Module Title: B9DA102 Data Storage Solutions for Data Analytics

Technical Report

for

Flight On-Time Performance

**BY:**

**Akshay Kalmetar (10525530)**

**Jishnu Vattakattil (10522789)**

**Shandilya Sutar (10526703)**

**Srikanth Tangirala (10522903)**

Contents

[**1. INTRODUCTION** 3](#_Toc27420977)

[**1.1 Reason for selecting the Subject Area** 3](#_Toc27420978)

[**1.2 Vision and Goals** 3](#_Toc27420979)

[**1.3 Key Stakeholders** 3](#_Toc27420980)

[**1.4 Business Requirements** 3](#_Toc27420981)

[**2. SCHEMA** 4](#_Toc27420982)

[**3. ETL (Extract, transform, load)** 5](#_Toc27420983)

[**4. VISUALIZATIONS AND REPORTS** 9](#_Toc27420984)

[**4.1 VISUALIZATIONS** 9](#_Toc27420985)

[**4.2 Reports** 13](#_Toc27420986)

[**4.2.1 Airline Cancellation Report:** 13](#_Toc27420987)

[**4.2.2 Airline Cancellation Monthly Report** 13](#_Toc27420988)

[**4.2.3 Airline Diversion Report** 14](#_Toc27420989)

[**4.2.4 Airline on-time percentages** 14](#_Toc27420990)

[**4.2.5 Airline Quarterly Delay Report** 15](#_Toc27420991)

[**7. CONCLUSIONS** 20](#_Toc27420992)

[**8.** **BIBLIOGRAPHY** 21](#_Toc27420993)

[**APPENDIX A –VISUALIZATIONS CODE** 22](#_Toc27420994)

[**APPENDIX B – NEO4J CODE** 24](#_Toc27420995)

# **1. INTRODUCTION**

## **1.1 Reason for selecting the Subject Area**

As the civil aviation industry is growing, the traffic in airspace is also growing. Due to this the airspace is becoming crowded. This crowdedness is causing flight delays in most major airports across the world. This situation is causing problems for airports, airlines as well as passengers. Every year, many flights get delayed which involves a lot of cost for the airlines and the passengers. This causes passengers time and money; the airline’s reputation is at stake.

## **1.2 Vision and Goals**

Nowadays delay is considered as one of the most important parameter to judge an “Airline”. There are many reasons which can cause delay such as weather conditions, air trafficking and many more, but we can take on this delay issue by considering process improvement. Hence, statistics of Flight delays becomes crucial factor in understanding the airlines working. Our study shows presents the analysis of flight delay data for the city of New York from USA for the year 2018. We have identified various factors responsible for and associated with flight delays for different airlines.

## **1.3 Key Stakeholders**

The major stakeholders for this assignment are passengers, airport officials, crew members, etc. who get directly associated to the airline industry and the data analysts who work on these datasets for analysis.

