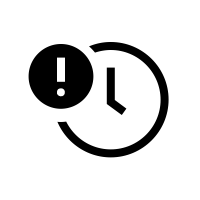
FLIGHT DELAY DATA INFORMATION

USING TABLEAU





Group – 5

Done by

Yu-Ting Chang

Sri Sai Vivek Cherukuri

Lingyu Ma

CONTENTS

Executive Summary ……………………………………………………………………………… 2

Design and Findings ……………………………………………………………………………… 4

Conclusion ……………………………………………………………………………….11

EXECUTIVE SUMMARY

This dataset is obtained from the RITA website which contains information about flight delays and performance. The dataset used ranges from 2012-2017 from all airports and airlines within the Unites States. It has 22 variables and around 70000 rows.

We created the visualizations in Tableau that answer the following questions:

1. What are the percentages of different types of flight delays or cancelled/diverted arrivals and how they are distributed over months, carrier types and airports?
2. What are the carriers that cause most delays when all the airports are taken into account and what are the airports that contribute for these delays the most for different carriers?
3. What are the airports that cause the most delays when all the carriers are taken into account and also what are the airports that cause most delays for a particular carrier?
4. What are the top 10 airports and carriers that are responsible for the delays and when combined together how do they affect each other?

DESIGN AND FINDINGS

The initial design choice for the first point of the story was to break down different delay types to see a comparison between them and how much each was contributing towards the total delay. Also, parallelly we wanted to break down even further by months to see if the delays had a high and low point during a particular time of the year.

For this reason, we created a derived column called 'All Delays' where we took every delay and added them up. And then in this first plot, we placed total delay also to make a point of reference. Here we can also filter by airport, carrier and year to drill down even further.

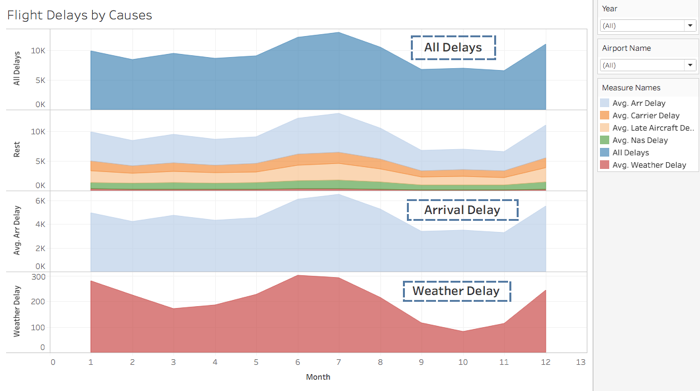


Fig.1

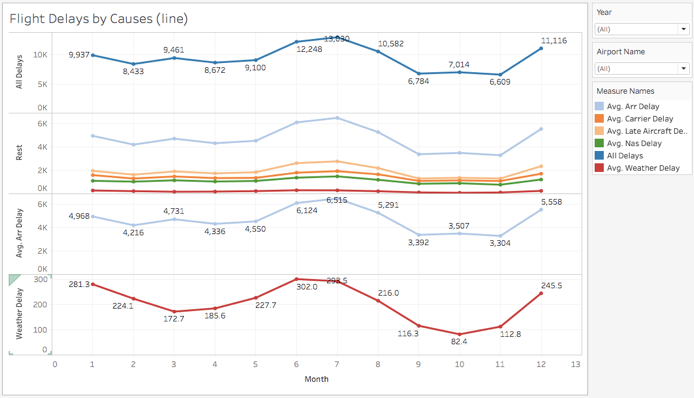


Fig.2

Our second design consideration was if there's any way to make a geolocation plot for all the airports and then map the delays via a heatmap to that. But in the original csv file, there was no field for the countries, cities or states. But we saw there was an airport name field which had all the abbreviations of the airports. So, we changed the field type from string to geolocation and chose the option 'Airport' which actually made me a field which can be plotted in maps. The size and also the darkness of the points showed a redundant plot of how much delay was happening in a particular airport. The more the delay, the bigger and darker the point is.

