

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE**  
**Departamento de Matemáticas y Física**

**Sustentabilidad y tecnología**

**PROYECTO DE APLICACIÓN PROFESIONAL (PAP)**  
**Programa de Análisis de Datos COVID-19**



**ITESO, Universidad**  
**Jesuita de Guadalajara**

**1732Q**

**Programa de Modelación Matemática para el Desarrollo de Planes y Proyectos de**  
**Negocio**

**PRESENTAN**

Ingeniería Financiera. Miguel de Jesús Aguirre Orozco

Ingeniería Financiera. Moisés Flores Ortiz

Ingeniería Financiera. Sean Nicolás González Vázquez

Ingeniería Financiera. Miriam Eunice Rosas Medellín

Profesor PAP: Diana Paola Montoya Escobar

Tlaquepaque, Jalisco, mayo de 2022

## Tabla de contenido

Presentación Institucional de los Proyectos de Aplicación Profesional .....	1
Resumen .....	2
1. Ciclo participativo del Proyecto de Aplicación Profesional.....	3
1.1 Entendimiento del ámbito y del contexto .....	3
1.2 Caracterización de la organización.....	4
1.3 Identificación de la(s) problemática(s).....	6
1.4. Planeación de alternativa(s).....	8
1.5. Desarrollo de la propuesta de mejora .....	9
1.5.1. Bloque 1. Entender el problema .....	9
1.5.1.1. Actividad 1. Entender los datos.....	9
1.5.2. Bloque 2. Análisis de datos .....	14
1.5.2.1. Actividad 2. Positividad y Rt .....	14
1.5.2.2. Actividad 3. Análisis comparativo de municipios y olas de contagio ....	24
1.5.2.3. Actividad 4. Distribución de edad por olas .....	38
1.5.2.4. Actividad 5. Análisis de texto .....	40
1.5.2.4.1. Covid-19 Jalisco .....	41
1.5.2.4.2. Cubrebocas Jalisco y México .....	44
1.5.2.4.3. Pandemia Jalisco y México.....	49
1.5.2.4.4. Vacunación Jalisco y México .....	53
1.5.2.5. Actividad 6. Razones de cambio .....	58
1.5.3. Bloque 3. Modelado .....	63
1.5.3.1. Actividad 7. Análisis de separabilidad de variables t-SNE.....	63
1.5.3.1.1. Año 2020.....	64
1.5.3.1.2. Año 2021.....	66
1.5.3.1.3. Año 2022.....	67
1.6. Valoración de productos, resultados e impactos .....	70
1.7. Bibliografía y otros recursos .....	71
2. Productos .....	72
3. Reflexión crítica y ética de la experiencia.....	73
3.1 Sensibilización ante las realidades .....	74
3.2 Aprendizajes logrados .....	75

## Presentación Institucional de los Proyectos de Aplicación Profesional

*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) (coparticipación) 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 PAP Programa de Modelación Matemática para el desarrollo de Planes y Proyectos de negocio para el periodo primavera 2022 tiene como título Analítica avanzada para Covid-19 para el Estado de Jalisco y su propósito general consiste en el desarrollo de herramientas de visualización de datos, así como la implementación de técnicas de *machine learning* para realizar predicciones, clusters y análisis para toma de decisiones.

Los objetivos consisten en el desarrollo de la capacidad de trabajo en equipo; desarrollar habilidades de lectura y escritura en lenguaje matemático y ordinario; examinar información de diversas fuentes bibliográficas y manejar tecnologías de la información como herramientas auxiliares para la solución de problemas. Los objetivos antes mencionados se pretenden alcanzar a través de actividades específicas y relacionadas con el análisis del Covid-19 para el Estado de Jalisco.

La metodología se centra en un análisis exploratorio de datos históricos de Covid-19 recopilados en la base de datos del gobierno federal y del gobierno del estado de jalisco. Además del uso de herramientas estadísticas e informáticas: R, Python, Tableau y algunas otras. El resultado alcanzado es un dashboard/bot que muestre la situación actual de la pandemia en Jalisco y la generación de un modelo de predicción a través de variables correlacionadas para la toma de decisiones.

Finalmente, este proyecto tiene como beneficio potencial mejorar la toma de decisiones por medio del modelo predictivo generado, para evitar la saturación hospitalaria, la propagación del virus y la afección de la economía del Estado de Jalisco.

## 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 construyen 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.

### 1.1 Entendimiento del ámbito y del contexto

El Covid-19 es un virus que ha afectado al estado de Jalisco y al mundo entero en términos de salubridad, económicos, sociales, financieros, entre otros. Es por esto, que con el fin de llevar un seguimiento de cómo se han desarrollado los casos del virus en el estado, se ha buscado generar bases de datos, tanto desde el ámbito federal (por medio de la base de datos de SINAVE), como en el estatal (con la base de datos de RADAR JALISCO). Para lograr entender como el virus ha afectado a la sociedad de Jalisco, en este proyecto buscaremos usar las bases de datos históricos para obtener *insights* acerca del virus.

Entendiendo los estragos del virus, podemos proponer una metodología para la realización de este proyecto, buscando resolver los problemas que serán planteados, se harán una serie de actividades que nos facilitarán entender el virus, los casos y todo lo relacionado específicamente en el contexto de México, con el fin de poder plantear y desarrollar modelos que nos permitan ver que tan separables son las variables además de poder predecir el comportamiento del virus.

Hablando del contexto, en México desde el inicio de la pandemia ha habido un total de 5.73 millones de casos, con 324,00 fallecidos totales, lo cual aproximadamente nos habla de una gran pandemia, en la cual gran parte de la población se ha contagiado y es probable que posteriormente se contagien aún más, si bien, queremos saber cómo se ha comportado el virus, podemos decir que las políticas públicas del gobierno de México han logrado reducir la probabilidad de contagiarse. Esto también se buscará comprobar en el proyecto por medio de los datos oficiales.

Este gran número de contagios, como se comentó previamente ha traído consigo estragos socioeconómicos en el país, con un 30% de la población total mexicana en condiciones de desempleo y el producto interno bruto del país sufriendo una contracción. En conclusión, es de vital importancia para el estado entender como se ha comportado el virus históricamente, además de cómo se espera que se comporte.

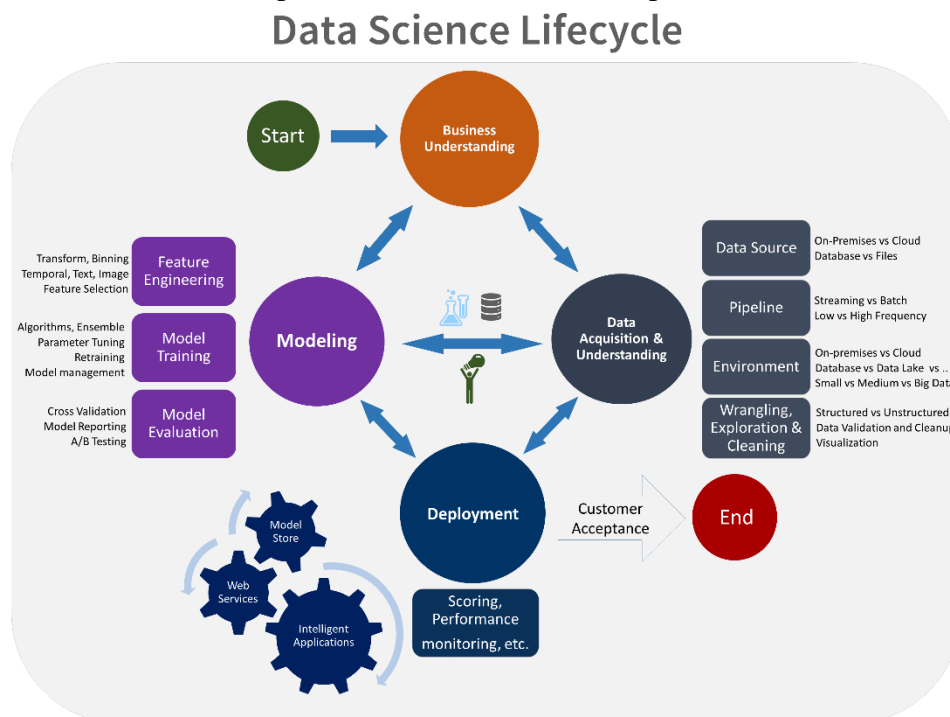
Conociendo el contexto y las condiciones en las cuáles se está desarrollando el proyecto podemos entender la importancia del mismo, buscamos ayudar al gobierno, en colaboración con el área de Inteligencia Artificial del gobierno de Jalisco a analizar los datos del Covid-

19 para así formular mejores políticas públicas en todos los ámbitos (social, económico, etc), para posteriormente buscar propiciar mejores condiciones para la población afectada por el virus.

## 1.2 Caracterización de la organización

En términos de la organización para el desarrollo del proyecto se tomó la metodología planteada por Microsoft denominada Team Data Science Process (TDSP) la cual permite proporcionar soluciones de ciencia de datos de manera eficiente, también ayuda a mejorar la colaboración y el aprendizaje de los miembros del equipo asignando a cada uno un rol específico que en conjunto genera un funcionamiento óptimo para la generación de soluciones.

De acuerdo con los componentes del TDSP la parte operativa para el desarrollo del proyecto se centra en el ciclo de vida de ciencia de datos donde están asignados cada uno de los roles de los participantes del proyecto, en este sentido las áreas funcionales son: conocimiento de negocio, adquisición y comprensión de los datos, modelado e implementación. Estas áreas están interconectadas en un flujo iterativo desde el comienzo hasta el final del proyecto. A continuación, se muestra la representación visual de dicho proceso:



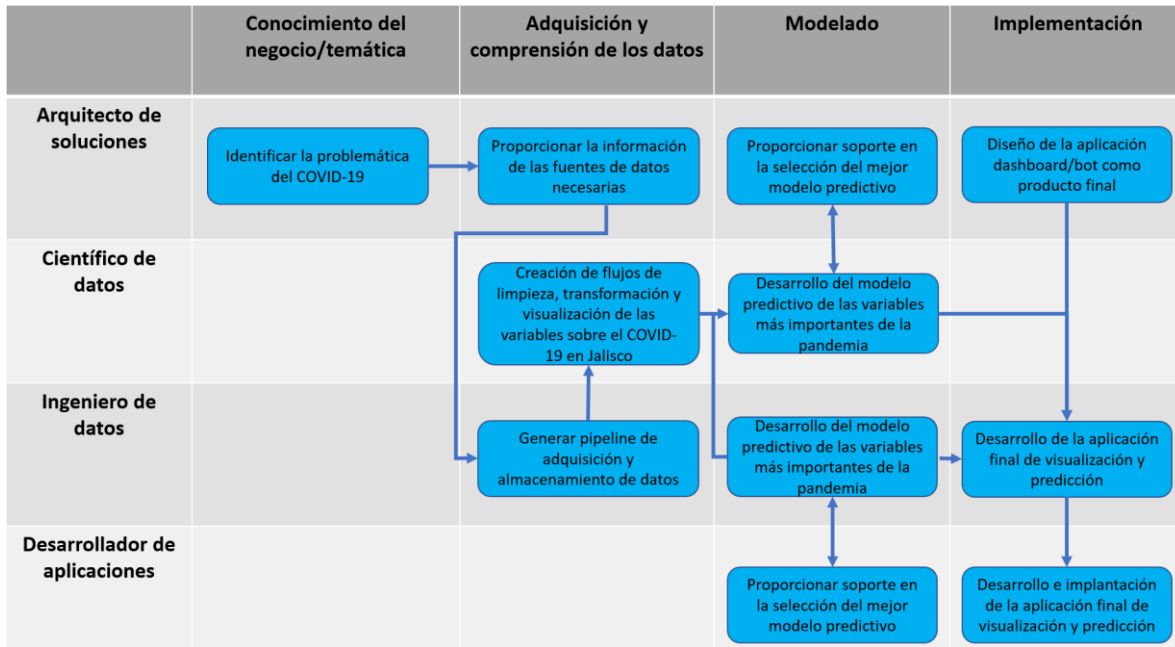
Las responsabilidades de las áreas funcionales descritas anteriormente aplicadas al proyecto están especificadas en la siguiente tabla:

<b>Rol</b>	<b>Función</b>
Conocimiento del negocio/temática	Identificar todos los elementos de la problemática del Covid-19 – 19, los antecedentes históricos y las implicaciones actuales de la situación en el Estado de Jalisco. Además, comprender los resultados de los análisis y modelos para proporcionar información que ayude a la toma de decisiones.
Adquisición y comprensión de los datos	Conocer las fuentes de datos públicos que proporciona el Gobierno Federal y del Estado de Jalisco. Implementar técnicas para la adquisición óptima de los datos, así como de limpieza y exploración de los datos de la pandemia. Y generar pipelines para la transformación de los datos y su posterior visualización.
Modelado	Aplicar técnicas estadísticas identificar las variables que tengan correlación en el comportamiento del Covid-19, así mismo, generar modelos predictivos de las variables más importantes en cuanto a la salud y la economía de la población.
Implementación	Desarrollar y diseñar la aplicación final (dashboard/ bot) en dónde se presentarán los datos más relevantes sobre el comportamiento de las variables del Covid-19 y los respectivos resultados de los modelos predictivos generados.

El equipo se conforma de 4 roles aginados a cada uno de los integrantes:

- **Arquitecto de soluciones:** Miguel De Jesús Orozco Aguirre.
- **Científico de datos:** Sean Nicolas González Vázquez.
- **Ingeniera de datos:** Miriam Eunice Rosas Medellín.
- **Desarrollador de aplicaciones:** Moisés Flores Ortiz.

El siguiente diagrama explica la intervención de cada uno de los roles en las áreas del ciclo de vida de ciencia de datos:



En cuanto a la estructura de proyecto estandarizada se utiliza como software GitLab donde se almacenan los entregables a través de una estructura numerada de *issues*, ahí se guardan los archivos documentales de visualización y el código fuente o cualquier recurso utilizado en el desarrollo de las etapas del proyecto.

Finalmente, los recursos e infraestructura utilizados en el proyecto se dividen en 3 ramas: almacenamiento, en la cual se utiliza un fichero de archivos delimitados por comas o Excel donde se encuentran los registros de la situación de la pandemia; la parte de código en la que se utiliza Jupyter para las implementaciones de Python y R Studio para las respectivas en R; y por último software de visualización: Tableau.

### 1.3 Identificación de la(s) problemática(s)

El Covid-19 es una enfermedad infecciosa que afecta las vías respiratorias, para entender la gravedad de la enfermedad, esta fue la primera pandemia de este milenio. El virus comenzó a expandirse desde China y debido a la globalización llegó a todos los continentes en un período aproximado de dos meses.

Debido a esto, a nivel mundial se han establecido medidas sanitarias como el confinamiento, la aplicación de vacunas, el uso de cubrebocas, el distanciamiento social y la digitalización en diferentes áreas. Debido al confinamiento diferentes vertientes socioeconómicas han sido afectadas, personas de bajos y medianos recursos perdiendo su estabilidad económica, al mismo tiempo afectando cadenas de suministro, transporte y actividades que requieren presencialidad. Para aterrizar este problema a nivel estatal, según un estudio del ITESO, el



51% de los habitantes económicamente activos han visto una disminución en el ingreso familiar, además de que casi un 30% se encuentra desempleada.

No solo el ámbito económico ha sido afectado, en el sector salud, la ocupación hospitalaria en el Estado se ha visto saturada varias veces, a finales del 2020 se registró en 9 de los 14 hospitales públicos más grandes de Jalisco una saturación por arriba del 50%. Posteriormente en agosto del 2021, la ocupación en lugares públicos era del 68.8%. También, se han visto afectados otros ámbitos como el psicológico, con el aumento de casos de depresión, ansiedad, entre otros.

En nuestro estado, nos encontramos actualmente en la cuarta ola de la pandemia, esto quiere decir que es la cuarta vez que existe un gran aumento en la velocidad de contagios, como se muestra en la siguiente gráfica.



En conclusión, esta pandemia ha afectado muchísimos sectores a nivel mundial y ha marcado un antes y un después en la dinámica de las actividades socioeconómicas. En específico en Jalisco, ha traído crisis en diferentes sectores, en salud, economía e incluso psicológicamente, por lo cual, nuestro equipo considera de alta importancia realizar este proyecto para entender diferentes aspectos del virus y de cómo ha afectado a la población de Jalisco.

### Principales preguntas a resolver

- ¿Las vacunas lograron disminuir el número de defunciones y hospitalizados en Jalisco?
- ¿Las vacunas han disminuido los contagios en esta cuarta ola en Jalisco?
- La última sepa del virus (ómicron), ¿tiene menos letalidad que sus predecesoras?
- ¿Cuál ha sido el sector económico más afectado por el virus en Jalisco?
- ¿Cuál ha sido la comorbilidad más común en los fallecidos en Jalisco?

- ¿Como le ha ido a Jalisco en la razón contagiados y defunciones contra los otros estados de la República?
- ¿Han ayudado las medidas de distanciamiento social en Jalisco a evitar contagios?
- ¿Disminuyó la contaminación ambiental en Jalisco durante la pandemia?

#### 1.4. Planeación de alternativa(s)

El Covid-19 en el Estado de Jalisco ha tenido grandes estragos en el ámbito de salud del estado, es por esto, que tanto el estado como el gobierno mexicano han desarrollado un sistema de seguimiento para los casos, para lograr esto, el Estado de Jalisco desarrolló la base de RADAR-JALISCO, mientras que, a nivel federal se desarrolló SINAVE, en este proyecto se usaron ambas herramientas para intentar responder las preguntas planteadas en el apartado anterior. Esto se hizo con el fin de tener un proyecto respaldado por bases históricas oficiales, en base a estas, se buscará solucionar nuestra problemática haciendo una serie de actividades de manera cronológica, como se muestra en la siguiente tabla.

ISSUES DEL PROYECTO	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10	Semana 11	Semana 12	Semana 13	Semana 14	Semana 15	Semana 16	Semana 17
	22-ene	29-ene	05-feb	12-feb	19-feb	26-feb	05-mar	12-mar	19-mar	26-mar	02-abr	09-abr	16-abr	23-abr	30-abr	07-may	14-may
<b>BLOQUE 1: Entender el problema y los datos</b>																	
Entender el problema de COVID, principales preguntas que están surgiendo.																	
Entender los datos de SINAVE																	
Limpieza y transformación de los datos																	
<b>BLOQUE 2: Análisis de datos</b>																	
Análisis general, compartativa 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																	
<b>BLOQUE 3: Modelados</b>																	
<b>Pronóstico de casos confirmados, hospitalizados</b>																	
Modelado de covid (regresión y/o clasificación)																	
<b>BLOQUE 4: Entrega del proyecto</b>																	
Reporte PAP																	
Presentación PAP																	

Como se puede observar, se dividió el proyecto en cuatro bloques, el primero consiste en entender el problema y las actividades consisten puramente en el entendimiento y la limpieza de los datos, mientras que el segundo bloque consiste el análisis de los datos dividido en seis problemáticas específicas (Análisis General, Análisis entre olas de contagio, Análisis de casos hospitalizados, Razones de cambio, Análisis por grupos de edad y finalmente el análisis de texto), entender estas problemáticas se resolverá con una serie de actividades explicadas a detalle en el siguiente apartado, en el bloque tres, tendremos el modelado de los casos de Covid-19, para entender mejor como estos datos se agrupan y cómo se comportan. El último

bloque consta de la entrega y la presentación del proyecto. Como se comentó anteriormente, en el siguiente apartado se buscará explicar con más detalle cada una de las actividades/entregas generadas y que preguntas se contestaron, así mismo que los problemas que se pudieron explicar.

## 1.5. Desarrollo de la propuesta de mejora

### 1.5.1. Bloque 1. Entender el problema

#### 1.5.1.1. Actividad 1. Entender los datos

#### **Introducción**

El análisis de datos es la ciencia encargada de examinar un conjunto de datos con el propósito de sacar conclusiones sobre la información para una correcta toma de decisiones o simplemente ampliar el conocimiento sobre un tema en específico. En este caso, usaremos el EDA (Exploratory Data Analysis) que genera visualizaciones y gráficos para generar *insights*.

Para esta segunda entrega, se buscará por medio de las bases de datos dadas de Radar Jalisco, hacer un EDA, en el cual se contestarán las diferentes preguntas planteadas en la actividad anterior, generando graficas e *insights* para esto.

#### **Preguntas a contestar**

- ¿Cuál ha sido el mes durante toda la pandemia en el cual ha habido más contagiados en Jalisco?
- ¿Cuántas olas de Covid-19 ha habido en el Estado?
- ¿Cuál es la tasa de Positividad histórica en Jalisco?
- ¿Cuál es el sistema de salud que ha detectado una mayor cantidad de casos?
- ¿Cuál es el sexo que ha predominado en contagios en Jalisco?
- ¿Cuál es el grupo de edad con mayor probabilidad de contraer el virus SARS-CoV-2?
- ¿Cuál es el tipo de distribución de la edad de los contagiados?
- ¿Qué porcentaje de personas hospitalizadas por síntomas de Covid-19 son confirmadas?
- ¿Cuántos pacientes hospitalizados ha habido en Jalisco por día?
- ¿Ha bajado el número de hospitalizados desde que se empezó la campaña de vacunación en Jalisco?
- ¿Cuál ha sido el mes durante toda la pandemia en el que ha habido más ocupación hospitalaria por Covid-19 en Jalisco? ¿Concuerda con la información presentada acerca de la saturación de hospitales en el estado?

## Procedimientos

- **Limpieza de datos de Radar Jalisco:** Se limpiaron los valores nulos, se renombraron columnas para mejorar la visibilidad, además de establecer un índice para todas las variables.
- **Transformación de los datos:** Se llenan espacios blancos y se modifica de ser necesario las variables de la información para que se pueda trabajar de forma más sencilla con ella.
- **Exploración de datos:** Se analiza como quedaron nuestros datos, y por medio de una lluvia de ideas se piensa en que *insights* y visualizaciones se pueden generar con los datos dados.
- **Visualización de datos:** Se generan gráficas e *insights* que busquen responder las preguntas planteadas en el punto anterior, con el objetivo de tener una forma más sencilla de entender los datos.

## Resultados

La primera visualización generada, fue con el objetivo de responder las preguntas ¿Cuál ha sido el mes durante toda la pandemia en el cual ha habido más contagiados en Jalisco? y ¿Cuántas olas de Covid-19 ha habido en el Estado?, como se observa a continuación, podemos ver que actualmente estamos en la cuarta ola, además de que ahorita la cantidad de casos positivos está en el pico histórico para México de la pandemia.

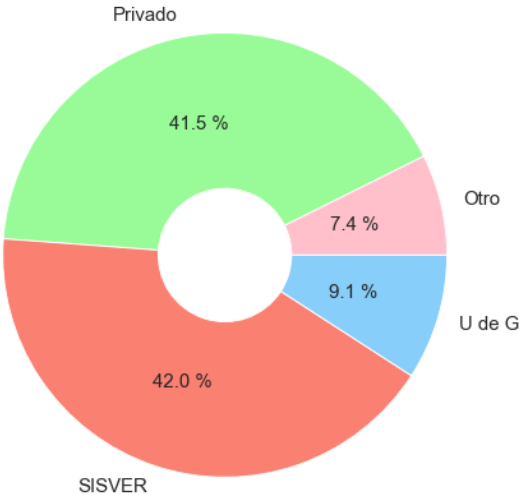


Para la siguiente visualización buscamos encontrar tanto la tasa positiva diaria de Jalisco como el *insight* de la tasa de positividad general en el estado, la última nos dio un valor de 20.62%, lo cual quiere decir que de acuerdo a la información histórica de las personas que se han hecho la prueba, el 20.62% sale positivo, pero, si segmentamos esta probabilidad de manera diaria, encontramos que la tasa de positividad ha sido la siguiente, lo cual va de acuerdo a la gráfica anterior, al tener en la actual ola un pico de positividad.

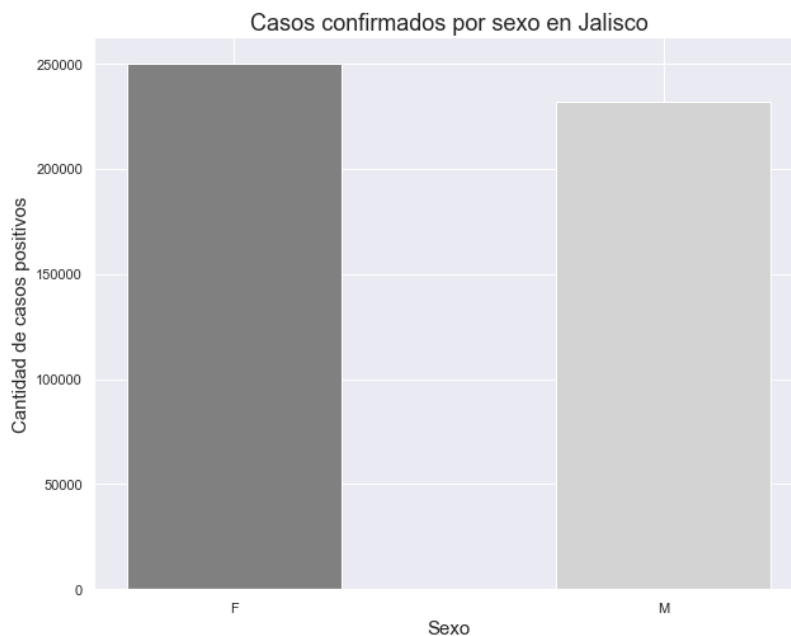


Al buscar contestar la pregunta de ¿Cuál es el sistema de salud que ha detectado una mayor cantidad de casos?, encontramos que el Sistema de Vigilancia Epidemiológica de Enfermedades Respiratorias es el que ha encontrado casos confirmados de Covid-19 en Jalisco.

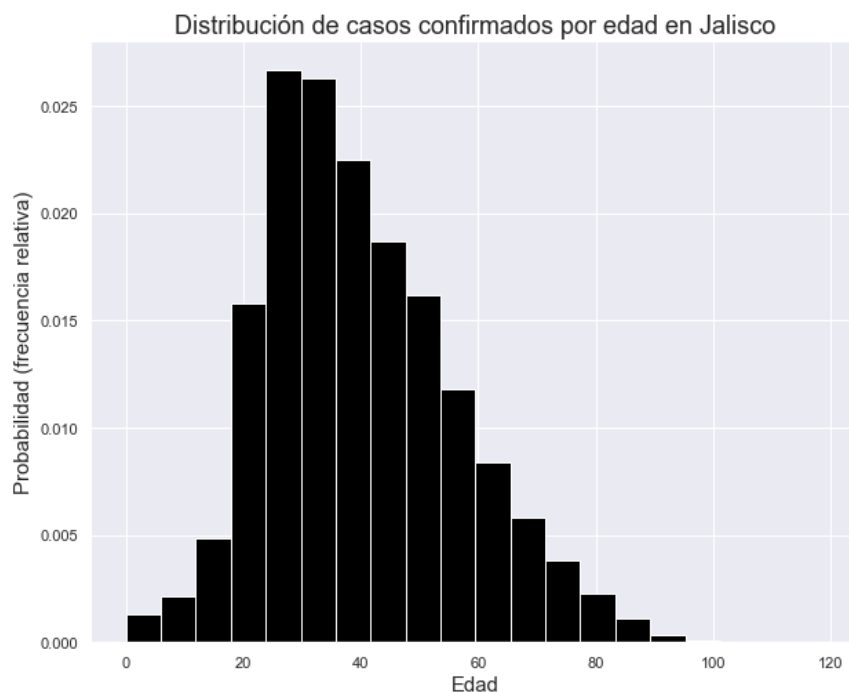
Casos confirmados por laboratorio en Jalisco



Posteriormente, se buscó dividir los casos por sexo y encontramos que las mujeres son el sexo que mayor número de casos positivos ha tenido. Es probable que esto se deba a que, según INEGI, existe una mayor cantidad de mujeres que de hombres en el estado.



