THE UNIVERSITY OF SYDNEY

School of Information Technologies

Faculty of Engineering & IT

ASSIGNMENT/PROJECT COVERSHEET - GROUP ASSESSMENT

Unit of Study: **COMP5048 – Visual Analytics**

Assignment name: **Assignment 2 –Final Report**

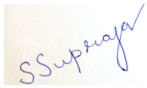
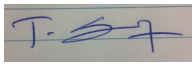
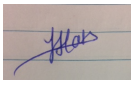
Tutorial time: **8pm** **Tutor name:** **Dr Quan Nguyen**

DECLARATION

We the undersigned declare that we have read and understood the *University of Sydney Academic Dishonesty and Plagiarism in Coursework Policy*, an, and except where specifically acknowledged, the work contained in this assignment/project is our own work, and has not been copied from other sources or been previously submitted for award or assessment.

We understand that failure to comply with the *Academic Dishonesty and Plagiarism in Coursework Policy* can lead to severe penalties as outlined under Chapter 8 of the *University of Sydney By-Law 1999* (as amended). These penalties may be imposed in cases where any significant portion of my submitted work has been copied without proper acknowledgement from other sources, including published works, the internet, existing programs, the work of other students, or work previously submitted for other awards or assessments.

We realise that we may be asked to identify those portions of the work contributed by each of us and required to demonstrate our individual knowledge of the relevant material by answering oral questions or by undertaking supplementary work, either written or in the laboratory, in order to arrive at the final assessment mark.

	Project team members			
Student name	**Student ID**	**Participated**	**Agree to Share**	**Signature**
1. Ruchita Manuja	460494470	20%	Yes	
2. Supraja Sridharan	460474472	20%	Yes	
3. Anirudh Sharma	460482684	20%	Yes	
4.Sundaram Thangaraj	470531015	20%	Yes	
5. Abhijeet Date	470299528	20%	Yes	

SIT Building, J12
The University of Sydney
NSW 2006 Australia

T +61 2 9351 3423
F +61 2 9351 3838
E sit.info@sydney.edu.au

ABN 15 211 513 464

CRICOS 00026A

FINAL REPORT

Group 14

Ruchita Manuja - 460494470

Supraja Sridharan - 460474472

Anirudh Sharma - 460482684

Sundaram Thangaraj – 470531015

Abhijeet Date - 470299528

Table of Contents

Introduction

Everybody seeks for the best experience when it comes to travelling on flights. “My flight was pleasant and on time” is a statement everybody wants to share with others. But this is not always the case, as there lie twists, tricks, and surprises that change a traveller's thought and manipulate their views for a certain airline merely based on an unpleasant experience such as delays. This report explains our analysis on the US airlines data and provides an insight to the visualization system created. Through this visualization system assist the travellers can make the best decisions on their next travel in terms of choosing the most reliable airline, airport or travel times.

Data Set

Airlines dataset comes from the Research and Innovative Technology Administration (RITA), from the United States Department of Transportation. Data is from 1987 to 2008 and consists of figures about the arrival and departure details for the commercial flights throughout the 3,376 airports in the USA. The data is split down into 22 yearly chunk files, year by year. The dataset is 12GB with nearly about 120 million records.

Additional data requirement

The dataset needed to be combined with the supplementary datasets provided to complete the analysis and visualization.

- **Airport data** - Required to get airport names, city, longitude, and latitude of the airports
- **Carrier data** - Required to get Airline's name

Data Cleansing

Though the data set was already formatted, it still required additional data cleansing activities such as:

- Date is split into 3 columns (year month, day of month), it had to be combined
- Some columns like Arrival delay and departure delay contained value as 'NA' wherever not applicable. This needed to be converted to zero before performing any arithmetic operation
- Some of the airport names contain ',' in their names which was replaced with spaces

Aims and Contribution

We have performed analysis and produced visualizations on the Airlines data in order to identify:

- Usual and unusual patterns in the data
- Best Month/day of the travel to avoid delay/cancellation
- Best and Worst Airlines
- Best and Worst Airports
- Identify reasons for Cancellations
- Analysis of Airport Routes based on the best and Worst Airports
- Factors that contributed to the airline’s delay/cancellation such as 9/11 attack, Global financial crisis, weather and public holidays

Design and Approaches

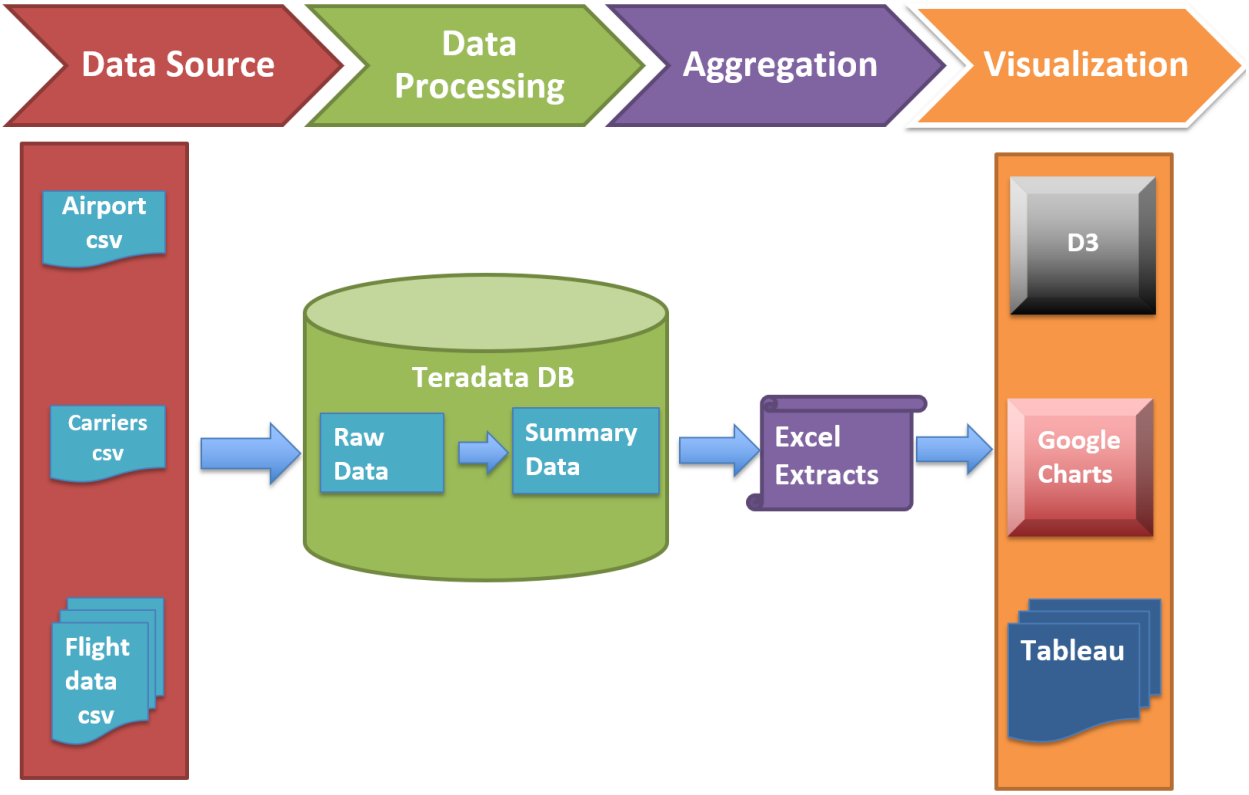


Figure 1 - System Overview [Created by Group 14 for COMP5048]

Data Extraction

The following data was extracted from http://stat-computing.org/dataexpo/2009/.

- Flights data (year wise)
- Carrier data
- Airports
- Planes

Data Loading

Data from CSV files was loaded into Teradata database using load scripts. Teradata was selected for data storage since Teradata load scripts are efficient for loading and processing large volume of data. Once the files were loaded into Teradata tables, they were validated for missing or invalid values.

Aggregation

Once the data was cleansed of missing/invalid data, it was aggregated and validated using SQL. The aggregated results were extracted into CSV files for creating visualizations.

Analysis

Following analysis techniques were performed.

Filter the Data:

As the data set is fairly large it can be categorized under Big Data. We filtered the records in order to focus on a smaller subset of data as below:

- Our main focus was to analyse the data from the 21st century.
- We used various filters such as Airports, Carrier and routes.
- Flights actual departure or arrival time exceeded by 15 minutes from the scheduled departure or arrival we flagged those flights as delayed.

Clustering the data

The flight network could be considered as a directed graph with multiple nodes as airports and edges as flight connections among the nodes. So, a “Graph Clustering” method has been used to identify the clusters in the flight network.

A general approach on this flight network is to discover the groups of airports based on connections between the nodes in the network. The idea is to form clusters of airports that have more connections (and count of flights) to one another than they do to outsiders.

Centrality

This technique was used to identify important nodes (airports or carriers) and crucial links as below:

1. Degree centrality

- in-degree centrality -> total incoming traffic on a day
- out-degree centrality -> total outgoing traffic on a day

2. Betweenness centrality -> to identify proportion of shortest path between 2 nodes

Visualisation

Visualisation 1 - Total flights vs Cancellation

A Continuous Line Chart for analysis of Total flights with Cancellation has been created to identify the factors affecting the Performance of all the airlines. The charts for the years 1998 to 2008 has been combined as a video to identify the usual and unusual patterns.

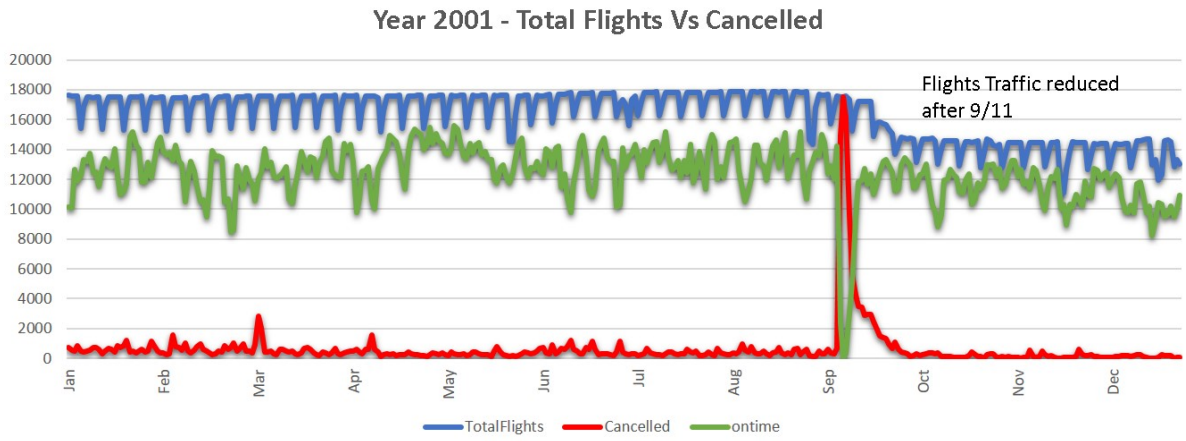


Figure 2 - Visualisation 1 a - Total flights Vs Cancellation for Year 2001[Created by Group 14 for COMP5048]

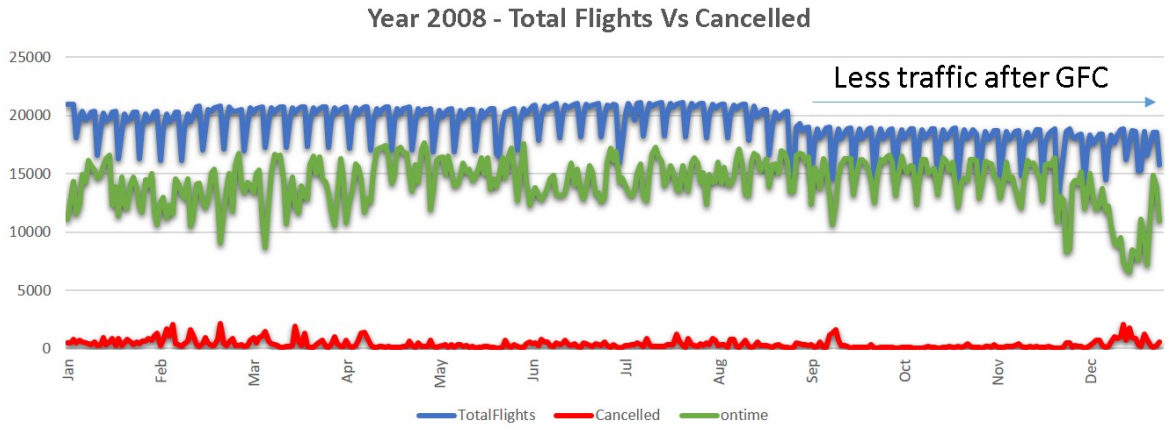


Figure 3 - Visualisation 1 b - Total Flights Vs Cancellation for Year 2008[Created by Group 14 for COMP5048]

Patterns Identified:

- Carriers operate less flights on Saturdays
- Less flights observed on 27th November which is thanksgiving holiday
- Airlines traffic reduced after 9/11 attack in 2001
- Airlines traffic reduced after Global Financial Crisis in 2008

Visualisation 2– Carrier Analysis

A ***bar*** visualisation for the carrier analysis has been created to identify Top and Bottom Carriers based on percentage of On Time, Delay and Cancelled from the Visualisation 2 a and 2 b