## **1.4 Business Requirements**

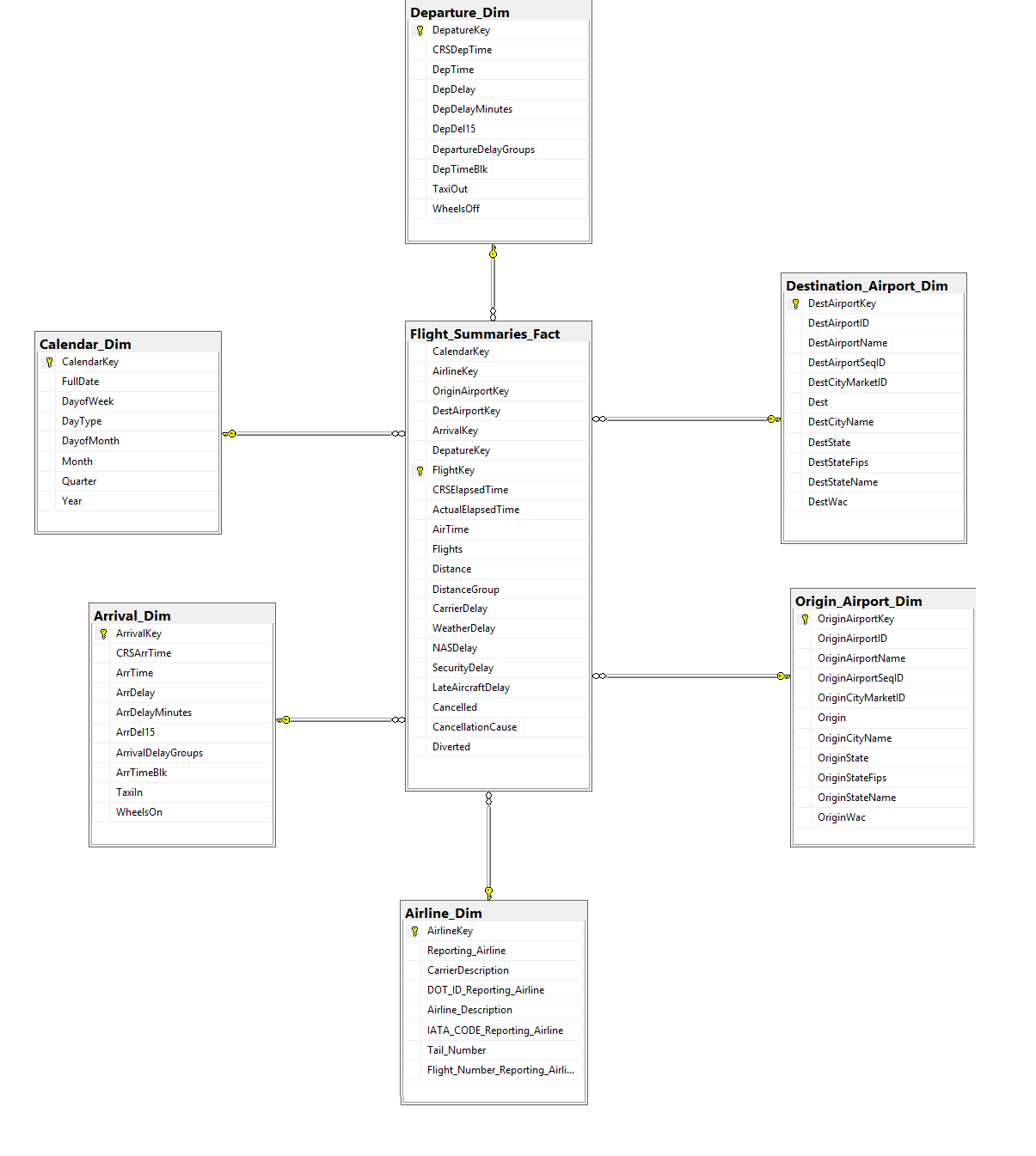
According to the 2018 report made by the US Federal Aviation Administration, the economic price of domestic flight delays entails a yearly cost of 28 Billion dollars to the passengers, lost demand and indirect costs. Most of these loses are beard by the passengers. They not only lose time while waiting for their flights, but they generally tend to miss the connecting flights, spend money on the living expenses during the transition period i.e. Food and Hotel.

*But what are the causes of this delay?*

After carrying out detailed study on the dataset, we made some sketches which could be represented visually using visualization tools. We built different types of sketches to observe the trends about the delays based on different types of delay Causes and Airlines. With the help of Visualization tools, we implemented our designs from the above prepared sketches. We used R, Neo4j and Visual Studio for the purpose of Visualization which made it easy and readable. We made sketches on below business requirement:

1. Number of flights cancelled during each month for the year 2018.
2. Calculate the arrival delay for all the airlines coming to New York during 2018.
3. Find out the Delay Reasons for all the flights during 2018.