En esta gráfica encontramos la división de casos por edad, como podemos observar la probabilidad de ser un caso confirmado sigue una distribución uniforme, el virus suele tener la misma probabilidad de aparecer en todas las generaciones, quitando los outliers de mayores de 100 años.



Finalmente, para responder el último grupo de preguntas siendo estas ¿Qué porcentaje de personas hospitalizadas por síntomas de Covid-19 son confirmadas?, ¿Cuántos pacientes hospitalizados ha habido en Jalisco por día?, ¿Ha bajado el número de hospitalizados desde que se empezó la campaña de vacunación en Jalisco?, ¿Cuál ha sido el mes durante toda la pandemia en el que ha habido más ocupación hospitalaria por Covid-19 en Jalisco?, ¿Concuerda con la información presentada acerca de la saturación de hospitales en el estado?. Realizamos una visualización con el número de pacientes hospitalizados en Jalisco por día. Además, encontramos que el porcentaje de personas hospitalizadas por Covid-19 que efectivamente son casos confirmados es de 60.82%.

Al buscar responder las otras preguntas, confirmamos que a pesar de que estamos en el pico de número de casos confirmados, al estar ya avanzada la campaña de vacunación el número de hospitalizados ha disminuido a comparación de otras olas. Además, la sobresaturación de hospitales que hubo en 2020 y 2021 cuadra con las olas vistas, siendo a principios de 2021 el periodo con más hospitalizados en Jalisco.



## Conclusiones

Los datos coinciden con las tendencias nacionales, además de comprobar varios de nuestros objetivos iniciales como qué tan eficientes han sido las medidas de aislamiento, el impacto que han tenido las vacunas y estadísticas de los casos confirmados, así como su segmentación por edad y sexo. Para los siguientes pasos buscaremos explorar más a detalle cada una de las bases e información que tenemos, además de aplicar algoritmos para encontrar tendencias, hacer *forecast* y agrupar los datos.

## 1.5.2. Bloque 2. Análisis de datos

### 1.5.2.1. Actividad 2. Positividad y $R_t$

#### **Introducción**

En esta actividad, buscaremos encontrar diferentes *insights* acerca de los números de defunciones, la tasa de mortalidad, positividad y “ $R_t$ ”, específicamente en el estado de Jalisco, con el fin de ver cómo está el estado para diferentes grupos ya sea por edad, sexo, intubados, entre otros. Haremos uso de las dos bases de datos (SINAVE y Radar Jalisco), debido a que esta última no contiene información necesaria para calcular el número de defunciones y la tasa de mortalidad. Buscaremos dividir este trabajo en dos, la primera en la cual se calculan las defunciones y la mortalidad haciendo uso de la base del SINAVE, y la segunda para calcular la positividad y la “ $R_t$ ” usando la base de Radar Jalisco. Se reitera que este análisis será únicamente para Jalisco.

Empezaremos con algunas definiciones básicas y necesarias, primero las defunciones son la cantidad de muertos por Covid-19, mientras que la tasa de mortalidad es la tasa de personas que fallecen por Covid-19, dividiendo este número entre la cantidad de positivos.

La tasa de positividad se refiere al porcentaje de personas que son confirmadas como positivas al Covid-19, de todas aquellas que se sometieron a una prueba de cualquier tipo. Este índice nos ayuda a identificar esas “olas de contagios” y poder tomar las medidas y restricciones necesarias para evitar una propagación prolongada. Mientras que el “ $R_t$ ” se refiere a la Tasa de Reproducción Efectiva del Virus la cual nos indica el número de personas que se contagia por cada persona que da positivo.

#### **Preguntas a contestar**

- ¿Cuáles son los periodos de tiempo donde mayor número de defunciones hubo por día?
- ¿Cuál es el sexo que tiene el mayor número de defunciones?
- ¿Cuáles son las enfermedades más comunes que pudieron agravar más la salud entre las defunciones?
- ¿Cuáles son los sectores de edad que contienen el mayor número de defunciones?
- ¿Cuál es la relación que existe entre estar intubado o no y la cantidad de defunciones?
- ¿Cuál es la tasa de mortalidad de los hospitalizados y no hospitalizados?
- ¿Cuál es la tasa de mortalidad diaria a nivel Jalisco?
- ¿Qué sexo tiene la mayor tasa de mortalidad en Jalisco?
- ¿Cuál es la tasa de positividad diaria a nivel estatal?
- ¿Cuál es la tasa de reproducción efectiva diaria en el estado?
- ¿Cuál es el sexo que tiene mayor representación en la tasa de reproducción efectiva del virus?



## Procedimientos

- **Limpieza de bases de datos de Radar Jalisco y de SINAVE:** Se eliminaron las columnas con información innecesaria, se renombraron las columnas y algunas de las variables de resultados.
- **Transformación de los datos:** Se agruparon los datos por fecha, sexo, grupo de edad, intubado y hospitalizado. Hicimos coincidir fechas y llenado de espacios blancos.
- **Cálculo de variables:** Con los datos agrupados se calcularon las variables de defunciones, mortalidad, positividad y la tasa de reproducción efectiva del virus.
- **Visualización de datos:** Se generan gráficas de los datos y con estas visualizaciones poder entender y responder las preguntas a contestar.

## Resultados

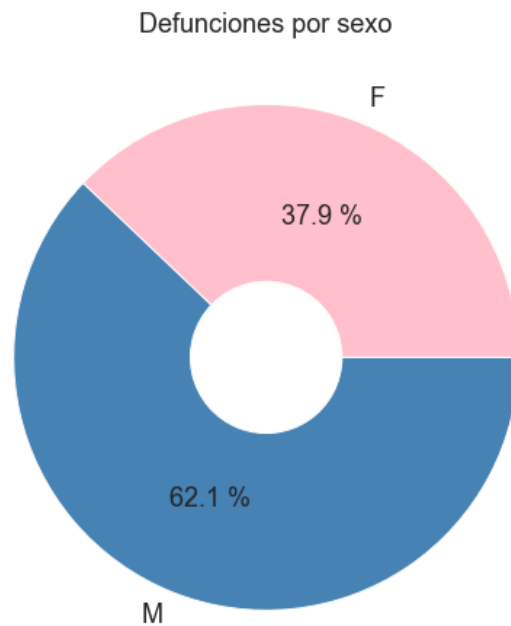
La primera gráfica desarrollada nos ayuda a responder cual es el número de defunciones diarias en Jalisco y cuáles son los periodos de tiempo en las que el número ha sido mayor y en cuales el menor. Como podemos observar estos coinciden con las olas de contagio y se puede observar un poco del impacto de las vacunas.



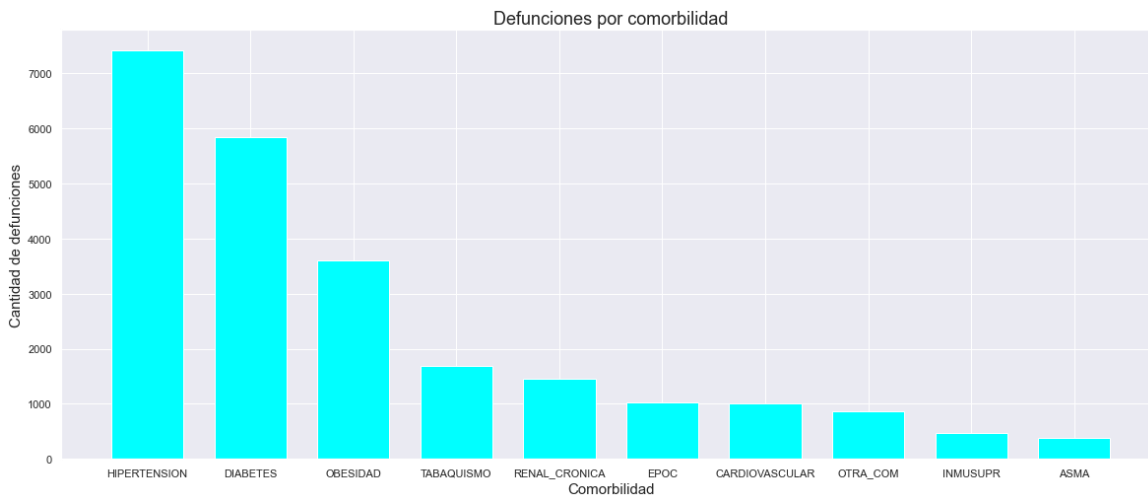
En la siguiente visualización añadimos la segmentación por sexo contra el tiempo, en la cual podemos observar que una mayor cantidad de hombres fallecieron por Covid-19 en el estado, a pesar de que un mayor número de mujeres se contagiaron (ver entrega número 2).



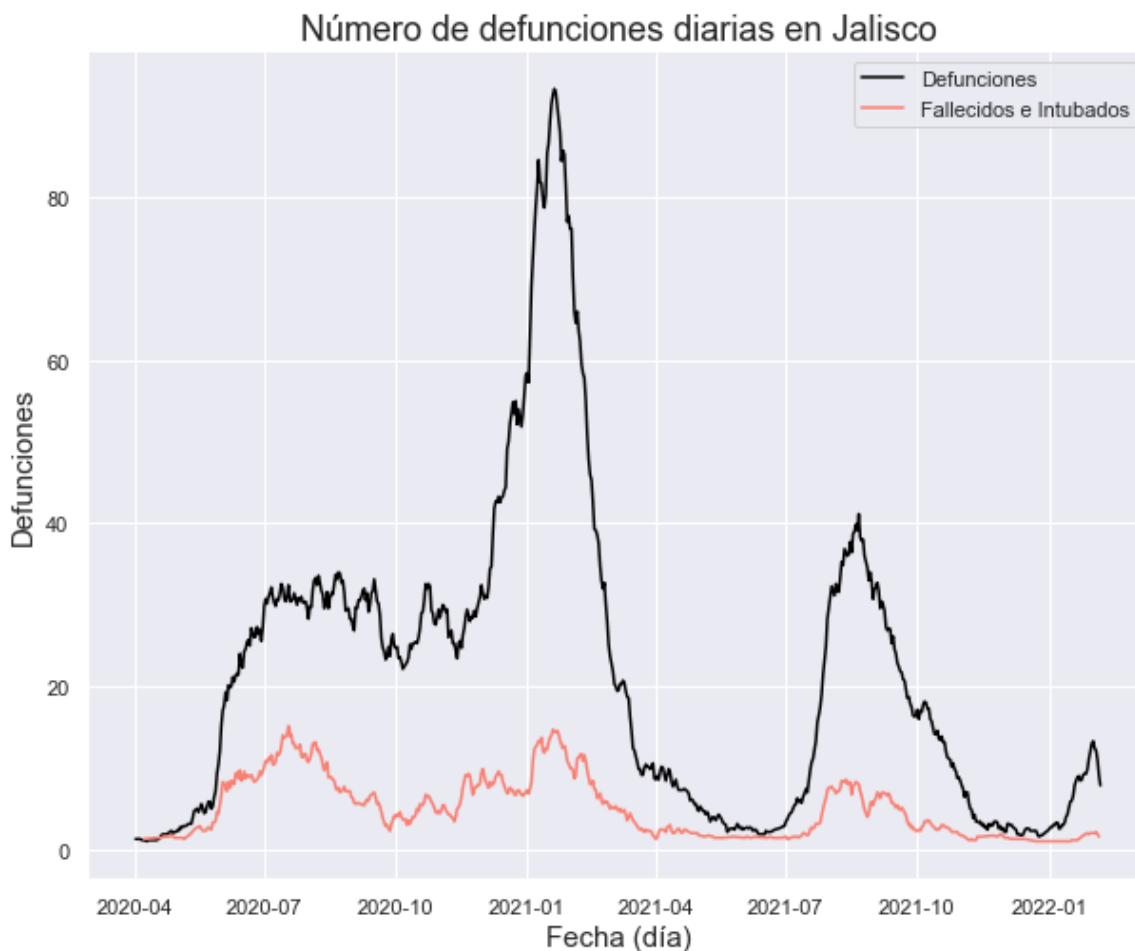
Posteriormente encontramos el porcentaje de defunciones por sexo. Además coincide con la gráfica anterior confirmando que el sexo masculino es que tiene mayor número de defunciones.



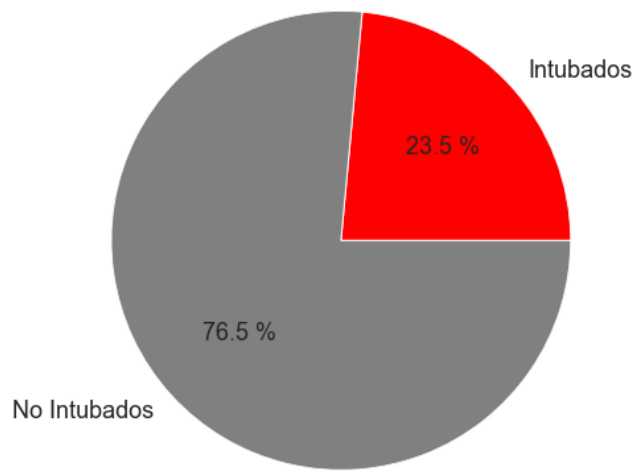
Buscamos encontrar la comorbilidad de otras enfermedades y su relación con la defunción por Covid-19. Como podemos observar la hipertensión, la diabetes y la obesidad son enfermedades que demuestran agravar la salud infectado por Covid-19.



También encontramos que la mayoría de los fallecidos por el virus no fueron intubados.

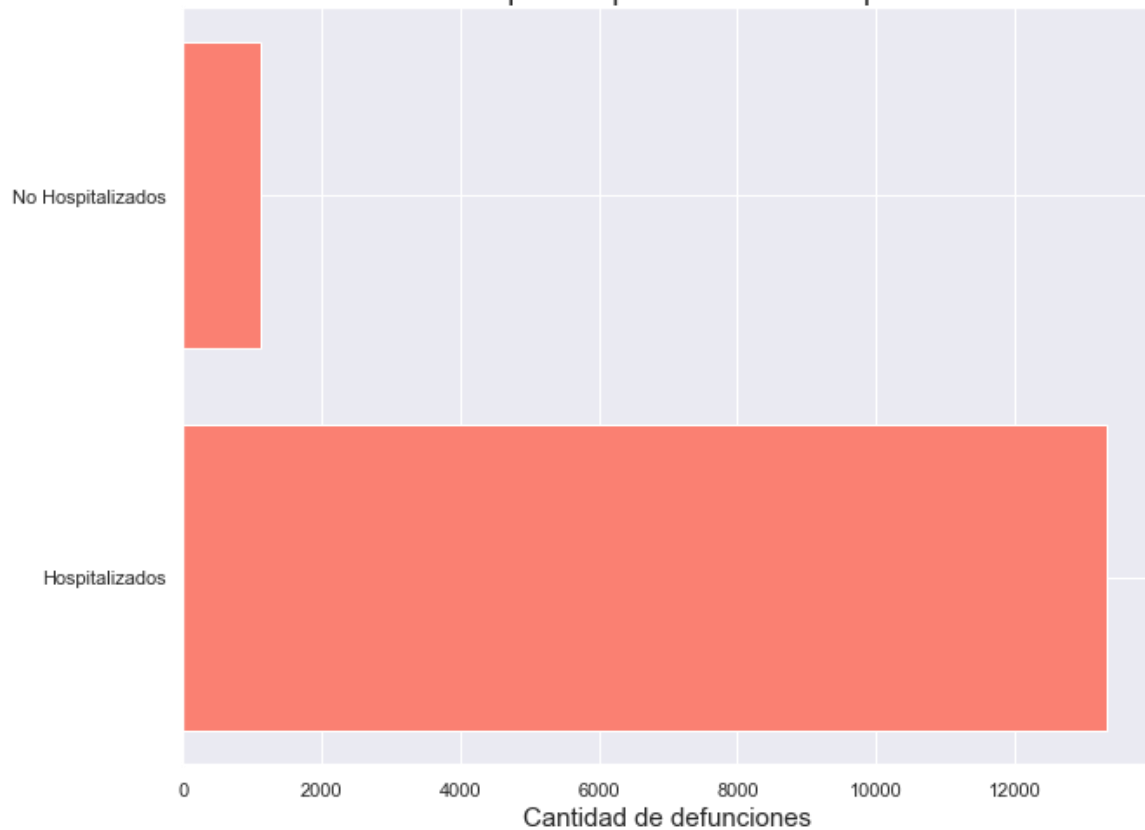


Defunciones por intubado/no intubado



Adicionalmente también graficamos la cantidad de defunciones y su relación con el estado de hospitalización. Como podemos observar, la mayoría de los fallecidos estuvieron hospitalizados en el estado.

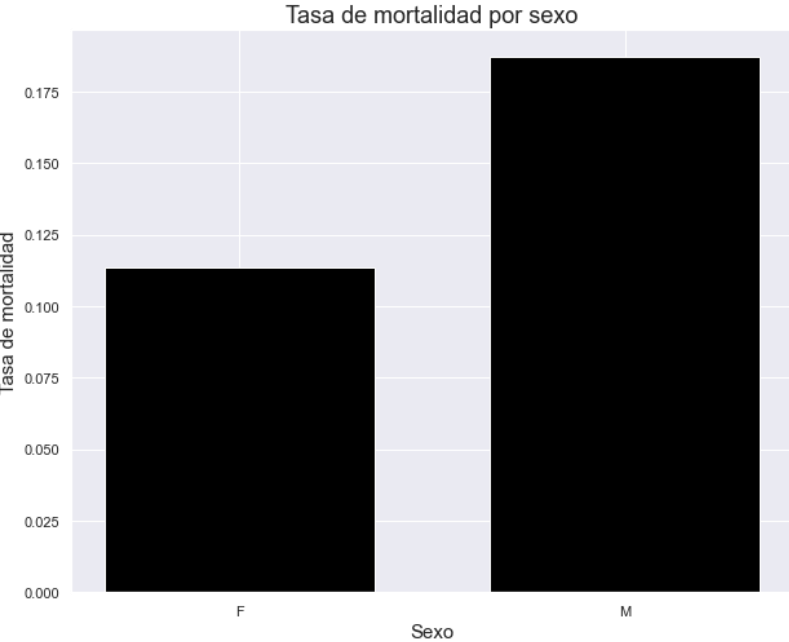
Defunciones por hospitalizados/no hospitalizados



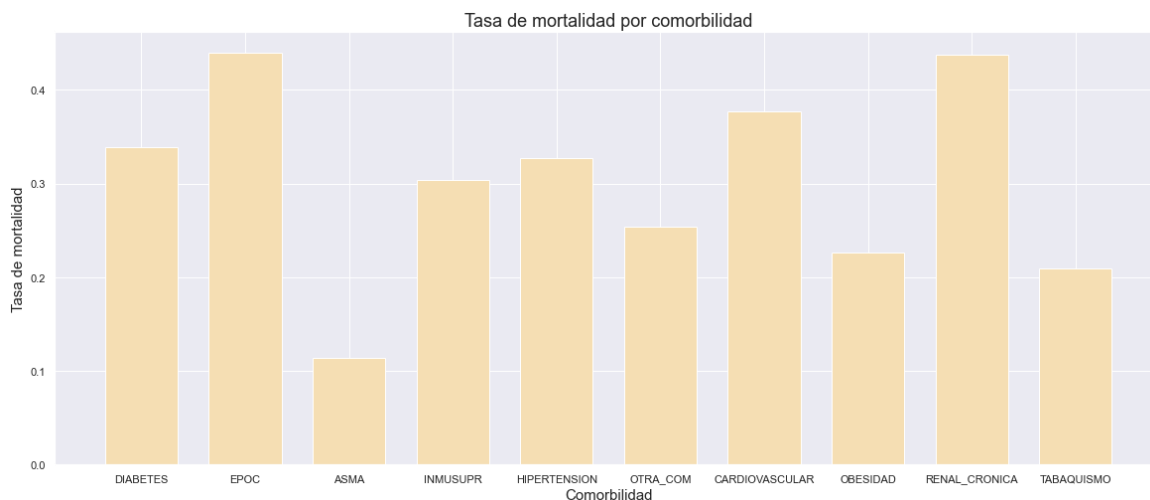
Analizamos las tendencias diarias en Jalisco con respecto a la tasa de mortalidad. Este fue calculado dividiendo el número de defunciones de casos confirmados entre el total de casos positivos. De aquí pudimos observar que los dos primeros picos de mortalidad coinciden con las olas de defunciones.



Con esta variable también buscamos confirmar la información acerca de las defunciones por sexo. Y como los resultados nos muestran existe una tasa de mortalidad superior en el sexo masculino.

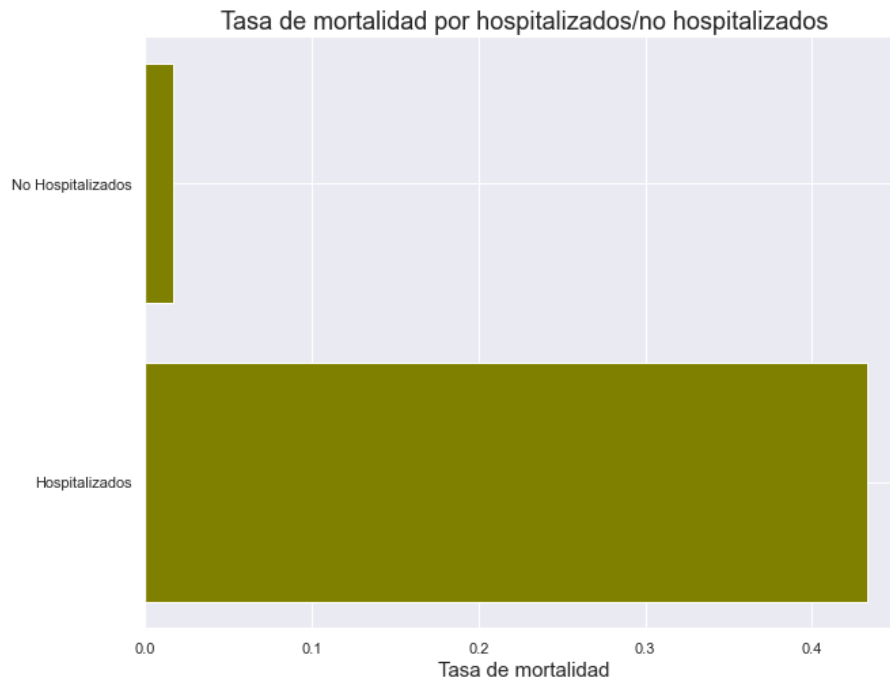


Seguimos analizando la comorbilidad de las enfermedades con respecto a la tasa de mortalidad del virus donde pudimos encontrar diferencias comparándola con el número de defunciones por comorbilidad, aquí encontramos que las enfermedades con las tasas más altas son las enfermedades pulmonares obstructivas crónicas (EPOC), las renales crónicas y cardiovasculares.



También graficamos la tasa de mortalidad de los intubados y no intubados y también la tasa de mortalidad por estado de hospitalización. Ambas coinciden con las visualizaciones con respecto al número de defunciones.

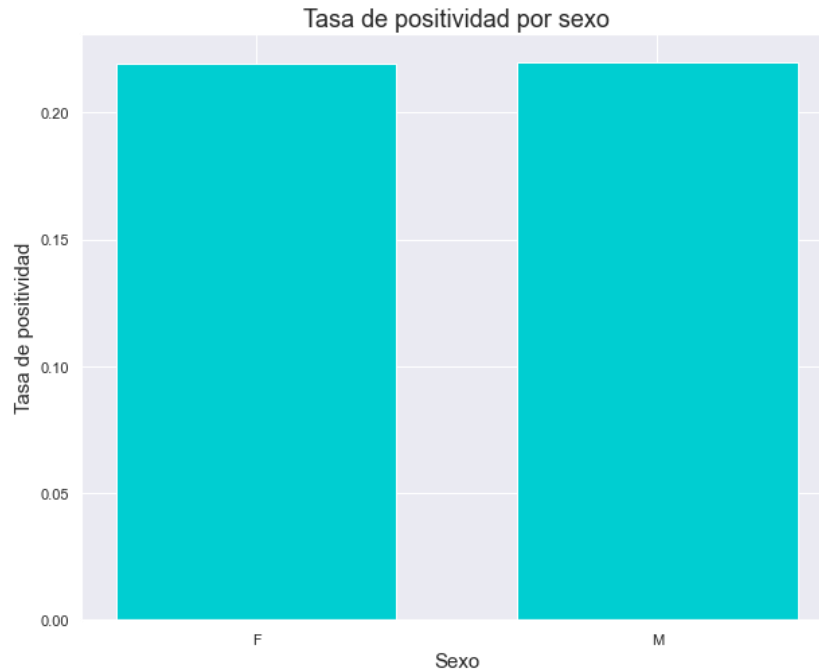




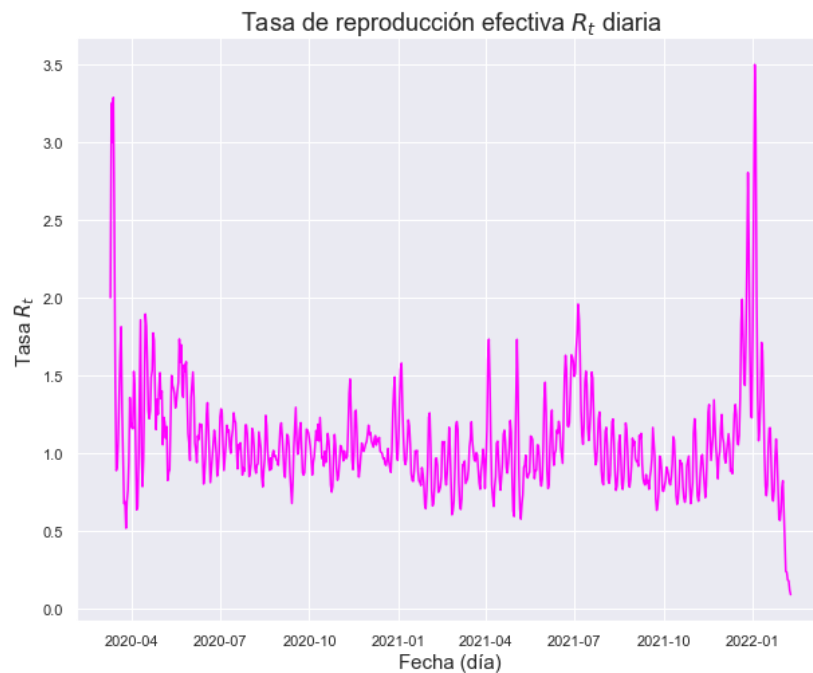
Pasando a la información obtenida de Radar Jalisco, hicimos el cálculo de la tasa de positividad diaria en el Estado y su relación con el tiempo. En esta podemos apreciar que al final hay un abrupto incremento en la gráfica, el cual coincide con todos los récords rotos de casos confirmados de Covid-19 en este inicio de año.



