



# SAN DIEGO STATE UNIVERSITY

**Assignment 2**  
**Tableau & MicroStrategy**

**Submitted by**

Harshal Sanap  
822044776

**MIS 691: Decision Support System**  
**Jan 27, 2019**

## Question 1: What did you do?

### Part Tableau:

- I installed tableau from the Teradata network and then watched introductory videos to get acquainted with the tool. I later checked for various data sources that can be connected to Tableau, I realized that apart from using static stored files like excel or csv, we can also use dynamic cloud storage files in tableau.
- Then I downloaded the bird strikes dataset for analysis and loaded the file into tableau as standard excel file upload. The file upload was quick, and I was able to get a quick overview of the data on loading it.
- I explored the data in tableau to check whether all the variables are maintained in correct format and tableau did a good job to read the variables in correct format automatically.
- Now to generate some insights, I went to the worksheet tab. I understood the difference between dimensions and measures. Dimensions are categorical and measures are numeric values from the dataset. I also understood different cards like marks, filters, pages etc. present in tableau.
- First insight needed to understand **which departure state has the highest total monetary costs** as a result of bird strikes. For this I first created a bar chart with states on x axis and total monetary cost on y axis but it made a very long bar chart visual thus I switched to a geographical visual as mentioned in the write up. After dragging the origin state onto the worksheet it created a map of USA and later I dragged total cost field onto the size marks and the visualization was ready.
- For the second insight I needed to **get the average speeds of the airplanes for different states**, for this I simply dragged speed of an airplane onto tooltip mark of the above created visual. The avg speed information was then available on hovering over any point of the visual.
- Lastly, to understand **the total damage cost associated with different species of birds**, I created a bar chart with different bird species on y axis and the cost associated with them as value on x axis. As we have many species, I limited it to see the top 10 species in terms of cost associated with them and I also sorted the results according to the cost. Finally, I added labels as cost and number of records.

### Part MicroStrategy:

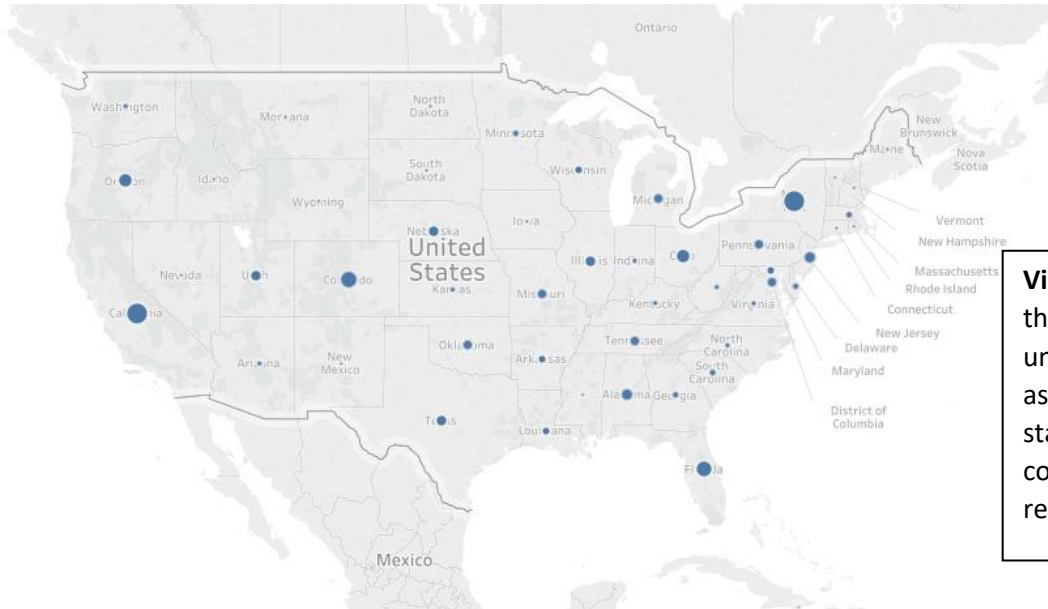
- After installing the software, I explored the software myself. I realized that it is quite like tableau with a different environment. I then watched quick tutorials on MicroStrategy website to get acquainted with the tool.
- For analysis I chose airline sample dataset. To **get the busiest airport**, I created a chapter in MicroStrategy and dragged airlines and number of flights on to the visual section of that chapter. Later I changed the visual type to bar chart. Visually it was seen that BWI had the highest number of flights, but I added a reference line to explicitly mark the highest number of flights.
- To **get the best hour for departure for BWI airport**, I first created a calculated metric as division of on time flights by total number of flights and set its aggregation type as average. Then I dragged hour and the above calculated measure in to the visual section to get a bar chart of hourly distribution of the calculated measure. Later as we wanted to analyze the metric only for BWI airport, I added that as a visual filter.
- To understand **the busiest month in year 2010 for southwest airlines**, I dragged month and total number of flights onto visual section and made visual type as bar chart. Then I added a filter as year 2010 and airlines as southwest to get the required result. I later also added a reference line for maximum number of flights.
- To get the **list of airlines with the least delays across different airports**, I selected the horizontal bar chart visual. Then I dragged airports and airlines onto y axis / verticals such that airport dimension was above airlines in order and dragged delays onto horizontal section of the chart.

