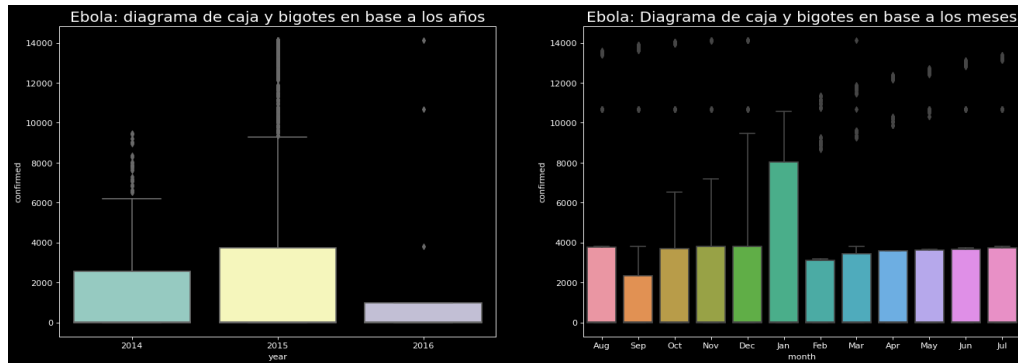


Figura 31: Gráfico de caja y bigotes para el caso de Ébola respecto a años y meses.

3. Resultados VIH SIDA

Rompiendo paradigmas (“Mayoría en contagios por relaciones homosexuales”) se encontró que la mayor cantidad de contagios en VIH SIDA se dan por uso de sustancias ilegales en aplicación por inyección. Seguida por casos de contagio en relaciones heterosexuales y luego por casos de contagio por relaciones homosexuales. El caso con menor proporción es por transmisión vertical.

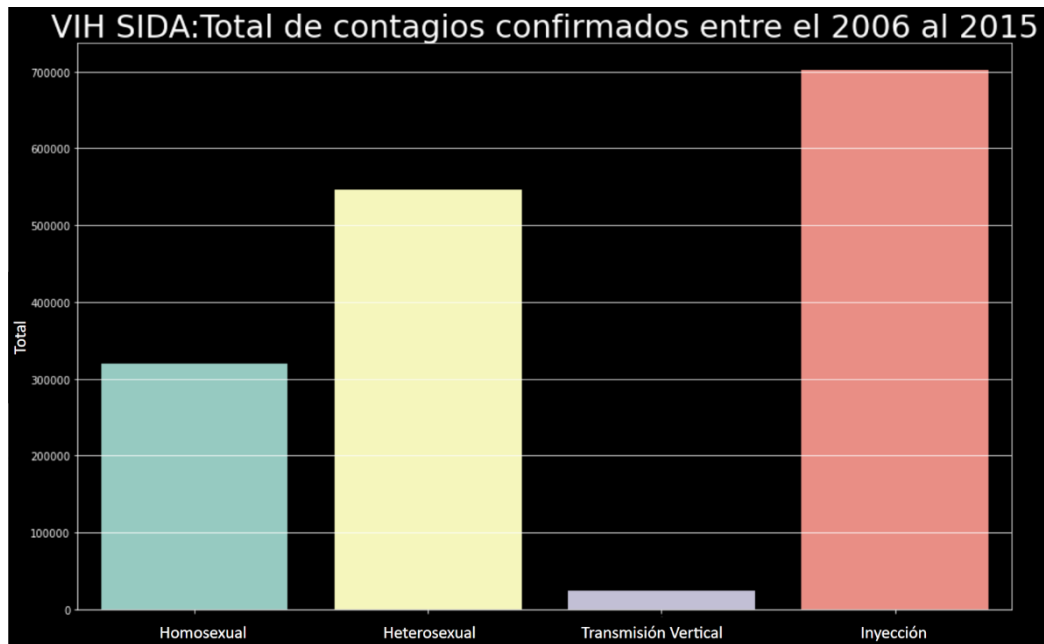


Figura 32: Gráfico comparativo sobre los diferentes tipos de contagios del caso de VIH SIDA siendo casos por contagios homosexuales, heterosexuales, transmisión vertical y por medio de inyecciones.

Adentrándonos en los casos, los casos de contagio por relaciones homosexuales y transmisión vertical se puede observar que hubo un descenso de contagio a lo largo de los años lo cual es opuesto a los casos de contagio por relaciones heterosexuales.

Por último, se realizó un diagrama "Pie Chart" o Gráfico Circular, en donde se pudo observar que la mayor parte de contagios son por parte del sexo masculino. En donde los casos indefinidos se refieren a datos que no recopilamos el sexo a la hora de transmitirlos.

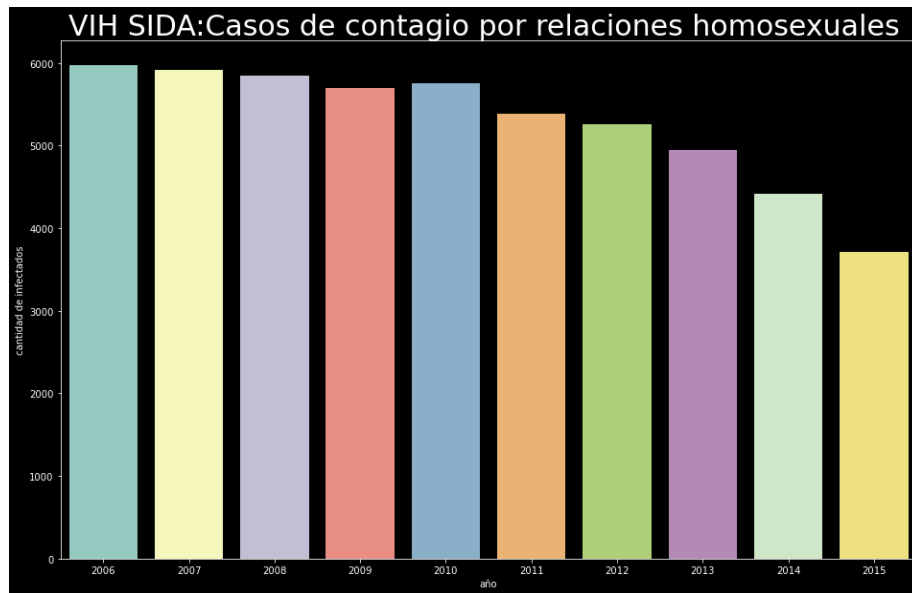


Figura 33: Gráfico de barras sobre el comportamiento a través del tiempo por contagios homosexuales.

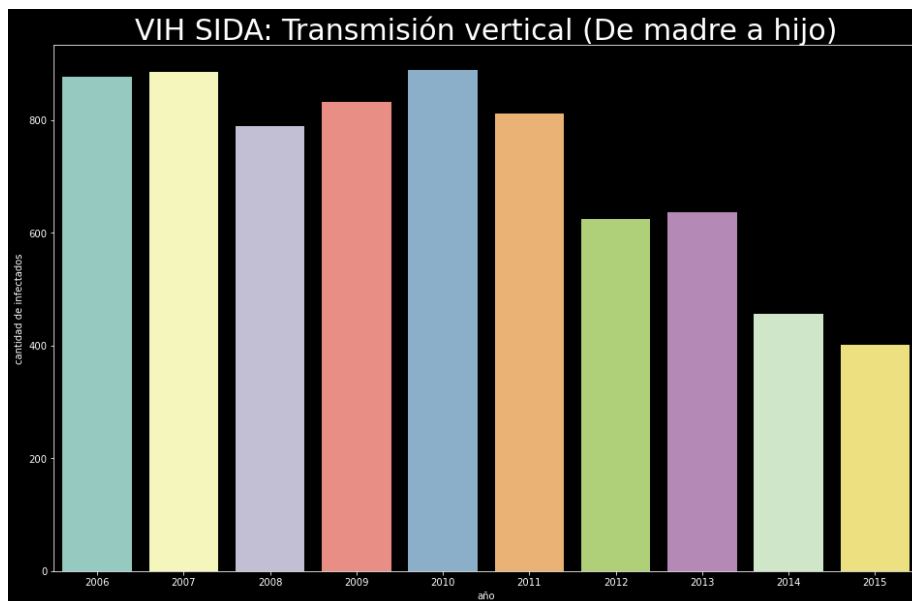


Figura 34: Gráfico sobre el comportamiento a través del tiempo por contagios de maternidad.

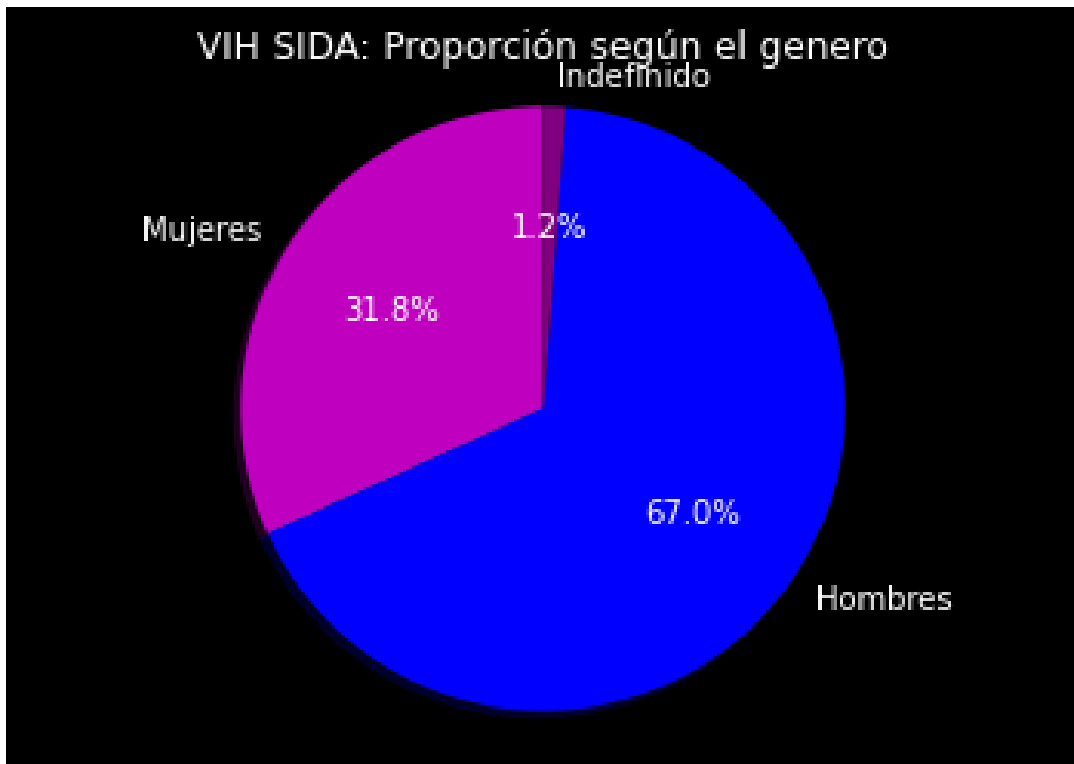


Figura 35: Gráfico circular que nos muestra la proporción de contagios de VIH SIDA según el género.

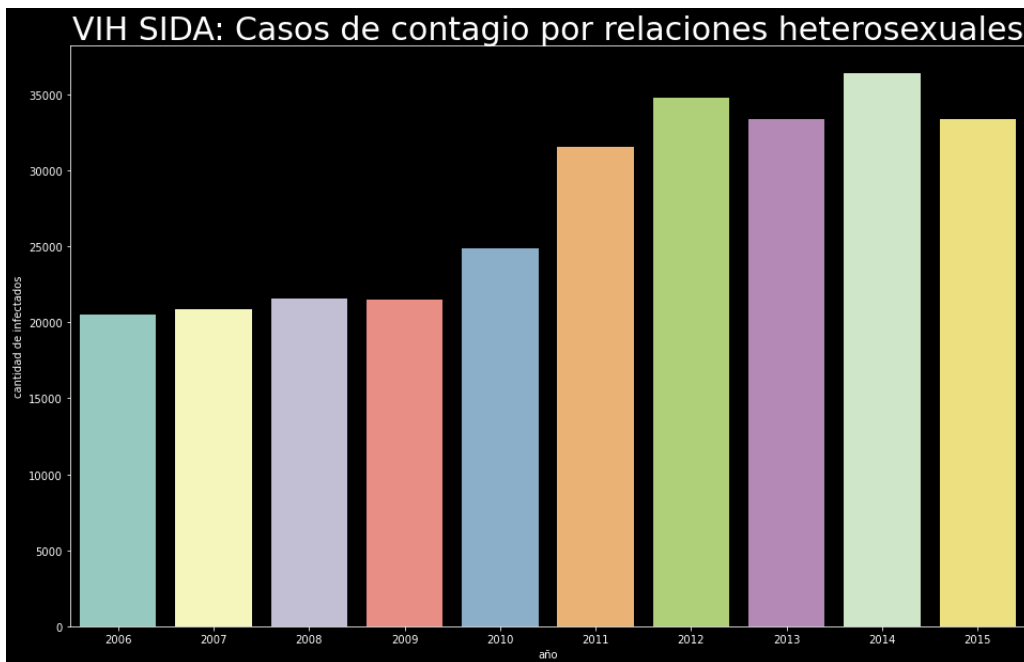


Figura 36: Gráfico de serie de tiempo sobre contagios por medio de una relación heterosexual.

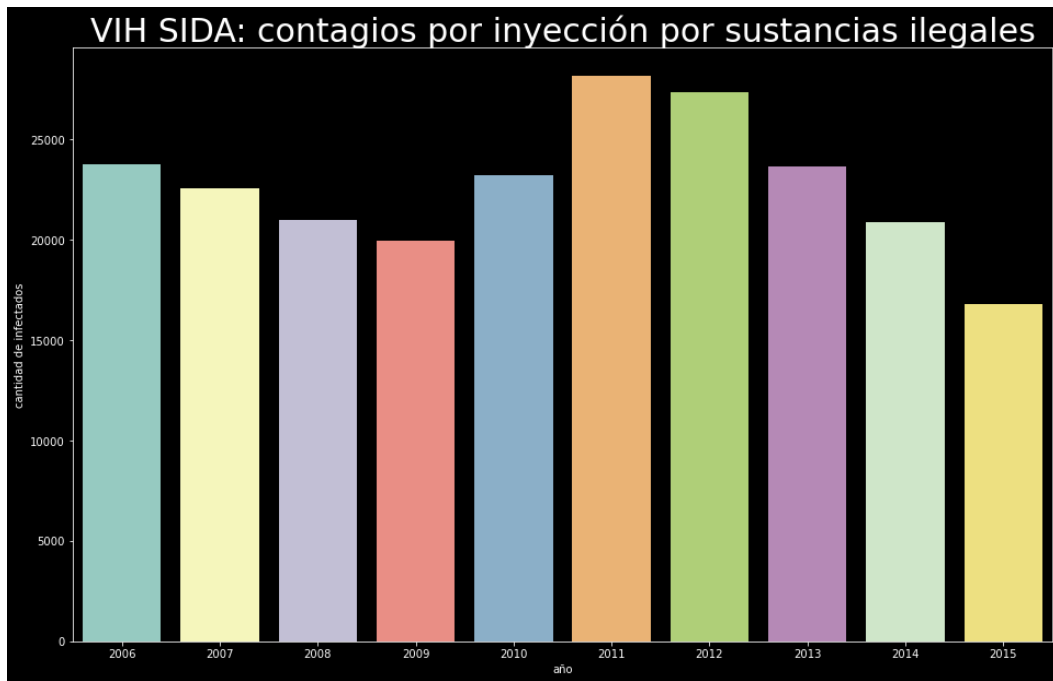


Figura 37: Gráfico sobre serie de tiempo por contagios por medio de inyección.

C. Cluster

Dentro de los resultados de clusters se utilizó un gráfico de codo para cada uno de los casos. En donde se logró apreciar un cambio brusco en la evolución de los datos, es decir su codo, se definió la cantidad de clusters que se deban utilizar. Siendo para ambos casos de COVID-19 el número determinado de 6 clusters. Mientras que para el caso de VIH Sida se determinó que se utilizarían 5 clusters.

Por otro lado, se muestran los resultados que muestran cómo fueron agrupados los países dentro de los clusters por medio de un query.