También fue calculada la misma tasa de manera histórica segmentada por sexo.

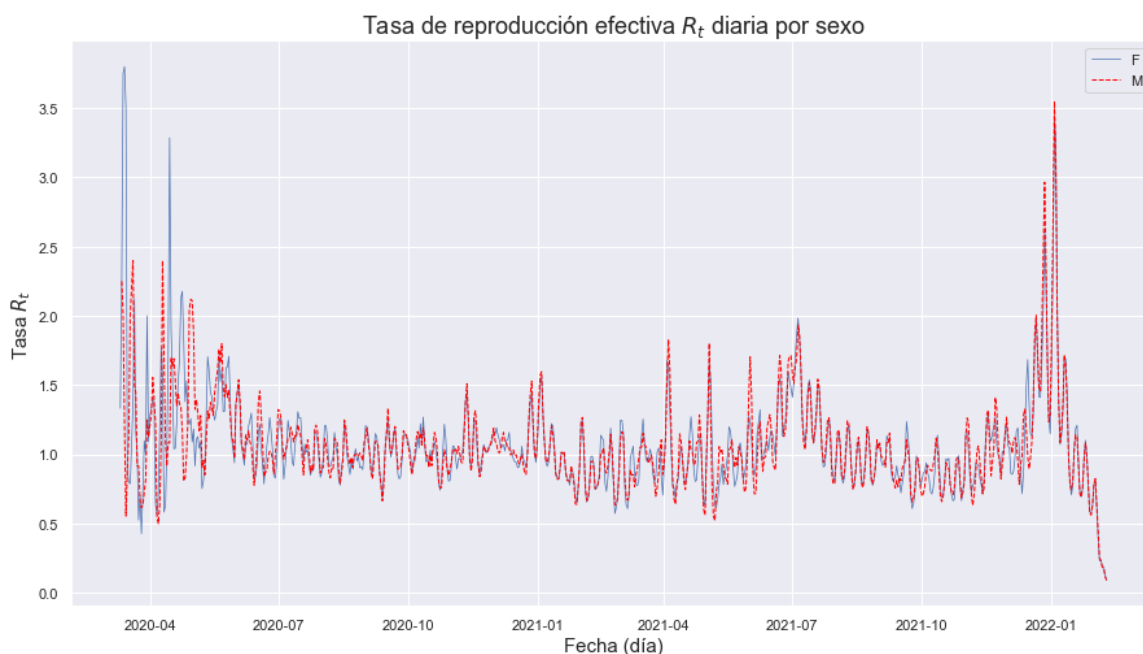


Finalmente se realizó una visualización de la tasa de reproducción efectiva diaria a lo largo del tiempo. El cálculo de esta variable nos indica que a pesar de que existen periodos muy estables, tanto al inicio como en este nuevo año 2022 hay periodos muy volátiles y que actualmente el virus como al inicio de la pandemia se reproduce en altas cantidades. Esto va de acuerdo con la información de salubridad proporcionada acerca del OMICRON, se comenta que esta es la variable más contagiosa de Covid-19





Esta variable también fue segmentada por sexo, dándonos información adicional del inicio de la pandemia que nos muestra que el virus históricamente ha sido más propenso a desarrollarse en el sexo masculino.



## Conclusiones

Como pudimos observar gracias a la información y visualización de estas variables, concluimos que las olas de contagio coinciden con los incrementos de defunciones y las tasas tanto de positividad como de mortalidad. Además, nos deja ver también que es más probable perecer ante esta enfermedad si se está intubado u hospitalizado y existen comorbilidades que aumentan aún más esta probabilidad. Algo interesante es que, si bien la positividad del virus ha aumentado, la tasa de mortalidad del virus ha ido en picada, esto se puede deber a dos factores, el primero, la campaña de vacunación del estado y la segunda puede ser que gran parte de la población pudo haber desarrollado una mayor resistencia al virus al estar en contacto con él.

Por último, observamos que el sexo masculino ha sido el que más probabilidad tiene de fallecer por Covid-19 en Jalisco, a pesar de que existe un mayor número de personas con sexo femenino en el estado. También se observó que la tasa de reproducción del virus se ha mantenido estable con el tiempo, con una media de aproximadamente 1, aun así, en los últimos meses, ha tenido una tendencia bajista.

### 1.5.2.2.Actividad 3. Análisis comparativo de municipios y olas de contagio

#### **Introducción**

En esta actividad buscaremos cumplir tres objetivos en concreto, el primero, hacer un análisis de los casos activos por fecha de inicio de síntomas, teniendo en cuenta que consideraremos un caso activo como aquel que cumpla con 14 días desde la fecha de inicio de síntomas. Además, haremos un análisis comparativo entre los 125 municipios existentes en el Estado de Jalisco, con la información obtenida de Radar Jalisco, finalmente haremos un análisis para buscar relaciones entre los distintos periodos de tiempo donde se presentaron los picos de las olas de pandemia en el Estado.

#### **Preguntas a contestar**

- ¿Cuál es la relación que existe a lo largo del tiempo y el número de personas consideradas activas?
- ¿Cómo influye la relación entre los picos de las olas y el número de personas activas?
- ¿Cuáles son los municipios con mayor cantidad de positivos per cápita?
- ¿Cuáles son los municipios con menor cantidad de positivos per cápita?
- ¿Cómo se ha visto afectada la ZMG por cantidad de positivos per cápita?
- ¿Cuáles son las cantidades de personas positivas, tasa de reproducción y tasa de positividad en los distintos picos de la pandemia?
- ¿Cómo se ven representadas las tendencias de las olas de contagio en los distintos periodos de tiempo?

#### **Procedimientos**

- **Limpieza de bases de datos de Radar Jalisco:** Se eliminaron las columnas con información innecesaria, se renombraron las columnas y algunas de las variables de resultados.
- **Transformación de los datos:** Se agruparon los datos por fecha, número de positivos, número de descartados y agrupaciones de los municipios que pertenecen a la ZMG. Hicimos coincidir fechas y llenado de espacios blancos, además, se buscó definir en la base el número de estado por nombre y no por código como estaba.
- **Cálculo de variables:** Con los datos agrupados se calcularon las variables de casos activos en el Estado, casos activos de cada municipio y casos activos de la ZMG a lo largo del tiempo. También hicimos una normalización de las distintas variables trayendo consigo los valores per cápita por cada 100,000 habitantes y finalmente la segmentación del periodo de pandemia en los distintos picos que esta ha tenido.
- **Visualización de datos:** Se generan gráficas de los datos y con estas visualizaciones poder entender y poder responder las preguntas a contestar.

## Resultados

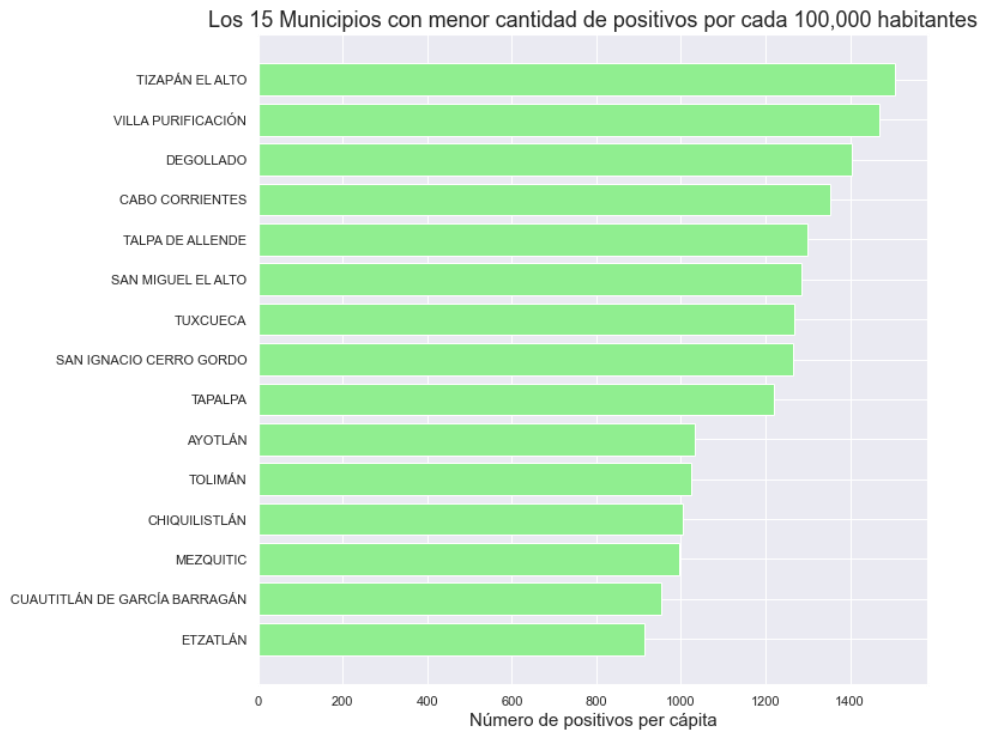
Para la primera parte de esta entrega, buscamos realizar un cálculo de los casos activos en Jalisco, en la siguiente gráfica podemos observar la cantidad de personas activas desde el inicio de la pandemia en 2020. Podemos observar que estos coinciden con las olas de contagio, también encontraremos que tiene una escala mayor al número de positivos, esto debido a que los casos activos se acumulan día con día por un periodo de 14 días.



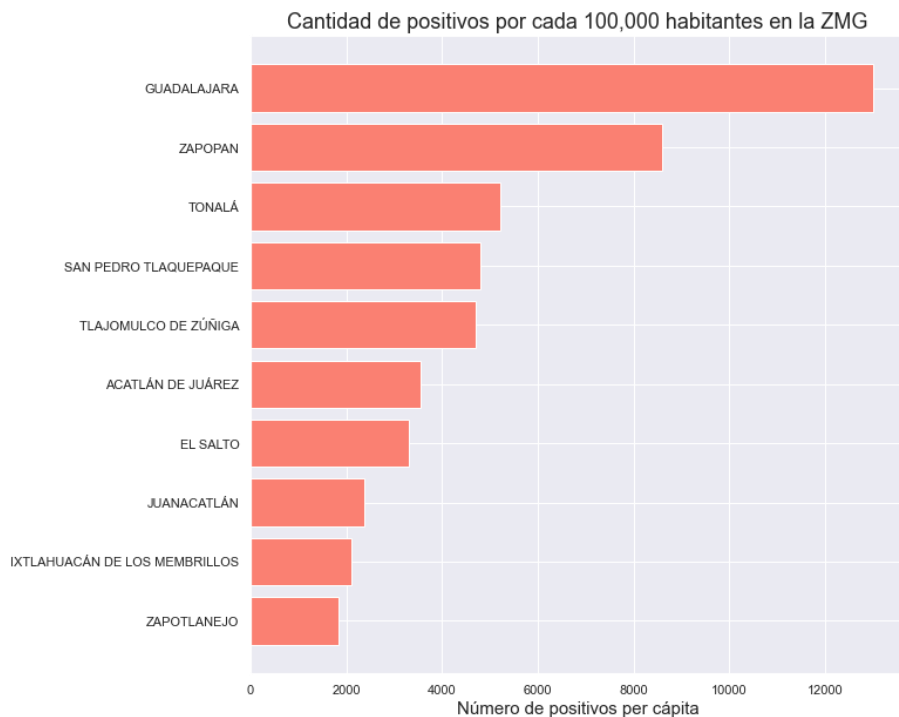
Posterior a esto hicimos un ranking con los municipios del Estado y su número de casos activos diarios. Primero encontramos los que presentan la mayor cantidad per cápita de positivos, destacando en primer lugar la ciudad de Guadalajara, seguida por Colotlán y Puerto Vallarta.



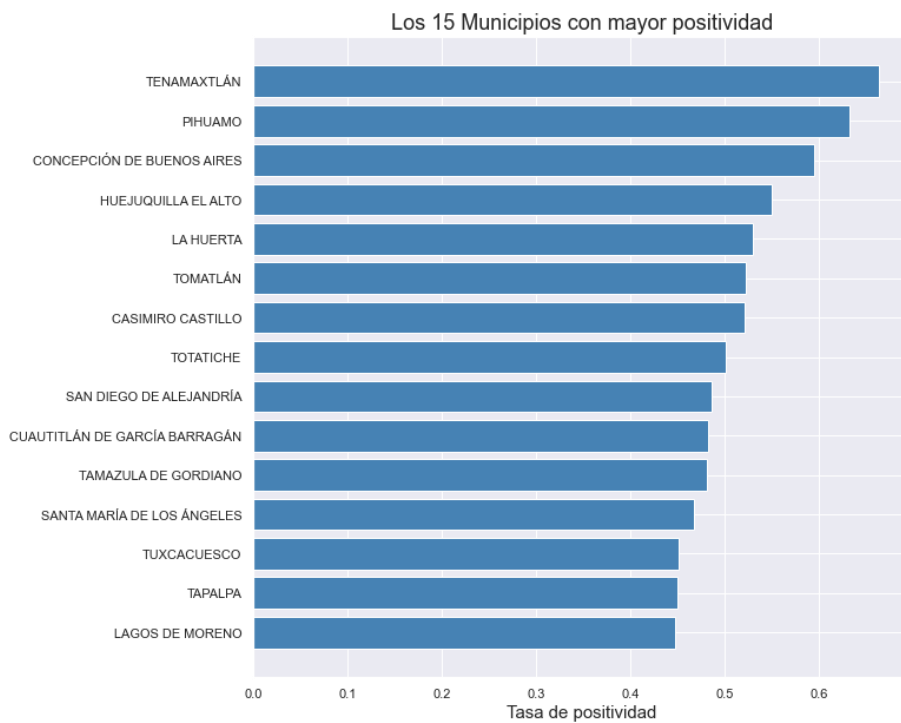
Si siguiendo con los resultados para este cálculo encontramos los municipios con la menor cantidad de positivos per cápita, resaltando a los municipios de Etzatlán, Cuautitlán de García Barragán y Mezquitic. Lo cual nos indica que son los que mejores medidas han tomado o también por ser municipios con poca rotación de personas.



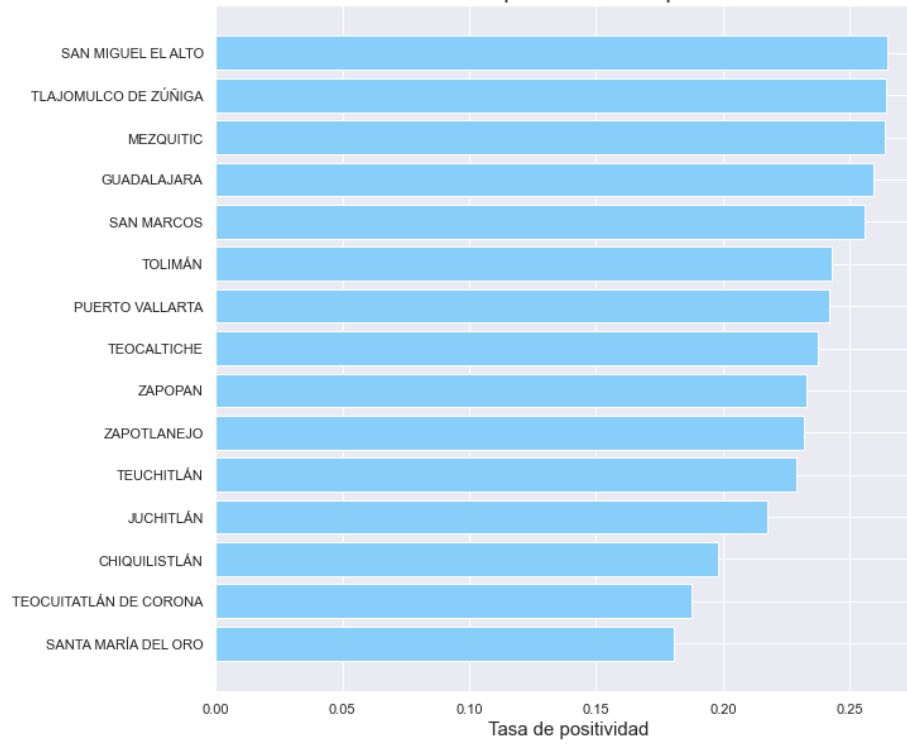
Se realizó una visualización agrupando a los municipios de la ZMG, para poder ver como se ha desarrollado el virus en la zona metropolitana.



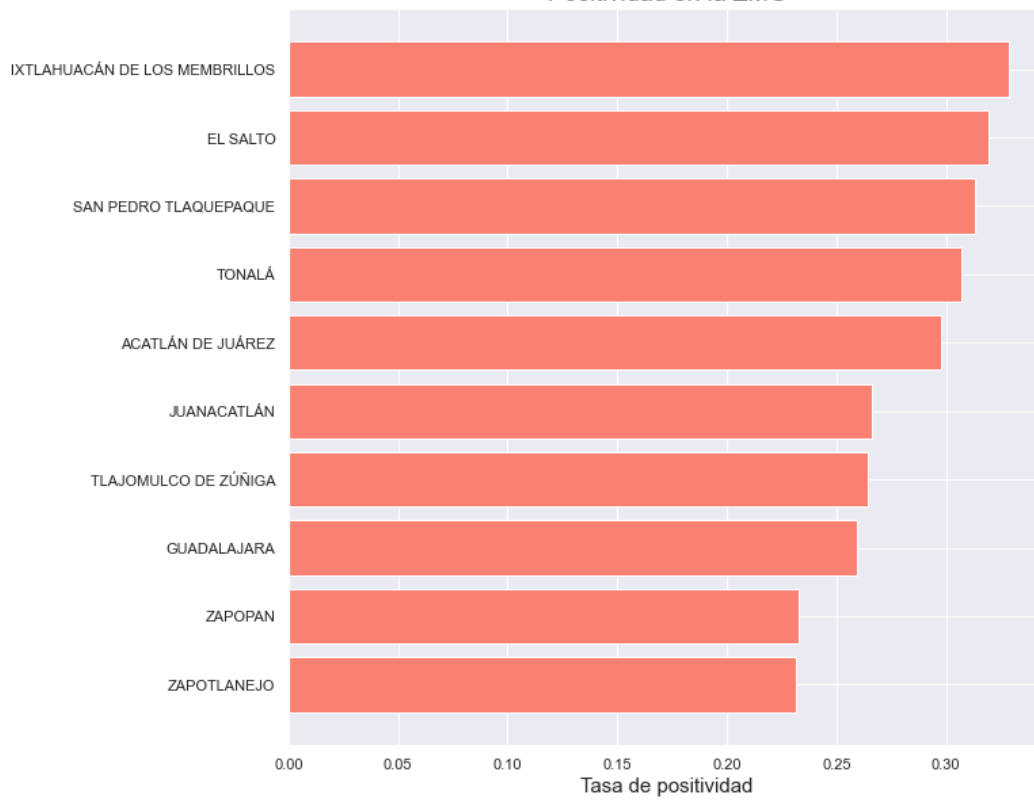
Buscando más generar más *insights* acerca del comportamiento de cada municipio, agregamos también la tasa de positividad para cada uno de estos.



Los 15 Municipios con menor positividad



Positividad en la ZMG

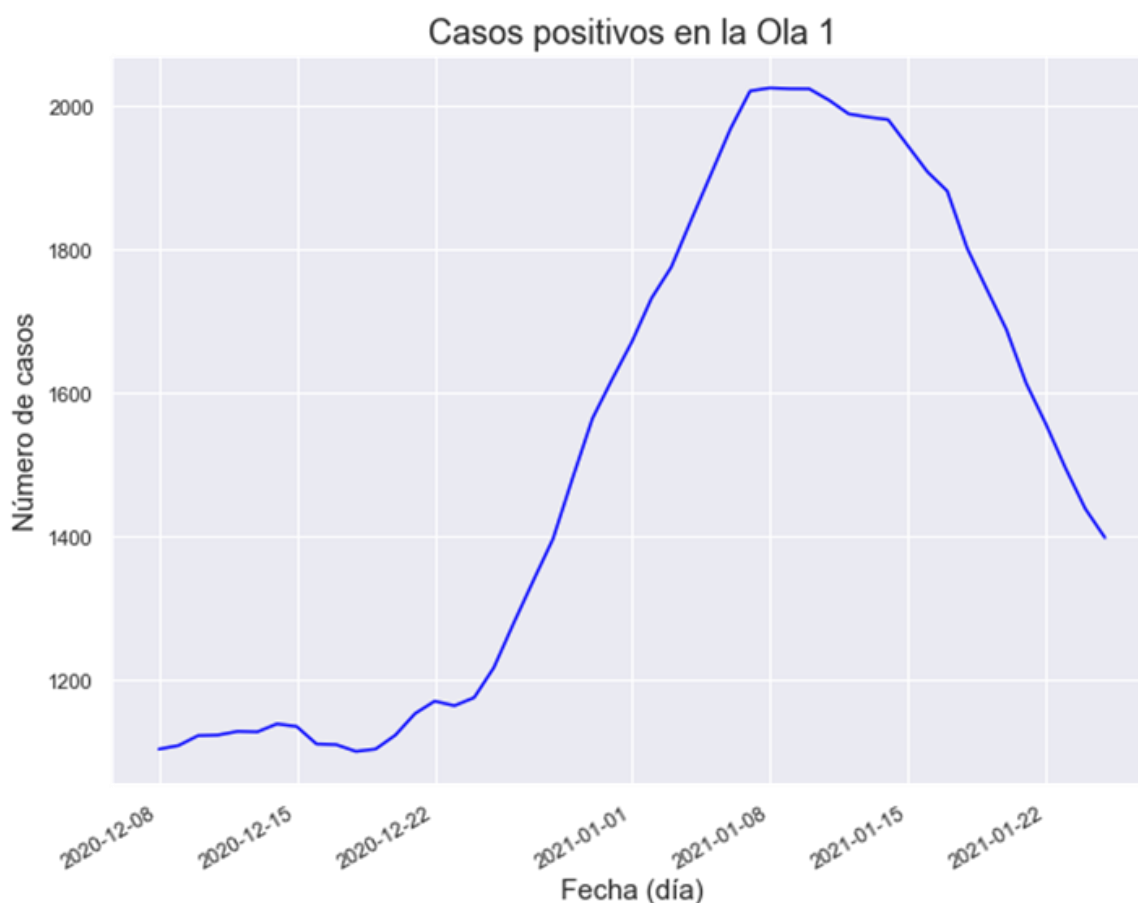


Como bien podemos observar, los resultados obtenidos no son parecidos a lo obtenido en la cantidad de positivos por municipio, esto es debido a que el cálculo de las tasas es hecho a través de casos confirmados entre el total de confirmados más el total de descartados, por lo que Guadalajara se puede encontrar en los de menor ya que se realizan una mayor cantidad de pruebas, a diferencia de Tenamaxtlán o Pihuamo donde puede ser que la gente tenga menos acceso a este recurso.