# **2. SCHEMA**



**Fig: 2.1 Star Schema for Airline Datawarehouse**

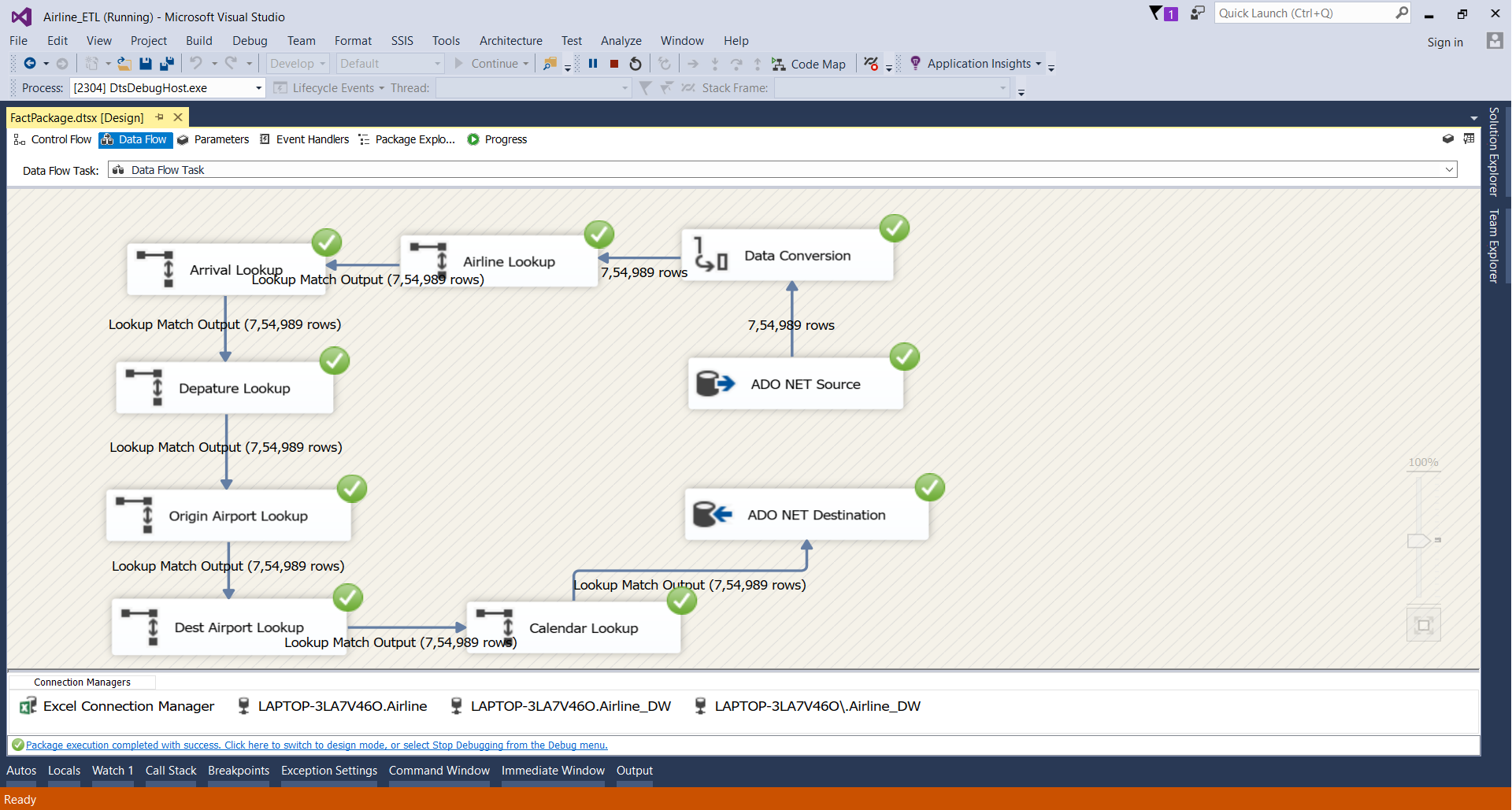
In this schema we tried to create a Fact Table for collecting data regarding Delay Causes occurring on different Airlines at New York. The main dimensions in the given database are:

1. Airline\_Dim
2. Departure\_Dim
3. Arrival\_Dim
4. Calendar\_Dim
5. Origin\_Airport\_Dim
6. Destination\_Airport\_Dim
7. Flight\_Summaries\_Fact

Flight\_Summaries\_Fact is the fact table for the above given dimensions.

