Write_up_tableau

February 19, 2018

0.1 Flight data visualization using Tableau

0.1.1 Author : Mithun Muralidhar

0.1.2 Date: February 18, 2018

0.1.3 My Tableau workbook link :

https://public.tableau.com/profile/mithun.muralidhar#!/vizhome/Project_versionupdated/Story1?publish=y

0.1.4 My Feedback link:

https://discussions.udacity.com/t/feedback-required-for-my-tableau-project-urgent-nearing-deadline/601757

0.1.5 Summary

The Flight dataset was downloaded from RITA website containing information on flight delays and performance. Here is the link for the dataset: https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp

I downloaded the data from December 2011 to December 2017.

Before beginning to dig into the visualization, I spent time in understanding the variables of the dataset and identified some of the interesting questions that I can explore such as:

1)What are percentage of different flight delays, which contribute the most and flight arrivals problems such as canceled/ diverted, these to be explored over different years, airports and carrier types?

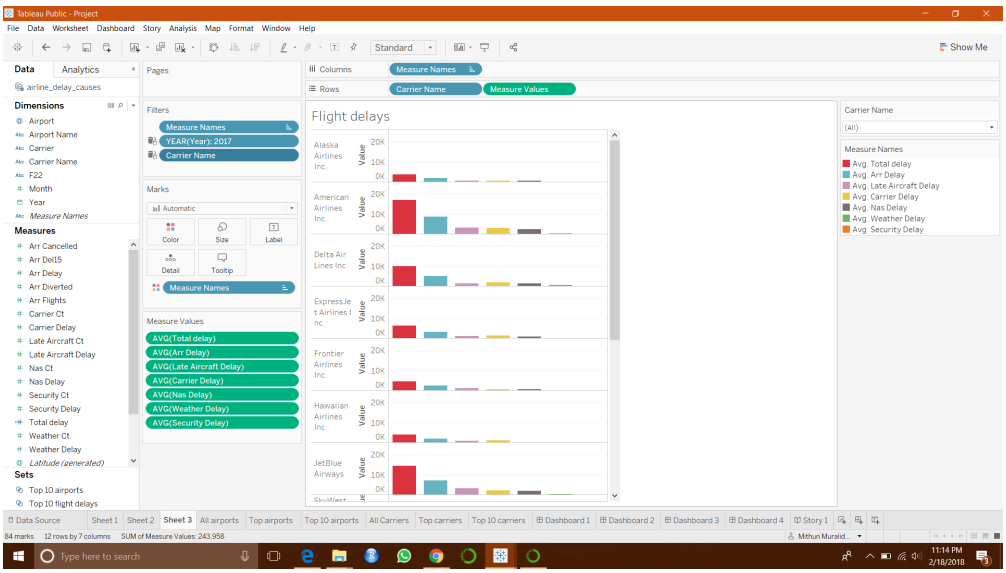
2)Which airports contribute the most for delays when all carriers are taken into consideration and for a particular carrier, which airports cause most delays and also what type of delay is the major contributor?

3)What are the carriers that causes most delays when all airports are considered?

4)What are the top 10 airports and carriers contributing for most delays and their relationship?

0.1.6 Design

In my first design, I wanted to explore different types of delays to see which are the major contributors. During this course, I created a new dimension 'Total delays' adding all other individial delays to see how they interact and appeal my visuals. I added years,carrier name and airport name to my filters to investigate more on delays related to a bunch of carriers/airports or for a particular one. Measure names with delay information was used for color to enhance the visual.



first

So my first design was a line plot of delays over months. Here is the image of the flight delays bar chart I first encountered

I received a feedback asked me to represent delays as a proportion of total delays. This was a meaningful feedback as it adds more clarity to the visuals. I measured values of all delays as a proprotion of total delays and created a new flight delay bar chart. Also I added carrier to filter to analyze the bar chart easily. The update now looks as follows