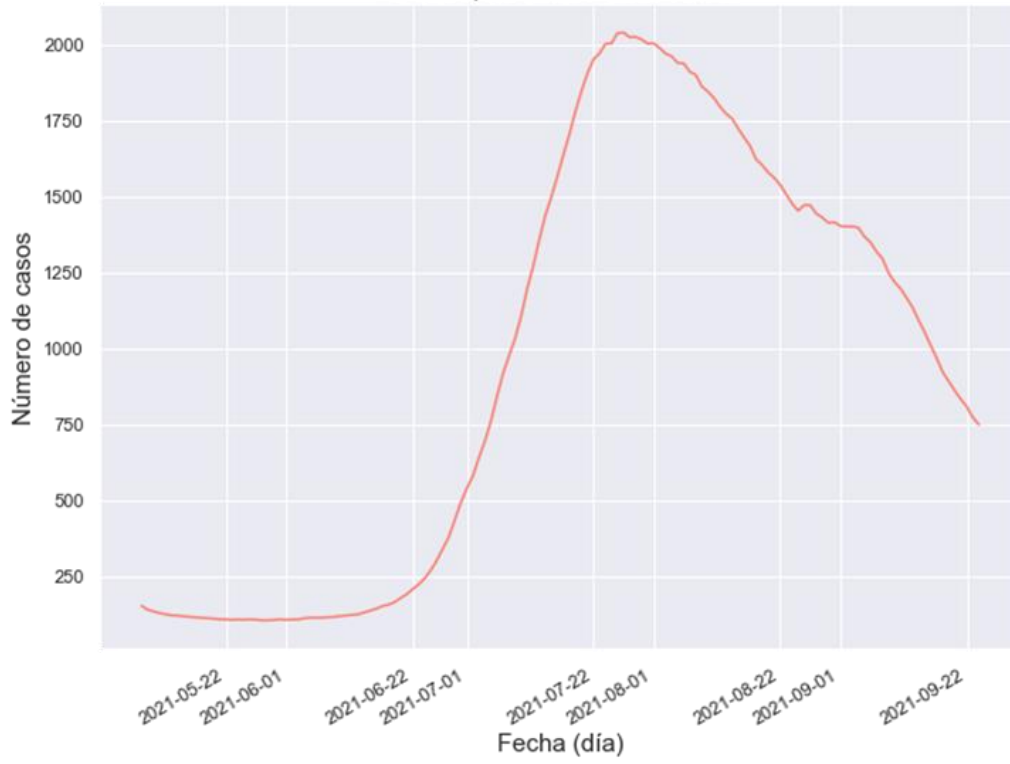
Pasando a la siguiente sección del reporte encontraremos el análisis entre las diferentes olas de contagio. Consideramos las olas de contagio con los siguientes periodos:

- **Ola 1:** Diciembre 2020 a Enero 2021
- **Ola 2:** Junio 2021 a Septiembre 2021
- **Ola 3:** Noviembre 2021 a Enero 2022

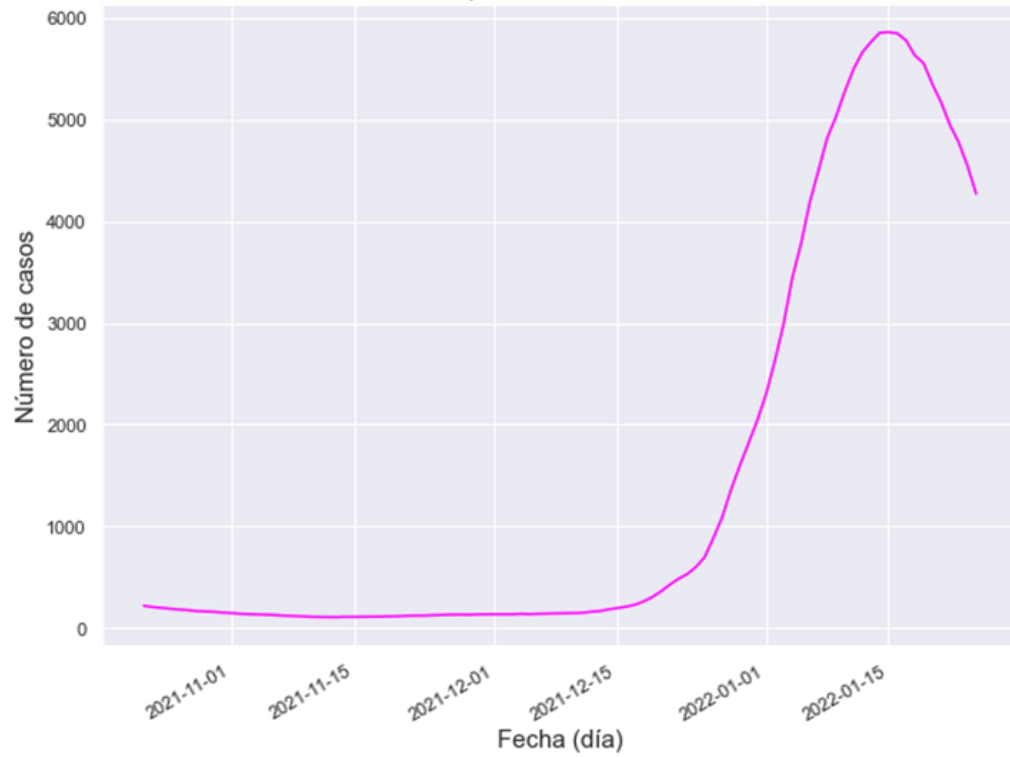
Para la comparativa, decidimos iniciar con los casos positivos entre olas:



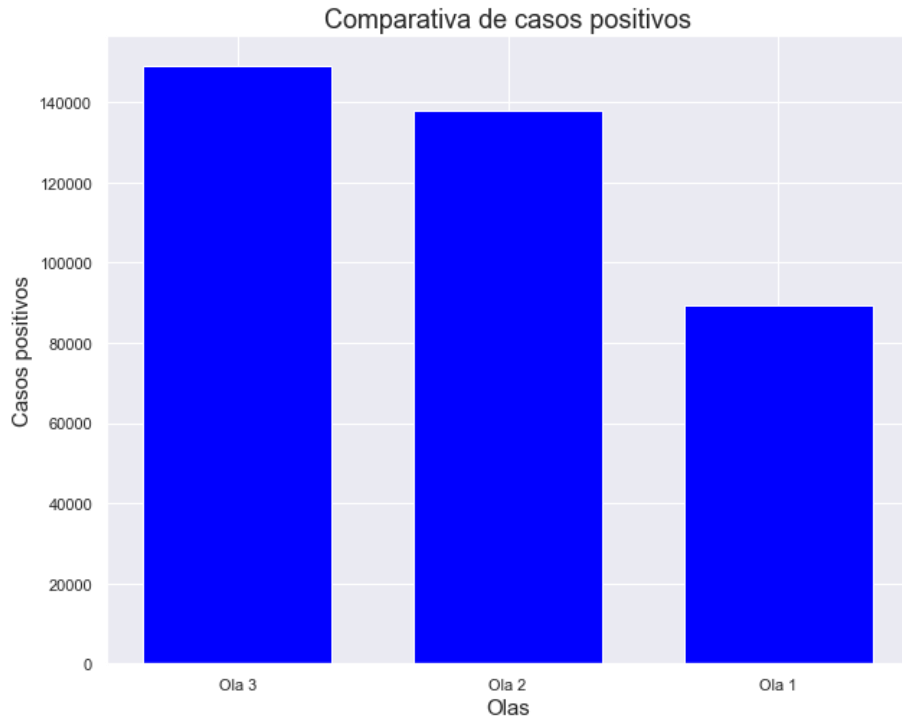
Casos positivos en la Ola 2



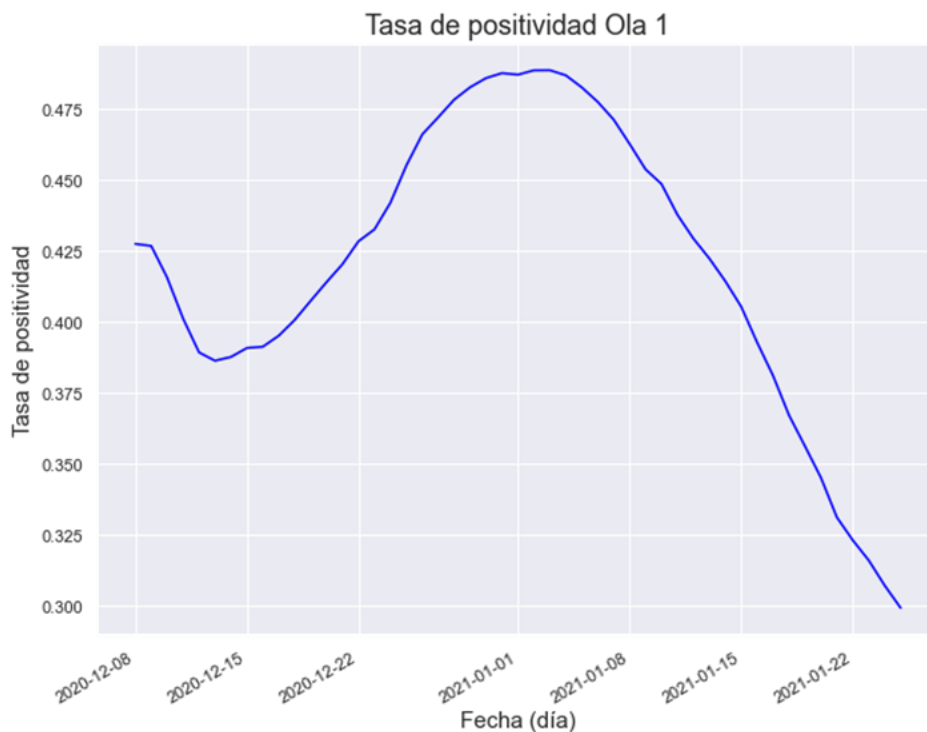
Casos positivos en la Ola 3

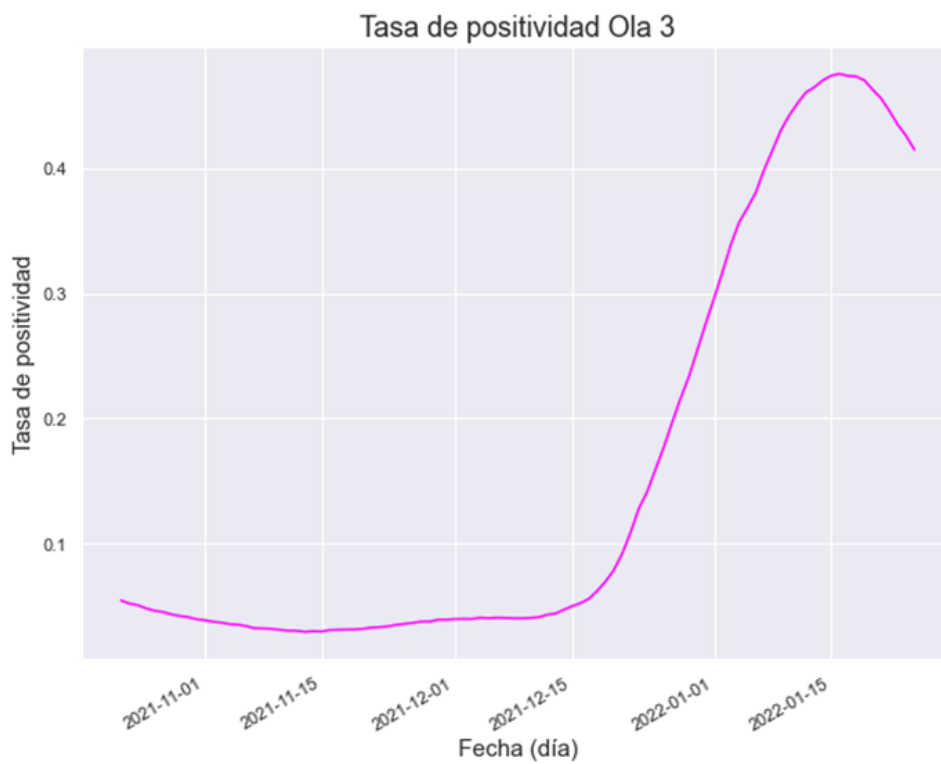
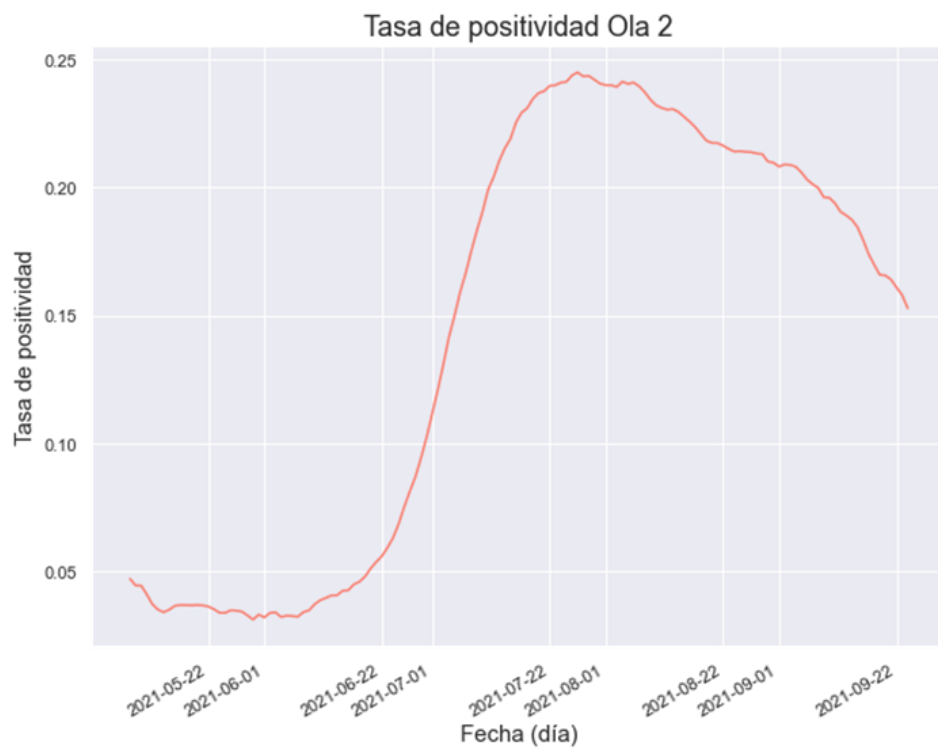


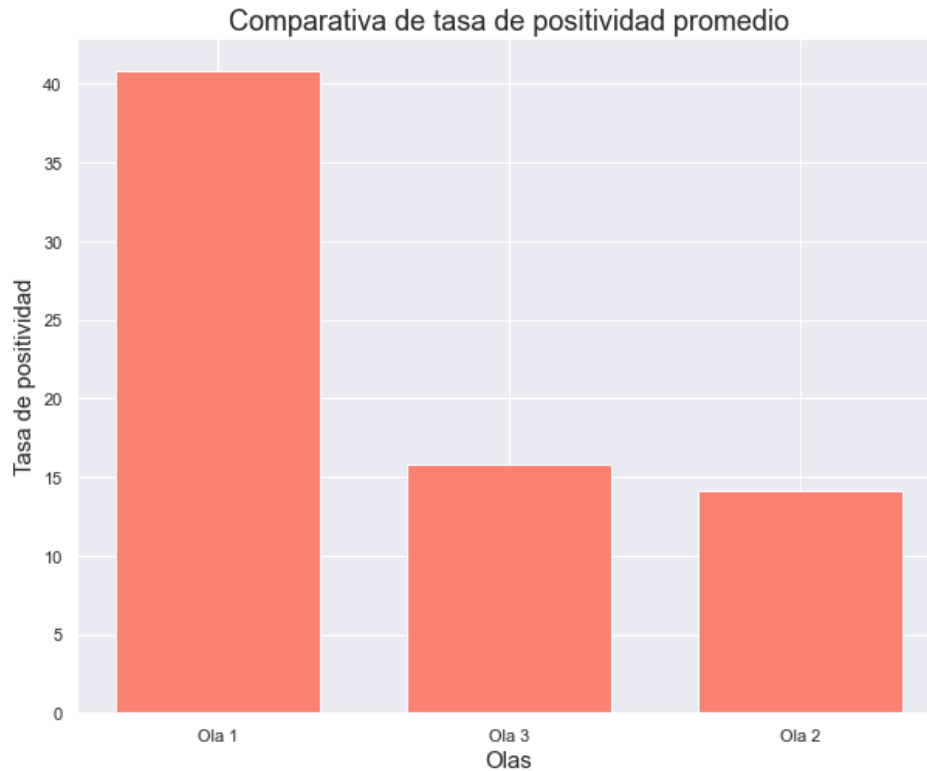




Como los datos y las gráficas nos indican la ola 3 fue la que mayores positivos tuvo, seguida muy de cerca por la ola número dos y finalmente la primera. El hecho de que la primera ola tuviera menores positivos se puede deber a la menor cantidad de pruebas que se hicieron en este período, además de la campaña de aislamiento hecha por el gobierno. Continuando con un análisis entre la tasa de positividad por ola.

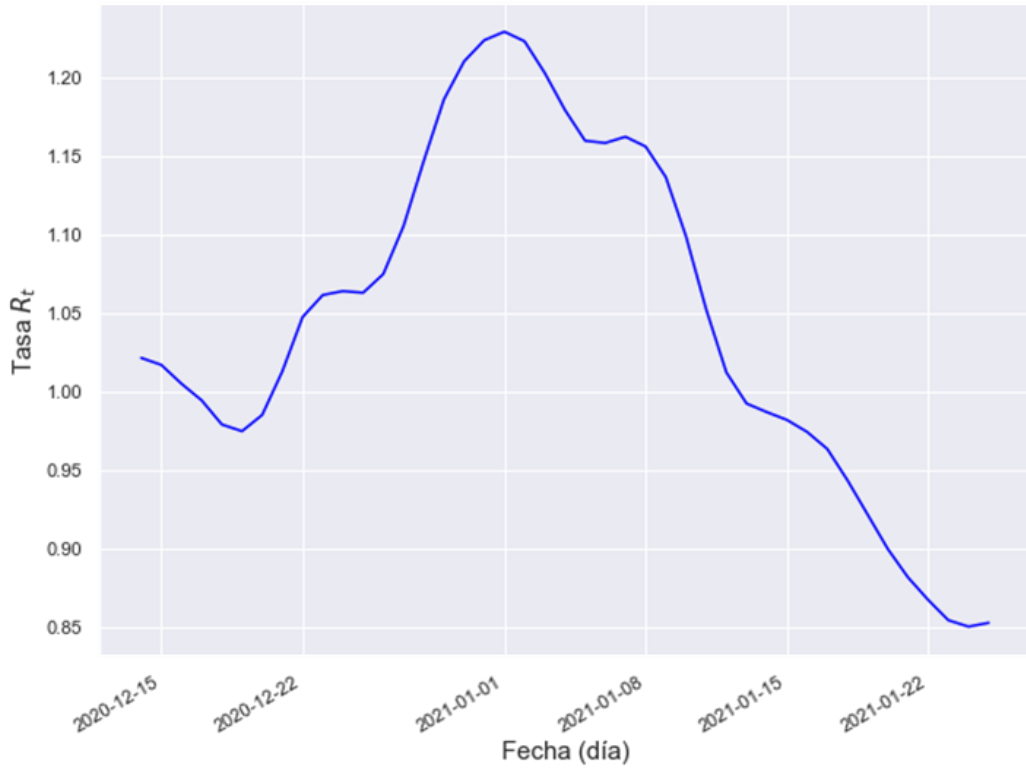




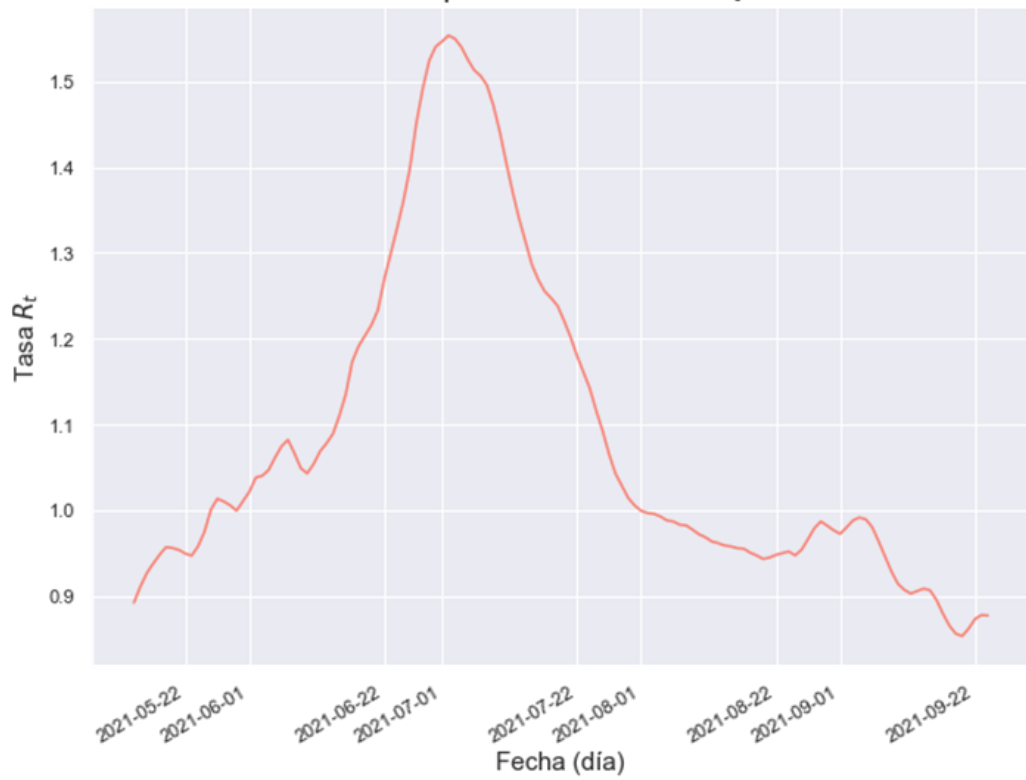


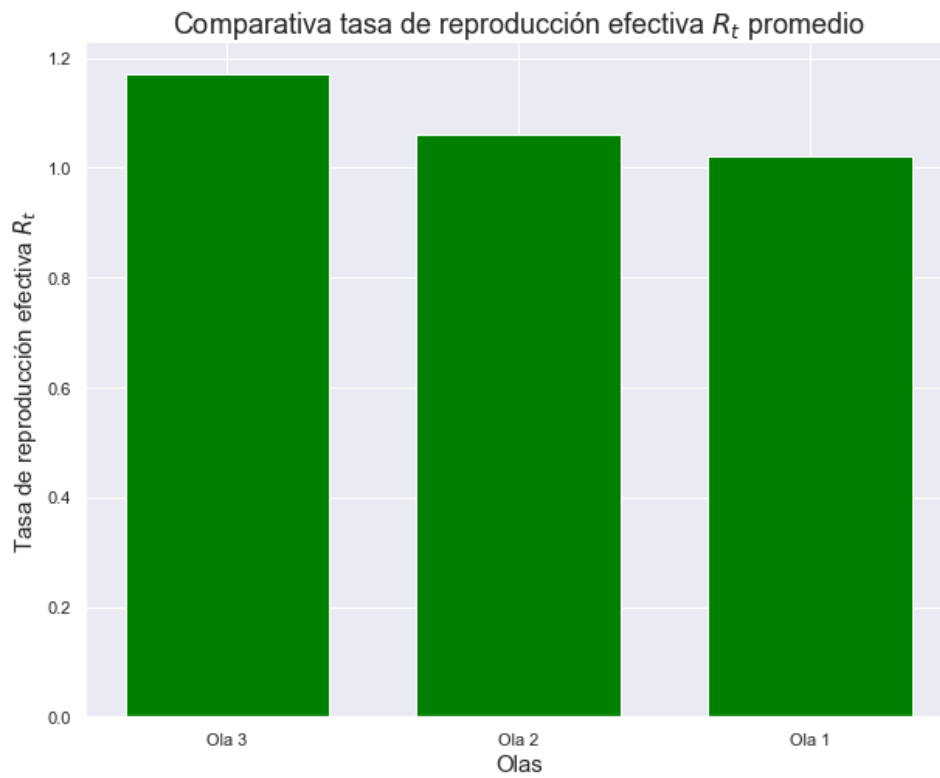
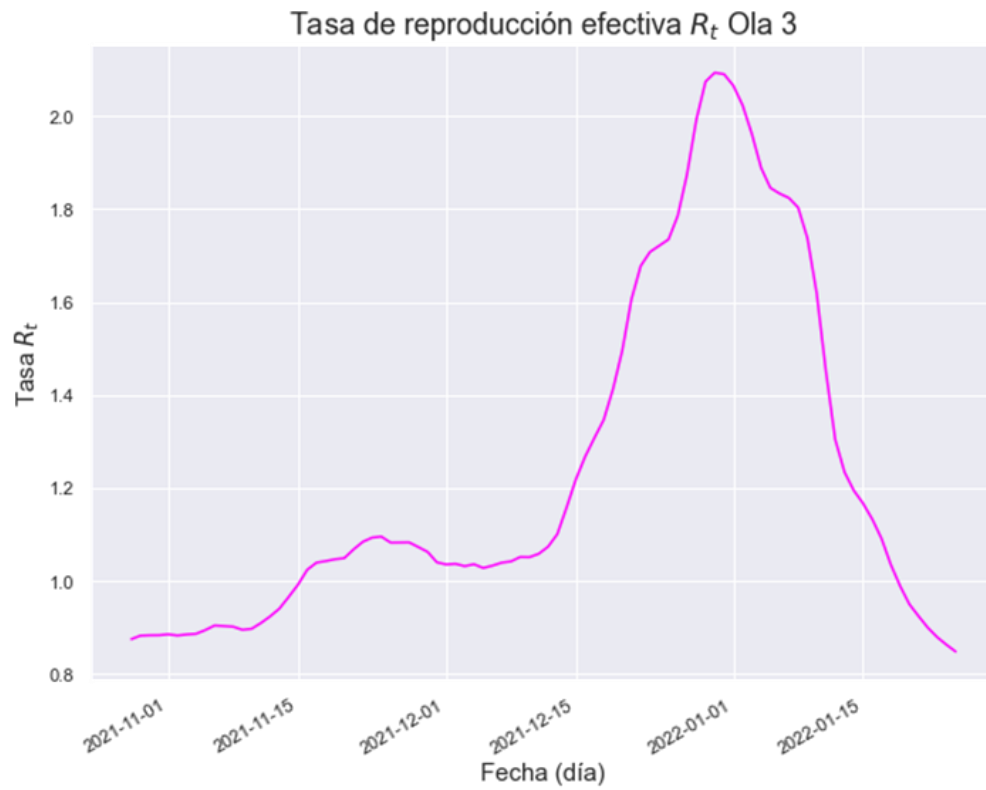
Podemos encontrar que la ola con mayor tasa de positividad fue la uno con una tasa que superaba el 40%, mientras que la segunda y tercera ola tuvo la mitad de estas tasas aproximadamente un 20%. También podemos ver que el orden de las olas cambio con respecto a las gráficas anteriores. Esto se debe a que en la primera ola se hacía un menor número de pruebas, además de que los síntomas eran mucho más visibles, además de que la facilidad para realizarse una prueba no era tanta a diferencia de ahora, actualmente en la ola 3 que existe un mayor número de personas que se han realizado pruebas. Seguimos, finalmente con la tasa de reproducción efectiva.