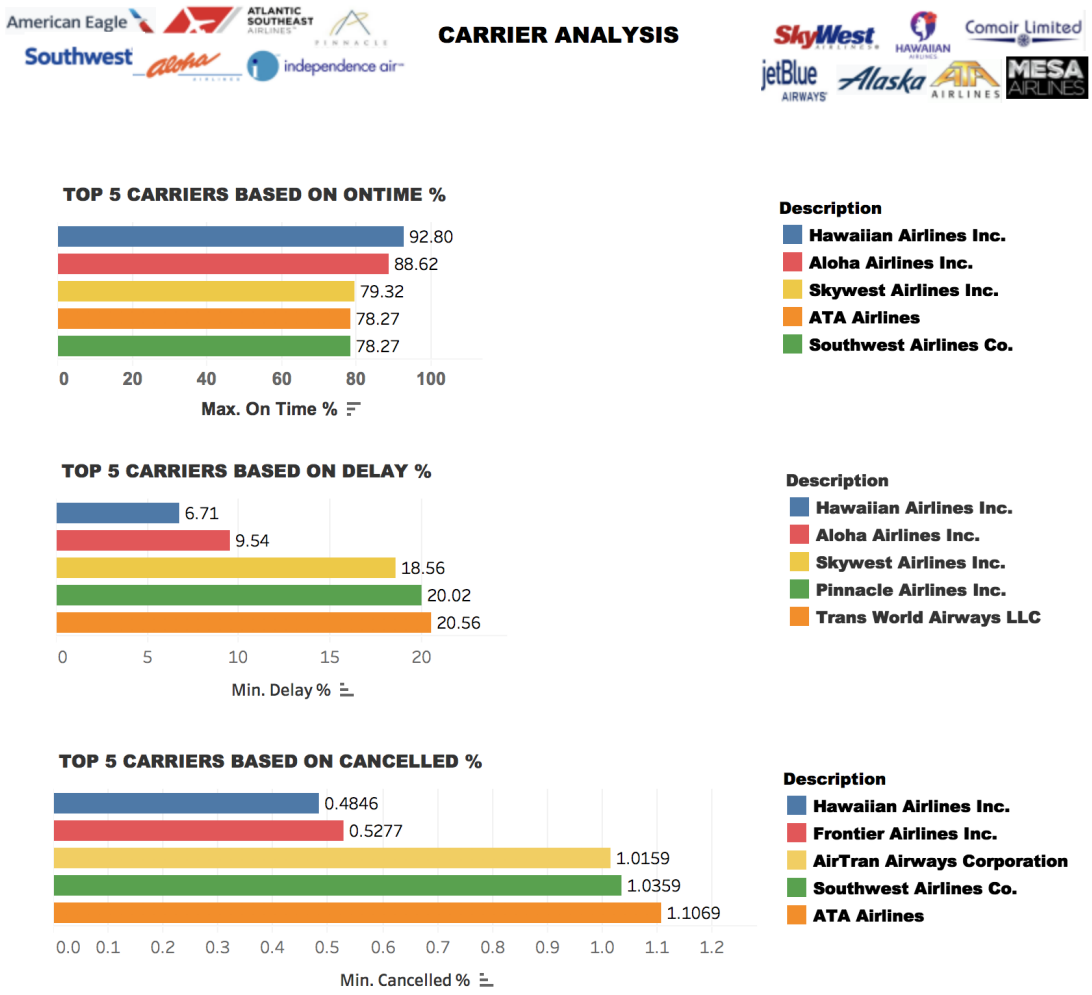


Figure 4 - Visualisation 2 a - Carrier Analysis [Created by Group 14 for COMP5048]

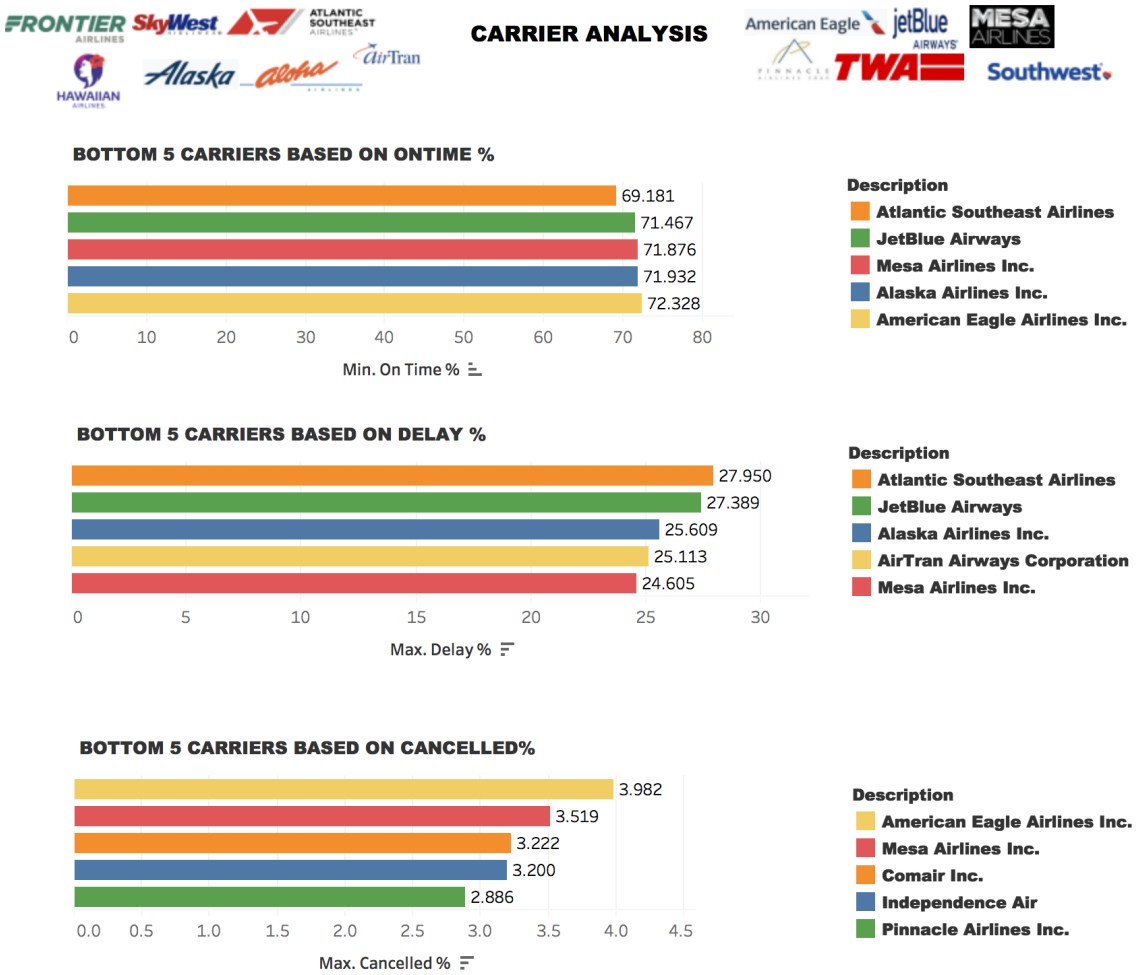


Figure 5 - Visualisation 2 b - Carrier Analysis [Created by Group 14 for COMP5048]

Patterns Identified:

Top Carriers:

- **Ontime** -Hawaiian Airlines
- **Delay Percentage**: Hawaiian Airlines
- **Cancellation Percentage**: Hawaiian Airlines

Bottom Carriers

- **Ontime** -Atlantic Southeast Airlines
- **Delay Percentage**: Atlantic Southeast Airlines
- **Cancellation Percentage**: American Eagle Airlines

Visualisation 3 - Number of flights by carrier per year

An interactive D3 donut chart has been created for the number of flights for each carriers over a ten-year period (1998 - 2008). This visualization assisted us in finding the most operating carrier for each year based on the Airline and the Total flights for each airline. *(Code files have been supplied for this visualisation)*

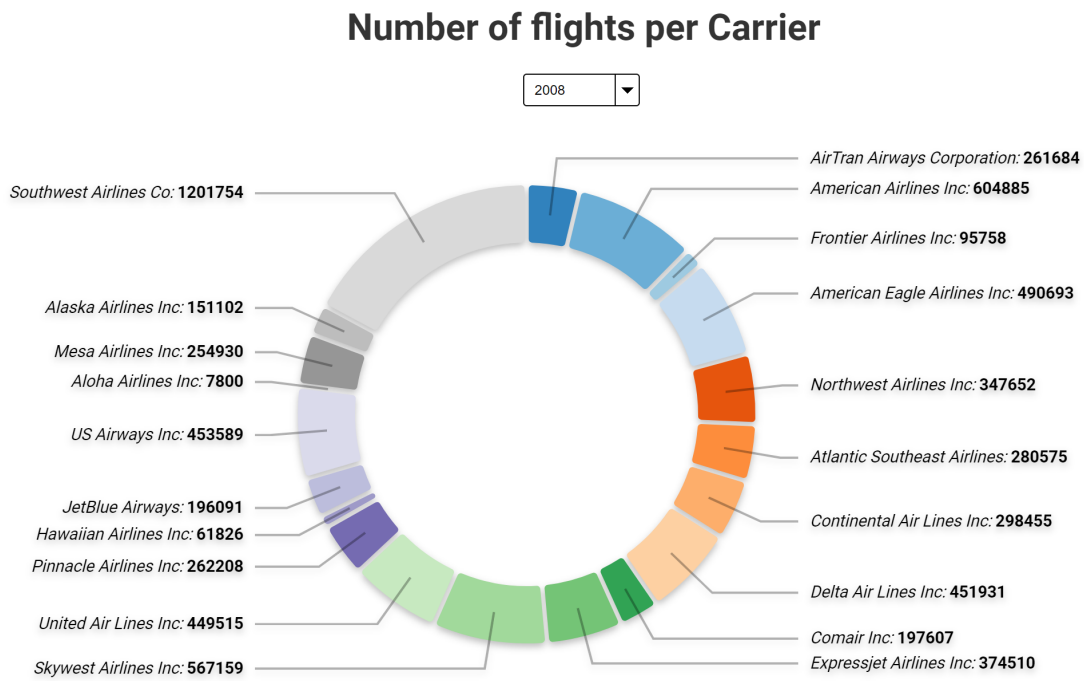


Figure 6 - Visualisation 3 - Number of flights per Carrier [Created by Group 14 for COMP5048]

Code Reference: A static, reusable donut chart for D3.js v4 [5]

Pattern identified:

Through this visualization, we identified that Delta Airlines used to be the most operative airline prior to 2000. However, since 2000 Southwest airline took over Delta Airlines and since then has been operating the most. Possible reasons for the decline in the number of airlines for Delta can be as the offerings that each airline offers to their passengers. Some of these are as below:

- Southwest offer no fees for the first two free checked in bags whereas Delta charge for each airline
- With Southwest passengers can purchase a 24-hour Wi-Fi pass for $8 whereas with Delta they have to pay $16 for the same service [1]
- Southwest also differentiates themselves by keeping low fuel costs and flying on cheaper and older planes thereby offering cheaper flights to their passengers and attracting more passengers [2]

Visualization 4 - Performance of Top 5 operating airlines

An interactive chart was prepared using Google charts to visualize the top 5 operating airlines for the period 2004-2008.

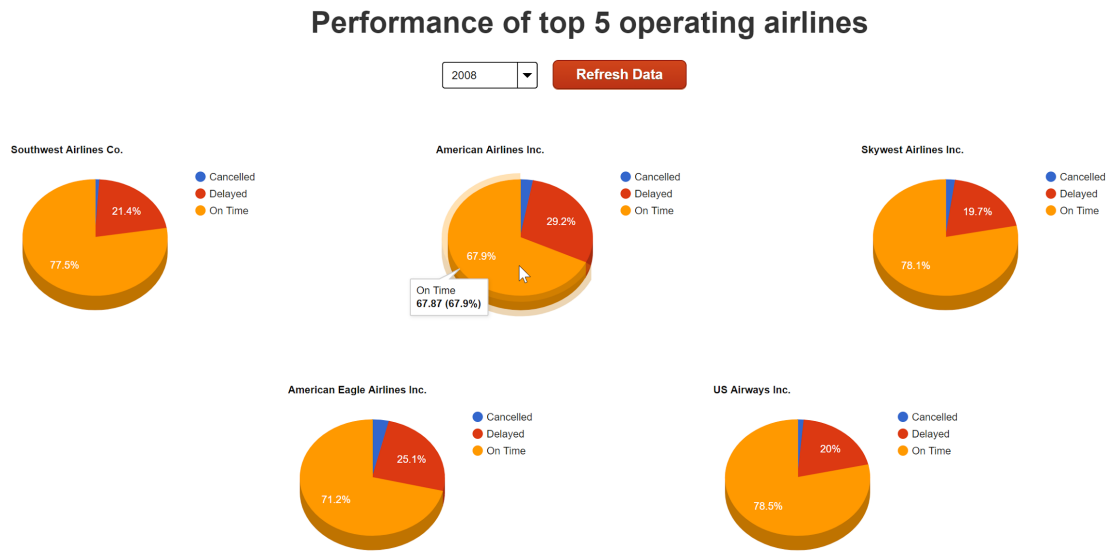


Figure7 - Visualisation 4 -Performance of Top 5 Operating Airlines [Created by Group 14 for COMP5048]

Code Reference: Visualization: Pie Chart | Charts | Google Developers [6]

Pattern identified:

- There have been no major changes in the performance of Southwest Airlines despite being the most operative airline during the period 2004 -2008
- SkyWest Airlines has an upward trend on the number of flights. In year 2005, they were the 5th operative airline whereas in 2006 they moved to the 4th operative airline. However, they experience a negative impact on their performance as the number of on time flights declined by of 5.6%

Visualisation 5 - Total Cancellations and Reasons during a 5-year period

A ***lines (continuous)*** visualisation have been created for the total cancellations VS the reason of flight cancellation. The information depicted by this visualisation will assist us in identifying the reason which led to the cancellation of a particular flight at a specific period of time. The X-axis has been plotted as the month and the Y-axis as the cancelled flights. The colour of the lines in the graph represents the reason for cancellation of flight. These reasons are categorized as below:

- A → cancellation of flight due to carrier
- B → cancellation of flight due to the weather
- C → cancellation of flight due to NAS
- D → cancellation of flight due to security reasons

