

Análisis De Datos Relacionados Con Covid19 En México

José Gabriel Rodríguez Rivas

Tecnológico Nacional de México. Instituto Tecnológico de Durango

gabriel.rodriguez@itdurango.edu.mx

Francisco Javier Favela Nájera

Tecnológico Nacional de México. Instituto Tecnológico de Durango

16041221@itdurango.edu.mx

Jorge Alberto Hernández Castro

Tecnológico Nacional de México. Instituto Tecnológico de Durango

15041220@itdurango.edu.mx

Resumen

La investigación que se realizó es en base a la información sobre COVID-19 en México por lo cual se utilizó los datos libres proporcionados por el gobierno de México por lo tanto se optó por el tema de analizar las enfermedades relacionadas con la tasa de infectados de COVID-19 así de igual manera los infectados por rango de edades, así como también los estados con mayor tasa de infectados y defunciones, como también el sector que ha registrado la mayor tasa de infectados atendidos, el objetivo de dicha investigación es mostrar que tiene una relación las enfermedades ya existentes con la posibilidad de ser más vulnerable de contraer COVID-19 de igual manera el mostrar cuales son los estados con mayor tasa de infección como defunciones, por lo tanto los resultados principales que se van a mostrar son algunas enfermedades que se cree sufren mayormente los infectados, los sectores con mayor tasa de infectados atendidos así como defunciones, los estados con mayor tasa de infección y defunciones, de igual manera se presentará un mapa regional donde se mostrará en base a colores el nivel de infectados por estado.

Palabras clave: Análisis de datos, Python, COVID19, Big Data, México

Abstract

The research carried out is based on information on COVID-19 in Mexico, which used the free data provided by the government of Mexico so opted for the topic of analyzing diseases related to the cup of COVID-19 infected as well as those infected by age range, as well as states with higher rate of infected and deaths , as well as the sector that has registered the highest rate of infected people served, the objective of this research is to show that it has a relationship with the existing diseases with the possibility of being more vulnerable to contract COVID-19 in the same way showing which are the states with the highest infection rate and deaths, therefore the main results that will be shown are some diseases that are believed to suffer mostly those infected , the sectors with the highest rate of infected cares as well as deaths, the states with the highest rate of infection and deaths, in the same way a regional map will be presented showing on the basis of colors the level of infected by state.

Keywords: Analysis of data, Python, COVID19, Big Data, Mexico

Introducción

¿Qué es 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 la COVID-19 es una pandemia que afecta a muchos países de todo el mundo.

Situación en México

Este martes 11 de agosto del 2020 se dieron a conocer las nuevas cifras oficiales de coronavirus en México, donde desgraciadamente ya hay 53,929 muertos a causa del Covid-19. Además, en toda la república mexicana se han confirmado 492,522 casos positivos, 6,686 de ellos en las últimas 24 horas. El número de defunciones en las últimas 24 horas es de 926.

Situación global

En la actualidad hay muchos países que se ven severamente afectados por la situación de la pandemia ya que en la mayoría de los países que se presentó dicho altercado ha llevado a que su economía de desplome siendo un gran

impacto, de igual manera el que sea tan fácil de ser contagiado hace que se propague de una forma muy rápida y se tenga que tomar las medidas necesarias para poder tratar de parar las infecciones.

De esta forma se pretende dar respuesta a la siguiente pregunta de investigación: ¿Hay enfermedades que te hagan más propenso a tener COVID-19?, ¿Cuáles son los estados con mayor contagios y defunciones?, ¿Cuáles son los sectores con mayor atención a infectados y cuál es su tasa de defunciones?, ¿Podemos ver en el mapa regional estados con mayor infección?

Objetivo general: Poder saber si hay algunas enfermedades que puedan hacer más propenso a contraer COVID-19, de igual manera cuales son los estados con mayor número de casos y con mayor número de defunciones.