Later I set the aggregate type of the measure as average. I also added reference line to highlight the lowest delays across all the airports and sorted the charts ascendingly according to average delay time.

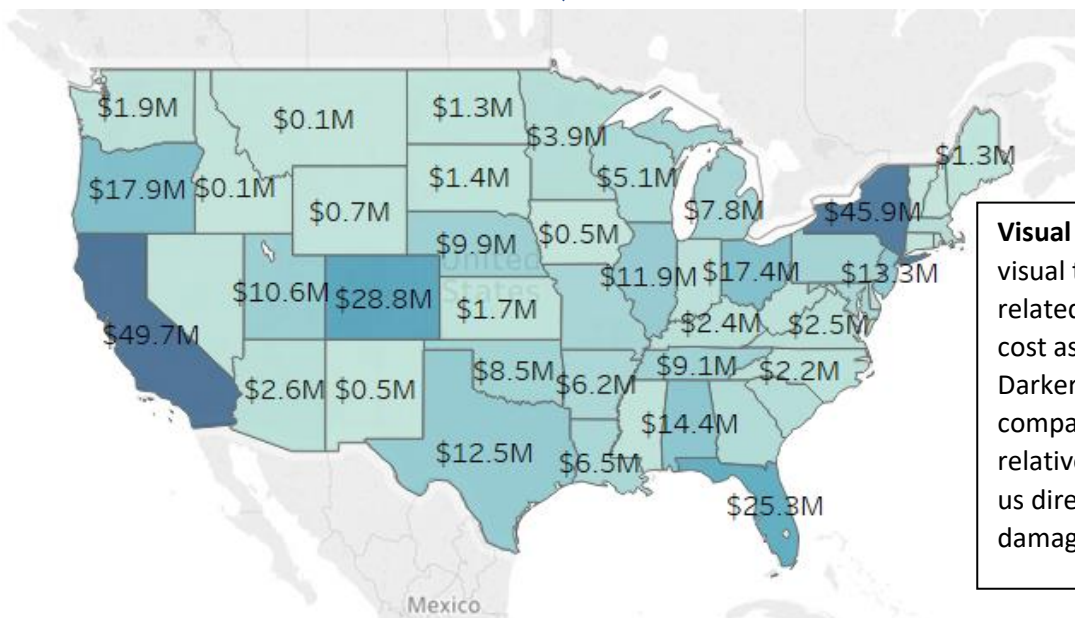
## Question 2 : What were the results?

### Tableau:

- State of **California** has highest cost of almost **\$49.7 Million** associated as damage caused by bird strikes. The visual created by following the steps mentioned in the instruction document was not interpretable enough thus I first switched the visual mark type from automatic to map and dragged total cost to title mark and changed the total cost format to currency in millions to get the below visual 2.

