PROJECT 4

BUILD A DASHBOARD

[FLIGHT DELAYS AND CANCELLATIONS]

Introduction

- This document explores visualizations generated from a Kaggle data-set, which tracks the on-time performance of the US domestics flights operated by large air carries in 2015.
- The visualizations are targeted at stakeholders who are part of the management in the US domestic flights space.
- The visualizations should provide sufficient management information to the relevant management stakeholders.
- Kaggle link : Flight-Delays

How :

- Prior to creating the visualisation for this project, a considerable amount of time was spent in optimising the data set.
- An aggregated date column , **DATE** , was created using the query :
  - `MAKEDATE([YEAR],[MONTH],[DAY])`
- This column converted the Number Data Type (Year, Month and Day) into Date Data Type.
- Another aggregation number column,**TOTAL**, was created using the query:
  - `[DEPARTURE_DELAY]+[ARRIVAL_DELAY]`

Visualisation 1 : Which airlines or airports have the worst delays and what causes these delays?

1. Tableau Link: Airline & Airport Dashboard
2. Design and Summary:
   - The aim of this dashboard is to give management a consolidated view airport and airlines that have the worst delays along with the causes in each quarter of the year 2015.
   - The Dashboard is divided into three quadrants.
   - The first quadrants (top left) calculates the average total delay for each airport and displays the airports with the highest delays using a packed-bubble chart.
   - The size of each bubble is proportional to the Total Delay with the airport with the highest delay highlighted in orange.
   - GMU – Guam International Airport - as seen on the dashboard has the worst delay with Spirit Airlines having the worst delay across USA in 2015.
   - The second quadrant (top right) shows is a bar graph of the average total delay for each Airline ordered by the average total delay in descending order.
   - The last quadrant (bottom) shows how the possible causes varies across the year 2015 in each quarter.
   - **Late Aircraft Delay** was the top possible cause across all the quarters of the year 2015
   - The dashboard is also dynamic as it allows the stakeholders to click on each bubble to get insight on which airline/s in that airport had the worst delay and as well as observing the possible cause trend.

Visualisation 2 : Which inland states has the worst delay and is there any relationship between arrival and departure delay ?

1. Tableau Link: Arrival & Departure Relationship Dashboard
2. Design and Summary:
   - The aim of this dashboard is to give management a national view of the total average delay and the corelation between arrival and departure delays.
   - The Dashboard is divided into two quadrants.
   - The first quadrants (left) calculates the average total delay for each state and is presented using a heat map. As seen on the heat map the inland state , IL – Illinois, has the worst delay out of all the inland states
   - The second quadrant show the corelation of the average arrival and departure delay. The scatter plot shows the on a national level there is a positive relationship between average arrival and departure delay with a moderate-strong strength.
   - The scatter plot also show management the four different clusters that the various airports can be clustered in.
   - Cluster 1 – Low Delay Airports , Cluster 2 – Low Medium Delay Airports , Cluster 3 High Medium Delay Airports , Cluster 4 High Delay Airports ,
   - If we dive deeper and click on the state of IL (that has the worst delay) , we see that the scatter plot indicates a positive relationship with a moderate strength.

Visualisation 3: Further MI Story

1. Tableau Link: [Further Insight Story]

2. Design and Summary:
   - The aim of this story is to show management the distribution of the airline trip (in minutes) and to observe the relationship that between distance and total delay for each airline trip.
   - The story has two pages.
   - The first page shows us how the "Elapsed Time" for each airline trip is distributed by using a histogram.
   - The dataset is grouped in bins of 30 min.
   - From the graph we can see that the distribution is positively skewed with the most frequent airline trip time lying between 60 min and 90 min (mode).
   - The graph also tells us median "elapsed time" of this dataset is most like around 180 min with the average frequency lower than that on the mode and mean.
   - If we move on to the second page, we can see a blue line that shows a negatively skewed distribution of the average distance of each airline trip for various airline ranges.
   - The orange line graph shows us the average delay in each "elapsed time" bin. We can see that there was a significant spike that tells us that there was a high delay for airlines trips between 510 min – 540 min
   - This graphs shows us that the average airline trip distance has minimal effect the amount of time the airline spend between origin to destination.

3. It is suggested that management , to improve passenger experience , find ways to decrease the average total delays at airports as this has a significant impact on the "elapsed time" for each airline trip.

References

1. Eric van Holm, P., 2022. *11 Displaying Data | Introduction to Research Methods*. [online] Bookdown.org. Available at: <https://bookdown.org/ejvanholm/Textbook/displaying-data.html> [Accessed 20 January 2022].
2. Help.tableau.com. 2022. *Create Maps that Show Ratio or Aggregated Data in Tableau*. [online] Available at: <https://help.tableau.com/current/pro/desktop/en-us/maps_howto_choropleth.htm> [Accessed 22 January 2022].