

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE

Departamento de Matemáticas y Física

Sustentabilidad y tecnología

PROYECTO DE APLICACIÓN PROFESIONAL (PAP)

PAP Programa de Modelación Matemática para el Desarrollo de Planes y
Proyectos de negocio



ITESO, Universidad
Jesuita de Guadalajara

Modelos de predicción en empresas y gobierno mediante aprendizaje estadístico

4J07A Analítica avanzada para COVID-19 para
el Estado de Jalisco

PRESENTAN

Organización: Coordinación General de Inteligencia Gubernamental

Estudiantes:

Landaverde Mercado, Ivette

Navarro Silva, Jose Tonatiuh

Zuñiga Ortiz, Jorge Luis

Profesor: Montoya Escobar, Diana Paola

Tlaquepaque, Jalisco, mayo de 2022

ÍNDICE

Contenido

REPORTE PAP	2
Presentación Institucional de los Proyectos de Aplicación Profesional ¡Error! Marcador no definido.	
Resumen	¡Error! Marcador no definido.
1. Ciclo participativo del Proyecto de Aplicación Profesional	2
1.1 Entendimiento del ámbito y del contexto	5
1.2 Caracterización de la organización	5
1.3 Identificación de la(s) problemática(s).....	6
1.4. Planeación de alternativa(s)	6
1.5. Desarrollo de la propuesta de mejora.....	8
1.6. Valoración de productos, resultados e impactos	51
1.7. Bibliografía y otros recursos	51
1.8. Anexos generales	¡Error! Marcador no definido.
2. Productos	52
3. Reflexión crítica y ética de la experiencia	53
3.1 Sensibilización ante las realidades	53
3.2 Aprendizajes logrados	55

REPORTE PAP

Los Proyectos de Aplicación Profesional (PAP) son experiencias socio-profesionales de los alumnos que desde el currículo de su formación universitaria- enfrentan retos, resuelven problemas o innovan una necesidad sociotécnica del entorno, en vinculación (colaboración) (co-participación) con grupos, instituciones, organizaciones o comunidades, en escenarios reales donde comparten saberes.

El PAP, como espacio curricular de formación vinculada, ha logrado integrar el Servicio Social (acorde con las Orientaciones Fundamentales del ITESO), los requisitos de dar cuenta de los saberes y del saber aplicar los mismos al culminar la formación profesional (Opción Terminal), mediante la realización de proyectos profesionales de cara a las necesidades y retos del entorno (Aplicación Profesional).

El PAP es un proceso acotado en el tiempo en que los estudiantes, los beneficiarios externos y los profesores se asocian colaborativamente y en red, en un proyecto, e incursionan en un mundo social, como actores que enfrentan verdaderos problemas y desafíos traducibles en demandas pertinentes y socialmente relevantes. Frente a éstas transfieren experiencia de sus saberes profesionales y demuestran que saben hacer, innovar, co-crear o transformar en distintos campos sociales.

El PAP trata de sembrar en los estudiantes una disposición permanente de encargarse de la realidad con una actitud comprometida y ética frente a las disimetrías sociales. En otras palabras, se trata del reto de “saber y aprender a transformar”.

El Reporte PAP consta de tres componentes:

El primer componente refiere al ciclo participativo del PAP, en donde se documentan las diferentes fases del proyecto y las actividades que tuvieron lugar durante el desarrollo de este y la valoración de las incidencias en el entorno.

El segundo componente presenta los productos elaborados de acuerdo con su tipología.

El tercer componente es la reflexión crítica y ética de la experiencia, el reconocimiento de las competencias y los aprendizajes profesionales que el estudiante desarrolló en el transcurso de su labor.

RESUMEN

El propósito de este reporte es reflejar el análisis y trabajo dedicado a lo largo del semestre al proyecto de Analítica avanzada para COVID-19 en el estado de Jalisco el cual se lleva a cabo de la mano con el departamento de Matemáticas y Física del ITESO y la Coordinación General de Inteligencia Gubernamental, está dividido en secciones comenzando con la contextualización de la información y datos a trabajar, seguido por una limpieza profunda de los datos, análisis y modelado.

El proyecto fue llevado a cabo en su totalidad a través de un escenario real para que los estudiantes tuvieran una experiencia concreta sobre la vida profesional en el entorno que se desenvuelven además de acreditar su servicio social.

Los datos trabajados son del portal SINAVE (Sistema Nacional de Vigilancia Epidemiológica) los cuales contienen toda la información acerca del COVID-19 y su desarrollo a través del tiempo en el país, contabilizando por fecha desde casos activos hasta defunciones por lo cual como participantes del proyecto se usan con el fin de tomar mejores decisiones en estos tiempos de incertidumbre.

A través de conocimientos de analítica avanzada, modelado y programación se realizan análisis y visualizaciones para hacer que estas decisiones sean más claras y que sea posible realizar pronósticos y proyecciones a partir de la tendencia de información pasada.

A pesar de que se trabaja con información nacional también es posible clasificarla por estados, hacer comparativos y hacer énfasis especial en el estado de Jalisco.

Propósitos educativos

Propósito general de la asignatura:

Desarrollar herramientas de visualización y análisis de datos. Emplear técnicas de machine learning para realizar predicciones, clusters y análisis para toma de decisiones.

Propósitos transversales de la asignatura:

1. Desarrollar capacidad para el trabajo en equipo.
2. Desarrollar habilidades de lectura y escritura en lenguaje matemático y ordinario.
3. Reunir y examinar información de diversas fuentes bibliográficas.
4. Manejar tecnologías de información como herramientas auxiliares en la solución de problemas.

Propósitos Específicos de la asignatura:

Vincular de forma integral, transferir y aplicar los conocimientos y habilidades que se ha adquirido y desarrollado a lo largo de la carrera para proponer soluciones en materia de gestión de riesgos de tipo financiero.

1. Ciclo participativo del Proyecto de Aplicación Profesional

El PAP es una experiencia de aprendizaje y de contribución social integrada por estudiantes, profesores, actores sociales y responsables de las organizaciones, que de manera colaborativa construye sus conocimientos para dar respuestas a problemáticas de un contexto específico y en un tiempo delimitado. Por tanto, la experiencia PAP supone un proceso en lógica de proyecto, así como de un estilo de trabajo participativo y recíproco entre los involucrados.

El Programa de Modelado Matemática para el Desarrollo de Planes y Proyectos de negocio tiene una duración de 21 semanas las cuales están distribuidas de la siguiente manera:

Sesión	Etapas	Actividad con docente	Actividad alumnos	Indicador de evaluación
1	Aprendizaje	Introducción al PAP, planeación del proyecto, división de equipos.		Generación de grupos de trabajo
2	Contextualización	Conociendo la base de datos, dónde encontrar la información de la base de datos y manual de descarga de datos.	Descargar bases de datos y conocer qué vamos a explorar.	Investigación de qué contienen las bases de datos y que significa cada columna.
3	Contextualización	Sesión con Eduardo Ulises Moya, director de Inteligencia Artificial	Contextualizar: Situación del COVID-19.	INVESTIGACIÓN Foro sobre qué está pasando
4	Aprendizaje/Aplicación	Introducción a manejo de bases de datos con R, Python o el software que vamos a utilizar.		Comenzar a manipular bases de datos pequeñas de los datos.
5	Aprendizaje/Aplicación	Análisis de texto en R. Conexión con Rtweet para analizar análisis de texto.	Conocer la paquetería de Rtweet y las bases de datos que se pueden descargar sobre lo que la gente comenta en twitter sobre la	Realizar análisis de texto

			situación actual de COVID.	
6	Avances de Proyecto	En esta etapa debe de ir en marcha la conexión a la base de datos, se muestran avances y resuelven dudas.		Avances de la conexión a bases de datos.
7	Aprendizaje/Aplicación	Sesión de conceptos básicos de series de tiempo y realización de código de pronóstico en R.	Conocer la paquetería de R para realizar pronósticos por medio de descomposición.	Realización de pronóstico y descomposición de series de tiempo.
8	Avances de Proyecto		Cada equipo debe mostrar el avance que lleva de la primera y segunda semana según la planeación del proyecto.	Actualizar diagrama de Gantt del proyecto según los avances de esta semana. - Cada equipo entrega un el resultado de la primera semana.
9	Aprendizaje/Aplicación	Introducción al manejo de software de análisis de datos Tableau	Análisis de datos en tableau y generación de Dashboards, conexión a bases de datos, elaboración de tablas, gráficos y pronósticos.	Elaboración de un primer Dashboard en Tableau para el análisis
10	Avances y preguntas	Sesión Eduardo Ulises Moya		Realizar los respectivos cambios dados por el feedback por parte de.
11	Avances de Proyecto		Cada equipo debe mostrar el avance que lleva de la tercera semana según la planeación del proyecto.	Actualizar diagrama de Gantt del proyecto según los avances de esta semana. - Cada equipo entrega un el resultado de la semana.
12	Avances de Proyecto			
13	Dudas	En esta sesión se resolverán las dudas que se tengan para poder		Resolución de dudas para continuar con el

		completar a tiempo la semana 3.		proyecto. Participación de todos los equipos de trabajo
14	Dudas	En esta sesión se resolverán las dudas que se tengan para poder completar a tiempo la semana 3.		
15	Avances de Proyecto		Cada equipo debe mostrar el avance que lleva de las cuatro semanas según la planeación del proyecto.	Actualizar diagrama de Gantt del proyecto según los avances de esta semana. - Cada equipo entrega un el resultado de la semana.
16	Avances de Proyecto			
17	Análisis de los datos	Realizar modelos y mostrar correlaciones. Lista de conclusiones e insights.		Lista de insights que nos ayudarán a obtener las conclusiones del proyecto.
18	Avances y preguntas			Realizar los respectivos cambios dados por el feedback.
19	Toques finales del proyecto	Ajustes del proyecto		
20	Reporte PAP	Exposición de cada equipo de lo realizado en el proyecto.	Cada equipo realizará una exposición de media hora sobre lo más relevante realizado en este PAP.	Exposición y repote PAP
21	Exposición PAP	Exposición final del proyecto a personas de la gobernación y profesores del ITESO.		Se tomará en cuenta las respuestas que den ante las preguntas generadas por el público de esta exposición.

1.1 Entendimiento del ámbito y del contexto

En diciembre del año 2019 apareció en China un nuevo virus denominado Coronavirus el cual es causante de la enfermedad COVID-19 mismo que tiene un alto grado de propagación por lo que la humanidad tuvo que entrar en un periodo de cuarentena, esta pandemia trajo consecuencias económicas y sociales por lo que el estudio de su comportamiento y propagación es clave para la toma de decisiones tanto de la Secretaría de Salud como otras dependencias en México y el mundo.

El proyecto nace con la finalidad de hacer un análisis exhaustivo de los datos relacionados a COVID-19 en México con visualizaciones y machine learning para tener participación en las decisiones de la Coordinación Nacional de Inteligencia Gubernamental y al mismo tiempo aplicar los conocimientos adquiridos a lo largo de los semestres cursados en la carrera de Ingeniería Financiera los cuales constan en habilidades técnicas y analíticas como ciencia de datos, estadística, programación orientada a objetos, entre otros.

1.2 Caracterización de la organización

El proyecto pertenece al departamento de Matemáticas y Físicas y la organización encargada es la Coordinación General de Inteligencia Gubernamental a través de la Doctora Diana Paola Montoya Escobar que es la directora de Representaciones Visuales de la Coordinación General de Análisis Estratégico en Gobierno del Estado de Jalisco.

También cuenta con la participación de Eduardo Ulises Moya como director de Inteligencia Artificial.

El grupo del PAP estuvo conformado a su vez por 10 alumnos de Ingeniería Financiera los cuales se dividen en 3 equipos.

Este equipo en particular está conformado por los alumnos Ivette Landaverde Mercado, José Tonatiuh Navarro Silva y Jorge Luis Zúñiga Ortiz.

1.3 Identificación de la(s) problemática(s)

Tras más de dos años en pandemia el covid-19 ha tenido grandes repercusiones en el país, desde rezago social y cultural hasta efectos económicos millonarios y un rotundo cambio en el estilo de vida de los mexicanos, ya que entre las estrategias implementadas ha sido la suspensión de clases presenciales y trabajos que se puedan desarrollar de manera remota, disminuir el aforo de lugares públicos o incluso hubo temporadas en las que no era posible realizar nada en espacios concurridos a menos que fuera de vital importancia. Es por esto que es primordial analizar la situación en México para tomar mejores decisiones tanto a nivel estatal como nacional.

Desde la semana 1 que fue dedicada al entendimiento de los datos y el contexto, como equipo desarrollamos algunas preguntas con el fin de tener una pauta o dirección ya que durante el análisis buscaríamos responderlas y de no ser así, cambiarlas conforme los datos lo indiquen. Las preguntas son las siguientes:

- ¿Qué porcentaje de los decesos tienen un esquema de vacunación completo?
- ¿Qué porcentaje de personas intubadas murieron?
- ¿Qué tan letal ha sido el virus con menores de edad?
- ¿Cuál fue el estado de México con menos contagios?
- ¿Qué medidas se tomaron en este estado con menos contagios?
- ¿Cuál es la tasa de mortalidad en Jalisco del COVID?
- ¿Cuál es la tasa de reproducción del COVID en México?
- ¿Cuáles son los factores de riesgo más comunes en las personas hospitalizadas?
- ¿Cuál ha sido el efecto de las jornadas de vacunación en las hospitalizaciones?
- ¿Qué tan viable es el regreso a clases y oficinas? ¿Cómo afectará la nueva variante la tendencia que se llevaba de contagios?
- ¿Qué porcentaje de infectados fueron realmente atendidos por una institución formal?
- ¿Qué entidades son más vulnerables y por qué?
- ¿Qué porcentaje de la población ha sido la más afectada? ¿Qué estados tienen menos disponibilidad en hospitales y de personal de servicio?

1.4. Planeación de alternativa(s)

Cronograma:

Además del plan de trabajo anexo en la sección 1 contamos con un cronograma de trabajo el cual divide el proyecto en cuatro bloques los cuales son:

- **Bloque 1: Entender el problema y los datos**

Entender el problema de COVID, principales preguntas que están surgiendo.
Entender los datos de SINAVE, limpieza y transformación de los datos

- **Bloque 2: Análisis de los datos**

Análisis general, comparativa entre estados, análisis entre olas de contagio, análisis de casos hospitalizados, razones de cambio y velocidades, análisis por grupos de edad, análisis de texto, análisis extras.

- **Bloque 3: Modelados**

Pronóstico de casos confirmados, hospitalizados
Modelado de covid (regresión y/o clasificación)

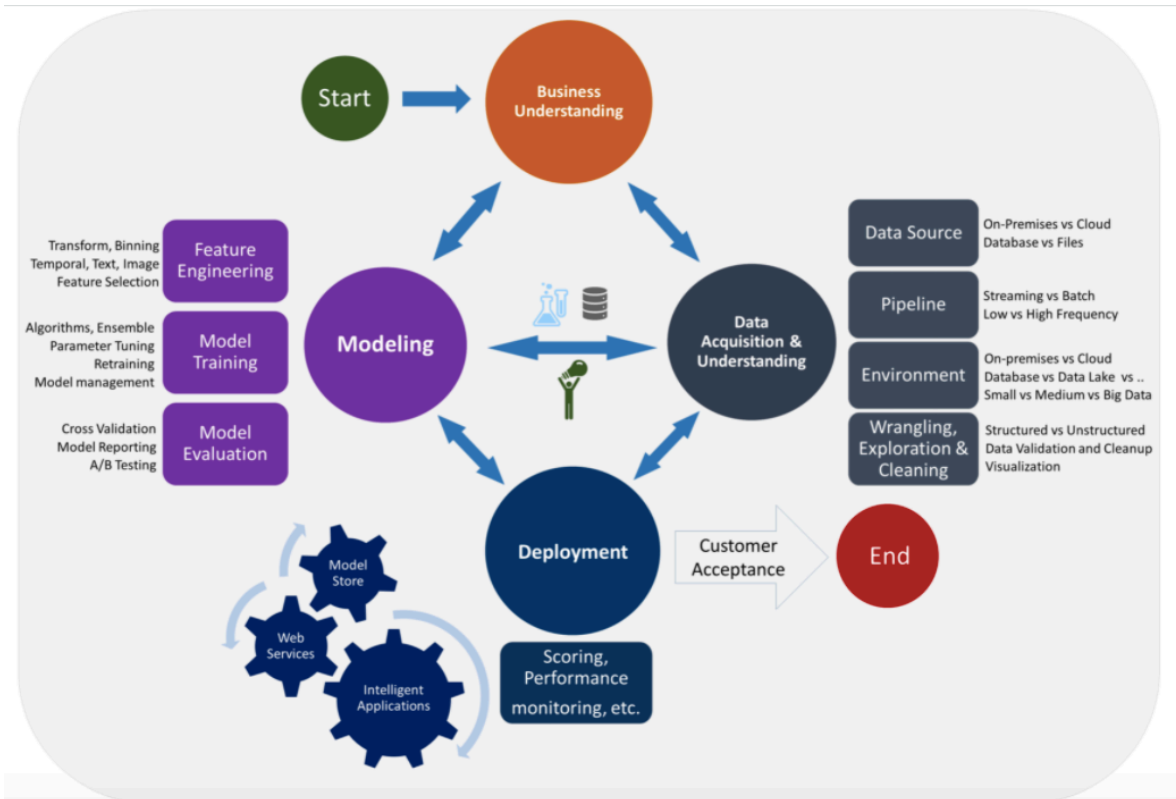
- **Bloque 4: Entrega del proyecto**

Reporte y presentación PAP

Metodología empleada:

La metodología empleada en el proyecto PAP fue TDSP (Team Data Science Process) que es iterativa y ofrece soluciones de análisis predictivo eficientes. Además, al ser un proyecto en equipo esta metodología facilita la asignación de roles ya que sugiere como funcionaría mejor.

El TDSP es un ciclo de vida de un proyecto de ciencia de datos y marca los pasos a seguir en orden. La siguiente imagen es una representación visual de las principales etapas que este tipo de proyectos suelen ejecutar.



Siguiendo esta metodología se espera conseguir visualizaciones y modelos predictivos que ayuden a una toma de decisiones más asertiva y a poder comparar de manera correcta y normalizada los estados en distintas categorías para saber que alternativas empleadas por cada estado están resolviendo mejor las necesidades de la población.

1.5. Desarrollo de la propuesta de mejora

Link Gitlab:

<https://gitlab.com/dpmontoya/pap-mmd-p2022>

Análisis exploratorio de los datos:

Como equipo trabajamos con los datos abiertos de SINAVE y antes de hacer cualquier otra modificación o modelar los datos nos enfocamos en realizar un Análisis exploratorio a través de una limpieza y filtrado de información, enfocándonos en la población con nacionalidad mexicana.

Nos aseguramos de que todos los registros con los que trabajamos fueran de personas mexicanas y que no se repitieran datos (duplicados), también nos cercioramos que no existieran valores nulos presentes.

Se realizó un diccionario para almacenar la información ya que las variables estaban de manera numérica y SINAVE cuenta con un documento excel con los significados de esta numerología.

