

Predicciones de Casos Positivos y Fallecidos Covid-19 Mediante Inteligencia Artificial Para Apoyar al Sector Turístico en México.

SocialData Challenge 4.0, Datalab Community

Integrantes:

Angel Gonzalez Espinosa

David Alejandro Ozuna Santiago

Juan Antonio Ríos Mercado

Nery Asaid Delgado Estrada

Paula Andrea Abad

GIT REPO: [TURISMEX](#)

20 de Septiembre de 2021

Descripción del Problema y Entendimiento del Contexto

Desde el inicio de la pandemia uno de los sectores más afectados fue el sector turístico, presentando según datos de [datatur](#), una pérdida del 49.3% en la llegada de turistas internacionales entre enero del 2020 y enero del 2021, así como una baja en la ocupación hotelera del 55.1% en el mismo periodo de tiempo, afectando esto a muchos empresarios y trabajadores de este sector en general.

Según [Naranjo, 2021](#), este sector representa casi el 8.7% del Producto Interno Bruto (PIB) y da empleo directo a más de 4.5 millones de personas. Si agregamos los empleos indirectos, el sector turístico da empleos a más de 10 millones de personas a lo largo y ancho del país. Según estimaciones del Consejo Empresarial Nacional Turístico (CNET) y la Universidad Anáhuac, en el 2019 esta industria generó 25,000 millones de dólares en ingresos, a través de más de 45 millones de turistas que nos visitaron. Antes de la pandemia, México era el séptimo destino turístico a nivel mundial. Desde la primera alarma de esta pandemia, toda la actividad económica relacionada con el turismo sufrió un golpe muy fuerte. El cierre de las economías, cancelación de vuelos y actividades no esenciales han tenido un impacto directo en innumerables negocios y regiones de nuestro país. El CNET estima que únicamente en el sector turismo se han perdido más de 1 Millón de empleos, han cerrado más del 80% de los hoteles y han dejado de operar más de 50,000 restaurantes. Todo esto ha llevado a que el sector turismo pierda más de la mitad de sus ingresos y pierda casi el 45% de su relevancia en la economía de México, donde se estima que ahora contribuye únicamente al 4.9% del PIB. Estamos ante la peor crisis de nuestra historia. Todo esto está poniendo en riesgo real lo que hemos construido en Turismo con tanto esfuerzo en décadas de trabajo.

Actualmente el consejo Nacional Empresarial Turístico de México pronosticó en un informe de enero que las llegadas de turistas internacionales podrían aumentar un 10% este año, aunque aún estarían un 40% por debajo de las de 2019 y prevé un "muy largo" camino para la recuperación del sector conforme a [Gonzales Diaz en 2021](#).

Alcance

Se plantea analizar los datos en los siguientes municipios ya que aparecen en el atlas turístico de México con mayor afluencia de turistas.

CIUDAD	ESTADO
Acapulco	Guerrero
Taxco	Guerrero
Los Cabos	Baja California Sur
Ciudad de México	Distrito Federal
Guadalajara	Jalisco

Cancun	Quintana Roo
Mérida	Yucatán
Puerto Vallarta	Nayarit

Definición de usuarios, clientes y/o beneficiarios de la solución.

¿Quién es el usuario?

- Turistas

¿Quién se beneficia?

- El sector turístico: Aerolíneas, agencias de viajes, Restaurantes, Transportadores, Guías, Operadores.

Preguntas detonantes de la investigación

¿Es posible predecir los brotes de Covid-19 en los principales destinos turísticos de México?

¿Estas predicciones permitirán establecer estrategias para orientar a los turistas nacionales e internacionales a visitar los lugares con menor volumen de contagios?

¿La ciencia de datos puede beneficiar al sector turístico?

Hipótesis

Los algoritmos de inteligencia artificial y los modelos de series de tiempo son eficaces para predecir la positividad y letalidad que podría tener el Covid-19.

Hipótesis particulares

- Mediante los algoritmos de series de tiempo son eficaces para predecir un brote de Covid-19.
- El sector turístico se ve beneficiado al tener una herramienta que les permita saber con anticipación en qué sitios turísticos se espera un incremento en el número de casos de Covid-19.

Objetivo

General

Usando los datos proporcionados por el gobierno mexicano sobre infecciones diarias de Covid-19, planteamos analizar mediante alguna técnica de machine learning, deep learning o series de tiempo aquellos destinos con un potencial turístico, para obtener una métrica que nos permita conocer a qué lugares sería posible viajar con el menor riesgo de contagio y/o letalidad.

Específicos

- Explorar y analizar los datos provistos por el gobierno de México en la página de [datos abiertos](#) de Covid-19 y el atlas turístico de México.
- Cargar y transformar los datos de Covid-19 y cruzar la información con los datos de turismo.
- Analizar los datos de mortalidad, contagio y esparcimiento de la enfermedad.
- Encontrar el mejor modelo para predecir el comportamiento de los siguientes días.
- Presentar estos datos de una forma amigable para el usuario.

Marco teórico

La existente relación entre turismo y crecimiento económico ha sido analizada desde múltiples enfoques teóricos ampliamente. Entender mejor cuales son los mecanismos de transmisión del turismo en la actividad económica, es preciso previamente comprender cuales son los factores determinantes del crecimiento económico.

Conforme a Lambogglia (2014) los primeros estudios sobre el crecimiento económico se presentan en el siglo XVIII con la teoría clásica del crecimiento económico, en la cual se plantean el crecimiento económico y sus factores de una manera poco flexible. Smith (1776) afirma que el crecimiento económico y la riqueza de las naciones obedecen a factores como el trabajo y el progreso técnico, factores que a su vez están determinados por la especialización del trabajo, por la tendencia al intercambio, tamaño del mercado y finalmente por la acumulación de capital, la cual se la considera el motor del crecimiento.

Posteriormente, Ricardo (1817) incorpora a la teoría de A. Smith los rendimientos decrecientes, manteniendo la posibilidad de estados estacionarios y sugiere como medida para contrarrestar este estado, el aumento del capital y de la eficacia productiva, definiendo al ahorro como un determinante del crecimiento.

Poco después, Malthus (1820) afirma, que el exceso de ahorro, el consumo escaso y el comportamiento de la población, que genera rendimientos decrecientes, son factores que afectan negativamente en el crecimiento económico por lo que es necesario incentivar la demanda. Para esto propone que adicionalmente al aumento de la inversión, se promueva el aumento de la oferta para que esta a su vez influya en la dinámica de la población.

Schumpeter (1912) incorpora al análisis económico factores como la ciencia y la tecnología ya que propone que es por medio de las innovaciones que se logrará la acumulación de capital y plantea además que el crecimiento económico es cíclico y tiene dos fases, la estacionaria y la del crecimiento económico. En la primera, la economía no crece debido a que no existe innovación en los procesos productivos y en cierto punto el desarrollo tecnológico está detenido; por lo tanto, para pasar a la segunda etapa es necesaria la innovación productiva y el aumento de la inversión con el fin de atraer beneficios para la empresa innovadora, siendo estos beneficios los que animarán a las empresas competidoras a innovar e invertir. Una vez que el proceso innovador y la inversión se detienen, la economía entra nuevamente en la fase estacionaria, al menos que nuevas innovaciones sean introducidas en el mercado y los competidores las reproduzcan.