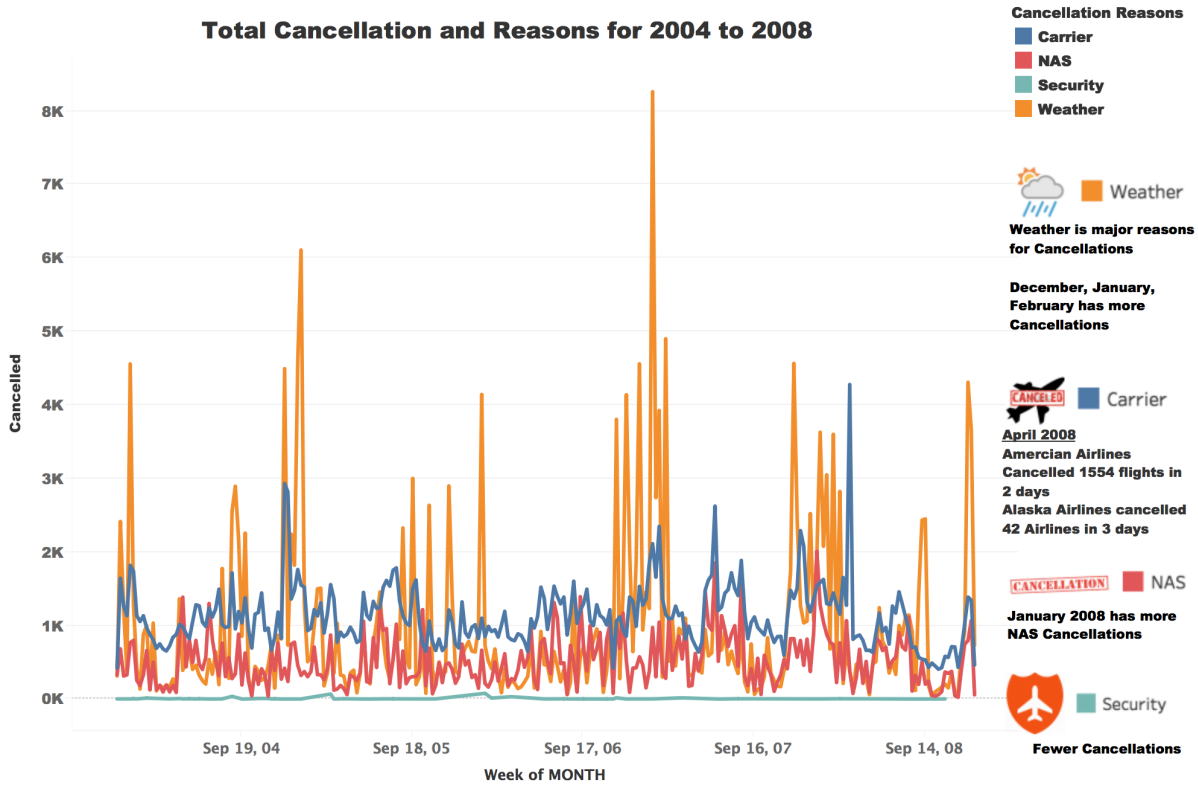


Figure 8 - Visualisation 5 - Total Cancellations and Reasons during a 5-year period [Created by Group 14 for COMP5048]

Pattern Identified:

- From the visualisation 5, We can observe that weather is a major reason for Cancellations from 2004 to 2008
- December, January and February has more number of Cancellations due to weather.
- The Second highest reason for cancellation is due to Carrier. In April 2008, American Airlines cancelled 1554 flights in two days. Alaska Airlines cancelled 42 Airlines in three days [3]
- January 2008 has more number of NAS Cancellations
- There are fewer Cancellations due to Security

Visualisation 6 - Total number of flights and cancellations per week during a 5-year period

A ***heatmap*** visualisation for the total number of cancellations/delays per week /month (Visualisation 6 a to 6 f) have been created, which will assist us in identifying the weeks/months with the maximum and the minimum number of cancellations/delays.

Pattern Identified:

Based on Cancellations

- Best Month: April, May, October, November
- Worst Month: January, February, September, December
- Best Days: Saturday, Sunday
- Worst Days: Monday, Tuesday

Based on Delays:

- Best Month: April, May, September, November
- Worst Month: June, July, August, December
- Best Days: Tuesday, Saturday
- Worst Days: Thursday, Friday

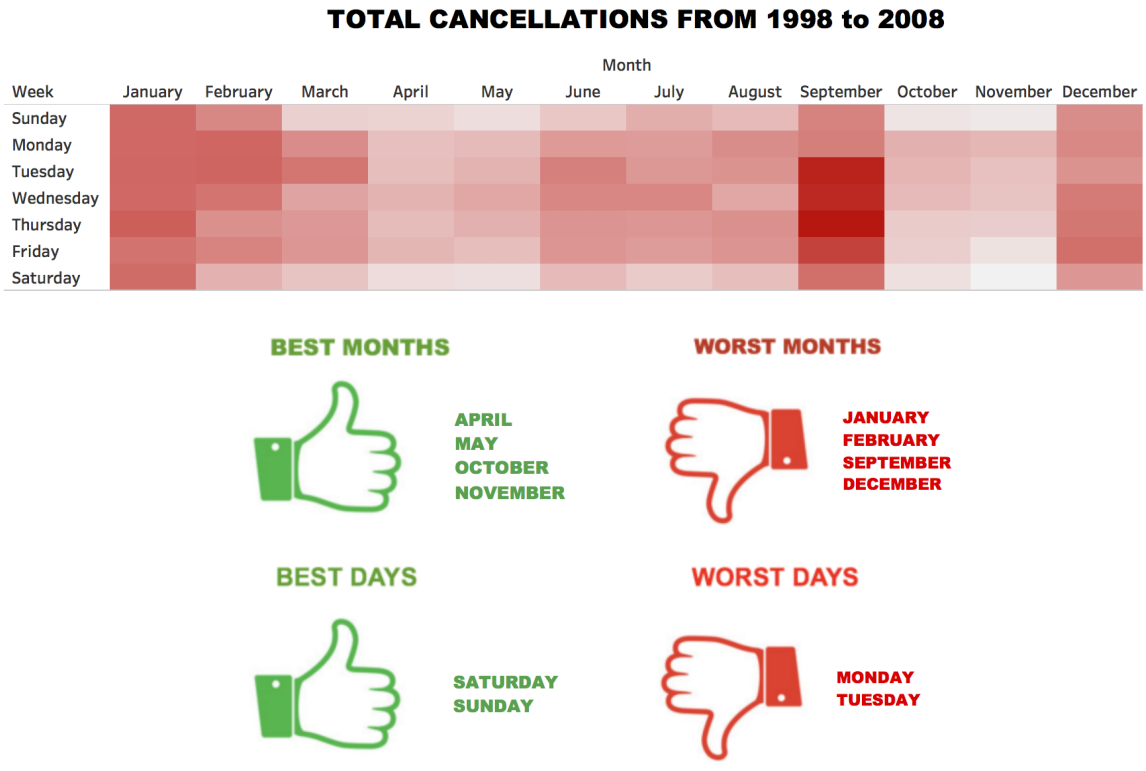


Figure 9 - Visualisation 6 a - Total Cancellations from 1998 to 2008 [Created by Group 14 for COMP5048]

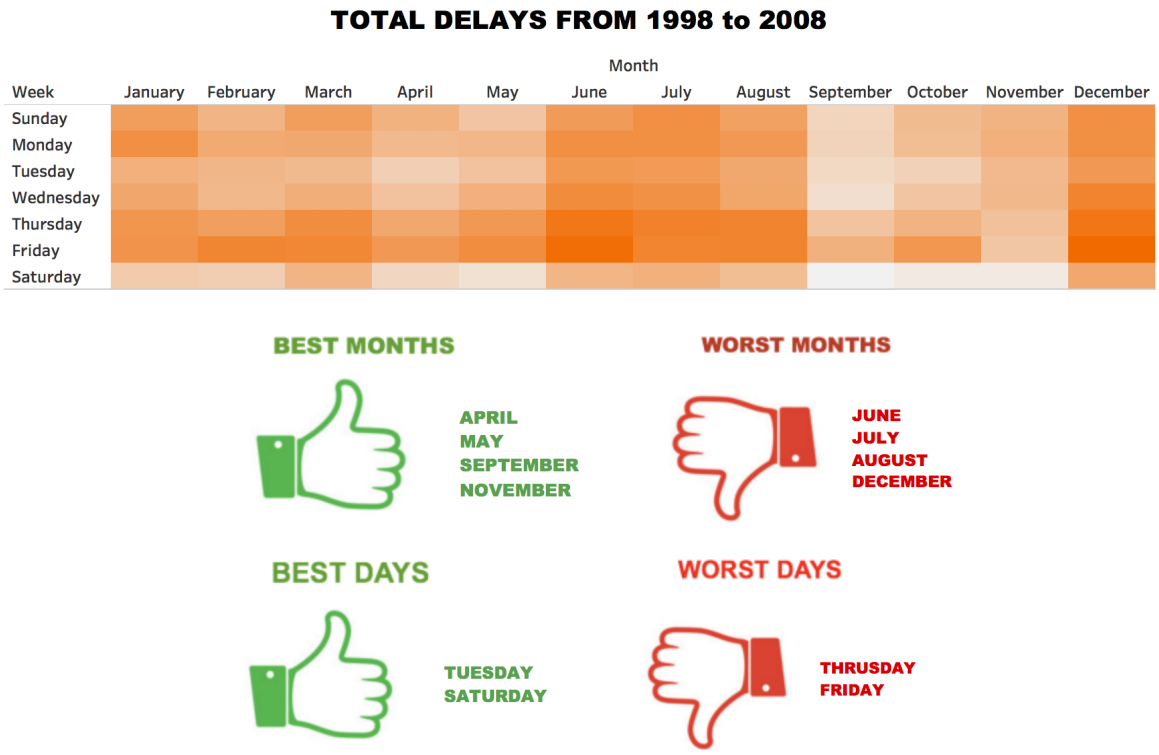


Figure 10 - Visualisation 6 b - Total Delays from 1998 to 2008 [Created by Group 14 for COMP5048]

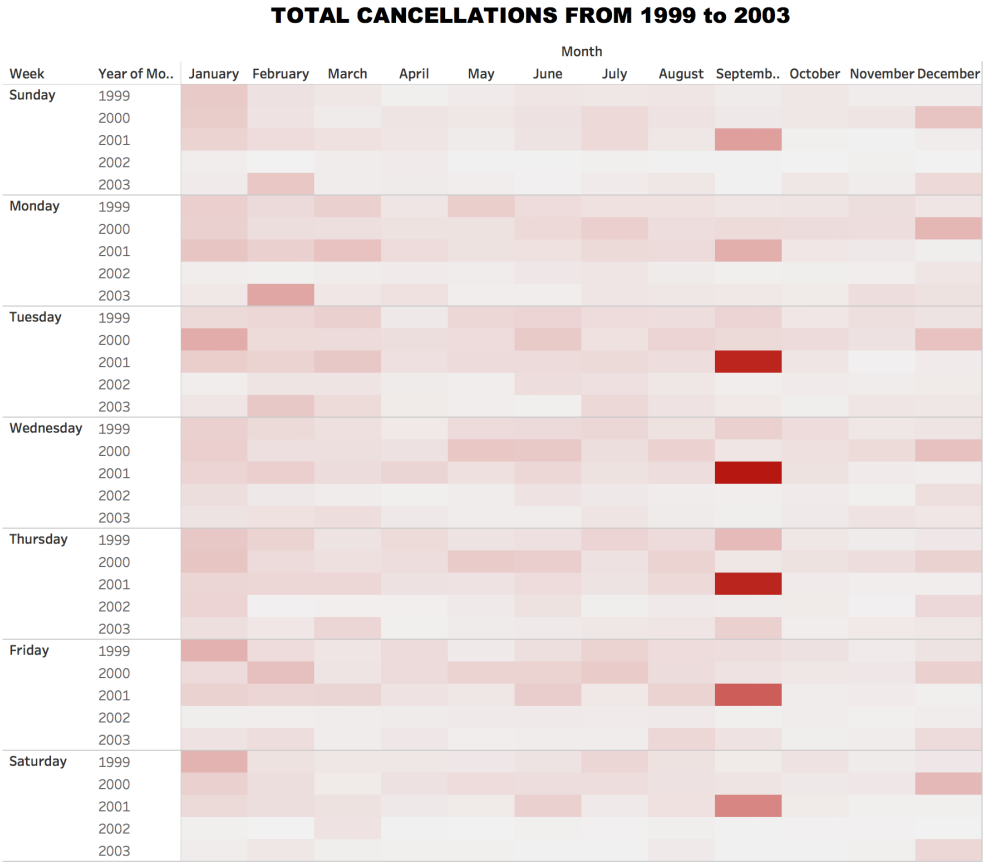


Figure 11 - Visualisation 6 c - Total Cancellations from 1999 to 2003 [Created by Group 14 for COMP5048]

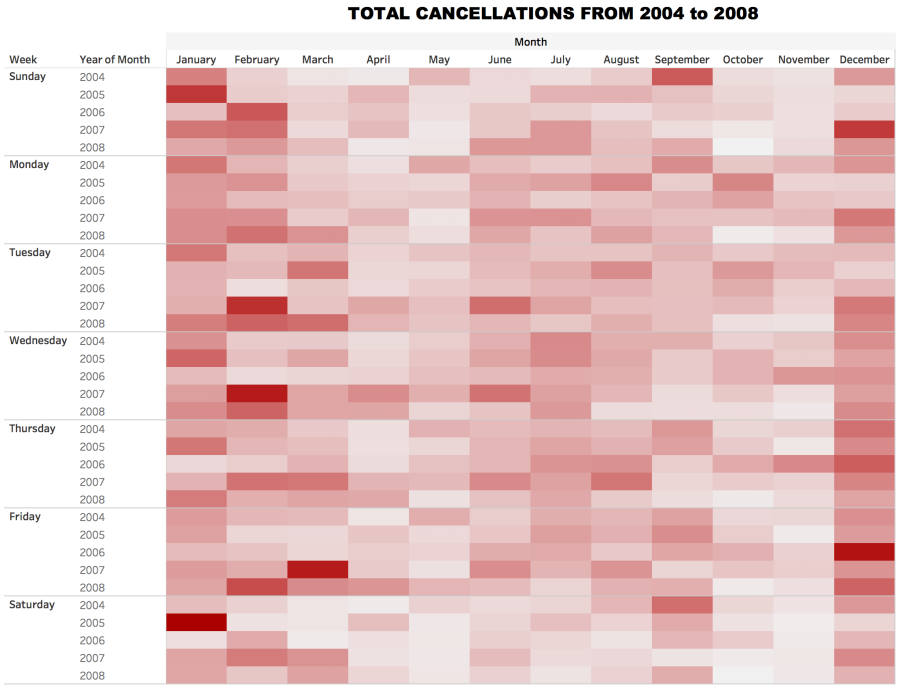


Figure 12 - Visualisation 6 d - Total Cancellations from 2004 to 2008 [Created by Group 14 for COMP5048]