Normalización:

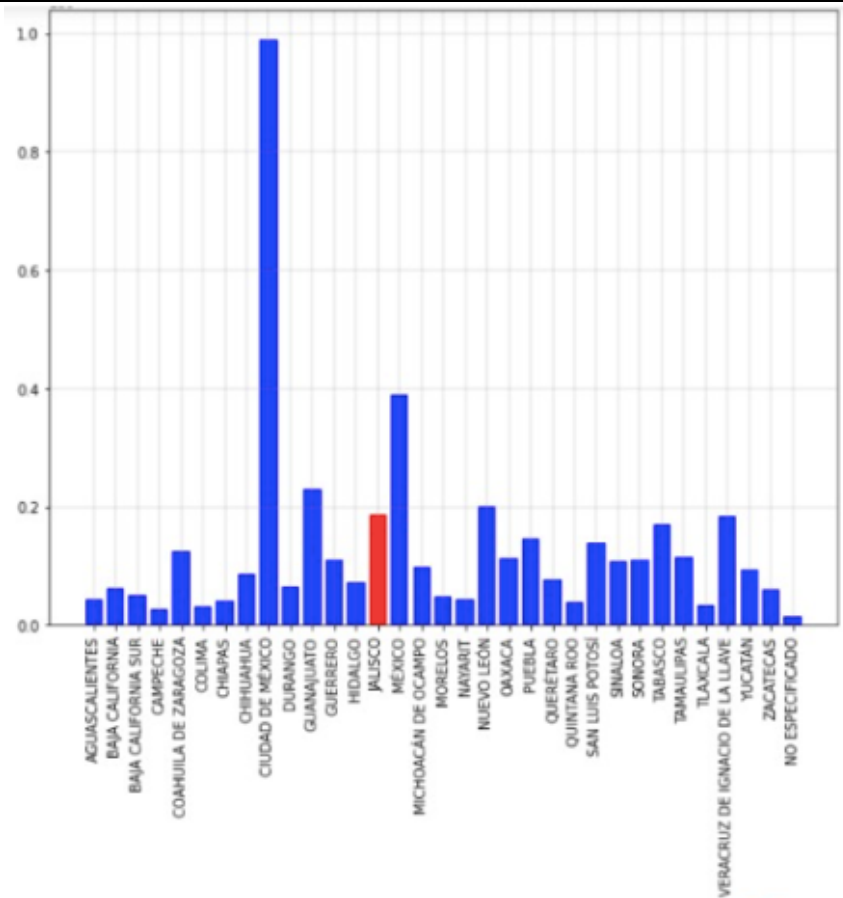
Ya que la meta es hacer comparaciones por estados y no se podía comparar directamente el número de contagiados, defunciones, etc de un estado a otro ya que estos no cuentan con el mismo número de población normalizamos los datos de cada estado, dividiendo las variables entre el número de habitantes por su respectivo estado y multiplicando por 10,000 de esta manera podemos saber el número de cada variable por estado de cada 10,000 habitantes.

Ejemplo: cuantos habitantes son positivos en Zacatecas de cada 10,000 y cuantos en Jalisco, de esa manera si es posible comparar porque están en la misma escala y se tiene en cuenta sus habitantes.

Visualizaciones:

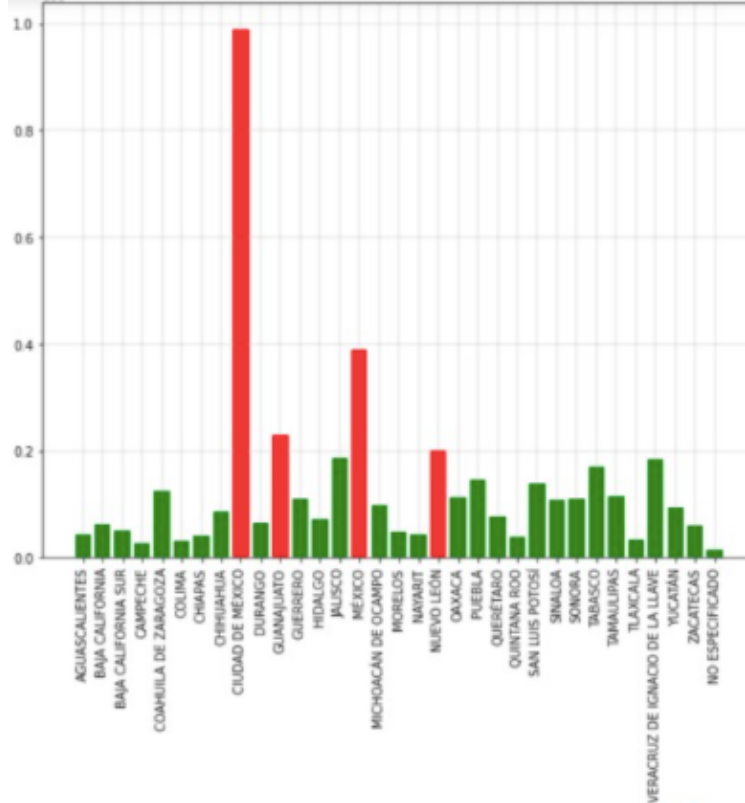
Ya con los datos limpios y normalizados es posible responder a algunas de las preguntas desarrolladas al inicio del proyecto a través de visualizaciones.

¿Qué tanto ha afectado el COVID a Jalisco en comparación de otras entidades y en relación a su población?



Como se puede observar en la gráfica, Jalisco no es de los estados más afectados pero tampoco de los que tienen menos contagios, sin embargo, sí se encuentra por encima del promedio.

Entidades con más casos de COVID que Jalisco:

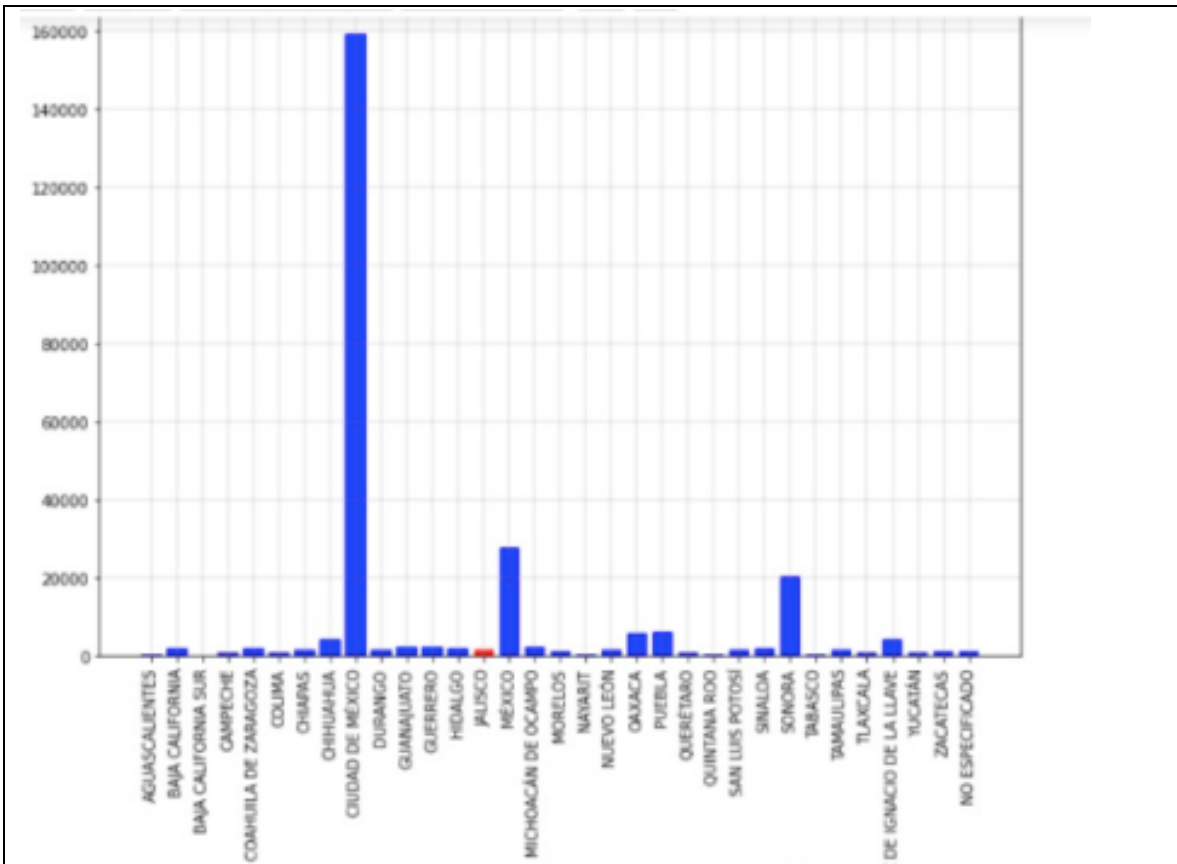


Ciudad de México, México y Nuevo León son las entidades con más casos que Jalisco.

¿Cuál fue el estado de México con menos contagios per cápita?

	Casos
NO ESPECIFICADO	15562
CAMPECHE	26227
COLIMA	32046
TLAXCALA	33493
QUINTANA ROO	38459

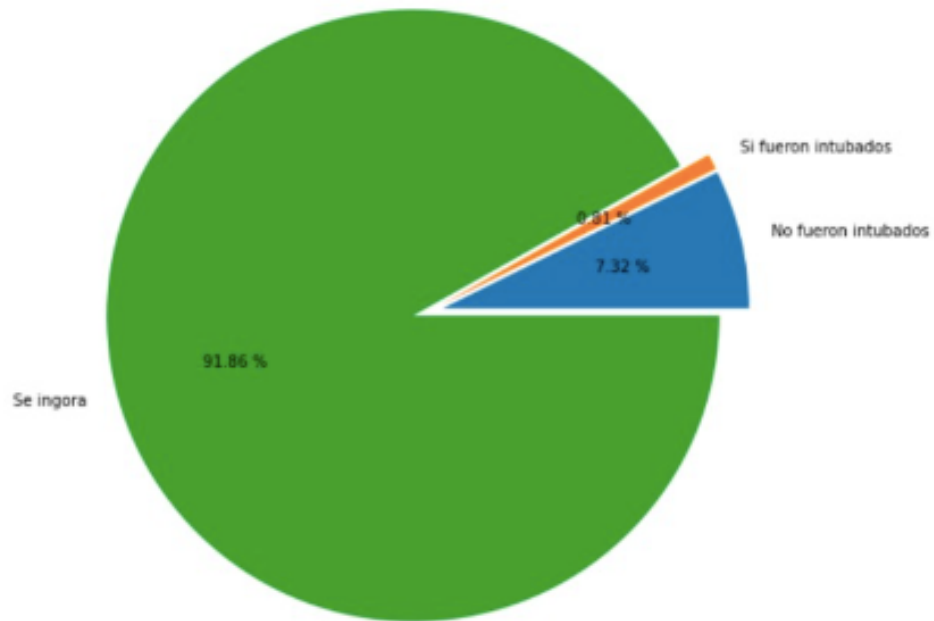
¿En qué estados ha habido menos restricciones para evitar la propagación del COVID?



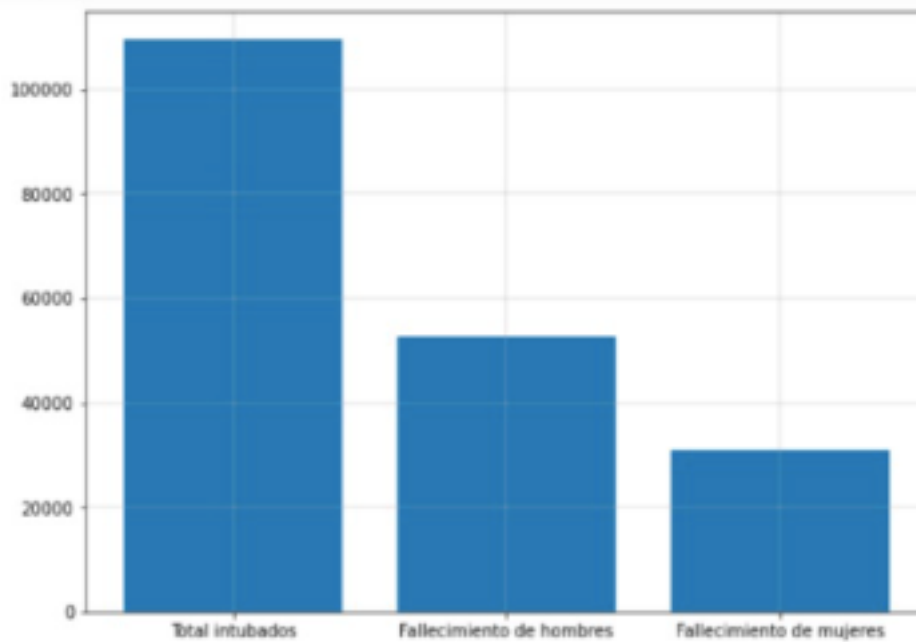
Los estados con la mayor cantidad de casos por asociación, es decir, la persona informa haber tenido contacto con un positivo a COVID-19 son:

	Casos
CIUDAD DE MÉXICO	159248
MÉXICO	27910
SONORA	20562
PUEBLA	6326
OAXACA	5789

¿Cuál es la relación de personas que fueron intubadas, las que no y las que se desconoce?

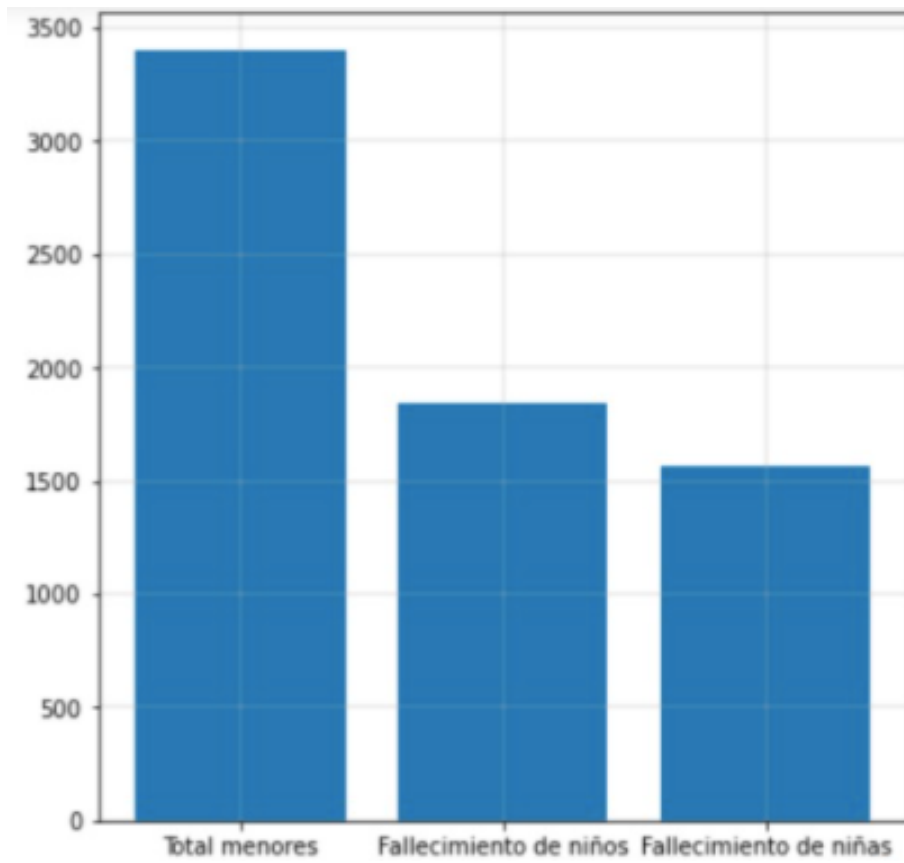


¿Qué porcentaje de personas que fueron intubadas murieron? ¿Cuántos fueron hombres y cuántas mujeres?



```
]: 1 print('el', np.round(((valores[1]+valores[2])/valores[0])*100,2), '% de
el 76.23 % de las personas que fueron intubadas murieron
```

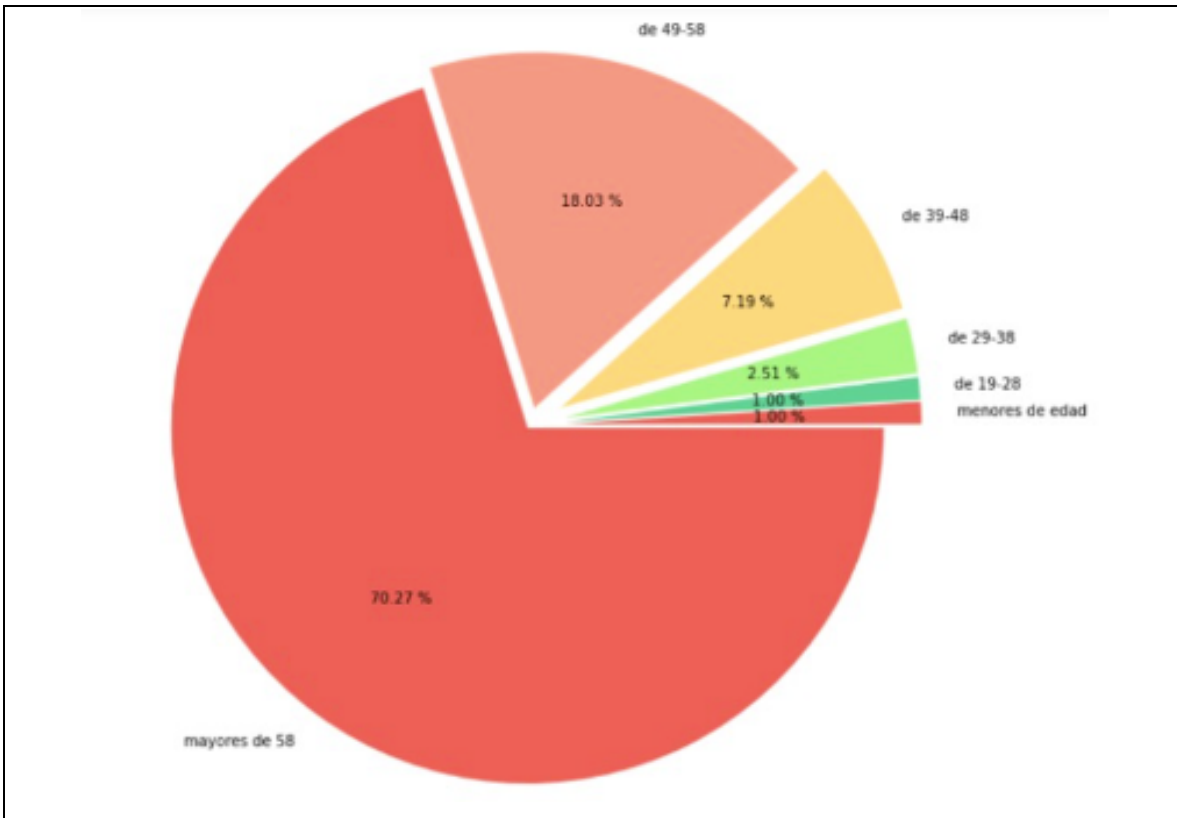
¿Cuántos menores de edad han fallecido?



Total fallecimientos

Total menores	3401
niños	1840
niñas	1561

¿Qué rango de edades ha sido la más afectada?



En esta gráfica se puede ver claro como el virus COVID-19 va subiendo su mortalidad conforme sube la edad de la persona infectada llegando a tener una mortalidad de alrededor del 35% en el grupo de las personas mayores a 57 años, lo que nos da a entender que la mortalidad del virus tiene una correlación con la edad del infectado y este grupo de personas ha sido el más afectado por la pandemia debido a que son el grupo más vulnerable.