Tasa de reproducción efectiva  $R_t$  Ola 1



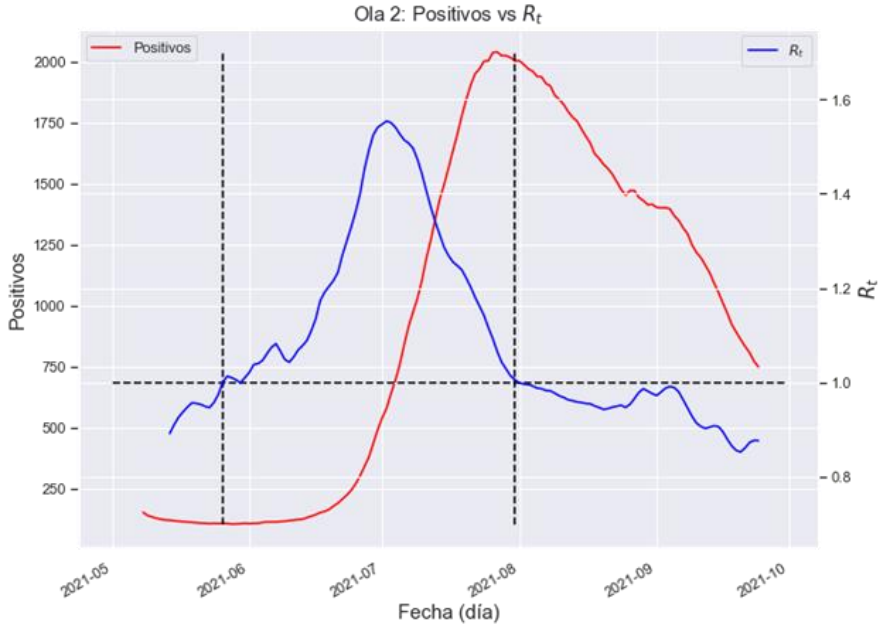
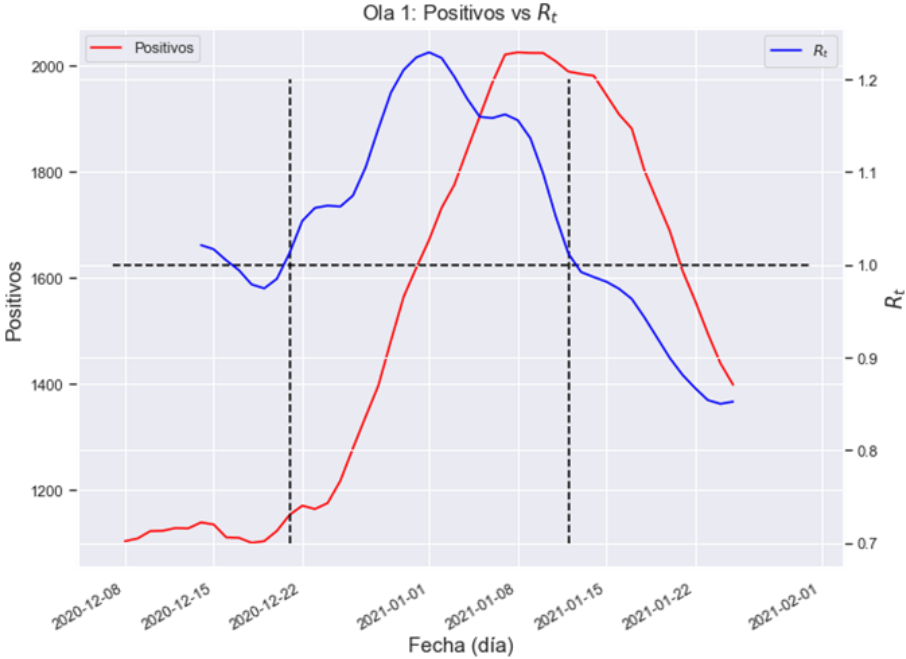
Tasa de reproducción efectiva  $R_t$  Ola 2

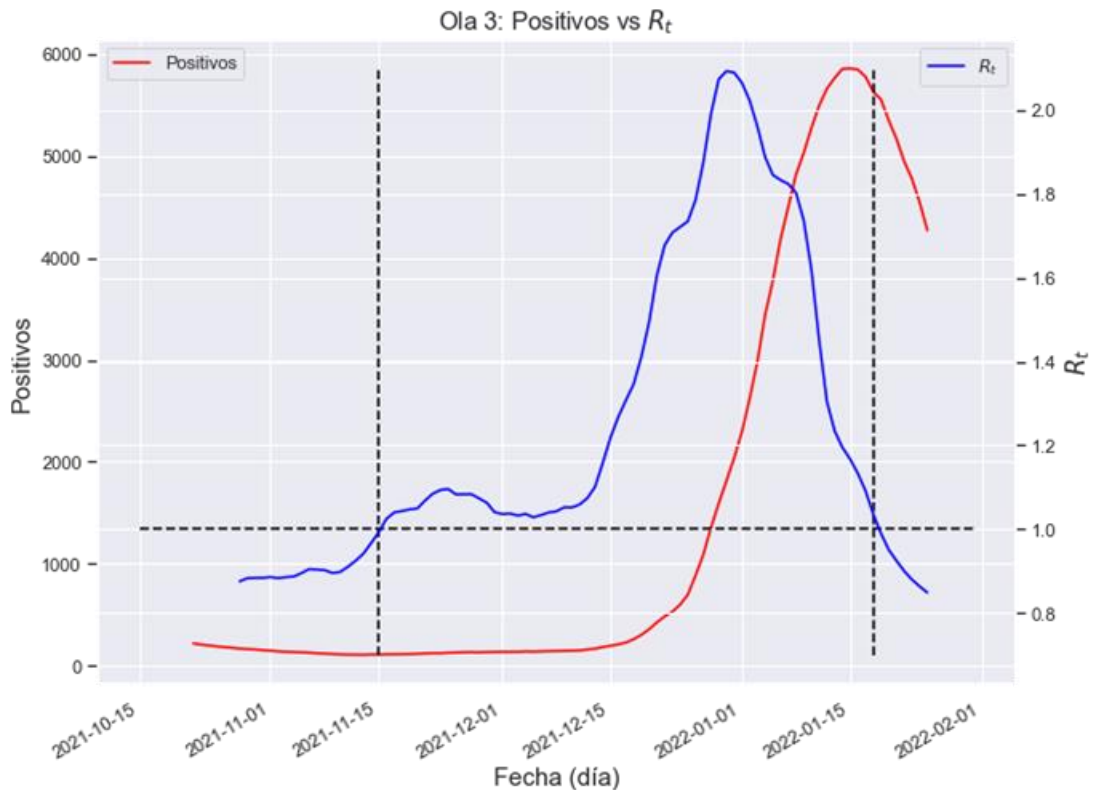




En este caso, la ola que mayor tasa de reproducción activa tuvo fue la tercera, la cual coincide con el número de casos positivos. Mientras que las otras dos olas quedan muy a la par. Esto coincide con lo comentado por las autoridades acerca de la actual sepa del Covid-19, el Ómicron ha sido la variante más contagiosa que ha llegado a Jalisco.

Además, hicimos una doble gráfica comparando el  $R_t$  contra los casos positivos, y observamos una relación cuando la  $R_t$  aumenta de 1, los casos positivos suben y cuando la  $R_t$  baja de 1, los casos positivos bajan, esto ocurre en las tres olas.





## Conclusiones

La relación del tiempo y el número de casos considerados activos coinciden con los periodos de picos de pandemia, además existe un mayor número de casos activos que de confirmados diarios ya que los casos activos se acumulan día con día por un periodo de 14 días.

Los municipios con mayor número de casos corresponden a los estados donde hay mayor rotación de personas, al ser los más importantes de la ciudad, encabezados por Guadalajara, Colotlán y Puerto Vallarta. Cabe destacar municipios como Colotlán que a pesar de ser un municipio pequeño tuvo una gran cantidad de casos activos. También es importante destacar que Etzatlán, Cuautitlán de García Barragán y Mezquitic tuvieron los menores números de casos activos.

Finalmente podemos apreciar que dentro de las olas la tercera podría considerarse la que mayor contagiosa fue y donde se tuvo una mayor cantidad de contagiados, gracias a la variante Ómicron, seguido por la primera ola por sus altos índices de contagios y positivos que tienen explicación en el contexto de síntomas, pruebas y positividad en el que se encontraba en ese periodo de tiempo. Además, en las tres olas se repite un patrón, cuando la tasa de reproducción va hacia arriba de uno, los casos positivos aumentan y cuando va hacia abajo estos disminuyen, lo cual es una observación interesante.

### 1.5.2.3. Actividad 4. Distribución de edad por olas

#### **Introducción**

En esta actividad haremos un análisis de las distribuciones de edad, busquemos encontrar información acerca de cómo los distintos rangos de edad han sido afectados durante este periodo y como han ido evolucionando tras las distintas olas, además no sólo estaremos utilizando información de los casos confirmados sino que también incluiremos a nuestro análisis información de la base de datos del SINAVE, para así conocer también el número de defunciones y así dar resultados más completos.

#### **Preguntas a contestar**

- ¿Cuál fue la ola que tuvo la más cantidad de positivos?
- ¿Cuáles son las diferencias entre las distintas olas y sus probabilidades de resultar como un caso confirmado del virus?
- ¿Cuál ha sido el rango de edad que ha sido más afectado de acuerdo con su probabilidad de ser positivo?
- ¿Han cambiado las tendencias de edad conforme el paso de los distintos picos de la pandemia?
- ¿Cuál es el periodo de tiempo que encabeza la probabilidad de ser positivo y fallecer?
- ¿Cómo ha sido el comportamiento de las defunciones en los distintos periodos de las olas de contagio?
- ¿Existe algún notable cambio entre los distintos picos de pandemia y los rangos de edad de la sociedad jalisciense?
- ¿Cuál es el sector de edad que presenta mayores probabilidades de perecer ante este virus?

#### **Procedimientos**

- **Limpieza de bases de datos de Radar Jalisco y de SINAVE:** Se eliminaron las columnas con información innecesaria, se renombraron las columnas y algunas de las variables de resultados.
- **Transformación de los datos:** Se agruparon los datos por fecha, número de positivos, número de descartados, número de defunciones. Hicimos coincidir los rangos de fechas y llenado de espacios blancos.
- **Cálculo de variables:** Filtramos la cantidad de positivos en cada una de las olas con la base de Radar Jalisco y para las defunciones se usó la base de SINAVE.
- **Visualización de datos:** Se generan gráficas de los datos, en este caso utilizamos histogramas y con estas visualizaciones poder entender y poder responder las preguntas a contestar.



## Resultados

Para iniciar, queremos mencionar que se tomaron los siguientes periodos para cada una de las olas:

- **Ola 1:** Diciembre de 2020 hasta Enero de 2021.
- **Ola 2:** Julio 2021 hasta Septiembre 2021.
- **Ola 3:** Enero 2022 hasta el día de hoy.

Iniciamos sacando los histogramas de positivos por edad, como se muestra a continuación. Se usó la base de **RADAR JALISCO**.



Como podemos observar en la primera ola tenemos una mayor cantidad de gente considerada como adulto mayor, la cual se ve disminuida en comparación las otras dos. Esto es debido a que el virus es una enfermedad respiratoria la cual como es bien sabido tiene una mayor letalidad para este rango de edad. También como podemos observar durante las 3 olas encontraremos que el mayor volumen de confirmados tiene entre 20 y 40 años edad. Es importante resaltar que los dos últimos periodos de tiempo tienen un comportamiento bastante similar.

Cabe recalcar que con la ayuda de las gráficas hemos podido apreciar que durante la primera ola las personas de 60 años tenían una probabilidad de casi 1% de ser positivos mientras que para el año actual durante la tercera ola esta tasa ha bajado por mitad dándonos un poco menos de 0.5%. Esto se puede deber a la campaña de vacunación del estado.

Continuando con las defunciones de la base de **SINAVE**:



Con ayuda de estas gráficas encontraremos información que apoye a los resultados obtenidos en el procedimiento anterior. Como bien mencionábamos el virus ha afectado principalmente a las personas con una edad mayor, como podemos observar durante el paso del tiempo, las probabilidades de ser positivo y fallecer han ido disminuyendo en la gente más adulta y predominan durante la primera ola. Esto quiere decir muchas de las personas pertenecientes a este sector de la población lamentablemente han perecido.

Es importante resaltar que durante el primer periodo de pandemia la gente con edad por arriba de los 60 años tenía una mayor probabilidad de ser positivo y fallecer mientras que durante el periodo actual podemos observar altos picos, pero con una probabilidad menor.

## Conclusiones

El virus del Covid-19 es una enfermedad de tipo respiratorio que afecta a varios órganos internos, de los cuales si no se tiene un sano sistema inmune puede llegar a incluso causar la muerte. El rendimiento del sistema inmune es afectado por la edad, por lo que incrementó la probabilidad de perecer ante esta enfermedad para los rangos de edad de adultos mayores. Como pudimos observar, el sesgo del histograma de las defunciones va cambiando, lo cual quiere decir, que, si bien pudieron haber fallecido varios adultos mayores, la campaña de vacunación también pudo haber ayudado a cambiar esto. También observamos las personas jóvenes (entre 20 y 40 años) son las que mayor probabilidad de salir positivos tienen.

### 1.5.2.4. Actividad 5. Análisis de texto

#### Introducción

En esta entrega se buscará hacer un análisis de texto, específicamente de *tweets*, con el tema relacionado con la pandemia en Jalisco y en algunos casos en México, los temas a tratar son los siguientes:

- Covid-19 Jalisco
- Cubrebocas Jalisco y México
- Pandemia Jalisco y México

- Vacunas Jalisco y México

Se dividirá el reporte en diferentes secciones a las comentadas anteriormente.

## Resultados

### 1.5.2.4.1. Covid-19 Jalisco

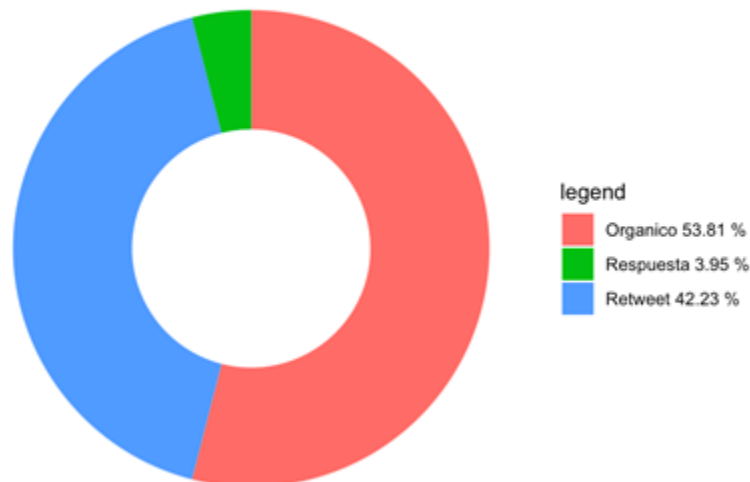
Al analizar los *tweets* de Covid-19 Jalisco, Covid-19 Jalisco y Covid-19 Guadalajara podemos obtener información como la opinión de las personas dentro del Estado, los sentimientos que han expresado en esta red social y cuáles han sido algunas de las tendencias que han existido respecto al tema.

Con estas palabras podemos encontrar:

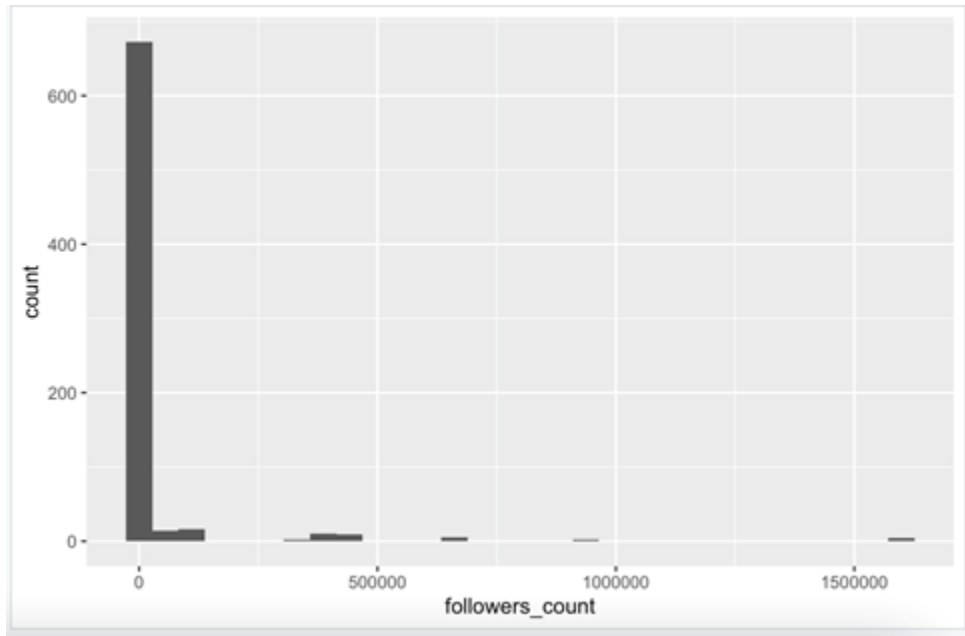
- Covid-19 Guadalajara : 176 *tweets*
- Covid-19 Jalisco: 316 *tweets*
- Covid-19 Jalisco: 242 *tweets*

Dándonos un total de 734 *tweets*, los cuales se dividen en orgánicos que son directamente las personas que publicaron desde su usuario, *retweets* que comparten la misma opinión o les gusta de más la misma publicación y las respuestas que la gente ha escrito esos *tweets*.

Tipos de tweets acerca de Covid en Jalisco

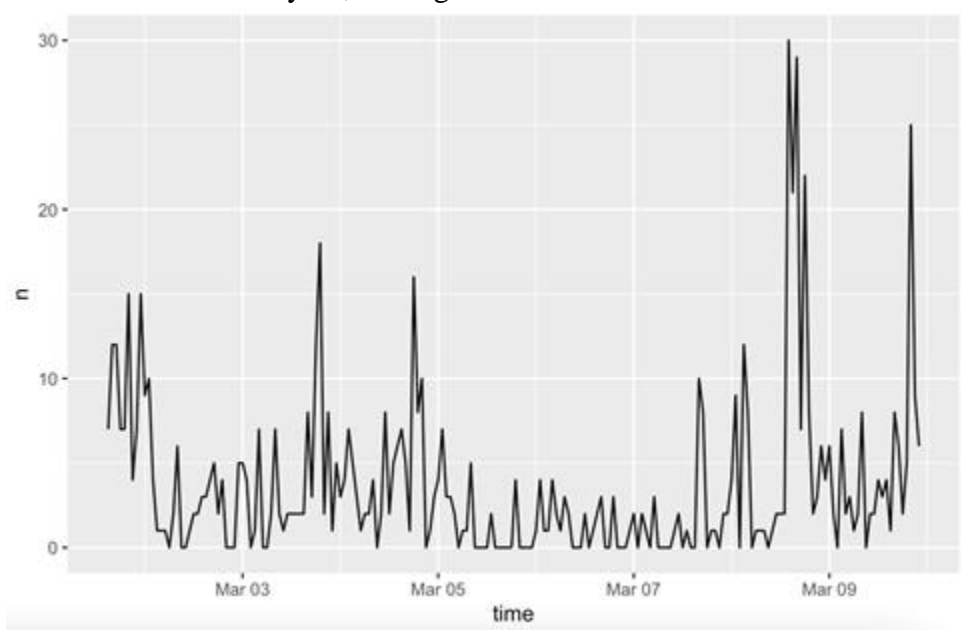


De las cuales podemos encontrar que solo una muy baja cantidad de publicadores cuentan con alguna fama por la red, lo cual nos indica que no son muy famosos.

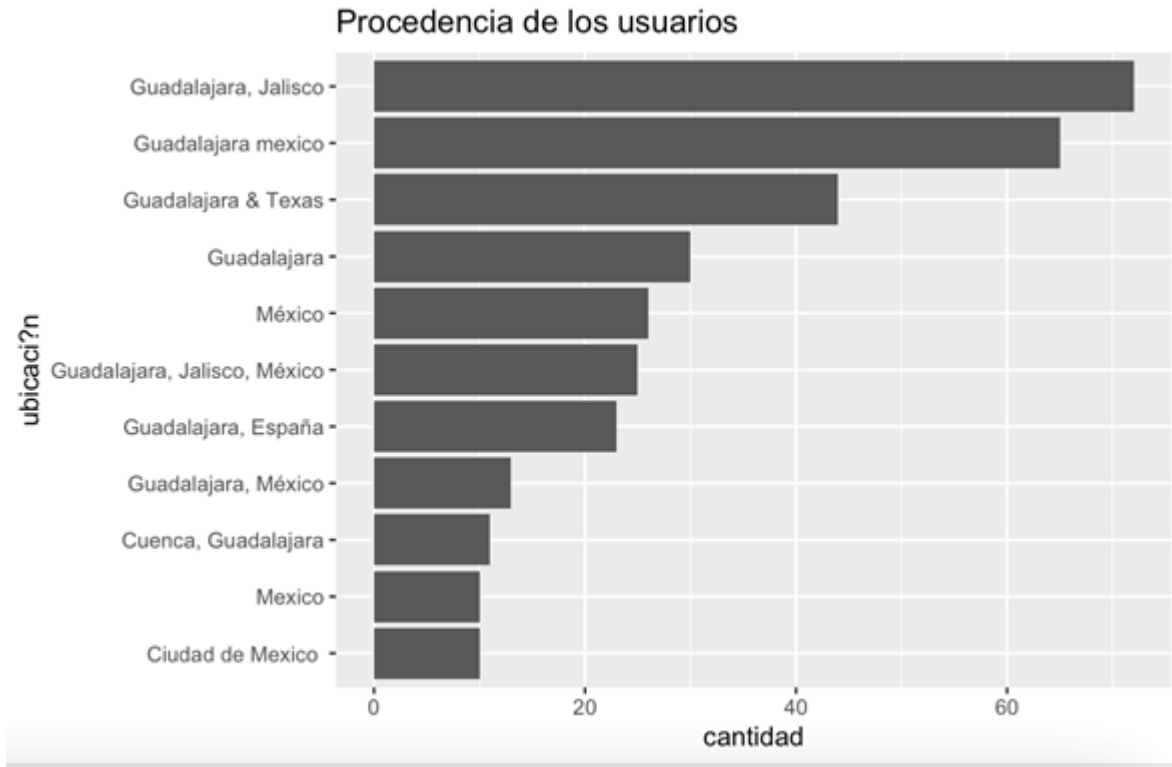


Encontramos que el 9 de marzo se creó una alta tendencia de *tweets* del tema, esto puede explicarse a las campañas de vacunación que empezaron a recorrer los distintos estados de la república y su impacto en la opinión de las personas.

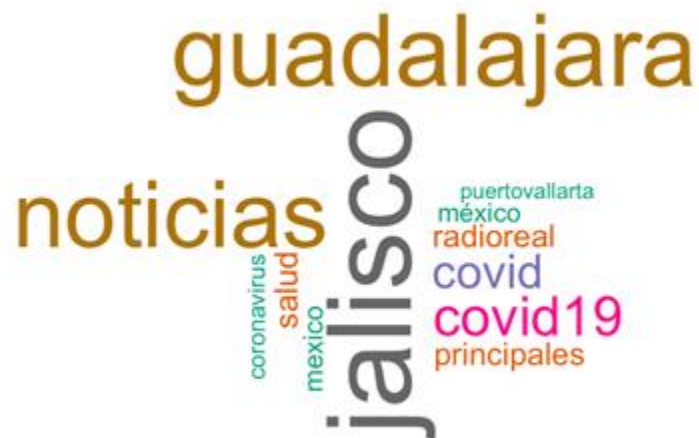
Los usuarios más populares fueron AcustikInforma con 1,598,753 seguidores y Trafico ZMG con 923,665 seguidores. La publicación más famosa respecto al tema fue una de la cuenta de Salud Jalisco con 33 *retweets* y 86,152 seguidores.



La mayor parte de estas publicaciones son publicadas desde la Ciudad de Guadalajara.



Los *hashtags* más utilizados fueron Guadalajara, Jalisco y noticias.

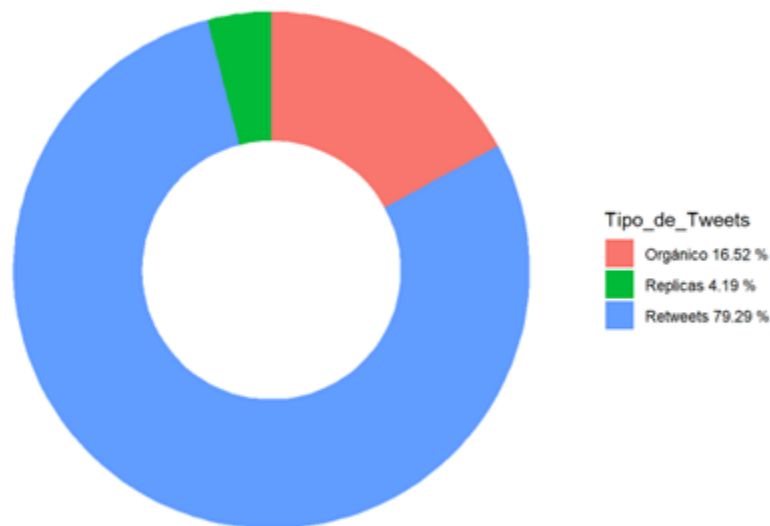


La cuenta que tiene la mayor presencia entre todos los *tweets* acerca del tema es Salud Jalisco. Quien además hemos visto es la cuenta con más *retweets*.



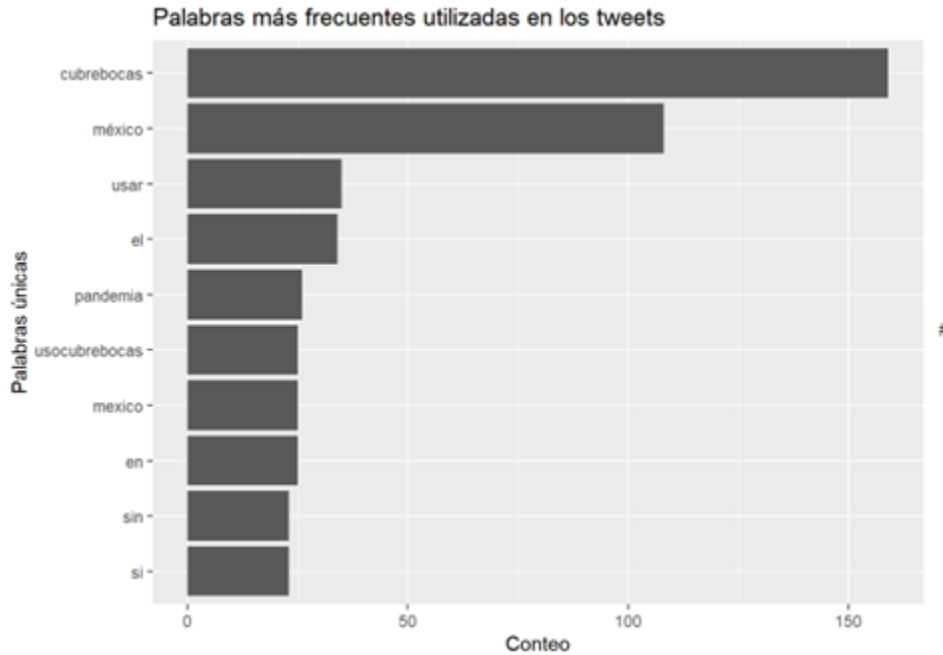
en el contexto de la pandemia por Covid-19. Se descargaron 1,014 *tweets* que contienen las palabras cubrebocas México y 517 con las palabras cubrebocas y Jalisco. Posteriormente, se unieron dichas bases de datos para comenzar a desarrollar el análisis.

El objetivo de este análisis es descubrir la actividad de la opinión sobre el cubrebocas como una de las medidas para hacer frente a la pandemia. Tiene bastante relevancia ya que el uso del cubrebocas ha sido controversial en términos de su efectividad para reducir el contagio. Primeramente, se analizan cuáles *tweets* fueron orgánicos, *retweets* y respuestas o *replies*.



Cerca el 80% de los *tweets* analizados son *retweets*, mientras que solo el 4% son replicas.

Al analizar las palabras más frecuentes encontradas en los *tweets* podemos sobresaltar las palabras “usar” que hace referencia al uso del cubrebocas, “sin” haciendo referencia al no uso del cubrebocas y “en” palabra que hace referencia a lugar, pues es importante recordar que la recomendación es el uso del cubrebocas sobre todo “en” lugares cerrados. Las palabras cubrebocas y México por consecuencia lógica de la descarga de datos se presentan aquí. La palabra “pandemia” hace referencia al contexto en el que se está discutiendo sobre el cubrebocas.