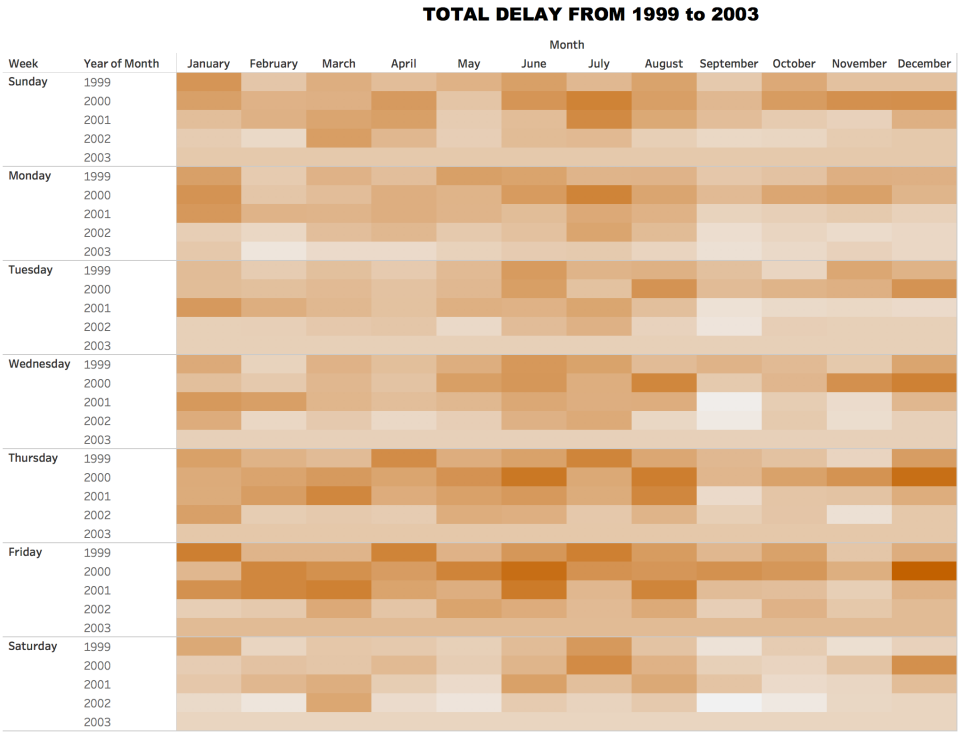


Figure 13 - Visualisation 6 e - Total Delays from 1999 to 2003 [Created by Group 14 for COMP5048]

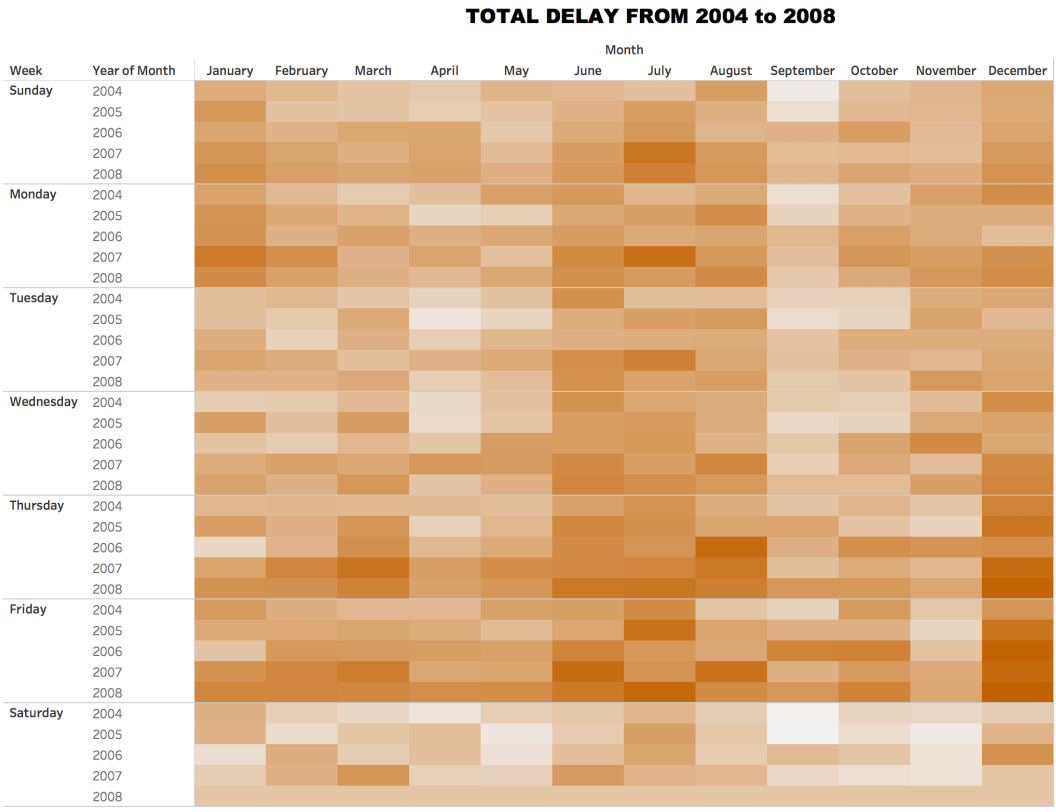


Figure 14 - Visualisation 6 f - Total Delays from 2004 to 2008 [Created by Group 14 for COMP5048]

Visualisation 7 - Airport Performance by State

A ***geo-maps*** visualisation for the airport performance in terms of on time, major delay, and cancellations has been created to assist us in figuring out the best and the worst performing state with respect to the total number of on time, major delay, and cancelled flights. The performance of the airports (on time/cancellation) was calculated and plotted on the USA map. The three visualizations below are from year 2003-2008.

Criteria to decide the best and the worst airport:

- The states with the on-time flight percentage above 85% are considered as best performing states and the states having on time flight percentage below 85% are considered to be the worst performing states.
- The states with the cancelled flight percentage below 2.5% are considered to be the best performing states while the ones having cancelled flight percentage above 2.5% are considered to be the worst performing states.

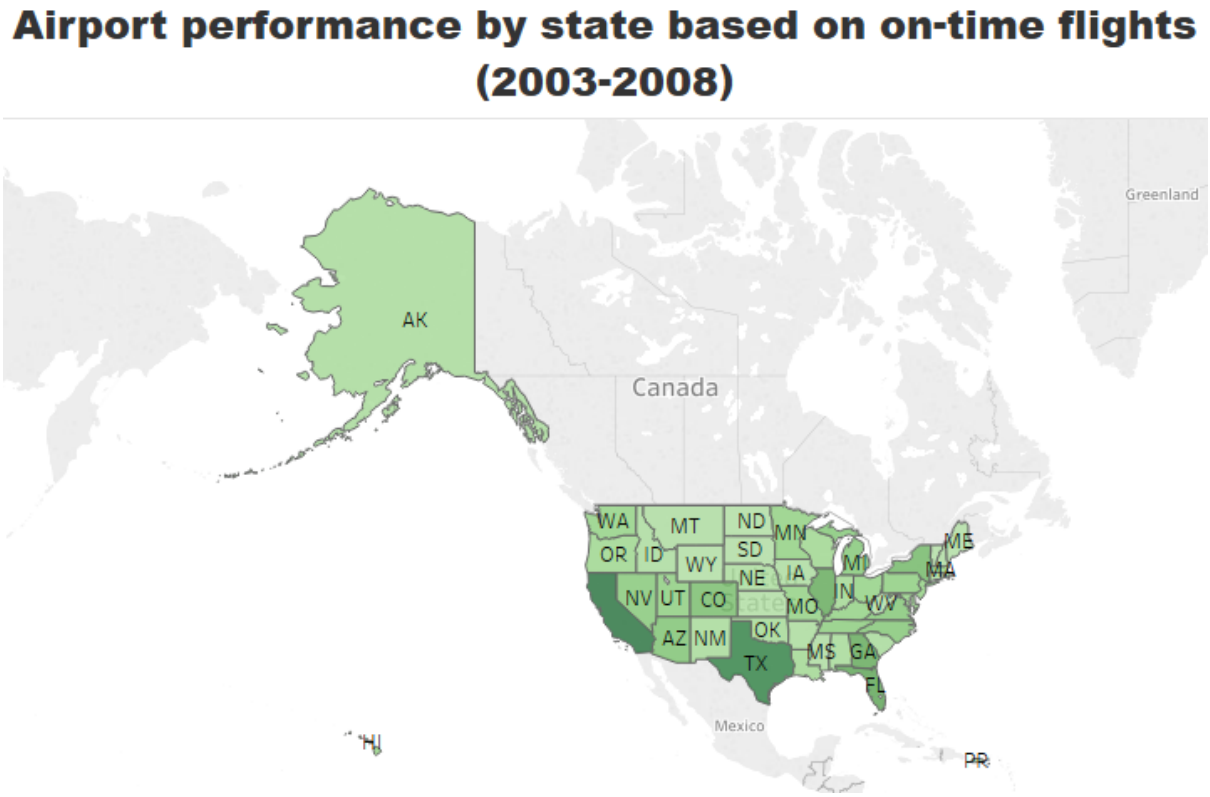


Figure 15 - Visualisation 7 a - Airport Performance by State based on On-Time Flights from 2003 to 2008 [Created by Group 14 for COMP5048]

Patterns identified

Through the visualization, we identified that California state within US has the maximum number of ontime flights of 700,599, from the period 2003-2008.

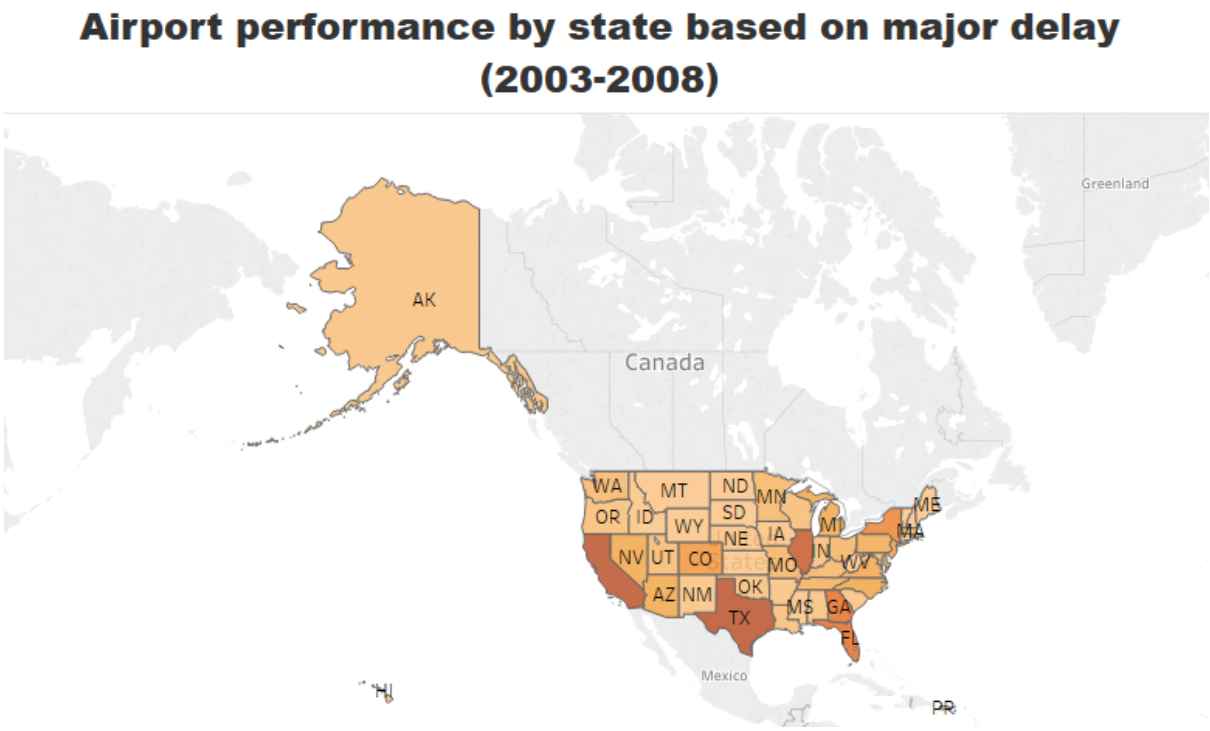


Figure 16 - Visualisation 7 b - Airport Performance by State based on Major delay from 2003 to 2008 [Created by Group 14 for COMP5048]

Patterns identified

The visualization helped us to identify that Texas within USA is the state which has the maximum number of flights with major delays. The number of major delays flights accounts to a number of 79,526 from the period 2003-2008.

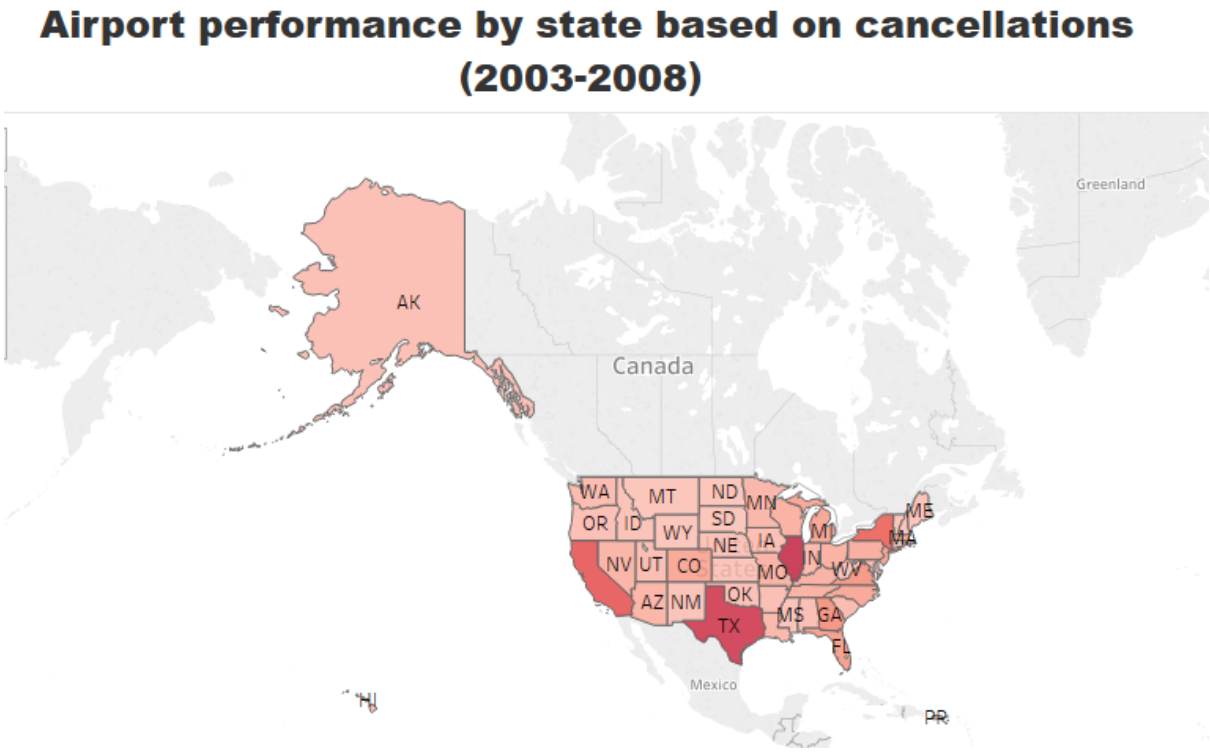


Figure 17 - Visualisation 7 c - Airport Performance by State based on Cancellations from 2003 to 2008 [Created by Group 14 for COMP5048]

Patterns identified

This visualization helped us to identify that Illinois state within USA deals with the maximum number of cancelled flights. The number of cancelled flights accounts to 14,234 over a period of 2003-2008.