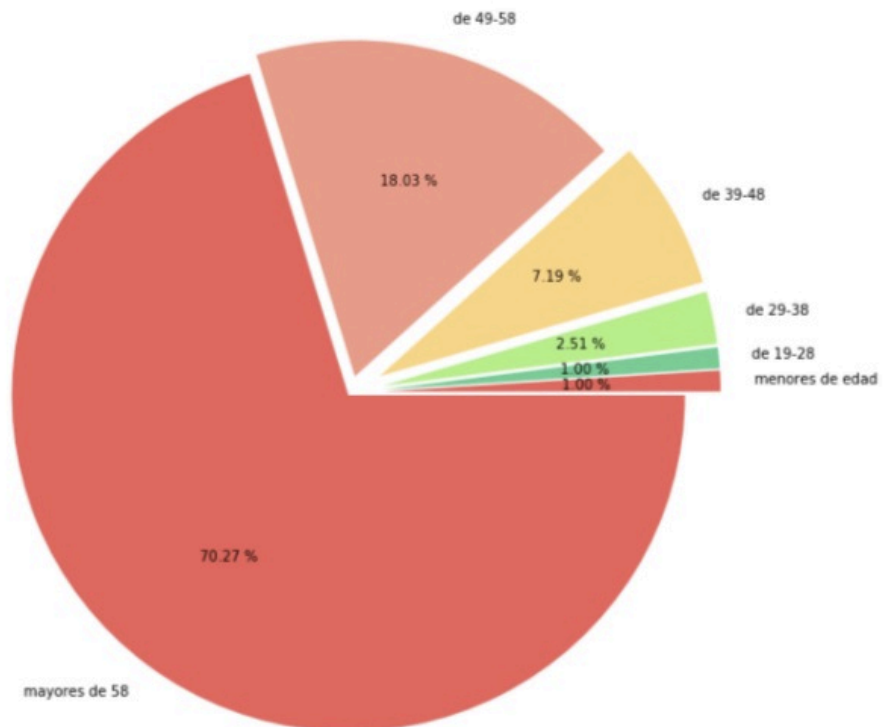
El realizar una limpieza y un análisis profundo nos ayudó a aligerar un poco la carga computacional ya que son diversos datos y además fue de gran utilidad ya que se lograron responder las preguntas centrales de una manera visual que facilita el entendimiento de la información.

Algunas de las preguntas que originalmente se habían planteado no se pudieron resolver como tal con los datos de SINAVE por lo que se modificaron.

Defunciones, Positividad y RT:

Defunciones:

La pandemia ha dejado varias defunciones alrededor del mundo, siendo México el lugar número 5 dentro de los países con mayores muertes debido al COVID. Totalizando 388,520 muertes dentro del territorio nacional. Al hacer un gráfico circular dividido por la cantidad de muertes por grupo de edad podemos notar que el grupo más afectado es el de las personas mayores a 58 años.

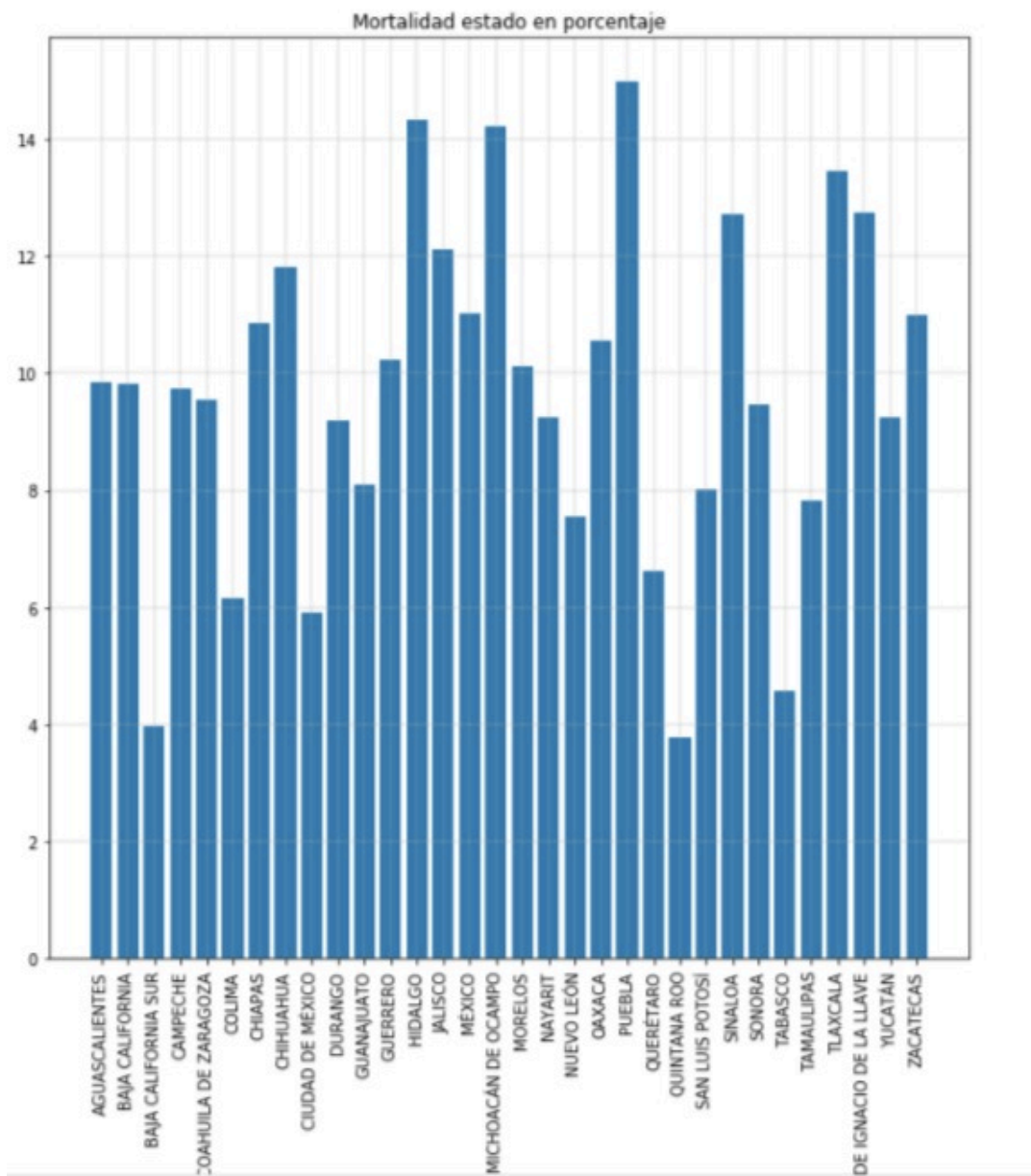


Mortalidad:

La palabra mortalidad está definida como la “Tasa de muertes producidas en una población durante un tiempo dado, en general o por una causa determinada” de acuerdo con la Real Academia Española. La mortalidad del COVID a nivel nacional se encuentra en 9.03%, la fórmula para obtener este número es la cantidad de muertes por COVID entre la cantidad de casos positivos.

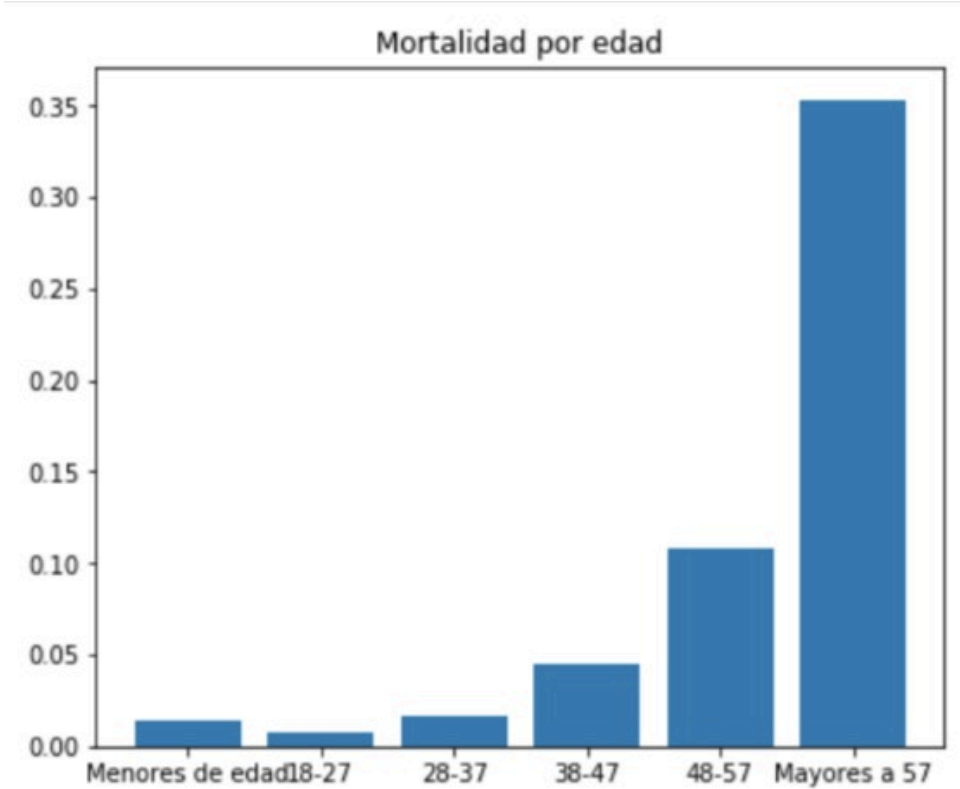
Desglosando la mortalidad por diferentes categorías podemos obtener varias conclusiones, las entidades con mayor mortalidad son las siguientes, llegando a tener hasta un 15% en el caso de Querétaro.

- Querétaro
- Michoacán
- Hidalgo



Al comparar la mortalidad de los hombres con las mujeres podemos ver que el COVID es más riesgoso para los hombres que para las mujeres debido a que la mortalidad en hombres es mayor un 4%.

La siguiente grafica de barras nos permite indagar un poco más en como el COVID ha afectado diferente dependiendo de a qué grupo de edad perteneces y podemos llegar a la conclusión de como aumenta el riesgo en la misma dirección que aumenta la edad y siendo el grupo menos afectado por este virus el de las personas que tienen entre 18 y 27 años.



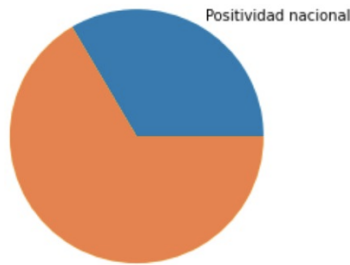
Positividad:

La positividad es una medida que se usa para saber si se están “escapando” casos de COVID. Lo que indica este porcentaje es si se está pudiendo encontrar adecuadamente a las personas infectadas en la población.

La Organización Mundial de la Salud (OMS) recomienda mirar esta variable para evaluar si la definición de caso sospechoso y la cantidad de rastreos de contactos es suficiente y sugiere mantenerla por debajo del 10%.

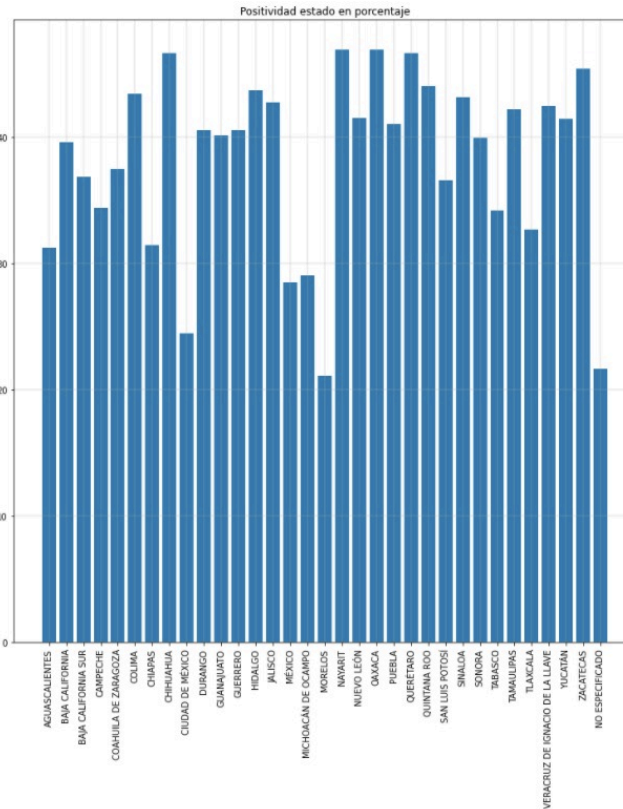
Se calcula dividiendo el número de casos positivos confirmados entre el número de casos positivos más negativos.

Positividad nacional:



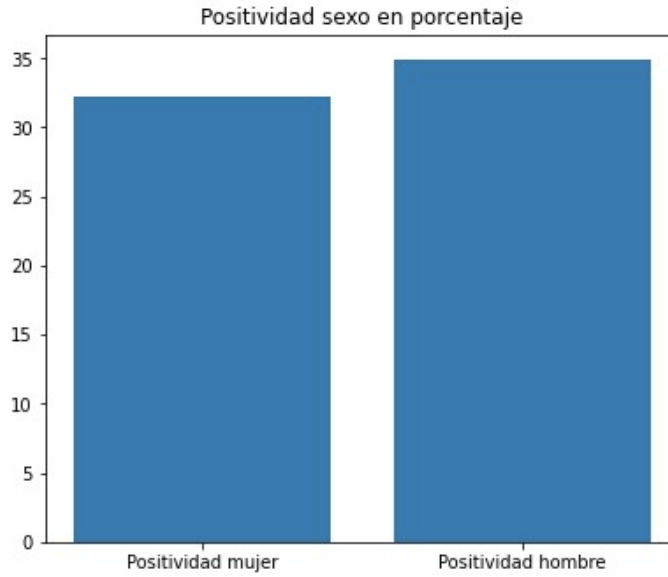
'La positividad a nivel nacional es de 33.484699030943425%'

Positividad por estados:

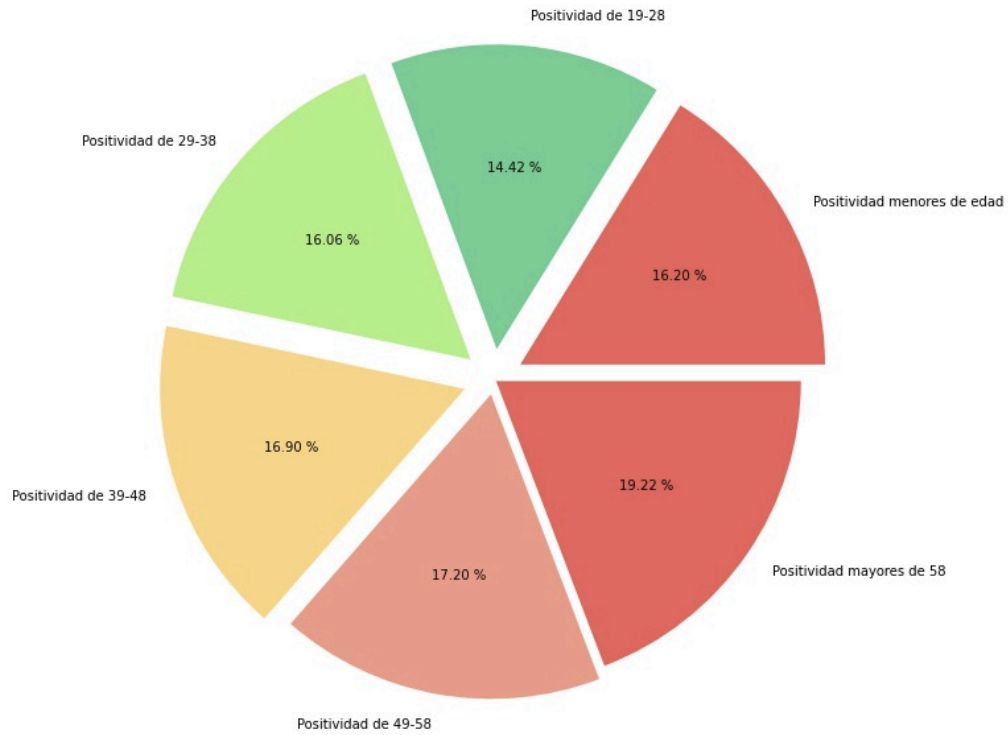


Positividad por sexo:

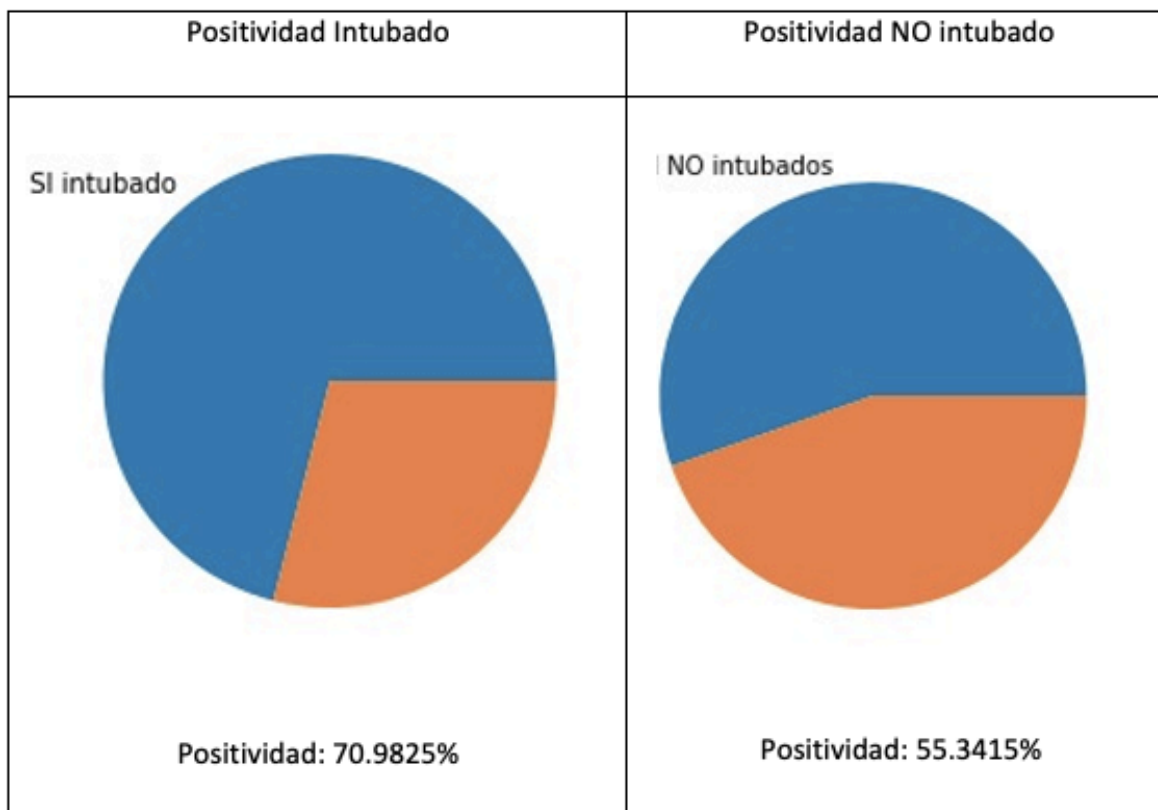
	Sexo	Positividad sexo
0	1.0	0.321892
1	2.0	0.349551



Positividad por edad:



Positividad intubados y no intubados:



A partir del porcentaje de positividad obtenido con los datos de SINAVE y calculado por grupos se puede percibir que la positividad supera por mucho el porcentaje recomendado por la OMS; incluso la positividad nacional triplica ese 10%.

