Tableau Story Project Report

Summary

In this project, I created visualizations for the flight delays data from 2010 to 2018, which is downloaded from The U.S. Bureau of Transportation Statistics. The purpose of my visualizations is to answer the following questions:

- How does flight delay time change over time in comparison to the number of arrived flights? Is there any variation in delay time in different quarters of the year?
- How much time is delayed by each delay cause? Does the proportion change over time?
- Which airports have more flight delays? Is the delay time related to the number of arrived flights?

Through the visualizations, I showed that:

- Flight delay time have greater fluctuations compared to number of flights. The total number of minutes delayed usually increases in the first and second quarter of the year and decreases in the third and fourth quarter of the year.
- Late aircraft contributes to the longest flight delays, followed by carrier, national air system, weather, and security reasons. This proportion remains fairly constant over time.
- Airports with vary in delay time regardless of the number of arrived flights.

Data Assessment and Wrangling

Before creating the visualizations, I assessed the downloaded data set using the pandas library and noticed that the year and month data are recorded on separate columns as integers. To enable plotting of trends over time, I created a date column using the year and month columns. After that, I saved the dataframe into an xlsx file. The details and codes can be found in the flights_data_assess_wrangle.ipynb notebook.

Design

First Draft

https://public.tableau.com/profile/xiang.hui.low#!/vizhome/Flightsvizworkbook1/FlightDelays

In the first page of my Tableau story, I plotted line charts of total flight delay time and number of arrived flights over date. I placed these two charts next to each other for the viewers to easily compare the trends.

In the second page, I plotted a bar chart to show the total time delayed by each cause and a line chart to show the variation in delay time by each cause over the years. The bar chart

helps viewers to understand the total and differences in delay time between each cause, while the line chart helps viewers to see whether the proportion changes over time.

In the third page, I encoded the delay time using color and the number of arrived flights using size. This helps to display a lot of information about the airports clearly, such as which airports have more arrived flights and/or longer delay time. I also added full airport names as detail marks so that viewers can immediately know the name of the airport when they position their mouse on the circles. In addition, I added a filter (slider type) for the viewers to choose the particular year that they would like to see.

Second Draft

https://public.tableau.com/profile/xiang.hui.low#!/vizhome/FlightsDataVisualizations-Draft2/FlightDelays

In my second draft, I added filters in the first and second page of my Tableau story for viewers to select and view the year or range of years that they are interested in. I also changed the column names of Arr Delay and Arr Flights to Total Time Delayed and Number of Arrived Flights, respectively. This helps to prevent confusion when the viewers read the details of the data by looking at particular data points.

On top of that, I converted all the delay time data from minutes to hours for easier understanding by creating calculated fields and re-plotting my charts using the created fields.

Third Draft

https://public.tableau.com/profile/xiang.hui.low#!/vizhome/FlightsDataVisualizations-Draft3/FlightDelaysTableauStory

According to the feedback I received on my second draft, the x-axis labels in the first two pages were confusing, because only the tick marks for the first quarter of the year are labeled. Therefore, in my third draft, I used major tick marks to show years and minor tick marks to show quarters of the year. I only labeled the major tick marks (i.e. the years) because labeling the quarters would make the x-axis labels too crowded and difficult to read.

The reviewer of the second draft also claimed that the y-axis of the bottom chart in the second page was truncated. Upon inspection, I confirmed that it is intact and the bottom chart can be viewed in its entirety by scrolling down. Nevertheless, I changed the size setting of the page to automatic so that the page will resize itself to fit into any screen, enabling viewers to see the whole page without scrolling. This would help to prevent such confusion in the future.

In the same feedback, I received the suggestion to change the filter titles in my first and second page from Year of Date to Year, which I followed accordingly. For consistency purpose, I also changed the filter style of the third page to be the same as those in the previous pages.

In addition, I was told to smooth out the lines in the line charts by doing transformations. However, I think that plotting the lines using actual values presents the data in a clearer way

by allowing viewers to see the exact delay time/number of arrived flights when they are looking at individual data points. Plus, the fluctuations in the lines are not very high and do not justify the need for transformations.

Fourth Draft

https://public.tableau.com/profile/xiang.hui.low#!/vizhome/FlightsDataVisualizations-Draft4/FlightDelaysTableauStory

Instead of plotting the total delay time of all flights, I modified the visualizations to display the average delay time per flight. This would help the viewers to draw more meaningful conclusions when looking at the visualizations, such as the expected time delayed per flight due to each delay cause.

Feedbacks

First Draft

- Feedback 1:
  - Provide filter for date or year in first dashboard and second dashboard.
- Feedback 2:
  - In the map, Arr Delay and Arr Flights in the details of circles are confusing – can be solved by changing the column names.
  - Delay time, which is in minutes, can be converted to hours for easier understanding.

Second Draft

- Feedback 1:
  - In the first and second page, the x-axis labels are not consistent. You have plotted points for only quarter 1 and quarter3. What about quarter 2 and quarter 4?
  - Y-axis of the second plot is truncated.
  - Filter name can just be Year instead of Year of Date.
  - In the second dashboard also, it seems you have plotted only for quarter 1. Either you plot for 4 quarters or for an entire year.
  - Please smoothen the lines by doing transformations.
  - In the third dashboard, why have you used slider instead of drop down for the filter?

Third Draft

- Feedback 1:
  - You've used a counting statistic, meaning you've used the totals for both number of flights and total time delay in hours. Counts definitely have their place, but often for comparisons, a better statistic is some measure of

central tendency, such as mean or median. Please change your statistic to a mean or median to show more meaningful results.