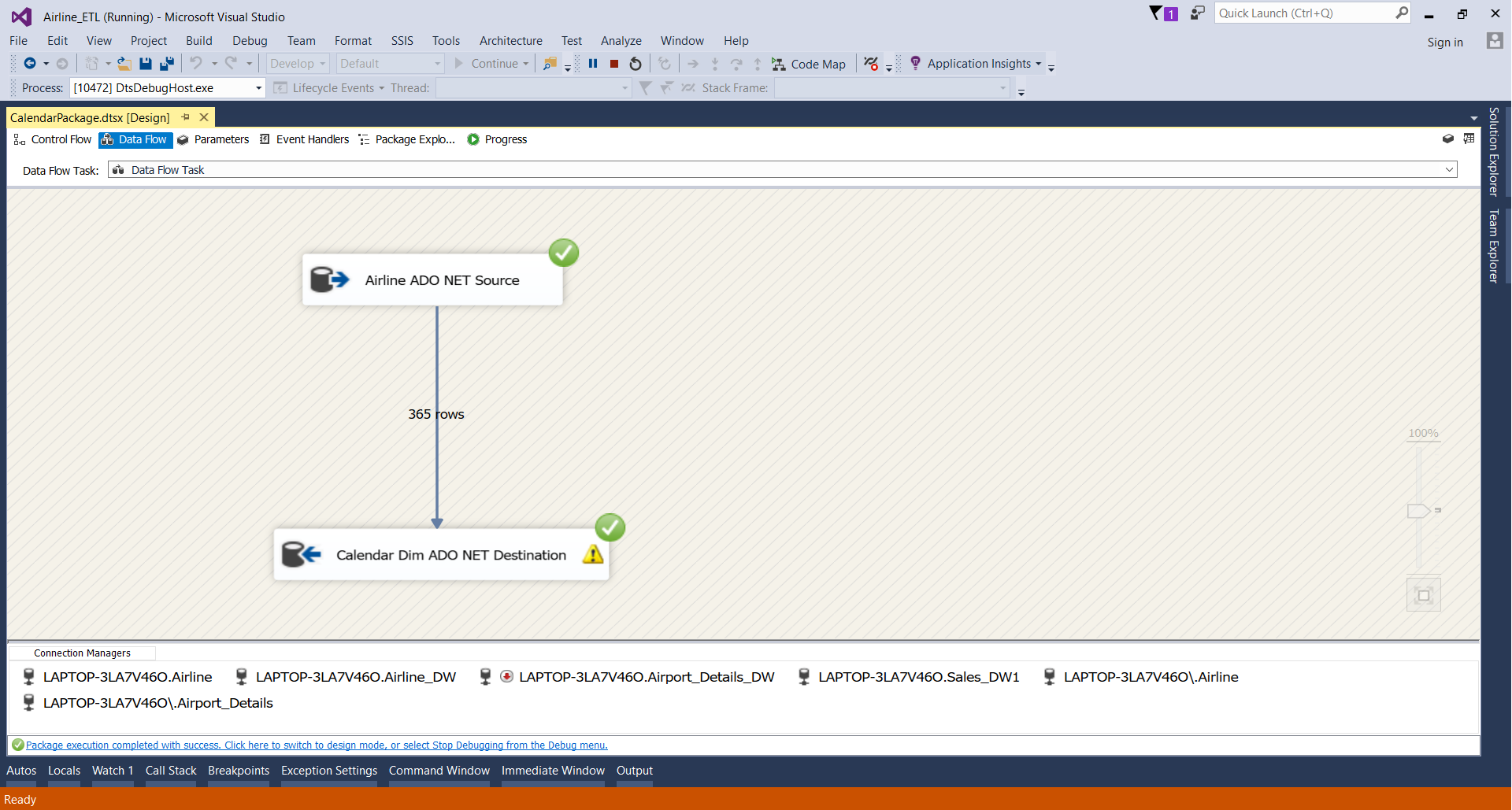
# **3. ETL (Extract, transform, load)**

ETL is basically the process in Data Warehousing and it does the process of Extracting, Transforming and Loading the data. It is one of the methods to move data from various sources into Data warehouse.