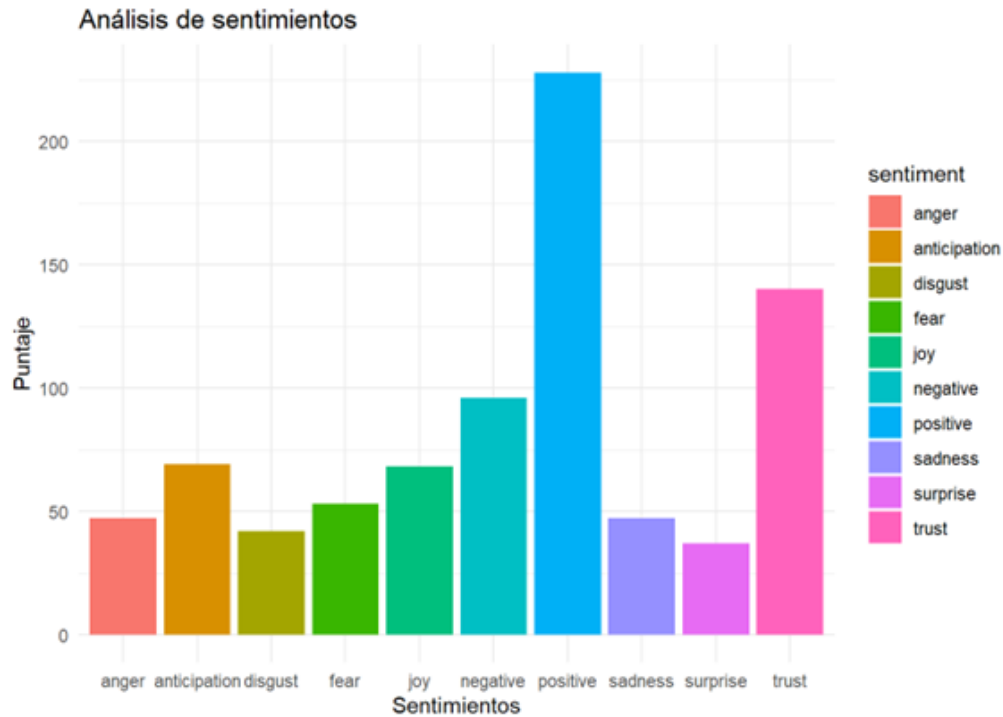
Los *hashtags* más utilizados por su parte son los siguientes:



En su mayoría hacen referencia a la pandemia en México, sin embargo, encontramos algunos que hacen referencia a recomendaciones de salud (#usacubre bocas, #salud, #nuevanormalidad, #cuidatusalud), a lugares específicos (#quintanaroo, #benitojuárez,

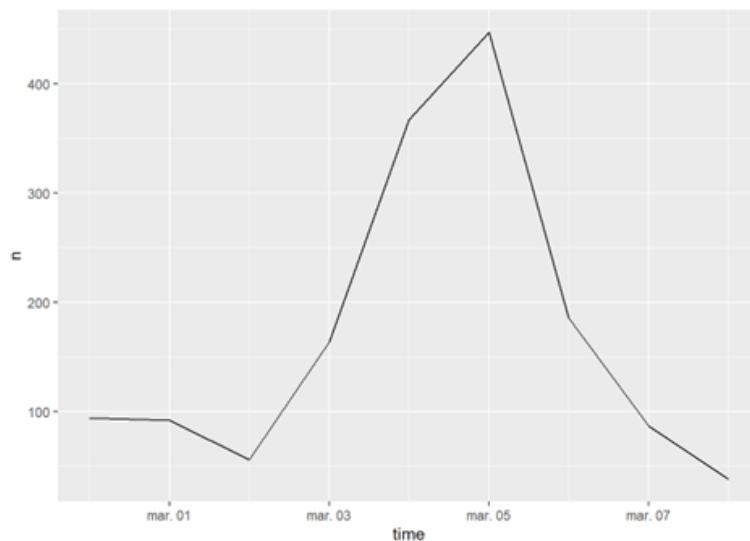






Los sentimientos con más puntaje son positividad y confianza que tienen sentido con el mensaje de la reducción de la pandemia por parte del Dr. Francisco.

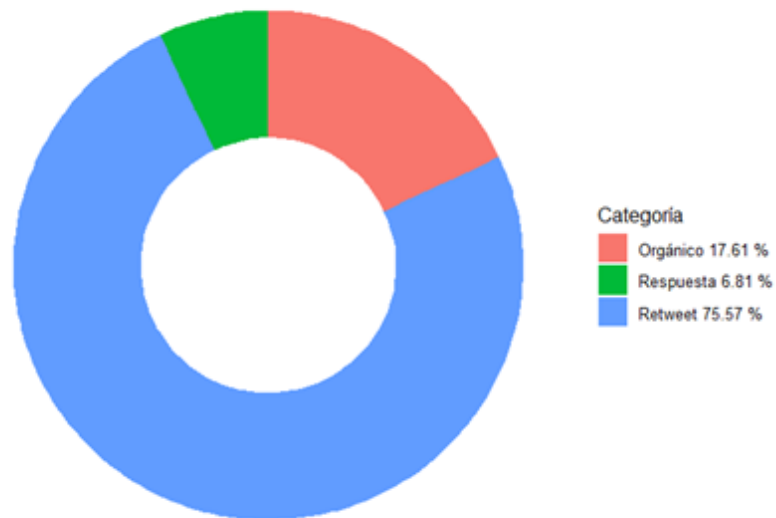
Finalmente, para visualizar como se ha desarrollado el tema del cubrebocas es posible observar la producción de *tweets* respecto del tiempo. Fue a comienzos de este mes que el tema tuvo un pico de relevancia, el cual puede tener relación con el anuncio del semáforo verde generalizado en México que viene acompañado con el refuerzo del uso del cubrebocas como medida para superar la pandemia del Covid-19.



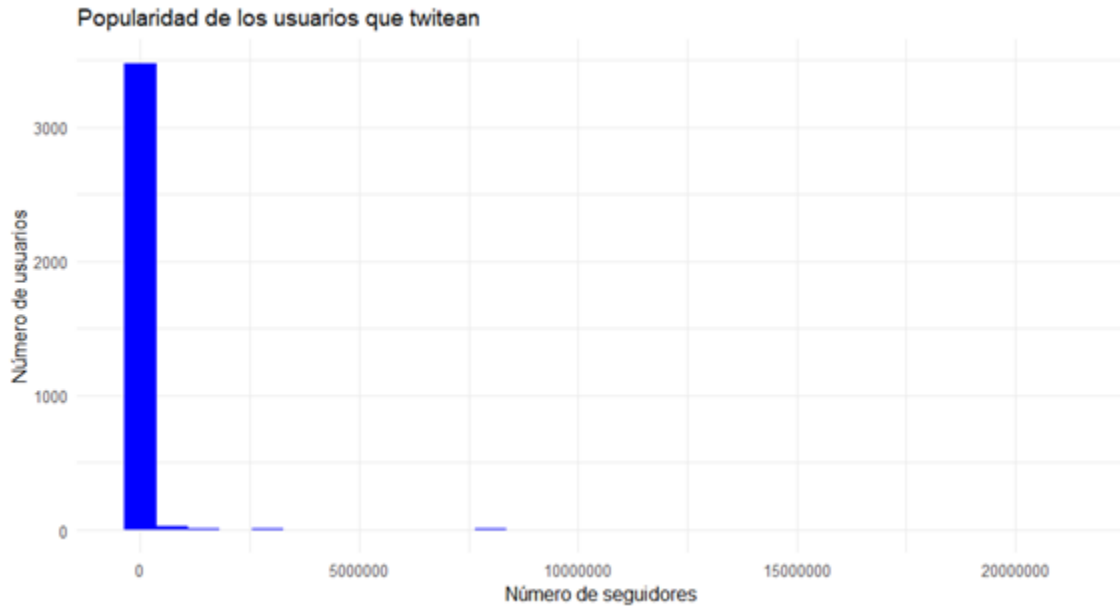
#### 1.5.2.4.3. Pandemia Jalisco y México

El objetivo de este análisis de texto es extraer información acerca de lo que se publica en la red social de Twitter con respecto a la pandemia, tanto en la entidad federativa de Jalisco como en México (esto debido a que sobre Jalisco hay muy pocos *tweets*). Más concretamente, se analizará qué es lo que la gente dice y opina acerca del tema, cuáles son los sentimientos que transmiten, qué *hashtags* utilizan, etcétera. En total, se analizaron 337 *tweets* sobre la pandemia en Jalisco y 3200 sobre la pandemia en México.

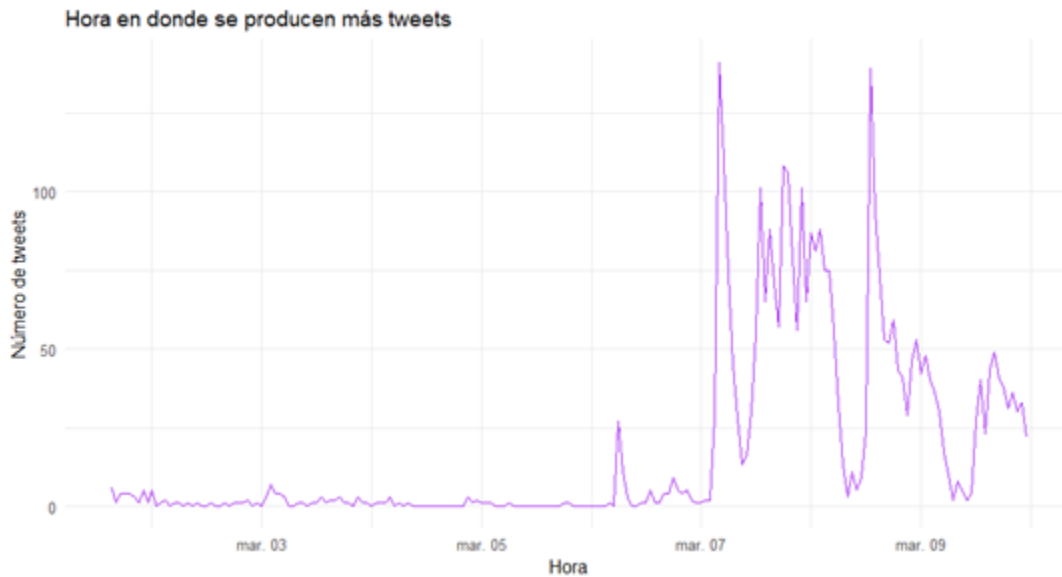
Los *tweets* se dividen en tres categorías: orgánicos, respuestas y *retweets*. La mayoría de los *tweets* analizados son *retweets*, representando el 75.57%. 17.61% son orgánicos y sólo 6.81% son respuestas.



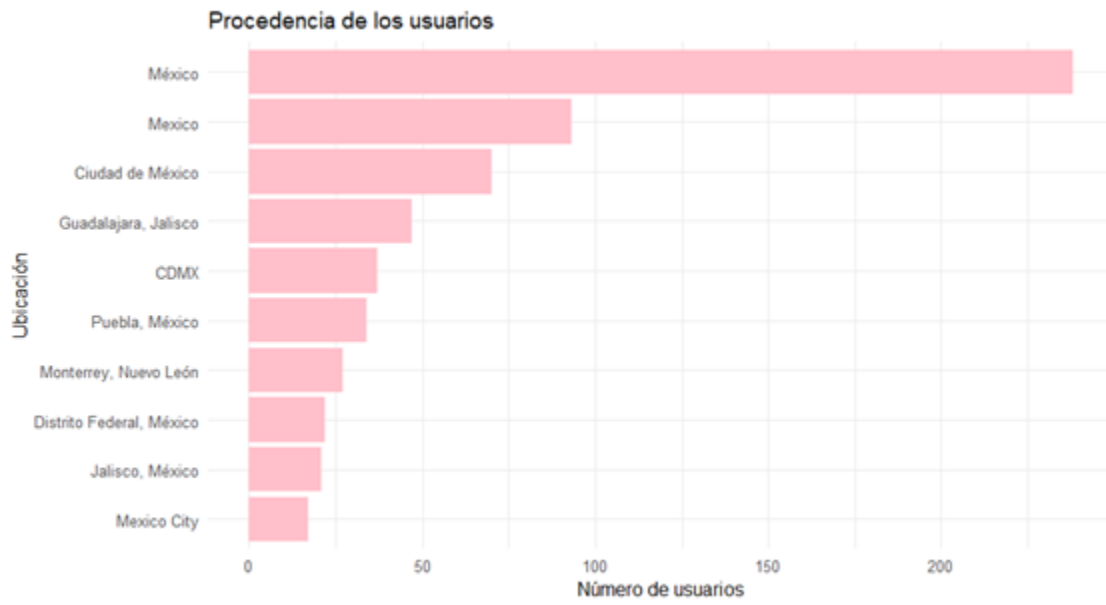
Además, la mayoría de los usuarios que twittean no son tan populares en la plataforma, como lo indica la siguiente gráfica:



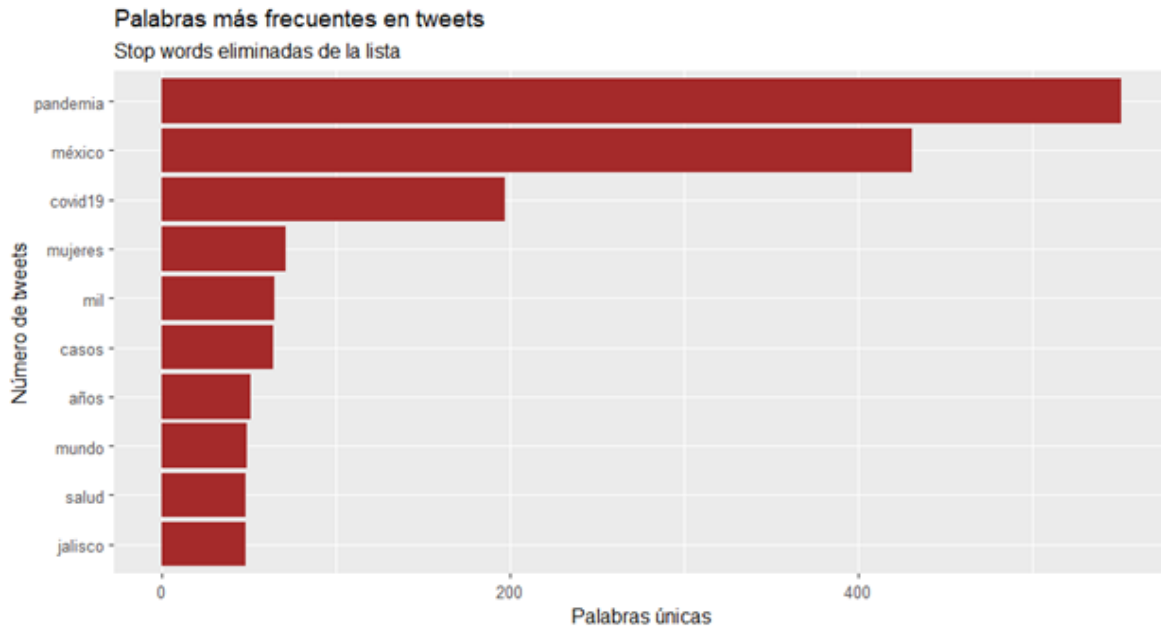
El usuario que más ha *tweeteado* sobre el tema es SerendipiaData, la cual es una cuenta de periodismo de datos (10,515 seguidores), seguida de la cuenta del periódico La Jornada, lajornadaonline (2,697,567 seguidores). Se encontró que durante los últimos tres días de este análisis (del 7 al 9 de marzo) la cantidad de *tweets* sobre el tema ha aumentado considerablemente.



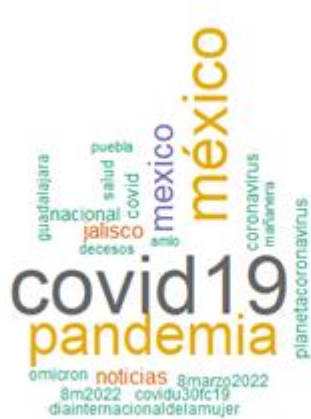
La siguiente gráfica muestra que la mayoría de los usuarios que *tweetean* del tema, lo hacen desde la Ciudad de México, seguida de Guadalajara y Puebla.



La gráfica que se muestra a continuación muestra además que las palabras más frecuentes utilizadas en los *tweets* sobre la Pandemia en Jalisco y México (además de las obvias) son “Covid19”, “Mujeres” y “Mil”.



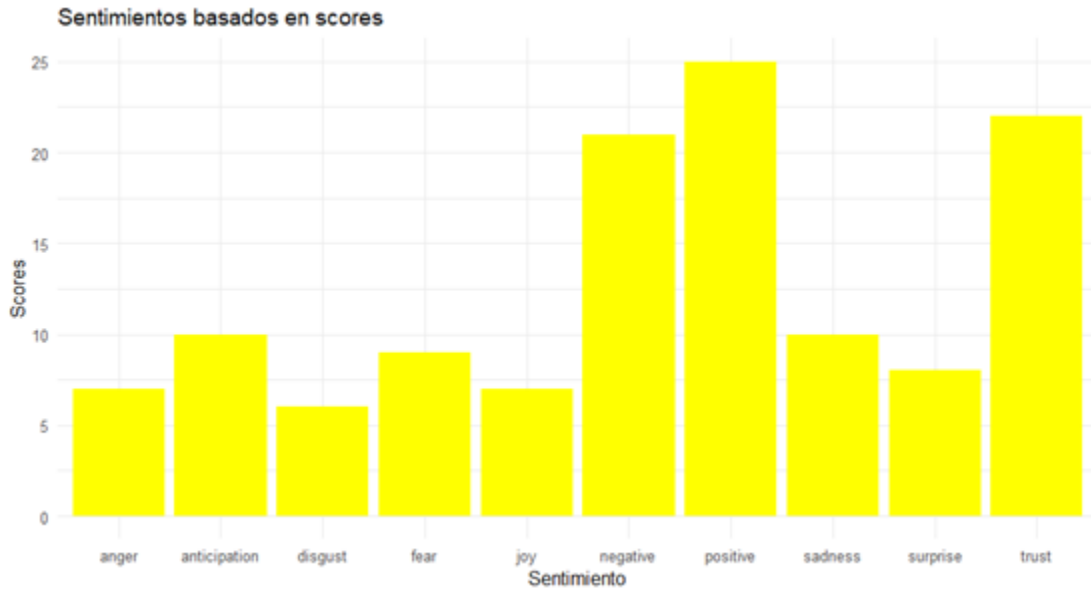
Además, el *hashtag* más utilizado es #covid19:



La persona que más ha *tweeteado* sobre el tema (tanto *retweets*, como respuestas como *tweets* orgánicos) es Javoac (16,872 seguidores), seguido del periodista Joaquín López Dóriga (8,059,831 seguidores) y el Secretario de Relaciones Exteriores, Marcelo Ebrard (2,3216,042 seguidores).



Finalmente, se realizó un análisis sobre los sentimientos que se transmiten en los *tweets* y es sorprendente que el que predomina es la positividad, seguida de Confianza. Eso puede significar que muchas personas confían en que el manejo del gobierno sobre la pandemia ha sido bueno. Sin embargo, el tercer lugar es el sentimiento de negatividad.

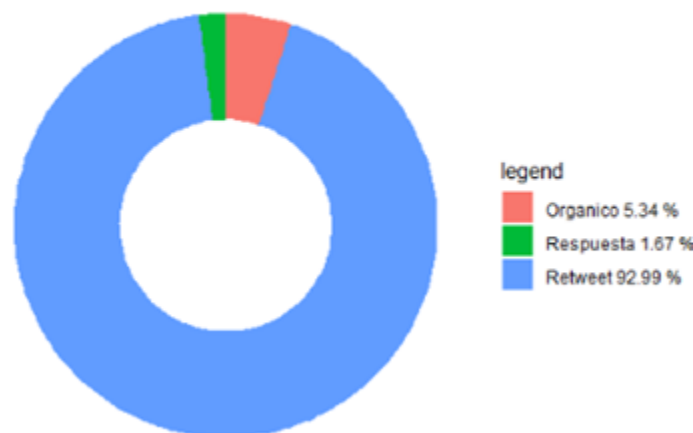


#### 1.5.2.4.4. Vacunación Jalisco y México

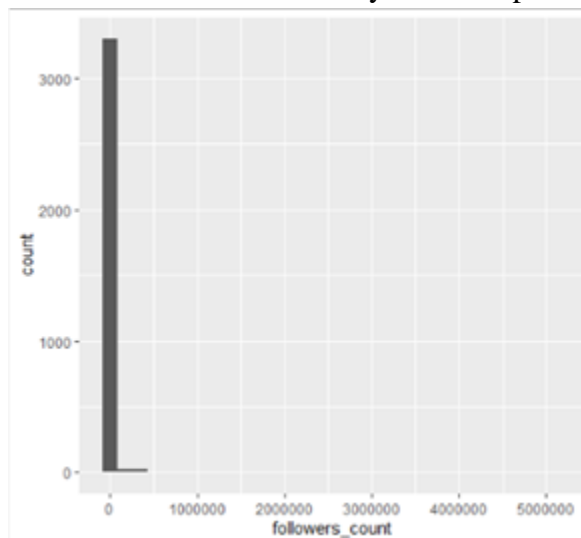
En este análisis de texto, buscaremos encontrar información acerca de lo que se publica en Twitter acerca de la vacunación en Jalisco y en México. Con el objetivo de hacer un análisis de texto, analizar que se comenta, cuáles son los sentimientos que genera la vacunación además de otros *insights* del tema. Se bajó un total de 179 *tweets* de Vacunación Jalisco y un total de 3200 de vacunación México.

Primero, encontramos que estos *tweets* se categorizan en orgánicos, respuestas y *retweets*, aproximadamente el 93% de los *tweets* del tema son *retweets*, lo que quiere decir que la mayoría de los *tweets* del tema son compartidos por usuarios de otro usuario que generó el *tweet* originalmente.

Tipos de tweets acerca de vacunación en Jalisco y MX



Además, también encontramos que la mayoría de los usuarios que twittean del tema no son tan famosos, como se muestra a continuación la mayoría tiene pocos seguidores.

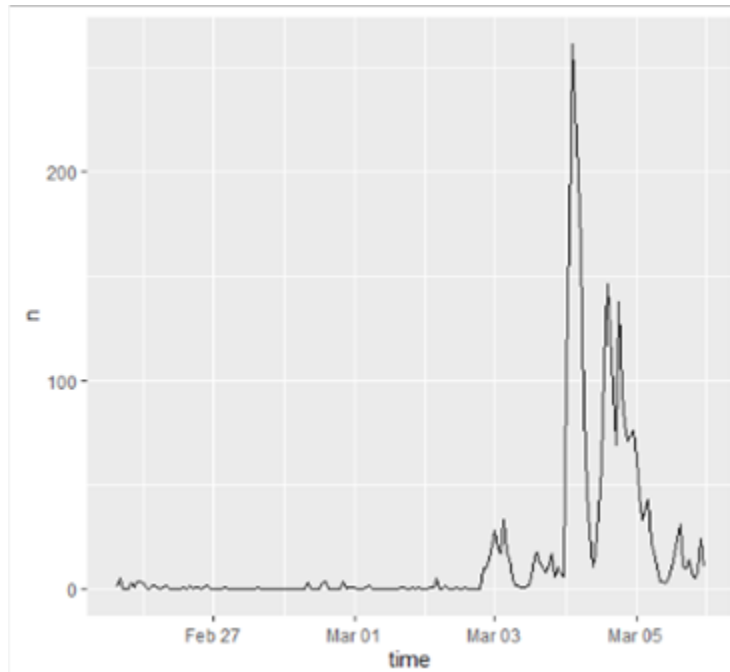


Los usuarios con más seguidores que han twitteado acerca del tema son de la revista proceso (5,088,080 seguidores), la politóloga Denise Dresser (4,461,440 seguidores) y del sub secretario de salud Hugo López-Gatell (1,829,387 seguidores).

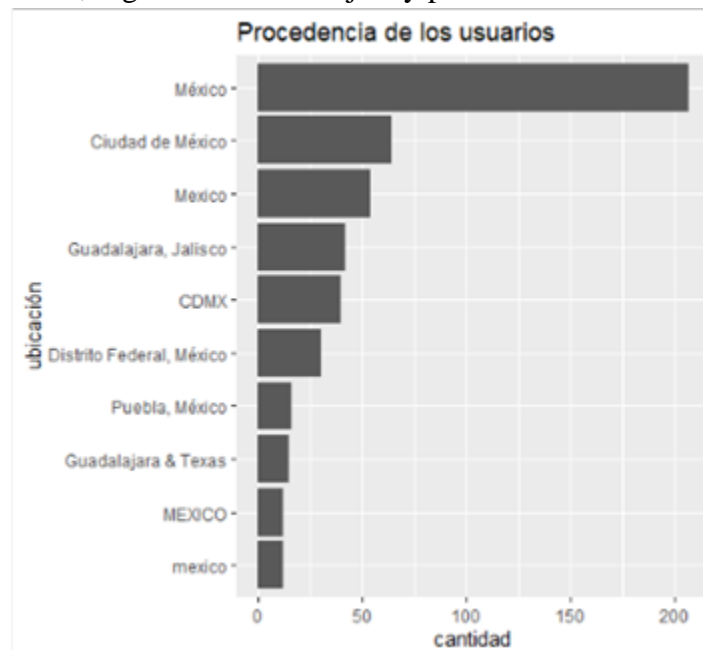
El *tweet* que más “*rts*” ha tenido es uno publicado por el columnista Hector Rossete, con 2,161 *retweets*, que comentaba: “Al 27 de febrero se han registrado 1,202 defunciones en niños, niñas y adolescentes por Covid-19 en México”.

Encontramos que el 4 de marzo del 2022 se realizaron una gran cantidad de *tweets* del tema, esto se puede deber a que en estos días se ha hablado mucho acerca de la tercera dosis de vacunación en el país, a continuación, se anexa la gráfica:





Algo interesante es que la mayoría de los *tweets* acerca del tema se hacen desde la capital del país, Ciudad de México, seguido de Guadalajara y posteriormente de la ciudad de Puebla.



Pasando al análisis de texto de estos *tweets*, encontramos que las palabras más frecuentes utilizadas en el tema de vacunación en Jalisco y México son: vacunación, México, Covid-19, dosis y Jalisco.





La mayoría de los que publican *tweets* al respecto tienen pocos seguidores, mientras que los usuarios que si tienen una gran cantidad de seguidores pertenecen a funcionarios públicos o periodistas. Las dos palabras más comunes en todos los análisis fueron Covid-19 y pandemia. Una gran cantidad de *tweets* vienen desde las ciudades más pobladas de la República: Ciudad de México y Guadalajara.

Para cerrar la conclusión pudimos observar comportamientos muy similares de los usuarios, lo cual nos provee buenos *insights* acerca de qué piensa la población de Jalisco y México acerca de la pandemia.

#### 1.5.2.5. Actividad 6. Razones de cambio

##### **Introducción**

En esta actividad haremos un análisis de las razones de cambio de los casos confirmados y de los hospitalizados, buscando las razones de su relación además de distintas comparaciones que nos puedan ayudar a tener más información acerca del tema y poder avistar y prevenir futuras olas de contagios. Para el apoyo de estos cálculos estaremos utilizando las correlaciones que existen con la tasa de reproducción efectiva del virus ( $R_t$ ) y la tasa de positividad. Estos serán hechos con la información proporcionada por Salud Jalisco y sus dos bases de datos de casos confirmados y de hospitalización en el Estado.

