Data Visualization Project – Flights Information 2015

Visualization 1:

https://public.tableau.com/profile/sameera.preethi#!/vizhome/Dash2_23/Dashboard2?publish=yes

Summary:

This visualization shows total weather delays across different states and % contribution of weather delay of their respective airports.

Insight:

Based on the data provided in flights.csv for 2015, Chicago has the most number of weather delays with a total of 28,766 minutes and Chicago O'Hare International airport contributes to close to 84.76% of total weather delays in that state. Following Chicago is Texas is with the next highest number of weather delays at 27,758 minutes and Dallas/Fort Worth International airport contributing to more than 58.64% of the total delay in that state.

Design:

Initially I made a map of states with sum of weather delay as measure in the size pill of the marks card. Then I wanted to make it clearer so I changed the map to 'shaded map'.

Then I created a bar chart with airport names and weather delays by taking the measure as sum. To make it better, I now added a quick table calculation to the sum of the weather delay to reflect 'percent of total' to show which airport within the state is contributing to most weather delays.

Now I created a dashboard using these two worksheets and used the map as a filter to filter the airports by state.

Resources:

https://eazybi.com/blog/data_visualization_and_chart_types/ for selecting appropriate charts

Visualization 2:

https://public.tableau.com/profile/sameera.preethi#!/vizhome/Dash1_60/CancellationDiversion?publish=yes

Summary:

This visualization shows total number of Diverted and Cancelled flights by Airline. This also shows the number of cancelled or diverted flights over different months in 2015.

Insight:

Based on the data provided in the flights and airports for 2015, Southwest Airlines had a total of 181 diverted flights and 818 cancelled flights.

July has most number of diverted flights (34) and February had most number of cancelled flights (171).

Design:

Initially I created a bar chart for displaying the total number of cancelled and diverted flights by airline. Since this has dual axis, I changed it to single axis and synchronized the values. When I changed it to dual axis the map automatically changed to scatter plot. However since we are only comparing the two variables, I changed it to a side by side bar chart.

To show the number of flights cancelled or diverted by months I used a line chart. I got two line charts with different axis for cancelled and diverted flights separately, but I wanted to show these two measures on the same axis so again I used the dual axis and synchronized it and now I have two measures on same axis with different colors. Initially I used only the month (datatype: number) to display the chart however since the numbers don't make lot of sense, I used the MAKEDATE () to create a new calculated field 'DATE'. I converted the DATE field to continuous and drilled down to month level data which now displays the names of months.

Now I created the dashboard with these two worksheets, removed the titles for worksheets and added a title for the dashboard. I used the bar chart worksheet as a filter so you can filter the data by airline which reflects the trend of cancelled and diverted flights over different months across 2015 of that particular airline. I also included the 'cancelled' and 'diverted' measure legends so we can filter out the data for only 'cancelled' or only 'diverted' flights and compare the values across various airlines and view the data.

Resources:

https://www.interworks.com/blog/jwright/2012/06/18/multi-measure-dual-axis-charts for dual axis charts

https://blog.bonzaiengineering.com/how-to-create-a-dual-and-synchronized-axis-chart-in-tableau-5808789c6df5 for creating a synchronized axis

https://community.tableau.com/thread/146656 for converting the year, month and date columns to a single 'Date' column with date datatype.

Visualization 3:

https://public.tableau.com/profile/sameera.preethi#!/vizhome/TheStoryofSouthWestAirlines/TheStoryofSW?publish=yes

Summary:

This story tries to uncover which airlines has most delays overall and what regions of USA were most affected by the delays and what is the reason behind this delay. Once we uncover the reason for delay it shows which month was most affected due to that delay.

Insight:

Based on the data provided in flights.csv for 2015, Southwest Airlines had the most delays. East coast was most affected by these delays and the main reason for the delay is 'Departure delay'. The month of July had the most number of departure delays.

Design:

I created a calculated field called 'Total delay' by summing up Airline delay, Arrival delay, Departure delay, Air system delay, Late aircraft delay, weather delay and security delay. Now I created a bar chart using the total delay measure by airline and then changed it to tree map for a better view.

Then I created a map of states and then grouped all the states on West coast to 'west coast group', all the east coast states to 'east coast group' and the remaining states to 'Central America group'. I named this new group filed as 'Regions USA'. I now moved the total delay measure to the color pill of the marks shelf and added the airline filter to work across selected worksheets.

Next I created a new worksheet with the new group 'Regions USA' in columns shelf and sum of measure values of all the different delays in the rows shelf like airline delay, departure delay, arrival delay etc.; This initially was displayed as a side by side bar but due to repetition of labels and legends and multiple colors, I removed the label name for measures and only retained the legends which resulted in a stacked column chart.

Now since I have identified the reason of the most delay (which is departure delay) I wanted to see the % change in departure delay across 2015. I used a line chart here. First I created a calculated field called 'DATE' by using MAKEDATE () and merging the Year, month and day columns. Now I used this in columns shelf and drilled down to month level data then I used Sum of departure delays in rows shelf. Now I added a quick table calculation on departure delay to show 'Percent of Total' across different months.

Resources:

https://onlinehelp.tableau.com/current/pro/desktop/en-us/multiple_measures.html for showing multiple measures on synchronized axis

http://www.vizwiz.com/2011/08/tableau-tip-7-easy-steps-to-create.html for overlapping line charts

https://community.tableau.com/thread/146656 for merging the year, month and day columns to a single 'Date' column with date datatype.