Data Visualization- Group Project

Annual Flight Data Analysis

(Delays & Carrier Performance across airports in US)

Submitted By: Group 1

Aarya

Rounak Burad

Meng-Lun Li

Gayathri Muraleedharan

**Table of Contents**

Introduction 3

About the data 3

SAS Analysis 4

Visualization using Tableau 5

Reasons for Flight Delay 5

Carrier Delay Across Months 6

Analysis of Flight Delays over Days of the Week 7

Delays Occurring on Days of the Month 8

Average Delay Across States 9

Departure Delays from Airport of Origin 10

Arrival Delays at the Destination Airport 11

Flight Carrier performance 12

Twitter Data for each Airline 13

Sentiment Analysis of Tweets 14

Analyzing Reasons for Negative Tweets by visualizing with Tree Maps 15

Tableau Story and Rhetoric 16

Conclusion 17

References 18

**Introduction**

Air travel and the decision making on when to book, which carrier to choose, how to plan the journey etc. is a cumbersome process with the increased number of carriers and umpteen agencies who come up with a variety of customized offers. While planning for the same we tend to forget a very important factor, delays, which costs us the most valuable element, our time. Calculation and prediction of delays is near to impossible. Through the analysis of a flight data set, we have tried to come up with a and precise model of delay prediction to choose a proper carrier and an appropriate time for travel. We based the report on previous data and utilizing visualization tools to illustrate our concept and analyzing the result.

**About the data:**

* Primary Data

Flight data for the calendar year 2016 was considered for analysis. Fields under analysis include Origin and Destination cities of flights by different carriers.

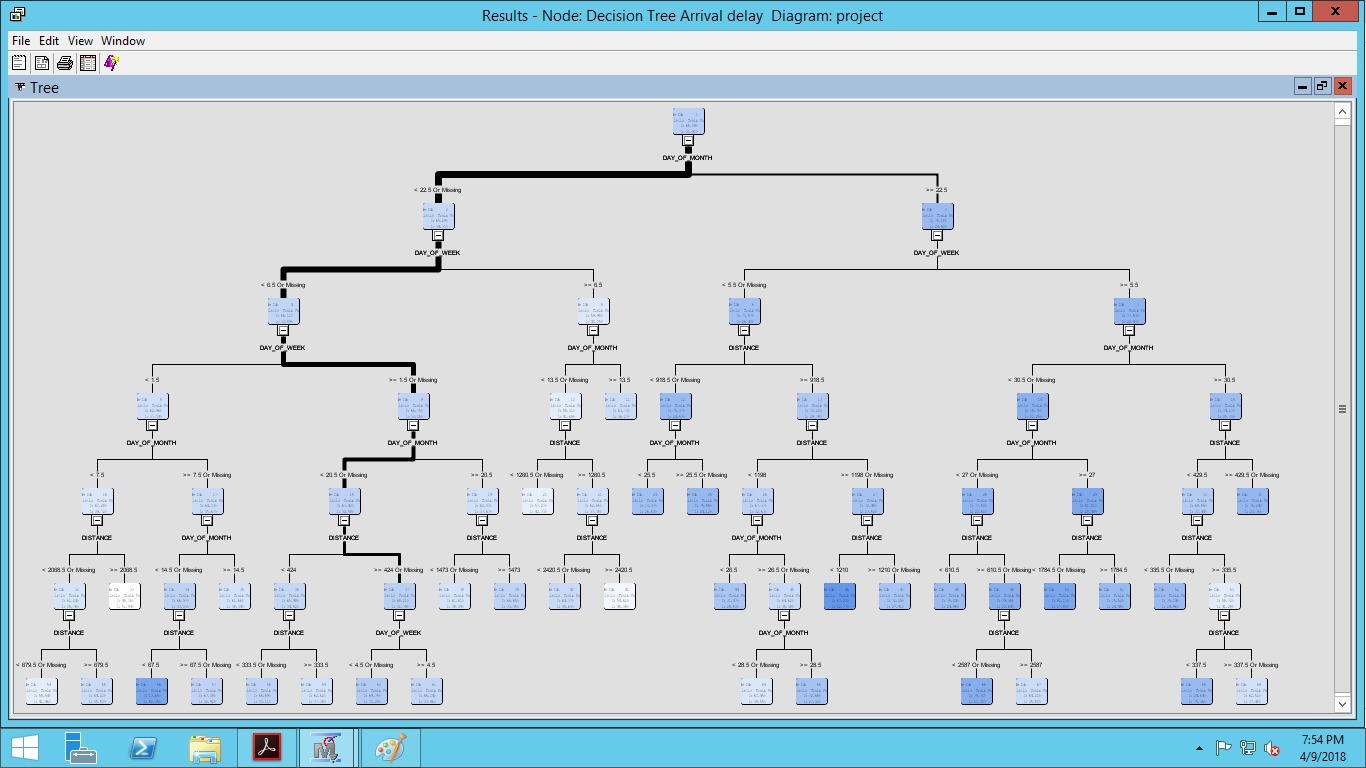
Fields added: Total delay for a flight was calculated by adding the arrival delay and departure delay for each row.

* Secondary Data

Social Media is a very powerful factor which influences every decision in a common man’s life today. We have added twitter data to perform a sentiment analysis about the different carriers.

**SAS Analysis:**

A SAS predictive analysis was done on the data to form a predictive model which could predict delays of different flights as shown.



The above decision tree predicts the chance of a flight getting delayed by analyzing various possibilities. The analysis further gives us model accuracy to tell us how correctly we could predict the flight delay. The model, however cannot be considered as a good visualization for a layman. For a non-technical person to understand the above analysis and several other similar analyses, we would have to keep a separate presentation to first explain to them terminologies like decision tree, p-value, event classification tables etc. And even after that they might still get confused. With the help of Tableau, the following analyses have been done to give better insights for the data.