Overall patterns identified from all the above three visualisations:

- Illinois and New York are the two worst performing states with large proportion of cancelled flights. In Illinois for every 30 flights, 1 gets cancelled and in New York, for every 31 flights, 1 gets cancelled.
- Illinois and New York have a flight cancelled percentage of 3.25% and 3.15% respectively. Also, Illinois and New York have an ontime flight percentage of 79% and 82%.
- California and Florida are the best performing states with the minimum number of cancelled flights. In California, for every 71 flights, 1 gets cancelled and in Florida, for every 77 flights, 1 gets cancelled.
- California and Florida have a flight cancelled percentage of 1.39% and 1.28% respectively. Also, California and Florida have an on-time flight percentage of 89.13% and 87.2%

Visualisation 8 - Top & Worst Airports

A bubble chart visualisation has been created using Tableau to find the top 5 and worst 5 airports based on the percentage of cancellations, major delays and ontime flights. In this visualization, the intensity of the colour is depicting the percentage of each of the performance measures i.e. cancellations, major delays and on time. Higher the intensity of the colour means the percentage is higher for each of the performance measure. The size of the circles depicts the number of flights for each airport. Larger the size of the bubble, larger the number of flights for that airport.

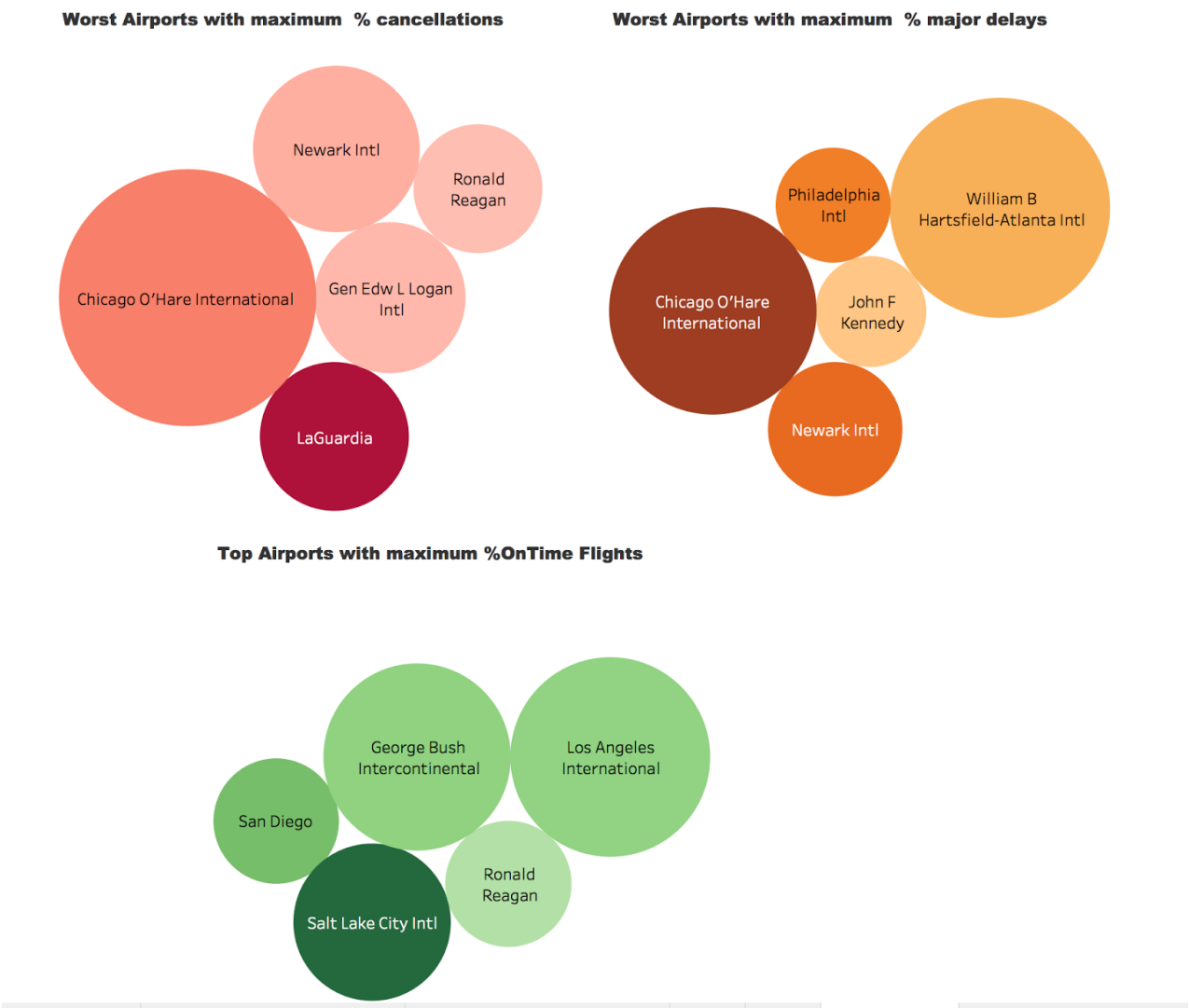


*Figure 18 - Visualisation 8 -**Top and Worst Airports** [Created by Group 14 for COMP5048]*

Patterns Identified:

- ***Top Airport based on percentage of cancellation:*** LaGuardia
- ***Top Airport based on percentage of major delays:*** Chicago O'Hare International
- ***Top Airport based on percentage of ontime flights:*** Salt Lake City International

Visualisation 9 - Airlines Routes

A **geomap** visualisation using Tableau is created to display all the routes for the three selected airports as below:

- Salt Lake City International - This airport has the maximum percentage of time flights
- LaGuardia - This airport has maximum percentage of cancelled flights
- Chicago - This airport has maximum percentage of delayed flights

To show the change in the number of routes a video capture from Tableau was also created. (Refer to the submitted video files)

Salt Lake City International

Year 2003:

Figure 19 - Visualisation 9 a - Airport Routes of Salt Lake City International for year 2003 [Created by Group 14 for COMP5048]

Year 2008:

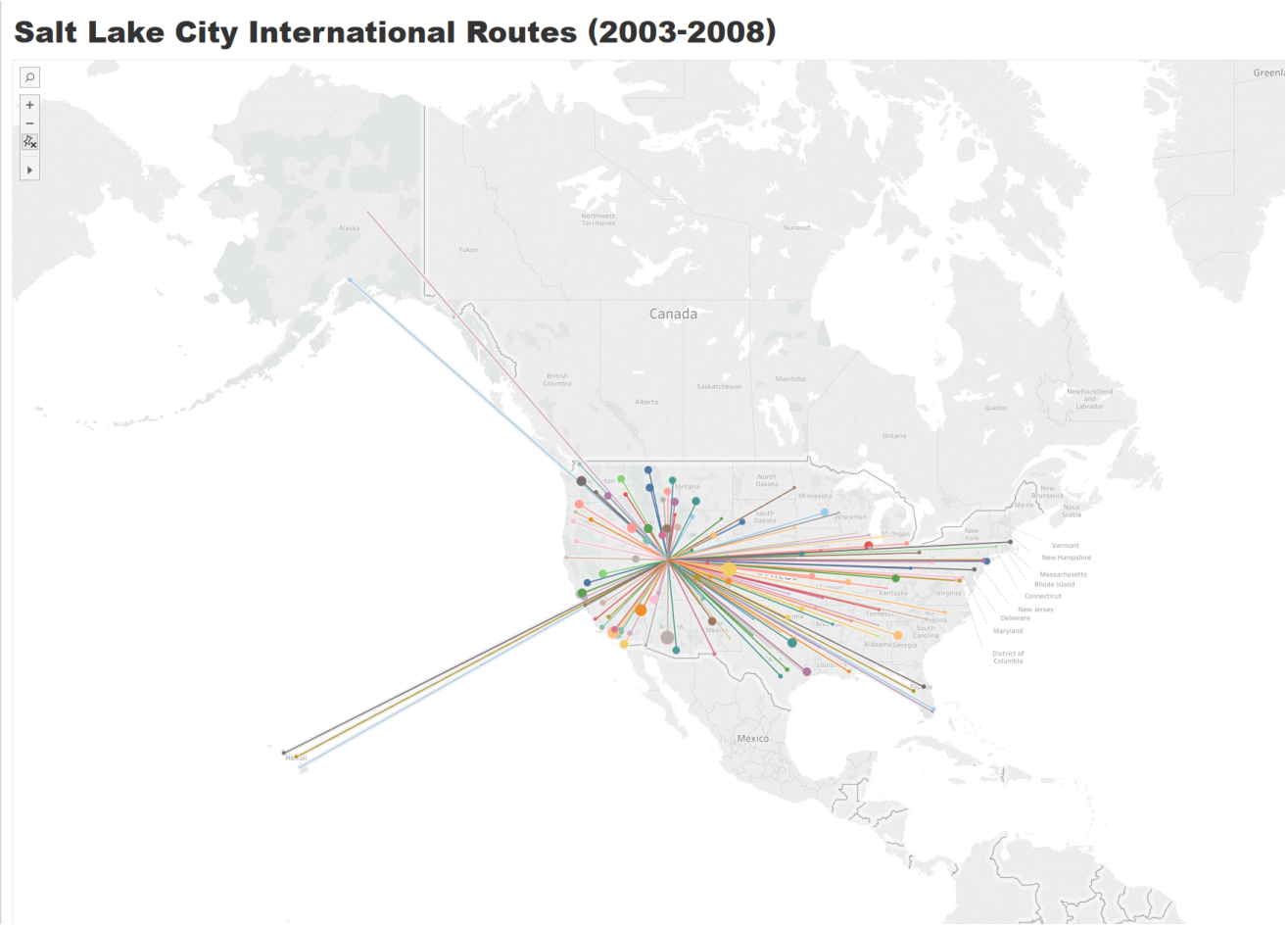


Figure 20 - Visualisation 9 b - Airport Routes of Salt Lake City International for year 2008 [Created by Group 14 for COMP5048]

Through the two visualisations we observed that the number of flights increased every year, however this did not affect the airport's performance and Salt Lake City International remained to be the airport with the maximum percentage of on time flights.

LaGuardia

Year 2003:

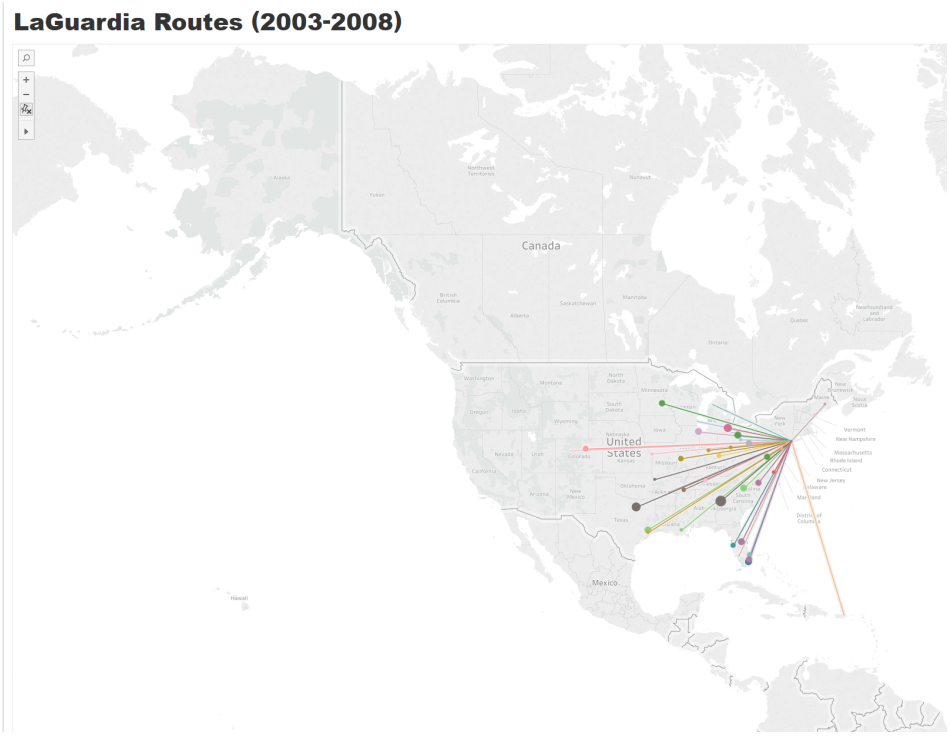


Figure 21 - Visualisation 9 c - Airport Routes of LaGuardia for year 2003 [Created by Group 14 for COMP5048]

Year 2008:

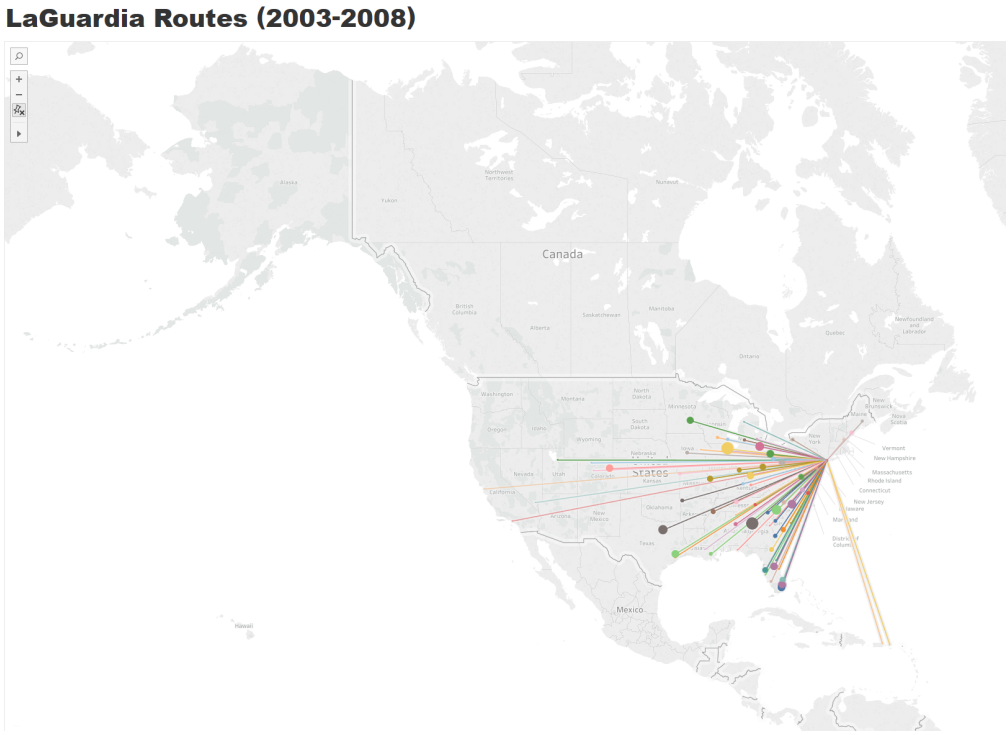


Figure 22 - Visualisation 9 d - Airport Routes of LaGuardia for year 2008 [Created by Group 14 for COMP5048]

Through the two visualisations we observed that the number of flights have increased every year. The increase in the number of flights every year could be a possible reason for why LaGuardia airport remained as the airport with the maximum percentage of cancelled flights.