Aun cuando el turismo es uno de los más grandes empleadores y exportadores de servicios, y que para la economía mundial es uno de los sectores principales, existen limitados estudios acerca de la contribución de este sector al crecimiento de la economía de los países (Balaguer, Cantavella, 2002).

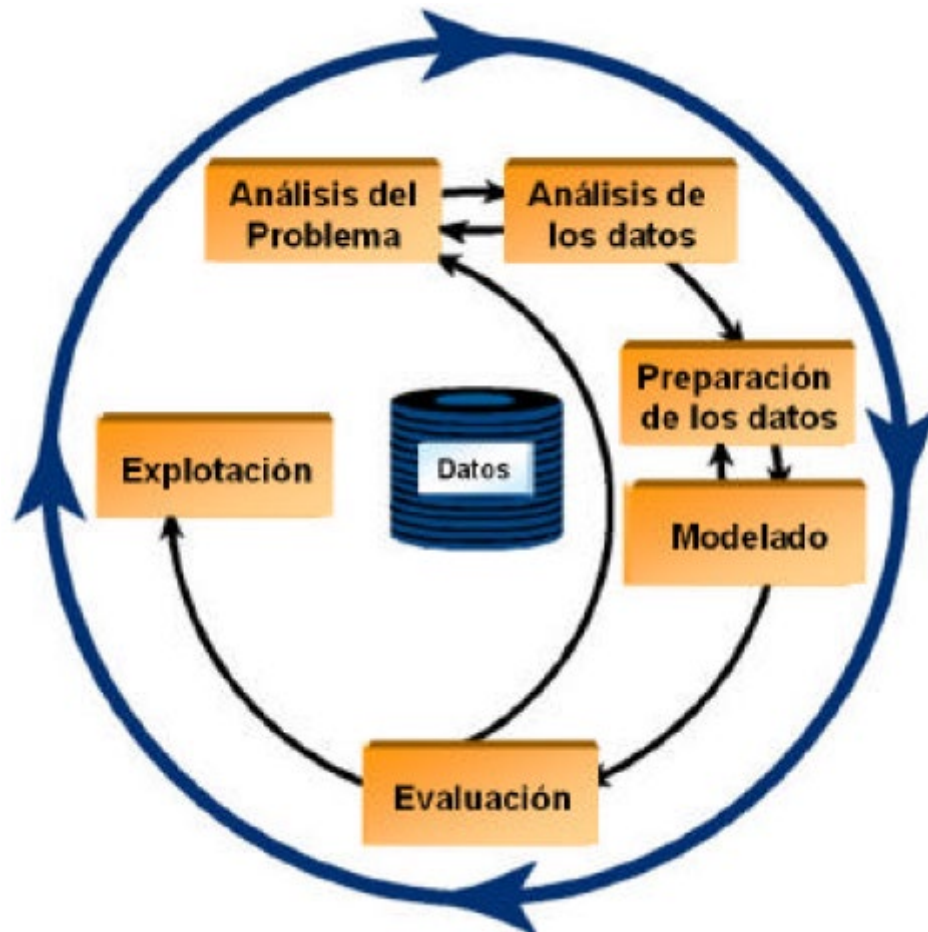
El turismo ha sido un sector en expansión en la economía de muchos países alrededor del mundo, debido a una serie de factores que afectan favorablemente a este sector. Entre ellos podemos mencionar, la curiosidad que motiva a los seres humanos a viajar y conocer nuevos lugares, nuevas culturas y personas; así como, el mejoramiento de la calidad de vida, del nivel educativo y del empleo, la mejora en los servicios de transporte en muchos países, y la continuamente creciente comunicación global (Fokiali, Xanthakou, Tatlidil, y Tatlidil, Kaila, 2006). Como es conocido, la experiencia turística además de las atracciones turísticas principales del destino depende de la infraestructura y facilidades con la que cuente el mismo (Sancho, 1998). Lo que permite el redescubrimiento de la herencia local, entendida como un conjunto de recursos cuya valencia económica muchas veces resulta poco explotada y tampoco reconocida.

Por otro lado, la pandemia de COVID-19, popularmente conocida la como pandemia de coronavirus o simplemente como el coronavirus, es una pandemia mundial y actualmente en curso derivada de la enfermedad ocasionada por el virus SARS-CoV-2.⁷⁸ Su primer caso fue identificado en diciembre de 2019 en la ciudad de Wuhan,⁹ capital de la provincia de Hubei, en la República Popular China, al reportarse casos de un grupo de personas enfermas con un tipo de neumonía desconocida. La mayoría de los afectados tenía vinculación con trabajadores del Mercado mayorista de mariscos de Huanan.¹⁰ La Organización Mundial de la Salud (OMS) la reconoció como una pandemia el 11 de marzo de 2020 (cuando informó que había 4291 muertos y 118 000 casos en 114 países).

Al 17 de septiembre de 2021, se ha informado de más de 228.4 millones de casos de la enfermedad en 258 países y territorios en el mundo, y 4 692 850 de fallecidos. Por otra parte, para octubre de 2020, la Organización Mundial de la Salud estimaba que al menos el 10 % de la población mundial ya se había contagiado de esta enfermedad (unos 780 millones de personas infectadas aproximadamente), debido al gran subregistro de casos a nivel mundial.

Metodología

La metodología que se utilizará será CRISP-DM (Cross-Industry Standard Process for Data Mining) la cual estructura el ciclo de vida de un proyecto de Data Mining en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo del proyecto.



La primera fase análisis del problema, incluye la comprensión de los objetivos y requerimientos del proyecto desde una perspectiva empresarial, con el fin de convertirlos en objetivos técnicos y en una planificación.

La segunda fase de análisis de datos comprende la recolección inicial de datos, en orden a que sea posible establecer un primer contacto con el problema, identificando la calidad de los datos y estableciendo las relaciones más evidentes que permitan establecer las primeras hipótesis.

Una vez realizado el análisis de datos, la metodología establece que se proceda a la preparación de los datos, de tal forma que puedan ser tratados por las técnicas de modelado. La preparación de datos incluye las tareas generales de selección de datos a los

que se va a aplicar la técnica de modelado (variables y muestras), limpieza de los datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato.