**Visualization using Tableau:**

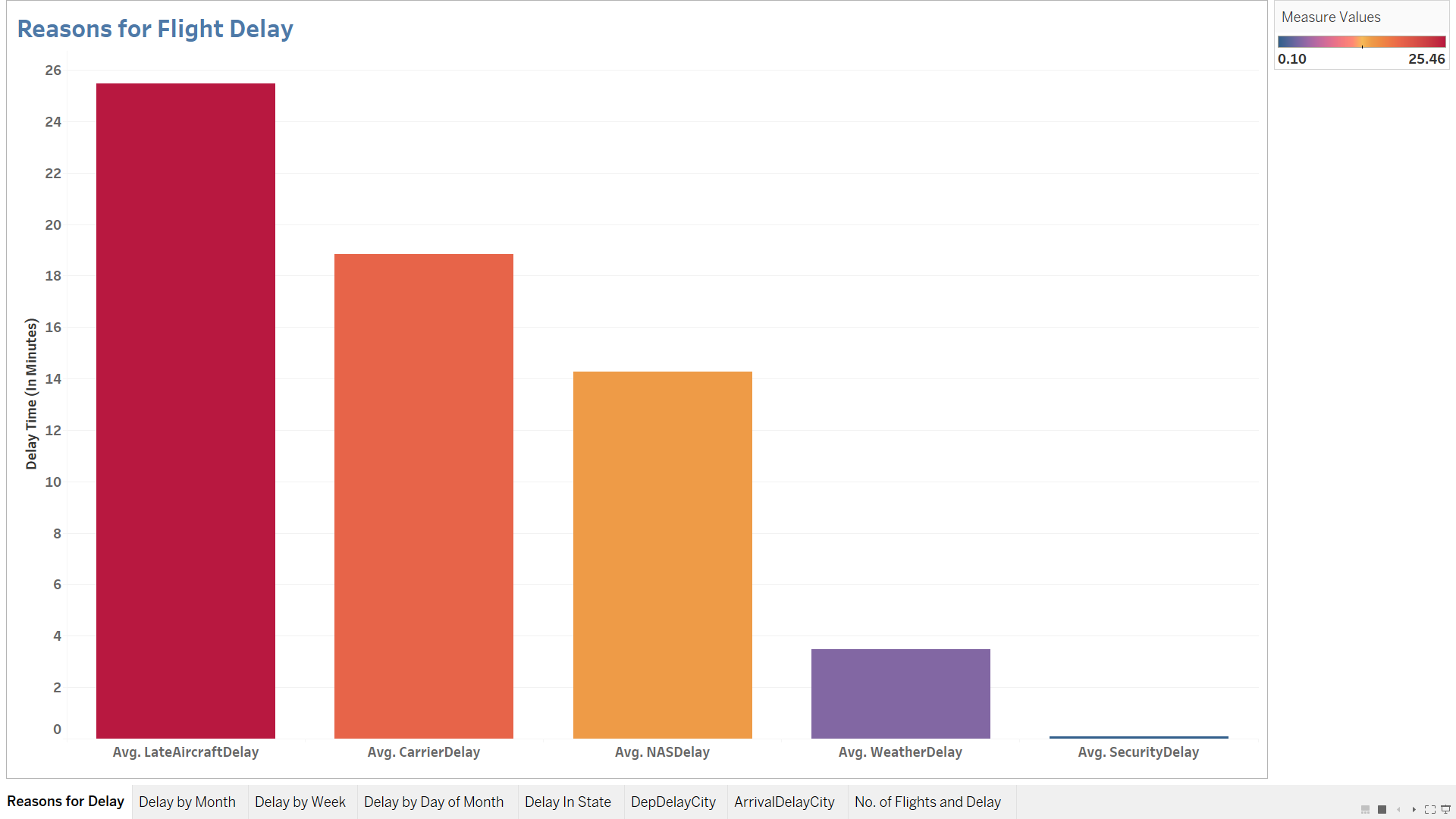
The following visualizations were done using the data.

* **Reasons for Flight Delay**

Flight delays were analyzed for different reasons.

* + Aircraft Delay
  + Average Carrier Delay
  + Average NAS (National Aviation System) Delay
  + Average Weather Delay
  + Average Security Delay

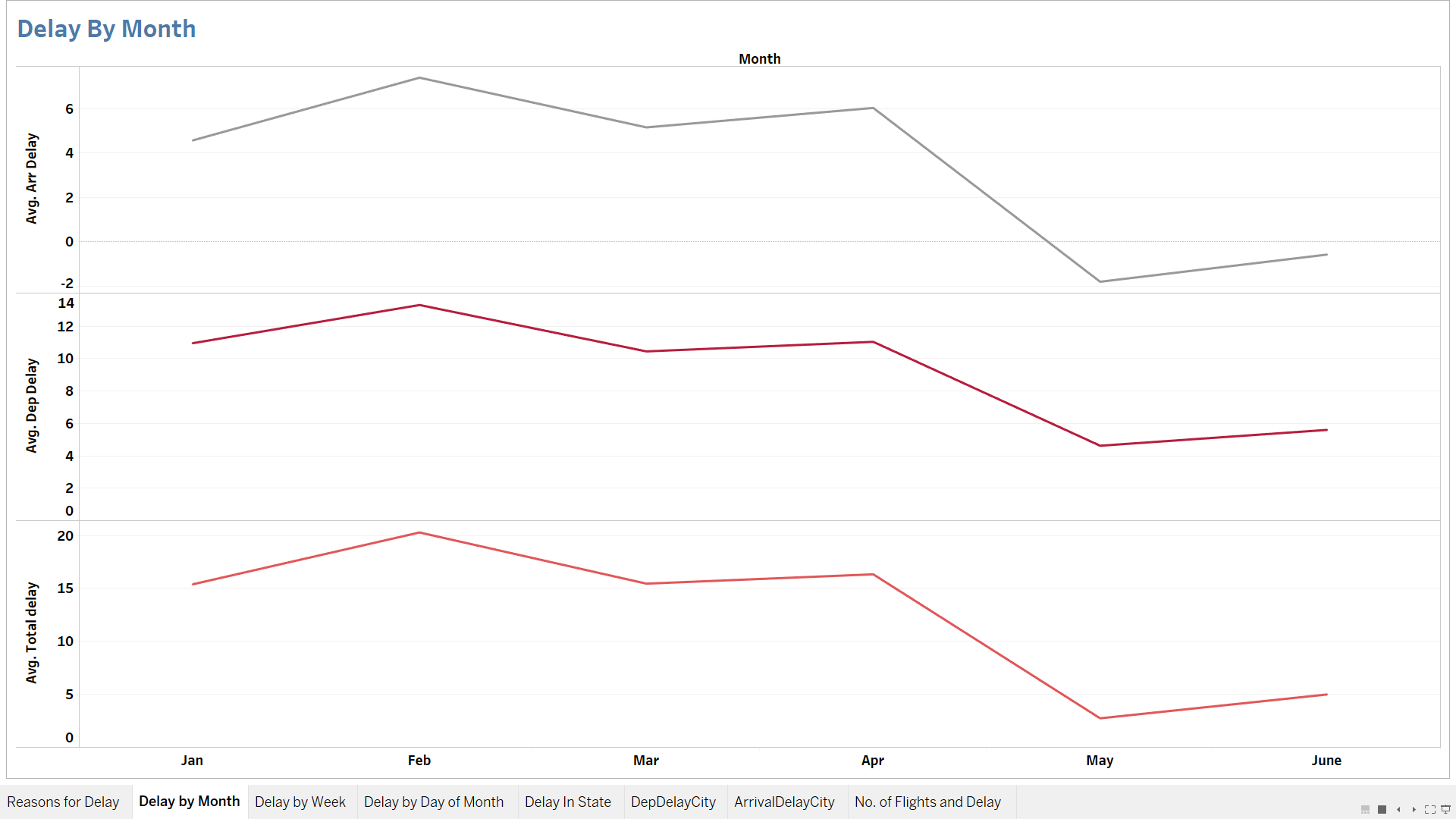
The delays due to the above reasons were plotted on a bar graph to analyze the reason which caused the delays and to find the gravity to which the reason affected the delays.