Metodología

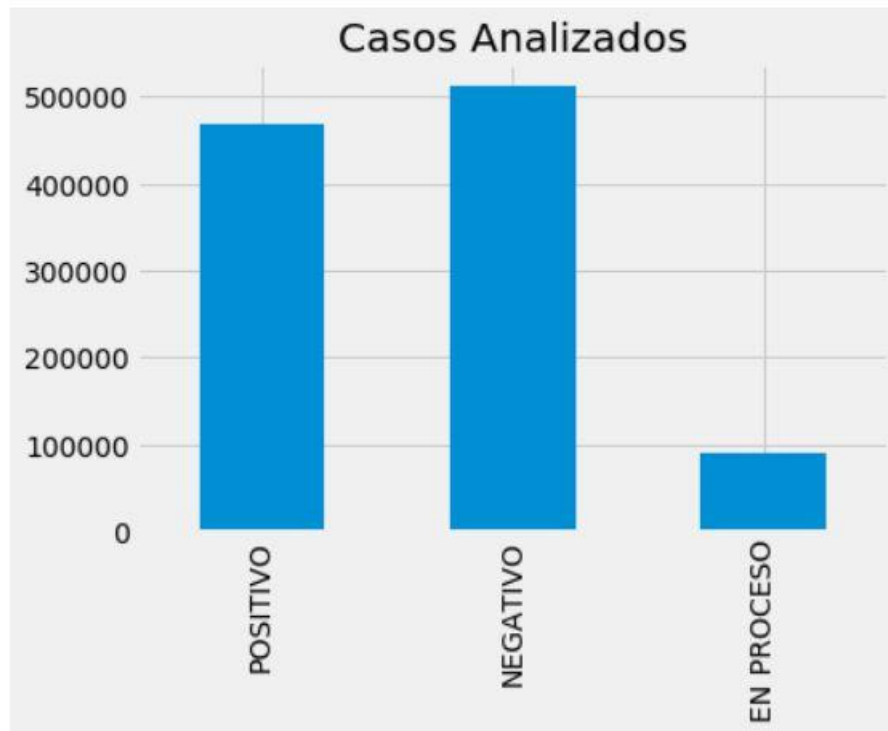
El presente estudio se desarrolló bajo un enfoque cuantitativo. En este enfoque se utiliza información cuantificable para describir o tratar de explicar los fenómenos que se estudian (Briones, 2002). El diseño de la investigación es no experimental y de tipo transversal que de acuerdo con Hernández, Fernández, y Baptista (2010) se utilizan para describir y analizar variables en un momento dado, y de tipo descriptivo para establecer la forma de distribución de una o más variables.

Para el análisis estadístico de la información se usó la plataforma Jupyter Notebook, la cual es una distribución libre y gratuita para análisis de grandes cantidades de datos con las librerías pandas, numpy, matplotlib y seaborn, usando el lenguaje de programación Python.

Para la recolección de los datos se utilizó la información disponible en la base de datos de la Dirección General de Epidemiología al día 2020-08-07, publicada en el portal de datos abiertos del gobierno de México referente a los casos asociados a COVID-19 en la dirección URL <https://www.gob.mx/salud/documentos/datos-abiertos-152127>. La base de datos contiene al momento 1071706 registros en un archivo de tipo csv (Archivo de texto separado por comas).

Resultados

Los datos totales de la información proporcionada 1071706 que representa el total de las personas que se han realizado la prueba del COVID-19 para lo cual los resultados se dividen en tres posibles opciones, NEGATIVO, POSITIVO y EN PROCESO

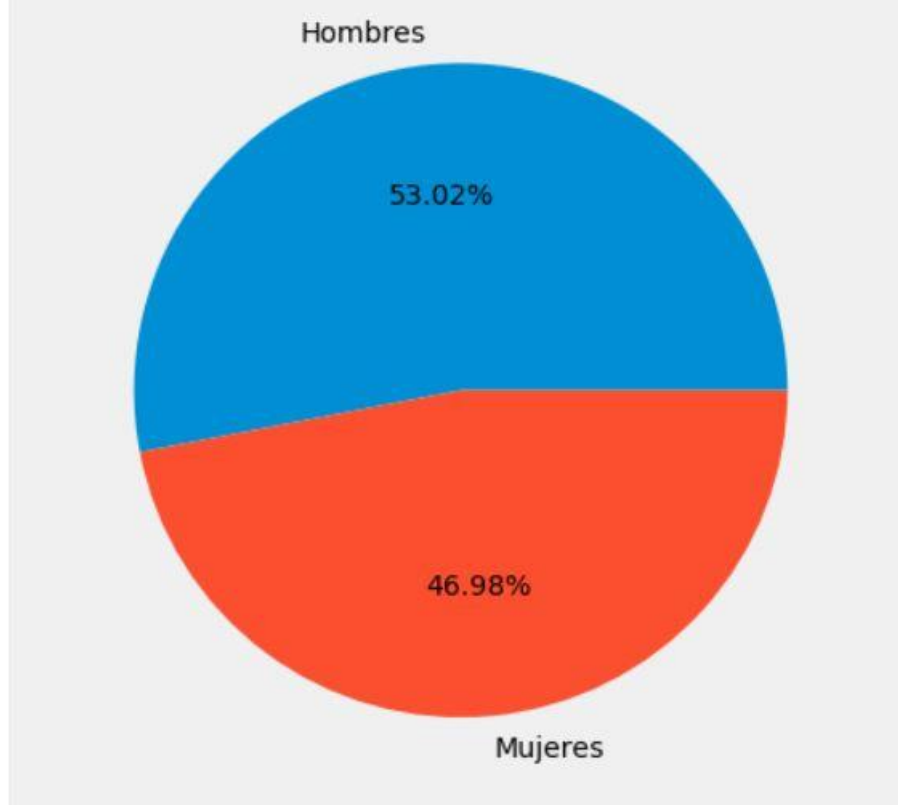


De todas las personas que se realizaron la prueba los resultados arrojaron que 469,407 dieron positivo mientras que el 513,144 dieron negativo a la prueba y aún están en proceso de resultados 89,155.