1. Cluster de COVID-19 para casos confirmados



Figura 38: Gráfico de codo para el caso de casos confirmados de COVID-19.


```
import json
df = df_classification.query('Label == 0')
primer_grupo_json = df.to_json(orient="split")
primer_grupo_json

'{"columns":["Country","Label"],"index":[2,3,7,9,10,12,13,16,19,20,23,28,29,37,39,45,47,51,54,58,59,62,64,70,76,79,80,82,84,87,89,90,92,94,98,101,103,105,106,107,108,110,114,115,123,128,132,133,136,138,142,144,151,159,161,164,166,173,177,181,182,183,185,189,190],"data":[[{"Country":"Algeria",0}, {"Country":"Andorra",0}, {"Country":"Armenia",0}, {"Country":"Austria",0}, {"Country":"Azerbaijan",0}, {"Country":"Bahrain",0}, {"Country":"Bangladesh",0}, {"Country":"Belgium",0}, {"Country":"Bhutan",0}, {"Country":"Bolivia",0}, {"Country":"Brazil",0}, {"Country":"Burundi",0}, {"Country":"Cabo Verde",0}, {"Country":"Colombia",0}, {"Country":"Congo (Brazzaville)",0}, {"Country":"Cyprus",0}, {"Country":"Denmark",0}, {"Country":"Dominican Republic",0}, {"Country":"El Salvador",0}, {"Country":"Eswatini",0}, {"Country":"Ethiopia",0}, {"Country":"France",0}, {"Country":"Gambia",0}, {"Country":"Guatemala",0}, {"Country":"Honduras",0}, {"Country":"India",0}, {"Country":"Indonesia",0}, {"Country":"Iraq",0}, {"Country":"Israel",0}, {"Country":"Japan",0}, {"Country":"Kazakhstan",0}, {"Country":"Kenya",0}, {"Country":"Kosovo",0}, {"Country":"Kyrgyzstan",0}, {"Country":"Lesotho",0}, {"Country":"Liechtenstein",0}, {"Country":"Luxembourg",0}, {"Country":"Madagascar",0}, {"Country":"Malawi",0}, {"Country":"Malaysia",0}, {"Country":"Maldives",0}, {"Country":"Malta",0}, {"Country":"Mexico",0}, {"Country":"Moldova",0}, {"Country":"Netherlands",0}, {"Country":"North Macedonia",0}, {"Country":"Panama",0}, {"Country":"Papua New Guinea",0}, {"Country":"Philippines",0}, {"Country":"Portugal",0}, {"Country":"Rwanda",0}, {"Country":"Saint Lucia",0}, {"Country":"Serbia",0}, {"Country":"South Africa",0}, {"Country":"Spain",0}, {"Country":"Suriname",0}, {"Country":"Switzerland",0}, {"Country":"Togo",0}, {"Country":"US",0}, {"Country":"United Kingdom",0}, {"Country":"Uruguay",0}, {"Country":"Uzbekistan",0}, {"Country":"Venezuela",0}, {"Country":"Zambia",0}, {"Country":"Zimbabwe",0}]}]'
```

```
df = df_classification.query('Label == 1')
segundo_grupo_json = df.to_json(orient="split")
segundo_grupo_json

'{"columns":["Country","Label"],"index":[8,24,30,36,48,60,69,75,91,95,104,113,124,126,143,147,168,170,171,172],"data":[[{"Country":"Austria",1}, {"Country":"Brunei",1}, {"Country":"Cambodia",1}, {"Country":"China",1}, {"Country":"Diamond Princess",1}, {"Country":"Fiji",1}, {"Country":"Grenada",1}, {"Country":"Holy See",1}, {"Country":"Korea, South",1}, {"Country":"Laos",1}, {"Country":"MS Zaandam",1}, {"Country":"Mauritius",1}, {"Country":"New Zealand",1}, {"Country":"Niger",1}, {"Country":"Saint Kitts and Nevis",1}, {"Country":"San Marino",1}, {"Country":"Taiwan*",1}, {"Country":"Tanzania",1}, {"Country":"Thailand",1}, {"Country":"Timor-Leste",1}]]}'
```

```
df = df_classification.query('Label == 2')
tercer_grupo_json = df.to_json(orient="split")
tercer_grupo_json

'{"columns":["Country","Label"],"index":[146],"data":[[{"Country":"Samoa",2}]]}'
```

```
df = df_classification.query('Label == 3')
cuarto_grupo_json = df.to_json(orient="split")
cuarto_grupo_json

'{"columns":["Country","Label"],"index":[1,4,6,11,17,21,22,25,27,41,43,46,65,68,73,77,86,88,96,97,100,102,118,119,120,121,122,134,137,140,155,156,162,167,174,175,178,179,187],"data":[[{"Country":"Albania",3}, {"Country":"Angola",3}, {"Country":"Argentina",3}, {"Country":"Bahamas",3}, {"Country":"Belize",3}, {"Country":"Bosnia and Herzegovina",3}, {"Country":"Botswana",3}, {"Country":"Bulgaria",3}, {"Country":"Burma",3}, {"Country":"Costa Rica",3}, {"Country":"Croatia",3}, {"Country":"Czechia",3}, {"Country":"Georgia",3}, {"Country":"Greece",3}, {"Country":"Guyana",3}, {"Country":"Hungary",3}, {"Country":"Jamaica",3}, {"Country":"Jordan",3}, {"Country":"Latvia",3}, {"Country":"Lebanon",3}, {"Country":"Libya",3}, {"Country":"Lithuania",3}, {"Country":"Montenegro",3}, {"Country":"Morocco",3}, {"Country":"Mozambique",3}, {"Country":"Namibia",3}, {"Country":"Nepal",3}, {"Country":"Paraguay",3}, {"Country":"Poland",3}, {"Country":"Romania",3}, {"Country":"Slovakia",3}, {"Country":"Slovenia",3}, {"Country":"Sri Lanka",3}, {"Country":"Syria",3}, {"Country":"Trinidad and Tobago",3}, {"Country":"Tunisia",3}, {"Country":"Uganda",3}, {"Country":"Ukraine",3}, {"Country":"West Bank and Gaza",3}]]}'
```

```
df = df_classification.query('Label == 4')
quinto_grupo_json = df.to_json(orient="split")
quinto_grupo_json

'{"columns":["Country","Label"],"index":[0,5,14,15,18,26,31,32,33,34,35,38,40,42,44,49,50,52,53,55,56,57,61,63,66,67,71,72,74,78,81,83,85,93,99,109,112,116,117,125,127,129,130,131,135,139,141,145,148,149,150,152,153,154,157,158,160,163,165,169,176,180,186,188],"data":[[{"Country":"Afghanistan",4}, {"Country":"Antigua and Barbuda",4}, {"Country":"Barbados",4}, {"Country":"Belarus",4}, {"Country":"Benin",4}, {"Country":"Burkina Faso",4}, {"Country":"Cameroon",4}, {"Country":"Canada",4}, {"Country":"Central African Republic",4}, {"Country":"Chad",4}, {"Country":"Chile",4}, {"Country":"Comoros",4}, {"Country":"Congo (Kinshasa)",4}, {"Country":"Cote d'Ivoire",4}, {"Country":"Cuba",4}, {"Country":"Djibouti",4}, {"Country":"Dominica",4}, {"Country":"Ecuador",4}, {"Country":"Egypt",4}, {"Country":"Equatorial Guinea",4}, {"Country":"Eritrea",4}, {"Country":"Estonia",4}, {"Country":"Finland",4}, {"Country":"Gabon",4}, {"Country":"Germany",4}, {"Country":"Ghana",4}, {"Country":"Guinea",4}, {"Country":"Guinea-Bissau",4}, {"Country":"Haiti",4}, {"Country":"Iceland",4}, {"Country":"Iran",4}, {"Country":"Ireland",4}, {"Country":"Italy",4}, {"Country":"Kuwait",4}, {"Country":"Liberia",4}, {"Country":"Mali",4}, {"Country":"Mauritania",4}, {"Country":"Monaco",4}, {"Country":"Mongolia",4}, {"Country":"Nicaragua",4}, {"Country":"Nigeria",4}, {"Country":"Norway",4}, {"Country":"Oman",4}, {"Country":"Pakistan",4}, {"Country":"Peru",4}, {"Country":"Qatar",4}, {"Country":"Russia",4}, {"Country":"Saint Vincent and the Grenadines",4}, {"Country":"Sao Tome and Principe",4}, {"Country":"Saudi Arabia",4}, {"Country":"Senegal",4}, {"Country":"Seychelles",4}, {"Country":"Sierra Leone",4}, {"Country":"Singapore",4}, {"Country":"Solomon Islands",4}, {"Country":"Somalia",4}, {"Country":"South Sudan",4}, {"Country":"Sudan",4}, {"Country":"Sweden",4}, {"Country":"Tajikistan",4}, {"Country":"Turkey",4}, {"Country":"United Arab Emirates",4}, {"Country":"Vietnam",4}, {"Country":"Yemen",4}]]}'
```

```
df = df_classification.query('Label == 5')
sexto_grupo_json = df.to_json(orient="split")
sexto_grupo_json

'{"columns":["Country","Label"],"index":[111,184],"data":[[{"Country":"Marshall Islands",5}, {"Country":"Vanuatu",5}]]}'
```

Figura 40: Query que muestra la agrupación de países de casos confirmados por COVID-19.

2. Cluster de COVID-19 para casos de mortalidad

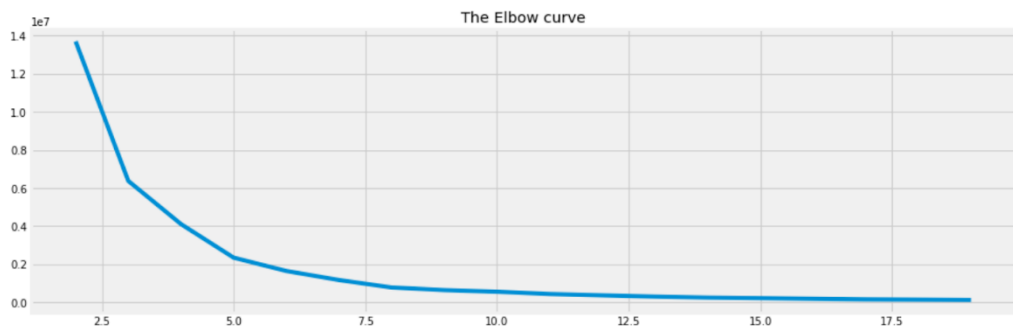


Figura 41: Gráfico de codo para el caso de casos de mortalidad sobre el caso de COVID-19.

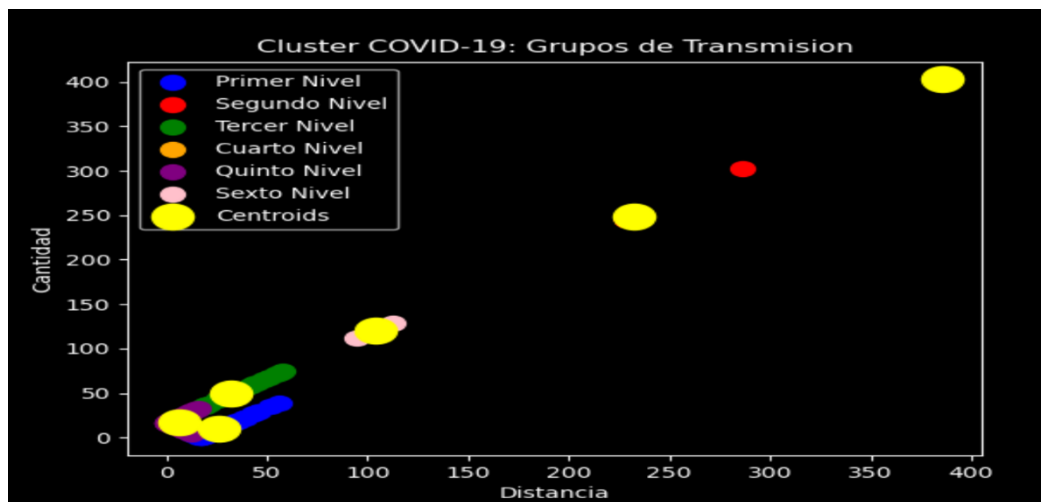


Figura 42: Cluster para el caso de casos de mortalidad sobre el caso de COVID-19.

```

import json
df = df_classification.query('Label == 0')
primer_grupo_json = df.to_json(orient="split")
primer_grupo_json

'{"columns":["Country","Label"],"index":[2,5,31,34,38,40,42,44,49,63,68,77,81,87,96,99,102,109,131,138,147,150,152,155,157,162,168,175,176,179,181],"data":[{"Algeria",0}, {"Antigua and Barbuda",0}, {"Cameroon",0}, {"Chad",0}, {"Comoros",0}, {"Congo (Kinshasa)",0}, {"Cote d'Ivoire",0}, {"Cuba",0}, {"Djibouti",0}, {"Gabon",0}, {"Greece",0}, {"Hungary",0}, {"Iran",0}, {"Japan",0}, {"Latvia",0}, {"Liberia",0}, {"Lithuania",0}, {"Mali",0}, {"Pakistan",0}, {"Portugal",0}, {"Sao Tome and Principe",0}, {"Serbia",0}, {"Sierra Leone",0}, {"Slovenia",0}, {"Somalia",0}, {"Sudan",0}, {"Tajikistan",0}, {"Turkey",0}, {"US",0}, {"United Arab Emirates",0}, {"Uruguay",0}]}'

df = df_classification.query('Label == 1')
segundo_grupo_json = df.to_json(orient="split")
segundo_grupo_json

'{"columns":["Country","Label"],"index":[19,30,32,50,56,69,75,95,111,117,143,144,145,151,156,171],"data":[{"Bhutan",1}, {"Cambodia",1}, {"Canada",1}, {"Dominica",1}, {"Eritrea",1}, {"Grenada",1}, {"Holy See",1}, {"Laos",1}, {"Marshall Islands",1}, {"Mongolia",1}, {"Saint Kitts and Nevis",1}, {"Saint Lucia",1}, {"Saint Vincent and the Grenadines",1}, {"Seychelles",1}, {"Solomon Islands",1}, {"Timor-Leste",1}]}'

df = df_classification.query('Label == 2')
tercer_grupo_json = df.to_json(orient="split")
tercer_grupo_json

'{"columns":["Country","Label"],"index":[4,6,11,17,20,22,27,29,37,41,58,59,64,65,73,79,82,86,88,89,92,97,98,100,105,118,119,120,121,122,130,134,142,158,163,166,173,174,177,182,183,184,185,188,189],"data":[{"Angola",2}, {"Argentina",2}, {"Bahamas",2}, {"Belize",2}, {"Bolivia",2}, {"Botswana",2}, {"Burma",2}, {"Cabo Verde",2}, {"Colombia",2}, {"Costa Rica",2}, {"Eswatini",2}, {"Ethiopia",2}, {"Gambia",2}, {"Georgia",2}, {"Guyana",2}, {"India",2}, {"Iraq",2}, {"Jamaica",2}, {"Jordan",2}, {"Kazakhstan",2}, {"Kosovo",2}, {"Lebanon",2}, {"Lesotho",2}, {"Libya",2}, {"Madagascar",2}, {"Montenegro",2}, {"Morocco",2}, {"Mozambique",2}, {"Namibia",2}, {"Nepal",2}, {"Oman",2}, {"Paraguay",2}, {"Rwanda",2}, {"South Africa",2}, {"Suriname",2}, {"Syria",2}, {"Trinidad and Tobago",2}, {"Tunisia",2}, {"Uganda",2}, {"Uzbekistan",2}, {"Venezuela",2}, {"Vietnam",2}, {"West Bank and Gaza",2}, {"Zambia",2}, {"Zimbabwe",2}]}'

df = df_classification.query('Label == 3')
cuarto_grupo_json = df.to_json(orient="split")
cuarto_grupo_json

'{"columns":["Country","Label"],"index":[1,4,6,11,17,21,22,25,27,41,43,46,65,68,73,77,86,88,96,97,100,102,118,119,120,121,122,134,137,140,155,156,162,167,174,175,178,179,187],"data":[{"Albania",3}, {"Angola",3}, {"Argentina",3}, {"Bahamas",3}, {"Belize",3}, {"Bosnia and Herzegovina",3}, {"Botswana",3}, {"Bulgaria",3}, {"Burma",3}, {"Costa Rica",3}, {"Croatia",3}, {"Czechia",3}, {"Georgia",3}, {"Greece",3}, {"Guyana",3}, {"Hungary",3}, {"Jamaica",3}, {"Jordan",3}, {"Latvia",3}, {"Lebanon",3}, {"Libya",3}, {"Lithuania",3}, {"Montenegro",3}, {"Morocco",3}, {"Mozambique",3}, {"Namibia",3}, {"Nepal",3}, {"Paraguay",3}, {"Poland",3}, {"Romania",3}, {"Slovakia",3}, {"Slovenia",3}, {"Sri Lanka",3}, {"Syria",3}, {"Trinidad and Tobago",3}, {"Tunisia",3}, {"Uganda",3}, {"Ukraine",3}, {"West Bank and Gaza",3}]}'

df = df_classification.query('Label == 4')
quinto_grupo_json = df.to_json(orient="split")
quinto_grupo_json

'{"columns":["Country","Label"],"index":[0,5,14,15,18,26,31,32,33,34,35,38,40,42,44,49,50,52,53,55,56,57,61,63,66,67,71,72,74,78,81,83,85,93,99,109,112,116,117,125,127,129,130,131,135,139,141,145,148,149,150,152,153,154,157,158,160,163,165,169,176,180,186,188],"data":[{"Afghanistan",4}, {"Antigua and Barbuda",4}, {"Barbados",4}, {"Belarus",4}, {"Benin",4}, {"Burkina Faso",4}, {"Cameroon",4}, {"Canada",4}, {"Central African Republic",4}, {"Chad",4}, {"Chile",4}, {"Comoros",4}, {"Congo (Kinshasa)",4}, {"Cote d'Ivoire",4}, {"Cuba",4}, {"Djibouti",4}, {"Dominica",4}, {"Ecuador",4}, {"Egypt",4}, {"Equatorial Guinea",4}, {"Eritrea",4}, {"Estonia",4}, {"Finland",4}, {"Gabon",4}, {"Germany",4}, {"Ghana",4}, {"Guinea",4}, {"Guinea-Bissau",4}, {"Haiti",4}, {"Iceland",4}, {"Iran",4}, {"Ireland",4}, {"Italy",4}, {"Kuwait",4}, {"Liberia",4}, {"Mali",4}, {"Mauritania",4}, {"Monaco",4}, {"Mongolia",4}, {"Nicaragua",4}, {"Nigeria",4}, {"Norway",4}, {"Oman",4}, {"Pakistan",4}, {"Peru",4}, {"Qatar",4}, {"Russia",4}, {"Saint Vincent and the Grenadines",4}, {"Sao Tome and Principe",4}, {"Saudi Arabia",4}, {"Senegal",4}, {"Seychelles",4}, {"Sierra Leone",4}, {"Singapore",4}, {"Solomon Islands",4}, {"Somalia",4}, {"South Sudan",4}, {"Sudan",4}, {"Sweden",4}, {"Tajikistan",4}, {"Turkey",4}, {"United Arab Emirates",4}, {"Vietnam",4}, {"Yemen",4}]}'

df = df_classification.query('Label == 5')
sexto_grupo_json = df.to_json(orient="split")
sexto_grupo_json

'{"columns":["Country","Label"],"index":[111,184],"data":[{"Marshall Islands",5}, {"Vanuatu",5}]}'