It is clear that maximum average delay was 25 minutes caused by aircraft delays. Security delays whenever happened, was for an average span of less than a minute.

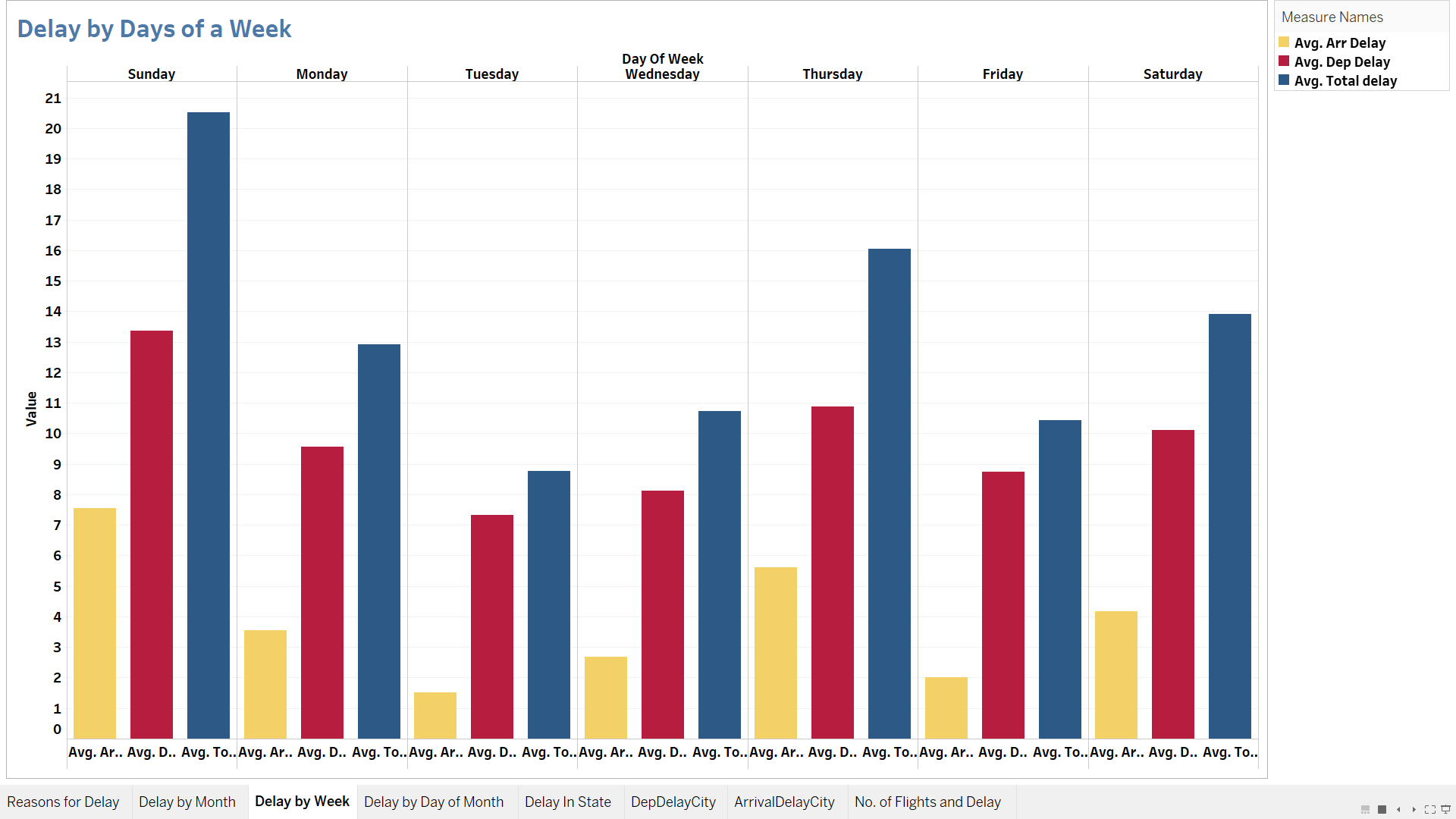
* **Carrier Delay Across Months**

Average flight delays across months of the year was taken into consideration next. The average arrival, departure and total delays across months were analyzed and plotted as line graphs. The arrival and departure delays followed the same pattern depicting that the departure delays were not always covered up during the flight. February marked maximum delays while May had the least delays in the first half of the year.