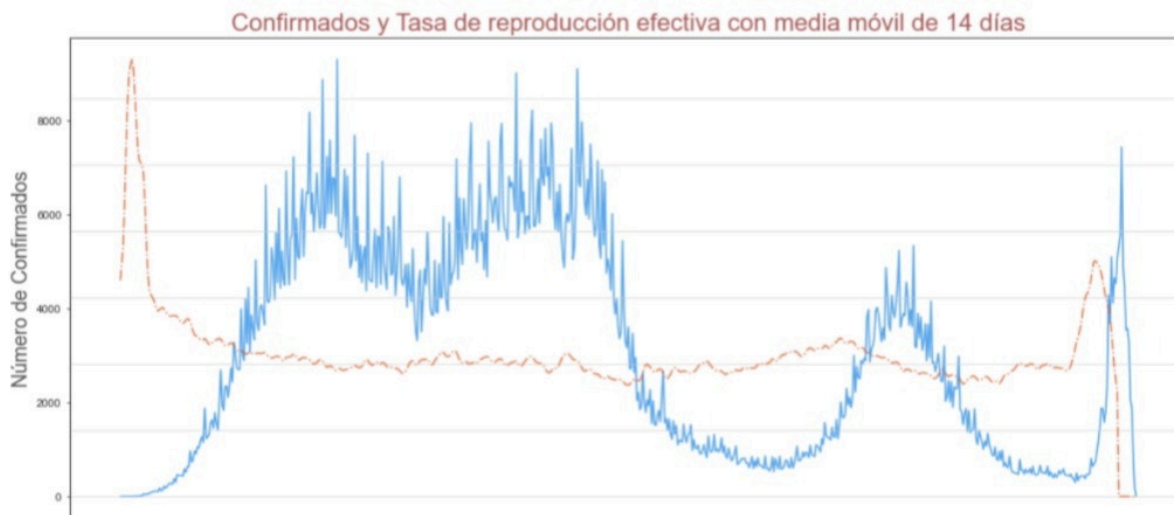
La clave para disminuirla es llevar a cabo un correcto rastreo de contactos ya que no se puede simplemente realizar pruebas COVID masivas a toda la población.

El hecho de que el porcentaje de positividad sea tan alto se debe en gran parte a que no se cumple este lineamiento de “rastreo de contactos” a pesar de ser requerido por la OMS, que indica que lo ideal sería realizar una prueba en personas que tuvieron contacto con infectados.

Tasa de reproducción efectiva del virus (RT):

1. Nivel Nacional

La tasa de reproducción efectiva a nivel Nacional ha llegado a tocar hasta el punto de 5.5 en Marzo del 2020 pero como podemos ver en el siguiente gráfico en el cual se le aplicó una media móvil de 14 días, ha mostrado que se ha mantenido la mayor parte del tiempo en un valor de 1.

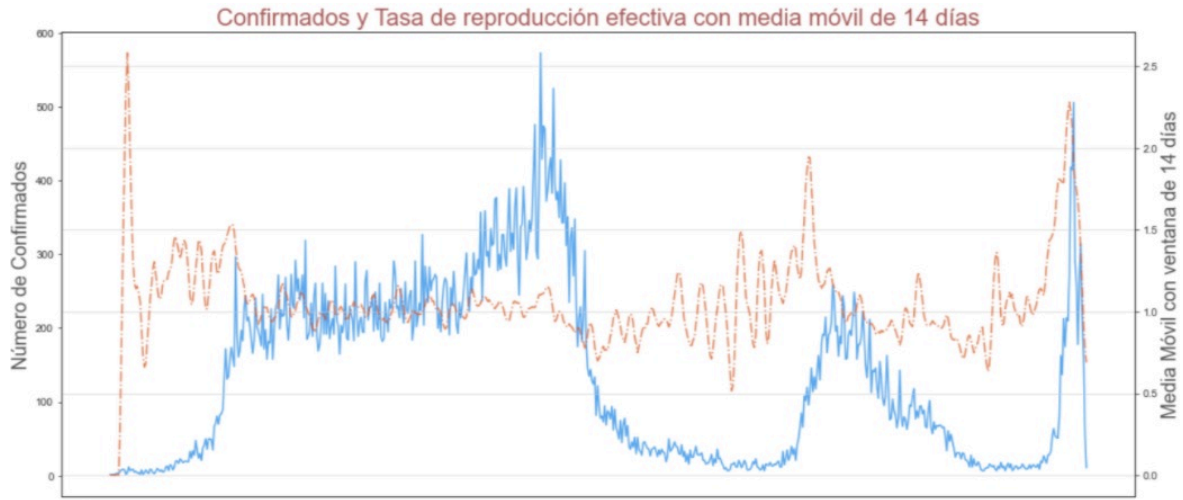


2. Por Estados

Realizamos las visualizaciones de la tasa de reproducción efectiva diaria a nivel estatal para cada uno de los 32 estados, sin embargo el que más nos interesó analizar es Jalisco.



Podemos observar que la R_t ha oscilado a lo largo del tiempo entre 0 y 3.5, pero para poder eliminar el ruido de esta gráfica y poder observar los movimientos con un poco más de claridad, decidimos de igual manera aplicar una media móvil de 14 días.



3. Por Sexo:



Como podemos apreciar en el gráfico, no existe una notable disparidad entre el comportamiento que ha tenido la tasa efectiva de reproducción entre hombres y mujeres.

Análisis comparativo Municipio/ Estados y entre olas de contagio

Casos activos:

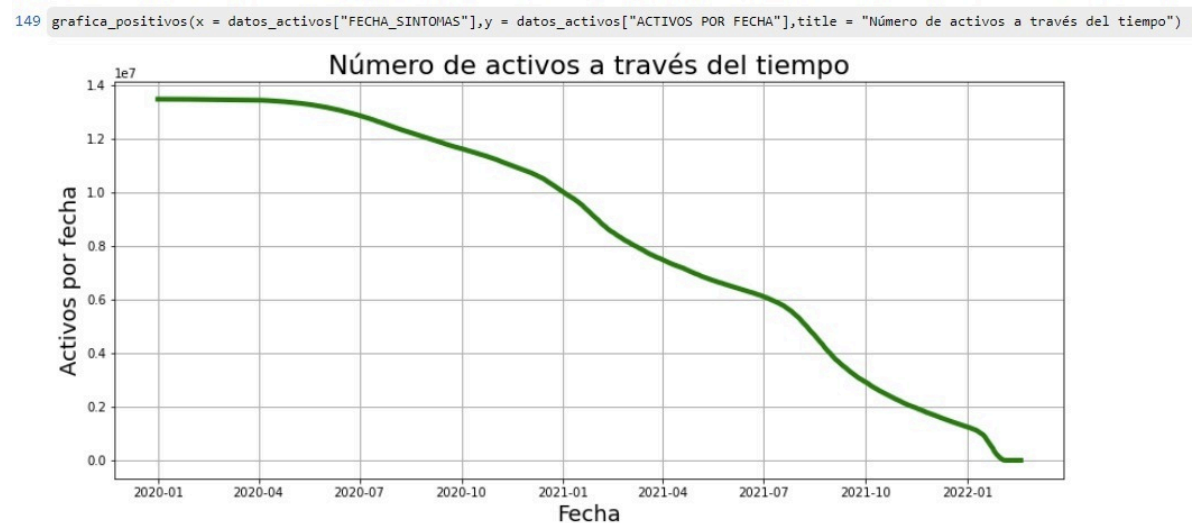
Se realizó el cálculo de los casos activos filtrando 14 días después del inicio de síntomas, después se creó un nuevo dataframe en donde se guardaron los valores de los pacientes que cumplían esta característica por fecha.

El número de casos positivos varía mucho con el tiempo, este es un ejemplo representativo de algunas fechas:

```
117 datos_activos.sample(5)
```

	FECHA_SINTOMAS	ACTIVOS POR FECHA
741	2022-01-11	1062211
193	2020-07-12	12715430
442	2021-03-18	7736835
734	2022-01-04	1195968
753	2022-01-23	489235

Graficando los activos de todas las fechas reportadas en la base de datos de SINAVE, el comportamiento se ve de la siguiente manera:



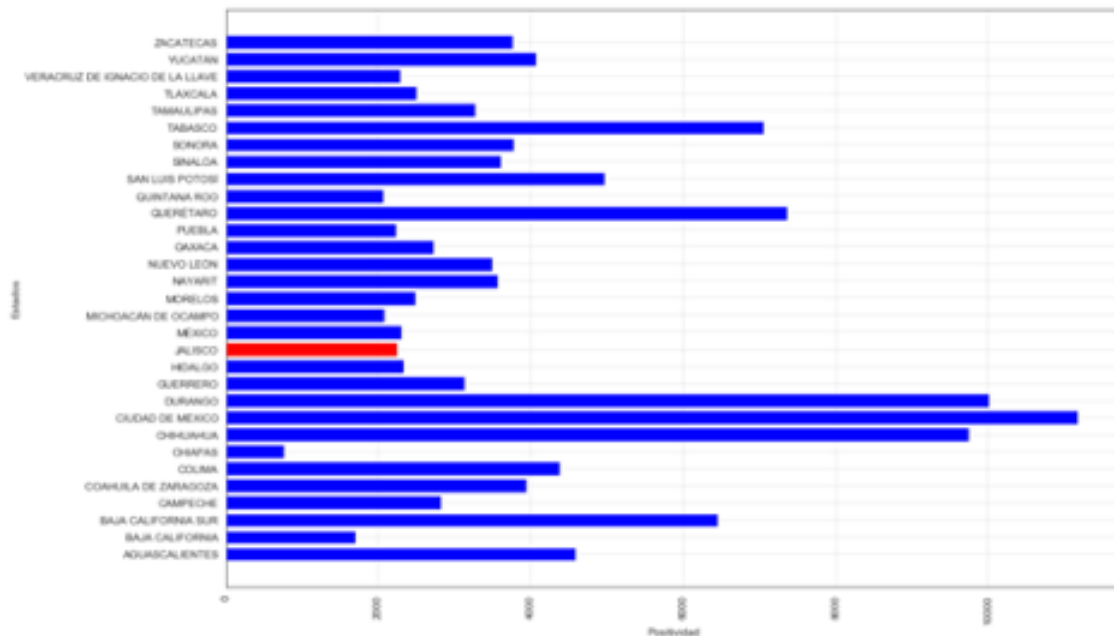
Evidentemente ha ido disminuyendo a través del tiempo y en cuanto a casos activos no es posible observar las olas como en otros indicadores a excepción de que en el 2020 si había muchos más casos activos y esto se debe a que el número de contagios a nivel nacional era mucho mayor y mientras unos pacientes terminaban sus 14 días activos muchos más los comenzaban.

Análisis comparativo por municipio/estados

Positividad:

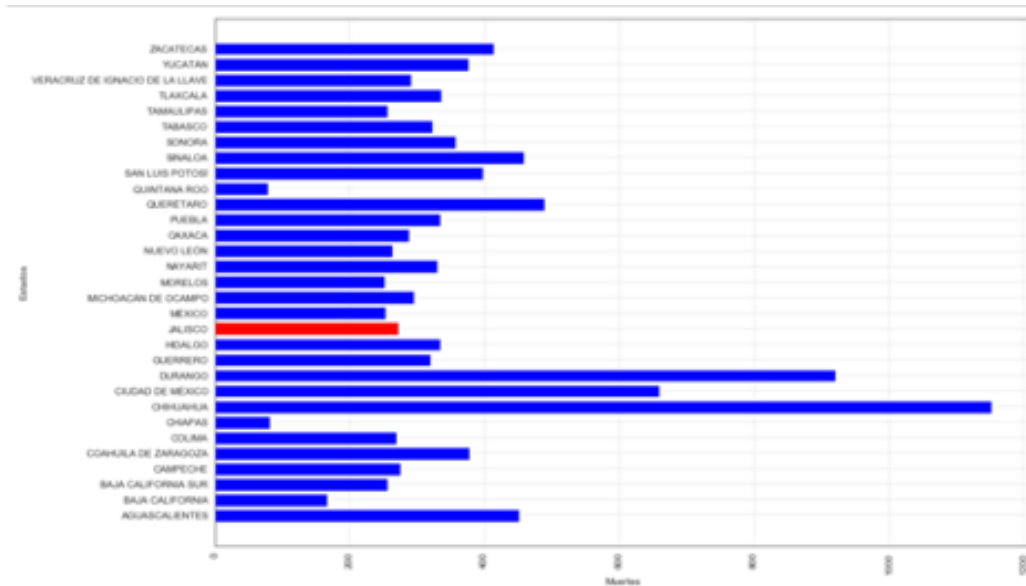
Para poder realizar un análisis comparativo entre distintos estados es necesario primero normalizar los datos debido a que cada uno de estos tiene diferente población por lo que no es lo mismo 3 positivos en Colima que es un estado con poca población a esos mismos positivos en el estado de México que es el estado más poblado de México.

En este caso se dividió la cantidad de positivos por el número de habitantes de cada estado y se hizo una multiplicación por 100,000 dando como resultado el número de positivos por cada 100,000 habitantes. La tasa de positividad de Jalisco esta subrayada en rojo y no es de las más altas de México.



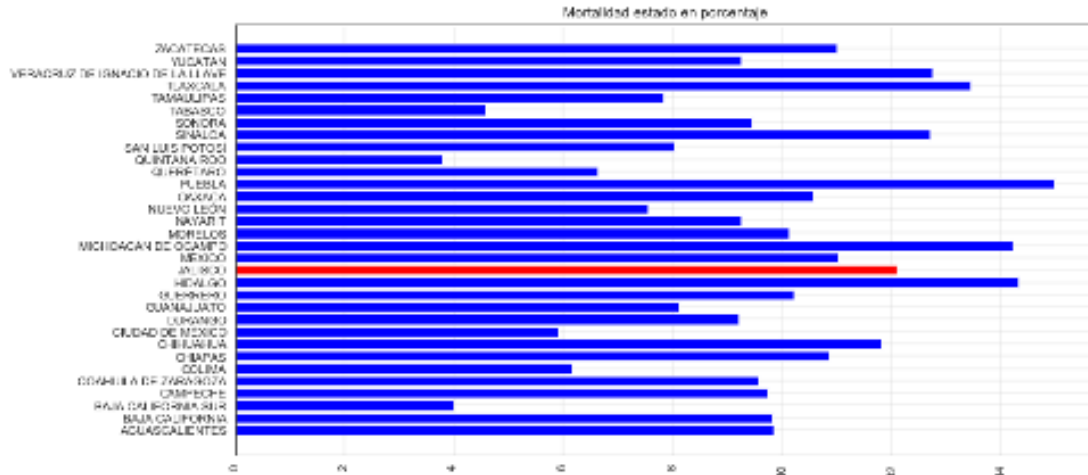
Muertes:

En la siguiente gráfica se muestran el número de muertes por cada 100,000 habitantes de cada estado y Jalisco resaltado de tono rojo, cómo podemos notar los estados que más se han visto afectados por el COVID en muertes son Durango, Ciudad de México y Chihuahua, este último teniendo más de 1,000 muertes por cada 100,000 habitantes del estado.



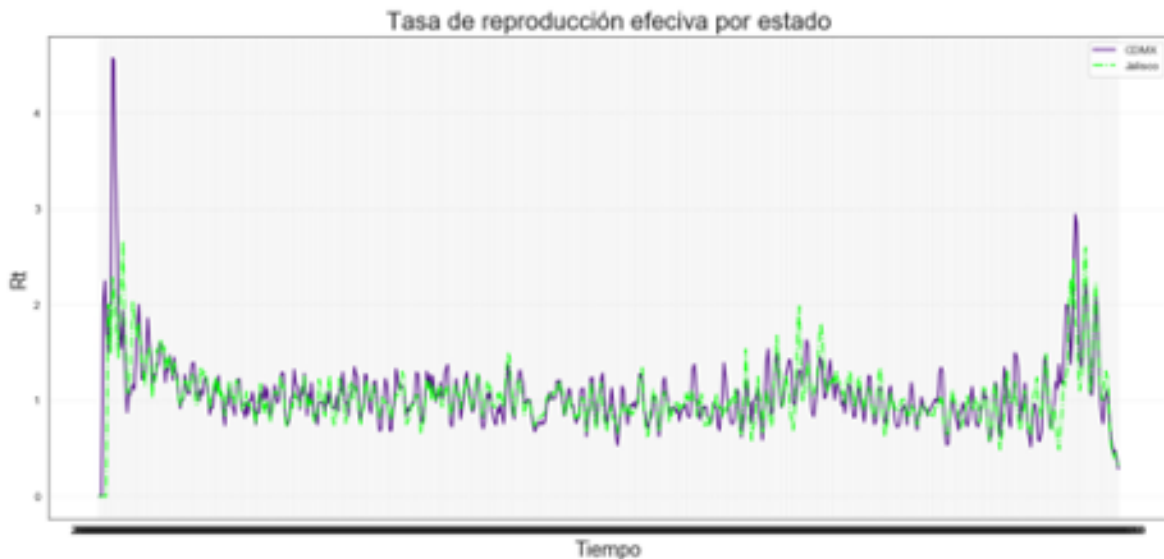
Mortalidad:

Mientras que en la siguiente gráfica se muestra la mortalidad, la cantidad de muertos que hay entre la cantidad de positivos en cada estado, se puede ver que la mayoría de los estados se encuentran alrededor del 10% siendo Jalisco un estado con esta métrica bastante alta cuando se compara con otras entidades, se encuentra como el estado número 7.



RT:

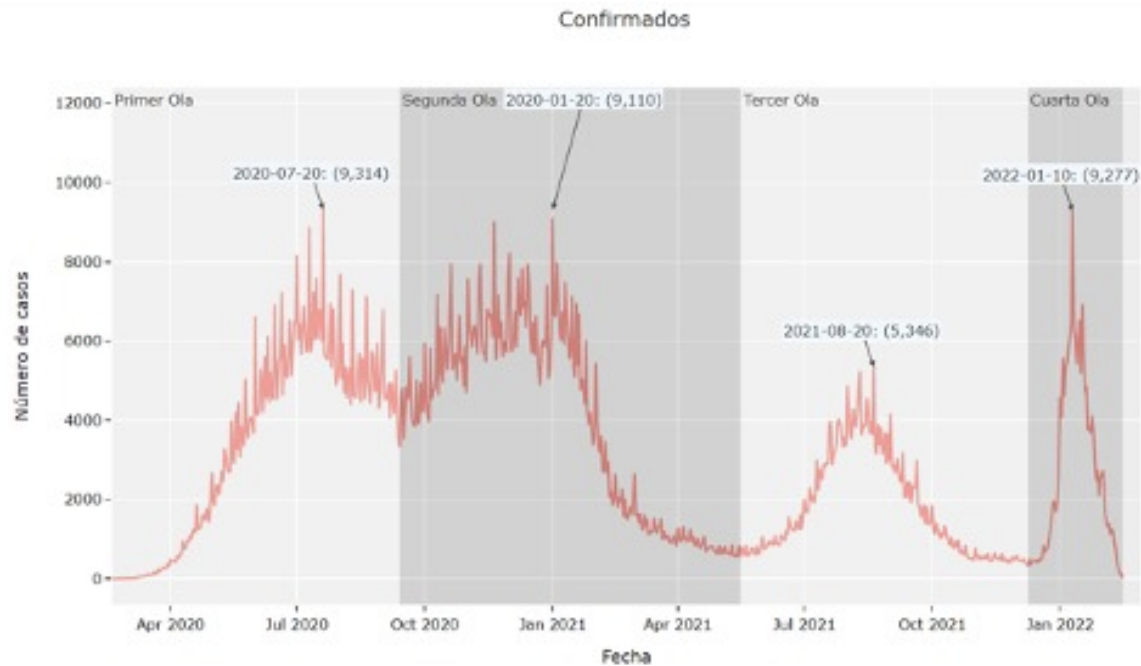
En la gráfica de tasa de reproducción efectiva se pueden ver las diferencias que ha habido entre diferentes estados, en este caso cuando comparamos las tasas entre Jalisco y CDMX se muestra como la CDMX (morado) aunque se comporta de manera muy similar a Jalisco, esta ha llegado a tener picos más altos lo que podría ser un indicador de que las medidas para evitar el propago de COVID han sido más duras o más eficientes en el estado de Jalisco.