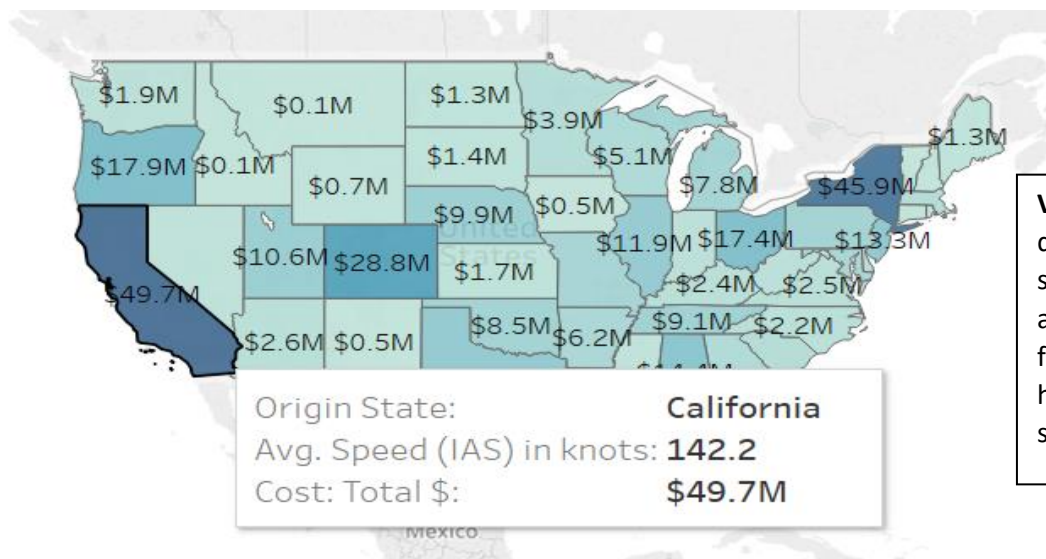


**Visual 1:** The visual created by following the document allowed us to simply understand the total relative cost associated with bird strikes at different states. Bigger circle stating higher cost compared to smaller circle stating relatively lower cost.



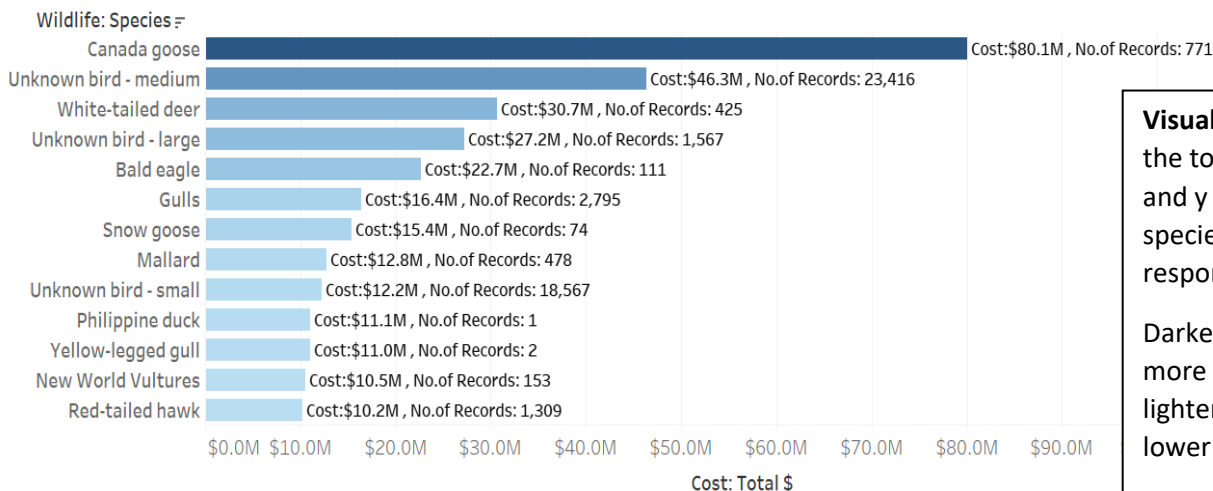
**Visual 2:** This is an update over above visual that gives more information related to the total bird strike damage cost associated with different states. Darker color signifies higher cost compared to lighter color stating relatively lower cost. And this visual tells us directly that California has the highest damage cost.