* **Analysis of Flight Delays over Days of the Week**

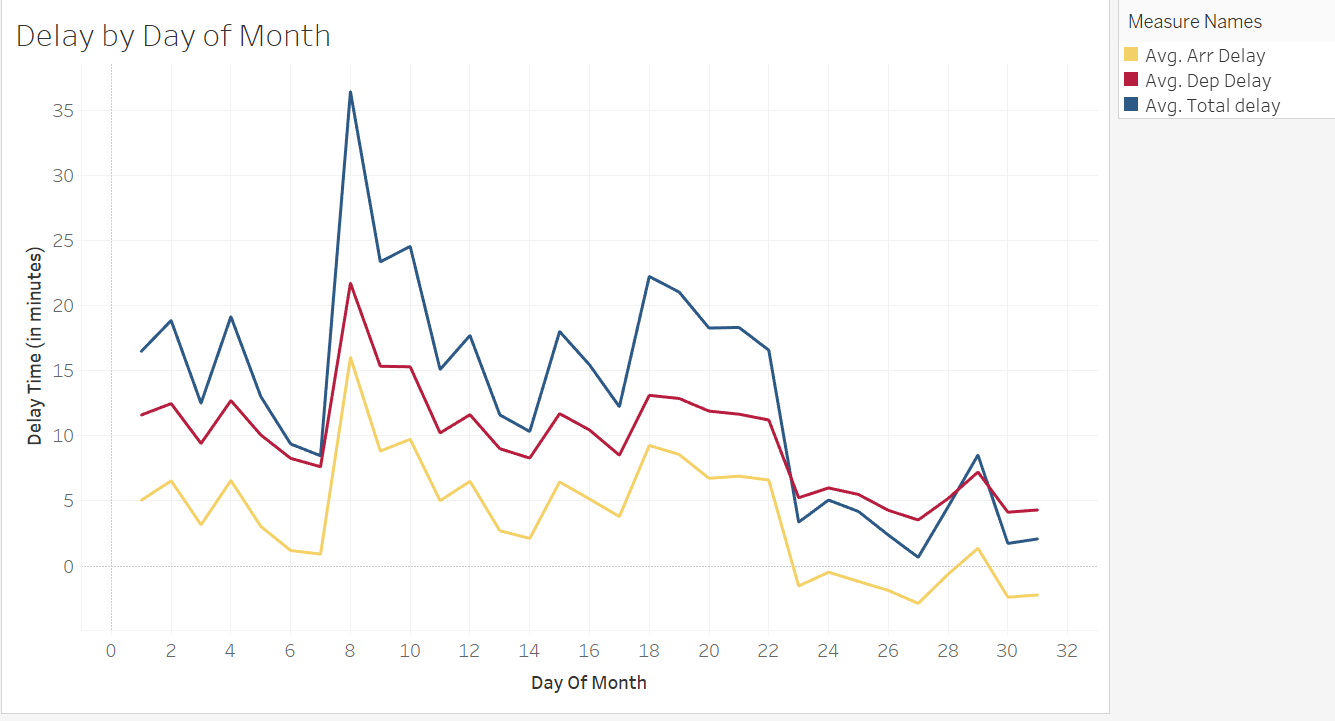
Average flight delays varied across days of the week. The maximum average total, arrival and departure delays were on Sundays. Arrival delays were least on Tuesdays and departure delays were least on Tuesdays followed by Wednesdays and Mondays with a very small difference.



The delays on Sunday might be caused by the end of weekend or beginning of the week. Tuesday delays were the lowest probably being the mid of the week.

* **Delays Occurring on Days of the Month**

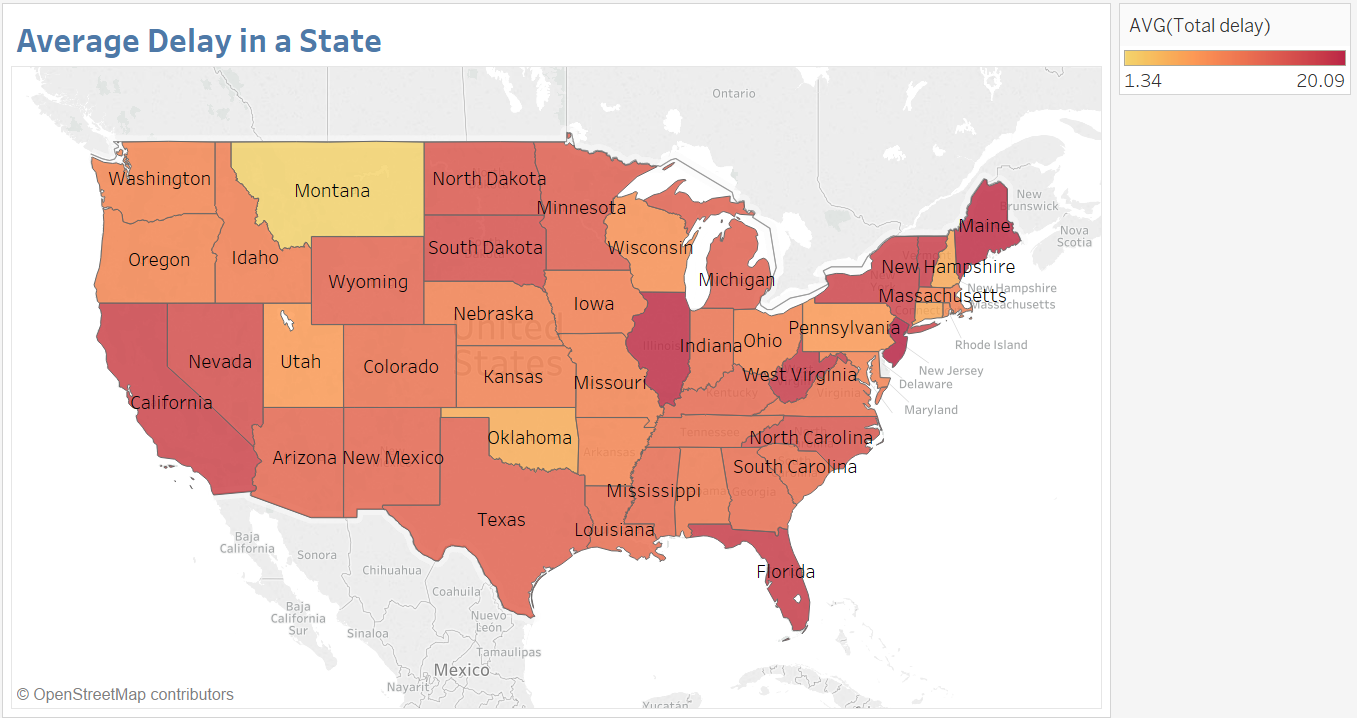
While considering the delays across the days of the month, maximum average delays were at its peak on the 8th of the month, on the contrary to the expectation that delays would be high because of the crowd at the beginning or end of each month



The least delays were observed in the last week of the month which could be due to financial constraints, being the end of the month, or maybe because travel plans happen at the beginning of the month and the general tendency is to travel at the beginning of the month. Arrival delays show negative values over this week indicating that flights were on or before time which maybe because of lesser crowd at the airport.

* **Average Delay Across States**

The average delays varied geographically across states. In 2016, Montana marked the least average delay among all the states while New Jersey had the maximum delay.