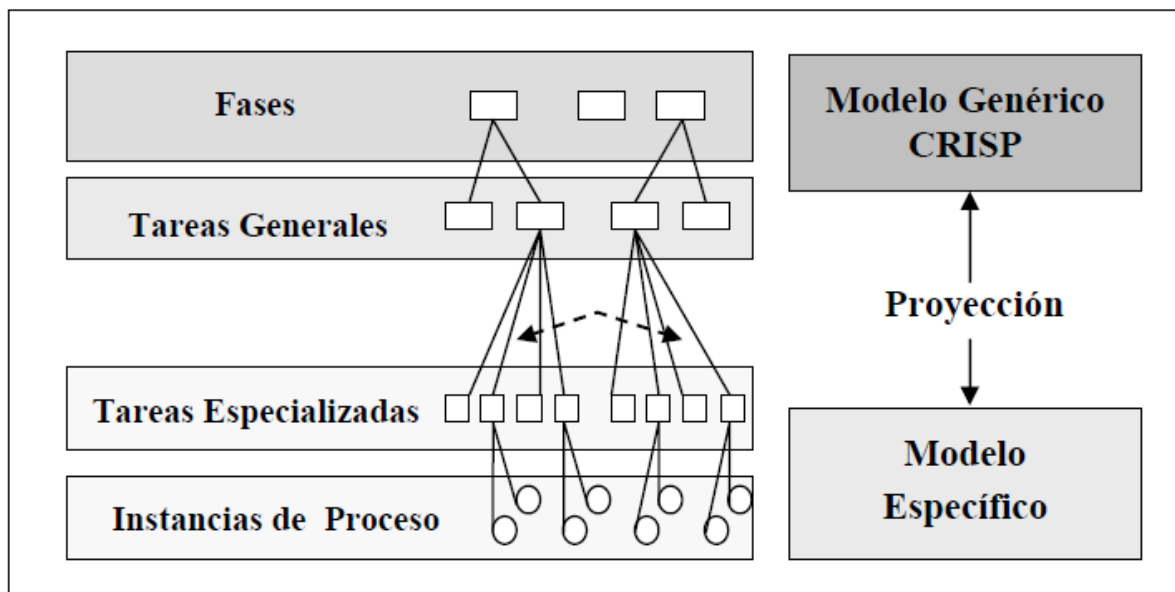
La fase de preparación de los datos, se encuentra muy relacionada con la fase de modelado, puesto que en función de la técnica de modelado que vaya a ser utilizada los datos necesitan ser procesados en diferentes formas. Por lo tanto las fases de preparación y modelado interactúan de forma sistemática.

En la fase de modelado se seleccionan las técnicas de modelado más apropiadas para el proyecto de Data Mining específico.

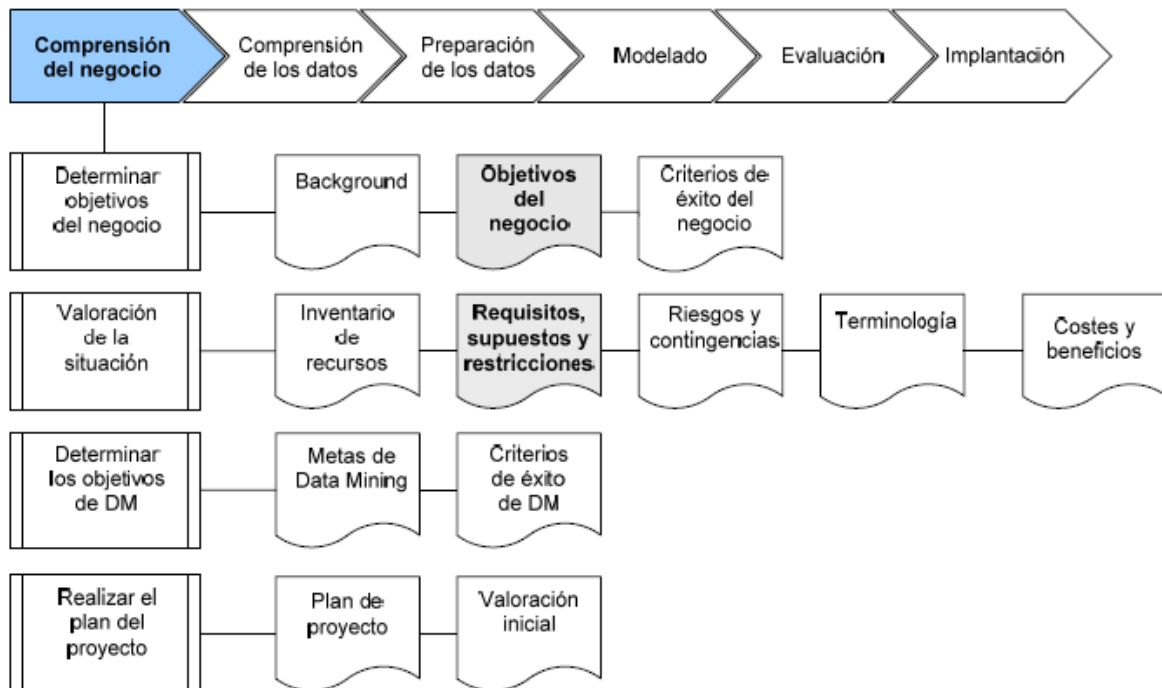
En la fase de evaluación, se evalúa el modelo, no desde el punto de vista de los datos, sino del cumplimiento de los criterios de éxito del problema. Se debe revisar el proceso seguido, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso en el que, a la vista del desarrollo posterior del proceso, se hayan podido cometer errores. Si el modelo generado es válido en función de los criterios de éxito establecidos en la primera fase, se procede a la explotación del modelo.

Primera fase: Análisis del problema

CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases como se muestra en la siguiente figura.



En la primera fase de análisis del problema también denominada análisis del negocio se pueden desarrollar las tareas que se muestran en la siguiente figura:



Para el caso específico de este proyecto a continuación se mencionan cada una de las tareas que se han desarrollado:

Objetivos del proyecto (negocio)

Mediante alguna técnica de análisis de datos, machine learning, o deep learning planteamos investigar aquellos sitios turísticos a los que los turistas podrán viajar con el menor riesgo de contagio y/o letalidad.

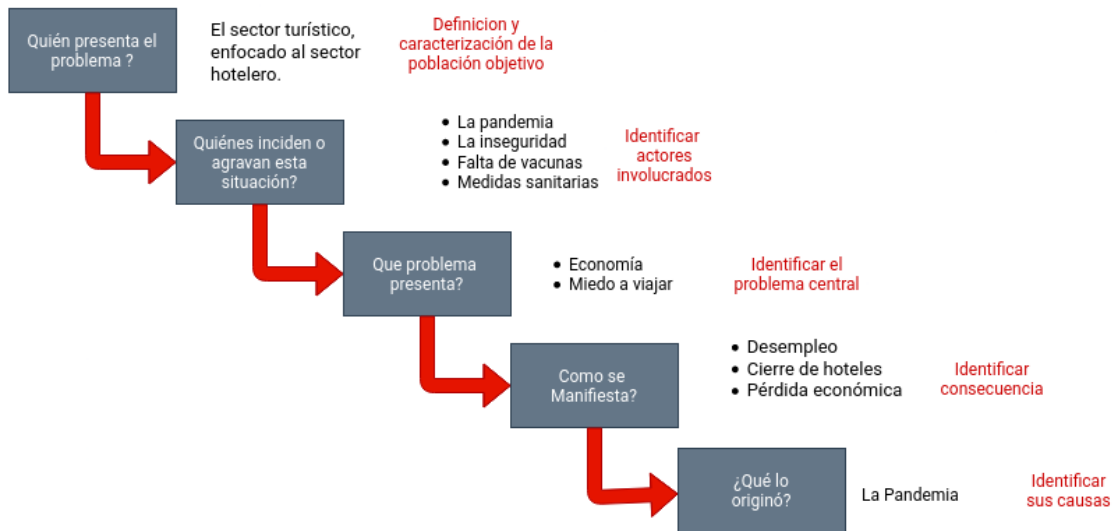
Valoración de la situación

Para realizar esta tarea se utilizó parte de la metodología del marco lógico, a continuación, se muestra el análisis del problema o situación.