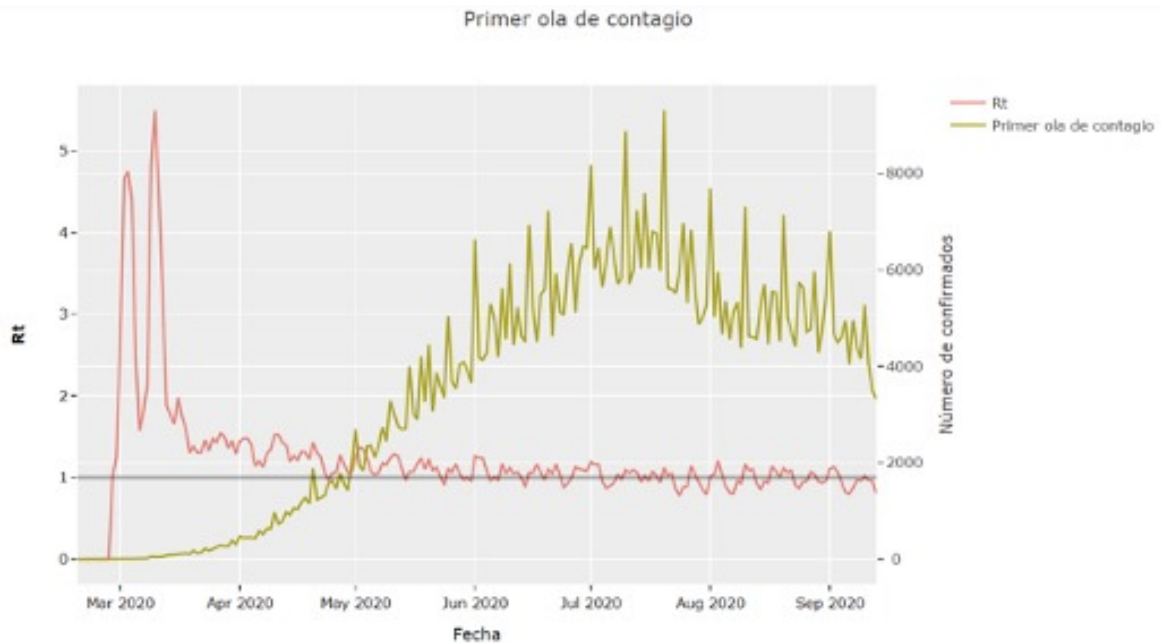
Olas de contagio (Nivel nacional):

En esta primera gráfica podemos observar el comportamiento del número de confirmados a lo largo de estos últimos dos años, diferenciando el inicio y fin de cada ola con los mínimos

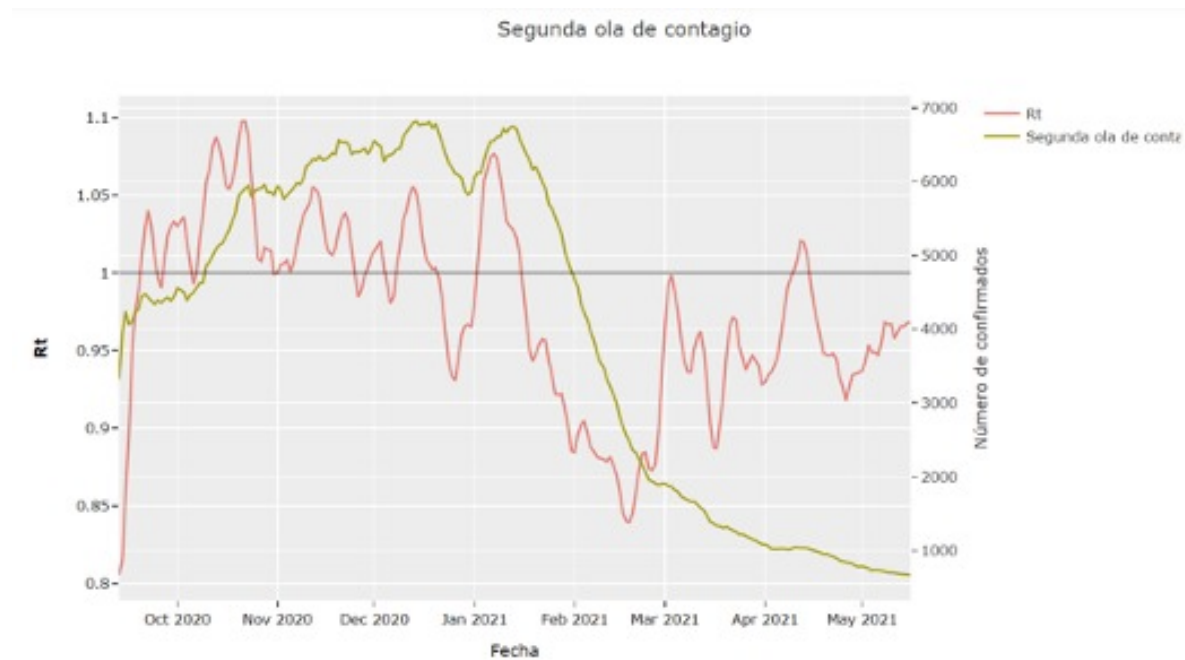
locales de esta serie de tiempo. Podemos observar el máximo número de confirmados en cada ola y la fecha en que ocurrieron.



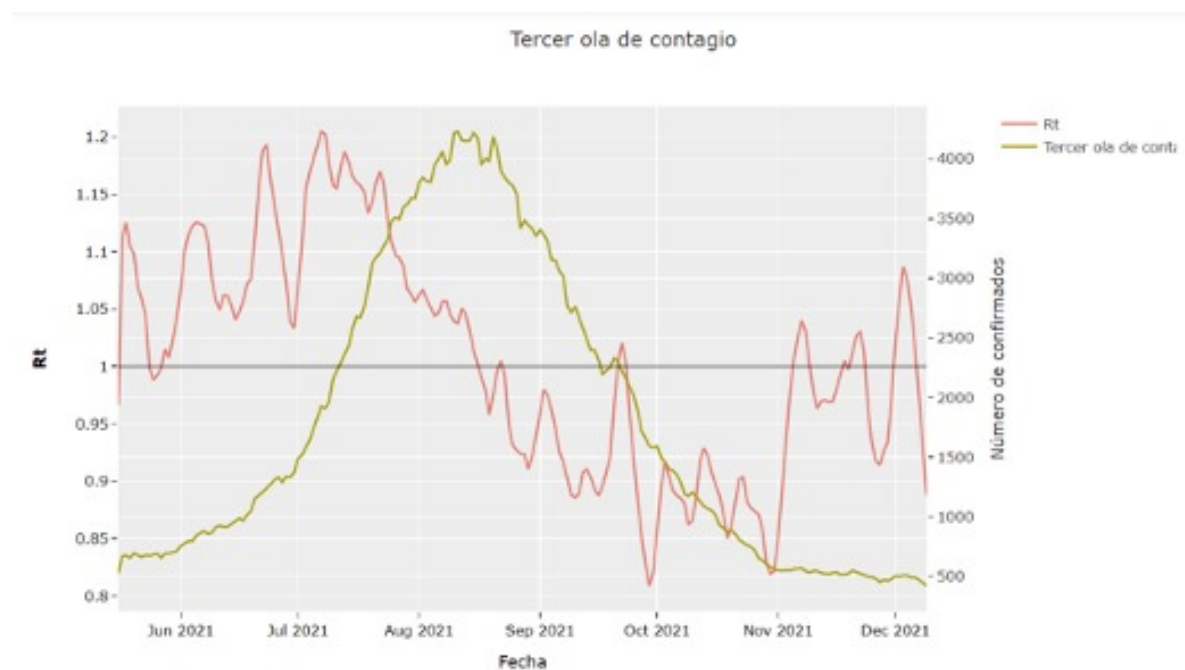
Enfocándonos en la primera ola de contagio decidimos compararla contra la tasa de reproducción efectiva, resaltando cuando la tasa es mayor o menor a uno debido a que son los momentos en que nuestros casos confirmados aumentan o disminuyen.



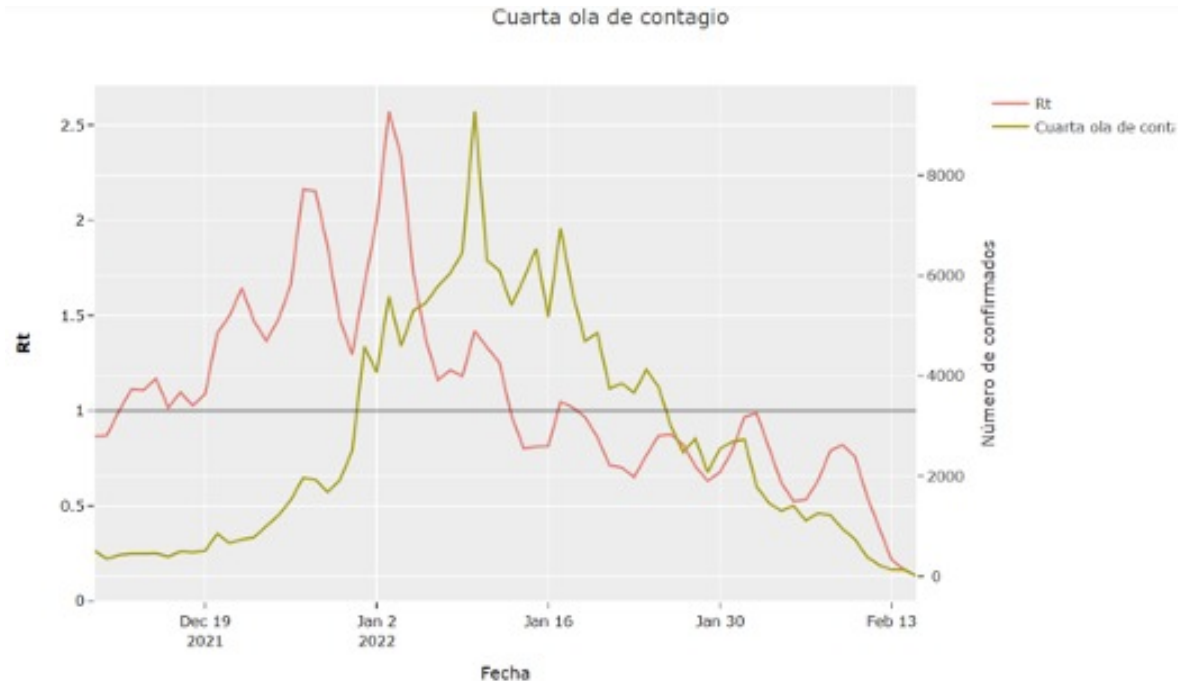
En cuanto a la segunda ola decidimos utilizar una media móvil de 14 días debido a que el gráfico presentaba demasiado ruido y de esta manera obtuvimos una mejor visualización.



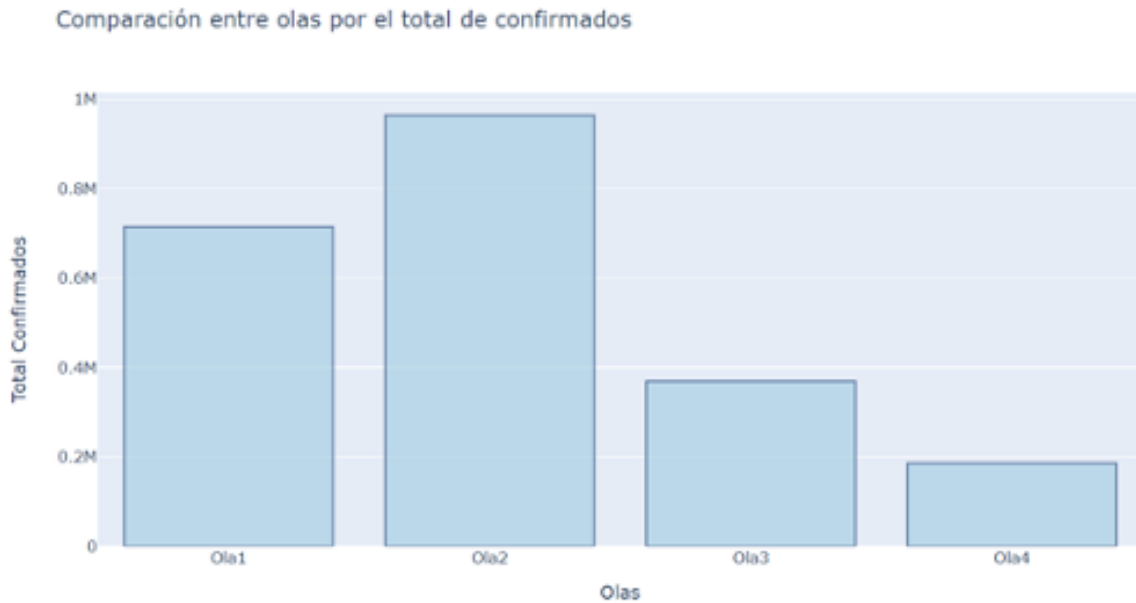
Para la tercera ola también aplicamos media móvil, pero en este caso fue con ventana de 7 días.



Y para nuestra cuarta y última ola al no presentar ruido, no le aplicamos media móvil.

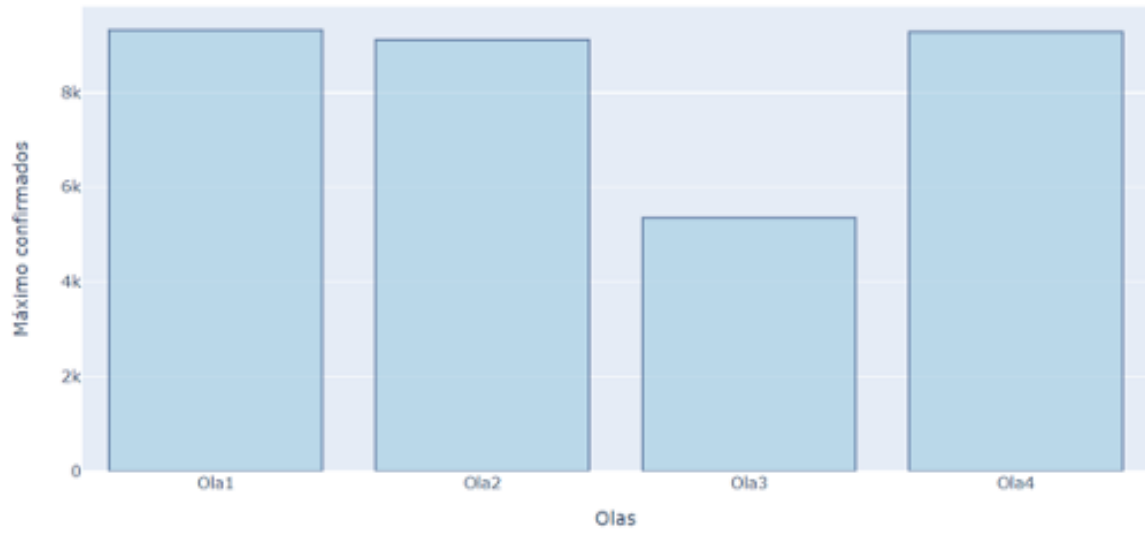


También decidimos realizar otros tipos de comparaciones entre estas olas, siendo la primera comparación sobre el total acumulado de casos confirmados de cada ola.



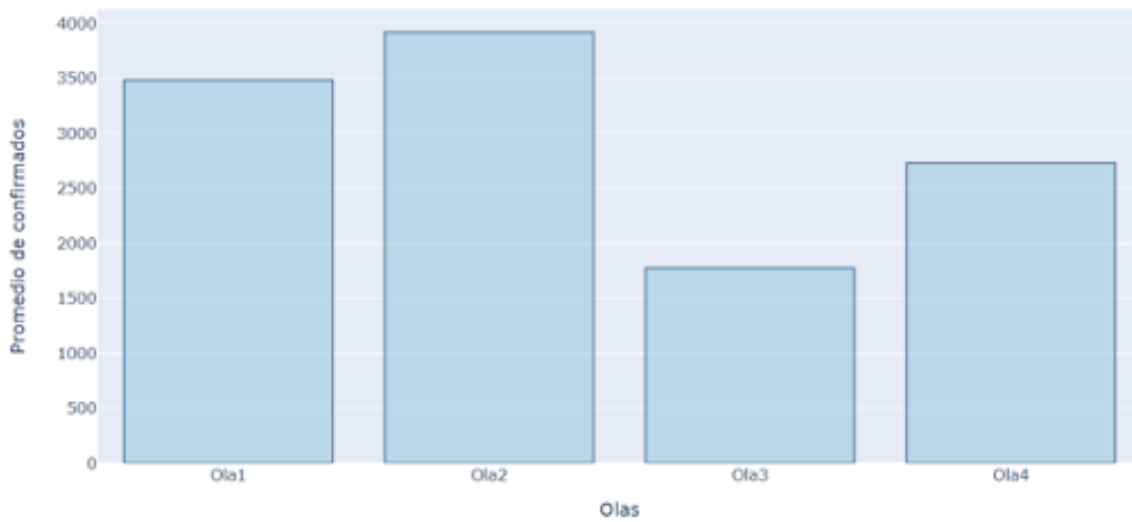
Además, comparamos por el máximo número de contagios registrado en 1 día

Comparación entre olas por máximo de confirmados



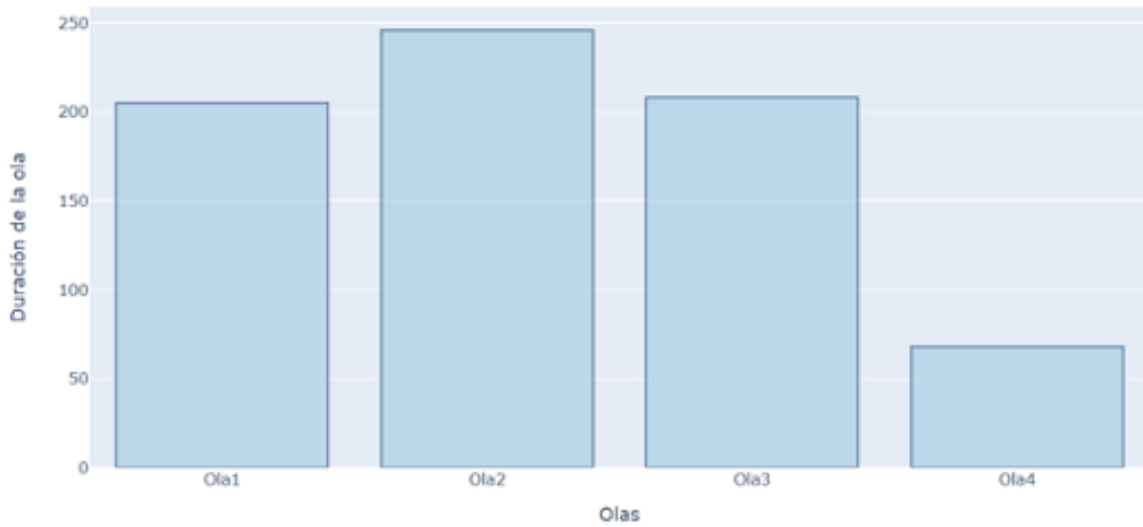
Por promedio de casos confirmados:

Comparación entre olas por promedio de confirmados



Y el último criterio para este análisis fue el de la duración en días que tuvo cada ola.