The variation in delay geographically can be attributed to the demographics of the state. Montana, a state with one of the least population densities across US, has the lowest delays because of lesser crowd. The map shows that the highest delays are in the states which have higher population densities like New Jersey, Maine etc.

* **Departure Delays from Airport of Origin**

Analysis on the average departure delays based on airport location shows that following cities have the maximum delay

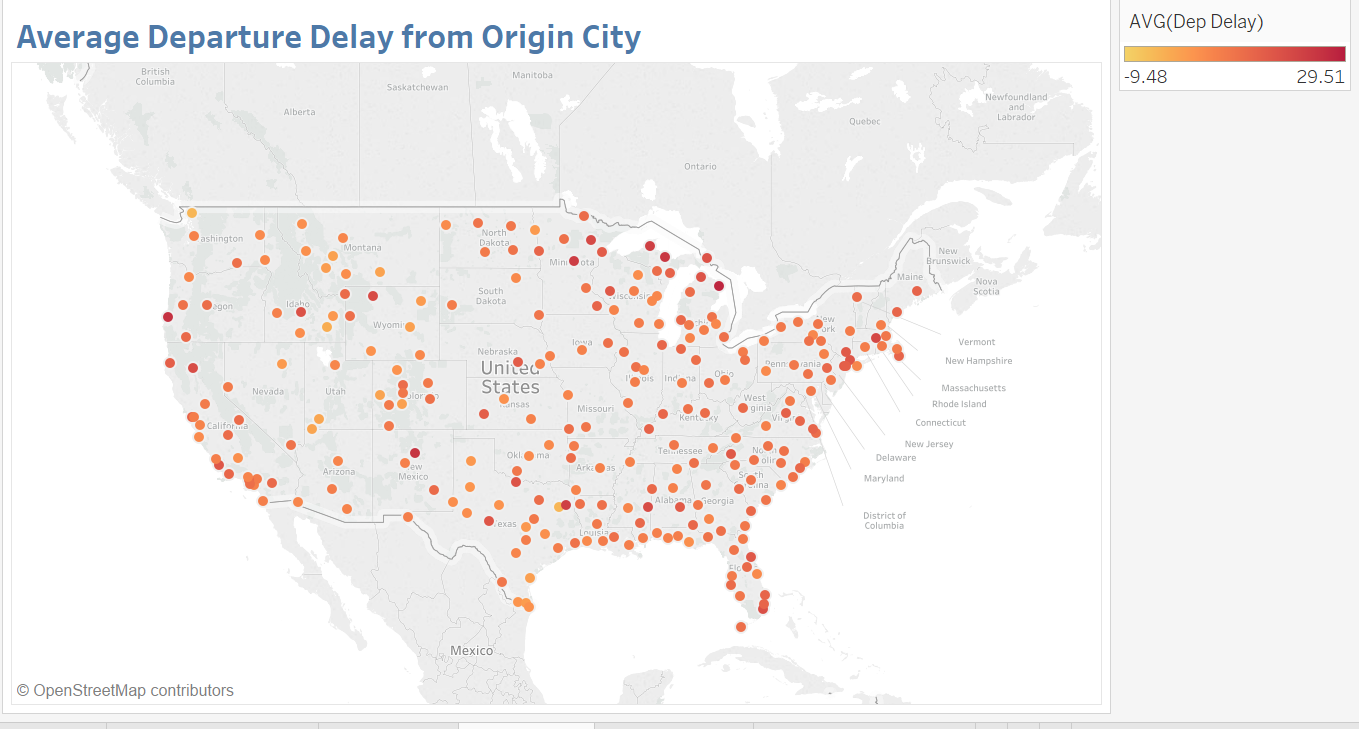
1.Chicago

2. Alpena

3. Santa Fe

4. North Bend

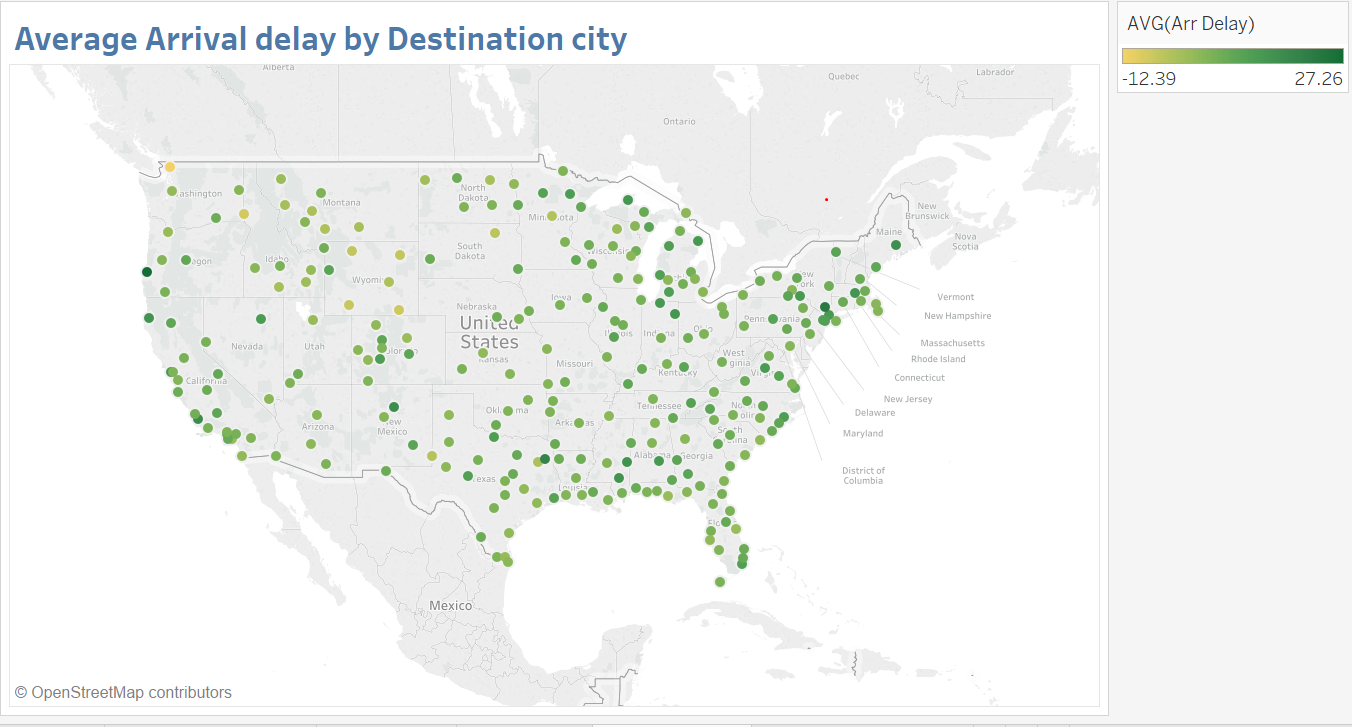
5. Longview



* **Arrival Delays at the Destination Airport**

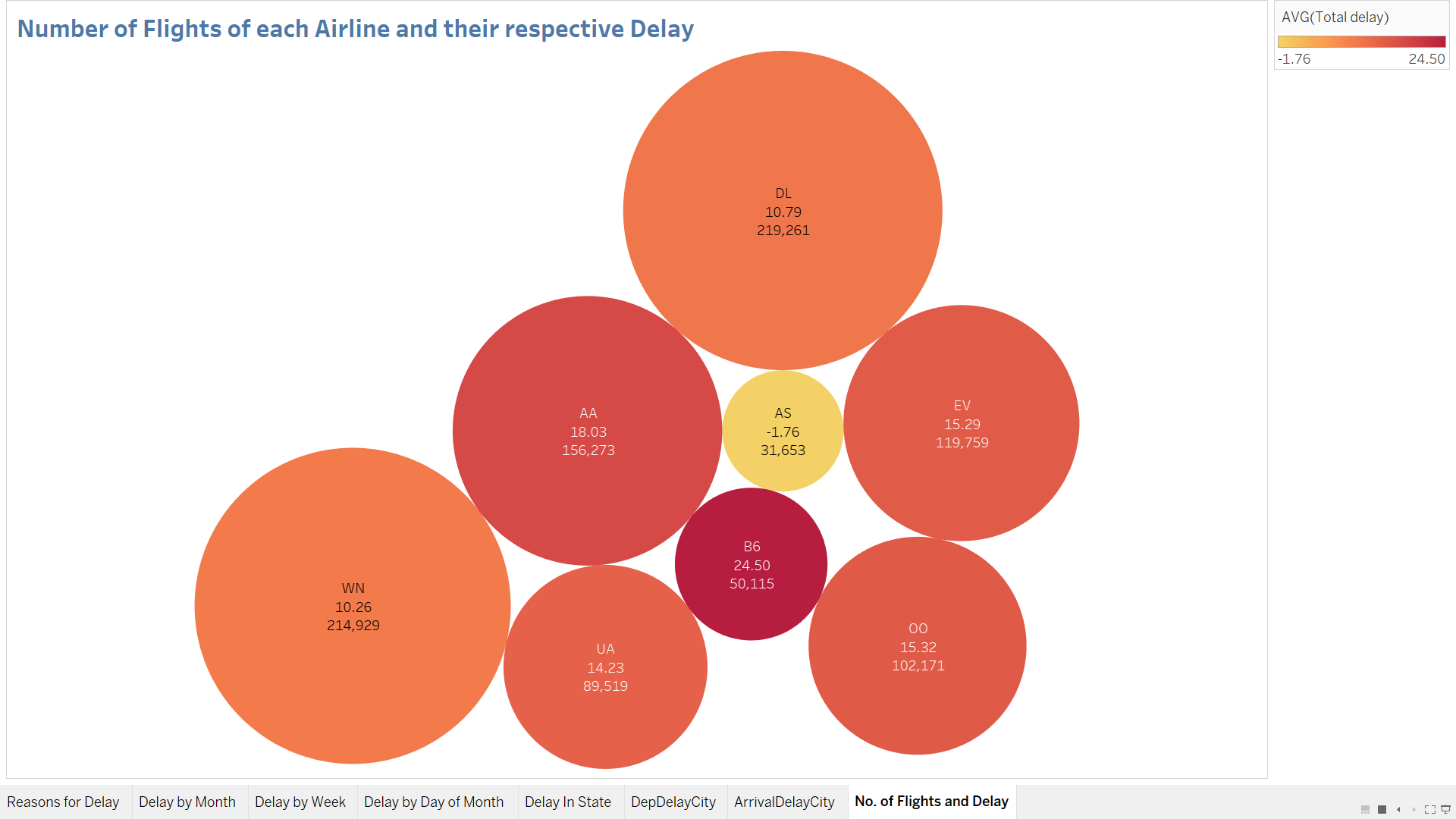
For arrival delays, following destination cities have maximum delay:

1. North Bend
2. Long view
3. Santa Maria
4. New Burgh



* **Flight Carrier performance**

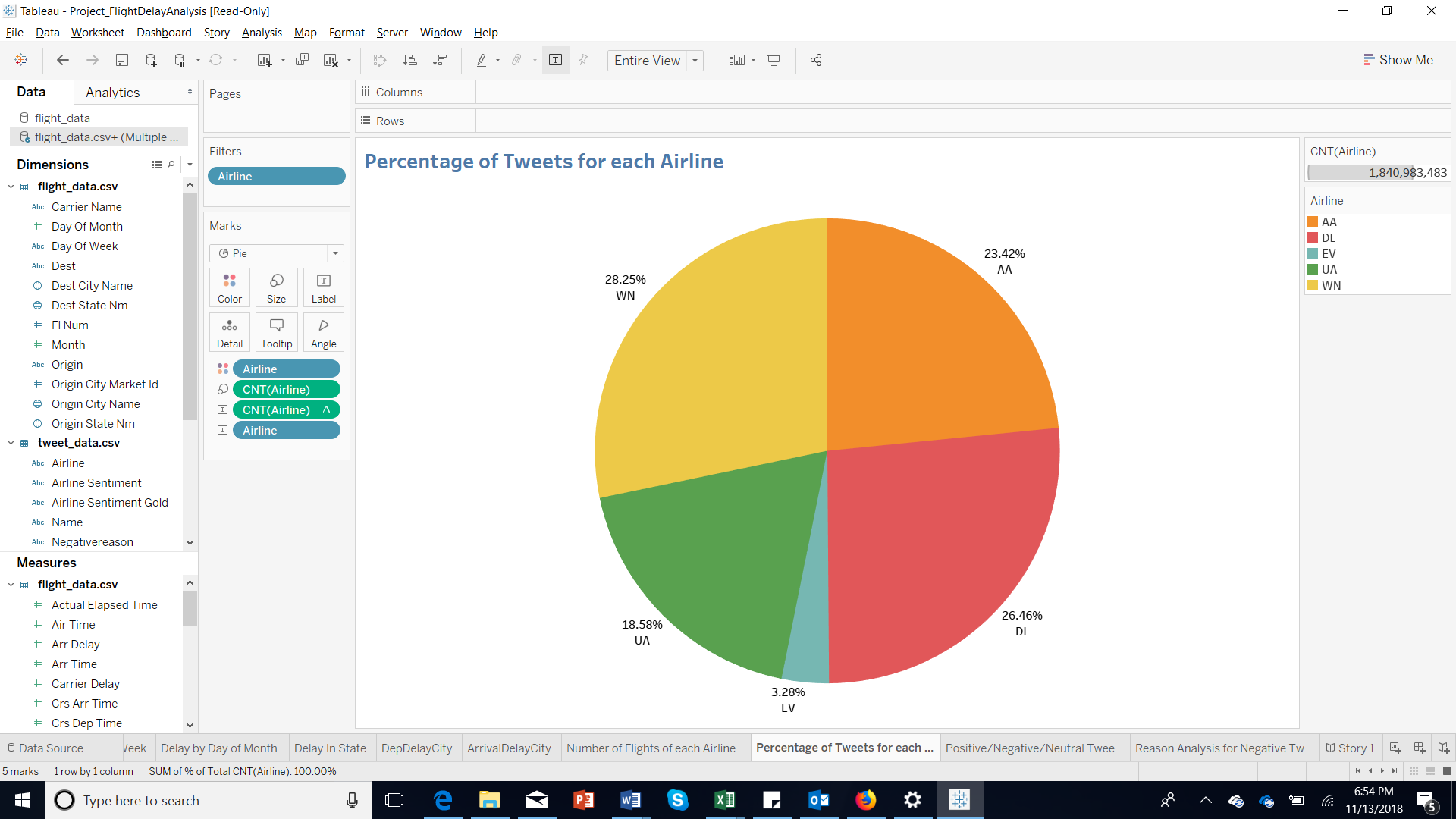
Delays varied for different service providers. Alaska Airlines have a negative value for average delay indicating they are before time. However, the number of flights made by Alaska Airlines is the lowest compared to all the other service providers across the year.



The number of flights made by Delta Airlines is the highest in a year. But they have an average delay of around 11 minutes. JetBlue has the maximum delays even with one of the smallest number of flights.

* **Twitter Data for each Airline**

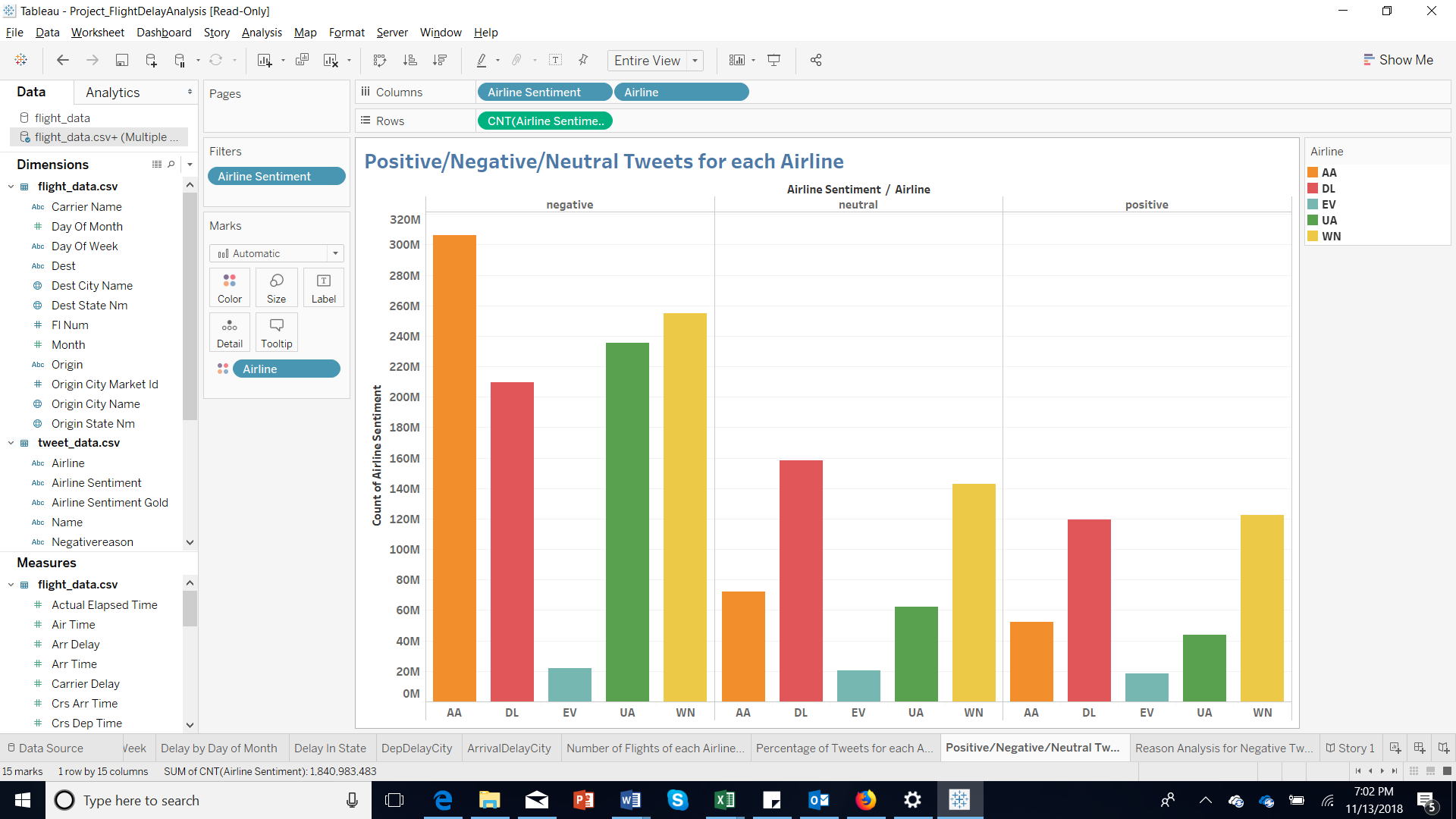
We got another “Twitter US Airline Sentiment” dataset from Kaggle which had various sentiment analysis data based on positive, negative and neutral feedbacks that customers have provided for a particular airline service.