In my second design, I generated geolocation plot of airports. Initially there wasn't a variable to enabale maps, I converted aiports variable from string to geolocation to create maps. To highlight airports with high delays, I entered avergae total delay measure to size marks, also added it to color marks. So top airport with highest delay would look something like dark orange color with big circle seperated from the rest. Here is the image of my first gelocation of airports

I got a feedback "Enhance the visual in geolocation plot of airports by adding airport details". I guess he meant by adding airport name details because that is what missing in my plot. I added airport name to the details section in marks to allow us to see which aiports caused more delay. Here is the updated version

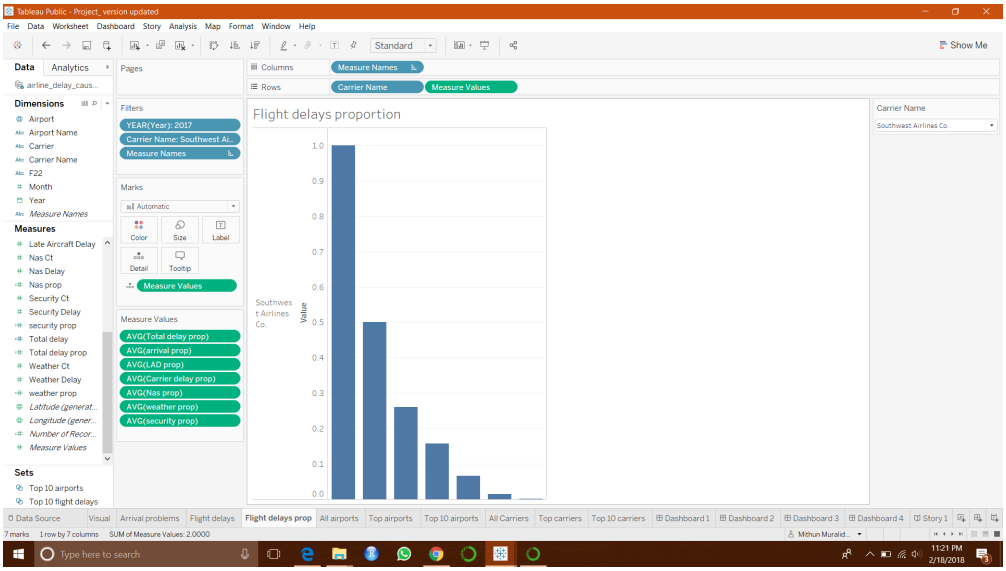
In my third design, I wanted to find which were the carriers that got delayed the most considering all airports and years. Also for a particular carrier, which airport caused the highest delay. This would give me valuable information in a single plot. I created heatmap for carriers and mapped it to geolocation plot. I added delay information details so that when we hover over the map, we get delay information directly. Here is the image from my dashboard

A feedback for this plot was to represent percentage of delays across different carriers. I updated the same and here it is

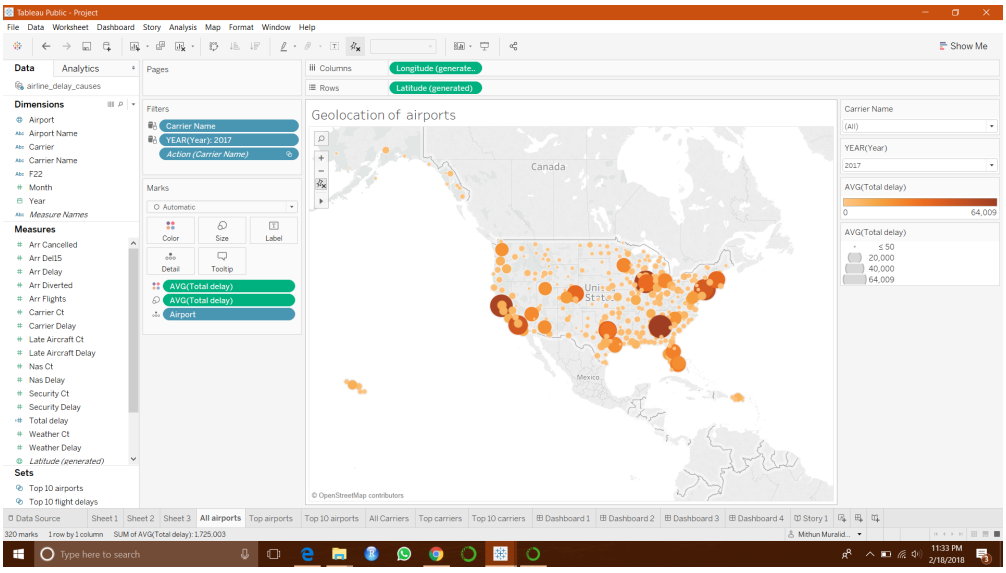
My final desgin involves creating plots for top 10 airports and carriers for most delays. I created two sets for the same. I created top 10 airports in geoplot and top 10 carriers in heatmap.