Chicago O'Hare International Airport:

Year 2003:

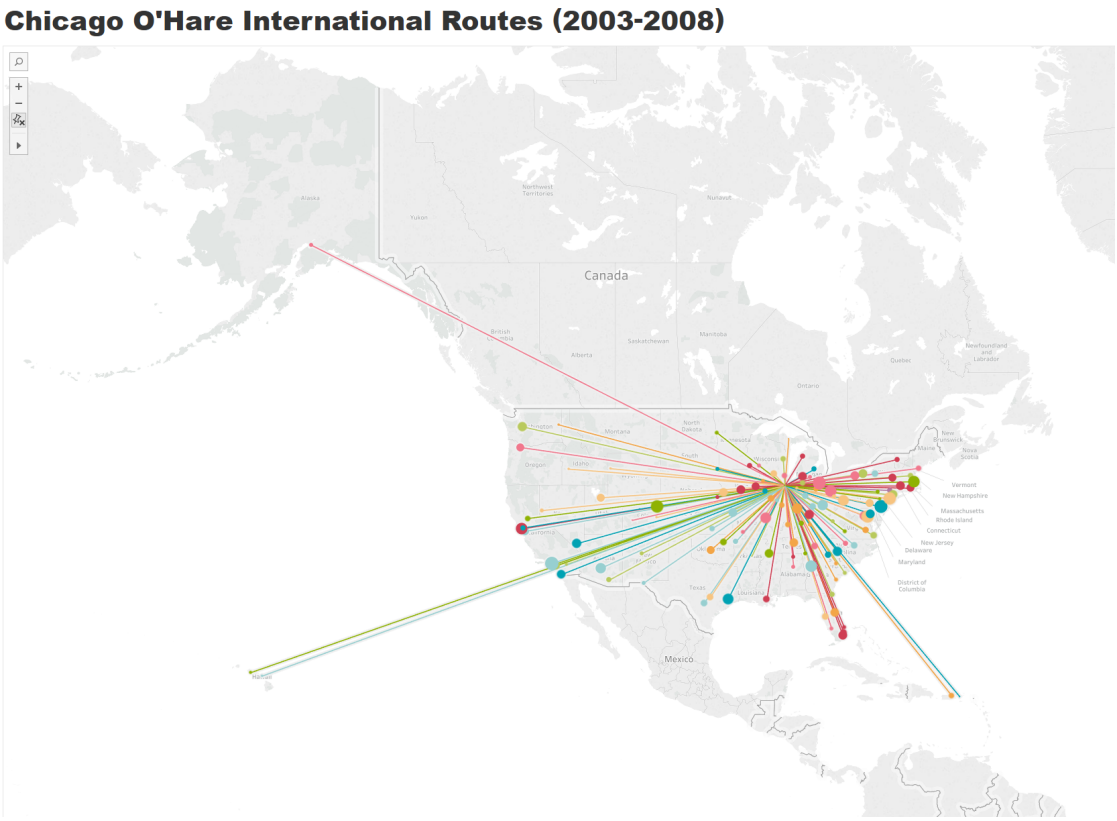


Figure 23 - Visualisation 9 e - Airport Routes of Chicago O'Hare International for year 2003 [Created by Group 14 for COMP5048]

Year 2008:

Figure 24 - Visualisation 9 f - Airport Routes of Chicago O'Hare International for year 2008 [Created by Group 14 for COMP5048]

Through the two visualisations we observed that the number of flights have increased every year. The increase in the number of flights every year could be a possible reason for why Chicago O'Hare International airport remained as the airport with the maximum percentage of delayed flights.

Visualisation 10 - Trend Analysis

A ***lines (continuous)*** visualisation have been created for the based on the Top Five Airlines and Airports with respect to Cancellation and Delays to analyse the Trends across years.

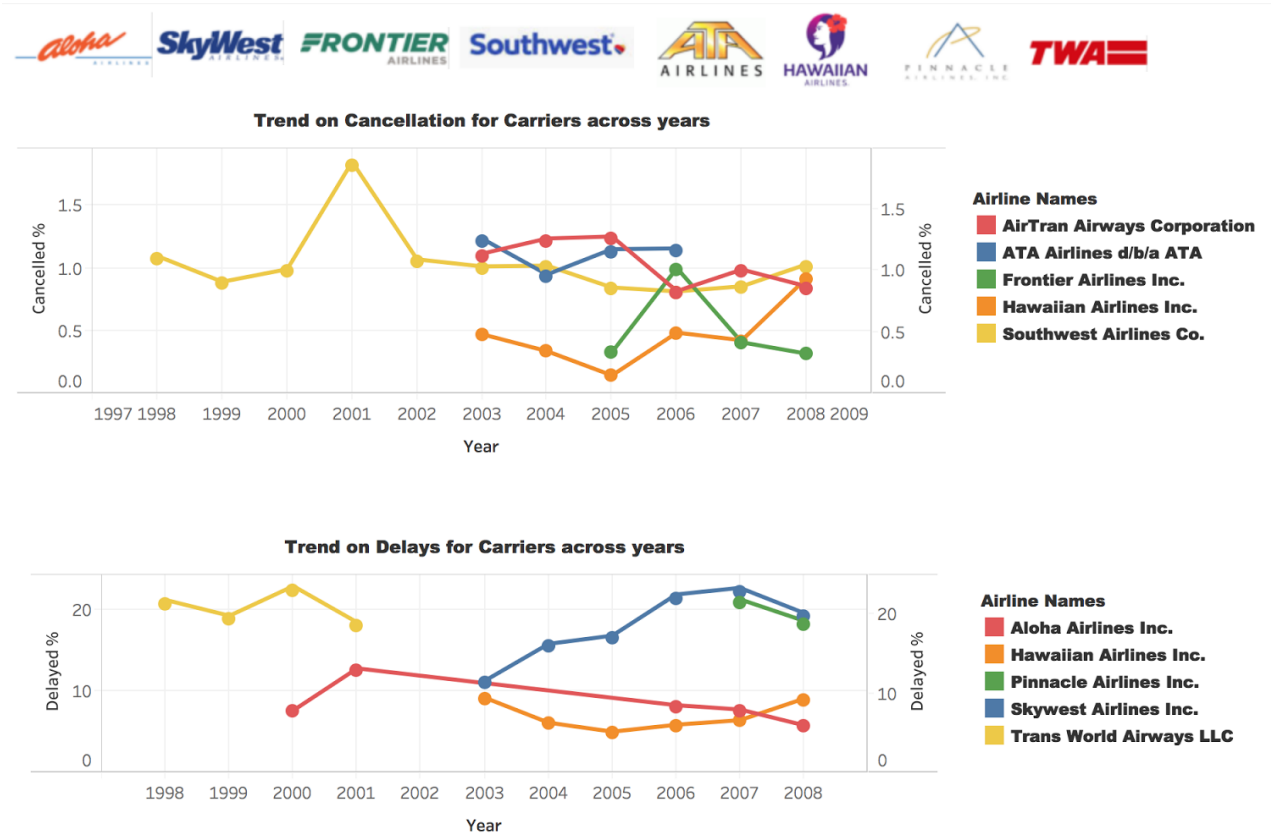


Figure 25 - Visualisation 10 a - Trend analysis on Cancellation/Delays for Carriers [Created by Group 14 for COMP5048]

From the Visualisation 10 a we can observe the trend For Cancellation and Delays for Carriers. We have Identified the Top 5 Carriers from Visualisation 2 a and 2b-Carrier Analysis. We can observe that South-west Airlines experienced higher Cancellation percentage and Aloha Airlines had higher delay percentage in 2001 due September 11 terrorist attack. On 26th Nov and 17th December of 2006, there was 3805 and 4140 cancellations respectively due to weather (From Visualisation 5). Frontier, Southwest and ATA Airlines has higher Cancellations percentage on 2006.After 2006, the cancellation percentage of Frontier had gradually decreased.

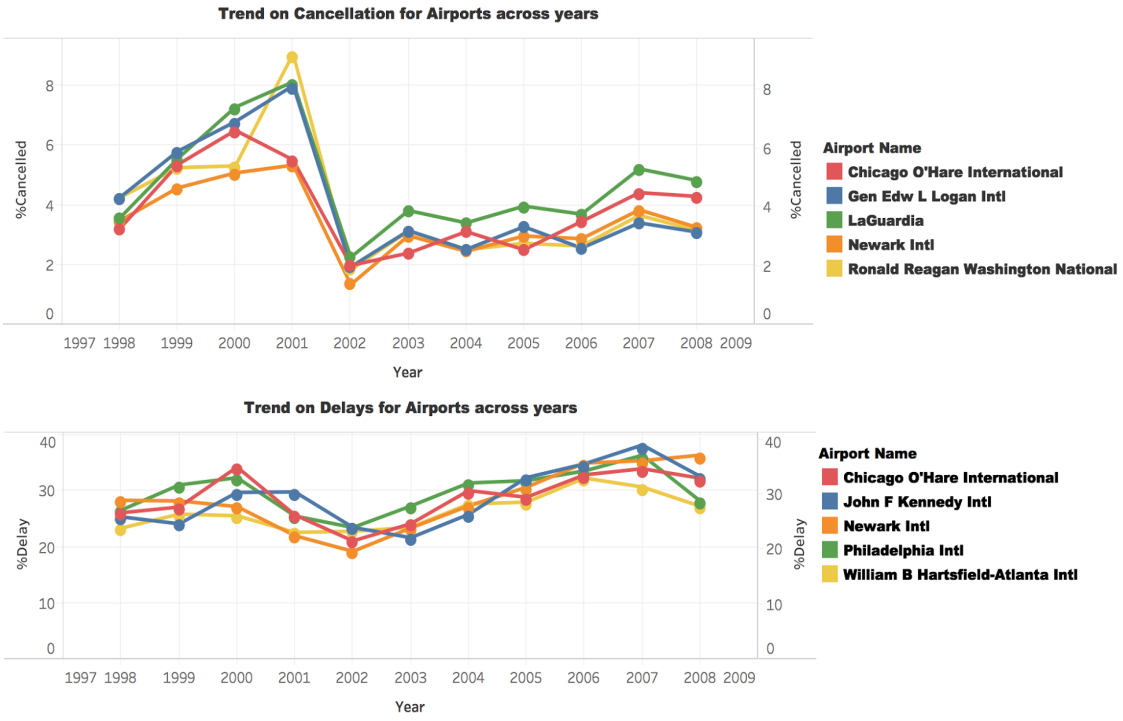


Figure 26 - Visualisation 10 b - Trend analysis on Cancellation/Delays for Airports [Created by Group 14 for COMP5048]

From the Visualisation 10 b we can observe the trend For Cancellation and Delays for Airports. All the five airports (Chicago O'Hare International, Gen Edw L. Logan International, LaGuardia, Newark International and Ronald Reagan Washington National) had higher cancellation Percentage in 2001 due September 11 terrorist attack. In 2002, all the five airports experienced a sudden dip in the cancellation Percentage. In 2007, all the 5 airports had highest delay Percentage and comparatively higher Cancellation percentage. According to USA Today Article published in 2007, the main reasons for the cause of delays are due to shortages of pilot, longer time to refuel and mechanical breakdowns. [4]

Implementation

The following tools were utilized for data processing, storage and visualizations:

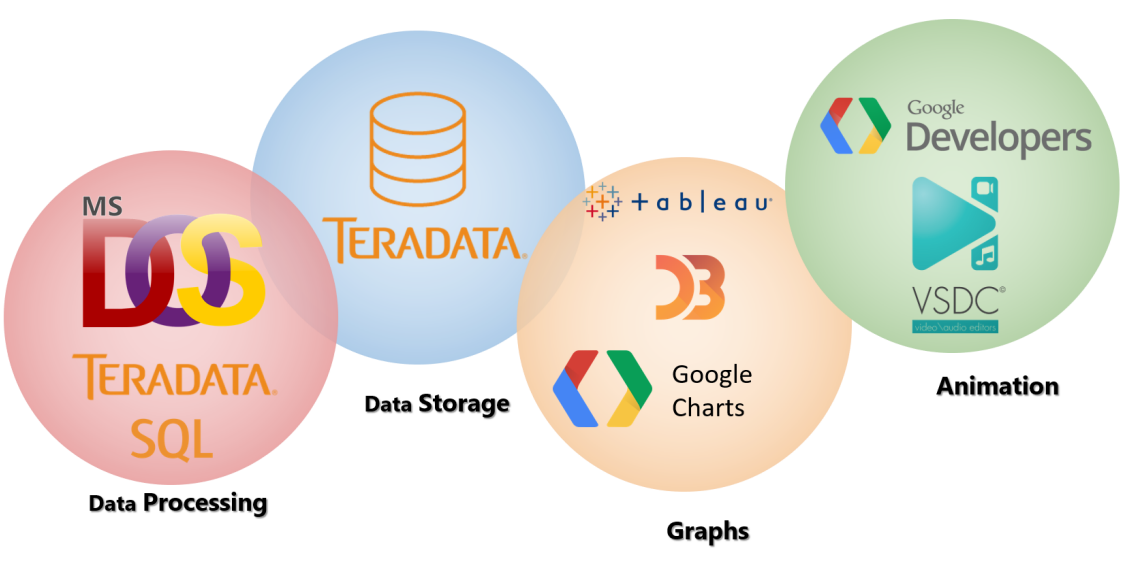


Figure 27 - Implementation Tools [Created by Group 14 for COMP5048]

Evaluation and Results

To achieve the aim of our analysis, we evaluated based on the following tasks and we have summarised the results

Task Name	Visualization Number	Visualisation Name	Results
Factors affecting travel	*Visualisation 1 a* *Visualisation 1 b*	*Total flights Vs Cancellation for Year 2001* *Total flights Vs Cancellation for Year 2008*	• Carriers operate less flights on Saturdays • Less flights observed on 27th November which is thanksgiving holiday • Airlines traffic reduced after 9/11 attack in 2001 • Airlines traffic reduced after Global Financial Crisis in 2008
Top 5 Carriers	*Visualisation 2 a* *Visualisation 2 b*	Carrier Analysis	**Top Carriers:** • On time -Hawaiian Airlines • Delay Percentage: Hawaiian Airlines • Cancellation Percentage: Hawaiian Airlines