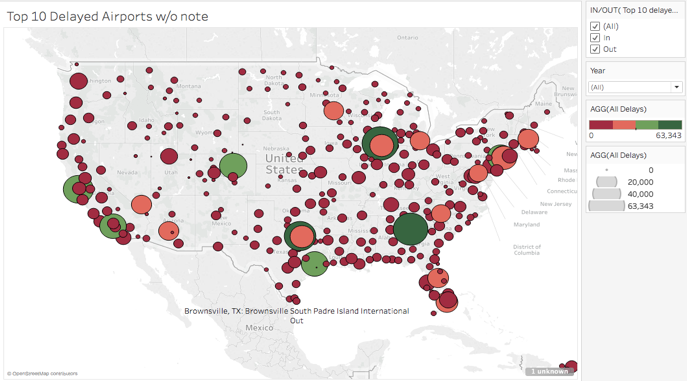


Fig.3

Now our next target was to see if there's a way to find out which were the carriers that got delayed the most when all the airports and all the years were taken into account. Also, in the same pair of plots, we wanted to see that for a particular carrier, what are all the airports that contributes the most for its delay. So, we made a heatmap for the carriers and mapped it to the above geolocation plot so that we can see for a particular year, which were the carriers that got delayed the most and what were the airports that contributed most for the delay. We also created an action here such that if only one particular flight was clicked in the heatmap, the corresponding airports from which it got delayed were shown on the map.



Fig.4

When a carrier is clicked:



Fig.5

Similarly, we wanted to create another plot where we can see which airports contributed for the most delays when all the carriers were taken into account or any one of the carriers is considered. We created a similar heatmap-geoplot mapping for this. When a particular carrier is selected from the dropdown, it shows all the airports that contributed the most for its delay. In this plot if you hover over an airport on the heatmap, you see its exact geolocation in the map.



Fig.6

For Spirit Airlines:



Fig.7

The last thing we wanted to do was to find out the top 10 airports and carriers which caused the most delay and we wanted to see if these two sets are combined together, was there any new observations? So, we created two sets:

1. Top 10 Delayed Airports
2. Top 10 Delayed Flights

These two sets only contained the top 10 airports and carriers respectively. So, by using these two sets, again created a similar plot. First, plotted all the top 10 airports in the geoplot. Then plotted the top 10 carriers in a customizable heatmap. Whenever one of the top 10 airports is chosen, it will show us the carriers (which are also in the top 10) that got most delayed from that particular airport.

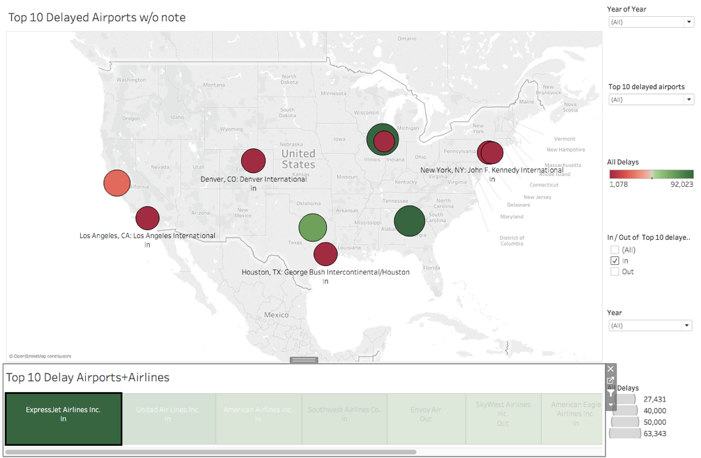


Fig.8

CONCLUSION

We see that Southwest moves back to the 4th position and now ExpressJet Airlines take the top position of most delayed airlines. This means the most delay of Southwest is not from the most delayed airports but rather spread over all the airports. When we click Southwest airlines in the plot, we see among the top 10 airports, the top 3 airports causing the delay are Chicago, Denver and Los Angeles.

Top 3 Delayed Airlines:

1.ExpressJet Airlines

2.United Airlines

3.American Airlines

Top 3 Delayed Airports:

1.Chicago O’Hare International Airport

2.Hartsfield Jackson International Airport Atlanta

3.Dallas Fort Worth. International Airport