```

Figura 43: Query que muestra la agrupación de de casos de mortalidad de COVID-19.

3. Cluster de VIH SIDA según las personas que se encuentran en tratamiento de VIH SIDA

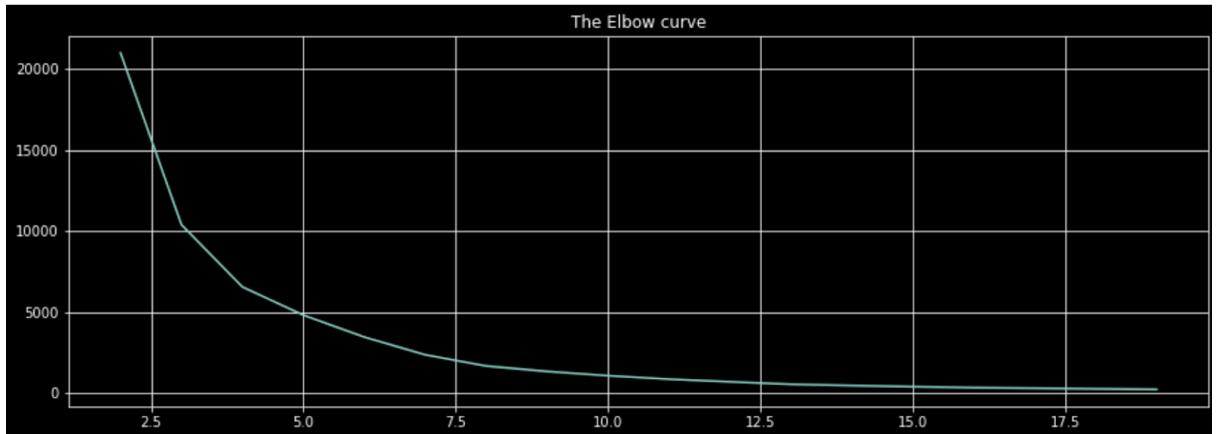


Figura 44: Gráfico de codo para el caso de VIH SIDA.

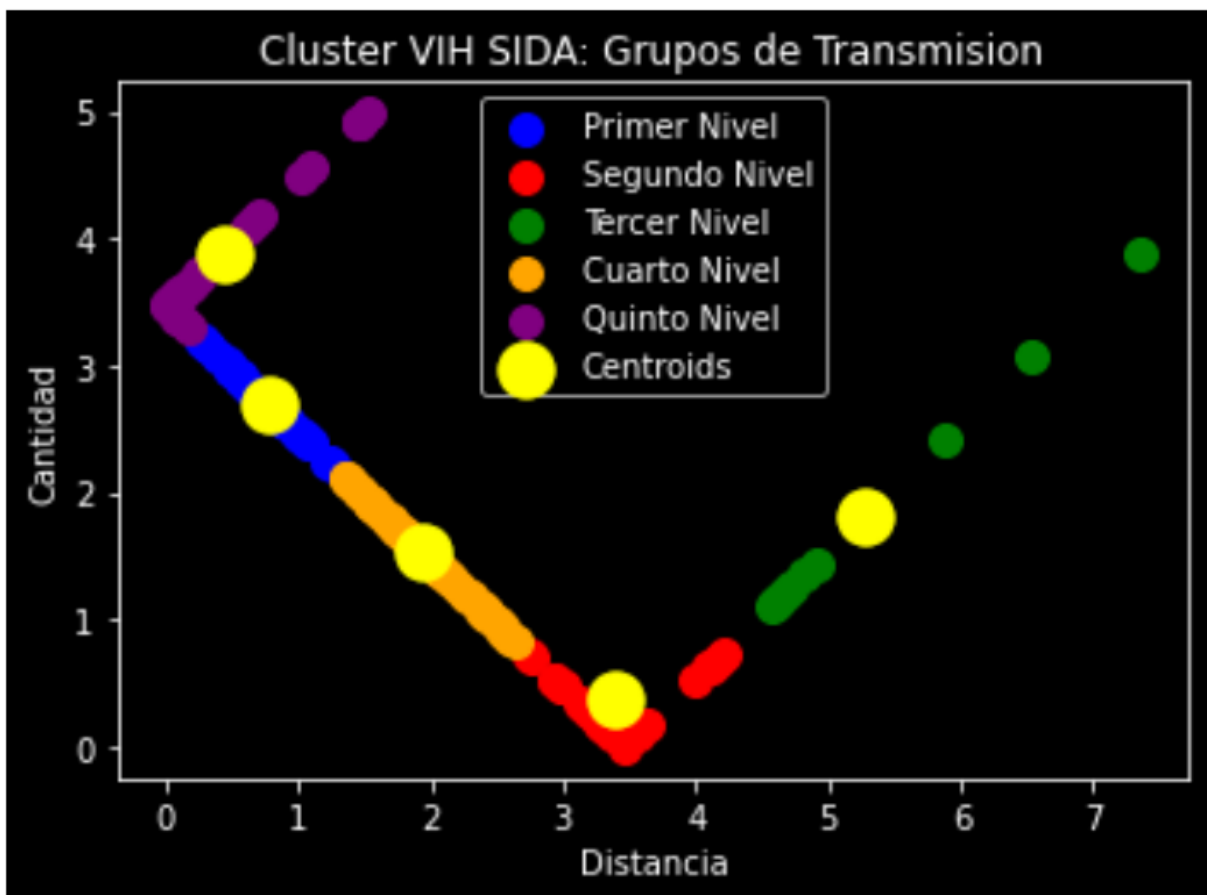


Figura 45: Cluster para el caso de VIH SIDA

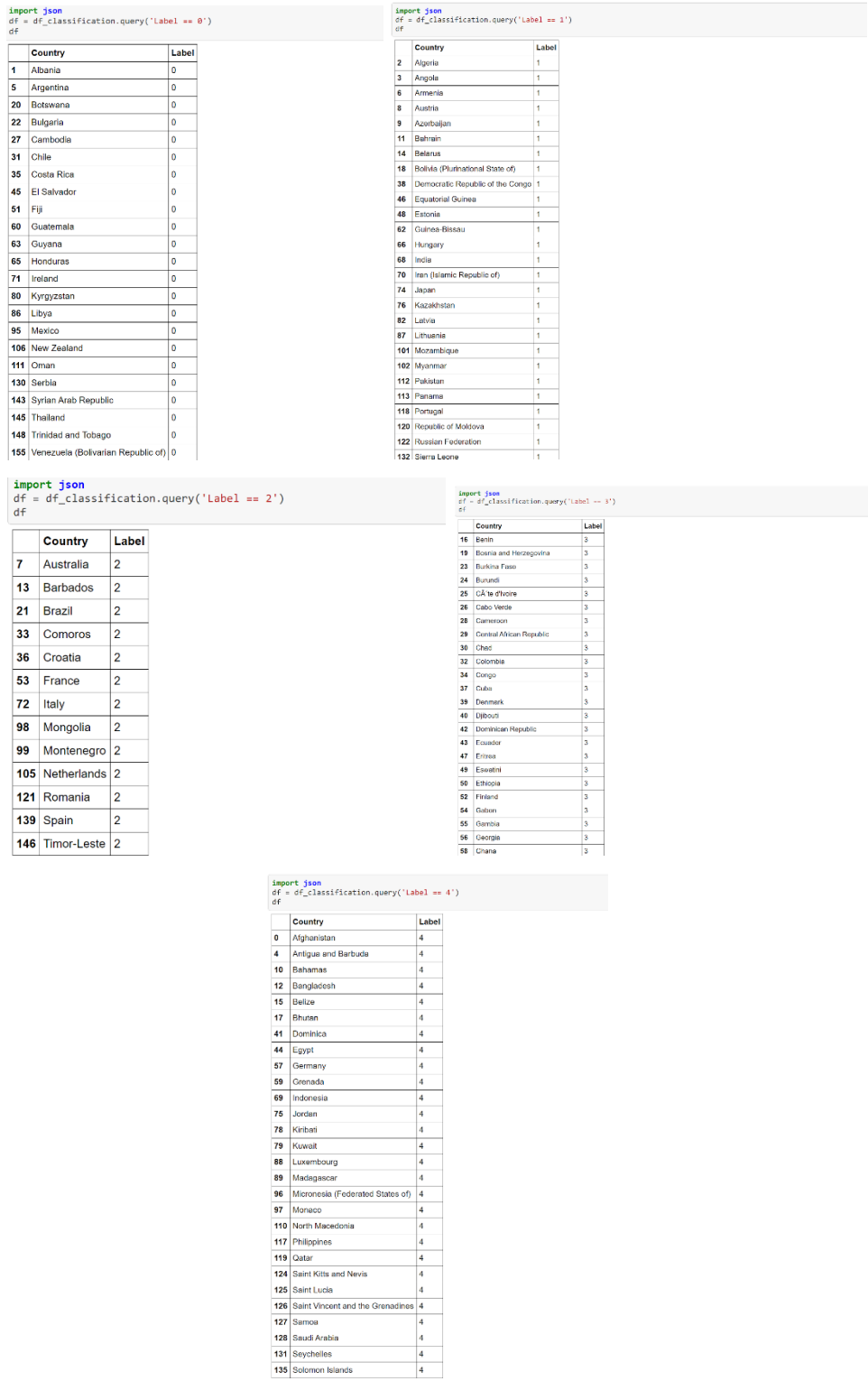


Figura 46: Query que muestra la agrupación de países de VIH SIDA.

D. Predicciones

1. Predicciones COVID-19

Se realizaron predicciones con redes neuronales para el caso de COVID-19 tomando en cuenta una predicción general en donde se toma en cuenta todos los países para los casos confirmados, casos de mortalidad y casos de recuperación. Al mismo tiempo, que se incluye la opción de realizar una predicción por país. La línea celeste nos muestra cómo se encuentran los datos actualmente mientras la línea punteada de color amarillo nos muestra la predicción. Debido a que estas predicciones se ejecutan diariamente la precisión suele variar. Sin embargo, las predicciones suelen oscilar entre 60 a 85%. No se toma en cuenta predicciones superiores a 85% debido a que se puede dar el caso de overfitting. Las predicciones se generan cada día en base a la nueva data obtenida y luego guarda la mejor predicción.

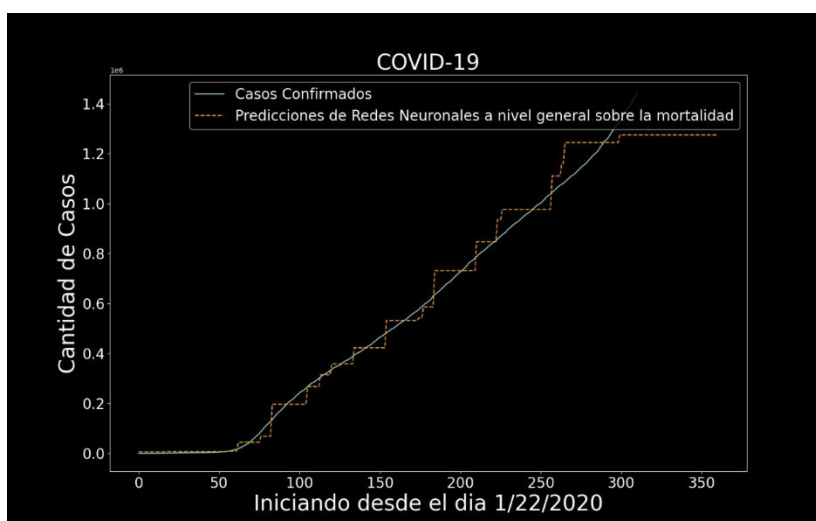


Figura 47: Predicciones con Redes neuronales a nivel general para casos de mortalidad.

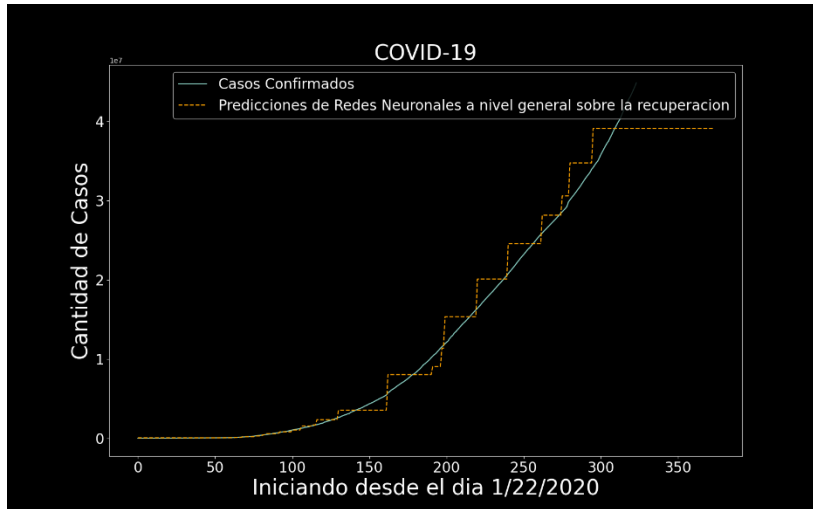


Figura 48: Predicciones con Redes Neuronales a nivel general para casos de recuperación.