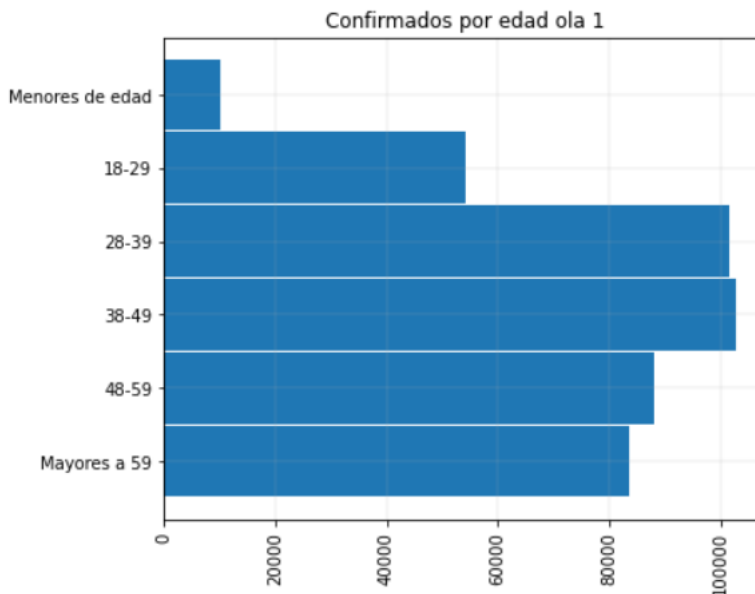
Comparación entre olas por días que duro cada una



Distribución de edad por olas

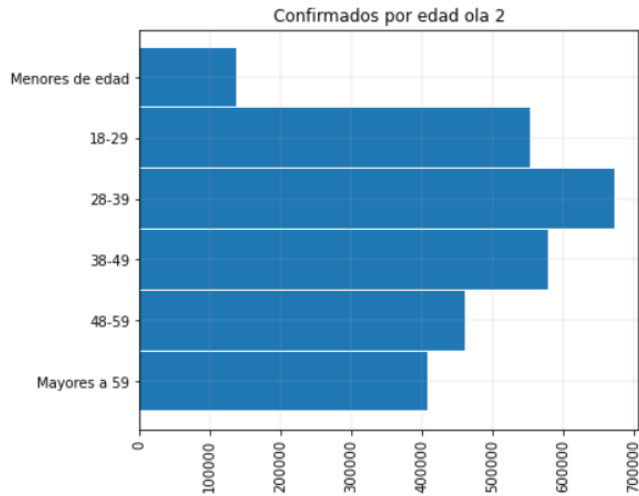
Primer Ola

En esta primera ola podemos observar que el mayor número de confirmados se concentró en los rangos de edad de entre 28 a 49 y siendo los menores de edad junto con los jóvenes menores de 29 años los que presentaban un menor número de confirmados.



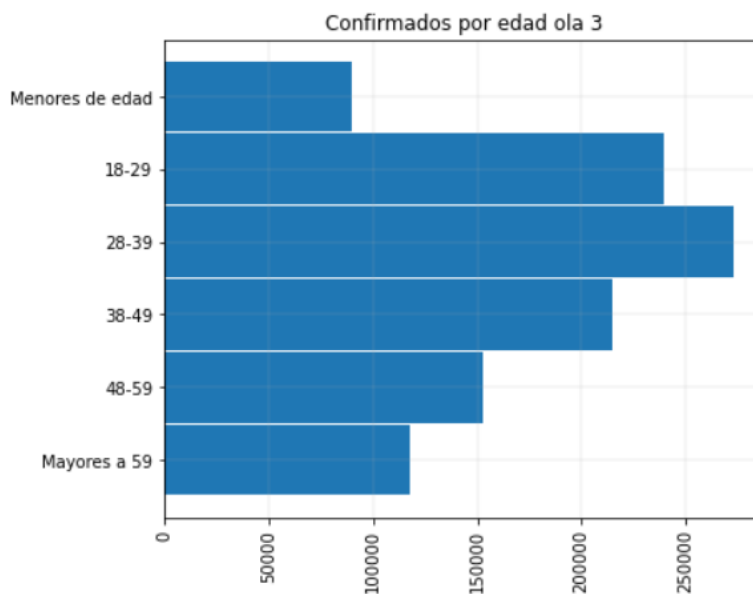
Segunda Ola

En la segunda ola fue bastante significativo el incremento de casos confirmados en todos los rangos de edad, sin embargo, las personas que se encontraban en un rango de edad de entre 28-39 fue el que tuvo el mayor incremento de todos los demás.



Tercer Ola

En cuanto a la ola número 3 observamos primeramente que hubo una reducción importante de número de casos confirmados en adultos mayores de 48 años en adelante y que los jóvenes de 28-39 años por segunda ocasión fueron los que presentaron la mayor concentración de confirmados.



Cuarta Ola

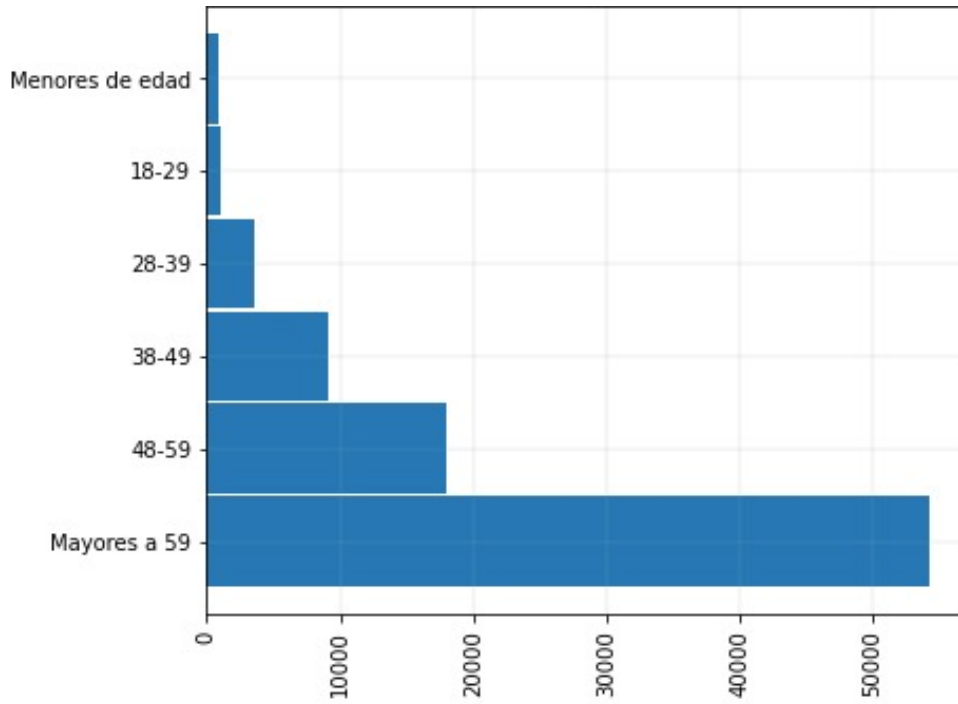
En la cuarta y última ola se presentó una distribución de casos altamente parecidos a los de la tercera ola, siendo menores de edad y adultos mayores de 48 años los que presentaron menos confirmaciones.



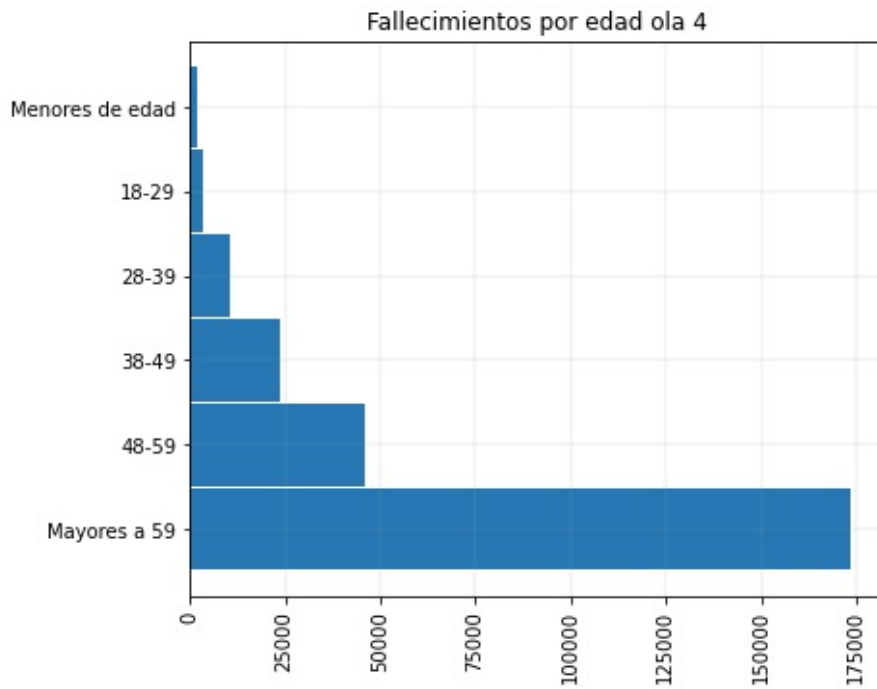
Defunciones por edad (Nacional)

En cuanto a las defunciones en todo México podemos apreciar que el comportamiento es prácticamente idéntico en las 4 olas de contagios, en las cuales se presenta que las defunciones están positivamente correlacionadas con la edad de las personas, mientras mayor es su rango de edad mayores defunciones se presentan.

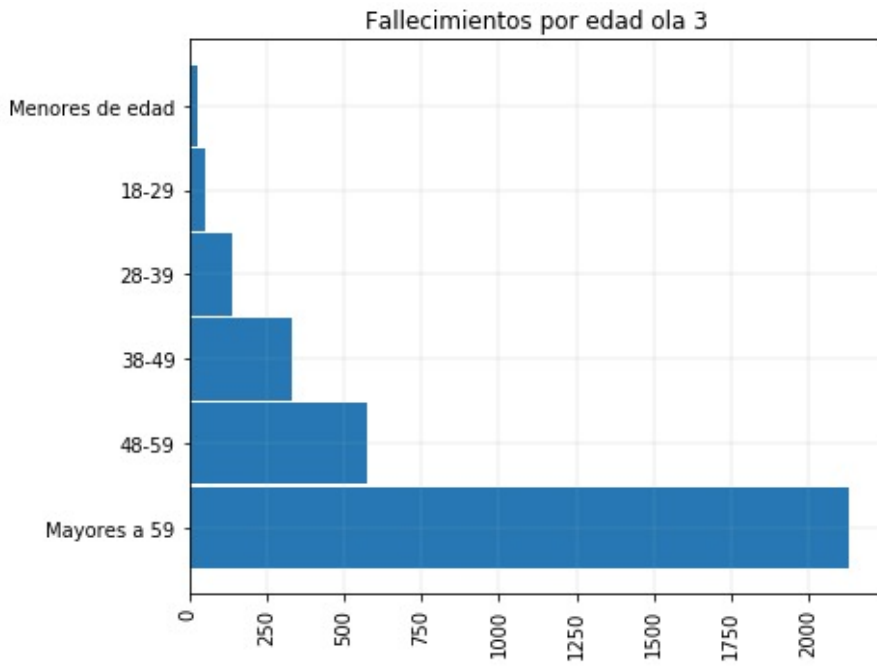
Primer Ola



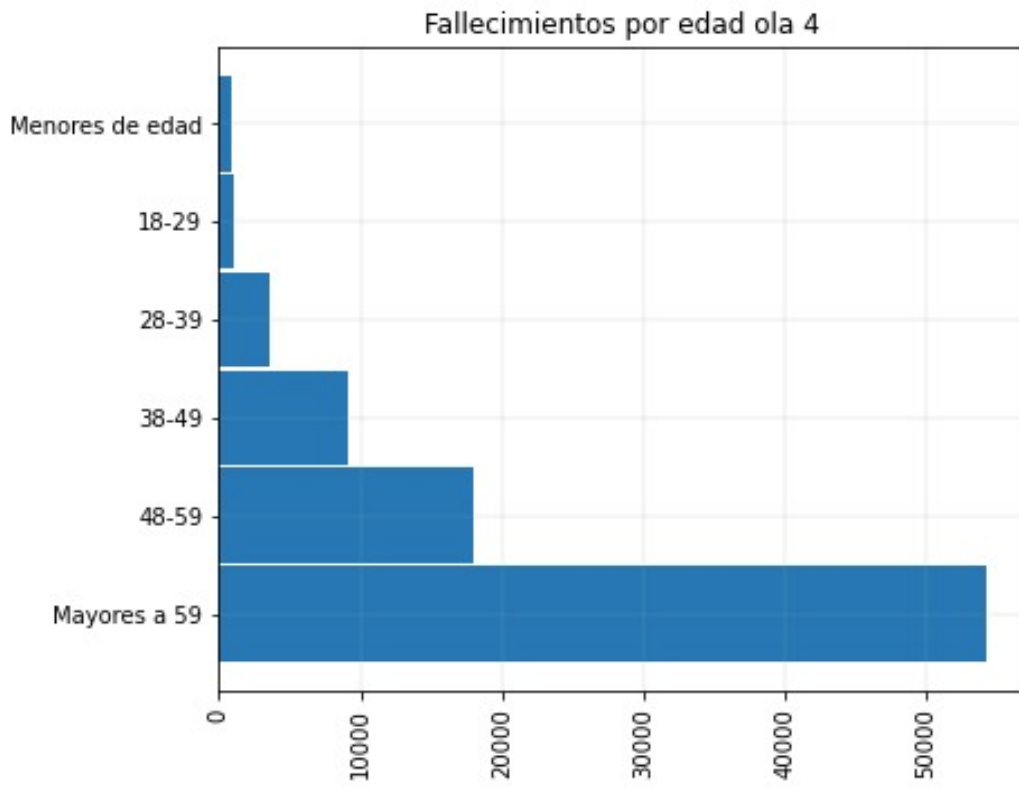
Segunda Ola



Tercer Ola



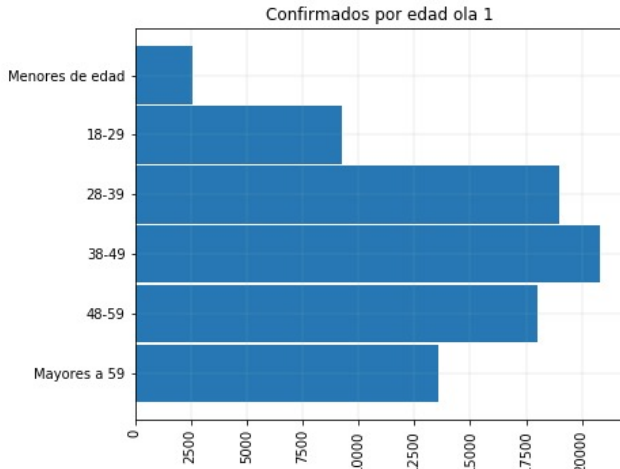
Cuarta Ola



Confirmados por edad (Ciudad de México)

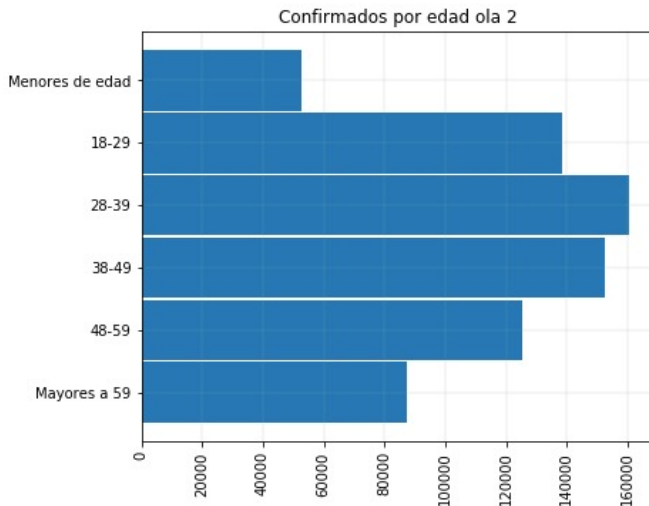
Primer Ola

Respecto a la primer Ola lo que es notable resaltar es que el rango de edad que presentó un mayor número de confirmados es el de 38-49 y los que menor número presentaron fueron los menores de edad.



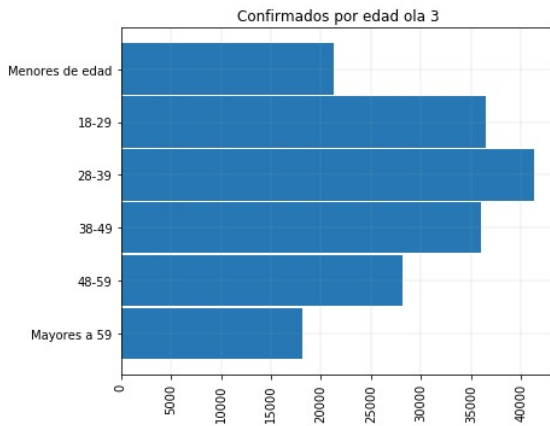
Segunda Ola

En la segunda ola al igual que a nivel nacional los números de casos confirmados aumentaron respecto a la primera ola, sin embargo, se presentó una distribución bastante similar.



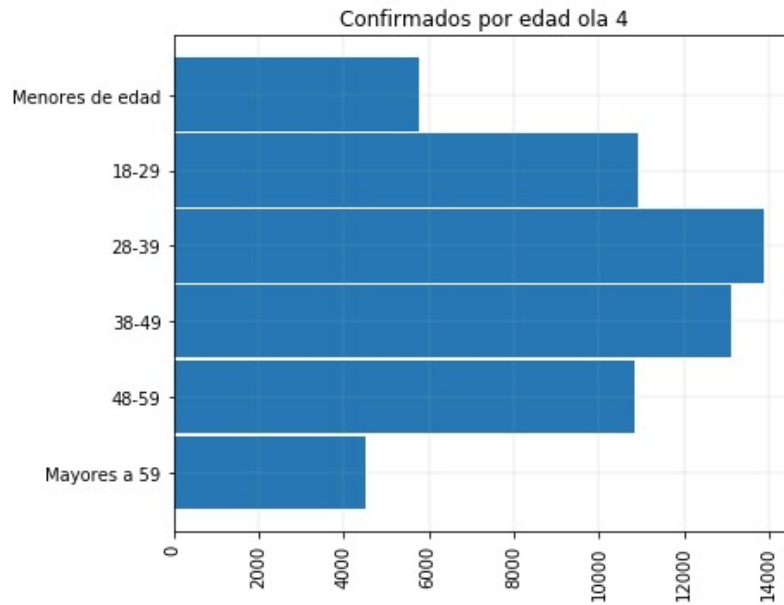
Tercer Ola

En la tercer ola en la ciudad de México pudimos notar que sucedió algo que a nivel nacional no, los números de confirmados en los menores de edad superaron a los casos en los adultos mayores a 59 años.