- **Average speed of airplanes** in California that were involved in bird strikes is **142.2 knots**.



**Visual 3:** The visual displays the total damage cost associated with each origin state. And hovering over each state gives additional information of average speed for that state. Darker color signifies higher cost compared to lighter color stating relatively lower cost.

- **Canada goose** has caused the **maximum amount of damage** in terms of cost worth over **\$80 million** and there are total 771 entries from the same species.



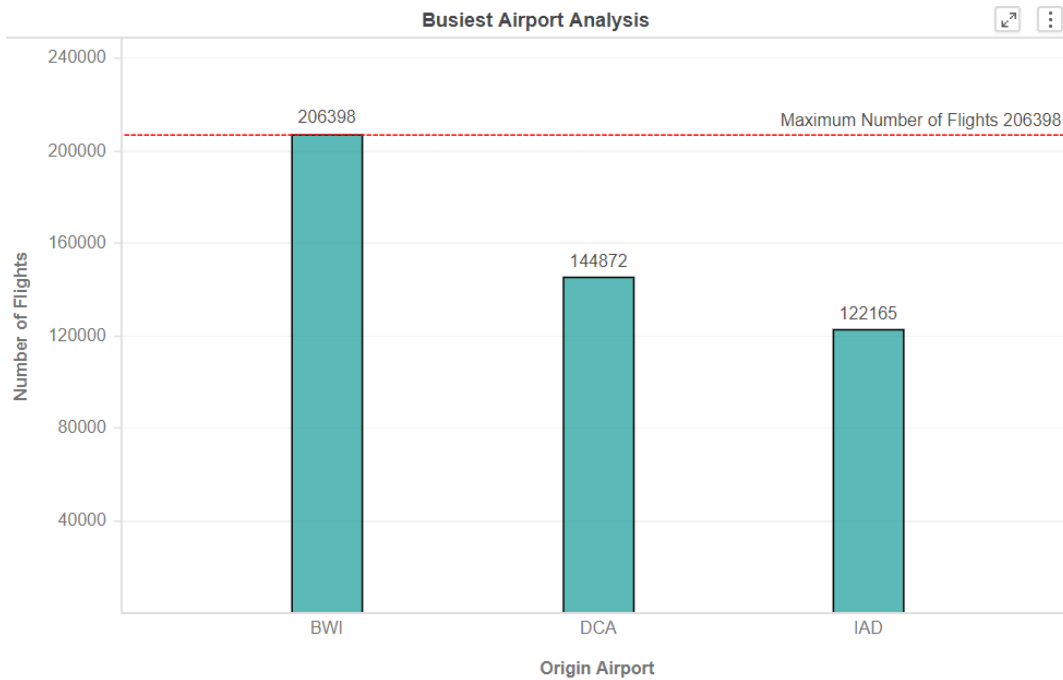
**Visual 4:** X axis represents the total cost of damage and y axis represents the species that were responsible behind it.

Darker color represents more associated cost and lighter color is for relatively lower associated cost.

Every bar label gives total cost and total number of instances on record.