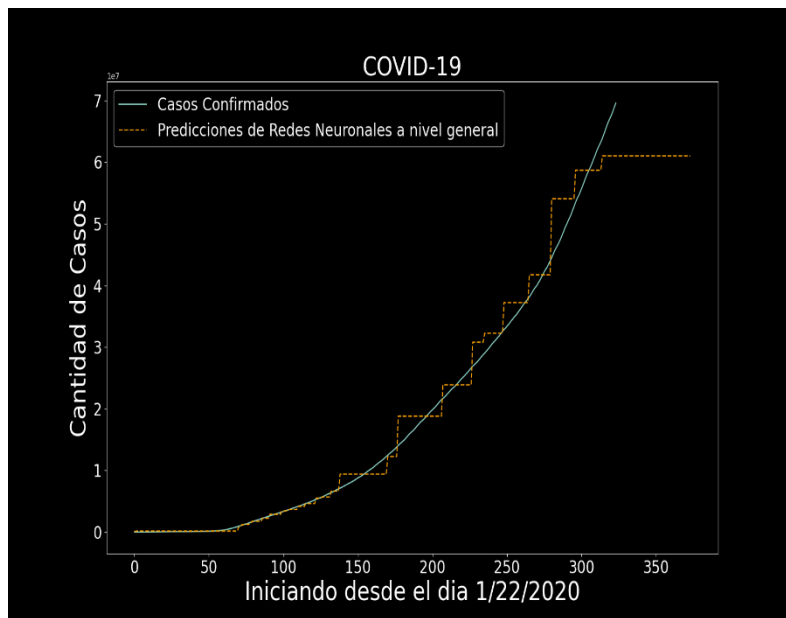


Figura 49: Predicciones con Redes Neuronales a nivel general para casos confirmados.

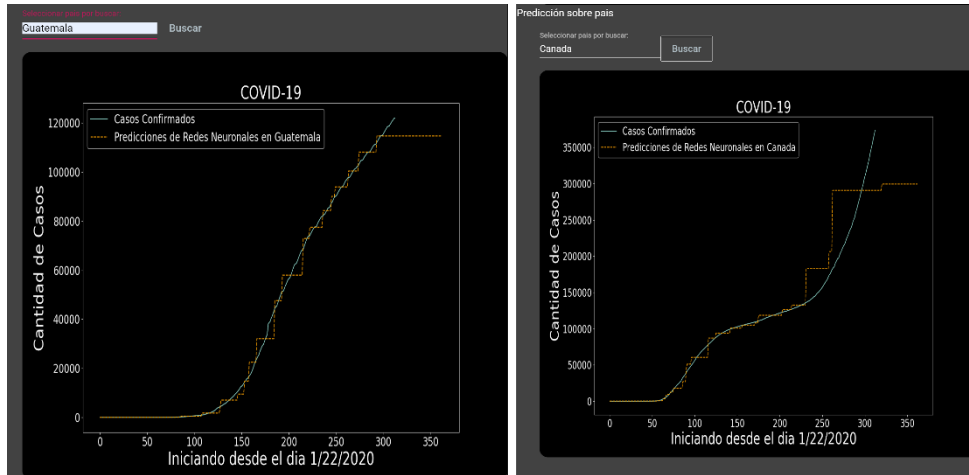


Figura 50: Predicciones con Redes Neuronales sobre casos confirmados por país dentro de los cuales se compara Guatemala vs Canadá.

2. Predicciones VIH SIDA

Se realizaron predicciones con Lasso en el cual se obtuvo un MAE de 0.025.

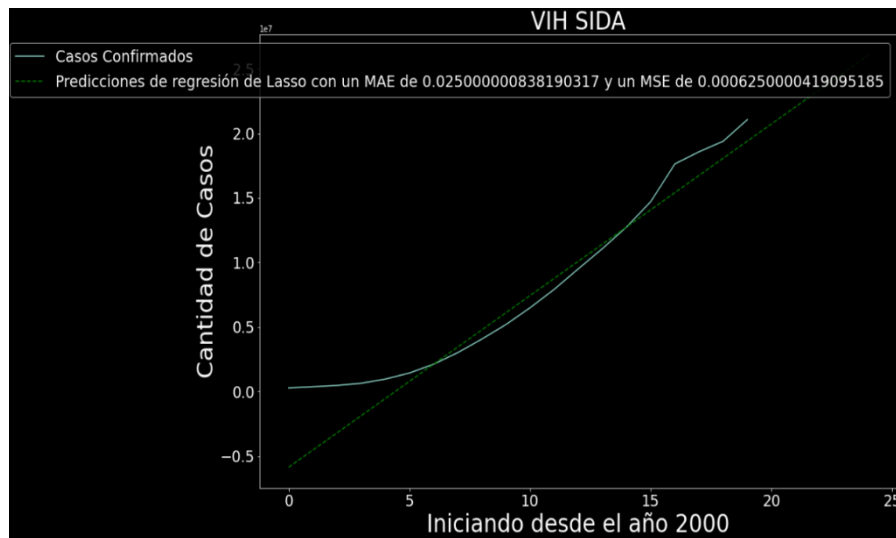


Figura 51: Predicciones con Lasso sobre el caso de VIH SIDA.

Se realizaron predicciones de series de tiempo con ARIMA con un error medio cuadrático (MSE) de 157828897625579.84 el cual se encuentra bastante alejado de cero.

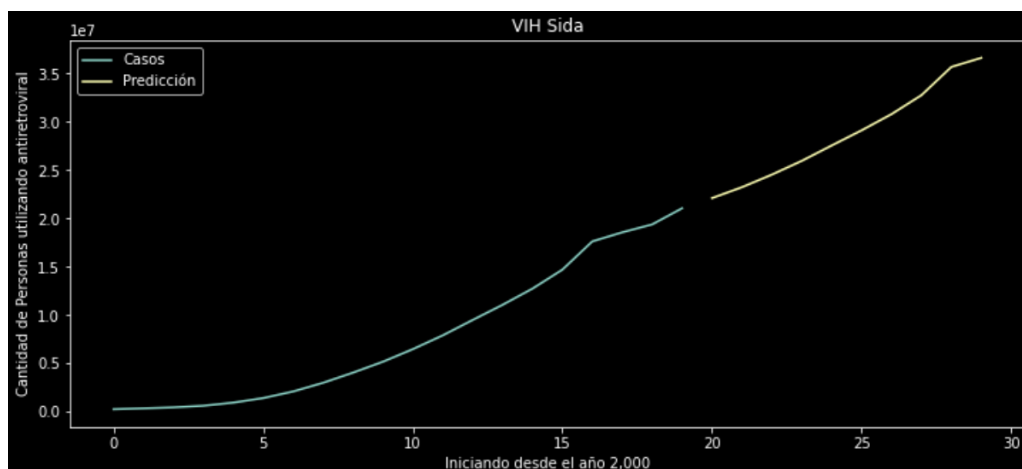


Figura 52: Predicciones con ARIMA sobre el caso de VIH SIDA.

E. API

Dentro del API se definieron distintas rutas que ejecutan distintos resultados de COVID-19 en base 64. Se utilizó base64 ya que elimina las solicitudes de HTTP request para la carga de imágenes al envolver el código de imagen codificando dentro de css o HTML. (Korvin, 2020)

Aunque inicialmente se había determinado que las predicciones se generan al ser solicitadas al momento de realizar pruebas en donde se realizaron varias consultas al mismo tiempo se mostró que el API mostraba lentitud y que llegar a presentar una caída al servicio. Fue debido a esto que se realizó una limpieza completa al API ya sea de prints como de funciones repetidas para cada uno de los diferentes casos. Por otro lado, se generaron nuevas rutas en las cuales no pueden ser consultadas por el frontend sino que un usuario externo que tenga control del API, como administrador, pueda reproducir las predicciones si estas no muestran el resultado deseado, guardando la mejor predicción de la semana en la carpeta del servidor.

Para el caso de redes bayesianas, aunque estas ya no se muestran en el frontend debido a su baja seguridad de predicción se decidió dejarlas con el objetivo de mantenerlas como histórico.

A continuación, se brinda una lista de las diferentes rutas:

Ruta	Función
<code>/covid/generate_confirmed_neural_network</code>	genera una predicción de red neuronal de casos confirmados
<code>/covid/confirmados_red_neuronal</code>	muestra la predicción de red neuronal en base 64 sobre casos confirmados
<code>/covid/confirmados_red_bayesiana</code>	genera una predicción de red bayesiana para casos confirmados

Ruta	Función
/covid/mortalidad_red_bayesiana	genera una predicción de red bayesiana para casos de mortalidad
/covid/generate_mortality_neural_network	genera una predicción de red neuronal de casos de mortalidad
/covid/mortalidad_red_neuronal	muestra la predicción de red neuronal en base 64 sobre casos de mortalidad
/covid/recuperacion_red_bayesiana	genera una predicción de red bayesiana para casos de recuperación
/covid/generate_cured_neural_network	genera una predicción de red neuronal de casos de recuperación.
/covid/recuperacion_red_neuronal	muestra la predicción de red neuronal en base 64 sobre casos de recuperación
/covid/confirmados_red_bayesiana_pais/<pais>	genera una predicción de red bayesiana de casos confirmados por país
/covid/confirmados_red_neuronal_pais/<pais>	genera una predicción de red neuronal de casos confirmados por país
/covid/mortalidad_red_bayesiana_pais/<pais>	genera una predicción de red bayesiana de casos de mortalidad por país
/covid/mortalidad_red_neuronal_pais/<pais>	genera una predicción de red neuronal de casos de mortalidad por país
/covid/recuperacion_red_bayesiana_pais/<pais>	genera una predicción de red bayesiana de casos de recuperación por país
/covid/recuperacion_red_neuronal_pais/<pais>	genera una predicción de red neuronal de casos de recuperación por país
/covid/get_cluster	muestra un cluster en base 64
/covid/total	muestra el total de contagios
/covid/fecha_actual	muestra la fecha actual
/covid/proporcion_de_contagios_covid	muestra la proporción de contagios de covid en un gráfico circular.
/covid/crecimiento_diario_covid	muestran el crecimiento de contagios con respecto al día anterior.

Tabla No 2: Rutas y funciones del API

El API se encuentra en el url: <https://mochi.software/>. Un ejemplo de su ejecución se presenta en la imagen posterior en donde se está buscando obtener la predicción sobre Guatemala en base 64. En donde luego se prueba la imagen utilizando un convertidor de base 64 a imagen.

Mostrando la siguiente imagen al decodificarlo:

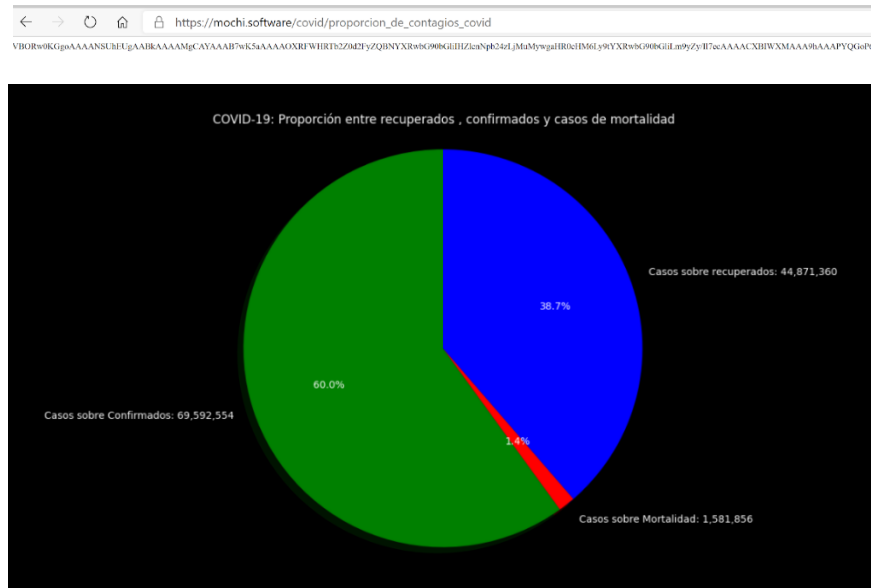


Figura 53: Codificación y decodificación de una gráfica en base 64.

F. Sistemas en la nube

Se cuenta con una máquina virtual en la nube, en la cual se encuentra corriendo el API.

Name	Instance ID	Instance state	Instance...	Status check	A..	Availability...	Public IPv4 DNS	Public IPv4...	Elastic IP
mochi.software	i-0481c91...	Running	t2.micro	2/2 checks...	+	us-east-2c	ec2-3-138-213-1...	3.138.213.189	3.138.213.189

Figura 54: Máquina EC2 en AWS.

Dentro de la máquina se ejecuta el siguiente comando para levantar el servicio Web con Gunicorn.

```
Welcome to Ubuntu 18.04.5 LTS (GNU/Linux 5.4.0-1030-aws x86_64)

* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage

System information as of Fri Dec 11 21:13:56 CST 2020

System load:  0.0          Processes:      97
Usage of /:   15.5% of 19.32GB   Users logged in:  0
Memory usage: 18%          IP address for eth0: 172.31.41.183
Swap usage:   0%

* Introducing self-healing high availability clusters in MicroK8s.
  Simple, hardened, Kubernetes for production, from RaspberryPi to DC.

  https://microk8s.io/high-availability

* Canonical Livepatch is available for installation.
  - Reduce system reboots and improve kernel security. Activate at:
  https://ubuntu.com/livepatch

25 packages can be updated.
0 updates are security updates.

New release '20.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Fri Dec 11 21:13:07 2020 from 190.56.74.39
ubuntu@ip-172-31-41-183:~$ cd Pandemias-API/
ubuntu@ip-172-31-41-183:~/Pandemias-API$ gunicorn -w 2 -b 127.0.0.1:5000 pandemias_api:app --daemon --reload --timeout 600
```

Figura 55: Máquina EC2 en AWS ejecutándose.

```
gunicorn -w 2 -b 127.0.0.1:5000 pandemias_api:app --daemon --
reload --timeout 600
```

Ya teniendo el API accesible se colocan las funciones de calendarización con CRON para generar las predicciones diariamente a las 2am, así evitar que se generen las predicciones cada vez que se ingresa a la aplicación Web y prevenir la saturación del servidor. El comando de CRON utilizado es el siguiente:

```
0 2 * * * curl -s
http://3.138.213.189/covid/generate_confirmed_neural_network >
/dev/null
```

Este comando realiza la llamada al API, la cual genera una imagen y la guarda en la misma carpeta del código del API para luego solamente leerla cada vez que alguien la requiere por parte de la aplicación Web.

E. FRONT-END

Se realizó una aplicación en Angular que se encarga de integrar los diferentes tableros GIS por medio de diferentes iframes, estos se cargan al presionar el botón la enfermedad dentro del menú, del mismo modo todas las consultas al API cargan al presionar este botón específicamente para el caso de COVID-19. Se manejaron diferentes iframes ya que en el primer prototipo se observó que la integración de todas las enfermedades en un solo tablero era

demasiado desordenado. Para el caso de VIH SIDA y de Ébola se usaron imágenes estáticas que simplemente son mostradas dentro del tablero. Es de mencionar que se utilizó una barra lateral para cada una de las enfermedades que muestra un breve análisis exploratorio, clusters y predicciones.

1. Área del menú para el caso de COVID-19

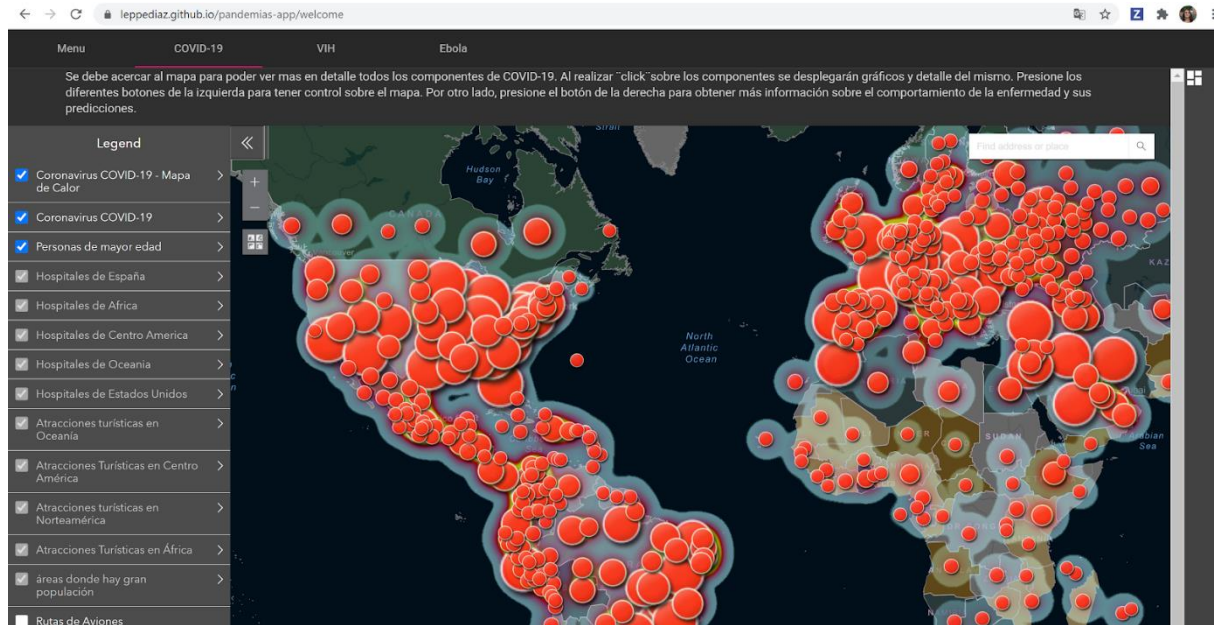


Figura 56: Menú y leyenda para el área del tablero de COVID-19.

