Project: Create a Tableau Story

Task : Create an **explanatory** data visualization from a data set that communicates a clear finding or that highlights relationships or patterns in a data set. Your work should be a reflection of the theory and practice of data visualization, and your final deliverable will be a write up along with a Tableau Public workbook.

Steps:

1. **Choose a Data Set**
    a. A Data Set of flight delays was chosen from the Bureau of Transportation Statistics website https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp
    b. Data provided in the data set was data from 1990-2008. A more recent data set was required for more pertinent analysis
    c. The website has options to select specific fields and download by month and year.
    d. Flight delay data from 2017 was downloaded month-by-month. Other reference data such as airport codes, carrier names etc. were also separately downloaded

2. **Get Organized**
    a. Summary: Flight departure data for over 5.4m flights that originated in the US in 2017 was analyzed to determine insights that could be obtained through underlying data. Data which was in the form of 12 monthly csv files and additional reference files were then cleansed with Python using the Jupyter notebook and os (to work with file operations), pandas (to work with dataframes merge, concatenate, data type manipulation), numpy (for certain array operations) and geopy (to obtain the geocodes for the airports from addresses) libraries. Data was then visualized on Tableau worksheets to determine patterns
    b. Design: These flights could vary due to multiple reasons – by the airline, by the airport, by the time of the day, by the month, by season (incl holiday periods), by flight volumes, etc. The metric itself lends to 2 distinct metrics 1.
        i. The actual delay in millions of flight minutes,
        ii. The average delay in minutes for each flight and
        iii. The average delay in minutes vs the flight time for each flight
        iv. So, I decided to have just the actual delay time metric and the avg delay per flight so as to not confuse with 2 different average delay metrics for the user

c. Individual worksheets were created that looked at such variations i.e. by carrier, airport, hourly time blocks. Flight delays could be visualized by airport, carrier, hourly blocks of time during the day as well as the overall monthly variation
    i. Flight data showing origin airports across the US would be interesting to plot on a map. Hence, the latitudes and longitudes were plotted to give a view of all the 320 airports being covered by the data. It would also be useful to show the carriers that operate out of these airports – so the carriers have been provided as a filter so users could then play with the data to understand which carriers operated out of which airports
    ii. Carriers, Airport Names, and time blocks could be defined as categorical data and the metric could be overall flight delays or avg delay per flight. So, I decided to use flight delay overall as a bar chart while avg delays would be lines on a Gantt chart and these would be displayed side by side – so users could understand the nuances – you could have large number of flights and hence delays from the hubs but you would also use the avg flight delay as another metric to understand the dynamic better
    iii. I also decided to add another view with filters on airport name, carrier and month to encourage users to do their own analysis and draw their own insights. Flight delays were coded to millions of minutes and average flight delays were coded to minutes.
    iv. Flights are scheduled to address demand and hence there could be seasonal variations in the number of flights, and that could also result in delays. So, a viz was created to show weekly variations in the number of flights, delay minutes and avg delay per flight that could pick up spikes around holiday periods
    v. A viz was also created to show the components of the delay by month
d. Feedback: The preliminary Tableau public workbook was then loaded and people were asked to give feedback on content and opportunities for improvement. The feedback received from the mentors were around naming convention and labels being cut in the visualization
e. Data Files used have been attached to this submission.

Initial Visualization:
https://public.tableau.com/profile/ravi.gopalan#!/vizhome/FlightDelayAnalysis-2017/FlightDelayAnalysis

Final Visualization:
https://public.tableau.com/profile/ravi.gopalan#!/vizhome/Analysisof2017USflightdelays-Final/FlightDelayAnalysis-Final?publish=yes