Cuarta Ola

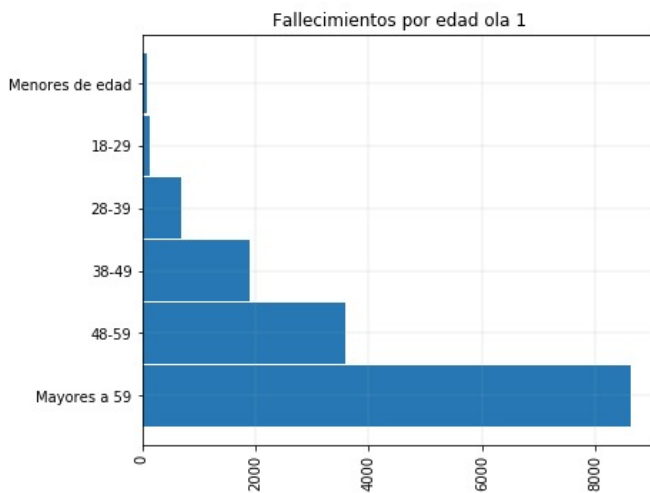
Y en nuestra cuarta ola se volvió a presentar el mismo fenómeno que en la tercera ola



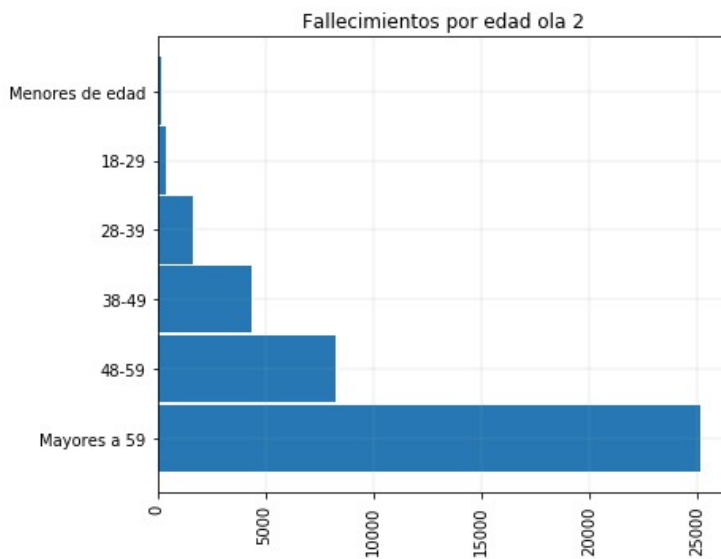
Defunciones por edad (Ciudad de México)

En el tema de las defunciones en la Ciudad de México obtuvimos como resultados los mismos que a nivel nacional, donde los adultos mayores de 59 años fueron los más afectados en las 4 olas, sin embargo, algo que se presentó particularmente en Ciudad de México fue que en la última ola el número de menores fallecidos superó a los jóvenes de 18 a 39 años.

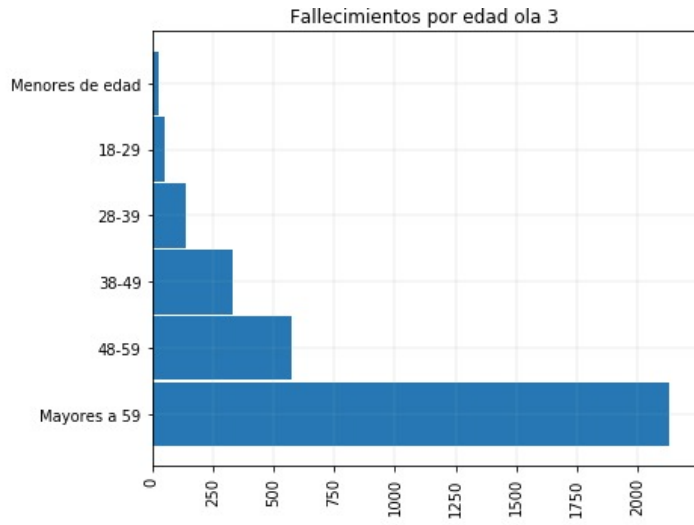
Primer Ola



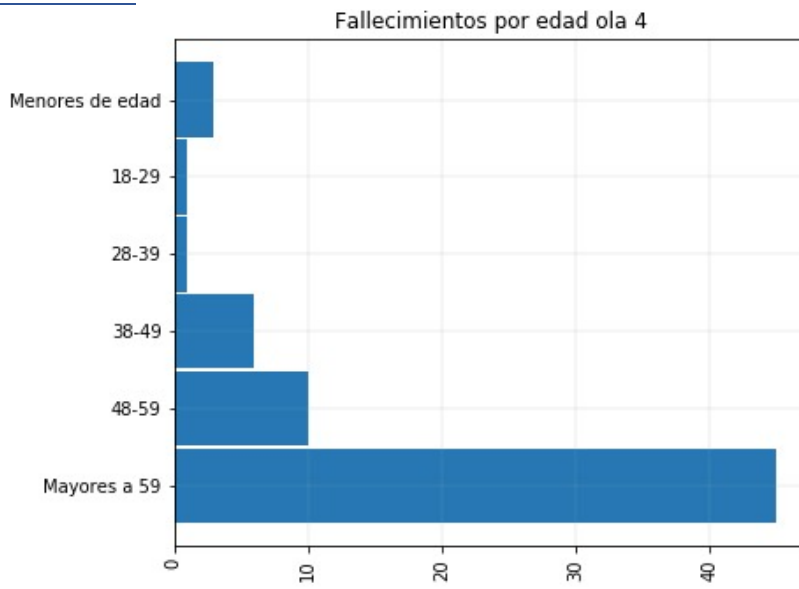
Segunda Ola



Tercer Ola



Cuarta Ola



Razones de cambio

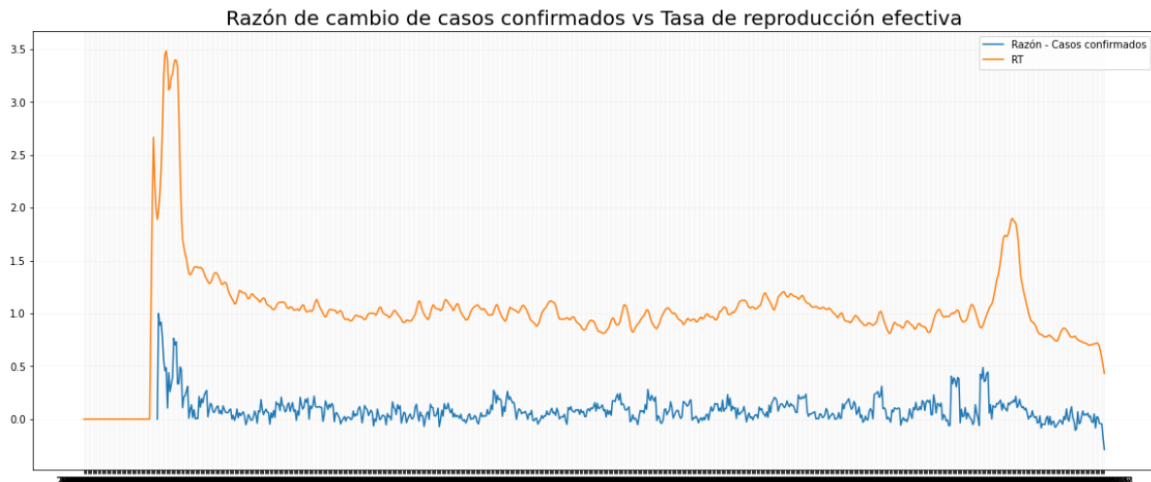
Se realiza un análisis visual de las razones de cambio de los casos confirmados y las defunciones en relación a la positividad y la tasa Rt.

Comenzamos creando una nueva base de datos con cada una de las variables que nos interesa

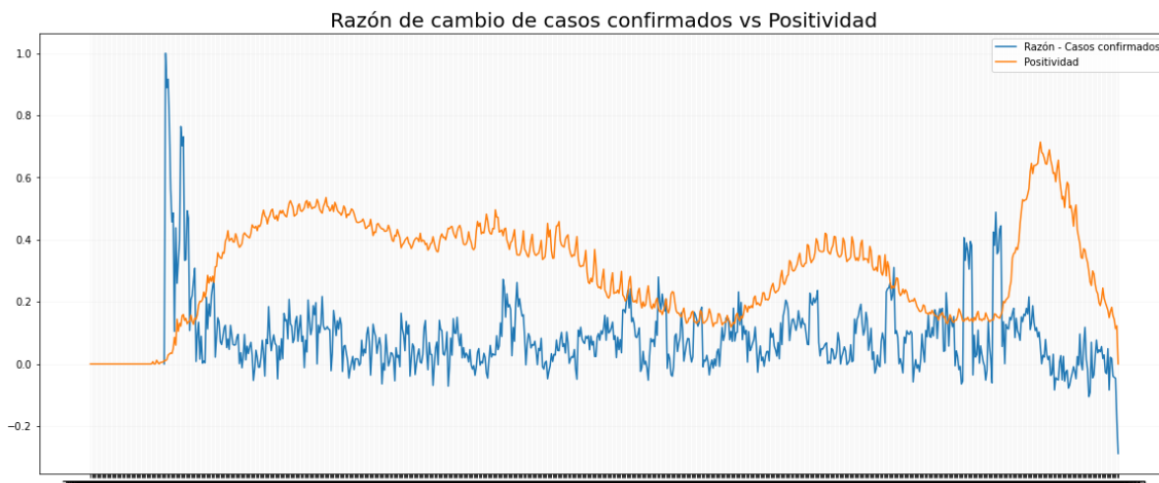
- Número de positivos: los que en nuestra base de datos están clasificados con un 1 en el resultado de laboratorio y se sumaron los totales por fecha sin repetir las fechas para que esta posteriormente nos sirviera de eje x.
- Número de confirmados: los confirmados en la base de datos de SINAVE son los que tienen 1 en su clasificación oficial ya que no se consideraron los no oficiales.
- Número de defunciones: se sumaron las defunciones por fecha también
- Positividad: Número de casos positivos (confirmados) dividido entre el número de casos positivos más negativos, no se consideraron los invalidos o que no se realizó prueba.

para posteriormente calcular sus cambios, es decir, el porcentaje que aumentó o disminuyó un elemento en relación al anterior, en este caso las variables estaban acomodadas por fecha ascendente así que se calculó el cambio de un día a otro.

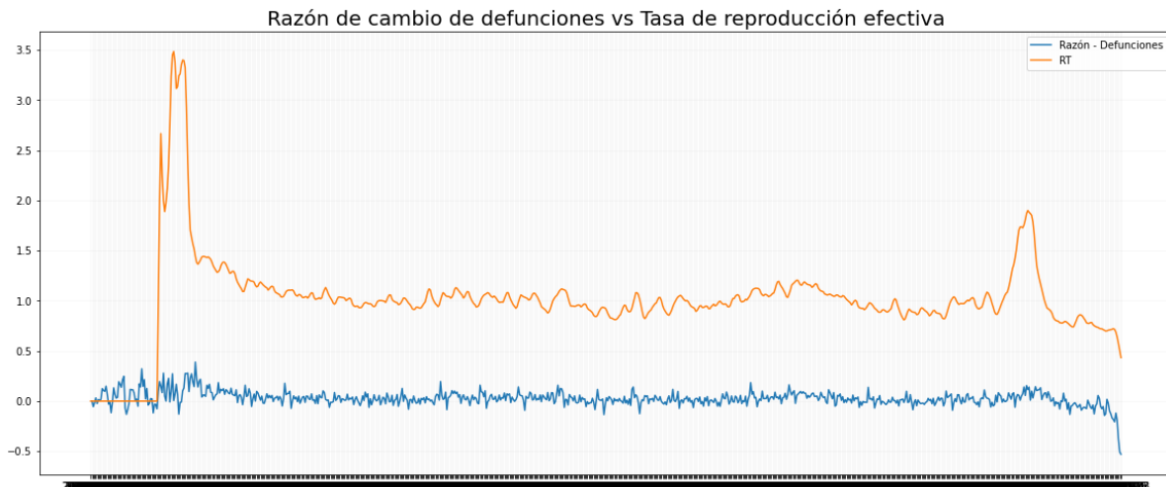
Razón de cambio de confirmados vs Rt



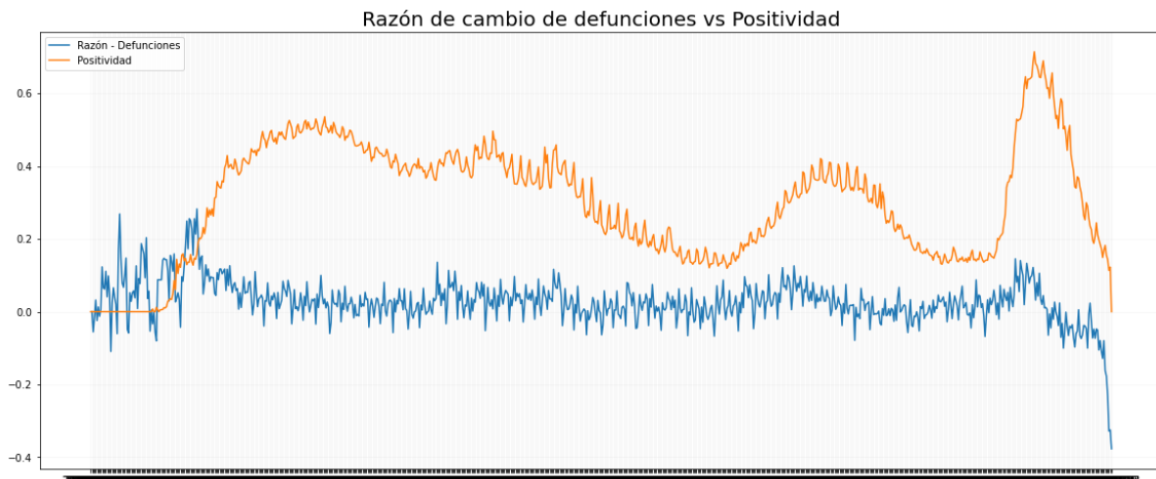
Razón de cambio de confirmados vs Positividad



Razón de cambio de Defunciones vs Rt



Razón de cambio de Defunciones vs Positividad



Análisis de clasificación de intubación

1. Filtrar México:

2. Reporte de Calidad de Datos:

Este reporte de calidad de datos nos ayuda bastante para poder identificar los tipos de datos de cada una de nuestras variables, si tenemos valores nulos y la cantidad de valores únicos por cada una.

3. Entendimiento de las variables:

Para poder comprender mejor nuestros datos decidimos hacer un diccionario que funcionara como catálogo para poder consultar el significado de cada categoría de nuestras variables.

Tratamiento de los datos:

Primero que nada lo que hicimos fue filtrar nuestros datos y quedarnos con un DataFrame que contuviera únicamente personas que en su variable de 'INTUBADAS' tuvieran un 1 o un 2, lo que quería decir que con seguridad podíamos decir que fueron o no intubadas, ya que dentro de esta variable teníamos otras categorías que no eran de nuestro interés, Después lo que decidimos hacer fue el **balance** de nuestros datos y cómo podemos observar nuestros datos estaban totalmente desbalanceados.

```
Total intubados: 114803 Total no intubados: 1069719
Porcentaje de intubados: 9.69 %
Porcentaje de no intubados: 90.31 %
```

Debido a esto fue que decidimos quedarnos con el 9.69% de nuestros datos donde si resultaron intubados y nos quedamos con la misma cantidad de datos donde no resultaron intubadas las personas para hacer tener un balance del 50% de los datos para cada resultado.

A continuación, lo que hicimos fue eliminar variables de acuerdo a los siguientes criterios:

Variables de tipo 'object' y que por criterio propio no consideramos relevantes

```
no_numericas = (dqr['Variable Name'][(dqr['Variable Type'] == 'object') | (dqr['Variable Type']=='category')]).values
# Debido a que estas variables son irrelevantes no las tomaremos en cuenta
df_total = df_total.drop(columns = list(no_numericas))
#df_total = df_total.drop(columns = 'MUNICIPIO_RES')
```

```
list(no_numericas)

['FECHA_ACTUALIZACION',
 'ID_REGISTRO',
 'FECHA_INGRESO',
 'FECHA_SINTOMAS',
 'FECHA_DEF',
 'PAIS_NACIONALIDAD',
 'PAIS_ORIGEN']
```

Variables con varianza igual a 0

```
df_var = pd.DataFrame({'Variable':np.array(df_total.var().index), 'Varianza':df_total.var().values})
df_var[df_var['Varianza']==0]
```

	Variable	Varianza
7	TIPO_PACIENTE	0.0
11	NACIONALIDAD	0.0
31	MIGRANTE	0.0

Variables altamente correlacionadas

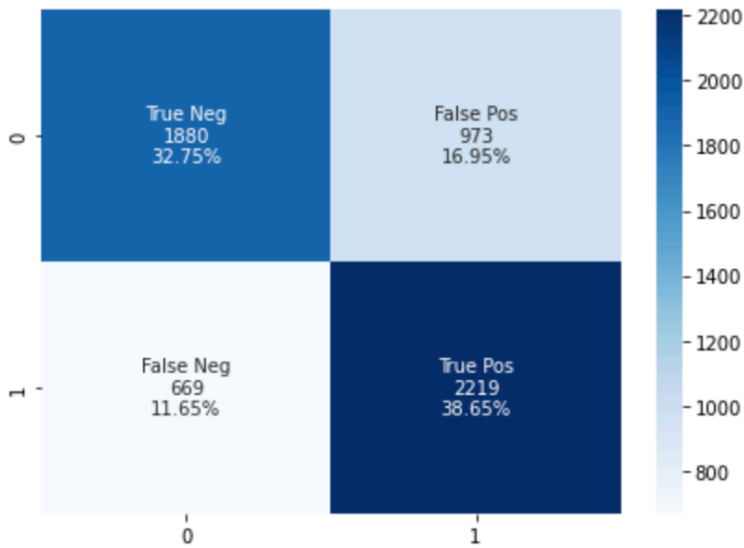
Una vez nos deshicimos de las variables que consideramos únicamente le estarían metiendo ruido a nuestros modelos decidimos volver algunas variables de tipo 'categoricas' (**ENTIDAD_RES, MUNICIPIO_RES, CLASIFICACION_FINAL, SECTOR**).

Modelos:

El primer modelo que decidimos probar fue el de **regresión logística**, este modelo tiene muy pocos hiperparámetros por lo que realmente no hicimos demasiadas pruebas y para este modelo en específico no utilizamos una grid para probar distintas combinaciones de hiperparámetros.