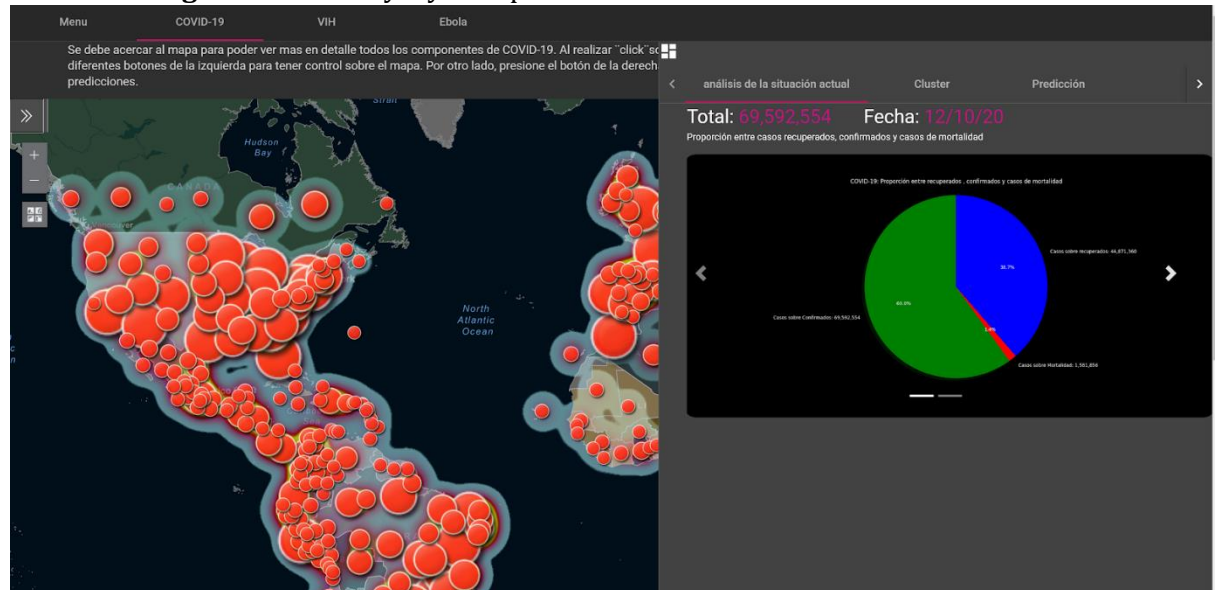


Figura 57: Mapa y menú para el área del tablero de COVID-19 sobre el estado actual.

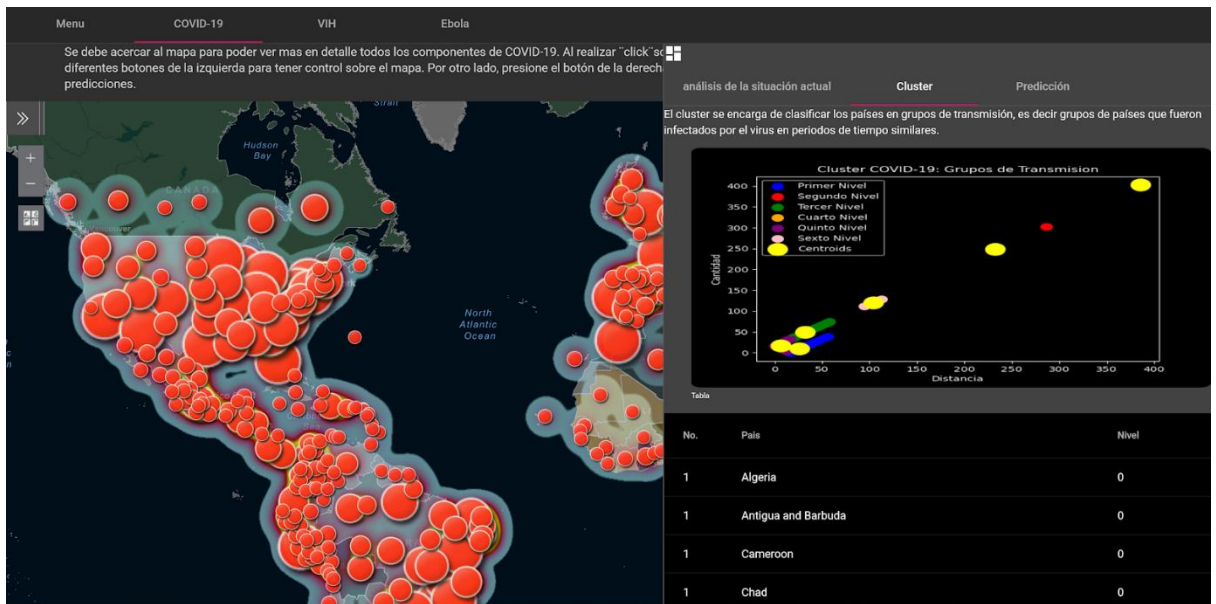


Figura 58: Mapa y menú para el área del tablero de COVID-19 mostrando el área de clusters.

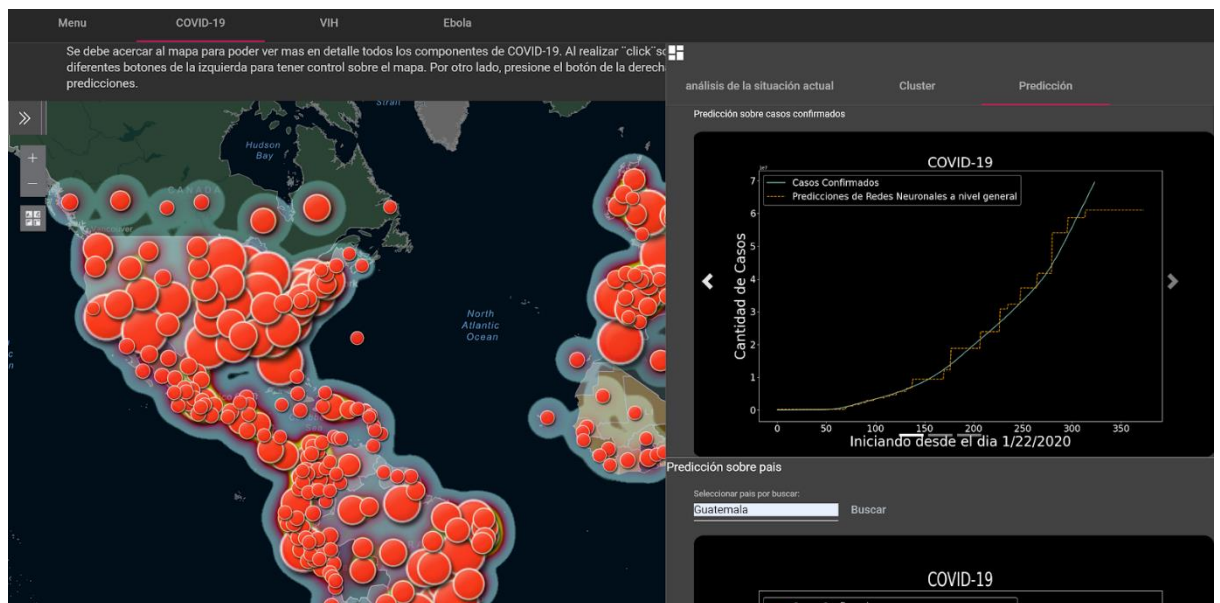


Figura 59: Mapa y menú para el área del tablero de COVID-19 mostrando el área de predicciones

2. Área del menú para el caso de VIH SIDA

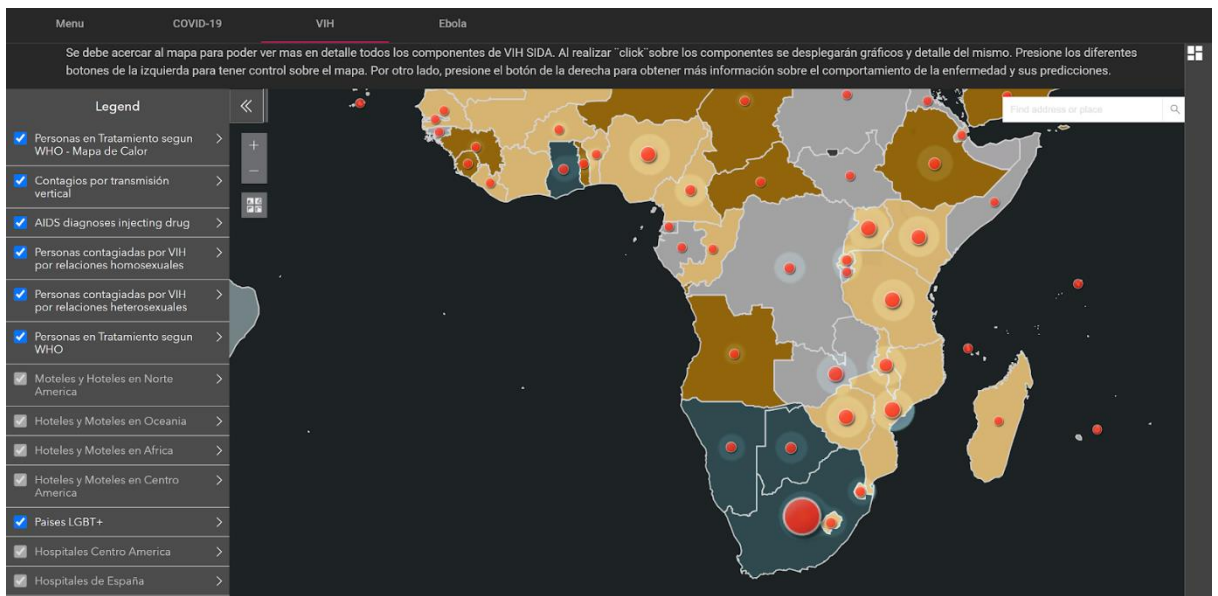


Figura 60: Menú y leyenda para el área del tablero de VIH SIDA.

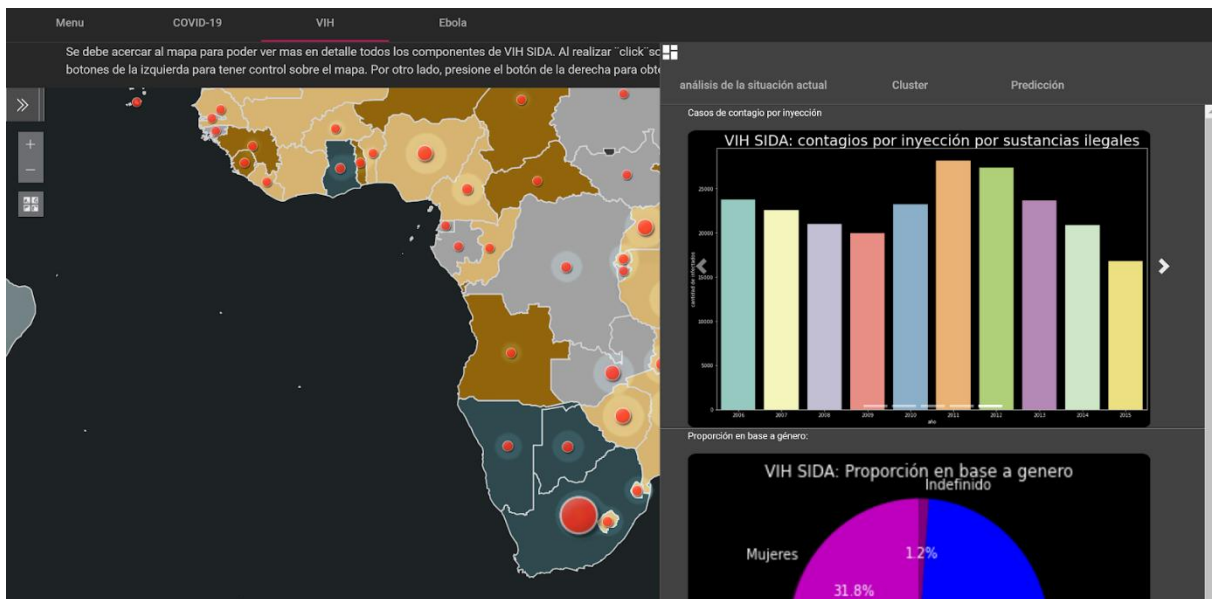


Figura 61: Mapa y menú para el área del tablero de VIH SIDA mostrando el estado actual.

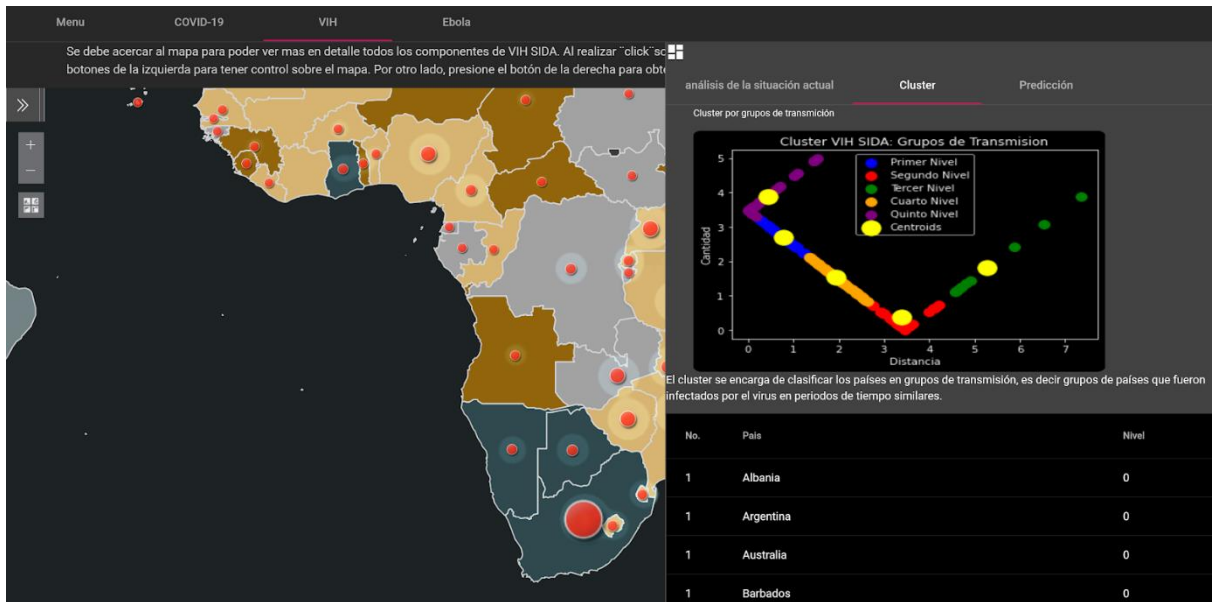


Figura 62: Mapa y menú para el área del tablero de VIH SIDA mostrando el área de clusters.