0.1.7 Feedback

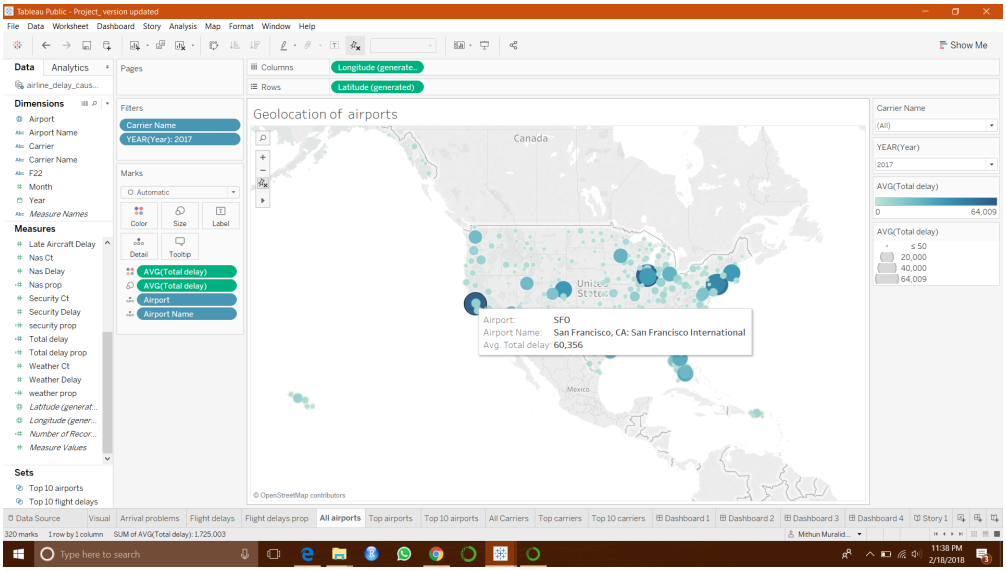
Hey Mithun, Here's my Feedback on your project based on my understanding:



