Communicate Data Findings

REVIEW | CODE REVIEW | HISTORY

readme.md

```
# <font color=blue>Data Expo 2008 «Airline on-time performance»</font>
## by [<font color=orange>Mostafa Abobakr</font>](https://cutt.ly/MostafaALinkedIn)


## Dataset

> This dataset consisting of __7,009,724 rows__ or data points after removing 4 duplicated points, reports

[Download the dataset from here](http://ww2.amstat.org/sections/graphics/datasets/DataExpo2009.zip) (1.6 GB

Features documentation:
- [2009 - Joint Statistical Computing and Statistical Graphics Section](https://community.amstat.org/joints
- [United States Department of Transportation - BUREAU OF TRANSPORTATION STATISTICS: Airline On-Time Perfor
- [The data. Data expo 09. ASA Statistics Computing and Graphics](https://web.archive.org/web/2019122014471
- [United States Department of Transportation - BUREAU OF TRANSPORTATION STATISTICS: Reporting Carrier On-T

Supplemental data:
- [Supplemental data. Data expo 09. ASA Statistics Computing and Graphics](https://web.archive.org/web/2019
- [airports.csv](https://web.archive.org/web/20191229040110/http://stat-computing.org/dataexpo/2009/airport
- [carriers.csv](https://web.archive.org/web/20191229040110/http://stat-computing.org/dataexpo/2009/carrier
- [plane-data.csv](https://web.archive.org/web/20191229040110/http://stat-computing.org/dataexpo/2009/plane

We could use the first two supplemental datasets, but we will not use the last plane-data.csv in our invest

You also can use this [Google drive link](https://drive.google.com/drive/folders/1qIX3ZNEzrd4bijppXIjsmn0yK

Other resources:
- [Data Expo 2009 - Airline on-time Performance Analysis | by Carlson Hoo | Medium](https://carlson-hoo.med
- [Airlines Delay | Kaggle](https://www.kaggle.com/giovamata/airlinedelaycauses)

## Summary of Findings

* **1**<sup>**st**</sup>: <u>**related to arrival delays and delay causes in general**</u>➢
    <br>🔧 'Carrier' as well 'Weather' were the most common to cancel a flight.
    <br>🔧 Flights-cancellation due 'Security' doesn't seem to be the common.
    <br>🔧 Most arrival delays were of 15 minutes to about 78 minutes.
    <br>🔧 More higher frequencies were for "Carrier" and "National Air System" lower-delays values than ot
    <br>🔧 'Carrier' delays exceeded with the highest spread and outliers.
    <br>🔧 'NAS' delays had more lower-values frequencies.
    <br>🔧 In general, Weather-delays had the highest mean, median, the wider IQR, and a wider range from i
    <br>🔧 Flights that had weather-delays were probably to have the most arrival delays in general.
    <br>🔧 Arrival delays due to 'Weather' had a higher mean, median (which is about 62 minutes), with a la
    <br>🔧 Much more frequencies for arrival delays due to 'NAS'.
    <br>🔧 For the whole flights of 15+ min. arrival delay, in general, arrival-delays due to 'Security' ha
    <br>🔧 For all delayed or flights of 15+ arrival delay, there were some kind of linearity between arriv
    <br>🔧 In general, arrival delays had more stronger moderate correlation with carrier-delays.
    <br>🔧 'Carrier' and 'NAS'-delays had more higher correlation than 'Weather'-delays with arrival delays
<br>

* **2**<sup>**nd**</sup>: <u>**related to 'Months'**</u>➢
    <br>🔧 All 2008 months had close-ratios of recorded total flights frequencies.
    <br>🔧 Months 'Feb', 'Dec', 'Jan', and 'Mar', had the highest flights-cancellation counts.
    <br>🔧 'May', 'Nov', and 'Oct' had the lowest flights-cancellation counts.
    <br>🔧 'Feb' exceeded in the ratio of cancelled flights, then 'Dec' and 'Jan',  which are notable to be