Población objetivo:

- Población de Referencia: El sector turístico de México
- Población Afectada: Hoteles, Agencias, Operadores, Guías, Restaurantes, Aerolíneas, transporte.
- Población No afectada: De acuerdo a nuestro planteamiento todo el sector turístico se vio afectado en cierta manera.
- Población Objetivo: Turistas
- Población Postergada: Ciudades que no hacen parte o no son relevantes en el contexto turístico o tienen nivel bajo de arribo de turistas.

Pasos para el análisis de problema



Caracterización de la población:

Población Objetivo: Turistas.

Factor Socioeconómico:

- Pérdida de empleos.
- Disminución del PIB, (crisis económica / sanitaria).
- La pandemia afectó el estado económico de los turistas.

Factor Cultural:

- Delincuencia.
- Cierre de fronteras lo que disminuye el arribo de turistas.

Factor Individual:

- El estado mental de las personas debido a la pandemia.
- El miedo a ser contagiados se abstiene de viajar y están esperando que la situación mejore.
- Las personas están esperando su turno para ser vacunadas y esto influye que no pueden viajar todavía.

Factor familiar:

- Desempleo
- El fallecimiento de familiares por Covid-19
- Pasaporte Sanitario.



- Cronograma de Vacunaciones para el PEA (Población Económicamente Activa)

Factor Tecnológico:

- Falta de información de datos abiertos sobre vacunas.

Inventario de recursos

Los recursos que se utilizarán principalmente en el proyecto son:

- 5 computadoras personales (una por cada integrante del equipo)
- Repositorios de datos en la nube: github
- Libros electrónicos sobre minería de datos, inteligencia artificial, programación, estadística
- Fuentes de datos públicas de turismo y COVID
- Software de desarrollo de aplicaciones: R, Rstudio, Phyton
- Software de colaboración: Slack, Google docs
- Software de análisis de datos: Excel

Riesgos y contingencias

Los principales riesgos del proyecto son:

- Falta de información confiable
- Poco tiempo para el análisis de datos y modelos predictivos
- Poco tiempo para el desarrollo de la aplicación

Costos y Beneficios

Los principales costos del proyecto están dados por el tiempo que cada uno de los integrantes del equipo pueda aportar al mismo. El equipo del proyecto se estará reuniendo cada lunes y jueves de 8 a 10 de la noche y adicionalmente en la medida de las posibilidades al menos 2 integrantes del equipo asistirán a los talleres impartidos por Datalab.

Los principales beneficios del proyecto serán para el sector turístico de México y en específico para Hoteles, Agencias, Operadores, Guías, Restaurantes, Aerolíneas, transporte y turistas quienes tendrán una aplicación en internet donde podrán consultar las predicciones de nuevos casos de COVID por cada Estado de México y de los principales lugares turísticos de México y con ello podrán planear de mejor forma su estancia.

Determinar los objetivos técnicos de la Minería de Datos

Los principales objetivos técnicos son:

- Explorar y analizar los datos provistos por el gobierno de México en la página de [datos abiertos](#) de Covid-19 y el atlas turístico de México.

- Cargar y transformar los datos de Covid-19 y cruzar la información con los datos de turismo.
- Analizar los datos de mortalidad, contagio y esparcimiento de la enfermedad.
- Encontrar el mejor modelo para predecir el comportamiento de los siguientes días.

Criterios de Éxito de la Minería de Datos

Los criterios de éxito estarán dados por el menor error en las predicciones que se realicen, para ello se evaluarán diferentes modelos predictivos y se evaluarán principalmente los siguientes indicadores en cada modelo:

RMSE: Root Mean Squared Error (Raíz del error cuadrático medio)

Es una medida de uso frecuente de las diferencias entre los valores (valores de muestra o de población) predichos por un modelo o un estimador y los valores observados.

La RMSE es siempre no negativa, y un valor de 0 (casi nunca alcanzado en la práctica) indicaría un ajuste perfecto a los datos. En general, una RMSE más baja es mejor que una más alta.

MAE: Mean Absolute Error (Error absoluto medio)

Es una medida de la diferencia entre dos variables continuas. Considerando dos series de datos (unos calculados y otros observados) relativos a un mismo fenómeno, el error absoluto medio sirve para cuantificar la precisión de una técnica de predicción comparando por ejemplo los valores predichos frente a los observados.

Realizar Plan de Proyecto

El plan del proyecto está determinado principalmente por el tiempo definido en el Social Data Challenge el cual tiene un periodo de aproximadamente 3 meses de Agosto a Octubre como se muestra en la siguiente figura:

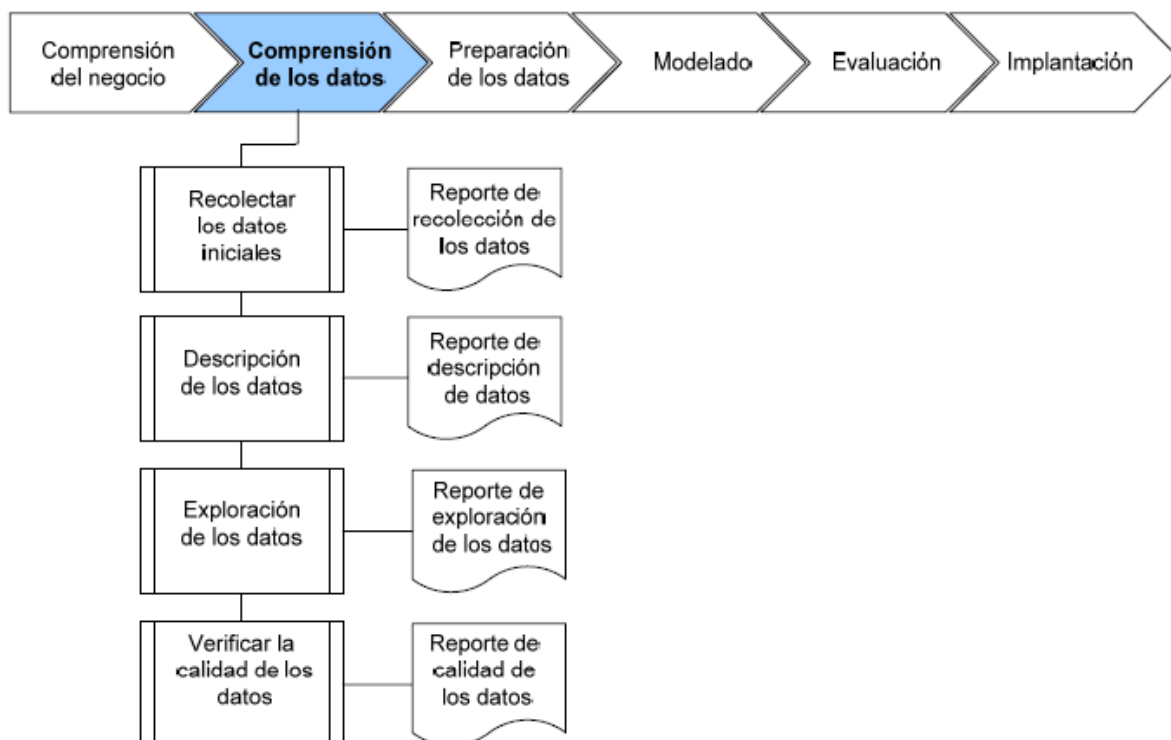


En estos 3 meses se estarán desarrollando las 6 fases de la metodología CRISP-DM divididas aproximadamente de la siguiente forma:

Fase	Periodo
Análisis del problema	9 Agosto - 23 Agosto
Análisis de los datos	16 Agosto - 27 Agosto
Preparación de los datos	30 Agosto - 24 Septiembre
Modelado	6 Septiembre - 1 Octubre
Evaluación	1 Octubre - 13 Octubre
Implantación	13 Octubre - 29 Octubre

Segunda fase: Análisis de los datos

En la segunda fase de análisis de los datos también denominada comprensión de los datos se pueden desarrollar las tareas que se muestran en la siguiente figura:



Para el caso específico de este proyecto a continuación se mencionan cada una de las tareas que se han desarrollado:

Recolectar los datos iniciales

En el apartado de Bases de Datos es donde se definen las fuentes de datos que se utilizaran en el proyecto.