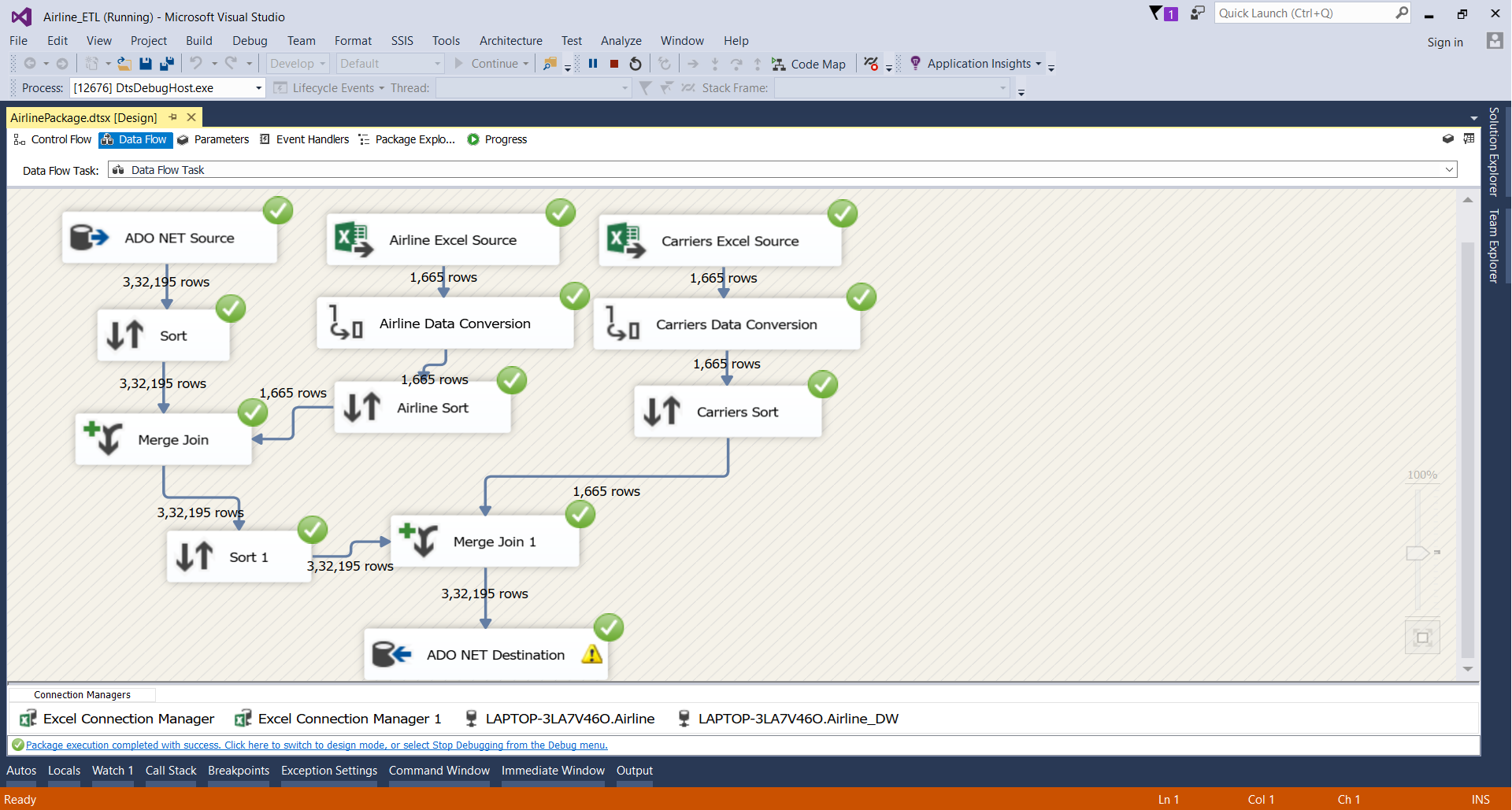
**Fig 3.1 Fact Table ETL Diagram**

We took the data for the year 2018 as the website contains the real time data from 1975 to present. As the amount of data was huge, we downloaded the data for only 2018. The Source data consist of various Lookup Tables. We took that data from SQL Server and combined the other packages to form Fact Table. Fact table is formed by combining all the lookup tables. The lookup tables are as follows: Arrival, Airline, Departure, Origin Airport, Destination Airport, and Calendar.

****

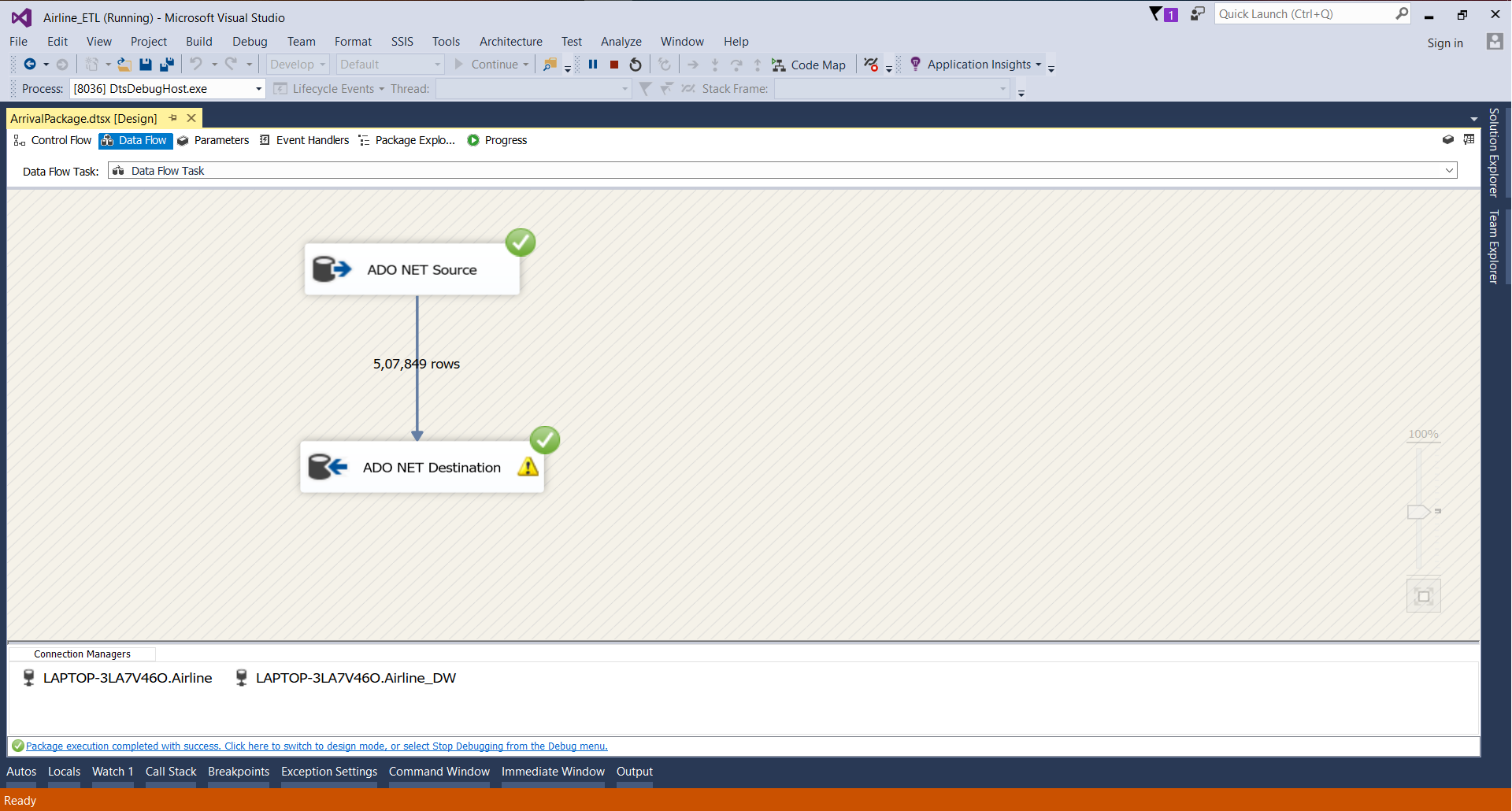
**Fig 3.2 Calendar Dimension ETL Diagram**