## MicroStrategy:

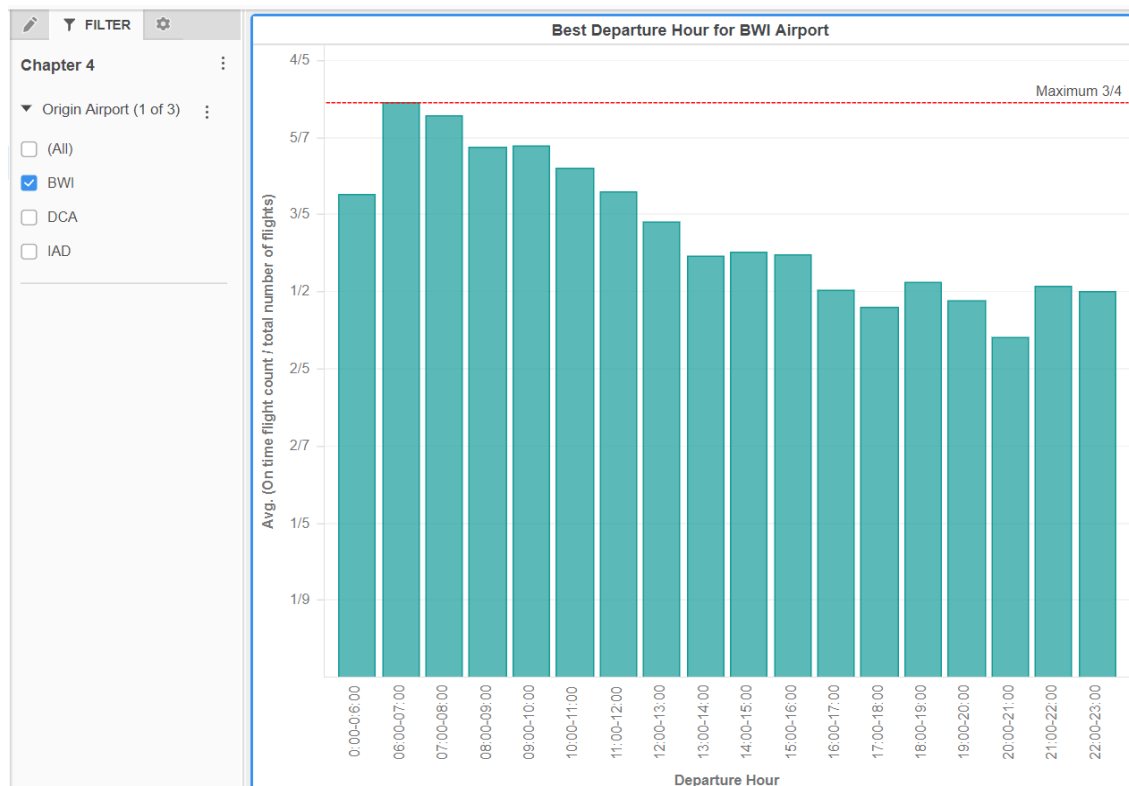
- **Baltimore Washington International airport is the busiest airport in Washington DC area** having over 200k number of flights in year 2009-11.



**Visual 5:** Values on x axis represents different airports in Washington and y axis represents the total number of flights for those airports respectively.

The red reference line marks the airports with highest number of flights and that's BWI.

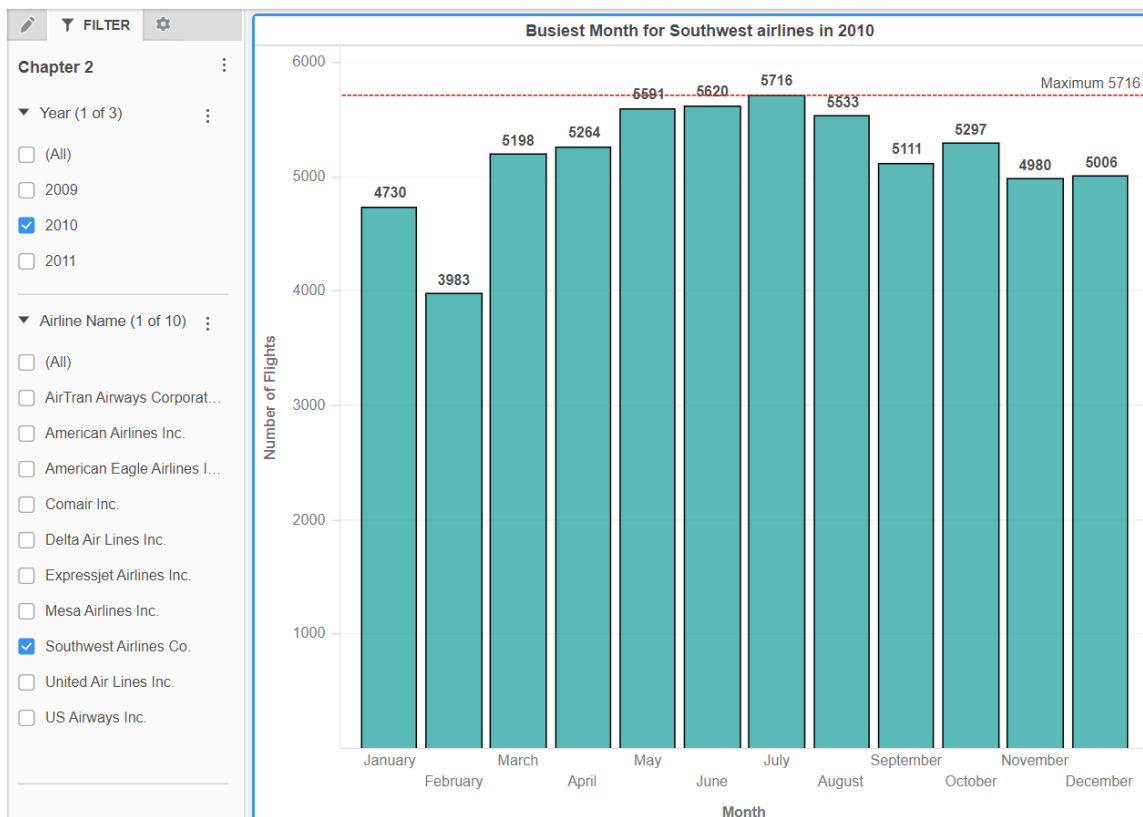
- **Best hour for departure at Baltimore Washington International airport is 6-7 am** where on an average 3 out of 4 flights are on time.



