Introduction:

In this project, I created visualizations to reveal insights from a data set. This data comes from a Kaggle dataset, it tracks the on-time performance of US domestic flights operated by large air carriers in 2015. You can find the dataset in supporting materials at the bottom of this page.

Q1: Which airlines or airports have the worst delays?

A bar chart was used for that visualization. The viz is comparing the different airlines to the delays only caused by airlines and the different airports to the delays caused by the summation of both (arrival and departure delays). The Southwest Airlines (WN) had the worst delays with 182,670 minutes of delays while in further analysis it turned out that Chicago O'Hare International Airport (ORD) had the highest delays with 316,610 minutes of delays.

A map was used in the airports delays to represent the geographical date with only one colour and the amount of delays was represented by the size of the circles. However, I used a bar chart for airlines with degrading blue colour for colour blind people while the darker the colour the more there are delays by that airlines while bar charts compared the difference in the amount of delays.

Q2: What causes delays?

A pie chart represented this viz. The viz was to compare the different reasons of delaying. A new csv file was introduced for easier visualization of the reasons that simply included the delay type and the explanation (a screenshot is shown below). It turned out that 54% of the reason of delays are caused by the weather as expected. 28% by Airlines while 18% by National Air Systems. 0% of delays were caused by security in this dataset.

I used a pie chart to represent the percentage of delays caused by each reason.

Q3: Which day or month had the highest delays?

A line chart represented both vizs here. Both vizs were used to compare days of the week to overall delays (summation of both arrival and departure delays) and months to overall delays as well. It turned out that the worst day to deliver is on the start of the week (Day 1) with 709,190 minutes of delays while the best day to travel is the 6th day with almost only half the minutes of delays with 349,042 minutes of delays (can be considered as half the chances to be delayed). Moreover, and as expected there is higher possibility of being delayed around summer and winter (June and December with 604,992 and 455,165 minutes of delays respectively) while the chances of being delayed drops around spring and autumn (April and September with 249,131 and 96,274 minutes of delays respectively).

I used a line chart because I wanted to plot a time series continuous data (days and months) over time.

Tableau Public Link:
https://public.tableau.com/views/FlightDelays_16559440949720/Delaysvs_DaysandMonths?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link