Informe Técnico: Retraso de vuelos

Daniel Perea Mercado

Estudiante de Ingeniería Industrial

Andrés Gómez Ardila

Docente

Curso: Introducción a la analítica de negocios

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

2024

# Introducción

El análisis de datos permite realizar un examen detallado de un asunto o problema para conocer su naturaleza, su estado, sus características y los factores que intervienen en ellos. También, permite responder rápidamente a tendencias del mercado ofreciendo la posibilidad de una mejor ventaja frente a los competidores. Para esto la empresa debe tener la capacidad organizacional de procesar data con el fin de encontrar conocimiento para lograr una mejor adaptabilidad al mercado.

El análisis de datos ayuda a las empresas a mejorar la visibilidad y el conocimiento más profundo de sus procesos, productos y servicios. Proporcionando información detallada de las ventas, pero también de posibles problemas que se puedan presentar en la operación en sus operaciones.

En esta oportunidad analizaremos una base de datos con el registro de los vuelos de algunas aerolíneas en su operación en Estados Unidos. Exploraremos los orígenes y destinos de los vuelos, la frecuencia de los vuelos en el tiempo, los retrasos y cancelaciones de vuelos y mucho más. Finalmente este informe ténico contiene código en Python para ilustrar como se hizo la analítica y como se encontraron los hallazgos.

# 

Acceso a los Datos

Se importaron librerías para la lectura y analítica de la base de datos

import pandas as pd

import numpy as np

Se cargó la base de datos a Google drive y se accedió al sito para su lectura definiéndola como df

from google.colab import drive

drive.mount("/content/drive")

archivo= "/content/drive/MyDrive/Delay\_2023\_12.csv"

df=pd.read\_csv(archivo, sep= ",", header=0)

df

Realizamos una copia de los datos por seguridad de la información original y con el fin de trabajar sobre la copia a la cual la llamamos delay

delay= df.copy()

delay

# Calculamos información de los datos, observaciones y variables

print(f"la base de datos tiene {delay.shape[0]} observaciones y {delay.shape[1]} variables")

la base de datos tiene 570394 observaciones y 110 variables

## Calculamos datos duplicados

print(f"la base de datos tiene {delay.duplicated().sum()} observaciones duplicadas")

La base de datos tiene 0 observaciones duplicadas

## Calculamos datos nulos

delay.isnull().sum().sum()

la base de datos tiene 30373485 observaciones nulas

También, se calcularon las columnas con datos nulos y los datos nulos por variable

delay.isnull().any()

delay.isnull().sum()

# 

## Entendiendo los datos

Encontramos la información de los datos: 570.394 registros, 110 columnas, 70 variables de tipo float64, 21 variables de tipo int64 y 19 tipo object

delay.info()

Se hace una análisis de las variables numéricas 8 registros 91 columnas; encontrando en la base de datos al año 2023 como único registro, al igual que el mes 12 que corresponde a Diciembre, el mínimo día del mes es 1 y el máximo es 31 de diciembre, esto da cuenta que la base de datos tiene registros del año 2023 entre el 1 y el 31 del mes de diciembre y que este periodo de tiempo corresponde a vuelos realizados.

delay.describe()

Identifico las columnas de la base de datos y selecciono con cuales variables puedo trabajar: 'Reporting\_Airline'

delay.columns

Identifico las aerolíneas que hacen parte de la base de datos, con la variable 'Reporting\_Airline', encontrando que son 15 en total y que solo se registró su sigla: ['OO', 'UA', 'B6', 'DL', 'F9', 'G4', 'HA', 'MQ', 'NK', 'OH', 'WN','YX', '9E', 'AA', 'AS']

delay['Reporting\_Airline'].unique()

## Origen y destino de los vuelos

Encuentro desde y hacia donde vuelan las aerolíneas con la columna 'OriginStateName' y 'DestStateName'

delay['OriginStateName'].value\_counts()

Top 5 de Estados de Estados unidos con mayor número de vuelos y su origen:

Estado Total vuelos

1. California 61792
2. Texas 60822
3. Florida 56986
4. Georgia 29881
5. New York 29364

delay['DestStateName'].value\_counts()

Top 5 de Estados de Estados unidos con mayor número de vuelos y su destino:

Estado Total vuelos

1. California 61806
2. Texas 60779
3. Florida 57050
4. Georgia 29904
5. New York 29343

# Aerolíneas en operación

Top 5 de las aerolíneas con mayor número de vuelos

delay.Reporting\_Airline.value\_counts()

# WN 127773 Southwest Airlines

DL 81022 Delta Airlines, Inc

AA 75408 American Airlines

UA 59844 United Airlines, Inc

OO 55996 SkyWest Airlines

# Análisis de los vuelos

## Conteo total de vuelos

delay['Reporting\_Airline'].value\_counts().sum()

Se contaron un total de 570394 vuelos

## Conteo de Vuelos

Se trabajo con la variables “DayofMonth”, “DayOfWeek”

Encontrando el total de vuelos por días del mes

delay.DayofMonth.value\_counts()

Top 5 de los días del mes con más vuelos

Día del mes Total de vuelos

1. 22 19840
2. 21 19752
3. 15 19558
4. 14 19484
5. 7 19432

Ahora, por día de la semana

delay.DayOfWeek.value\_counts()

Top 5 de los días de la semana con más vuelos

Día de la semana Total de vuelos

1. 5 (Viernes) 97191
2. 7 (Domingo) 89536
3. 6 (Sábado) 85997
4. 4 (Jueves) 77853
5. 3 (Miércoles) 73891
6. 1 (Lunes) 73316
7. 2 (Martes) 72610