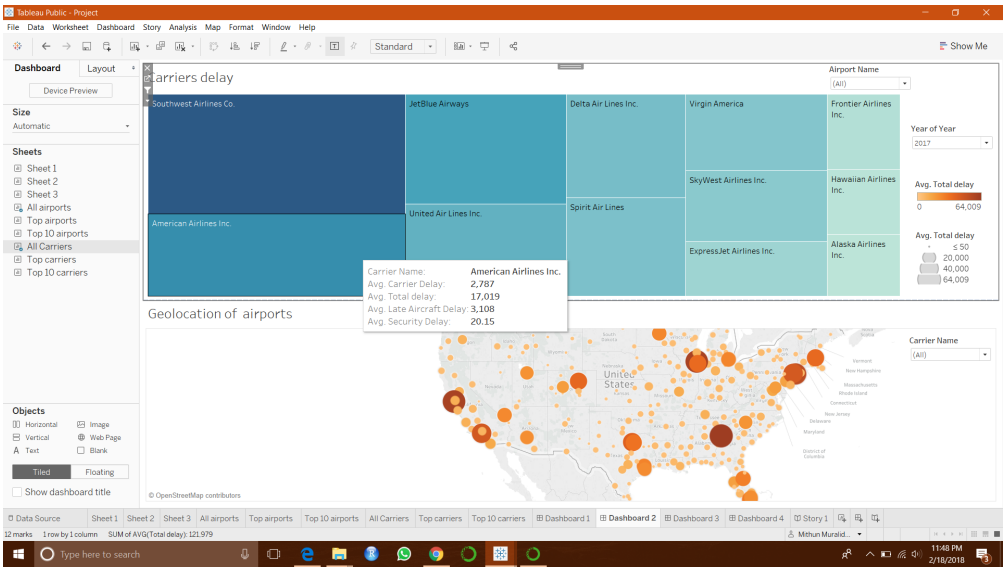
second



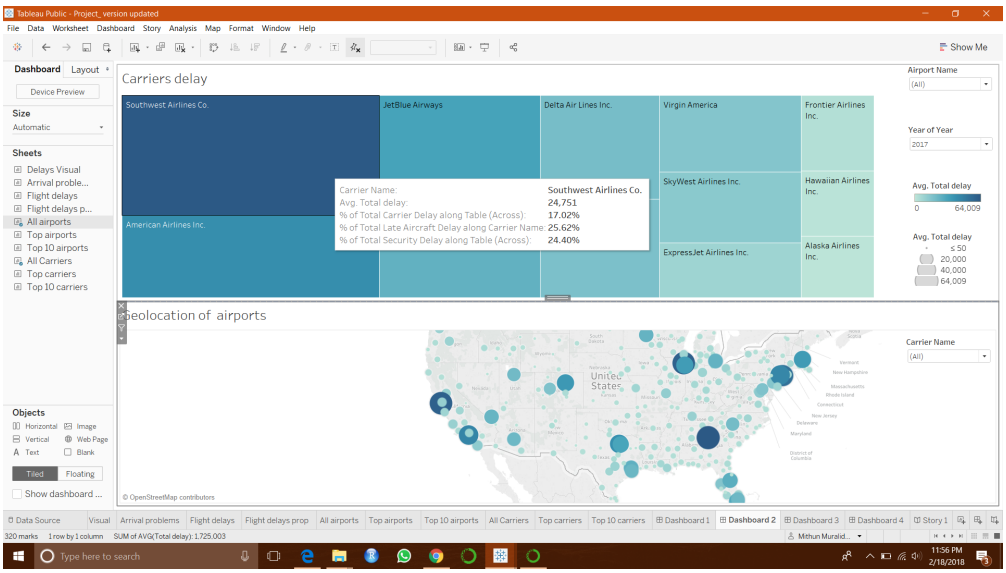
three


Geolocation of airports
Airport: SFO
Airport Name: San Francisco, CA: San Francisco International
Avg. Total delay: 60,356

four


Carriers delay
Southwest Airlines Co.
JetBlue Airways
Delta Air Lines Inc.
Virgin America
Frontier Airlines Inc.
American Airlines Inc.
United Air Lines Inc.
Spirit Air Lines
SkyWest Airlines Inc.
Hawaiian Airlines Inc.
ExpressJet Airlines Inc.
Alaska Airlines Inc.
Carrier Name: American Airlines Inc.
Avg. Carrier Delay: 2,787
Avg. Total delay: 17,019
Avg. Late Aircraft Delay: 3,108
Avg. Security Delay: 20.15
Geolocation of airports

five



image

1)In Flight delays, represent rest of the delays as proportion of total delay: Yes I made this change to my plot giving me a better visual details. Thank you

2)Enhance the visual in Geo location plot of airports by adding airport details: I made this change, this gives more insight to the data plot. Because as we hover oveer the airports, we get to see the names thereby helping us to identify easily.

3)Generate percentage of delays across different carriers: I made this change as well. A good suggestion because percentage is always a better metric to showcase it to your audience.

4)Refine the filter access to worksheets for better visuals: Works out!!

5)Filter alignments in story is not structured: Delivered this !!

0.1.8 Resources

Tableau documentation:https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields.html Udacity tableau lecture videos