			Bottom Carriers ● **Ontime** -Atlantic Southeast Airlines ● **Delay Percentage**: Atlantic Southeast Airlines ● **Cancellation Percentage**: American Eagle Airlines
Identify Top 5 Operative Airlines	*Visualisation 3*	Number of flights by carrier per year	● Delta Airlines used to be the most operative airline till 2000. ● Since 2000 Southwest airline has been most operative Airlines
Analyse the top 5 Operative Airlines based on Delay and	Visualisation 4	Performance of Top 5 operating airlines	● Southwest operates more flights during the period 2004 -2008 ● SkyWest Airlines has upward trend on the number of flights ● Northwest Airlines had steady decline during the period 2004-2008
Identify Various reasons for Cancellations	Visualisation 5	*Total Cancellations and Reasons during a 5-year period*	● Weather is a major reason for Cancellations from 2004 to 2008 ● More Cancellations on December, January and February due to weather ● Second highest reason for cancellation is due to Carrier ● January 2008 has more number of NAS Cancellations ● Fewer Cancellations due to Security
Best Month/Days to travel	*Visualisation 6 a* *Visualisation 6 b* *Visualisation 6 c* *Visualisation 6 d* *Visualisation 6 e* *Visualisation 6 f*	*Total Cancellations from 1998 to 2008* *Total Delays from 1998 to 2008* *Total Cancellations from 1999 to 2003* *Total Cancellations from 2004 to 2008* *Total Delays from 1999 to 2003* *Total Delays from*	**Based on Cancellations** ● **Best Month**: April, May, October, November ● **Worst Month**: January, February, September, December ● **Best Days:** Saturday, Sunday ● **Worst Days:** Monday, Tuesday **Based on Delays:** ● **Best Month**: April, May, September, November ● **Worst Month**: June, July, August,

		2004 to 2008	December ● **Best Days**: Tuesday, Saturday ● **Worst Days**: Thursday, Friday
Identify Airport Performance by State	*Visualisation 7 a* *Visualisation 7 b* *Visualisation 7 c*	*Airport Performance by State based on On-Time Flights from 2003 to 2008* *Airport Performance by State based on Major delay from 2003 to 2008* *Airport Performance by State based on Cancellations from 2003 to 2008*	• Illinois and New York are worst performing states with more percentage of cancelled and less On-time flights • Texas is the state which has the maximum number of flights with major delays • California and Florida have less percentage of Cancelled flights and higher on-Time flights
Identify Top & Worst Airports	*Visualisation 8*	***Top and Worst Airports***	● ***Top Airport based on percentage of cancellation:*** LaGuardia ● ***Top Airport based on percentage of major delays:*** Chicago O'Hare International ● ***Top Airport based on percentage of ontime flights:*** Salt Lake City International
Analysis of Airport Routes	*Visualisation 9 a* *Visualisation 9 b* *Visualisation 9 c* *Visualisation 9 d* *Visualisation 9 e* *Visualisation 9 f*	*Airport Routes of Salt Lake City International for year 2003* *Airport Routes of Salt Lake City International for year 2008* *Airport Routes of LaGuardia for year 2003* *Airport Routes of LaGuardia for year 2008*	The increase in the number of routes does not appear to the affect the performance of the top & worst airports. This result was achieved through following observations over a six-year period: ● Salt Lake City International, the airport with the maximum percentage of ontime flights remained to be the airport with the maximum percentage of on time flights ● LaGuardia airport the airport with the maximum percentage of cancelled flights remained as the airport with the maximum

		Airport Routes of Chicago O'Hare International for year 2003 *Airport Routes of Chicago O'Hare International for year 2008*	percentage of cancelled flights. ● Chicago O'Hare International airport the airport with the maximum percentage of delayed flights remained as the airport with the maximum percentage of delayed flights.
Trend Analysis of Carriers and Delays based on Cancellation/Delay	*Visualisation 10 a* *Visualisation 10 b*	*Trend analysis on Cancellation/Delays for Carriers* *Trend analysis on Cancellation/Delays for Airports*	***Airlines*** ● *Due to terrorist attack (2001 September 11) South-west Airlines experienced higher Cancellation percentage and Aloha Airlines had higher delay percentage* ● *Frontier, Southwest and ATA Airlines has higher Cancellations percentage on 2006.After 2006, the cancellation percentage of Frontier had gradually decreased.* ***Airports*** ● Chicago O'Hare International, Gen Edw L. Logan International, LaGuardia, Newark International and Ronald Reagan Washington National had higher cancellation percentage i*n 2001 due to terrorist attack* ● *In 2002, all the five airports experienced a sudden dip in the cancellation Percentage.* ● *Main reasons for the cause of delays are due to shortages of pilot, longer time to refuel and mechanical breakdowns.*

Discussion

Strengths

- We tried to maintain the colour uniformity in the visualisations we prepared. For instance, we choose green colour for the on-time flights, orange colour for the delays, and red colour for the cancellations.

- We tried to provide a detailed analysis and, the patterns we identified from each of the visualisation side by side.
- The heat map visualizations we prepared displays huge data in a compact manner.
- Initially, we prepared a bubble chart with a large number of dataset, but its drawback was that it seemed to be dense and cluttered and so, we were not able to see the names of some of the airport. So, to avoid this, we displayed only the top and the worst five airports in the bubble chart which turned out to be a good visualization.
- One of our task was to display the performance of only the top five operating airlines and so we choose the pie chart representation for this task as it is suitable to display less amount of data in a pie chart.
- We chose Teradata for Data loading and processing as it supports huge volume of data seamlessly and supports database capacity expansion if we need to scale up to cover more global flight data in the future.

Weakness

- In the heat map visualizations, we created, it is difficult to observe the changes in the colour of the cells immediately.
- There can be some level of difficulty in reading the code of the state names in the geo maps we created.
- Visualization 1 - Our analysis identified only the usual pattern that occurred across the ten-year period such as less flights on Saturdays and thanksgiving and the effect of 9/11 incident and Global Financial Crisis. In addition to this, there were unusual cancellations and drop in airlines traffic. Relating these unusual patterns to the root cause would provide more insights.

Conclusion

We used rigorous data analysis and visualization techniques to not just learn various methodologies, but also, we tried to deduce reasonable and useful insights which are relevant to the actual problems associated with flight industry and the passengers.
Some key questions such as best time (Day or month) to travel, best route for some specific source and destinations, carriers and the airports which are top performers or bad performers, were tried to be solved which would actually help real business case rather than just a proof of concept.
We also tried to build a grid based data storage option which can hold years of data and can even scale up to cover more global flight data and not just limited to US airline industry.

In a nutshell, it was a very good learning experience where we learnt technical method of data visualization, importance of iterative development, collaboration, task prioritization, presentation approach to tell complex designs into simple user friendly stories.
This will indeed help us to be a better analyst in future with focus to not just problem solution but to an overall n fold approach which will actually help to do things which actually matter

References

[1]"How to Compare US Airlines (Southwest, Delta, American, Jet Blue, United, and Virgin) | Me Want Travel", *Me Want Travel*, 2016. [Online]. Available: http://mewanttravel.com/compare-us-airlines/

[2]"Southwest Airlines: Delta's Worst Enemy? (LUV, DAL)", *Investopedia*, 2015. [Online]. Available: http://www.investopedia.com/articles/markets/080515/southwest-airlines-deltas-worst-enemy.asp.

[3]J. Bailey, "American Airlines Cancels 922 More Flights", *Nytimes.com*, 2008. [Online]. Available: http://www.nytimes.com/2008/04/10/business/11aircnd.html.

[4]"Airline glitches top cause of delays - USATODAY.com", Usatoday30.usatoday.com, 2007. [Online]. Available: https://usatoday30.usatoday.com/travel/flights/2007-12-20-flight-delays_N.htm

[5]M. Hall, "A static, reusable donut chart for D3.js v4.", *Bl.ocks.org*, 2017. [Online]. Available: https://bl.ocks.org/mbhall88/b2504f8f3e384de4ff2b9dfa60f325e2.

[6]"Visualization: Pie Chart | Charts | Google Developers", Google Developers. [Online]. Available: https://developers.google.com/chart/interactive/docs/gallery/piechart

Appendix 1 - Meeting Minutes

Meeting Number	1 - Kick off		
Date of Meeting:	14/Sep/17	**Minutes Prepared By:**	Ruchita
1. Meeting Objective and Outcome			
The project was kicked off with team introduction. Group Leader and Vice Group Leader were allocated. At the end of the meeting each member was asked to go through all the data sets provided and come up with views on each to determine the best data set for our project. Next meeting was scheduled - 18/Sep/17			
2. Attendance at Meeting			
Ruchita Manuja, Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date			

3. Role Allocations for Project		
Ruchita Manuja	Group Leader	
Supraja Sridharan	Vice Group Leader	
Anirudh Sharma	Group Member	
Sundaram Thangaraj	Group Member	
Abhijeet Date	Group Member	

Meeting Number	2 - Dataset Selection		
Date of Meeting:	18/Sep/17	**Minutes Prepared By:**	Supraja

1. Meeting Objective and Outcome

There was an online meeting initiated by our Group Leader. Every group member was asked to explain their preferred dataset along with their brief understanding. We have narrowed down our data selection to two- Flights data set and Mini Challenge Data Set. At the end, all the group members chose Flights data set unanimously.

Google Drive was created to collaborate and work together.

All the members were asked to go through the airport data in depth and come up with a set of aims before the next meeting.

Initial report tasks were also divided as per the role allocations

Next meeting was scheduled - 23/Sep/17 (Location: School of IT)

2. Attendance at Meeting

Ruchita Manuja, Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date

3. Role Allocations at the end of the meeting		
Ruchita Manuja	Introduction, Related Work, Visualization Explanation, Planning	
Supraja Sridharan	Analysis, Related Work, Visualization Samples	
Anirudh Sharma	Data Set, Related Work, Visualization Explanation	
Sundaram Thangaraj	Visualization Samples, Evaluation	
Abhijeet Date	Design & Approaches, Evaluation	

Meeting Number	3 - Work Collaboration and working on Initial Report		
Date of Meeting:	23/Sep/17	**Minutes Prepared By:**	Ruchita

1. Meeting Objective and Outcome

Face to face meeting was conducted in University. Each team member discussed their approaches and then collaborated the work done so far for the Initial Report. At the end of the meeting the following work was completed:

- Introduction
- Data Set
- Visualization Samples
- Related Work - Winner 1
- Implementation

The following work was pending at the end of the meeting

- Related Work - Winner 2
- Design & Approaches
- Explanation of the Visualization Samples
- Analysis
- Evaluation

● Detailed Planning of the tasks
2. Attendance at Meeting
Ruchita Manuja, Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date
3. Role Allocations at the end of the Meeting

Ruchita Manuja	Future Task Planning	
Supraja Sridharan	Analysis	
Anirudh Sharma	Explanation of the Visualization Graphs	
Sundaram Thangaraj	Evaluation	
Abhijeet Date	Design & Approaches	

Meeting Number	4 - Finishing off Initial Report		
Date of Meeting:	27/Sep/17	**Minutes Prepared By:**	Ruchita

1. Meeting Objective and Outcome
A group call was conducted and all tasks were reviewed as a group. Initial Report was finalized for submission Tasks for next week were assigned
2. Attendance at Meeting
Ruchita Manuja, Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date
3. Role Allocations at the end of the Meeting

Ruchita Manuja	Data Processing	

Supraja Sridharan	Data Preparation using DOS commands	
Anirudh Sharma	Data Preparation using DOS commands	
Sundaram Thangaraj	Data Processing	
Abhijeet Date	Setup Teradata environment	

Meeting Number	**5 - Consolidating data extracts and preparation for Presentation**		
Date of Meeting:	06/Oct/17	**Minutes Prepared By:**	Ruchita

1. Meeting Objective and Outcome

A face to face meeting was conducted. Processed data was consolidated and added on the shared drive for the team to start creating visualizations.

Presentation tasks were divided

2. Attendance at Meeting

Ruchita Manuja , Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date

3. Role Allocations at the end of the Meeting

Ruchita Manuja	Presentation Slides, Routes Visualization & Analysis	
Supraja Sridharan	Presentation Slides, Airlines Visualization & Analysis	

Anirudh Sharma	Presentation Slides, Day Visualization & Analysis	
Sundaram Thangaraj	Presentation Slides, 5 year Visualization & Analysis	
Abhijeet Date	Presentation Slides, Total flights/delays Visualization & Analysis	

Meeting Number	**6 - Finalize Presentation**		
Date of Meeting:	**11/Oct/17**	**Minutes Prepared By:**	Ruchita

1. Meeting Objective and Outcome

A hangouts call was conducted with the team to finalise the presentation slides.
More task ideas were created and allocated.
A google spreadsheet was created for task allocation

2. Attendance at Meeting

Ruchita Manuja , Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date

3. Role Allocations at the end of the Meeting

Ruchita Manuja	Number of Flights for each carrier visualization	
Supraja Sridharan	Total Flights on daily basis	
Anirudh Sharma	Airport Performance Visualization	
Sundaram Thangaraj	Dataset Creation	

Abhijeet Date	Trend Analysis for Carriers, Airports and Airlines	

Meeting Number	**7 - Presentation Practice**		
Date of Meeting:	17/Oct/17	**Minutes Prepared By**:	Ruchita
1. Meeting Objective and Outcome			
The team met in SIT and practiced for Presentation. Future task allocations remained as per previous meeting			
2. Attendance at Meeting			
Ruchita Manuja , Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date			
3. Role Allocations at the end of the Meeting			
Ruchita Manuja	Number of Flights for each carrier visualization		
Supraja Sridharan	Total Flights on daily basis		
Anirudh Sharma	Airport Performance Visualization		
Sundaram Thangaraj	Dataset Creation		
Abhijeet Date	Trend Analysis for Carriers, Airports and Airlines		

Meeting Number	**8 - Preparation of Visualization for Final Report**		
Date of Meeting:	25/Oct/17	**Minutes Prepared By:**	Supraja

1. Meeting Objective and Outcome

All the members met in SIT and discussed the outstanding tasks of each member.
We decided to finish all the allocated tasks before 31st of October

2. Attendance at Meeting

Ruchita Manuja, Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date

3. Role Allocations at the end of the Meeting

Ruchita Manuja	Number of Flights for each carrier visualization	
Supraja Sridharan	Total Flights on daily basis	
Anirudh Sharma	Airport Performance Visualization	
Sundaram Thangaraj	Dataset Creation	
Abhijeet Date	Trend Analysis for Carriers, Airports and Airlines	

Meeting Number	**9 -Revision of Final Report**

Date of Meeting:	1/Nov/17	**Minutes Prepared By:**	Ruchita

1. Meeting Objective and Outcome

All the members met in SIT and review all the entire report together. We consolidated all the codes and videos of the team members and prepared a zip file.