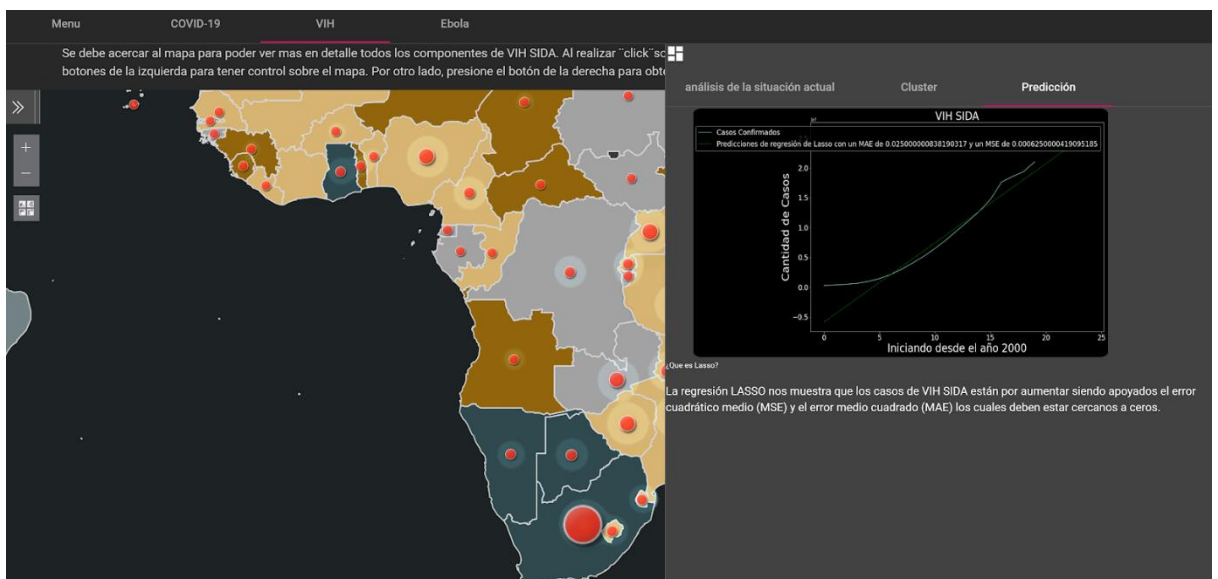


Figura 63: Mapa y menú para el área del tablero de VIH SIDA mostrando el área de predicciones.

3. Área del menú para el caso de Ébola

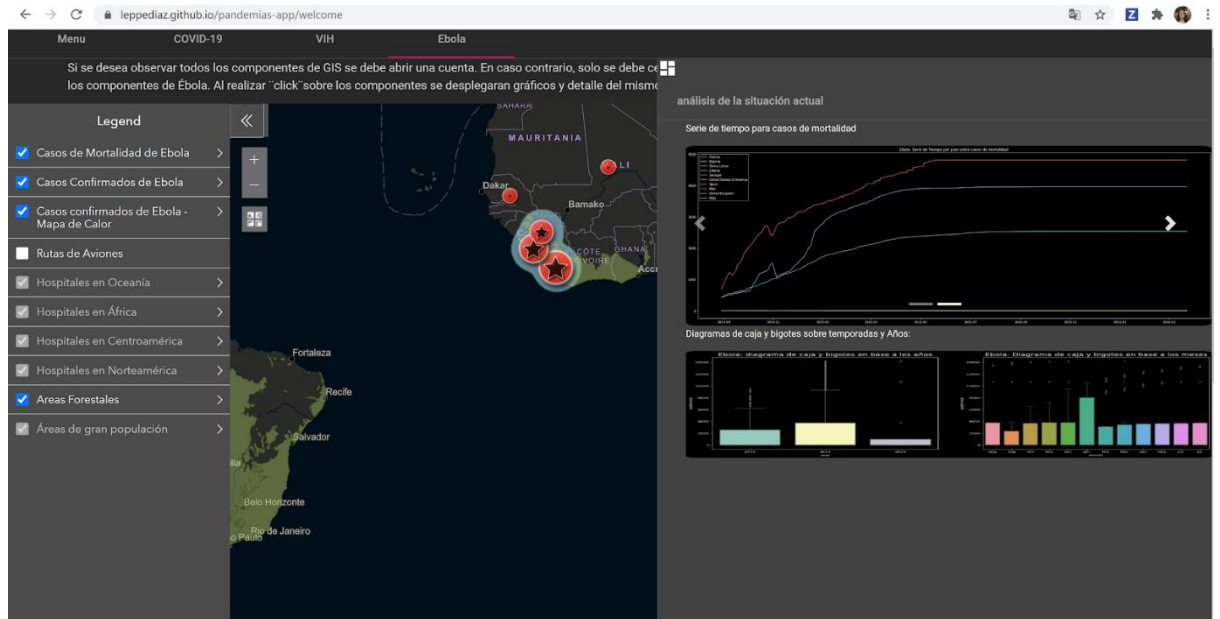


Figura 64: Mapa y menú para el área del tablero de VIH SIDA sobre su análisis de estado actual .

F. Pruebas de Usuario

Para validar el proceso del proyecto se realizaron varias encuestas en Google Forms, en la que se encontró que los diferentes tipos de usuario lograron desenvolverse bien dentro de la aplicación dentro de la aplicación, encontrando en mayoría un gran porcentaje de las funcionalidades según la lista de funcionalidades que les fue otorgada. Al igual que se aceptaron recomendaciones y de definir las áreas del dashboard que no fueron de su agrado. El dashboard tuvo una alta aceptación por el público general especialmente en las áreas de personal médico y profesionales de diferentes áreas.

Durante las primeras pruebas de prototipo se compartió una lista de funcionalidades del usuario por separado determinando si logró encontrar la funcionalidad específica, en respuesta se pudo obtener casi 100% en todas las funcionalidades con excepción a realizar una búsqueda de localización. Por otro lado, en la segunda prueba de prototipo se puede observar que mientras el usuario fue capaz de encontrar casi las funcionalidades en el COVID-19 no exploró a tanto detalle el caso de Ébola y VIH Sida. Siendo la funcionalidad menos utilizada el iniciar una ruta con "OpenStreetMap".

Dentro de la tercera prueba de prototipo se pudo observar que para los profesionales en medicina, químicos farmacéuticos, ingenieros en bioquímica y microbiología , ingenieros en administración, desarrolladores de software, ingenieros industriales y analistas de datos se tuvo

una aprobación del 81.3% siendo 20% el que resolvió que el tablero resuelve parcialmente este objetivo que había sido escrito como pregunta, el 20% menciono que les gustaría ver datos referentes a la edad, sexo de las personas y que tan ocupados se encuentran los hospitales solicitados. Mientras los usuarios que se encuentran ejerciendo en el área de medicina mencionaron que les sería de utilidad gracias a que al obtener de forma inmediata información del mapa sobre áreas afectadas y sus estadísticas respectivas, pueden obtener información de donde vinieron los pacientes para recibir el contagio y conocer las tasas de infecciones en dichos países, por lo que termina sirviendo a sí mismo como una recolección de la historia clínica de los pacientes y de esta manera orienta mejor el diagnóstico. Por otro lado, mencionan que les servirá específicamente para tomar decisiones en base a preparación y prevención, mencionando su gran aceptación por hospitales y rutas de acceso, aunque solicitando información del estado actual de cada hospital. En el área de ingeniería industrial y administración mencionaron su utilidad para tomar medidas predictivas sanitarias en las logísticas de transporte y surtido, mencionando en un caso de usuario que este servirá para las medidas sanitarias para Walmart. Por otro lado, mencionaron también su utilidad para realizar predicciones de ventas y exportación e importación de medicinas. Para los usuarios que son desarrolladores de software, analistas de datos, consultores jr y analistas de aplicaciones mencionaron su utilidad para obtener una idea bastante confiable de cómo va a seguir el estado de la enfermedad, el comportamiento que tendría, lugares de aglomeración y para asuntos personales como tomar un viaje. Dentro del área de farmacéutica e ingeniería en bioquímica y microbiología se mencionó la importancia de romper los paradigmas y el tabú del caso de VIH SIDA y los contagios tipo zoonóticos.

1. Primera prueba de prototipo

Has encontrado el botón de menú? (en cual despliega un análisis exploratorio de los datos)
17 respuestas

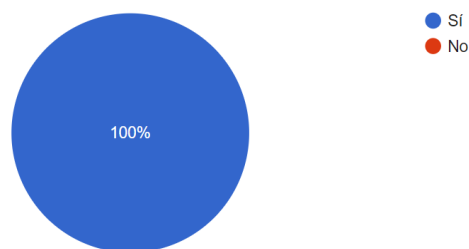


Figura 65: Gráfico circular que busca la funcionalidad del botón de menú.

Has buscado un país en específico?

17 respuestas

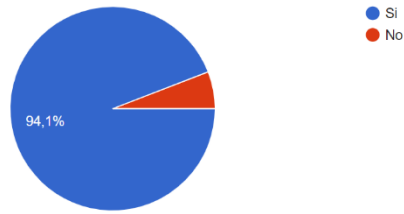


Figura 66: Gráfico circular que busca la funcionalidad de encontrar un país en específico.

Has visto la curva de crecimiento

8 respuestas

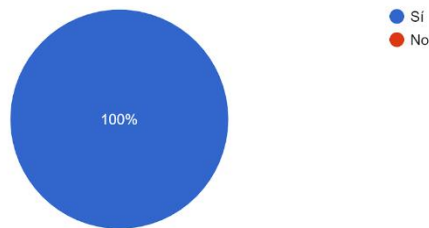


Figura 67: Gráfico circular que busca la funcionalidad de ver la curva de crecimiento por país.

2. Segunda prueba de prototipo

Marca temporal	Que funciones veas en el Dashboard?
16/09/2020 23:54:21	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. Encuentre hospitales y clínicas en el país que visualice. Encuentre moteles, hoteles etc en el mapa de VñH. visualice los puntos de población de cada país. encuentre el boton que desplaza
17/09/2020 1:13:41	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que visualice. Encuentre moteles, hoteles etc en el mapa de VñH. visualice los puntos de población de cada país. encuentre
17/09/2020 4:53:48	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que visualice. visualice los puntos de población de cada país. Observe las areas verdes en cada país en el mapa de Eñola
17/09/2020 9:21:04	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que visualice. visualice los puntos de población de cada país. Observe las areas verdes en cada país en el mapa de Eñola
17/09/2020 14:32:16	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. visualice los puntos de población de cada país. encuentre el boton que explica que se está visualizando y como utilizarlo. Encuentre la informacion Day of COVID presionando sobre
17/09/2020 19:04:22	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que presione (boton rojo) con el nombre. Encuentre moteles, hoteles etc en el mapa de VñH. visualice los puntos de poblab
17/09/2020 23:18:59	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. visualice los puntos de población de cada país. Observe las areas verdes en cada país en el mapa de Eñola
18/09/2020 14:03:46	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que presione (boton rojo) con el nombre. Encuentre moteles, hoteles etc en el mapa de VñH. visualice los puntos de poblab
18/09/2020 14:21:46	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que presione (boton rojo) con el nombre. Encuentre el grafico circular del C
18/09/2020 14:28:56	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. Logre hacerle zoom al mapa
18/09/2020 19:03:39	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que presione (boton rojo) con el nombre. Encuentre moteles, hoteles etc en el mapa de VñH. visualice los puntos de poblab
18/09/2020 19:47:21	Observa el mapa del COVID-19. Eñola y VñH. Realice click dentro de las diferentes características para encontrar graficos de cada enfermedad. busque su localizacion. Encuentre hospitales y clínicas en el país que presione (boton rojo) con el nombre. Encuentre moteles, hoteles etc en el mapa de VñH. visualice los puntos de poblab
22/09/2020 15:01:08	Observa el mapa del COVID-19. Eñola y VñH

Figura 68: Resultados de la encuesta sobre el porcentaje funcionalidades encontradas.

Que te perdio del dashboard?

12 respuestas

- Ver abajo
- Al darle click a la pestaña de VIH por ejemplo estando en Covid, te sigue mostrando Covid hasta que le das a ver mapa.
- No me di cuenta de los hospitales, hoteles y demás puntos hasta que hice zoom accidentalmente. Tal vez especificarlo en como utilizar la herramienta
- No sé lo que significa la bandera de LGTB en el mapa de VIH
- Cuando presioné COVID-19 esperé que se abriera el mapa automáticamente, pero hay que seleccionar "mapa" después, eso puede confundir un poco.
- Me perdí al momento de cambiar de VIH a Ebola tenía que volver a dar click en mapa de *nombre* para poder cambiar.
- donde se encontraban las diferentes enfermedades, porque al bajar para ver las instrucciones se pierde la parte superior donde están.

Figura 69: Resultados de la encuesta sobre comentarios en donde se podría mejorar el tablero.

Que te gustaria ver en el dashboard?

12 respuestas

- Los nombres de los hospitales para que las enfermos puedan dirigirse rápidamente; información de la situación diaria por país y departamento. Es posible agregar información del Seguro Social para Guatemala? Dónde se realizan los isopados, aquí en Guatemala.
- Me encanto la ventana de Que estoy viendo porque en si todo esta muy completo, entonces el empezar con una explicación cada vez que entras a un tab seria super
- Datos de resumen por país
- Me parece una herramienta bastante útil e interesante, por el momento no puedo pensar en algo que agregar más que otras posibles enfermedades endémicas algún día.
- Una forma de regresar al menú de seleccionar qué enfermedad quiero ver en lugar de tener que seleccionar "atrás" en mi navegador.
- Una tablita de estadística mundial de cada pandemia (no tiene que ser grande)
- Vinculación a noticias de relevancia.

Figura 70: Resultados en la encuesta sobre sugerencias de funcionalidades.

Que errores crees que se deben arreglar en el dashboard?

13 respuestas

- Es posible ampliar el tamaño de la letra y el espacio que ocupa en la pantalla y/o proporcionar información más amplia de cómo manejar e interpretar las pantallas para que sea fácil de manejar para todas las personas? No logré encontrar la gráficas.
- Desaparecieron las leyendas de los mapas, me aparece como No Legend
- Cuesta diferenciar los tamaños de acuerdo a la leyenda de VIH
- Ortográficos principalmente.
- Tiene faltas de ortografía (le hacen falta tildes). Tampoco marca todos los hospitales, por ejemplo El Pilar en zona 15 no sale pero eso ya debe ser que no lo tiene registrado en su base de datos.
- No encuentre ningún error (cualquier error te aviso)
- Colocar el como utilizar el mapa en cada uno de estos, optimizar esa opción ya que no pude observar la parte inferior.

Figura 71: Resultados en la encuesta sobre sugerencias de arreglos dentro del tablero.

3. Tercera prueba de prototipo

¿Cuál es tu profesión?

16 respuestas

Estudiante
Estudiante de medicina
Software Developer
Química Farmacéutica
Consultor jr
Bioquímica y Microbióloga
Administradora
Analista
medico

Figura 72: Pregunta sobre la profesión de los encuestados.

¿Te serviría este tablero para apoyar en la preparación, respuesta y control de enfermedades?

16 respuestas

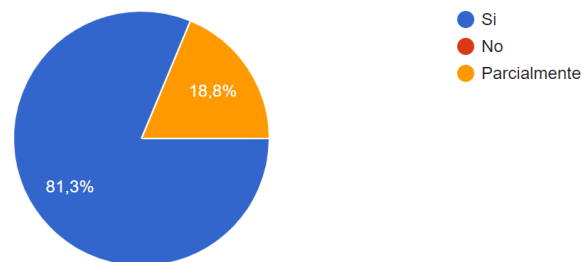


Figura 73: Resultados en la encuesta en un gráfico circular del cumplimiento del objetivo general.

¿Qué funciones lograste encontrar?

16 respuestas

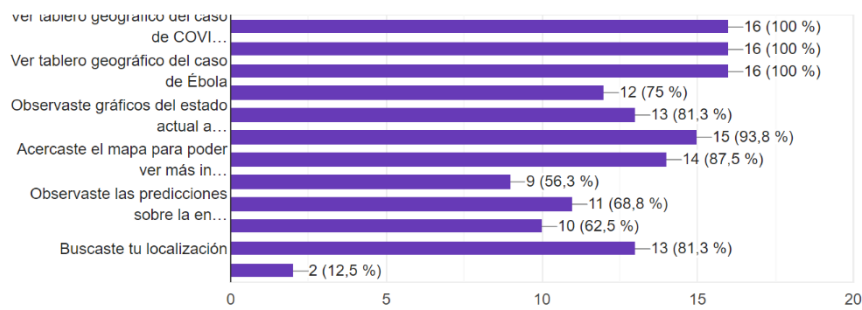


Figura 74: Resultados en la encuesta sobre las funcionalidades finales encontradas.

¿Para qué te serviría este tablero?

16 respuestas

Podría servir para conocer de donde vinieron los pacientes (algun lugar endemico de alguna enfermedad) y conocer las tasas de infecciones en dichos países, y esto nos sirve para la recolección de la historia clínica de los pacientes y así orientar el diagnóstico

Para visualizar la situación actual del Covid y poder tener una idea bastante confiable de como va a seguir el estado de la enfermedad, el comportamiento que tendra, lugares de aglomeración y también interesante el ver a donde viajar y donde abstenerse. De parte del VIH y Ebola sirve mucho en romper paradigmas de contagio y de areas de contagio, asociar algunos factores de ambiente como la diversidad de fauna y el ebola, y también para poder visualizar ciertas causas de contagio, así como hostiles cercanos.