The above pie chart shows how prominent each carrier is on twitter. We see that Southwest(WN) is the airline which is most talked about on a social platform while only less than 4% of the tweets are about SouthWest Airlines.

* **Sentiment Analysis of Tweets**

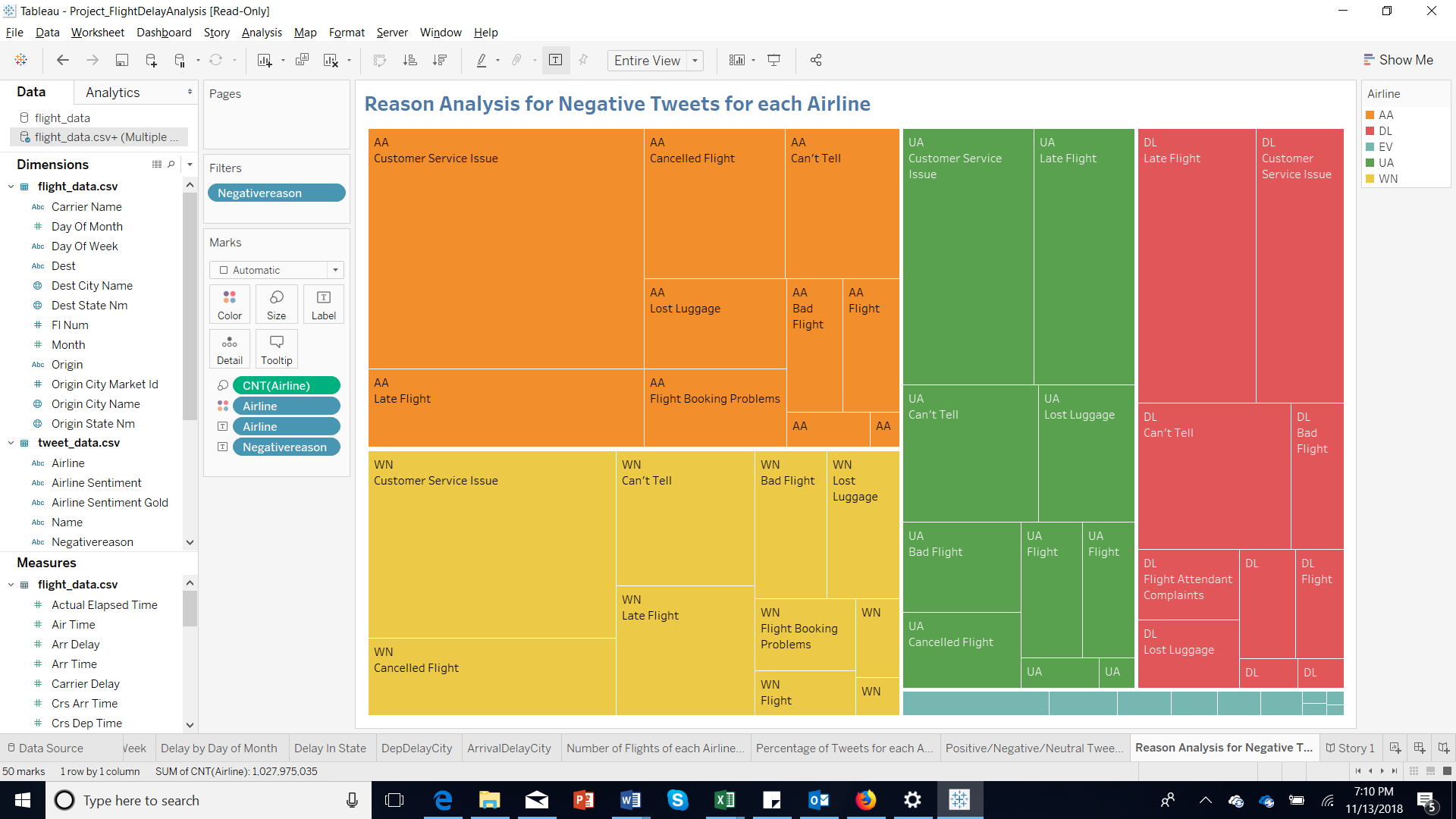
The above visualization showed the most discussed airlines. But that did not give us a clear picture on whether they were popular in a positive way or a negative way. To analyze this, sentiment analysis was done, and the nature of the tweets were classified as positive, negative and neutral tweets.



The above clustered bar graph shows that American Airlines were tweeted the most in a negative way, followed by Southwest; while these two also scores the most number of positive tweets. Delta Airlines received the maximum number of neutral tweets.

* **Analyzing Reasons for Negative Tweets by visualizing with Tree Maps**

The next step involves the analysis of the negative tweets which each airline received.



For American Airlines, SouthWest and United Airlines, the most number of negative tweets were about customer service issues indicating they are not satisfying the customer expectations in terms of service. The second major reason was late flight or cancelled flight which indicate that customers are definitely worried about delays. Delta Airlines, though the customer service seems better than the other 2 competitors, are accused of being late according to tweets.

* **Tableau Story and Rhetoric**

From the above visualizations, we have built a story in tableau using logical rhetoric which uses reasoning to make a case. Here we have used reasoning and data as proof to come up with a conclusion.

The initial part says that maximum delays were caused by aircraft delays. Delays were more in February when months were considered and on Sundays as far as days of the week were concerned. Origin and destination location also caused delays. Carrier options also influenced delays to a large extent. To know the extent to which this affected the customers, a twitter analysis was done. Sentiment analysis was used to know the attitude of the customers to the airlines and the negative reviews were further grouped to understand the reasons.

* **Conclusion**

According to the analysis,3 major factors should be taken into consideration while choosing a flight - time/date, location, and airline. For optimizing our flight experiences, the following factors, if kept in mind, can save a lot of time and money:

* avoiding peak period like vacation season
* attempting fling at off peak day
* considering your destination

For example, big cities like Chicago or Los Angeles are not good choices for flight transfers but heading to these sites would be better during off season on weekday.

Our inferences are listed below:

* May would have least delay among months
* Sunday would be the most delayed day and Tuesday is the least among weekday
* 8th is the most delayed date in a month
* Population density is one factor of affecting flight density in one city
* Delta Airline has least performance considering punctuality and number of flights.
* Customers expect good service and punctuality
* **References**
* <https://www.kaggle.com/niranjan0272/us-flight-delay/>
* <http://mrpassenger.com/dont-sweat-it-patience-is-key-during-flight-delays/>
* <https://mashable.com/2014/12/10/cost-of-delayed-flights/>