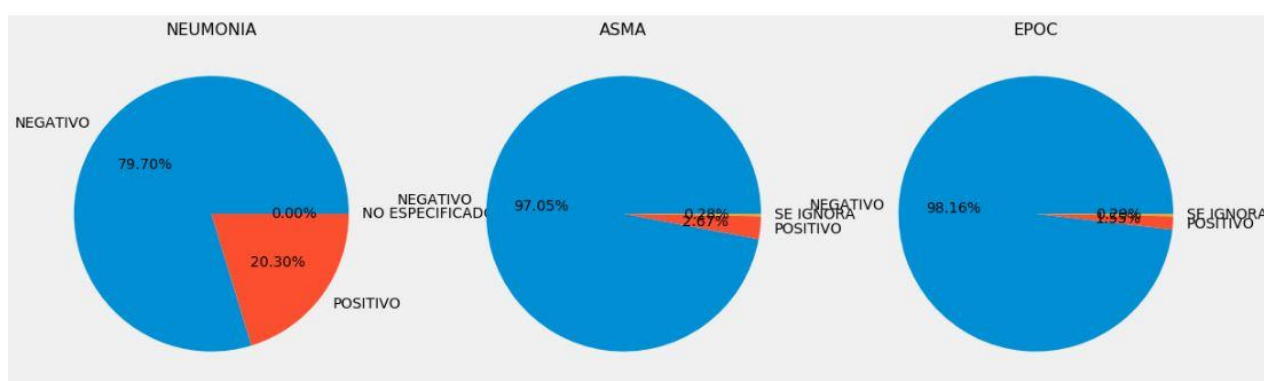
De estos datos analizados nos centraremos en los casos positivos ya que estos son los que nos ayudaran a la investigación.

De los casos positivos es necesario saber cuántos son hombre y cuantos son mujeres con fines estadísticos, para lo cual se observó que 248,893 lo que equivale a 53.02% de los infectados son hombre mientras que 220,514 son mujeres lo que equivale a 46.98%.

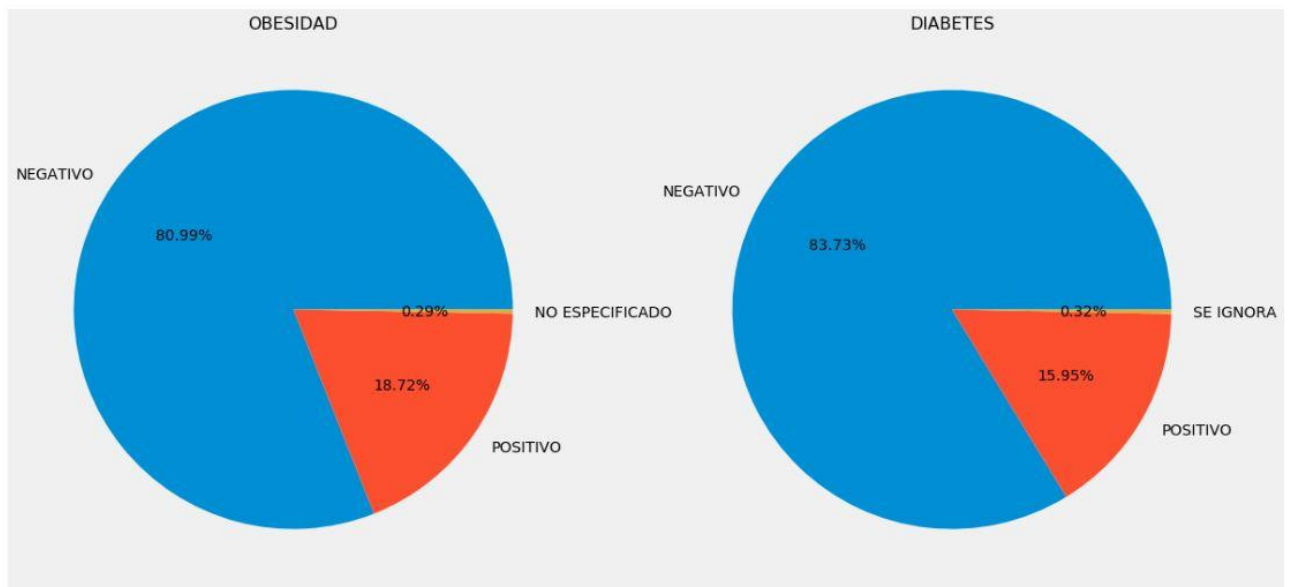
Porcentaje de hombres y mujeres positivos



Una vez recabado los datos, vamos a seleccionar tres enfermedades respiratorias que hayan salido positivo para resultado de COVID-19, las enfermedades seleccionadas son neumonía, asma y epoc.