Conocer datos actuales de pacientes confirmados de COVID, ebola y VIH según el país.

Ver predicciones de las enfermedades para tomar decisiones en cuanto a preparación y prevención. Además entender el comportamiento de las mismas.

Para el monitoreo de contagios, ver predicciones de contagios y determinar la mortalidad del virus.

Figura 75: Resultados en la encuesta sobre la utilidad del tablero en sus diferentes profesiones.

¿Qué más le gustaría ver en este tablero?

Sinceramente, está muy completo. Posiblemente una comparación entre curvas de crecimiento de ciertos países, así como en el caso de la enfermedad de Ebola específicamente en su visibilidad.

Para el VIH podría ser bueno tener un cuadro de diálogo con las medidas de prevención y tratamiento. Esta enfermedad sigue siendo tabú así que ayudar a informar a la población.

Tener una opción de exportar resultados

Disponibilidad de espacios y pruebas de diagnóstico en hospitales (En caso de COVID).

Con lo que veo me siento conforme

Datos en tiempo real

Tendencias históricas

ver algunos datos en español, las gráficas tienen la información con letras muy pequeñas

ver tendencia de casos por sexo y edad

Me gustaría ver algún pequeño detalle sobre las leyendas en la parte de VIH ya que algunas no se explicaban muy bien o también están sobrepuestas. De ahí todo está bastante completo

Que haya menos superposición de iconos de leyenda, ya que al estar sobrepuestos varios no se pueden ver con claridad.

Porcentaje de casos por departamentos que contiene cada país.

ir de donde vinieron los pacientes (algun lugar endemico de alguna enfermedad) y conocer las tasas de infecciones en dichos países, y esto nos sirve para la recolección de la historia clínica de los pacientes y así orientar el diagnóstico

icas. Visualización de predicciones. Exponer estadísticas sociales

Alarmas de seguridad o avisos que se activen con el GPS del celular a medida de estar más precavidos y tomar las medidas sanitarias correspondientes.

Me gustaría que quizás no se usara el sistema de puntos que tiene. Lo digo porque era demasiado agobiante tener que ver gran cantidad de puntos en área y tener que acercarse. Sería mejor que dependiendo que tan cerca del mapa este uno, ver

Figura 76: Resultados en la encuesta sobre las necesidades a cubrir con sus profesiones.

VII. DISCUSIÓN

Para validación del objetivo general en donde se determinó que el proyecto ayuda en la preparación, respuesta y control de enfermedades se llevó a cabo una encuesta para los diferentes casos de uso seleccionados. Esto fue validado debido a la alta aceptación en la pregunta de validación de este objetivo y los diferentes casos de uso mencionados en las encuestas. En donde se pudo observar futuras utilidades como medidas sanitarias y para planificar, proyectar, instalar y poner en marcha diferentes productos. Por otro lado, para el personal médico es de utilidad ya que permite al médico orientar el diagnóstico ya que le permite ver los riesgos de contagio que existe en una posición geográfica determinada y de esta manera poder orientar el diagnóstico en base a donde ha visitado su paciente. Por último, para el usuario general, es decir cualquier usuario, el tablero permite al usuario tener un análisis y entendimiento de distintas enfermedades pandémicas para uso personal y laboral. Para permitir que el tablero fuera fácil de utilizar para los diferentes usuarios se llevó a cabo dos entrevistas que transformaron el tablero en un tablero fácil de utilizar para el usuario, el cual es validado debido a que los usuarios lograron encontrar una gran cantidad de las funciones.

Los tableros de información que fueron unificados en Angular presentaron la información sobre los casos y predicciones de las distintas enfermedades incluidas en el proyecto. Debido a que el COVID-19 es una enfermedad que se actualiza diariamente los tableros de GIS no permitieron crear una capa discreta por medio de recuentos y cantidades que muestra puntos geográficos enfocados por país y provincias que tiene un nombre que cambia constantemente por lo que el dato actual se actualiza diariamente dentro de la capa con el mismo nombre pero sin guardar la información anterior y manteniendo su estructura inicial, es decir sin agregar nuevas columnas dentro del dataset. Sin embargo, se puede observar la curva de crecimiento gracias al API realizado en Flask consultando el país que se desea visualizar, permitiendo de esta manera obtener curvas de crecimiento por país y en resultados generales es decir ver la curva de crecimiento de todos los países juntos. Por otro lado, los "Heatmap" o mapas de calor nos permitieron observar la gravedad de las áreas de contagio por continentes mientras la capa "Counts and Amounts (Size)" o Recuentos y cantidades, nos permiten ver la expansión por países, con el objetivo de que usuario pueda tener un mayor alcance de la expansión de la enfermedad al mismo tiempo que va cumpliendo los requisitos del primer objetivo específico que deseaba presentar la información sobre casos de personas confirmadas de contagio, casos de personas recuperadas, casos de mortalidad y distintos gráficos sobre los datos de cada enfermedad que incluyen curvas de crecimiento y gráficos de contraste en un tablero actualizado. Del mismo modo, el tablero muestra un análisis descriptivo de la data que

muestra datos actualizados, clusters y gráficas predictivas que proporcionan al usuario un conjunto de gráficos que describen la información de manera objetiva, estos tuvieron una gran aceptación por parte de los diferentes usuarios entrevistados.

En los resultados obtenidos se puede observar a primera vista dentro de los tableros de GIS que existe una alta tendencia a enfermedades dentro de África, especialmente en el país de Sudáfrica. Según los resultados, se procedió a investigar a Sudáfrica el cual cuenta con una esperanza de vida baja siendo 66.2 años para la mujer y 59.3 años para el hombre según los datos de World Health Organization. Al mismo tiempo, una de las mayores causas de muerte de Sudáfrica es la tuberculosis la cual es 8.8% de las muertes dentro del país seguido de influenza y neumonía con 5.2% y diabetes mellitus con 4.8%. Todas estas enfermedades vuelven los síntomas, especialmente de COVID-19 y VIH SIDA, enfermedades mortales, aumentando de esta manera la tasa de mortalidad (The World's Largest HIV Epidemic in Crisis: HIV in South Africa, 2020) Por otro lado, se pudo observar una relación clara entre los brotes de enfermedades infecciosas humanas de tipo "zoonóticas" lo cual significa que se originan a partir de un virus que infectan a animales que luego terminan infectando a las personas. Esto se validó, al observar dentro del mapa de las tres enfermedades. Dentro del caso de Ébola y VIH SIDA, los países más afectados se encontraban en un área con alto nivel de vida salvaje, flora y fauna. Mientras el inicio de COVID-19 en Wuhan Hubei en China no se originó en un área con alto nivel de vida salvaje, si se encontraba cerca de las montañas del sureste de China y de Indo Burma los cuales son diferentes reservas naturales.

Dentro del caso del tablero de COVID-19 tanto las predicciones como su curva de crecimiento siguen un patrón de sigmoide, esto se puede observar en su curso tanto en su borde como de decadencia. Esto fue comprobado con la Universidad de Helsinki, en su artículo de Epidemias como un proceso natural, que compara como una epidemia seguirá el comportamiento de una sigmoide. Se observa una enfermedad del tamaño de una epidemia y cómo se relaciona con los portadores potenciales, la población susceptible, y los mecanismos que la enfermedad pueda explotar y difundir, el cual cuando se impide que el agente infeccioso acceda a más recursos de energía libre que obtiene de una población sana la propagación de la enfermedad empezará a disminuir. Por lo tanto, este iniciará de manera exponencial mientras va provocando que su pendiente vaya decreciendo y que se acerque a 1. (Helsinki, 2020) Para este sistema se utilizaron redes neuronales, estas se vuelven a calcular diariamente y guardan el resultado que obtuvo mejor exactitud por medio de Cross-Validation. Debido a que no se desea saturar el servidor se dejó como opción al administrador en donde él puede volver a entrenar la red neuronal por medio de volver a ejecutar el url. Se investigó y se decidió trabajar con regresión de cresta bayesiana, ya que este al igual que redes neuronales funcionan para datos

que se encuentran cambiando diariamente. No se pudo tomar en cuenta la cresta bayesiana debido que cuenta con un error absoluto medio (MAE) y un error medio cuadrático (MSE) de que se sobrepasa de 10,000 cuando estos deberían estar cerca de 0. Aunque no existen límites aceptables para MSE, en donde solo se establece que mientras menor sea el MSE mayor será la precisión de la predicción este muestra un resultado demasiado alto. Por esta misma razón se prosiguió a realizar Cross-Validation sin embargo se obtuvo un resultado negativo. Esto se puede deber a que el algoritmo no guarda los datos, sino que se ajusta a ellos al mismo tiempo que maximiza iterativamente la probabilidad logarítmica marginal de las observaciones dando resultados ligeramente más altos que los que se observan originalmente, debido a que los datos se actualizan de manera diaria se decidió no prescindir de ningún dato y por realizar la predicción lo más acertada posible tomando todos los valores lógicos que proporciona la base de datos.

Para analizar la situación actual de la enfermedad se incluyeron varios gráficos como un gráfico que nos permite ver cuánto creció la enfermedad en comparación con el día anterior, un "pie chart" o gráfico circular que se despliega tanto a nivel general como de país, que nos muestra la proporción actual entre casos confirmados, mortalidad y recuperados. Por último, se incluyó un cluster en el cual se encargó de clasificar los países en grupos de transmisión similares en base a sus casos de mortalidad y casos confirmados. El número de clusters fue determinado por un gráfico de codo, el cual determinó que el número óptimo de clusters fue el número 6 ya que este fue el punto en donde se estabilizó la gráfica. Dentro de los clusters del caso de los primeros contagios en cada país se pudo observar una diferencia de contagio en el país de Samoa el cual tuvo su primer contagio el 18 de noviembre y las islas Marshall el 28 de Octubre (grupo 3 y 6). Al observar los países de Samoa y las Islas Marshall, se pudo observar que ambas son islas del pacifico de oceanía en donde se puede observar que muchos países que lo rodean no han tenido contagios reportados debido a que habían cerrado temprano y todos se encuentran como islas ("COVID-19 Is Reaching the Last Coronavirus-Free Nations on Earth", 2020).

Dentro del área GIS se pudo observar gracias a los datos de provincias en que mientras mayor densidad poblacional haya en un área mayor será la cantidad de infectados. Esto se debe a que el COVID-19 es una enfermedad altamente contagiosa y se contagia por medio del contacto con otra persona, por lo que al estar en un lugar con una alta densidad poblacional se corre más riesgo de contagiarse de esta enfermedad, esto se puede observar comparando el este y el oeste de estados unidos en donde se puede observar que las provincias donde existe una mayor densidad poblacional existe una mayor cantidad de contagios por ejemplo Washington comparado con Montana y Wyoming. Por otro lado, en el caso de las áreas turísticas no se logró encontrar ninguna correlación debido a que no se tiene datos de lugares específicos de contagio,

sino que de provincias o países por lo que no se pudo verificar la correlación. En el caso de los vuelos aéreos se pudo crear una leve relación observando el área de Estados Unidos debido a que un alto porcentaje de los casos presentaba en que mientras menor sean los vuelos aéreos menor fueron la cantidad de casos; esto es fácilmente observable ya que Alaska, Wyoming y Montana son mucho menores que los casos de California, Florida y Nueva York los cuales tuvieron alta concurrencia de vuelos aéreos, al mismo tiempo que se puede observar esta comparación por medio de observar el este y oeste de estados unidos. Sin embargo, es de mencionar que esto no se cumplió para todas las provincias. Por último, se pudo observar que no existe una tendencia alta entre la cantidad de personas de mayor edad y mortalidad que existen en los diferentes países ya que esta tendencia no se presentó en todos los países aunque sí pudo influir en la alta mortalidad de Bélgica, Francia, Polonia, Rumania, Inglaterra, Puerto Rico, Argentina, Canadá y España ya que presentan una alta tasa de personas mayores a comparación de los demás países y tuvieron una significativa tasa de mortalidad.

En el caso del VIH se pudo observar una alta concentración en África, especialmente en Sudáfrica que ya se encontraba en una batalla contra la tuberculosis y tuberculosis multirresistente al igual que una extrema desigualdad según BBC. Ambas enfermedades pueden hacer el caso de VIH SIDA una enfermedad mortal. Se pudo observar que la mayor cantidad de contagio se obtuvo por medio de inyecciones seguido de contagios por relaciones heterosexuales. El hecho de que las relaciones homosexuales no sean una causa principal se debe a que los contagios disminuyen con ayuda del uso y accesibilidad de retrovirales, esto funcionó de manera contraria para las relaciones heterosexuales los cuales han aumentado mientras se avanza los años, esto puede ser influenciado por el hecho de una menor población homosexual y el hecho de que las parejas homosexuales han ido adquiriendo medidas para evitar la propagación tales como realizarse un examen de VIH SIDA de manera recurrente, esto se validó al enviar el proyecto a personas que pertenecen a la comunidad LGBTQ+ y obtener retroalimentación de su parte. Por el caso contrario, los casos de transmisión vertical es decir de madre a hijo disminuyeron en el tiempo, esto se debe a que la causa de que en los casos de transmisión vertical (madre a hijo durante el parto o lactancia) se pudo reducir utilizando antirretrovíricos para la madre e hijo durante la gestión y periodo de lactancia. Por otro lado, se pudo realizar una comparación en base a género en donde se dio un aumento en los casos del sexo femenino esto se debe a que el sexo femenino tiene un mayor riesgo de contraer VIH SIDA durante una relación sexual. (Who HIV/AIDS, 2020) al mismo tiempo que mientras se ha dado un aumento a lo largo del 2006 al 2011 de caso por medio de inyección por drogas según CDC uno de cada día diagnósticos de VIH ocurren entre personas que se inyectan drogas siendo el caso femenino el 21% de los casos (CDC, 2017). Por otro lado, se realizó un cluster en donde se

puede observar que dentro del grupo 3 existe una gran porcentaje de países africanos mientras en el grupo 4 se pudo observar países africanos y países europeos. Esto se debe a que el VIH SIDA se inició en África específicamente dentro del Congo (Origin of HIV & AIDS, 2020). Dentro del área de las predicciones se utilizaron dos diferentes algoritmos los cuales pueden manejar predicciones a pesar de tener una cantidad limitada de datos siendo el presentado dentro del tablero el algoritmo Lasso el cual tuvo un error absoluto medio MAE de 0.025 y error medio cuadrático MSE de 0.00062, ambos se mostraron ser cercanos a 0 por lo que se puede esperar que las predicciones de VIH SIDA sigan creciendo a lo largo de los años por lo que se debe seguir dando concientización de esta enfermedad y mostrarle a la población la importancia de los retrovirales. Por otro lado, se probó el algoritmo de series de tiempo con ARIMA en donde se obtuvo un MSE superior a 1 millón por lo que se decidió no tomarlo en consideración.

Dentro del Tablero de Ébola se pudo crear una relación clara entre las regiones o países más afectadas siendo Sierra Leona y Liberia y la tendencia en los brotes de enfermedades infecciosas humanas de tipo "zoonóticas" ya que ambos se encuentran en áreas con una alta cantidad de vida salvaje. Estudiando a fondo estos países, se pudo observar que dentro de Sierra Leona existe un alto nivel de pobreza siendo considerado como el segundo país más pobre de África, tuvo una guerra civil de 11 años de la cual todavía vive con las consecuencias y se subsidia de agricultura y caza. En la vida salvaje de Sierra Leona, es bastante diverso teniendo desde zebras hasta diferentes felinos y grandes extensiones de tierra. El problema del gran contagio que existe en este país recae en que hay una gran alimentación en la fauna silvestre en las comunidades rurales que cazan para obtener carne para suplementar sus dietas e ingresos. Estas comunidades se arriesgan a propagar la enfermedad a través de especies que contienen el virus, especialmente los murciélagos de fruta, primates y duiqueros. (Organización de las Naciones Unidas para la Alimentación y la Agricultura, 2020). Explicando la relación que hay entre los países subdesarrollados y el contagio del Ébola. En el caso de Liberia, es una situación similar ya que también cuenta con una gran cantidad de vida salvaje y una alta pobreza ya que vive 51% de la población en pobreza y 16% en pobreza extrema. Se pudo observar que la temporada del año con mayor cantidad de contagios fue en invierno esto se puede deber a que es el tiempo en el cual el sistema inmunológico se encuentra más débil al igual que las temperaturas bajas aumentan lo cual concuerda con el ambiente de supervivencia del virus del Ébola ya que el Ébola puede desaparecer con calor, sol y luz ultravioleta los cuales no se dan abasto en temporadas de invierno.

