US Air Traffic Data Analysis

by Soner Nefsiogullari

Summary

The dataset is taken from Bureau of Transportation Statistics webpage and contains information about flight delays in US. I preferred to use data of the years between 2013 and 2017 for my visually analysis. The link of the dataset can be seen from Resources part of this report.

Design

At the beginning, I made some changes when Tableau reading the raw data. I changed the airport abbreviations in to string and then select Airport from the geographic role list to make Tableau understand the longitude and latitude.

First, I plotted the arrival delays; however, I realized that there are other types of delays and decided to sum them up to find total delay. I also changed the names of attributes for nice looking.

For initial dashboards I used bar chart since it provides a great overview of the distribution between the different types of delays. I also plotted delay types with respect to carriers, total delays by carriers and delay trend by months, I prefer line plot since it helps us to see the trends. Moreover, I chose to represent airports in a geographical map as a map makes easy to understand locations.

The initial graphs and dashboards can be seen from link below:

https://public.tableau.com/profile/soner.nefsiogullari#!/vizhome/project_initial/Initial_Dashboard1

After collecting the feedback, I made changes in bar graphs and make them readable with single color. I decided to use colors if only details needed. I watched the lecture video again and learned to make graphs “use as filter”. Moreover, I cleaned some filters that shown twice in the dashboard. I also put the filters and legends in appropriate corner of the graphs and changed all the filters to “single value (dropdown)”

The final story can be seen from link below:

https://public.tableau.com/profile/soner.nefsiogullari#!/vizhome/FlightDelaysinUS/FlightDelaysinUS

There are various reasons of Flight Delays in US. When we group them to analyze the major groups will be: Arrival Delay, Late Aircraft Delay, Carrier Delay, Nas Delay, Weather Delay and Security Delay.

When we looked deep into delay types by month, we see that there is an obvious seasonality in Arrival Delays, Late Aircraft Delays and Carrier Delays. Total delays peaks in between June and July and also in December. We see that arrival delays has the largest percentage and the security check delay has the least percentage in total delay. We also see that canceled flights have peak at February and diverted fights have peak in between May-July period. This might be due to weather conditions or holiday seasons.

Chicago O'Hare International Airport has the highest total delay in between 2012 and 2017. Moreover, we interpret from the graphical analysis that Southwest Airlines Co. has the highest total delay. Also it is clear that large amount of the total delays are due to Arrival Delay.

I used geographic map to help us to analyze delays by size and color. When we click on Chicago O'Here Airport in the map, we see that in fact Southwest has no flights in that airport, instead they use Chicago Midway International Airport. In addition, they have the largest delay percentage compared to other carriers using same airport.

Moreover, we may see the type of delays when click on a significant airport. To illustrate, when we click on Atlanta (ATL) airport, we see that Delta Air Lines has the highest delay mostly due to arrival delay and carrier delay.

Finally, if we want to investigate top 5 carriers with regard to total delays we see that Southwest, American Airlines, United Air Lines, JetBlue Airways and Spirit Air Lines are in the top 5 respectively. If you are in New York (JFK), the riskiest carrier will be JetBlue. In Dallas (DAL) actually delays comes from only Southwest Airlines. American Airlines also have largest delay on Dallas but in different airport called Dallas Fort Worth International (DFW). In addition, United Air Lines has major of their delays in Chicago and San Francisco. To conclude, it is better to be aware and make alternative plans in case of delays in those airports and carrier combinations.

Feedback

The first feedback was about dashboard1: “too much information per graph, and it is too colored, I suggest you to make it simple and understandable with small effort. I couldn’t interpret results from delay circle graph.”

Second feedback came for dashboard2: “very good choice for visualization, but when I click on the bar of particular carrier name, I would like to see the map changes at the below”

Finally for dashboard3: “there are 2 filters for carrier name, I get confused”

Resources

Raw Data: https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?pn=1

Data Discription: https://www.transtats.bts.gov/Fields.asp?Table_ID=236

http://www.vizwiz.com/2016/07/tableau-tip-tuesday-layout-tips-for.html

https://www.thedataschool.co.uk/rachel-costa/customised-map-tableau/