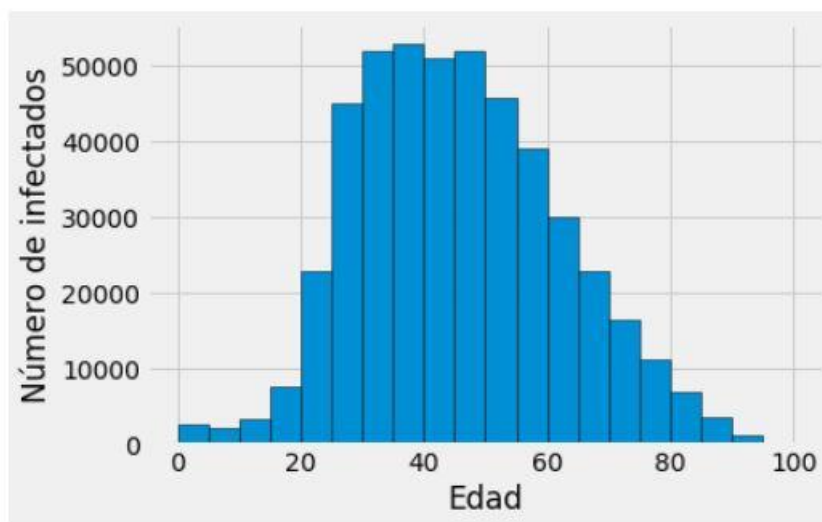


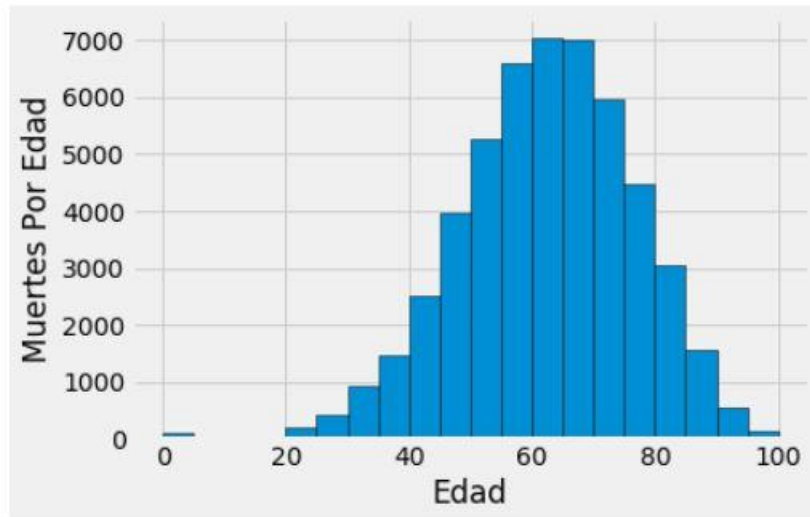
Lo cual muestra que el total de los casos positivos a COVID-19 solo el 20.30% muestra neumonía, solo el 2.67% presentan asma y el 1.55% presenta epoc También se pensó realizar la misma prueba para diabetes y obesidad lo cual mostro que 87,864 tienen obesidad y que 74,881 padecen de diabetes marcando un porcentaje de obesidad 18.72% y para 15.95%



Infectados y muertes por edad

Para la investigación de muertes e infectados debemos saber cuáles son los rangos de edades con COVID-19 a lo cual se muestra la siguiente tabla que muestra el número de infectados por edades.