##### **Preguntas a contestar**

- ¿Cuáles fueron las correlaciones más altas entre la positividad, hospitalizados, las tasas de positividad y reproducción efectiva?
- ¿Existe alguna relación entre la razón de cambio de los casos confirmados y la tasa de positividad y qué significa?
- ¿Cómo se ve representada la relación entre la razón de cambio de los casos hospitalizados y la tasa de positividad y qué significa?
- ¿Existe alguna relación entre la razón de cambio de los casos confirmados y la tasa de reproducción efectiva y qué significa?
- ¿Cómo se ve representada la relación entre la razón de cambio de los casos hospitalizados y la tasa de reproducción efectiva y qué significa?

##### **Procedimientos**

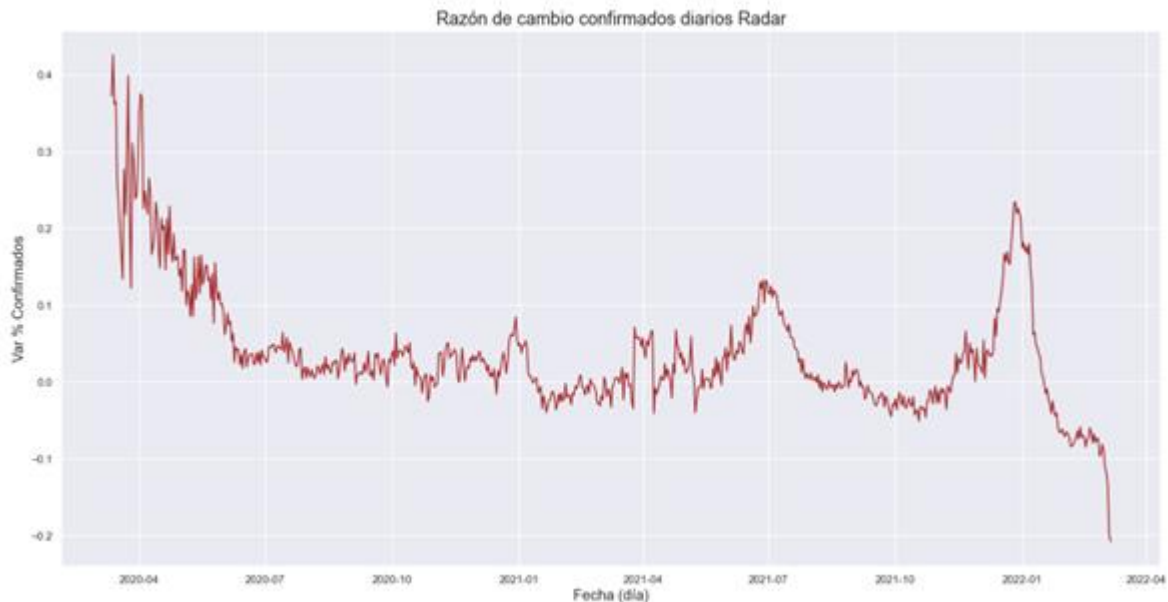
- **Limpieza de bases de datos de Radar Jalisco (Casos Confirmados y Hospitalizados):** Se eliminaron las columnas con información innecesaria, se renombraron las columnas y algunas de las variables de resultados.
- **Transformación de los datos:** Se agruparon los datos por fecha, número de casos positivos y número casos hospitalizados, número de defunciones. Hicimos coincidir fechas y llenado de espacios blancos.
- **Cálculo de variables:** Calculamos la probabilidad de ser positivo dividiendo el número total de pruebas entre el total de casos confirmados dándonos la tasa de

positividad. Para el cálculo de la tasa de reproducción efectiva del virus tomamos en cuenta un cociente entre la suma del día actual hasta dos días atrás y de 4 días hasta 6 días. Indicándonos el número de personas que se contagia por cada persona que da positivo. Además, agrupamos todo en un dataframe y calculamos las correlaciones con una función de la librería Numpy.

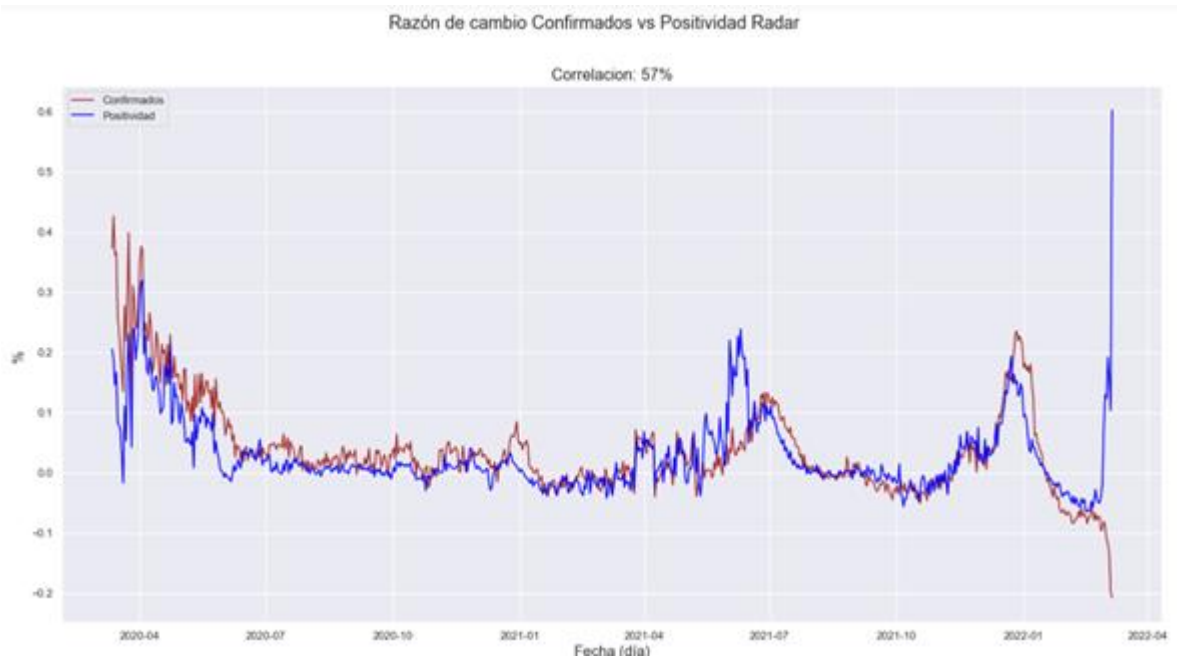
- **Visualización de datos:** Se generan gráficas de los datos, en este acaso utilizamos histogramas y con estas visualizaciones poder entender y poder responder las preguntas a contestar.

## Resultados

En esta primera gráfica encontraremos la razón de cambio de los casos confirmados diarios de la base de datos de Radar Jalisco que incluye dentro de su escala un rango de variación que va de entre el -20% hasta el 40%.



En esta segunda gráfica podremos encontrar que existe una correlación bastante alta entre la razón de cambio de los casos confirmados y la tasa de positividad diaria. Lo que quiere decir que en ciertos periodos se comportaron de manera bastante similar.



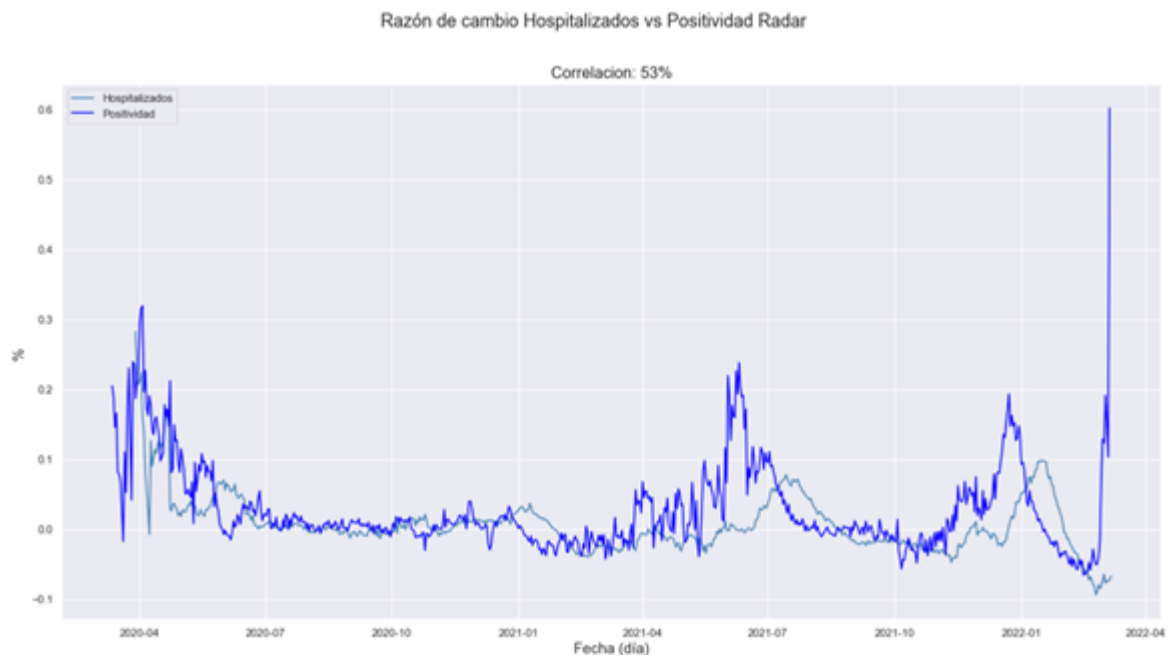
Si siguiendo una visualización que nos muestra los casos confirmados contra la  $R_t$ , encontraremos que la correlación es baja (10%), debido a que la razón de cambio de ambos en distintos periodos tiene un comportamiento totalmente diferente.



En esta representación se muestra la razón de cambio de los casos hospitalizados diarios de la base de datos de Radar Jalisco que incluye dentro de su escala un rango de variación que va de entre el -10% hasta el 30%. Se observa una menor variabilidad que los casos confirmados.



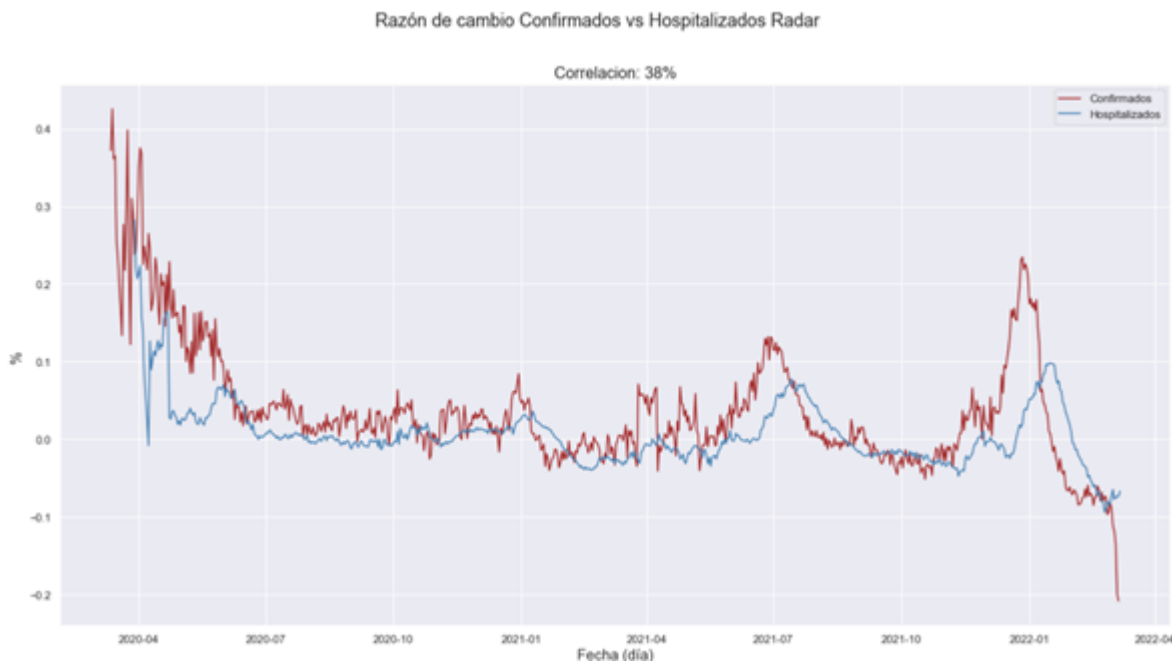
Retomando la imagen anterior y haciéndole una comparativa contra la tasa de positividad de Radar Jalisco podemos encontrar que tienen una alta correlación, la cual también se puede visualizar en esta comparativa. Podemos ver que cuando existe un incremento en la tasa de positividad, poco después encontraremos otro incremento en el número de casos hospitalizados, comportamiento similar al que mencionamos previamente entre la tasa de positividad y la razón de cambio de los hospitalizados.



Avanzando con la siguiente gráfica que muestra la razón de cambio entre hospitalizados y la  $R_t$ , podemos encontrar ahora una correlación negativa, dicho en otras palabras, encontraremos comportamientos en cierto sentido inversos.



Finalmente tenemos las razones de cambio de confirmados y hospitalizados comparadas y cómo podemos ver tienen comportamientos muy similares así como una correlación alta, esto da explicación a la similitud que existe entre las gráficas. No tenemos una correlación más alta porque como podemos ver la naturaleza de los casos es primero ser confirmado y después hospitalizarse.





## Conclusiones

Las razones de cambio tanto de los casos confirmados y los de hospitalizados tienen una fuerte correlación con la tasa de positividad, así como pudimos ver es probable que tengan comportamientos similares. Mientras que el  $R_t$  por su naturaleza no tiene una muy alta correlación con los cambios de confirmados y hospitalizados.

Además, algo interesante que pudimos observar es que la razón de cambio entre los casos confirmados y los hospitalizados tienen una correlación alta, esto se puede deber a que entre más casos confirmados hay, existe una mayor probabilidad de que haya un alza en los hospitalizados. Es por esto el comportamiento tan similar que existe entre ambas métricas.

### 1.5.3. Bloque 3. Modelado

#### 1.5.3.1. Actividad 7. Análisis de separabilidad de variables t-SNE

### Introducción

En esta actividad haremos una exploración de la separabilidad de las distintas variables de la base de datos de SINAVE por medio de un análisis de componentes a partir del algoritmo t-SNE. Este método fue diseñado con la finalidad de reducir la dimensionalidad no lineal para extraer datos de alta dimensión. Con esto obtendremos la separabilidad de las variables. Mostraremos las mejores tres iteraciones en cuanto a separabilidad por año.

### Preguntas a contestar

- ¿Cuál es la relación que existe entre las comorbilidades y si falleció o no falleció?
- ¿Existe alguna relación entre si pereció la persona o no y su edad?
- ¿Cómo se ve representada la relación entre el estado de defunción y si el paciente fue intubado?
- ¿Cuáles son los resultados de este algoritmo a través de los años 2020, 2021 y 2022?
- ¿Hay alguna similitud en estos mismos periodos de tiempo y sus variables?

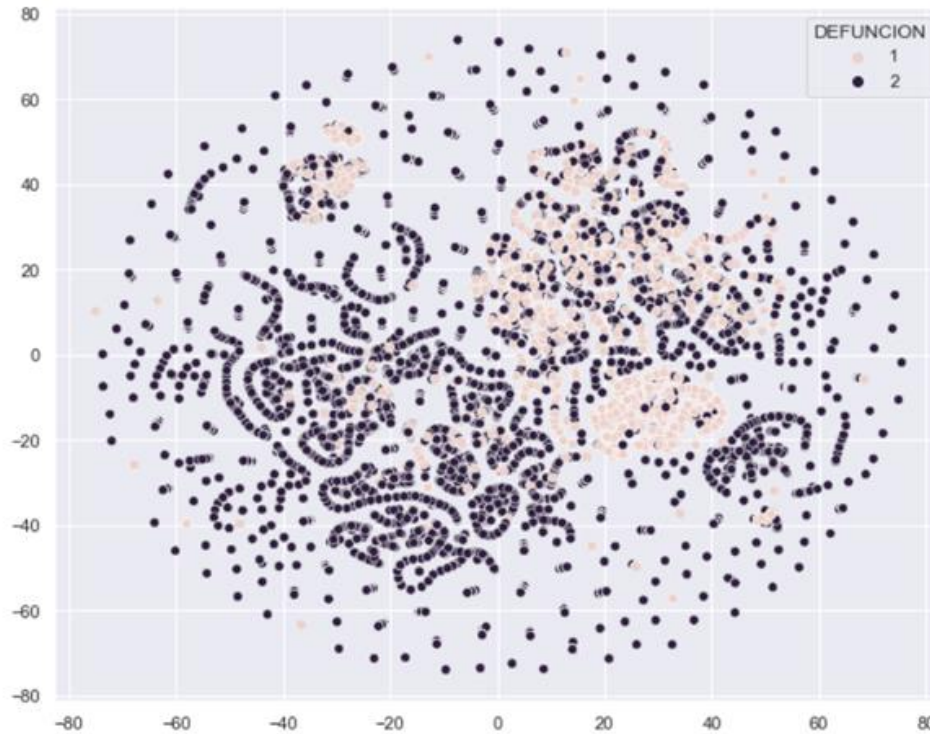
### Procedimientos

- **Limpiezas de bases de datos de SINAVE:** Se eliminaron las columnas con información innecesaria, se renombraron las columnas y algunas de las variables de resultados.
- **Transformación de los datos:** Se estandarizan los datos de la columna de edad, esto ya que tenemos pacientes con más de 100 años cumplidos, mejorando así el accuracy del modelo. Los demás datos no necesitan esta transformación ya que tienen naturaleza booleana.
- **Cálculo de t-SNE:** Con la ayuda de la librería *sklearn* de Python realizamos el algoritmo y la clasificación con las variables de defunción contrastada con las comorbilidades, intubación, edad y sexo. También realizamos distintas iteraciones y dejamos solamente las que nos traen información de utilidad.

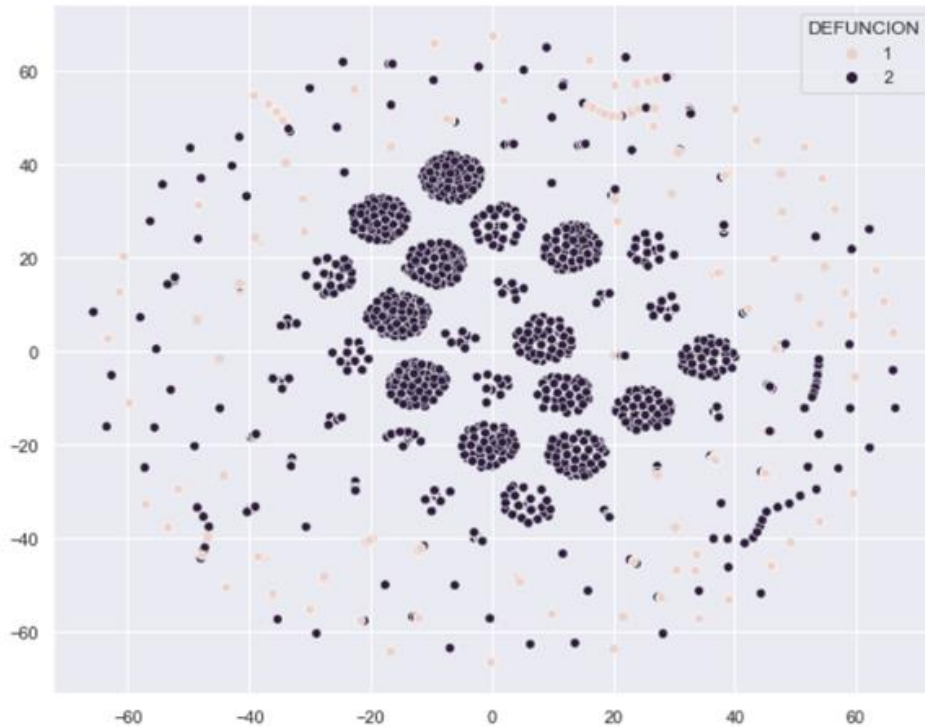
- **Visualización de datos:** Se generan gráficas de los datos, con la ayuda de la librería *seaborn* de Python realizamos gráficos de puntos con el fin de poder visualizar de una manera útil la relación de las variables.

#### 1.5.3.1.1. Año 2020

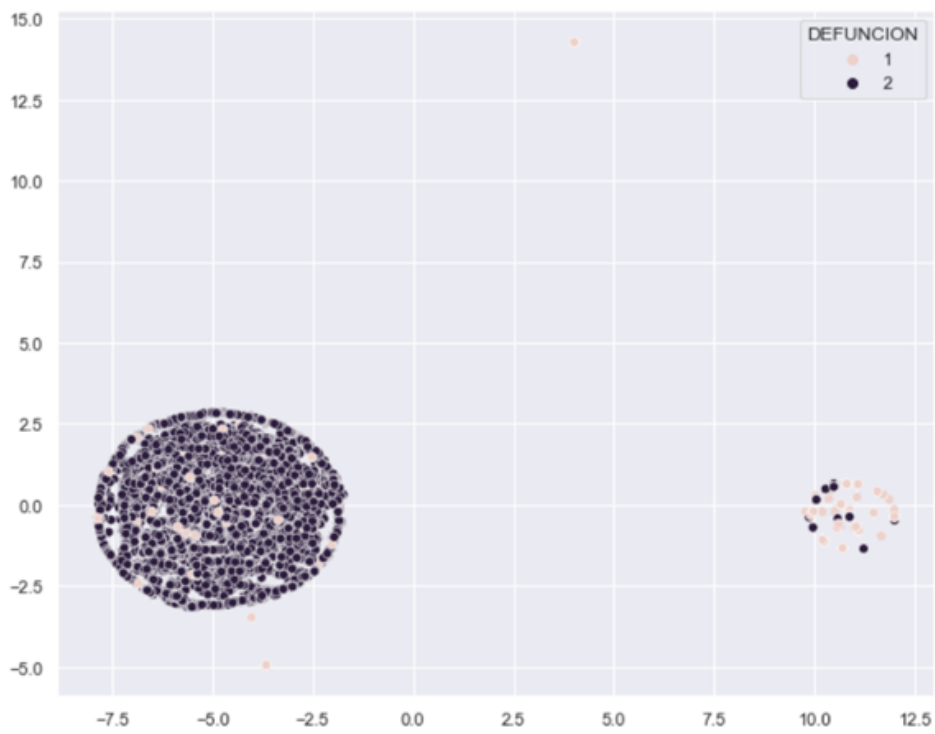
En esta iteración, tomamos todas las variables, siendo estas las comorbilidades, sexo, edad e intubado siendo si falleció o no las variables booleanas: 1 (pereció) y 2 (sigue con vida). Podemos observar que, si bien no se observan clústeres separados entre sí, se puede ver que una parte específica del conjunto de datos, en el lado central derecho del círculo se concentran las defunciones.



Para esta otra iteración, utilizamos únicamente la edad, intubado y tipo de paciente. Como se puede observar se muestra una mayor dificultad para ver las defunciones y las no defunciones.



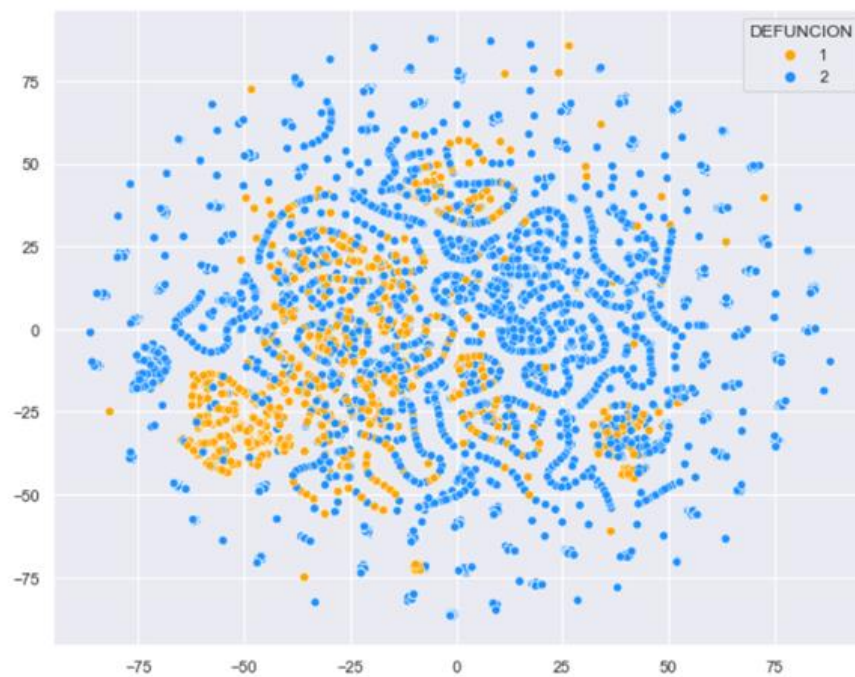
Para la segunda mejor iteración dejamos intubado y tipo de paciente. Como se puede observar al parecer existen dos clústeres en los que existe una mayor cantidad de no fallecidos en el lado izquierdo y en el derecho una mayor cantidad de fallecidos.



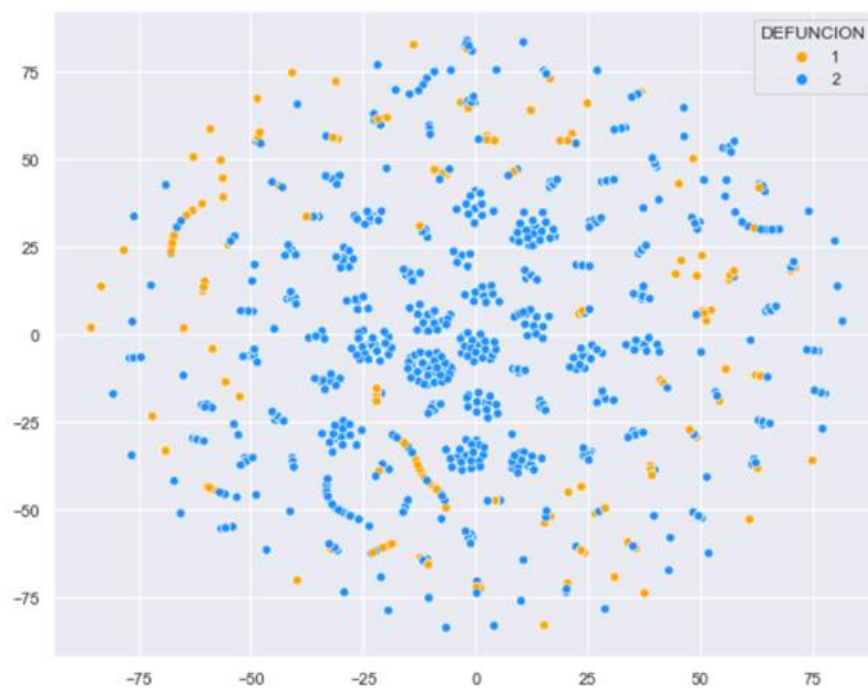
Realizamos el proceso nuevamente en los distintos años.