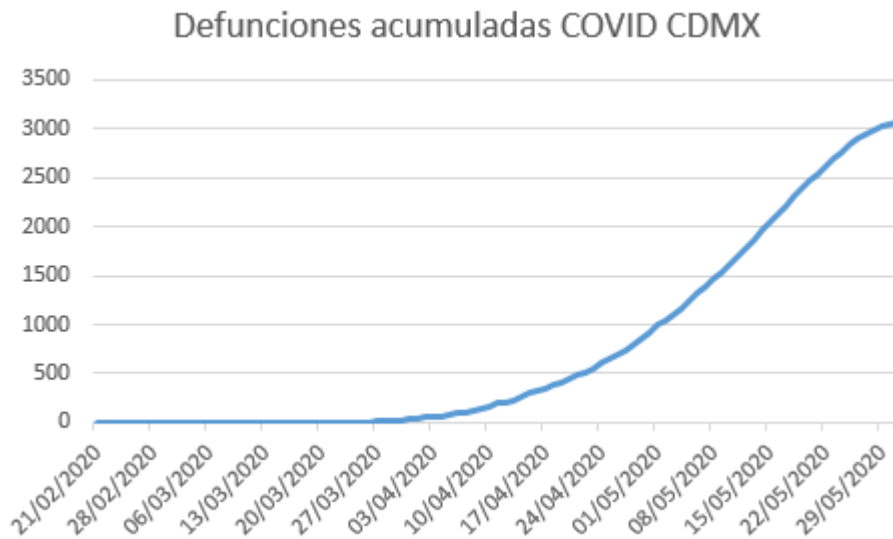
Descripción de los datos

Igualmente, en el apartado de Bases de Datos es donde se definen las fuentes de datos que se utilizaran en el proyecto.

Exploración de los datos

Los datos obtenidos del COVID-19 fueron analizados inicialmente en Excel ya que provienen de archivos CSV (valores separados por comas). Excel solo permite cargar aproximadamente 1,048,076 filas o registros por lo que no pudieron cargarse la totalidad de los registros y se cargaron únicamente hasta el mes de mayo de 2020 para hacer un análisis preliminar como se muestra en la siguiente gráfica filtrando registros de la CDMX.





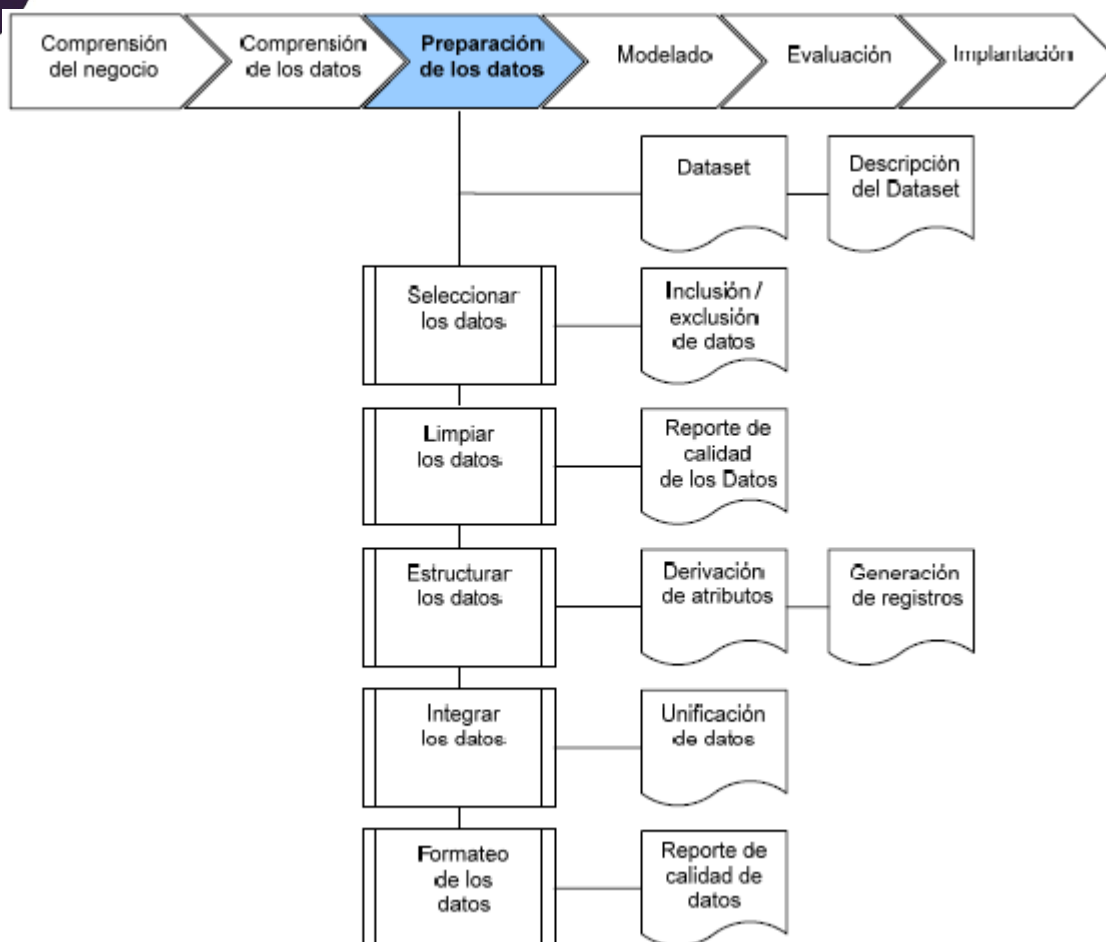
Verificar la calidad de los datos

Se realizó la calidad de los datos verificando que cumplan con algunas de las características siguientes: exactitud, completitud, integridad, actualización, coherencia, relevancia, accesibilidad.

No se encontraron datos nulos o fechas no exactas para los datos.

Tercera fase: Preparación de los datos

En la tercera fase de preparación de los datos se pueden desarrollar las tareas que se muestran en la siguiente figura:



Diccionario de Datos (Descripción del Dataset)

El archivo de casos diarios a nivel nacional que proporciona el sitio: <https://datos.covid-19.conacyt.mx/#DownZCSV> es un archivo CSV que contiene una fila por cada Estado de la República. El archivo puede descargarse para los casos confirmados o para los fallecidos. A continuación se muestra la pantalla del archivo abierto desde Excel.

