PROJECT – DATA VISUALIZATION IN TABLEAU

1) Flight Delays and Cancellations in the US

Summary

The **first visualization** suggests there exists a relationship between the weather conditions and the flights cancellations, particularly in January and February where airlines and airports, as a measure of precaution, cancel almost all flights that get delayed by the traditionally bad weather conditions of those two months. Through the rest of the year, the weather is relatively more favourable, so we can see that the number of cancellations is minor, even if more flights get delayed by the weather. A line chart was used herein to represent this relationship monthly, in the course of the year. A dual axis on the rows and a colour legend for the two variables were added to better read the visual and understand the variance of the two variables.

With the **second visualization**, I wanted to understand the distribution of taxis waiting outside of airports nationwide by region. (Thus using the Count aggregation to illustrate the "Taxi Out" measure on the map). Texas and California are, far and away, the two states where the most taxis wait outside of airports. A map chart was appropriate to visualize this occurrence by region, and a filter with a dropdown list of the States was pinned to the map to manage the visualization better.

On the **third visualization**, I decided to sum up arrival delays with departure delays (Calculated Field: Delays = [Arrival Delays] + [Departure Delays]) in order to get the aggregate amount of delays for each airport, being thus able to check their delay performance on a whole. The bar chart was the appropriate chart to present here because we are working with categorical data and non-sequential quantitative data. Filters of the airlines and the airports were added, through dropdown lists, to better handle the data.

The **forth visualization** suggests that, by comparing the amounts on the y axis of the chart with each other, departure delays, arrival delays and late aircraft delays are the worst kinds of delays on record. In addition, we can realize that the delays do vary depending on the time of the year – we see a peak in June and another peak at around the New Year period. A line chart was used here because it is of interest to see how each type of delay performed during the year (sequential quantitative data). Filters of the airlines and the airports were added, through dropdown lists, to better handle the data.

Nextly, I decided to investigate the most severe of the delays: departure delays, to understand its genesis and how it can be tackled by airlines and airports.

Hence, on the **fifth visualization**, via a map chart, we can see, across the country, that this particular type of delay was most severe (largest percentages) in the states of Texas, California, Illinois and Florida. A map chart was the obvious choice to visualize this phenomenon. Table calculations of Percentage Totals and their labels and the colouring intensity for each state were included to better represent and understand the situation at hand. Also, a filter of the states was added, through a dropdown list, to improve the data handling.

For these last three visualizations, I looked into how these large and diverse datasets could be most pleasing to visualize and handle and easier to understand in the charts. On the first one, the horizontal bars seemed more intuitive than the vertical ones as well as each airline and each airport name being better displayed this way on the y axis. On the second, the time period of interest

(months) was displayed on the x axis to improve the perception of the performance of each type of delay through time; also, I thought it would make it less perceptible to do a line chart where all the seven delay variables would crossed-over each other with multiple (and probably conflicting) colours, so they stayed compartmentalized. The third visualization was more direct than the others in a way; no special design improvements were needed for it to render itself intuitive to the reader.

Building the **dashboard** for the last three visualizations, all three of them were united, automatically sized and all the data sources (States, Airlines and Airports) were related to understand the "big picture" in a particular case of the reader's choosing. The State dropdown filter and the Percentage of Total Departure Delays colour spectrum, on the map chart, were floated to make the visualization slicker and more intuitive.

Lastly, building the **story** we want to convey, in line with the dashboard.

So, firstly, having the amount of delays for each airport and the respective airline, we will want to look closer at each type of delay to see how we can manage one or several of them in a specific or in multiple airports and/or airlines. By realizing the severity of the delays on the departure stage, we can investigate in which regions of the country it was most grave, in order to tackle it more effectively. Texas and California, in particular, clearly are the sloppiest on departure delays.

Therefore, with the insights presented above, what airlines and airports could improve on, and maybe work together on, is to act proactively on these aforementioned states by, for instance, increasing personnel and their benefits, and by implementing seasonally extraordinary organizational and managerial methods, since they know the most critical times of the year in terms of these delays. If planned and executed correctly, airlines and airports, with only this type of delay considered, will potentially see a rapid increase in customer satisfaction and operational efficiency, resulting in increasing revenues and decreasing costs.

Visualizations in Tableau Public

- Worksheets

1. What is the relationship between the weather conditions and the cancellation of flights in the course of the year?

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/WeatherCancellationsRelationship?publish=yes

2. What does the distribution of the number of taxis outside of Airports look like at a state-level?

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/OutsideTaxisbyRegion?publish=yes

3. Which airlines or airports have the worst delays?

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/QuestionOne?publish=yes

4. What causes delays?

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/QuestionTwo?publish=yes

5. What does the distribution of departure delays (worst delays) look like at a state-level? In which is this occurrence most severe?

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/QuestionThree?publish=yes

- Dashboard

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/Dashboard?publish=yes

- Story

https://public.tableau.com/profile/bernardo.vicente#!/vizhome/PROJECT_15904404963210/Story

Resources

https://knowledge.udacity.com/

https://classroom.udacity.com/nanodegrees/nd098/parts/05a8ba39-eb63-426e-9b98-755883bc81d6/modules/f922e0f7-d718-4d00-a0e7-13b8498cf7d3/lessons/bc51e4f0-dd29-49f9-87fb-982581e93583/concepts/1c43be37-bc7e-40a0-bf71-31efa6e200dd

https://www.youtube.com/watch?v=Xbg2WabUIKI

Bernardo Vicente, 29/5/2020