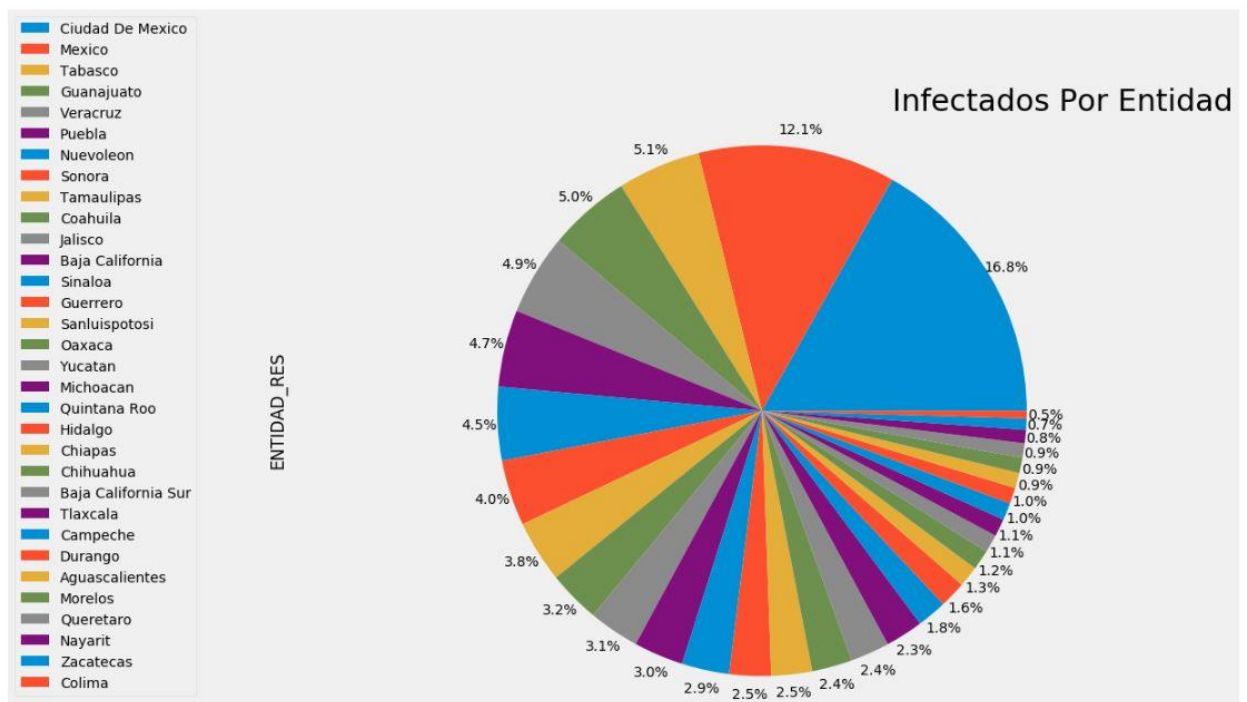




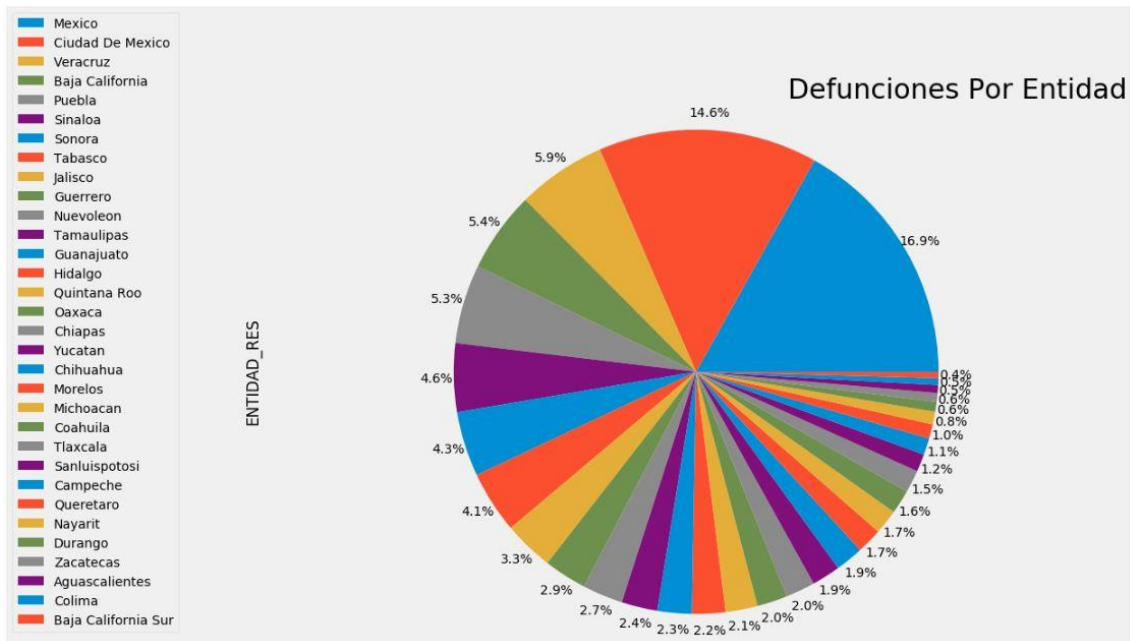
Dentro de la muestra de rangos de edad con COVID-19 se sacó una gráfica para poder saber cuáles son los rangos de edad con mayor cantidad de defunciones.

% de Entidades con contagios

Para la investigación sobre las entidades se requiere obtener el porcentaje de infectados por entidades lo cual se representa en la siguiente grafica de pastel



En la siguiente barra se muestra los porcentajes de defunción por entidad lo cual se muestra en la siguiente grafica.



Sectores de salud

En esta parte de la investigación se busca analizar el sector salud por lo tanto se analizaron cada uno de los sectores lo cual se observa que los primeros tres lugares los ocupa el ISSTE el IMSS y el SSA.



Una vez obtenido el número de infectados se van a filtrar los casos con defunción en cada uno de los sectores y se obtiene la siguiente gráfica, lo cual se puede observar que los tres primeros siguen siguiendo ISSTE, SSA e IMSS siendo el IMSS el que registra la mayor cantidad de muertes a pesar de que tiene menos infectados que el sector SSA.