The Calendar is one of the Dimensions which form the Airline Fact data. The data for this directly comes from the SQL Server.

****

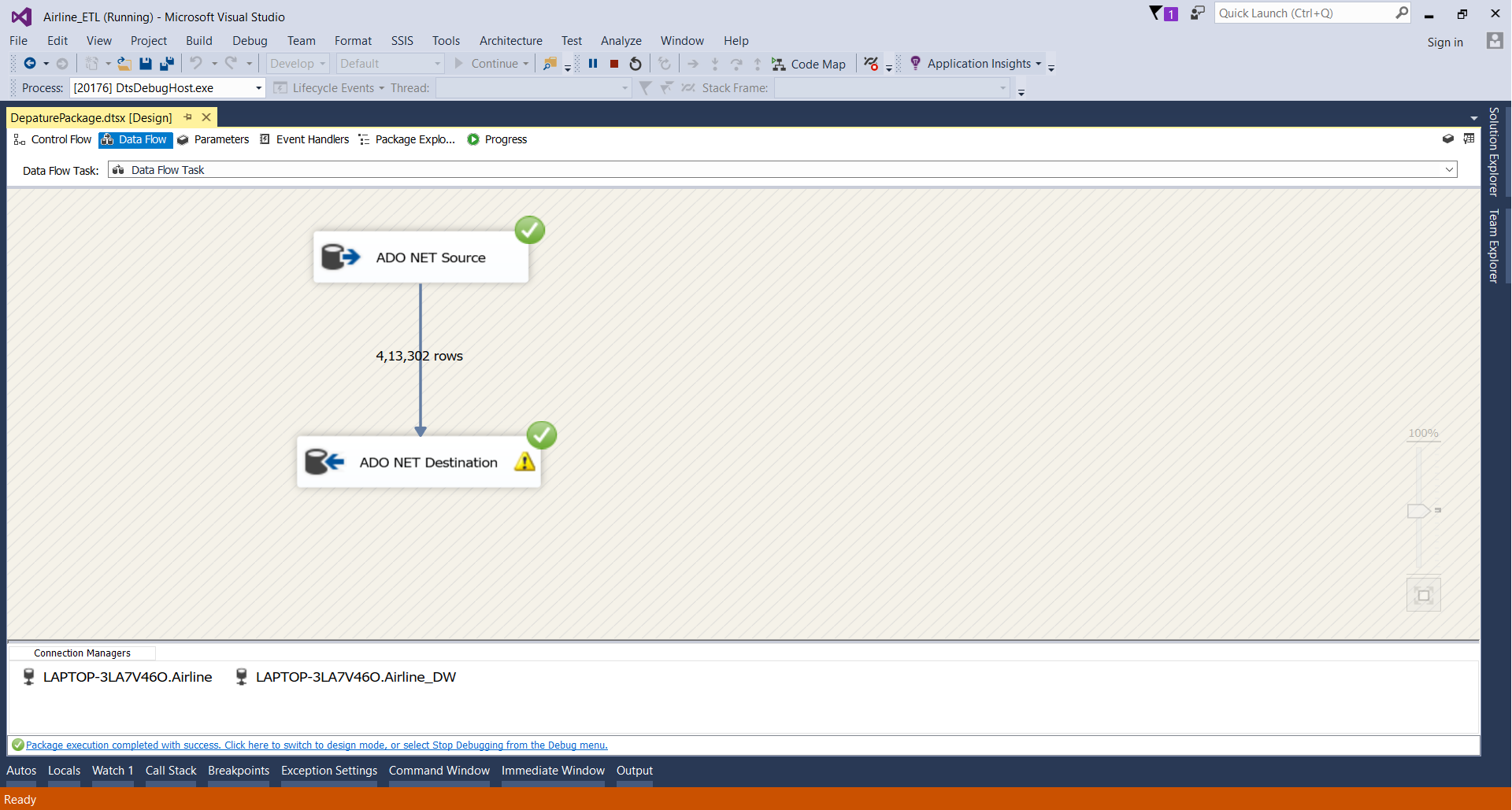
**Fig 3.3 Airline Dimension ETL Diagram**