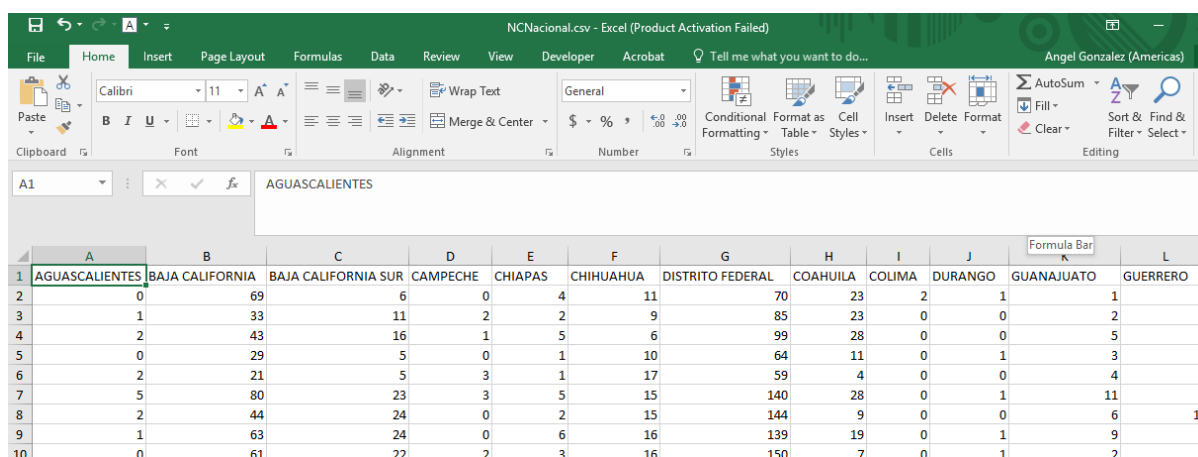
A	B	C	D	E	F	G	H	I	J
cve_ent	poblacion	nombre	18/02/2020	19/02/2020	20/02/2020	21/02/2020	22/02/2020	23/02/2020	24/02/2020
1	1434635	AGUASCALIENTES	0	0	0	0	0	0	0
2	3634868	BAJA CALIFORNIA	0	0	0	0	0	0	0
3	804708	BAJA CALIFORNIA SU	0	0	0	0	0	0	0
4	1000617	CAMPECHE	0	0	0	0	0	0	0
7	5730367	CHIAPAS	0	0	0	0	0	0	0
8	3801487	CHIHUAHUA	0	0	0	0	0	0	0
9	9018645	DISTRITO FEDERAL	0	0	0	0	0	0	1
5	3218720	COAHUILA	0	0	0	0	0	0	0
6	785153	COLIMA	0	0	0	0	0	0	0
10	1868996	DURANGO	0	0	0	0	0	0	0
11	6228175	GUANAJUATO	0	0	0	0	0	0	0
12	3657048	GUERRERO	0	0	0	0	0	0	0
13	3086414	HIDALGO	0	0	0	0	0	0	0
14	8409693	JALISCO	0	0	0	0	0	0	0
15	17427790	MEXICO	0	1	0	0	0	0	0
16	4825401	MICHOACAN	0	0	0	0	0	0	0

Las columnas que contiene el archivo son las siguientes:

Nombre de la columna	Descripción
cve_ent	Clave de la entidad (estado)
poblacion	Población (cantidad de habitantes en el estado)
nombre	Nombre del Estado
fecha (dd/mm/aaaa)	Cantidad de casos confirmados o fallecidos según el tipo de archivo descargado

Utilizando Excel se transformaron los datos para poderse cargar desde la plataforma Rstudio y poderlos utilizar en la fase de modelado. El archivo de fallecidos se le puso el nombre: FallecidosNacional.csv

A continuación se muestra un ejemplo con los datos de los casos confirmados por cada Estado de México por cada día.



	A	B	C	D	E	F	G	H	I	J	K	L
1	AGUASCALIENTES	BAJA CALIFORNIA	BAJA CALIFORNIA SUR	CAMPECHE	CHIAPAS	CHIHUAHUA	DISTRITO FEDERAL	COAHUILA	COLIMA	DURANGO	GUANAJUATO	GUERRERO
2	0	69	6	0	4	11	70	23	2	1	1	3
3	1	33	11	2	2	9	85	23	0	0	2	1
4	2	43	16	1	5	6	99	28	0	0	5	1
5	0	29	5	0	1	10	64	11	0	1	3	2
6	2	21	5	3	1	17	59	4	0	0	4	3
7	5	80	23	3	5	15	140	28	0	1	11	8
8	2	44	24	0	2	15	144	9	0	0	6	11
9	1	63	24	0	6	16	139	19	0	1	9	7
10	0	61	22	2	3	16	150	7	0	1	2	7

Análisis Preliminar de los datos

Para cargar los datos en Rstudio se utiliza la función:

```
datos2 <- read.csv(arch2, header=TRUE, sep=";")
```

Así quedaron cargados como tabla de datos en Rstudio:

	AGUASCALIENTES	BAJA.CALIFORNIA	BAJA.CALIFORNIA.SUR	CAMPECHE	CHIAPAS	CHIHUAHUA	DISTRITO.FEDERAL
134	5	6	7	5	7	11	
135	5	12	6	7	6	5	
136	6	23	9	9	1	7	
137	8	14	7	10	2	7	
138	6	18	10	3	2	6	
139	4	29	10	7	2	9	
140	5	22	12	4	3	4	
141	4	12	9	5	4	9	
142	5	11	8	1	3	9	
143	6	13	8	4	2	10	
144	3	16	4	4	1	9	
145	2	13	7	5	1	6	
146	16	17	8	2	0	12	
147	7	11	10	3	3	8	

Ya cargados en la plataforma Rstudio se puede analizar como por ejemplo usando la función summary:

```
> summary(datos)
```