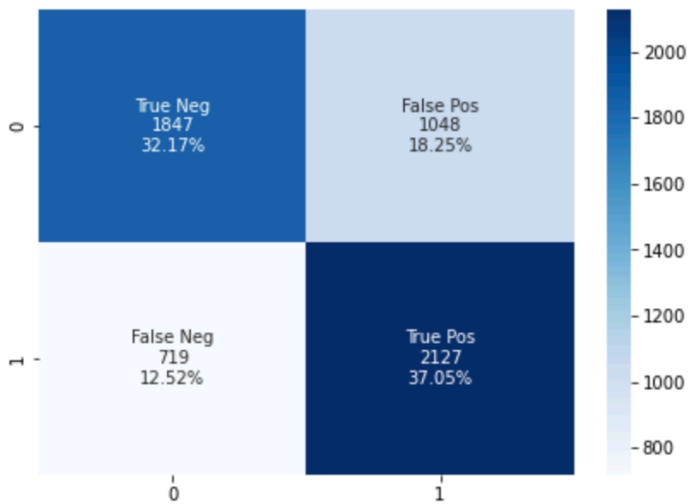
hicimos dos pruebas, la primera fue escalando nuestras variables y los resultados son los siguientes:

Exactitud: 0.714
Precisión: 0.7375
Recall: 0.659



Segunda prueba con nuestros datos sin escalar:

Exactitud: 0.6922
Precisión: 0.7198
Recall: 0.638



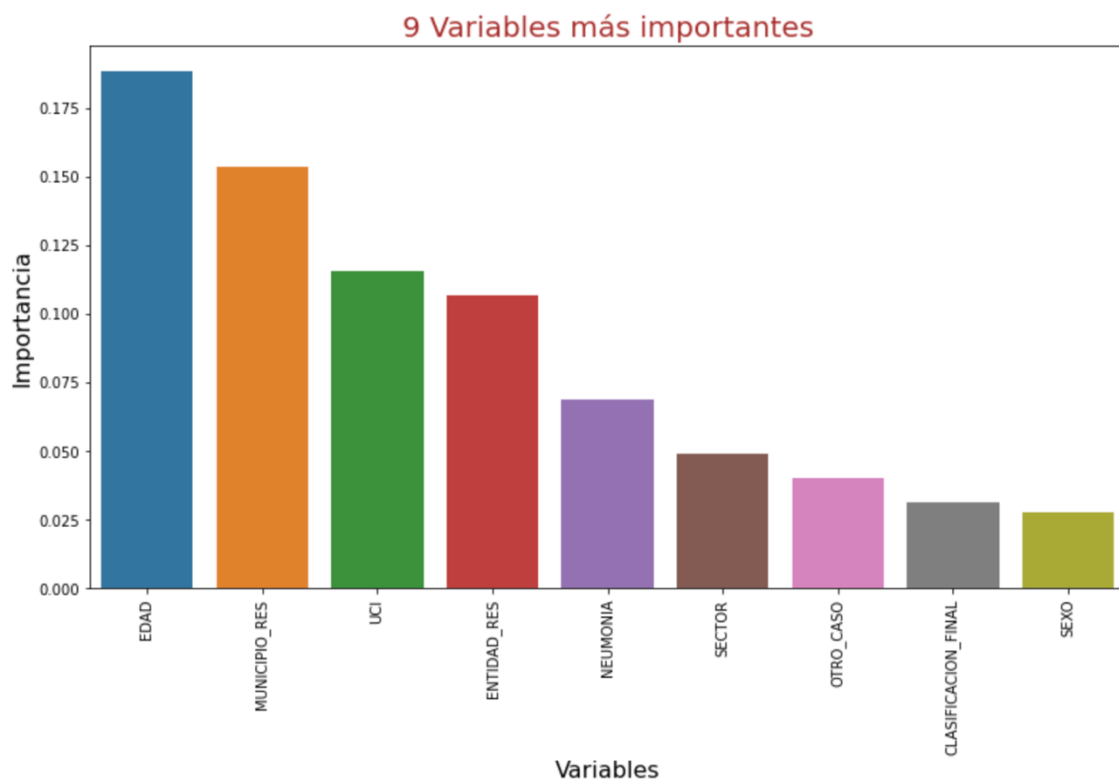
Como podemos apreciar nuestros resultados tienen una exactitud bastante baja y no varían prácticamente nada. El segundo modelo con el que decidimos intentar fue el de **Random forest** para este modelo si aplicamos una grid con una lista de valores para cada

hiperparámetro, este paso fue bastante tardado debido a la carga computacional, una vez el código terminó de correr obtuvimos que los peros valores eran los siguientes:

```
{'criterion': 'entropy', 'max_features': 10, 'n_estimators': 200, 'warm_start': True}
```

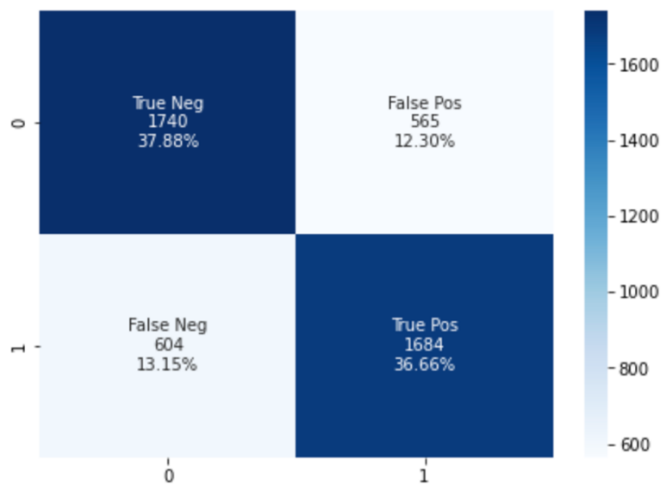
con estos parámetros la exactitud nos dió **0.7453**

y a continuación de este mismo modelo consultamos cuáles fueron las 9 variables más importantes, las cuales son las siguientes:



después decidimos correr nuevamente el código, pero esta vez tomando en cuenta solo las 9 variables con mayor importancia:

Exactitud: 0.7455
Precisión: 0.7423
Recall: 0.7549



como podemos observar mejoró el modelo muy ligeramente.

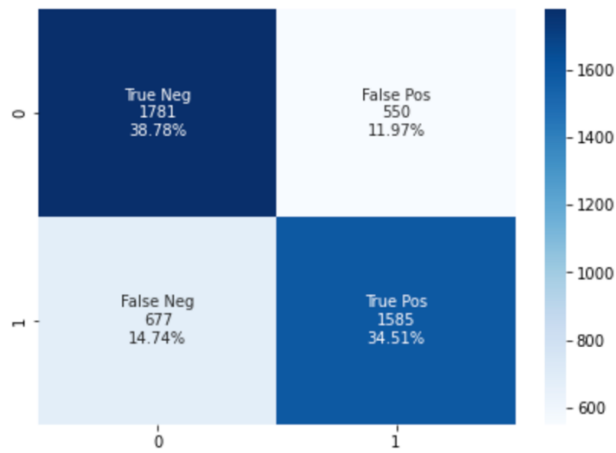
Por último el modelo que probamos fue el de **Perceptrón multicapa** y para este modelo los mejores parámetros que nos arrojó la grid fueron los siguiente:

```
{'activation': 'logistic', 'alpha': 1e-10, 'hidden_layer_sizes': (70, 2), 'max_iter': 1000, 'solver': 'adam', 'warm_start': True}
```

y estas fueron sus medidas de desempeño:

Exactitud: 0.7329
Precisión: 0.7246
Recall: 0.764

: <AxesSubplot:>



Otras pruebas:

1. Como podemos apreciar una vez nos deshicimos de todas las variables que no nos estaban aportando gran valor, realmente nos quedamos con **puras variables categóricas** a excepción de la variable de edad, por lo que decidimos volverlas de tipo 'category' (equivalentes a variables de tipo 'factor' en R) y decidimos correr el modelo de perceptrón, desafortunadamente nuestro modelo no mejoró.

2. Realizamos una técnica llamada 'Target Encoding' que lo que hace es codificar nuestras variables de tipo categóricas reemplazando cada una de las categorías de cada variable con el promedio de la variable objetivo. Como ya teníamos todas nuestras variables como tipo 'category' le aplicamos la técnica a todas nuestras variables a excepción claro de la edad. Nuevamente obtuvimos valores cercanos a los que ya habíamos conseguido

2.1 Decidimos volver al tipo de dato int a nuestras variables a excepción de: [SECTOR, ENTIDAD_RES, MUNICIPIO_RES, RESULTADO_LAB, CLASIFICACION_FINAL] (mantuvimos la técnica de target encoding en estas variables). Nuestros resultados no variaron

3. Pensamos que los valores de [97,98,99] que contenían algunas de nuestras variables podrían estar afectando nuestros modelos por lo que reemplazamos esos valores con unos más cercanos a los otros.

Probamos estos cambios con **Random Forest** y con **Perceptrón**, pero no obtuvimos resultados favorables.

Conclusiones:

El modelo que nos dio mejores medidas de desempeño fue **RANDOM FOREST** sin embargo no fueron las medidas que nosotros esperábamos alcanzar, ya que la máxima exactitud que obtuvimos fue de **.7521**, esta exactitud tan baja se la atribuimos considerablemente a que realmente es difícil predecir si una persona va a terminar INTUBADA o no cuando filtramos nuestros datos.

Ya que si recordamos una de las variables que eliminamos debido a que tenía varianza 0 fue la de 'TIPO DE PACIENTE' y si revisamos su catálogo nos podemos dar cuenta que es la variable que nos indica sobre si la persona está hospitalizada o no y al haber filtrado de esa manera la única categoría que tenemos dentro de esta variable es la categoría '2'

```
catalogo[ 'tipo_paciente' ]
```

	CLAVE	DESCRIPCIÓN
0	1	AMBULATORIO
1	2	HOSPITALIZADO
2	99	NO ESPECIFICADO

Por lo que una hipótesis es que si tomamos más clases para nuestra variable objetivo y en vez de hacer una clasificación binaria hacemos una clasificación multiclase podríamos obtener mejores medidas de desempeño ya que tendríamos otras variables que nos podrían aportar información valiosa.

1.6. Valoración de productos, resultados e impactos

El PAP nos permitió profundizar en como la pandemia ha impactado la vida de todos los mexicanos y prácticamente de todo el mundo. Y poder estudiar más a fondo como la enfermedad afecta a cada grupo de edad y a cada estado y con este poder ver como el virus se relaciona diferente dependiendo de las diferentes condiciones, no se va a reproducir igual el virus en una población donde hay pocas o nulas restricciones a donde hay un mandato obligatorio de cubrebocas, medida preventiva contra COVID y tampoco va a comportarse igual el virus en una persona adulta mayor a un joven de 25-30 años, así como las características cambian también cambia el comportamiento del virus.

Cada vez podemos conocer más a fondo como este virus afecta a la población en general y esto nos permite cada vez crear modelos más eficientes y que reflejen mejor la realidad, sin embargo, todavía tenemos camino por correr y es necesario obtener más datos con el fin de lograr mejores modelos que tengan valores más altos de atributos de desempeño, lo cual nos permitirá poder acercar los modelos más a la realidad.

1.7. Bibliografía y otros recursos

- How Charts Lie: Getting Smarter about Visual Information, Alberto Cairo, W. W. Norton & Company, 2019.
- Learning Tableau 2019: Tools for Business Intelligence, data prep, and visual analytics, 3rd Edition, Joshua N. Milligan, Packt Publishing Ltd, 2019.
- Storytelling Como Estrategia de Comunicación: Herramientas Narrativas Para Comunicadores, Creativos Y Emprendedores, Guillaume Lamarre, Edit Gustavo Gili, 2020.
- Visual Analytics with Tableau, Alexander Loth, John Wiley & Sons, 2018.
- Visual Data Storytelling with Tableau: Story Points, Telling Compelling Data Narratives, Lindy Ryan, Addison-Wesley Professional, 2018.
- Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole nussbaumer knaflic, John Wiley & Sons, 2015

Base de datos:

http://datosabiertos.salud.gob.mx/gobmx/salud/datos_abiertos/datos_abiertos_covid19.zip

La Salle. (2022). Plataforma para el monitoreo y pronóstico de la evolución del COVID-19 en México utilizando Inteligencia Artificial. Recuperado de: <https://monitoreocovid.lasalle.mx/#mx,us>

Secretaría de Salud Jalisco. (2021). Tasa de positividad de COVID-19 es de 34.2% al arrancar noviembre. Recuperado de: <https://ssj.jalisco.gob.mx/prensa/noticia/9464>

2. Productos

1. Reporte de la base de datos de SINAVE

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/7>

2. Reporte con visualizaciones para entender como el COVID ha afectado a los diferentes estados.

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/10>

3. Reporte comparando las diferentes olas de contagio que ha habido en el país así como el cálculo de RT.

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/13>

4. Reporte con visualizaciones de distribuciones por edad de cada ola.

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/16>

5. Reporte con análisis de tweets sobre palabras clave de la pandemia.

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/19>

6. Reporte con análisis de la razón de cambio de los casos confirmados.

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/22>

7. Reporte con análisis de clasificación de intubación.

<https://gitlab.com/dpmontoya/pap-mmd-p2022/-/issues/26>

3. Reflexión crítica y ética de la experiencia

El RPAP tiene también como propósito documentar la reflexión sobre los aprendizajes en sus múltiples dimensiones, las implicaciones éticas y los aportes sociales del proyecto para compartir una comprensión crítica y amplia de las problemáticas en las que se intervino.

3.1 Sensibilización ante las realidades

El análisis de datos siempre se me ha hecho muy interesante y es algo que me gusta hacer, pero creo que muchas veces dejamos de lado el componente humano a la hora de hacer las gráficas y/o modelos, lo que quiero decir es que cada punto, cada dato, cada positivo, cada defunción es una vida humana con todo lo que involucra, un caso positivo más en esa base de datos es una familia que está angustiada porque esa persona logra combatir el COVID, una defunción más es un papa, un hermano, un abuelo que ya no está en este mundo.

Una vez que ves con un lado más humano las cosas, resulta más complicado ver los datos y las estadísticas sobre todo de este proyecto en el que vimos los resultados que ha tenido la pandemia en la población mexicana, ver la tasa de mortalidad de los adultos mayores me hizo darme cuenta de lo afortunado que soy en que ninguno de mis abuelos haya fallecido por COVID durante esta pandemia.

Este proyecto me hizo darme cuenta que nunca hay que dejar de lado las realidades de las otras personas y no porque a mi alrededor no estén pasando cosas eso significa que tampoco estén pasando cosas en otros lados.

- (Jorge Zuñiga)

Realizar análisis de datos siempre trae consigo dilemas éticos porque durante el proceso se van reflejando situaciones injustas, en este caso trabajando con los datos de México sobre el COVID-19 y su propagación es fácil percibir que no todos los estados están actuando de la misma manera y estas decisiones afectan directamente en su población y la manera en que conllevo los contagios pero no sólo se refleja desde el punto de vista que los contagios crecen o disminuyen según el caso sino que también las defunciones y la tasa de positividad.

También me hizo reflexionar realizar el análisis por edades porque es evidente que las personas mayores son las que sufren más (mayores a 60 años) y por lo tanto es el grupo de la población que más muere, entiendo que no estén en su edad productiva pero desde mi punto de vista ninguna vida vale más que otra y es muy cruel que en la mayoría de los estados no se les este dando el énfasis que necesitan al tener más necesidades que cubrir.

- (Ivette Landaverde Mercado)

Al ir realizando las actividades fue bastante interesante e impactante ir indagando y respondiendo diferentes tipos de preguntas, ya que en distintos análisis pudimos darnos cuenta que nuestra sociedad fue muy vulnerable ante la situación del COVID-19 y que la mayor parte de la población fue afectada en varias ocasiones.

Al haber llegado a distintas conclusiones y al haber respondido numerosas preguntas mi sentimiento fue algo negativo ya que todo apuntaba a que las consecuencias que cada una de las personas pagó se debió a algo que se salió de control mayormente porque no se hizo un tratamiento adecuado de la situación y porque fue algo nuevo para cada uno de nosotros y nos tomó de improviso.

Todo este sentimiento se lo atribuyo a mis razonamientos al final de cada uno de los análisis que realicé y cada vez que mis compañeros fueron exponiendo de igual manera sus hallazgos.

- Tonatiuh Navarro

3.2 Aprendizajes logrados

La experiencia del PAP me ayudó a aterrizar ideas y conceptos aprendidos durante los semestres previos de Ingeniería Financiera, además de adquirir nuevos conocimientos técnicos gracias a las sesiones de aprendizaje.

El trabajar en equipo siempre resulta ser un reto porque es pensar en conjunto y no solamente como un individuo, habilidad que me llevo para desempeñar una mejor vida laboral.

Además, todo el proceso del proyecto y las sesiones para compartir lo trabajado cada semana fueron muy enriquecedoras, aprendí a ver el mismo problema desde distintas perspectivas y fue interesante ver como cada equipo tenía una propuesta o una manera de trabajar la misma información.

En cuanto a los propósitos del proyecto descritos en la primera sección del reporte PAP, honestamente creo que cumplí todos, desde desarrollar visualizaciones de información y análisis de datos a través de machine learning hasta desarrollar habilidades de escritura y comunicación para poder transmitir los conocimientos y el trabajo realizado en cada reporte semanal.

- (Ivette Landaverde)

Durante la mayoría del PAP he aprendido a aplicar algunas cosas que hemos visto en la carrera de Ingeniería financiera, como el análisis de datos, pero también aprendí nuevas cosas y técnicas como análisis de texto o programas como gitlab. A mi parecer lo más complejo es trabajar en equipo debido a que cada persona tiene sus tiempos y es necesario mucha comunicación para poder entregar los trabajos en tiempo y forma, pero esto es lo más cercano que hay a la realidad laboral donde constantemente estas trabajando de manera colaborativa.

Como lo dice el nombre son practicas de aplicación profesional donde abordamos todo un proyecto desde el inicio, desde la limpieza de los datos, la exploración de los mismos y ahí buscar extraer el mayor valor posible de la información que tenemos actualmente, hasta desarrollar modelos de predicción para el problema que es presentado, me pareció muy interesante tanto y gracias a este

proyecto me di cuenta que me gustaría profundizar más profesionalmente en el ámbito de los datos, ya que actualmente trabajo más en la parte de finanzas.

- (Jorge Zuñiga)

Cada uno de los trabajos en equipo que realizamos a lo largo de este proyecto de aplicación fueron resultado de horas de estudio de diferentes materias que cursamos donde estudiamos la teoría, realizamos algunos ejercicios, pero al fin pudimos implementar estos saberes en algo de la vida real y en un tema de gran interés en estos momentos, este PAP me fue de gran valor ya que más que sentirlo como una materia lo sentí como si hubiese estado involucrado en un proyecto laboral, y me parece excelente y bastante valioso el hecho de que este proyecto enriquezca nuestro currículum y podamos utilizarlo como referencia en futuras entrevistas de trabajo.

Mis intereses laborales están principalmente enfocados en el área de ciencia de datos y todas las actividades que realicé junto con mi equipo son bases en esta área, por lo que a pesar de que en algunas ocasiones me resultaban un poco retadoras, me sentía motivado y siempre me esforcé para que los trabajos quedaran lo mejor posible.

A pesar de que en la mayor parte del proyecto únicamente apliqué conocimientos que ya tenía, también aprendí y reforcé otras habilidades como lo es el trabajo en equipo, la buena comunicación, el tener una escucha activa y el liderazgo.

Tonatiuh Navarro