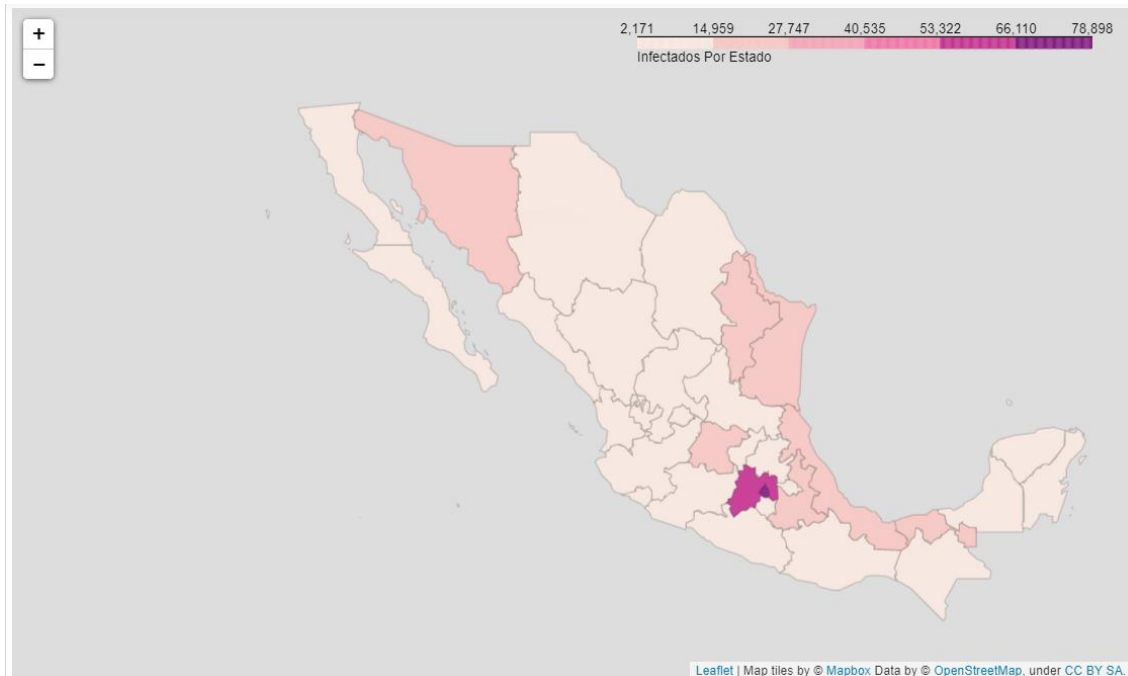


Mapas

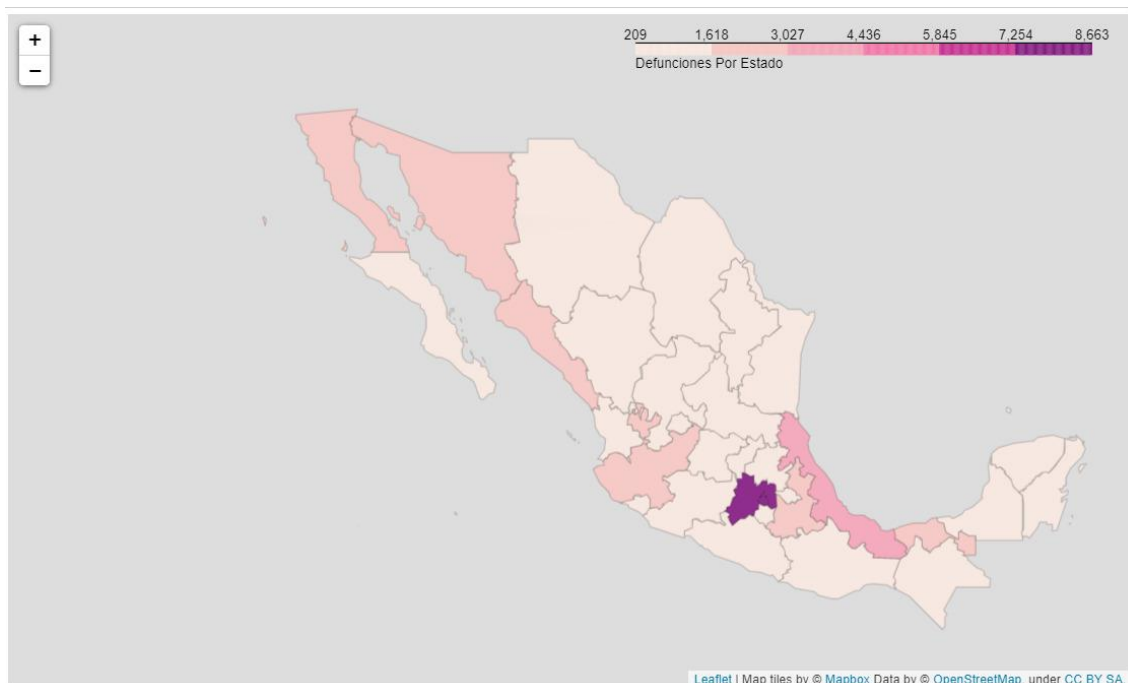
Se crearon mapas a partir de geo coordenadas usando la librería folium para poder visualizar de mejor manera los datos por entidad y tener una mejor comprensión de la magnitud de infecciones y decesos,



En este mapa consultamos a todos los infectados por estado y los agrupamos usando la función Choropleth de folium resaltamos los estados con mayor índice de infectados mediante una gama de colores que representan la densidad de infectados.

