Project 4: Storytelling with Tableau - "On-time performance of US domestic flights operated by large air carriers in 2015" dataset

In this report, I draw my insights on various questions that I wondered about and wanted to capture the answers through data visualizations in Tableau. I used a variety of visualizations according to whatever I deemed it to be suitable for such insights. I briefly touch on what are the next step(s) to be taken after sharing the insights drawn.

1st insight: Which states/cities have the highest number of cancellations?

Link: https://public.tableau.com/profile/hayat7638#!/vizhome/BAND-UdacityP4-DB/ofcanceledflights-DB?publish=yes

Summary: In this Dashboard, I included a map, packed bubble and bar chart to showcase the same information (from a broad to a narrower view on data). Starting with the map from the far left, it showcases the states with the most cancelled number of flights. We can see that Texas has the highest number of cancelled flights as that amounts to a total of 668 cancelled flights. Following closely behind, we have Illiones with a total of 563 cancelled flights. Next in line comes California with a total of 408 cancelled flights. Moving on to the bubble chart, we get to see a more specific view on the data as this chart includes the cities and not only the states (Something like a breakdown). Over here we can see that Chicago from the state of Illiones, IL is exceeding other cities with the number of cancelled flights (a total of 541).

Linking this information with the information mentioned previously, we can say that the number of cancelled flights in Illiones occurs mostly in the city of Chicago (563-541 = 22 for other cities in IL). This means that in order to minimize the total number of cancelled flights in IL, the focus has to be directed on tackling the cancellation reasons occuring in Chicago airports. Next in line is Dallas-Fort Worth from Texas (a total of 343). Furthermore, what is interesting here is that we can see that New York, NY is showing as the third city/state with

the highest number of cancellations (a total of 327) where as somoene might think it would be another city from the state of Texas or California (Though, those do follow closely behind in terms of numbers). Lastly, we have a bar chart showcasing only the cities in descending order for the number of cancellations (The simplest intrepertation out of all the charts).

Design: As mentioned earlier, my intention was to showcase the data from a broad to a narrower view, so that influenced my choice on selecting the visuals. The map was the most broad chocie as it showcases all the states and then to narrow down more on the data, I chose to show both the state and city as a clustter of data and this is where I found the packed bubble chart to be the most useful. Also, those two choices helped me in my analysis mentioned above. Furthermore, I chose the bar chart to display only the cities and the purpose of that chocie was to showcase the data in a simple, understandable manner without deciving the eyes with different sizes or distracting the reader with different colors. Also, the dashboard colors stir away from any combination of Red-Green colors as those are unsuitable for color blindness. I have included a filter where the user can see how the number of cancellations vary across the days of the week for states/cities. The insights I drew were based on selecting the whole week.

Resource: **https://www.kdnuggets.com/2018/05/6-tips-effective-visualization-tableau.html**

2nd insight: What causes the most delay?

Link: https://public.tableau.com/profile/hayat7638#!/vizhome/BAND-UdacityP4-PieChart/CancelationReasonby?publish=yes

Summary: Over here we have a simple pie chart that showcases the cancellation reasons in terms of percentages of total. We can see that "Weather" has the highest percentage (54.07%). Weather is an uncontrollable cancellation reason, so it is best to plan flights according to weather forecasts and to avoid having to incur a cancelled flight because of poor scheduling/alignment with forecasts. Next, we have "Airline/Carrier" (28.42%). This cancellation reason could be composed of multiple reasons as it is unclear to what exactly this cancellation reason is composed of and what should be the first problem to be tackled (Quality, management, staffing, etc.). Lastly, we have "National Air System" (17.51%), a cancellation reason that is also hard to tackle as this is a broad system and it includes multiple entities.

Design: I used the pie chart as my intention was to display the percentages of total and the data is of 6 categories or fewer, which made it suitable for using the pie chart. Also, the color palette chosen was the "Color-blind" palette.

Resource: N/A

3rd Insight: How do cancellation reasons vary across different airlines?

Link: https://public.tableau.com/profile/hayat7638#!/vizhome/BAND-UdacityP4-StackedBarChart/StackedBarChart?publish=yes

Summary: From this stacked bar chart, we can see how the cancellation reasons vary across different airlines and which airlines have the worst performance according to the accumulation of number of cancellation reasons. Southwest Airlines Co. have it the worst. It seems that this airline struggles mostly with the weather (a total of 446). Proper scheduling with weather forecasts can maybe help with the reduction of cancelled flights. Next, we have Atlantic Southeast Airlines, they struggle the most with National Air System (a total of 335). Now this is the opposite of the first airline and the reasons for that could be studied further if we were interested to see why different airlines struggle with different cancellation reasons in the same country. Generally, all airlines struggle with the weather mostly as that is logical and can be withdrawn from the previous pie chart mentioned.

Design: Again, I chose the "Color-Blind" color palette, and the stacked bar chart was the most suitable chart to showcase different categories (cancellation reasons) for one variable (airline).

Resource: N/A

4th Insight: Airline Delay vs Arrival Delay, which delay is more prominent?

Link: https://public.tableau.com/profile/hayat7638#!/vizhome/BAND-UdacityP4-ButterflyChart/ButterflyChart?publish=yes

Summary: For this insight, I used an interesting chart (Butterfly Chart) to showcase the contrast (side by side) between the two types of delays in terms occurrence. Generally, it seems that airlines have a higher number of airline delays than arrival delays (it is shown in the way the bar charts are higher on the left side of the chart). Therefore, the focus when trying to improve, should rely on addressing the airline delays as they occur more often. Southwest Airline Co. is topping the chart and that makes sense as it does have the highest number of cancellation reasons (From the previous insight) and those lead to incurring delays.

Design: As mentioned earlier, I wanted to showcase the contrast between the two types of delay and that influenced my choice of visualization. Moreover, the colors I have chosen for this chart do not conflict with the color-blindness color scheme (No combination of Red-Green colors).

Resource: https://www.youtube.com/watch?v=M74p3N3lnbI

https://www.pluralsight.com/guides/tableau-playbook-diverging-bar-chart