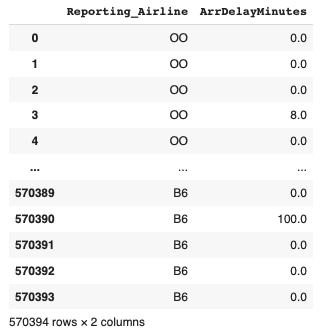
## Retrasos en los arribos de los vuelos

Se trabajo con la variables "ArrDelay", "ArrDelayMinutes",'Reporting\_Airline'

Encontrando los retrasos de los vuelos en los arribos por aerolíneas en minutos

delay[['Reporting\_Airline','ArrDelay']]

Ilustración 1-Retrasos arribos aerolíneas



## 

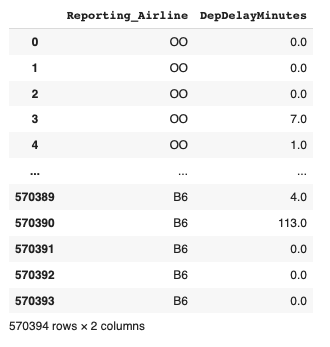
## Retrasos en las salidas de los vuelos

Trabajo con la variables "DepDelay", "DepDelayMinutes",'Reporting\_Airline'

Encontrando los retrasos de los vuelos en las salidas por aerolíneas en minutos

delay[['Reporting\_Airline','DepDelay']]

Ilustración 2-Retrasos salidas aerolíneas



Top 5 de lo peores retrasos en tiempo de los vuelos

delay["DepDelay"].sort\_values()

Tiempo

532815 -99.0

204073 -59.0

540096 -49.0

213487 -45.0

547895 -45.0

## Desvíos de los vuelos

Se encontraron 1.143 vuelos desviados

delay['Diverted'].value\_counts()

top 5 de aerolíneas con vuelos desviados

Aerolínea. Total desvíos

OO 1.0 258 SkyWest Airlines

# WN 1.0 178 Southwest Airlines

AA 1.0 124 American Airlines

DL 1.0 95 Delta Airlines, Inc

UA 1.0 85 United Airlines, Inc

Conteo de los vuelos desviados por días del mes

delay[['DayofMonth','Diverted']].value\_counts()

Top 5 de los días que más desvíos ocurrieron

Día Total desvíos

13 1.0 16

28 1.0 16

19 1.0 14

7 1.0 14

31 1.0 13

## Vuelos cancelados

Total de vuelos cancelados

delay['Cancelled'].value\_counts()

se encontraron 2310 vuelos cancelados

Conteo de vuelos cancelados según la aerolínea

delay[['Reporting\_Airline','Cancelled']].value\_counts()

Aerolínea. Total cancelados

WN 1.0 806 Southwest Airlines

OO 1.0 421 SkyWest Airlines

9E 1.0 156 Endeavor Air

YX 1.0 131 Republic Airlines

AS 1.0 128 Alaska Airlines

## Causas vuelos cancelados

Se encontró que por el código B se presentan las mayores cancelaciones de vuelos con 1392 seguido del código A con 789

delay.CancellationCode.value\_counts()

Código de cancelación Total de cancelaciones

B 1392

A 789

C 127

D 2

Conclusiones:

Luego de la analítica aplicada a la base de datos se encontraron importantes hallazgos: la base de datos tiene 570394 observaciones y 110 variables, tiene 0 observaciones duplicadas y 3037485 datos nulos. Así mismo, se encontraron 15 registros de aerolíneas, estas empresas vuelan en Estados unidos y la base de datos registro los vuelos hechos entre el 1 y 31 del mes de diciembre. También se pudo encontrar que los estados desde donde más volaron son California con 61792 vuelos y Texas con 60822 vuelos y sus destinos también son California con 61806 vuelos y Texas con 60779 vuelos. Por otra parte, las 5 aerolíneas que más vuelos registraron en el mes de Diciembre 2023 se halló que Southwest Airlines registró 127773 vuelos, Delta Airlines, Inc registró 81022 vuelos, American Airlines, Inc con 75408 y en el cuarto y quinto lugar aparecen United Airlines, Inc y SkyWest Airlines.

En cuanto, al análisis de los vuelos pudimos establecer que los días del mes donde más vuelos se procesaron fueron el día 22 con 19840 vuelos, el día 21 con 19752 seguidos de los días 15, 14 y 7 como los 5 primeros días que más registraron vuelos. Mientras que los días de la semana donde más volaron fueron Viernes con 97191 vuelos, domingo con 89536 vuelos y Sábado con 85997 vuelos.

Por otra parte, se analizaron los vuelos retrasados encontrando los tiempos máximos de retrasos en 99 y 59 minutos. Así mismo, Se encontraron 1.143 vuelos desviados, la aerolínea Skywest encabeza la lista con 258 registros, lo sigue Southwest con 178 desvíos y los días que se presentaron mayores desvíos fueron el 13, 19 y 28 de Diciembre. Ahora, los vuelos cancelados sumaron 2310, se registraron 806 en la aerolínea Southwest y Skywest con 421. Finalmente, pudimos hallar que los motivos por los cuales se presentan mayor número de vuelos cancelados; por el código de cancelación B se registraron 1392 cancelaciones que corresponde a condiciones meteorológicas adversas, por el A 789 cancelaciones que corresponde a razones atribuibles a la aerolínea y por el C 127 que corresponde a problemas de control de tráfico aéreo y el código D con 2 que corresponde a Razones de seguridad.