Udacity Junior Data Analyst Nanodegree

Project 5 (US Flight Delays Dataset)

by

Elohor Okpako

INTRODUCTION:

This dataset is hosted on the Kaggle database. The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations.

url: https://www.kaggle.com/datasets/usdot/flight-delays

Description:

The analysis for this project was pooled from three datasets which describe the flight-delay of airlines in the US in 2015. The three datasets were merged into one by their common columns. There are 274,964 rows and and 39 columns in the final dataset which resulted from merging the three datasets.

In exploring the dataset, certain insights were discovered which were answers posed to certain questions that were asked.

INSIGHT 1:

- **Which airlines or airports have the worst delays?**

Southwest Air Line was observed to be the airline with the worst arrival delays with a total sum arrival delay of 289,992.

https://public.tableau.com/views/Sumofairlineswithworstdelays/Arrivaldelaybyairlines?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

However, is the **SUM** a true measure for comparison? I decided to use the averages instead, as the total corresponding flights by each airline might be different. In comparing the averages, **Spirit Air Lines** was identified as the airline with the worst delays.

https://public.tableau.com/views/SumAverageArrivaldelaysbyairlines/Story1?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

In both charts, interestingly, the airline with the best arrival time was **Alaska airline**

- **Design Choices**

In choosing the colours for the design, I considered the high ink-data ratio. I use just three different colors to highlight the necessary parts of the data that needed attention. A bar chart was also used because it was the best choice for showing clear differences among fields/categories.

- **Links/Resources**

I followed the thread on the research gate link on the discussion on whether to use sum or averages for accurate comparison

https://www.researchgate.net/post/What_the_justification_is_for_using_a_sum_score_instead_of_an_average

INSIGHT 2:

- **Which combination of Origin and Destination airports had the worst delays?**

The cumulative sum of the worst arrival delays was observed in the combination of Origin airport, **LAX (Los Angeles International Airport)** and destination airport, **SFO (San Francisco International Airport)** with a total sum of **9,251**. It is interesting to also note that the second worst arrival delays was equally observed for **SFO** and **LAX**, as origin and destinations respectively.

https://public.tableau.com/views/sumarrivaldelaybyairport/Sumarrivaldelaybyairport?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

On comparing the averages, **OGG (Kahului Airport)** and **ORD (Chicago O'Hare International Airport)** had the worst delays as origin and destination airports respectively.

https://public.tableau.com/views/combinationoforigindestinationforarrivaldelay/Combinationoforiginanddestinationairporstwithworstdelays?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

- **Design Choices:**

I chose a bar plot for this analysis because the height of the bars was a good way of representing the summation/averages of the arrival delays. I applied a filter on the averages because the values obtained were too many to be displayed all at once.

- **Resources/Links: NA**

INSIGHT 3:

- **What causes delays?**

https://public.tableau.com/views/Arrivaldelaytrendwithmonth/DelayTrendwithmonth?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

From the chart above, it is observed that the peak arrival delays peaked in June while the lowest is observed in September. It is interesting to note that arrival delays created a trend all through the year.

What factors Influence Arrival Delays??

In attempting to answer this question, Security delay, weather delay, late aircraft arrival and Air system delay were considered and the pattern over the year was also observed.

https://public.tableau.com/views/Whatcausesdelay_16555486647920/Story2?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

In the chart above, It can be observed that **late aircraft** contributed more towards the delay than any other reason. It is also interesting to know that the delay fluctuated with the months of the year. **Air system delay** and **weather delay** were other contributing factors, while security delay seem not to be a reason.

- **Design Choices:**

A trend line was chosen for this visualization to show changes over time. In choosing the colors for identifying the reasons for the delay, the color choices for color blind persons were also considered. The blue-green comparison color choice was avoided.