Realizando comparaciones de mapa entre las distintas enfermedades se pudo observar que existe una tendencia de mayor contagio de estas enfermedades en donde se tiene un índice de progreso social bajo y una alta cantidad de niveles excepcionales de vida salvaje, flora y fauna.

Por lo que se hace una recomendación en poner especial atención a estos países especialmente en los países que se alimentan de murciélagos de fruta. (FAO, 2020). Esto es fácilmente observable dentro de la capa de índice de progreso social de VIH SIDA y la capa de puntos críticos de biodiversidad del mapa del Ébola.

Los resultados de Amazon en EC2 (Elastic Cloud Compute) mostraron ser satisfactorios. Sin embargo, se determinaron ciertas rutas dentro del API en donde el administrador se debe hacer cargo si llega a observar una predicción que no llega a obtener el resultado deseado esto se realizó con el objetivo de no llegar a saturar el servidor y mejorar la rapidez del mismo. En base a diferentes pruebas técnicas se decidió optimizar el API por medio de eliminar prints y optimizar el código, todo esto se realiza con el objetivo de tener el API mucho más rápido para el usuario y evitar errores de HTTPS en donde se sobrepasa el límite de tiempo. Mostrando de esta manera un servicio que realice predicciones diariamente y guarde la mejor predicción de manera exitosa que garantiza rapidez y disponibilidad a pesar de tener una gran cantidad de conexiones.

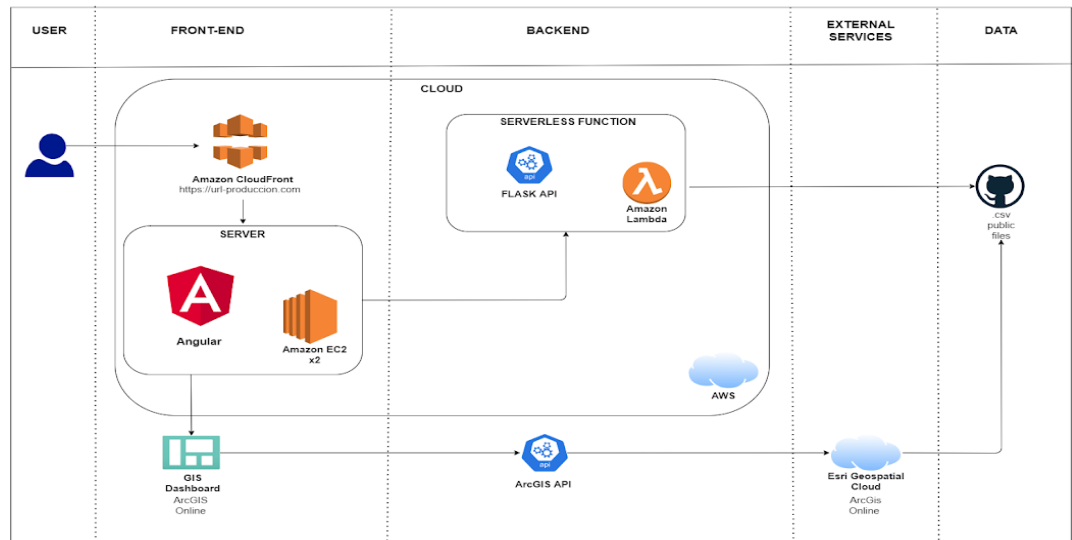
VIII. CONCLUSIONES

- Dados los resultados de las encuestas de aceptación de uso de la plataforma, se pudo obtener y validar las variables existentes, así como otras variables específicas, que serían de mayor utilidad en áreas de industria médica, farmacéutica, administrativa y logística. Algunas de las variables mencionadas anteriormente son: puntos críticos de biodiversidad, índice de progreso social global, cantidad de personas de la tercera edad por país y densidad poblacional. Las variables que indicaron los usuarios que podrían ser de utilidad para sus áreas de trabajo son: cantidad de pacientes contagiados de la enfermedad por hospital, venta y abastecimiento de medicamentos y sucursales de empresas en específico.
- Según los resultados de las encuestas se obtuvo que la aplicación si podría ser útil para brindar la información necesaria para preparación, respuesta y control de pandemias dado a su aceptación y las diferentes utilidades que fueron propuestas por los usuarios.
- Tomando como base las herramientas de visualización utilizadas en el presente proyecto, se logró realizar representaciones gráficas que ayudan a entender mejor el estado de los países afectados por las enfermedades endémicas, así como representaciones de regiones de contagio, curvas de crecimiento y poder presentar de manera geográfica los casos de mortalidad, recuperación y confirmación de infectados para cada enfermedad. Al igual que se incluyeron gráficos sobre variables externas que pueden ser de utilidad para formar una relación visual de cada enfermedad con la variable.
- Como resultado de los gráficos de concentración de las enfermedades, realizado sobre el Sistema de Información Geográfica (GIS) se puede concluir que uno de los continentes más afectados es el continente Africano seguido de Europa, con excepción a Europa del norte, y Norteamérica, para todas las enfermedades incluidas en el presente proyecto.
- Dadas las predicciones realizadas para COVID-19 y VIH Sida en base al crecimiento de contagios se pueden observar ciertos comportamientos de tendencia. En el caso de VIH Sida, se observa la cantidad de casos en crecimiento, mientras que en el caso de COVID-19 se puede observar como el crecimiento puede tender a una sigmoide.

- Los sistemas en la nube con Amazon EC2 permitieron tener predicciones actualizadas, listas para cualquier cambio del comportamiento de la enfermedad al igual que se buscó que estas mismas tuvieran rapidez y disponibilidad.

IX. RECOMENDACIONES

- Para poder realizar una continuación adecuada al proyecto y generar una aplicación lista para producción se deben de realizar pasos muy importantes, más en el tema de arquitectura y hosting, por ende, para ello se debe:
 - Cambiar el hosting de la aplicación Web hacia un servicio diferente al de github.pages, el cual es un servicio gratuito para páginas con poco tráfico, este podría migrar hacia un servidor EC2 de AWS por ejemplo.
 - Aumentar la capacidad del servidor que contiene el API, actualmente el ancho de banda es limitado y la capacidad de cómputo no es la adecuada para tolerar una gran cantidad de tráfico debido a que es un servicio gratuito de AWS, durante el proyecto y las dificultades encontradas se optimizó el código para utilizar lo menor de recursos de sistema posible, sin embargo, queda aquí en el hosting una oportunidad de mejora.
 - Actualmente se tiene el dominio mochi.software para el API por temas de costos, sin embargo, con una aplicación en producción se recomienda formalizar el nombre del mismo para que esté presente una mayor usabilidad de usuario.
 - Por último, para liberar carga del sistema, se puede realizar la ejecución de las predicciones por país semanalmente. Esto debido a que al momento de querer obtener la gráfica de la predicción por país está es generada, lo cual es un proceso tardado y a escala puede generar latencia del lado del servidor del API.
 - Se recomienda utilizar Amazon CLOUDFRONT ya que no necesita un servicio externo de seguridad. Al igual que transformar el API a Serverless que nos permite tener un aumento en rapidez. Por otro lado, es recomendable que se tengan dos servidores para el área del frontend en caso una página llegue a caer siempre se tenga el otro como backup.



- Se recomienda ampliar en el tema del análisis exploratorio y limpieza de datos para poder implementar un método de limpieza de los datos que ayude disminuir el error de las predicciones por medio de eliminar valores atípicos utilizando métodos como limpieza en base a la desviación estándar.
- Explorar en el futuro utilizar una herramienta que se especializa en proveer un tablero de datos con la posibilidad de incluir en ella el tablero generado por ArcGIS debido a que en este caso cualquier herramienta que unifique los dos tiene costo adicional.
- Realizar un análisis especializado de la situación actual de África y como se puede mejorar el sistema de salud del mismo.

X. BIBLIOGRAFÍA

Aprendizaje Automático y aplicaciones IMERL / FING / UdelAR, 2013. Estimación Y Selección De Variables En Grandes Dimensiones Regresión Ridge, GNN, Lasso, Elastic Net, SCAD .. [online] p.78. Available at: <http://www.iesta.edu.uy/wp-content/uploads/2014/05/CursoPosgrado_Aprendizaje_Automatico_SCastro_2013.pdf> [Accessed 10 June 2013].

Centers for Disease Control and Prevention (CDC). HIV Surveillance Report: Diagnoses of HIV Infection in the United States and Dependent Areas, 2017
<https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2017-vol-29.pdf>

Centers for Disease Control and Prevention (CDC). HIV in the United States: At A Glance.; 2017.
<https://www.cdc.gov/hiv/statistics/overview/ataglance.html>

Centers for Disease Control and Prevention (CDC). HIV in the United States: At A Glance.; 2017.
<https://www.cdc.gov/hiv/statistics/overview/ataglance.html>

Cost of the Ebola Epidemic Error processing SSI file. (2020). Retrieved 2 November 2020, from Mahanta, J., 2017. Introduction To Neural Networks, Advantages And Applications. [online] towardsdatascience. Available at: <https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207>> [Accessed 11 December 2020].

COVID-19: ¿quién está a un mayor riesgo para los síntomas graves?. (2020). Retrieved 2 November 2020, from mayo clinic

Crosier, S. (1999). ArcGIS 9. Redlands, CA: ESRI.

Epidemias y Pandemias(2020). Retrieved 2 November 2020, from
<https://policlinicametropolitana.org/informacion-de-salud/epidemias-y-pandemias/>

Evaluando el error en los modelos de regresión. (2021). Retrieved 1 October 2020, from <https://aprendeia.com/evaluando-el-error-en-los-modelos-de-regresion/#:~:text=Es%20el%20promedio%20de%20la,diferencia%20entre%205%20y%200.>

Mitchell, T. (1997). Machine Learning. New York: McGraw-Hill.

FAO - Noticias: La FAO alerta del riesgo de los murciélagos de la fruta en la epidemia de Ébola en África occidental. (2020). Retrieved 2 November 2020, from <http://www.fao.org/news/story/es/item/239215/icode/>

FAO - Noticias: La FAO alerta del riesgo de los murciélagos de la fruta en la epidemia de Ébola en África occidental. (2020). Retrieved 2 November 2020, from <http://www.fao.org/news/story/es/item/239215/icode/>

Free condoms in motel rooms may prevent AIDS: study | CBC News. (2021). Retrieved 5 January 2021, from <https://www.cbc.ca/news/technology/free-condoms-in-motel-rooms-may-prevent-aids-study-1.239920>

Free condoms in motel rooms may prevent AIDS: study | CBC News. (2021). Retrieved 5 January 2021, from <https://www.cbc.ca/news/technology/free-condoms-in-motel-rooms-may-prevent-aids-study-1.239920>

HIV/AIDS. (2020). Retrieved 30 November 2020, from <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>

HIV/AIDS. (2020). Retrieved 30 November 2020, from <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>, Centers for Disease Control and Prevention (CDC). HIV and Injection Drug Use.; 2017. <https://www.cdc.gov/hiv/risk/idu.html>

Ibm.com. 2020. What Are Neural Networks?. [online] Available at: <<https://www.ibm.com/cloud/learn/neural-networks>> [Accessed 12 December 2020].

Impact of COVID-19 on people's livelihoods, their health and our food systems. (2020). Retrieved 2 November 2020, from <https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people%27s-livelihoods-their-health-and-our-food-systems>

Information on the Survivability of the Ebola Virus in Medical Waste | Cleaning and Disinfecting | Clinicians | Ebola (Ebola Virus Disease) | CDC. (2020). Retrieved 13 December 2020, from <https://www.cdc.gov/vhf/ebola/clinicians/cleaning/ebola-virus-survivability.html#:~:text=Physical%20agents%20that%20can%20eradicate,%2DBeam%2C%20and%20Gamma%20Rays.&text=Heating%20to%2060%C2%B0C,boiling%20water%20for%20five%20minutes>.

Information vs data vs knowledge. (2021). Retrieved 20 January 2021, from <https://datarob.com/information-vs-data-vs-knowledge/#:~:text=Information%20is%20nothing%20but%20the,the%20meaningful%20context%20provides%20information>.

Information vs data vs knowledge. (2021). Retrieved 20 January 2021, from <https://datarob.com/information-vs-data-vs-knowledge/#:~:text=Information%20is%20nothing%20but%20the,the%20meaningful%20context%20provides%20information>.

Instituto Superior Tecnológico de Panuco. (2021). Minado de series de tiempo utilizando la metodología ARIMA (p. 11). México. Retrieved from <http://Instituto Superior Tecnológico de Panuco>

Korvin, R., 2020. Base64 Images Advantages & Disadvantages - Coderiddles. [online] CodeRiddles. Available at: <<http://www.coderiddles.com/base64-images/>> [Accessed 12 December 2013].

LONDOÑO, A. (2016). ALGORITMOS DE CLASIFICACION LINEAL PARA LA IDENTIFICACIÓN DE ZONAS CEREBRALES. In UNIVERSIDAD TECNOLÓGICA DE PEREIRA (p. 32). PEREIRA.

Medlineplus.gov. 2020. Medlineplus - Health Information From The National Library Of Medicine. [online] Available at: <<https://medlineplus.gov/>> [Accessed 22 April 2020].

Origin of HIV & AIDS. (2020). Retrieved 30 October 2019, from <https://www.avert.org/professionals/history-hiv->

[aids/origin#:~:text=HIV%20first%20began%20to%20spread,the%20Democratic%20Republic%20of%20Congo.](#)

Prabhakaran, S., 2020. ARIMA Model - Complete Guide To Time Series Forecasting In Python | ML+. [online] www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/ [Accessed 12 December 2020].

Qué es y para qué sirve el clustering - un ejemplo de aplicación práctico - Conceptos Claros. (2021). Retrieved 21 January 2021, from <https://conceptosclaros.com/que-es-clustering/>

Reddy, M. (2011). *API design for C++*. Burlington, MA: Morgan Kaufmann Publishers/Elsevier.

redhat. (2020). Retrieved 2 November 2020, from <https://www.redhat.com/es/topics/cloud-computing/what-are-cloud-services>

Smyth, P. (2020). *Creating Web APIs with Python and Flask*. Retrieved 2 November 2020, from <https://programminghistorian.org/en/lessons/creating-apis-with-python-and-flask>

Soriano Flores, A. (2019). *Regresión Lineal: Enfoque Bayesiano*. Retrieved 21 January 2021, from <http://www.dpve.iimas.unam.mx/soriano/BAYES/DOCUMENTOS/NOTAS/REGRESION%20LINEAL%20BAYESIANA.pdf>

The World's Largest HIV Epidemic in Crisis: HIV in South Africa. (2020). Retrieved 2 November 2020, from <https://www.csis.org/analysis/worlds-largest-hiv-epidemic-crisis-hiv-south-africa> (2020). Retrieved 2 November 2020, from https://projecteuclid.org/download/pdfview_1/euclid.ba/1339878896

University of Helsinki. (2017). *Epidemic as a natural process* (p. 6). Helsinki.

Vargas Domínguez, A. and Palacios Álvarez, V., 2018. *Educación Para La Salud* (2A. Ed.). Ciudad de México: Grupo Editorial Patria.

Vmq.tripod.com. 2020. *Ejemplos Didacticos De Redes Neuronales - Diinf*. [online] Available at: <http://vmq.tripod.com/> [Accessed 12 December 2020].

VU Amsterdam, 2020. *Feature Selection Using LASSO. Feature Selection using LASSO. [online]* Amsterdam: Dr. Eduard Belitser, p.26. Available at: <https://beta.vu.nl/nl/Images/werkstuk-fonti_tcm235-836234.pdf> [Accessed 30 March 2017].

Who.int. 2020. Home. [online] Available at: <<https://www.who.int/>> [Accessed 6 May 2020].

World Population Prospects - Population Division - United Nations. (2020). Retrieved 2 November 2020, from

[https://population.un.org/wpp/Download/Files/1_Indicators%20\(Standard\)/EXCEL_FILES/1_Population/WPP2019_POP_F13_C_OLD_AGE_DEPENDENCY_RATIO_2069.xlsx](https://population.un.org/wpp/Download/Files/1_Indicators%20(Standard)/EXCEL_FILES/1_Population/WPP2019_POP_F13_C_OLD_AGE_DEPENDENCY_RATIO_2069.xlsx)

¿Qué es Machine Learning? – Cleverdata. (2021). Retrieved 21 January 2021, from

<https://cleverdata.io/que-es-machine-learning-big-data/>