**Visual 6:** Values on x axis represent the hours and y axis represents the avg. count of on time flights for that respective hour. On left the filter shows that only BWI airport was considered for analysis.

As seen from the chart, the red reference lines show the maximum number of on time flights as a proportion of total number of flights. And it is the highest for time period from 6am to 7am at BWI airport.

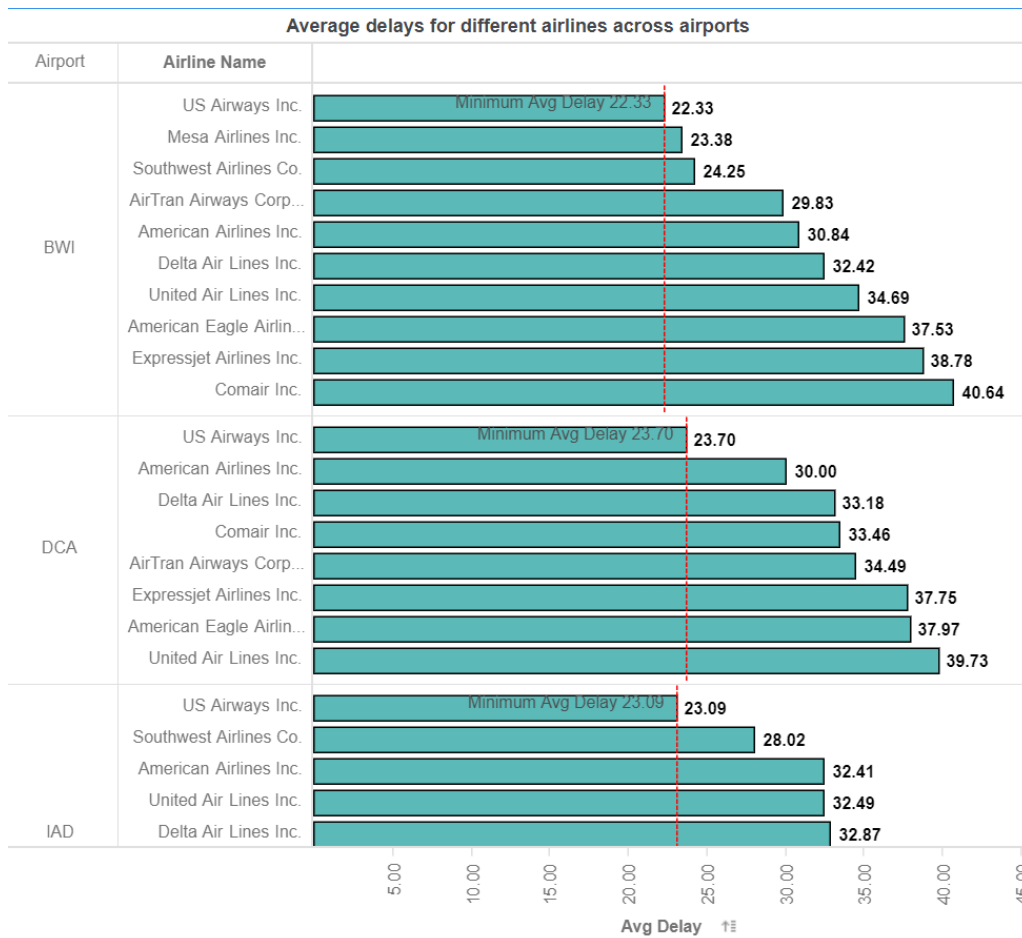
- The busiest month for southwest airlines in 2010 was July with 5716 number of flights.