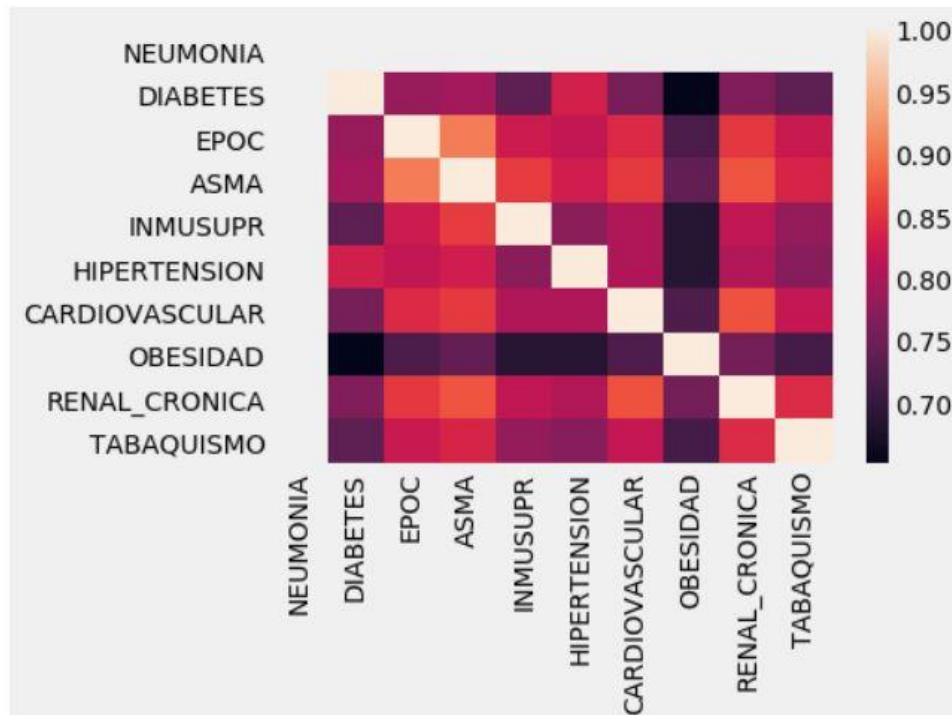


En este mapa consultamos a todos los decesos por estado y los agrupamos usando la función Choropleth de folium resaltamos los estados con mayor índice de decesos mediante una gama de colores que representan la densidad de infectados.

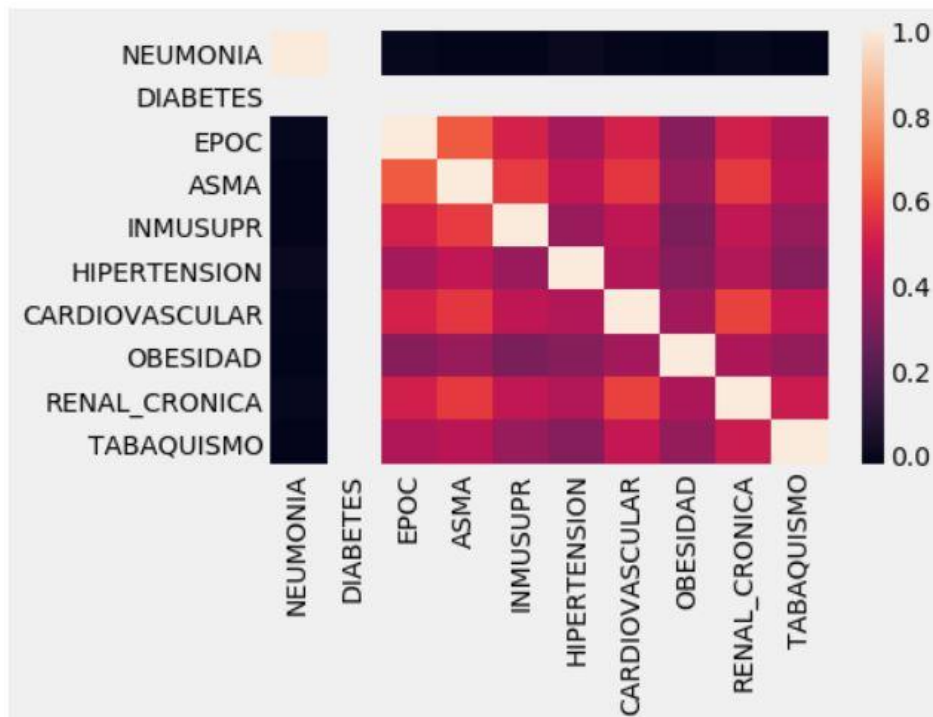


Correlación

En estos mapas de calor podemos ver la correlación que se da entre neumonía y diferentes enfermedades y su correlación con el covid como vimos en las gráficas anteriores la neumonía es un factor a la hora de contagiarse.



En este otro podemos ver la correlación que se da entre diabetes y las diferentes enfermedades.



Conclusiones

Una vez realizado la investigación podemos observar que las enfermedades existentes como lo son la neumonía, el asma y el epoc, no son factores concluyentes para poder determinar que estas enfermedades hacen a la población más propensa a poder contraer COVID-19, sin embargo, hay que resaltar que un 20.30% de las personas detectadas positivas tienen neumonía.

Con respecto a la situación que se maneja en México se decidió hacer un análisis con respecto a los padecimientos de obesidad y diabetes pudimos observar que el 18.72% de la población analizada que dieron positivo al COVID-19 padecen de obesidad mientras que el 15.75% padecen de diabetes siendo un porcentaje algo alto, pero lo que pudimos concluir respecto a estos porcentajes de acuerdo a dichos padecimientos es que en México gran parte de la población padecen obesidad o diabetes a lo cual podría ser mera coincidencia y no influyen en gran medida el que sea más susceptible a contagiarse si es que tienes estos padecimientos.