Airline Dimension is formed by taking the data from the SQL server as well as Excel Source. The Excel source is used for taking data for airline information and the Carrier information. The data taken is then sorted and merged. The merger takes place between the SQL data and the Excel data. This data is then stored in the ADO NET Destination.

****

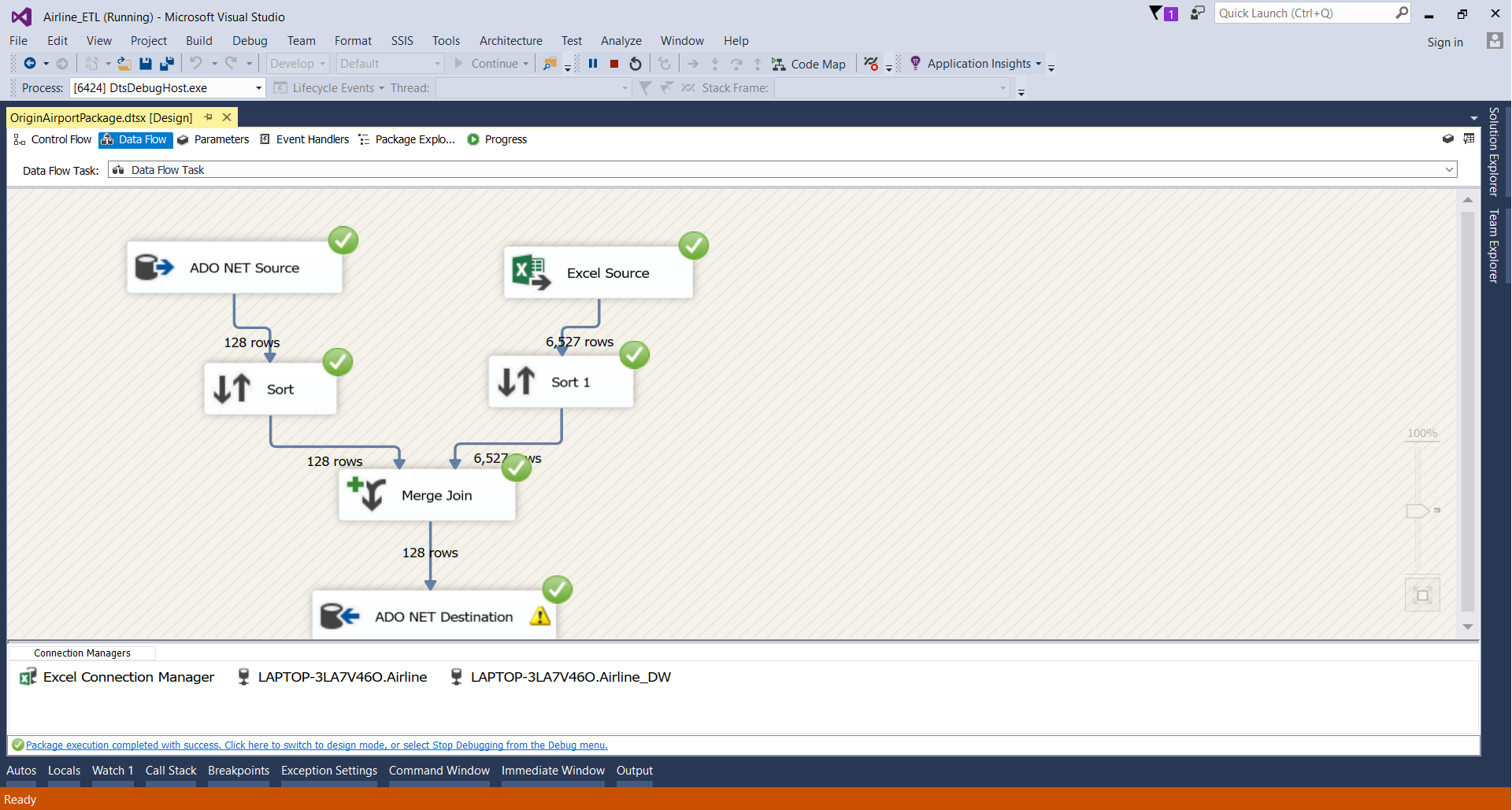
**Fig 3.4 Arrival Dimension ETL Diagram**

The Data for Arrival Dimension is taken from the ADO NET source and directly stored into the Destination Database.