    <br>🔧 "Weather" had the highest impact to cancel flights within months 'Feb', 'Dec', 'Jan', 'Mar', as
    <br>🔧 Months 'Dec', 'Jun', 'Feb', and 'Mar' had the highest arrival delays means.
    <br>🔧 'Sep' and 'Nov' had the lowest arrival delays means.
    <br>🔧 'Weather' had the highest impact on flights arrival delays within different 2008 months.
<br>

* **3**<sup>**rd**</sup> <u>**related to 'Carriers'**</u>➢
    <br>🔧 'Southwest Airlines Co.' recorded the highest total flights count within 2008; more than twice o
    <br>🔧 Carriers 'American Eagle Airlines Inc.', 'American Airlines Inc.', 'Skywest Airlines Inc.', 'Sou
    <br>🔧 'Hawaiian Airlines Inc.', 'Frontier Airlines Inc.', and 'Aloha Airlines Inc.' had the lowest fli
    <br>🔧 "Carrier" procedures had the highest impact to cancel flights within 11 Airlines carriers of 20.
    <br>🔧 "Weather" had the highest impact to cancel flights within 9 Airlines carriers of 20.
    <br>🔧 "NAS" or National Air System had the second highest impact to cancel flights within 5 Airlines c
    <br>🔧 'American Airlines Inc.'  had more cancellations due to 'Carrier'.
    <br>🔧 'Hawaiian Airlines Inc.'s flights-cancellation were because of 'Carrier', as a majority.
    <br>🔧 'Aloha Airlines Inc.'s 42 cancelled-flights were because of 'Carrier'.
    <br>🔧 Carriers 'American Airlines Inc.', 'Mesa Airlines Inc.', 'Comair Inc.', 'United Air Lines Inc.',
    <br>🔧 'Pinnacle Airlines Inc.', 'US Airways Inc.', and 'Hawaiian Airlines Inc.', had the lowest arriva
    <br>🔧 Arrival delays mean of 'Aloha Airlines Inc.' was about negative 3 minutes, indicating almost no
    <br>🔧 On average, 'Weather' had the highest impact on flights arrival delays within almost all carrier
    <br>🔧 Arrival delays due to 'Security' may had greater medians and more greater-values distributions,

#### Interpretations and conclusions not included in Explanatory presentation, from Exploratory analysis➢

- 'Carrier' and 'NAS' had the largest counts of lower-values delays for different months, though 'NAS' exce

- 'Weather'-delays had the highest medians, and the largest third quartile ranges all over months. Then car

- 'Weather'-delays were the highest on average within different 2008 months, notably in 'Jul' and 'Sep'. Th

- Greater-values distributions due to 'Weather'-delays were the largest all over months.

- 'Carrier', as well 'NAS'-delays, had the largest counts of lower-values for different carriers, especiall

- There weren't any 'Security'-delays in 'AirTran Airways Corporation'. Unlike in 'Frontier Airlines Inc.',

- Weather-delays distributions were the highest at all within 'Mesa Airlines Inc.', 'Expressjet Airlines In

- 'Weather'-delays were the highest on average within almost the majority of carriers. Then also, in genera

- Unlike usual regarding different-causes delays within all carriers, 'Northwest Airlines Inc.' had the hig

## Key Insights for Presentation

I used the dataset to gain insights that could help make improvements against the flights delaying's, or to

> After finishing some work of columns-structuring using SQL, I extracted the columns I thought as helpful

> __From 'df_inv', I derived:__<br>
__1__ _**'airline_cancelled'**_ dataframe for data points of cancelled flights data<br>
__2__ _**'on_time'**_ *dataframe for flights with less than 15 min. arrival delay and not null, or data poin*
__3__ _**'df_inv_15'**_ *for flights with arrival delays that equal 15 minutes or more, then i* ***sampled 'df
```

Rate this review

START

RETURN TO PATH