**Visual 7:** Values on x axis represent month and y axis represents total number of flights for that month. On left the filter shows that only southwest airlines and year 2010 was considered for analysis.

As seen from the chart, the red reference line marks the maximum total number of flights.

- The average delay for USA airways airlines is minimum across all the airports in Washington.



**Visual 8:** Values on x axis represents average time delay and y axis represents airports drilled down to individual airlines.

As seen from the chart, the red reference lines mark the minimum average delays across different airlines at all the airports.

### Question 3 : What did you learn?

- After creating the first visualization I realized that the information conveyed by it could have been achieved by a small code in SQL or python but the visuals made in Tableau or MicroStrategy turns your alphanumeric results into beautiful visualization. This was the **aha moment** for me.
- I learnt that the visuals created can be exported easily as an image to be used in presentations or documentations. And they also allowed **real time report or dashboard sharing** with the team which I found fascinating and an essential BI tool.
- I learnt that **formatting the variable** to right format makes it more interpretable. As seen above in visual 2, the change of format for total cost from number to currency in millions made it more interpretable and concise.
- Any dataset has two parts, they are either **categorical** in form which are then used for **grouping** or **numerical** which are used for **calculations**. In tableau the categorical variables are known as **dimensions** and in MicroStrategy it's called attributes. However, both software store numerical variables as **metrics**.
- I found **tableau outputs to be visually more appealing than MicroStrategy's**. However, I found MicroStrategy's feature of data wrangling to be more powerful than that of Tableau's. On doing further research I realized that tableau recently launched a new product specially for data preparation called as tableau prep.
- I also realized that same information can be conveyed through different visuals and choosing the **right visual** shall be more beneficial for correct data interpretation.
- I also learnt the concept of **aggregation** and **drill down** while creating a MicroStrategy visual where the airports were drilled down to individual airlines.

### Question 4: How does it relate to class?

- From Night 3, presentation data warehouse and mining intro, Slide 10 helped me understand the mining process and slides 12-18 gave me good idea on concept of aggregation and drill downs. Slides 32-35 from chapter 3 helped me understand the concepts of OLAP, OLTP, aggregation and drill downs even more. These concepts of aggregations and drill downs were applied while creating the last visual in MicroStrategy and general data mining process was followed while creating both the visuals.
- Slide 42-45 from chapter 3 also made me understand about the concept of real time BI and DW. In our case we had a static data source but in real life our dataset would mainly be on cloud and we would require a real time BI/DW.
- On night 4, the slides 2-5 and 24 of sharda\_dss10\_ppt\_04 taught me about business reporting and how to make a successful report. From this I thought of adding data labels and formatting in visuals created above. Slide 13 from the same ppt also taught me different types of graphs and charts that are possible.
- Slides 22-33 from DSS\_User\_Interface\_Dialogue\_and\_Visualization\_Design helped me understand how to make a right visual for the given audience.

### References:

- <https://www.teradatauniversitynetwork.com>
- <https://www.tableau.com/learn>
- <https://www.microstrategy.com/us/desktop/help#tutorial-videos>