Business Reporting Tools

Professor: Matthijs Meire

Group Assignment
NYC Flights Data

Fernando Delgado
Dil’da Zhaksybek
Juan José Medina

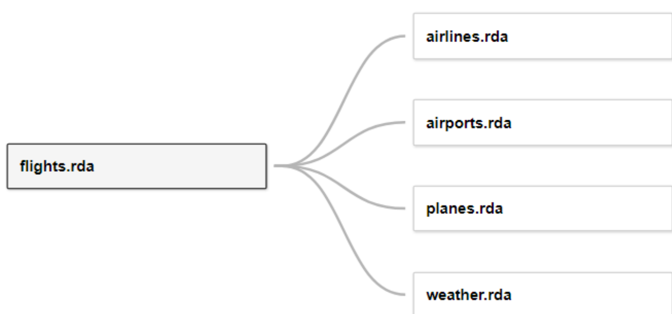
October 17, 2021

Introduction

The following report intends to explain insights on flight delays departing from New York airports, JFK, EWR, and LGA throughout the year of 2013. The analysis takes into account both internal and external factors such as, weather, type of aircraft, seasonality, origin airport, and destination airport.

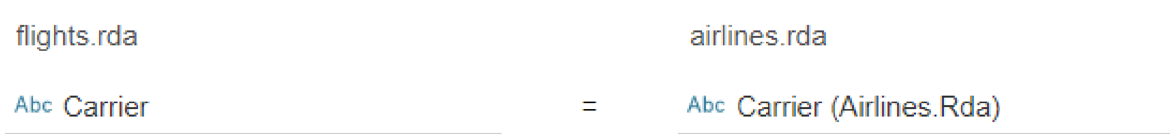
Annexed to this file, you can observe a Tableau file with a visual representation of the insights that we believe are the most valuable. It is important to mention that such file is built upon a 2012 New York flights database which is composed of 5 main tables:

- Flights
- Airlines
- Airports
- Planes



We are using *flights* as the main table because it enables us to create links with the rest of the data using their respective variables or keys.

To link *flights* with *airline* we use the variable of *carrier:*



To link *flights* with *airports* we use the variable of *destinations* from *flights* and *FAA* from *airports:*

flights.rda		airports.rda
Abc Dest	=	Abc Faa

To link *flights* with *planes* we use the variable *tailnum*, which is the unique code usually assigned for every airplane in order to identify them:

flights.rda		planes.rda
Abc Tailnum	=	Abc Tailnum (Planes.Rda)

To link *flights* with *weather* we are using 2 different variables. First, we use *origin* to link it to the origin airport. Then, *time hour* to know the weather during a certain period of time:

flights.rda		weather.rda
Time Hour	=	Time Hour (Weather.Rda)
Abc Origin	=	Abc Origin (Weather.Rda)

Insights

After creating the required links between the data, we were able to create graphs and dashboards that could show valuable insights:

Map-Avg Delay Dest

First, with this graph we can geographically observe the average arrival delay time taken from the by flight, considering the location of the destination airports. The average delay is shown with both size and color to help us visualize the difference from each destination airport. It is also possible to filter the Origin Airport as well as the respective month of operation. This shows a visual representation of the busiest destinations.

Flights by Airport

Then, we wanted to show the number of flights per airport filtered by month. To do so, we used the sum of the number of flights per origin airport to determine how the monthly variation.

Origin and Delay

Moreover, this bar graph displays the average delay time on departure and compares them to their origin. By using the departure time instead of the arrival time, we can assume that it is the origin airport the main reason that causes the delay rather than problem from the airplane.

Day of the Week Deps

This graph shows the number of flights that have different types of delays such as Delayed, Left Early or on Time, while analyzing how it evolves during each day of the week, in order to determine if any specific day have more flights that have departure delays.

Day of the Week Arr

Subsequently, this graph is similar to the previous one. It shows the number of flights that have different types of delays such as Delayed, Left Early or On Time, while analyzing the daily evolution, in order to determine if any specific day has more flights that have arrival delays.

Hour of the day

Further, with this bar graph we are able to determine if the hour of the day affects the departure delay, while considering the number of flights that happened during that particular hour and observing the evolution per day of the week. All the information is taken from the *flights* data source.

Model and Delay

Then, we wanted to determine if external factors that are not fault of the origin airport may influence on the arrival delays. This graph allows us to visualize if the model of the airport taken from the *planes* data source, affects the arrival delay.

Delay by Engine

Next, the following graph allows us to determine if the type of engine (lanes) has any effect on the arrival delay (flights). We use the arrival delay instead of the departure delay, because its more likely that a factor from the plane may result in a delay during the travel. We also consider the distance and speed of planes to see, which engines are used on long travels and if the engine affects the speed of the plane.

Visibility

Afterwards, the following area chart is a visual representation of how weather is affecting delays, particularly the Visibility. On the departure delay, the variable of visibility is measured in miles that the pilot is allowed to see in front of them without their vision obstructed. We can see that the origin airport is categorized into 3 colors, to aid us to determine which airport has the biggest delays.

Wind Speed

Similarly, this chart helps us visualize the effect from wind speed conditions on the departure delay. We can observe that the origin airport is categorized into 3 colors, for an easier understanding of which airport has the biggest delays related to the windspeed during that flight.

Worst Route

Furthermore, by using the Total Delay, which represents the sum of departure and arrival delay, we can observe the worst possible route, considering the origin and destination airports. Additionally, by applying a filter by origin airport we can observe the data by origin.

Avg Dep Delay by number of Flights

Moreover, this graph is a visual representation of the departure delays time in relation with the count of flights. We can visualize how most of the flights end in a group that is determined by the departure delay by using the count of flights as rows and size.

Month and Arrival Delay Correlation

Finally, we are able to determine if the month of the year has any impact on the departure delay, to see if busy months affect the departure times on the origin airports. Additionally, the flight count displays which months were busy or not.

Conclusion

To conclude, analysis of the data shows the following findings.

Newark airport, being the busiest of all three origin locations, has an average departure delay of 14.9 minutes, comparing to comparing to 6.62 in JFK and 5.64 in LGA.

While there is no clear correlation between average delay time and day of the week, on average, the highest departure delay happens during the time from 19 pm to 21 pm.

Another factor that is affecting delay time is the plane model. Average arrival delay is the highest for the Model 747-451, 120 minutes. Quite obviously, the delay is the shortest for those planes that have the fastest engine, which is turbo-jet. For those engines, arrival delays averaged to only 3.19 minutes.

Another observation that we consider to be an important departure delay factor is visibility. The higher the visibility, the shorter is the departure delay. Wind speed is affecting departure delay significantly starting from 33 miles per hour.

There is clearly a spike in arrival delays during the months of June and July, as well as December. Although the number of flights is almost evenly distributed throughout the year.

Analysis also shows that some of the destinations account for a longest delay than others. With the longest being Columbia Metropolitan Airport (CAE).

Overall, there is no single factor that contributes the most to the delays. It is rather a combination of several factors, such as weather conditions, plane and engine type, destination and seasonality.

The link to the tableau report:

https://public.tableau.com/views/GroupAssignmentTABLEAUFinal/Story1?:language=es-ES&publish=yes&:display_count=n&:origin=viz_share_link