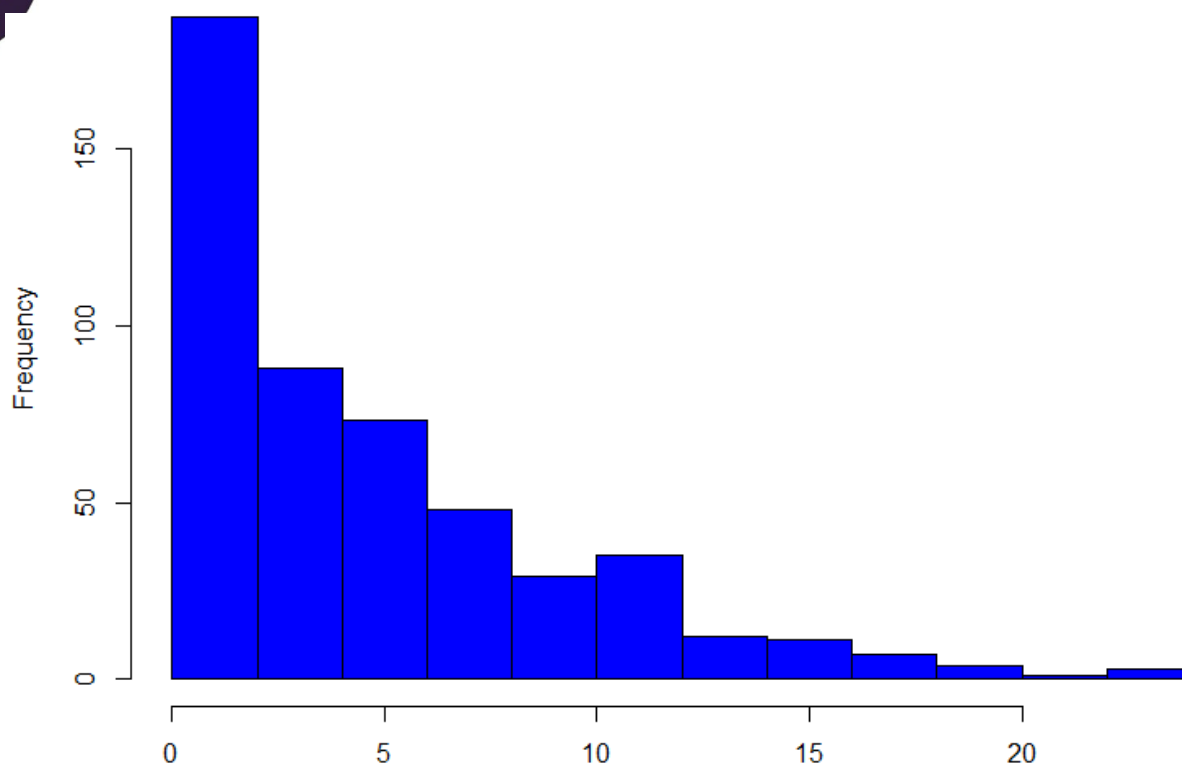
AGUASCALIENTES	BAJA.CALIFORNIA	BAJA.CALIFORNIA.SUR	CAMPECHE	CHIAPAS
Min. : 0.000	Min. : 0.00	Min. : 0.000	Min. : 0.000	Min. : 0.000
1st Qu.: 1.000	1st Qu.: 5.00	1st Qu.: 1.000	1st Qu.: 1.000	1st Qu.: 0.000
Median : 4.000	Median :13.00	Median : 3.000	Median : 2.000	Median : 2.000
Mean : 5.058	Mean :17.99	Mean : 4.195	Mean : 3.088	Mean : 3.715
3rd Qu.: 8.000	3rd Qu.:28.00	3rd Qu.: 6.000	3rd Qu.: 4.000	3rd Qu.: 4.000
Max. :24.000	Max. :67.00	Max. :20.000	Max. :21.000	Max. :30.000

CHIHUAHUA	DISTRITO.FEDERAL	COAHUILA	COLIMA	DURANGO
Min. : 0.00	Min. : 3.00	Min. : 0.00	Min. : 0.000	Min. : 0.000
1st Qu.: 7.00	1st Qu.: 36.25	1st Qu.: 2.00	1st Qu.: 0.250	1st Qu.: 1.000
Median :11.00	Median : 55.50	Median : 8.50	Median : 2.000	Median : 4.000
Mean :15.46	Mean : 73.07	Mean :13.27	Mean : 2.705	Mean : 5.209
3rd Qu.:17.00	3rd Qu.:107.50	3rd Qu.:22.00	3rd Qu.: 4.000	3rd Qu.: 8.000
Max. :99.00	Max. :264.00	Max. :58.00	Max. :12.000	Max. :24.000

GUANAJUATO	GUERRERO	HIDALGO	JALISCO	MEXICO
Min. : 0.00	Min. : 0.00	Min. : 0.00	Min. : 0.00	Min. : 1.0
1st Qu.: 3.00	1st Qu.: 4.00	1st Qu.: 5.00	1st Qu.: 8.00	1st Qu.: 38.0
Median :14.00	Median : 8.00	Median :11.00	Median :24.50	Median : 65.0
Mean :22.62	Mean :10.23	Mean :13.22	Mean :27.41	Mean : 78.6
3rd Qu.:34.00	3rd Qu.:15.00	3rd Qu.:17.00	3rd Qu.:35.00	3rd Qu.:109.0
Max. :122.00	Max. :32.00	Max. :53.00	Max. :128.00	Max. :252.0

Para estos mismos datos se puede generar un histograma:

```
hist(covid.ts, col="blue")
```



Y con la función `ts` se pueden convertir a una serie de tiempo:

```
covid.ts=ts(datos[1], start = c(2020, 4), freq = 365)
```

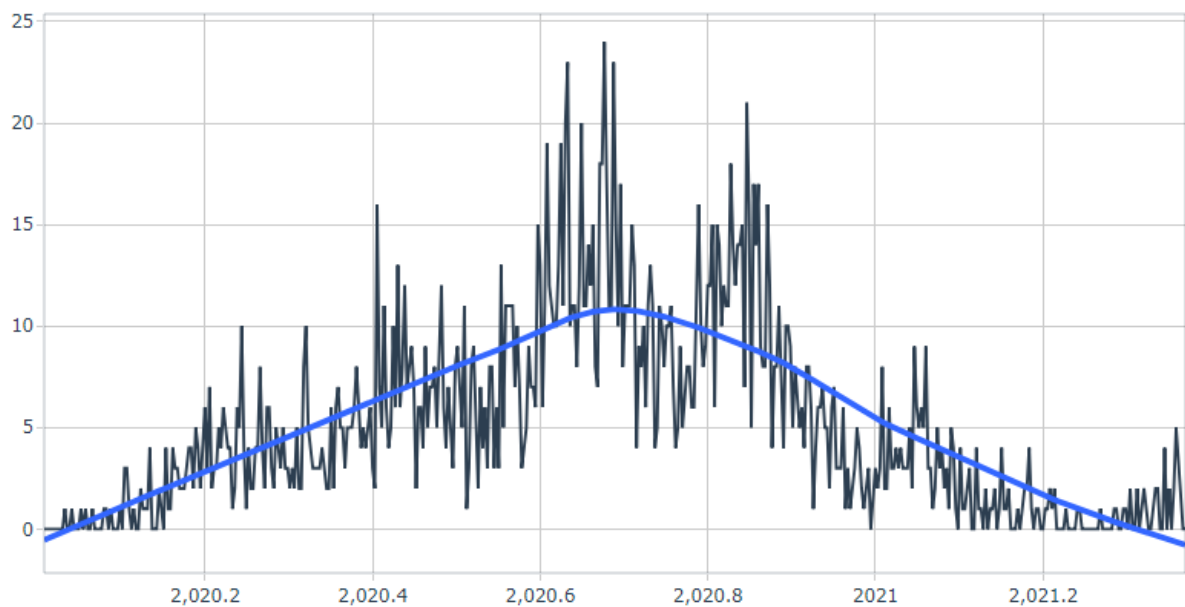
En este caso se hace una serie de tiempo para el Estado de Aguascalientes

Y con la librería “`timetk`” se pueden obtener la siguiente graficar:

```
timetk::tk_tbl(covid.ts) %>%
```

```
plot_time_series(index, AGUASCALIENTES)
```