2. Attendance at Meeting

Ruchita Manuja , Supraja Sridharan, Anirudh Sharma, Sundaram Thangaraj, Abhijeet Date

Appendix 2 - Code & Videos

Following TeraData SQL queries were used for processing the data:

Query to extract monthly summary

```
select Month1, dayofmonth, count(*)      as TotalFlights,
sum(case when status='cancelled' then 1 else 0
end) as Cancelled, sum(case when status='On Time'
then 1 else 0 end) as ontime from
(
select    Month1, dayofmonth,
case when cancelled = '1' then 'cancelled'
       when cancelled = '0' and (arrdelay <= 15 and depdelay
            <=15) then 'On Time' end as status

from ipstaging.year_2007_fmt  a
) a
group by 1,2
order by 1,2
;
```

Query to extract airlines route details

```
SELECT 1998, origin, airport, count(*) as TotalFlights, SUM(CASE WHEN
cancelled='0' then 0 else 1 end) as CancelledFlights,
SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1
else 0 end) as DelayedFlights,
```

```
SUM(CASE cancelled '0' and ( <= 15 depdelay
WHEN = arrdelay and <= 15)

then else end
1 0 ) as OnTimeFlights

from ipstaging.YEAR_1998_Fmt
a

inner join
tduser.airport b


on a.origin=b.iata

group by 1,2,3;


SELEC 1999, origin, count(* as TotalFlights, SUM(CASE
T airport, ) WHEN

cancelled='0 then 0 1 as
' else end) CancelledFlights,

SUM(CASE cancelled '0' and ( > or depdelay >
WHEN = arrdelay 15 15) then 1


else as DelayedFlights
0 end) ,

SUM(CASE cancelled '0' and ( <= 15 depdelay
WHEN = arrdelay and <= 15)

then else end
1 0 ) as OnTimeFlights

from ipstaging.YEAR_1999_Fmt
a

inner join
tduser.airport b


on a.origin=b.iata
```

```
group by 1,2,3;


SELECT 2000, origin, airport, count(*) as TotalFlights, SUM(CASE WHEN cancelled='0' then 0 else 1 end) as CancelledFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1 else 0 end) as DelayedFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay <= 15 and depdelay <= 15) then 1 else 0 end) as OnTimeFlights

from tduser.YEAR_2000_Fmt a

inner join tduser.airport b

on a.origin=b.iata

group by 1,2,3;



SELECT 2001, origin, airport, count(*) as TotalFlights, SUM(CASE WHEN cancelled='0' then 0 else 1 end) as CancelledFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1
```

```
else          as DelayedFlights
0      end)   ,

SUM(CASE      cancelled  '0' and (      <= 15     depdelay
WHEN          =          arrdelay       and       <= 15)

then   else   end
1      0      )     as OnTimeFlights

from
tduser.YEAR_2001_Fmt      a

inner join
tduser.airport            b


on a.origin=b.iata

group by 1,2,3;


SELEC  2002, origin,          count(*   as TotalFlights, SUM(CASE
T      airport,               )         WHEN

cancelled='0   then 0     1       as
'              else       end)    CancelledFlights,

SUM(CASE      cancelled  '0' and (      >     or depdelay >
WHEN          =          arrdelay       15    15) then 1


else          as DelayedFlights
0      end)   ,

SUM(CASE      cancelled  '0' and (      <= 15     depdelay
WHEN          =          arrdelay       and       <= 15)

then   else   end
1      0      )     as OnTimeFlights

from
tduser.YEAR_2002_Fmt      a

inner join
tduser.airport            b
```

```
on a.origin=b.iata

group by 1,2,3;


SELECT 2003, origin, airport, count(*) as TotalFlights, SUM(CASE WHEN
cancelled='0' then 0 else 1 end) as CancelledFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1
else 0 end) as DelayedFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay <= 15 and depdelay <= 15)
then 1 else 0 end) as OnTimeFlights

from tduser.YEAR_2003_Fmt a

inner join tduser.airport b


on a.origin=b.iata

group by 1,2,3;


SELECT 2004, origin, airport, count(*) as TotalFlights, SUM(CASE WHEN
cancelled='0' then 0 else 1 end) as CancelledFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1
```

```
else          as DelayedFlights
0      end)   ,

SUM(CASE      cancelled  '0' and (          <= 15     depdelay
WHEN          =          arrdelay           and       <= 15)

then   else   end
1      0      )     as OnTimeFlights

from
tduser.YEAR_2004_Fmt      a

inner join
tduser.airport            b


on a.origin=b.iata

group by 1,2,3;


SELEC  2005, origin,          count(*   as TotalFlights, SUM(CASE
T      airport,               )         WHEN

cancelled='0   then 0    1       as
'              else      end)    CancelledFlights,

SUM(CASE      cancelled  '0' and (          >     or depdelay >
WHEN          =          arrdelay           15    15) then 1


else          as DelayedFlights
0      end)   ,

SUM(CASE      cancelled  '0' and (          <= 15     depdelay
WHEN          =          arrdelay           and       <= 15)

then   else   end
1      0      )     as OnTimeFlights

from
tduser.YEAR_2005_Fmt      a
```

```
inner join
tduser.airport              b


on a.origin=b.iata

group by 1,2,3;


SELEC   2006, origin,          count(*    as TotalFlights, SUM(CASE
T       airport,               )          WHEN

cancelled='0    then 0      1       as
'               else        end)    CancelledFlights,

SUM(CASE        cancelle    '0' and (           >    or depdelay >
WHEN            d =         arrdelay            15   15) then 1

else            as DelayedFlights
0       end)    ,

SUM(CASE        cancelle    '0' and (           <= 15     depdelay <=
WHEN            d =         arrdelay            and       15)

then    else    end
1       0       )     as OnTimeFlights


from
tduser.YEAR_2006_Fmt        a

inner join
tduser.airport              b


on a.origin=b.iata

group by 1,2,3;



SELEC   2007, origin,          count(*    as TotalFlights, SUM(CASE
T       airport,               )          WHEN

cancelled='0    then 0      1       as
'               else        end)    CancelledFlights,
```

```
SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1 else 0 end) as DelayedFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay <= 15 and depdelay <= 15) then 1 else 0 end) as OnTimeFlights

from ipstaging.YEAR_2007_Fmt a

inner join tduser.airport b

on a.origin=b.iata

group by 1,2,3;

SELECT 2008, origin, airport, count(*) as TotalFlights, SUM(CASE WHEN cancelled='0' then 0 else 1 end) as CancelledFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay > 15 or depdelay > 15) then 1 else 0 end) as DelayedFlights,

SUM(CASE WHEN cancelled = '0' and (arrdelay <= 15 and depdelay <= 15) then 1 else 0 end) as OnTimeFlights

from tduser.YEAR_2008_Fmt a
```

```
inner join
tduser.airport              b


on a.origin=b.iata

group by 1,2,3;
```

Query to extract carrier data

```
SELECT uniquecarrier,                                    as TotalFlights,
description,                              count(*)       SUM(CASE WHEN

                then 0       1
cancelled='0'   else         end)    as CancelledFlights,

                                                                  or depdelay
                cancelled    '0' and (              >       > 15) then
SUM(CASE WHEN   =            arrdelay               15      1

               as DelayedFlights
else 0  end)   ,

                cancelled    '0' and (              <= 15     depdelay
SUM(CASE WHEN   =            arrdelay               and       <= 15)

        else
then 1  0        end)    as OnTimeFlights


from ipstaging.YEAR_1999_Fmt a
left join
tduser.carrier b
on a.uniquecarrier = b.carriercode
group by 1,2
```

```
Teradata load script

.SESSIONS 5;
.LOGON 192.168.28.129/dbc,dbc;

DATABASE IPSTAGING;

BEGIN LOADING year_1998
ERRORFILES year_ERR1, year_ERR2;

SET RECORD VARTEXT ",";

DEFINE
Year1 (VARCHAR(255)),
Month1 (VARCHAR(255)),
DayofMonth (VARCHAR(255)),
DayOfWeek (VARCHAR(255)),
DepTime (VARCHAR(255)),
CRSDepTime (VARCHAR(255)),
ArrTime (VARCHAR(255)),
CRSArrTime (VARCHAR(255)),
UniqueCarrier (VARCHAR(255)),
FlightNum (VARCHAR(255)),
TailNum (VARCHAR(255)),
ActualElapsedTime (VARCHAR(255)),
```

```
CRSElapsedTime (VARCHAR(255)),
AirTime (VARCHAR(255)),
ArrDelay (VARCHAR(255)),
DepDelay (VARCHAR(255)),
Origin (VARCHAR(255)),
Dest (VARCHAR(255)),
Distance (VARCHAR(255)),
TaxiIn (VARCHAR(255)),
TaxiOut (VARCHAR(255)),
Cancelled (VARCHAR(255)),
CancellationCode (VARCHAR(255)),
Diverted (VARCHAR(255)),
CarrierDelay (VARCHAR(255)),
WeatherDelay (VARCHAR(255)),
NASDelay (VARCHAR(255)),
SecurityDelay (VARCHAR(255)),
LateAircraftDelay (VARCHAR(255))
FILE=1998.csv;

INSERT INTO IPSTAGING.year_1998
                (Year1, Month1, DayofMonth, DayOfWeek, DepTime,
                                CRSDepTime, ArrTime, CRSArrTime,
                                UniqueCarrier, FlightNum, TailNum,
                                ActualElapsedTime, CRSElapsedTime, AirTime,
                                ArrDelay, DepDelay, Origin, Dest, Distance, TaxiIn,
                                TaxiOut, Cancelled, CancellationCode, Diverted,
                                CarrierDelay, WeatherDelay, NASDelay,
                                SecurityDelay, LateAircraftDelay)
```

```
VALUES
(:Year1,:Month1,:DayofMonth,:DayOfWeek,:DepTime,:CRSDepTime,:ArrTime,
                :CRSArrTime,:UniqueCarrier,:FlightNum,:TailNum,:ActualElapsedTime,
                :CRSElapsedTime,:AirTime,:ArrDelay,:DepDelay,:Origin,:Dest,:Distance,
                :TaxiIn,:TaxiOut,:Cancelled,:CancellationCode,:Diverted,:CarrierDelay,
                :WeatherDelay,:NASDelay,:SecurityDelay,:LateAircraftDelay
                );


END LOADING;
LOGOFF;
```

```
Routes data Extract


insert into ipstaging.routes_ranking
select 2002, origin, origin_airport, origin_city,
origin_state, origin_country, origin_lat,
origin_longt,
dest, dest_airport, dest_city, dest_state,
dest_country, dest_lat, dest_longt,
sum(case when status='cancelled' then 1 else 0 end) as
Cancelled, sum(case when status='On Time departure'
then 1 else 0 end) as ontimedeparture,
sum(case when status='Small departure delay' then 1
else 0 end) as smalldeparturedelay


from
(
select     a.origin, b.airport as origin_airport , b.city as
origin_city,
b.state as origin_state, b.country as origin_country, b.lat as
origin_lat, b.longt as origin_longt,
a.dest, c.airport as dest_airport, c.city as dest_city, c.state as
dest_state, c.country as dest_country, c.lat as dest_lat, c.longt as

dest_longt, case when cancelled = '1' then 'cancelled'
 when cancelled = '0' and (arrdelay <=15 and
      depdelay <= 15 )                                          then
Time departure '
```

```
        when      cancelled = '0' and (depdelay >15      or depdelay >
        15 )
'Small departure delay'
else 'no staus'
end as status    from tduser.year_2002_fmt   a

 'On

then

inner join tduser.airport b
on a.origin=b.iata
inner join tduser.airport c
on a.dest=c.iata
) a
group by 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15;



insert into ipstaging.routes_ranking_1
select year1, origin, origin_airport, origin_city,
origin_state, origin_country, origin_lat, origin_longt,
dest, dest_airport, dest_city, dest_state,
dest_country, dest_lat, dest_longt,
 Cancelled,
ontimedeparture,
 smalldepartureddelay,
 cancelled+ontimedeparture+smalldepartureddelay as totalflights,
 row_number()over( partition by year1 order by totalflights desc )
 as ranking from ipstaging.routes_ranking;



insert into ipstaging.routes_ranking_2
SELECT     year1, ranking , 4    as step , 'Destination' , dest,
origin_airport,
origin|| '_'||dest as pathid , origin_lat, origin_longt,
      cancelled,
ontimedeparture,      smalldepartureddelay
from ipstaging.routes_ranking_1
union all
SELECT year1, ranking , 1 as step , 'Origin' , origin,
origin_airport, origin|| '_'||origin as pathid , origin_lat,
origin_longt, 0 cancelled, 0 ontimedeparture, 0
smalldepartureddelay from ipstaging.routes_ranking_1

union all
SELECT year1 , ranking, 2 as step, 'Destination' , origin,
origin_airport, origin|| '_'||origin as pathid , origin_lat,
origin_longt, 0 cancelled, 0 ontimedeparture, 0
smalldepartureddelay from ipstaging.routes_ranking_1
```

```
union all
SELECT year1, ranking , 3 as step , 'Origin' , origin,
origin_airport, origin|| '_'||dest as pathid , origin_lat,
origin_longt, 0 cancelled, 0 ontimedeparture, 0
smalldepartureddelay from ipstaging.routes_ranking_1
```

Following code files have been submitted to supplement this report :

1. **Number of flights per Carrier (D3 Donut Chart) :**
   This folder contains the following files:
   - CarrierDetails.html
   - index.js
   - index.css
   - 10 csv files for Carrier data from 1998 -2008

2. **Performance of top 5 Airlines (Google Chart) :**
   - Airline Performance.html
   - airlines.css

In addition to the above codes files a few video files have also been submitted as below:

- Flights Visualisation.wmv / Flights Visualisation.mpg
- Airline Routes_Chicago O'Hare International Routes (2003-2008),mp4
- Airline Routes_LaGuardia Routes (2003-2008).mp4
- Airline Routes_Salt Lake City International Routes (2003-2008).mp4
- Airport Performance by State based on Cancellation Count (2003-2008).mp4
- Airport Performance by State based on Count of Major Delay (2003-2008).mp4
- Airport Performance by State based on Count of on-time flights (2003-2008).mp4