#### 1.5.3.1.2. Año 2021

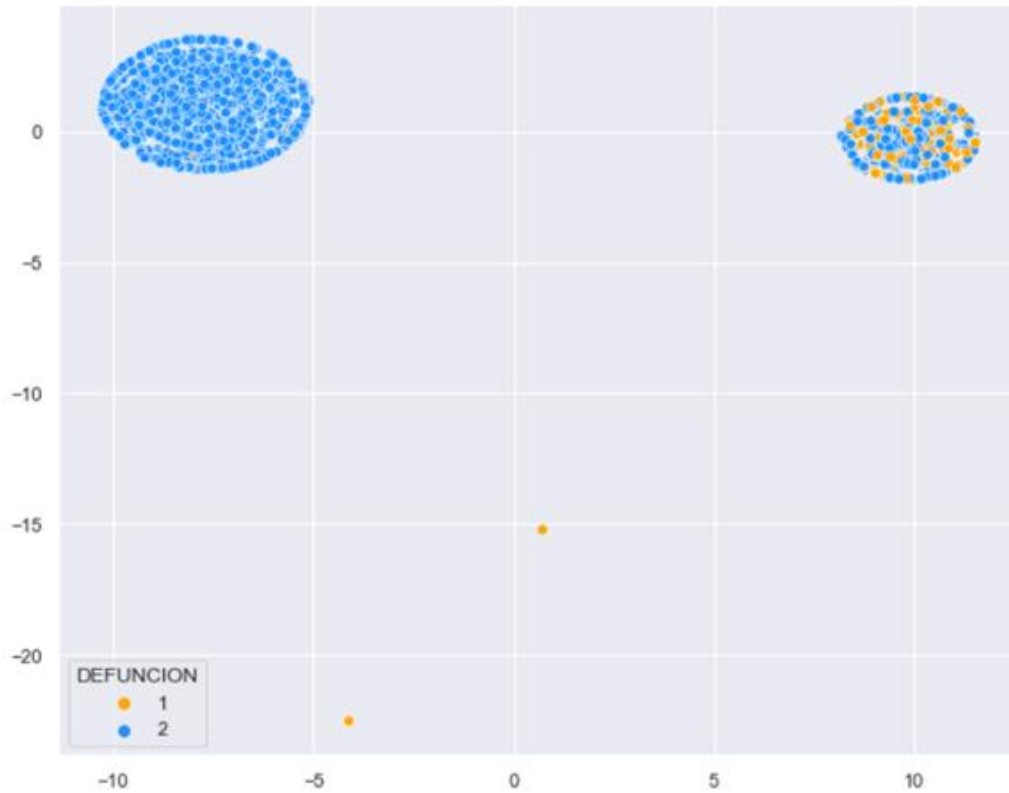
En este año observamos una menor separabilidad que en el anterior, aun así, mostraremos las gráficas en las cuales se observaba una mejor separabilidad.



Para esta primera gráfica tomamos todas las variables para tratar de explicar las defunciones.



Para esta segunda gráfica tomamos edad, tipo de paciente (Hospitalizado o no) e intubado.

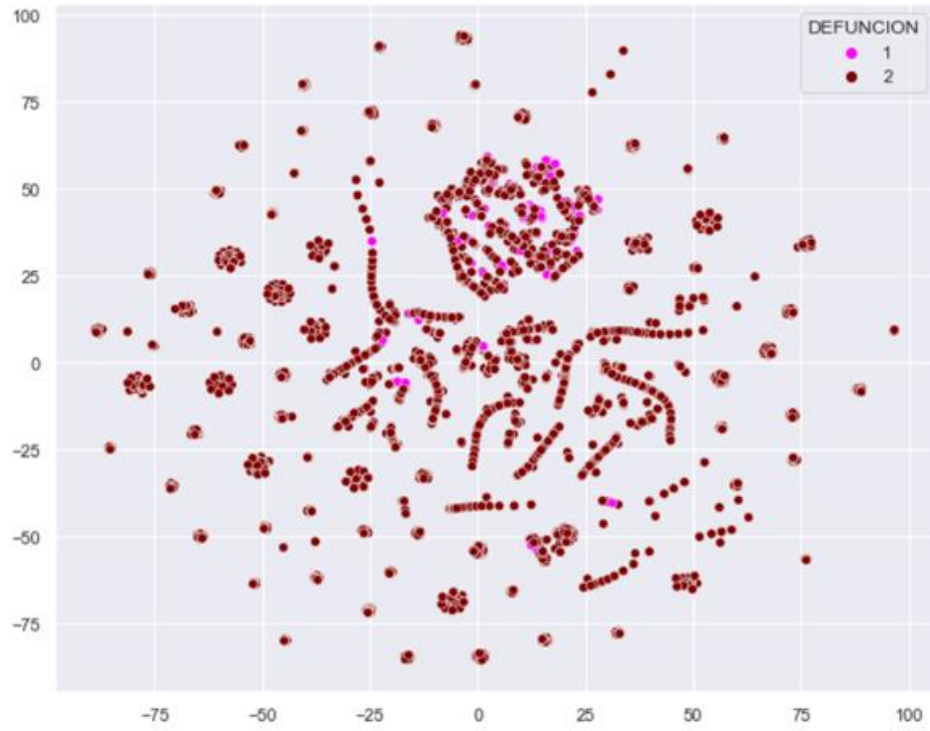


Finalmente, para la última gráfica, que fue en la cual se observa de mejor manera un clúster de no fallecidos en el lado izquierdo y de fallecidos en el lado derecho, tomamos solamente intubados y tipo de paciente.

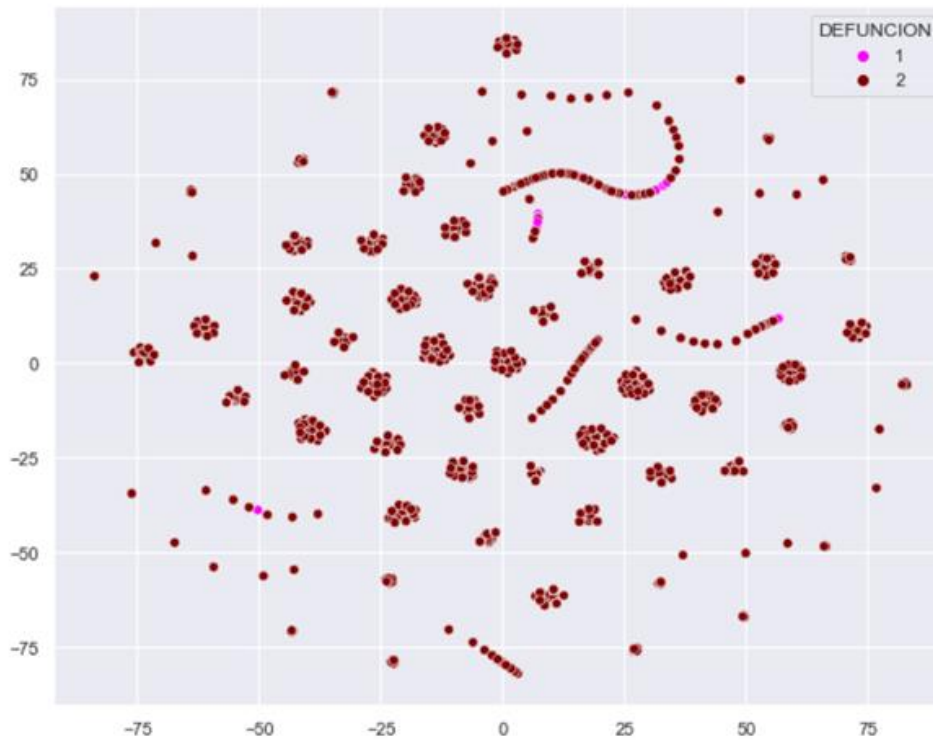
#### 1.5.3.1.3. Año 2022

Para este último año observamos todavía una menor separabilidad que en los años anteriores, esto se puede deber al decrecimiento de la cantidad de fallecidos en la República Mexicana.

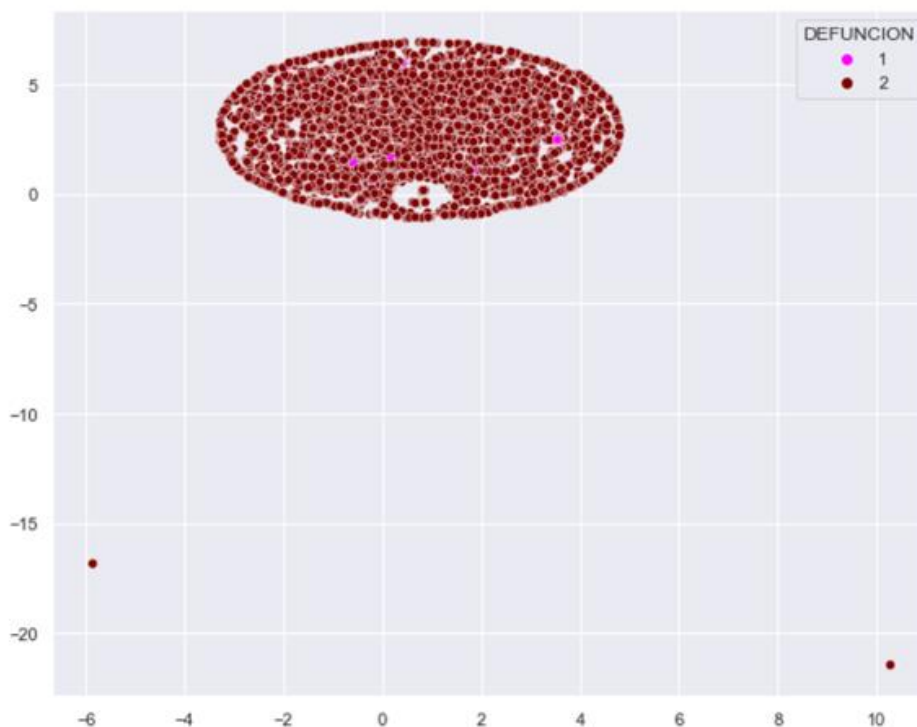
Para el primer análisis tomamos todas las variables:



Para el segundo tomamos edad, tipo de paciente e intubados:



En el último caso tomamos intubados y tipo de paciente.



### Conclusiones

Como fue mencionado durante los procedimientos hechos las gráficas mostradas son los resultados de las tres mejores iteraciones entre todas las variables establecidas y el estado de defunción. Podemos observar que tienen un comportamiento similar y el dato más importante relacionado al estado de defunción es el estado de intubación y el tipo de paciente. Podemos decir que si bien en varias partes se puede observar cierta separabilidad entre las variables utilizadas esta no es muy clara, por lo cual se concluye que no existe una separabilidad entre los datos que se quieren analizar, análogamente esto influye en no tener un score muy alto al realizar un modelo de clasificación o regresión.

Pudimos observar que la separabilidad con el paso de los años va decreciendo, esto se puede deber a que, con cada año, gracias a diferentes factores como la vacunación y la inmunidad natural generada por las personas, el número de defunciones ha disminuido a través del tiempo, y al ser esta la variable que queremos explicar, una menor cantidad de información sesga el resultado.

Finalmente es importante resaltar que durante las distintas iteraciones realizadas también fue considerado el gráfico 3D, pero dado que concluimos que estos datos no tienen separabilidad esta visualización no nos fue de gran utilidad.

## 1.6. Valoración de productos, resultados e impactos

La experiencia en general tuvo un balance positivo por la toda la práctica desarrollada en cuanto a limpieza de datos, análisis y creación de modelos predictivos, esto permitió visibilizar el comportamiento de la pandemia, sus variables y factores a través de los distintos reportes generados durante el PAP. También, es importante destacar que los análisis generados permiten realizar una toma de decisiones que tenga efecto en la pandemia, para disminuir su impacto en la salud, como en la economía.

En específico es importante destacar el análisis sobre el comportamiento de las Olas, pues fue posible identificar que grupo social fue más afectado en cada Ola, monitorear la tasa de positividad y comparar entre los positivos y la tasa de reproducción. En esta última comparación se llegó a la conclusión de que cuando la tasa de reproducción supera la magnitud de 1 se espera un incremento en las personas positivas, mientras que un decremento inferior a uno representa una disminución en el número de positivos. Lo más importante de este hecho es que se puede monitorear la tasa de reproducción e implementar medidas para mantenerla menor a 1.

Otro de los resultados a destacar es el análisis desarrollado sobre cuál es la población más afectada, con las comorbilidades que tienen más implicación para que haya un desarrollo fatal, pues de este análisis se pudo concluir la importancia de proteger a las personas de la tercera edad y aquellas con hipertensión, diabetes y obesidad. Otro desarrollo destacable es la creación del modelo t-SNE que permite ver la separabilidad de las variables y por tanto cuales tienen un mayor impacto con la variable de estudio: personas fallecidas, sin embargo, este análisis no otorgo una separabilidad muy clara.

Finalmente, como área de oportunidad entre los aspectos que se podrían incluir al PAP, es el desarrollo de un sistema de monitoreo continuo a través de un dashboard, el cuál demanda más tiempo, sin embargo, tendría un importante impacto en los resultados finales. Es un logro alcanzable si se delimita el análisis y se precisan variables de seguimiento. Si bien los resultados alcanzados son de un alto valor, este podría ser mayor si se llega a concretar un producto que permita dar seguimiento y presentación a los análisis desarrollados.



## 1.7. Bibliografía y otros recursos

1. Carapia, F. (2021). *Está ocupación hospitalaria en Jalisco al 68 por ciento*. Recuperado 24 enero 2022 de [https://www.mural.com.mx/aplicacioneslibre/preacceso/articulo/default.aspx?\\_\\_rval=1&urlredirect=https://www.mural.com.mx/esta-ocupacion-hospitalaria-en-jalisco-al-68-por-ciento/ar2245642?referer=-7d616165662f3a3a6262623b727a7a7279703b767a783a--](https://www.mural.com.mx/aplicacioneslibre/preacceso/articulo/default.aspx?__rval=1&urlredirect=https://www.mural.com.mx/esta-ocupacion-hospitalaria-en-jalisco-al-68-por-ciento/ar2245642?referer=-7d616165662f3a3a6262623b727a7a7279703b767a783a--)
2. CONACYT. (2022). *COVID-19 Tablero México*. Recuperado 24 enero 2022 de <https://datos.Covid-19.conacyt.mx/#DOView>
3. Gobierno de México. (2022). *Datos abiertos directorio de epidemiología*. Recuperado 25 abril 2022 de <https://www.gob.mx/salud/documentos/datos-abiertos-152127>
4. Gobierno del Estado de Jalisco. (2022). *Bases de datos del Covid-19*. Recuperado 25 abril 2022 de <https://coronavirus.jalisco.gob.mx/bases-de-datos/>
5. Martínez, G. (2020). *Crece ocupación hospitalaria en Jalisco*. Recuperado 24 enero 2022 de <https://traficozmg.com/2020/12/crece-ocupacion-hospitalaria-por-Covid-19-en-jalisco-2/amp/>
6. Velasco, E. (2021). *Miden el impacto del COVID-19 en Jalisco*. Recuperado 24 enero 2022 de [https://iteso.mx/web/general/detalle?group\\_id=25740926](https://iteso.mx/web/general/detalle?group_id=25740926)

## 2. Productos

- **Entender los datos:** En esta entrega se busca obtener *insights* a partir de un Exploratory Data Analysis (EDA). El objetivo es contestar preguntas generales sobre la situación del Covid-19 en Jalisco a partir de diferentes visualizaciones. Por lo tanto, se contrasta la importancia de saber comunicar adecuadamente los hallazgos, pues ellos repercuten directamente en la toma de decisiones.
- **Positividad y  $R_t$ :** Se busca responder a preguntas específicas sobre la evolución de la positividad y la tasa de reproducción efectiva en el estado de Jalisco, agrupando por diferentes variables como edad, sexo, entre otros. Estas agrupaciones se hacen con el objetivo de encontrar cómo afectan dichas variables a la evolución de la positividad y  $R_t$ .
- **Análisis comparativo de municipios y olas de contagio:** Este análisis contrasta las diferencias que existen entre municipios respecto a diferentes variables como la tasa de positividad, la tasa de reproducción efectiva y los casos positivos. También se realiza la misma comparación pero entre olas de contagio. Adicionalmente, se busca contrastar el impacto que tiene el número de casos activos en la evolución de las olas; es decir, cómo es que los picos cambian respecto a los casos activos a lo largo del tiempo.
- **Distribución de edad por olas:** En este análisis el objetivo es entender cómo afectan las diferentes variables a cada grupo de edad. Se responden preguntas como cuál es el grupo de edad que tiene mayor probabilidad de fallecer ante la enfermedad de Covid-19, cuál es aquel que tiende a contagiarse menos, etcétera. Esto se realiza con un enfoque en el estado de Jalisco.
- **Análisis de texto:** Esta entrega tiene la intención de saber qué es lo que opinan los usuarios de la red social Twitter acerca de diferentes aspectos de la pandemia en Jalisco y en México. Esto se hace metiendo como input diferentes palabras relacionadas (“cubre bocas”, “pandemia”, “covid-19” y “vacunación”) y obteniendo *insights* sobre qué sentimientos reflejan los *tweets* que contienen esas palabras, qué tan populares son, etcétera.
- **Razones de cambio:** Este entregable pretende encontrar relaciones entre las razones de cambio de diferentes variables y encontrar patrones con el objetivo de predecir comportamientos futuros. Se contestan preguntas cómo en qué punto del tiempo la correlación fue la más alta entre la razón de cambio de los casos positivos y la de la tasa de positividad.
- **Análisis de separabilidad de variables t-SNE:** Este producto pretende buscar relaciones entre diferentes variables con respecto a la pandemia y contrastar los resultados entre los diferentes años que han sido afectados por la misma (2020, 2021 y 2021). Por ejemplo, se contestan preguntas cómo qué tan relacionado está el hecho de que un paciente esté intubado a que fallezca.

### 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.

La pandemia de Covid-19 en México y en particular en Jalisco llegó a dar más visibilidad al problema sistemático del sector salud y las profundas desigualdades económicas de la sociedad. También, abrió un debate sobre el valor preponderante ante la situación: si la libertad por encima de la vida, si la vida por encima de la libertad o si la justicia sobre ambas. De igual manera hizo notar las capacidades distintas de la sociedad mexicana respecto a las que había tenido para enfrentar epidemias o pandemias pasadas, ahora existió la posibilidad de implementar el desarrollo tecnológico como un mecanismo de respuesta a la propagación del virus: desde el monitoreo y predicción de las variables más importantes hasta un pronto desarrollo de las vacunas.

Este proyecto es una herramienta más dentro del desarrollo tecnológico jalisciense para dar respuesta a la situación pandémica a través del análisis de los datos que se han ido recopilando y el modelado predictivo para una posterior toma de decisiones en la que se ve implicado el gobierno del Estado, por tanto, este proyecto tiene el potencial de influir en las decisiones ejecutivas frente a la pandemia, por lo que hay una gran responsabilidad social y se requiere un cuidado especial en cada uno de los análisis y modelos elaborados por la repercusión, que menor o mayor medida, podría tener en millones de personas.

Cuando se trata de datos es fácil perder la noción de lo que representan, justo en dicho punto entra el aspecto ético de quien los analiza pues se debe entender que detrás de los números hay personas que están siendo afectadas por una enfermedad lo que éticamente debe repercutir en quien genera información y más cuando pueden servir para tomar decisiones que influyen en el estado de la situación. Es importante, pues, generar información veraz y análisis imparciales para poder explicar la realidad con claridad y tomar acciones que sean efectivas.

Finalmente, este proyecto tiene la responsabilidad de presentar un análisis lo más certero posible, a través de la aplicación científica, en el tratamiento de los datos sobre la pandemia de Covid-19, con la intención de que haya una toma de decisiones oportuna con la información generada. El aporte social es la inclusión de la ciencia de datos a las capacidades que tenemos como sociedad para responder a fenómenos naturales que nos afectan, es una mayor capacidad para proteger la salud, la vida y la economía de los jaliscienses.

### 3.1 Sensibilización ante las realidades

Al trabajar en este PAP, encontramos algo difícil trabajar con bases de datos que no solo reflejan la situación actual del país y del estado de Jalisco en términos de casos, sino que también muestran defunciones de personas que tenían seres queridos y familiares, aunque al final todo se reduce a trabajar con números, al tener el trabajo de analizar datos de Covid-19 hay que estar muy sensibilizados y tener la mayor responsabilidad al responder preguntas y generar conclusiones en base a estos datos, debido a la naturaleza de los mismos, no son simples datos, si no que representan personas.

Esto nos dejó a todos un aprendizaje muy importante, al hacer una conclusión o un análisis de datos (que es muy común hacerlo en el área de la Ingeniería Financiera), siempre hay que preguntarse sobre que estamos concluyendo y no solo concluir en base a números, hay casos en los cuales estos números representan un caso muy sensible para alguna persona, y en el contexto en el cual nos encontramos actualmente, con una pandemia azotando todo el mundo, encontramos un claro ejemplo de la responsabilidad que tiene un científico de datos a la hora de reportar. Hay que sensibilizarse ante las personas que podemos afectar con nuestros análisis y tratar de entenderlas, no solo entender los números.

Además, el realizar este proyecto también nos ayudó a entender cómo es que las olas de contagio se generan, y la forma en la cuál en los periodos de gran convivencia social estas se van generando, es por esto por lo que, como equipo, hemos generado un mayor sentido de responsabilidad con la sana distancia y el uso del cubrebocas, para buscar tener una mejor conciencia social, siempre buscando el bien común, cuidándonos a nosotros y a nuestro entorno. Si bien nuestra comunidad en el ITESO tiene un poder adquisitivo en promedio mayor a otras y nosotros podemos acceder a mejores servicios, además de que si nos contagian podemos darnos el lujo de faltar al trabajo a la escuela, hay personas que no pueden darse este lujo, personas que viven al día, personas que laboran en condiciones malas y no tienen fácil acceso al sector salud, al cuidarnos, también apoyamos a que estas personas que residen en nuestro estado tengan una menor probabilidad de contagiarse al salir a trabajar.

Al final, el entorno en el cual nos desarrollamos y su bienestar está directamente relacionado con nuestro bienestar, es por esto que, apoyando desde nuestra propia responsabilidad, haciendo uso del cubrebocas y evitando reuniones con un gran número de personas en periodos que pudimos identificar como más propensos a generar olas de contagios (como por ejemplo, Navidad), podemos poner nuestro granito de arena para sensibilizarnos con las personas que han sido o pueden ser afectadas por esta pandemia, al final, nuestra comunidad es parte de nuestro mismo contexto y genera sinergia cuando los individuos de la misma son responsables y sensibles.

Para terminar, podemos concluir que no solo este proyecto nos trajo aprendizajes a nivel profesional, sino también a nivel personal, pero uniendo estos dos rubros para generar una

mayor conciencia social y ser más humanos a la hora de trabajar con datos que pueden ser sensibles, además de una mayor conciencia social y responsabilidad con nosotros y nuestro entorno.

### 3.2 Aprendizajes logrados

**Aguirre Orozco Miguel de Jesús:** Gracias a mi conocimiento adquirido con respecto a la problemática del COVID-19 pude ser capaz de proporcionar la información de las fuentes de datos necesarias y con esto dar un asertivo soporte y útiles puntos de vista para la selección del mejor modelo predictivo y clasificación de los datos.

Pude aplicar y conseguir más experiencia en los lenguajes de programación R y Python, además de poder identificar tendencias en los datos, pude mejorar mi relación de trabajo con mis compañeros ya que siempre fuimos muy organizados y delegamos adecuadamente las responsabilidades, dando fruto a un gran ambiente de trabajo productivo. Personalmente volvería a trabajar con el mismo equipo en búsqueda de desarrollar nuevos proyectos, además de incluso dejar ya hecho varios reportes de calidad que nos puedan ser de ayuda posteriormente.

**Flores Ortiz Moisés:** Mi rol dentro de este equipo me permitió utilizar y desarrollar mis habilidades de análisis mediante herramientas de programación en los lenguajes de R Studio y Python, lo cual me permitió dar soporte en la selección del mejor modelo predictivo y poder solucionar problemas derivados al gran tamaño del archivo de SINAVE. Además, también tuve la oportunidad de mejorar mis conocimientos e implementación enfocado en las gráficas derivadas de los datos. En el equipo también pude aprender mucho de mis compañeros y sus diferentes perspectivas que apoyaban para cada problema. Pudimos desarrollar un sano ambiente de aprendizaje y trabajo que sin duda me gustaría poder repetir, gracias a este proyecto pude afianzar y pulir mis enseñanzas como ingeniero financiero y también profesionalmente el desarrollo del proyecto me permitió aportar nuevas ideas y puntos de vista en los problemas que se presentan en el trabajo.

**González Vázquez Sean Nicolas:** El papel del cual me hice responsable durante el desarrollo de este proyecto fue la limpieza, transformación y visualización de las variables presentadas en las bases de datos de la enfermedad del COVID-19. Me permitió practicar mis conocimientos acerca del modelado y predicción de variables mediante la utilización de las habilidades tecnológicas aprendidas a lo largo de la carrera. Las herramientas de clasificación de datos como lo fueron R y Python nos permitieron desarrollar notebooks y visualizaciones para poder encontrar tendencias en los datos y con esto realizar predicciones de las variables encontradas como más importantes. El peso del archivo de SINAVE y la poca información proporcionada por Radar Jalisco hizo que la limpieza y transformación fueran complicadas, sin embargo, con un muy buen trabajo del equipo pudimos resolverlo. Me siento muy feliz con los resultados que pudimos encontrar y los pocos errores finales al momento de revisión,

para mí el equipo conformado es bastante valioso y mi aportación de delegación de responsabilidades hizo que pudiéramos ser muy organizados.

**Rosas Medellín Miriam Eunice:** Gracias al trabajo de mis compañeros pude proporcionar soporte en la selección del mejor modelo predictivo y el desarrollo e implementación de visualizaciones, además de que con ayuda de las enseñanzas por parte de la Dra. Montoya obtuve nuevas habilidades que complementan mis conocimientos de R y Python. Algunos de los retos que me fueron presentados fue que encontrábamos demasiada información y no siempre toda esta es útil, gracias a mi criterio pude reconocer las variables y gráficas más importantes y también poder optimizar de una mejor manera el desarrollo del modelo predictivo. El esquema de trabajo al cual nos adaptamos en este proyecto nos permitió poder comparar nuestros resultados con los demás equipos creando un gran ambiente de aprendizaje y trabajo.