En el análisis de infecciones y defunciones por rangos de edad pudimos observar que el pico de persona infectadas va desde los 25 años hasta los 55 años aproximadamente, lo cual podemos observar que la mayoría de las personas son las que están en el ámbito laboral lo cual los hace que no puedan cumplir la cuarentena completamente ya que la mayoría tiene que asistir al trabajo de manera presencial, con respecto a las defunciones se observa que el pico va desde los 60 años de edad hasta los 80, lo cual se llegó a la conclusión que son personas de la tercera edad lo cual los hace más vulnerables por que el sistema inmunológico es menos efectivo.

Una vez que llegamos a las conclusiones sobre las enfermedades y las edades con más infecciones con defunciones, pensamos el poder analizar los estados con mayor número de infectados y el número de defunciones, lo que pudimos observar en el número de infectados es que los tres primeros lugares resultaron ser Ciudad de México, México y Tabasco, para el número de muertes por estado de observo que casi se conservan los mismos estados que en la tabla de infectados solo con un pequeño cambio siendo Ciudad de México, México y Veracruz con la mayor cantidad de infectados, se llegó a la conclusión que estos estados pudieran estar dentro de los primeros lugares en infectados y defunciones porque estos tienen mayor cantidad de población que en otros estados lo cual si se analiza la muestra de número total de población y el número de infectados se podría apreciar que no tienen un % estadístico tan alto como lo muestran las gráficas.

Se desea también analizar los sectores para saber cuáles son los que reciben al mayor número de infectados así como cuales representan el mayor número de defunciones lo que se pudo observar que el SSA, IMSS y el ISSSTE son los que ocupan los primeros lugares con la atención a personas infectadas mientras que, Cruz Roja, DIF, Universitario, se llegó a la conclusión que estos últimos no son hospitales aptos para poder atender estos casos ya que en estos últimos son los que dan primeros auxilios y después son enviados para otra unidad de cuidados que este mejor preparada.

Cuando se realizó el análisis sobre el número de defunciones se puede ver que el ISSSTE supero al SSA en dicha cifra siendo que el SSA atiende a un mayor número de personas infectadas, se llegó a la conclusión de esta tabla que la mayoría de los trabajadores en México pertenecen a este sistema de salud y puede llegar a saturarlo lo que puede influir en que se pueda prestar un tratamiento totalmente seguro para los infectados.

Ya que se realizó toda la investigación con respecto al número de infectados y de defunciones se realizó dos mapas que representan dichos casos, en los cuales en el primero podemos observar que el número de infectados por estado como referencia a la intensidad del color se puede observar que Ciudad de México es de los estados con mayor número de infectados en relación a los demás estados, también se realizó el mismo tipo de mapa pero son la información sobre las defunciones lo cual hace que se aprecie los estados con mayor número de defunciones, cuando se llega a comparar los dos mapas se puede ver que incluso estados que llegan a tener un número de infectados alto el número de defunciones es bajo, al igual que en los estados que se tiene el número de infectados bajos tiene un mayor número de defunciones, esto puede indicar que hay estados donde la atención a personas infectadas puede ser mejor que en otros o incluso que se les detecta de una manera pronta para poder tomar los cuidados necesarios.

Para concluir la investigación se decidió hacer una correlación entre dos enfermedades con respecto al COVID-19 lo cual la primera correlación se mostró que la neumonía tiene una correlación alta ya que este es un síntoma que presenta el COVID-19, también se analizó la diabetes ya que se observó en el análisis de este padecimiento que llevan un gran porcentaje de las personas infectadas y lo que mostro una correlación baja ya que este padecimiento es porque en México gran parte de la población la sufre.

Referencias

- Capistrán, M. C. (23 de 09 de 2019). *json-estados-municipios-mexico*. Obtenido de <https://github.com/martinciscap/json-estados-municipios-mexico>
- Lo, C. (07 de 05 de 2020). *Using Folium to Generate Choropleth Map with Customised Tooltips (Python)*. Obtenido de <https://towardsdatascience.com/using-folium-to-generate-choropleth-map-with-customised-tooltips-12e4cec42af2>
- Mexico, G. D. (08 de 08 de 2020). Obtenido de <https://coronavirus.gob.mx/datos/>
- Claro, R. M. (2020, 12 agosto). Coronavirus México 11 de agosto; resumen de las últimas noticias, contagios y muertes. MARCA Claro México. <https://www.marca.com/claro-mx/trending/coronavirus/2020/08/11/5f329cd2ca4741940b8b4590.html>
- Briones, G. (2002). Metodología de la Investigacion Cuantitativa en las Ciencias Sociales. ARFO editores e impresores Ltda.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. (2010). *Metodología de la Investigación*. McGraw-Hill.
- Aparicio, L. (2020, 11 agosto). El bono social para autónomos, en riesgo a partir de septiembre. Cinco Días. https://cincodias.elpais.com/cincodias/2020/04/09/mercados/1586465563_966437.html