Time Series Plot

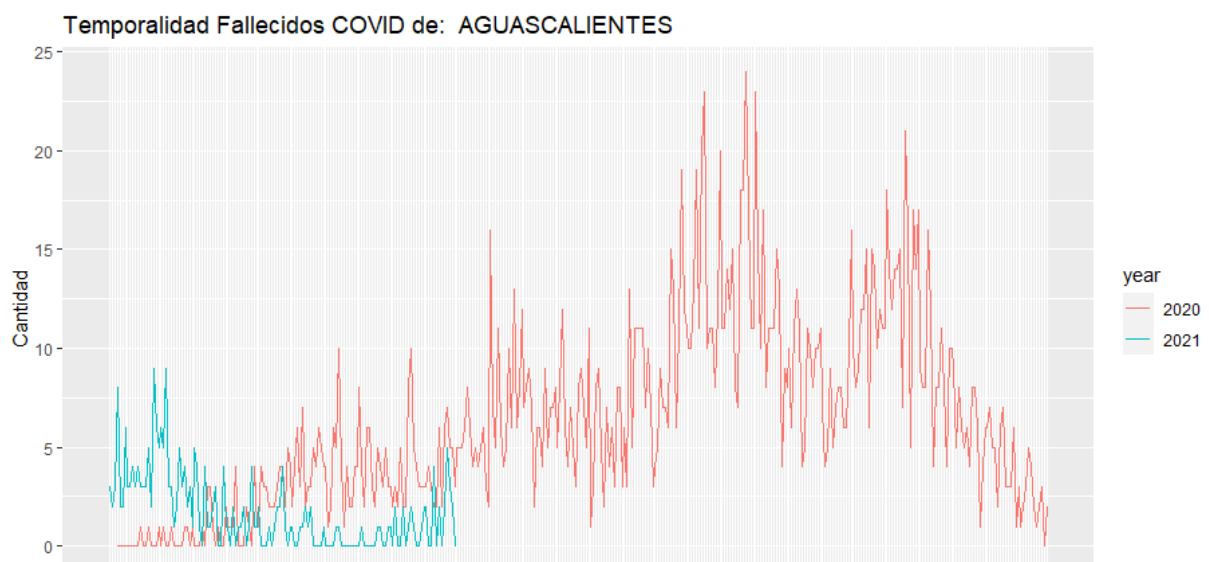


Y se puede generar una gráfica comparativa de casos de los años 2020 y 2021

```
nomedo = names(datos[1])
```

```
titgraf <- paste("Temporalidad Fallecidos COVID de: ", nomedo)
```

```
ggseasonplot(covid.ts, main = titgraf, xlab = "Fechas", ylab = "Cantidad")
```



Este mismo análisis se puede hacer por cada estado y por cada municipio y lugar del país.

Bases de Datos

Los datos que se utilizarán en el proyecto provienen de fuentes oficiales del gobierno de México entre las cuales se encuentran las siguientes que están disponibles en internet:

<https://datos.covid-19.conacyt.mx/#DownZCSV>

Esta fuente de datos contiene un resumen por fecha por Estado y Municipio de casos confirmados, fallecidos, negativos y sospechosos

<https://www.gob.mx/salud/documentos/datos-abiertos-152127>

Esta fuente de datos contiene el detalle de cada caso registrado, algunos de sus campos son la entidad y municipio, sexo, edad, fecha de ingreso al hospital, enfermedades preexistentes (diabetes, epoc, asma, hipertensión, etc.)

<https://datatur.sectur.gob.mx/SitePages/Inicio.aspx>

Esta fuente de datos contiene información estadística del Turismo como es: flujo de viajeros internacionales, turismo doméstico, flujos aéreos y marítimos, actividades de alojamiento, las cuales en su conjunto ofrecen una perspectiva de la dinámica del sector turismo en México.

<https://www.atlasturistico.sectur.gob.mx/AtlasTuristico/bienvenido.do>

Esta fuente de datos contiene los principales destinos turísticos y pueblos mágicos publicados por la Secretaría de Turismo de México.

Bibliografía

- Álvarez Cabal, J. Valeriano (2003), Metodologías para la Realización de proyectos de Data Mining, VII Congreso Internacional de Ingeniería de Proyectos, Pamplona
- Alocución de apertura del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de marzo de 2020». Organización Mundial de la Salud (OMS). 11 de marzo de 2020. Consultado el 27 de marzo de 2021.
- Balaguer, J y M. Cantavella-Jorda (2002). Tourism as a long-run of economic growth factor: the Spanish case. Applied Economics, 34(7), 877-884.
- Chaolin Huang; Yeming Wang; Xingwang Li; Lili Ren; Jianping Zhao; Yi Hu; Li Zhang; Guohui Fan; Jiuyang Xu; Xiaoying Gu (30 de enero de 2020). «Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China» [Características clínicas de los pacientes infectados con la nueva enfermedad de coronavirus de 2019 en Wuhan, China] (PDF). The Lancet (en inglés). p. 3. doi:10.1016/S0140-6736(20)30183-5. Archivado desde el original el 30 de marzo de 2020. Consultado el 2 de abril de 2020.
- Chapman, Pete (2000). CRISP-DM. Step by step data mining guide. SPSS Inc
- COVID-19 Virtual Press conference transcript - 12 October 2020. www.who.int (en inglés). Consultado el 27 de octubre de 2020.
- Fokiali, P., Y. Xanthakou, R. Tatlidil, E. Tatlidil, y M. Kaila (2006). "Tourism and sustainable development strategies in Rhodes: the awareness of the local societies". Ege Academic Review, Vol. 6, Issue 2: 25 - 46.
- Gonzales Diaz, M. (2021). Cómo México se convirtió en un oasis para turistas internacionales en medio de la pandemia (hasta para los que tienen que pasar cuarentenas) - BBC News Mundo. Retrieved 23 August 2021, from <https://www.bbc.com/mundo/noticias-america-latina-56035746>
- Lamboggia Ortiz, J. C. (2014). *Análisis del turismo y su importancia en el crecimiento económico en América Latina: el caso del Ecuador* (Master's thesis, Quito, Ecuador: Flacso Ecuador).
- Naranjo, S. (2021). La crisis del Turismo • Red Forbes • Forbes México. Retrieved 23 August 2021, from <https://www.forbes.com.mx/red-forbes-la-tesis-del-turismo/>
- Nuevo coronavirus - China». Organización Mundial de la Salud (OMS). 12 de enero de 2020. Consultado el 27 de marzo de 2021.
- Malthus, T. (1820). Principios de economía política, Fondo de Cultura Económica, Méjico.
- País, El (5 de octubre de 2020). «La OMS estima que el 10% de la población mundial se ha contagiado de covid, 22 veces más que los casos diagnosticado. EL PAÍS. Consultado el 27 de octubre de 2020.
- Ricardo D. (1817). Principios de Economía Política y Tributación, Fondo de Cultura Económica, Méjico.
- Sancho, A. (Direc.) (1998). Introducción al Turismo. Madrid: Organización Mundial de Turismo.
- Schumpeter, J. (1912). Teoría del desenvolvimiento económico, Fondo de la Cultura Económica, Méjico.



- Smith, A. (1776). Investigación sobre la naturaleza y causa de la riqueza de las naciones, Fondo de Cultura Económica, México