****

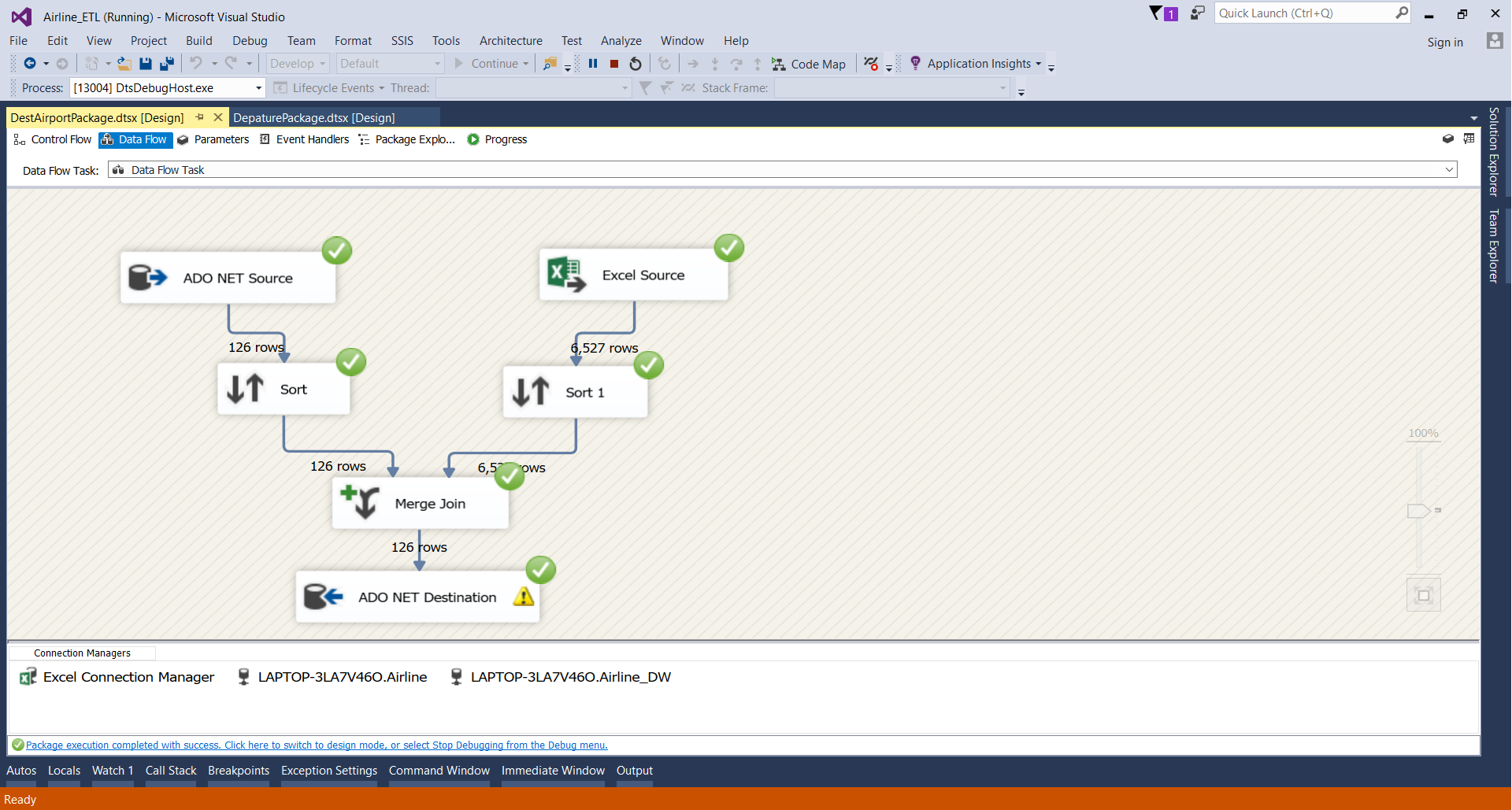
**Fig 3.4 Departure Dimension ETL Diagram**

The Data for Departure Dimension is taken from the ADO NET source and directly stored into the Destination Database.

****

**Fig 3.5 Origin Airport Dimension ETL Diagram**

For Origin Airport data, we take data from the SQL Server as well as the Excel Source then merge them and then into the Destination Database.

****

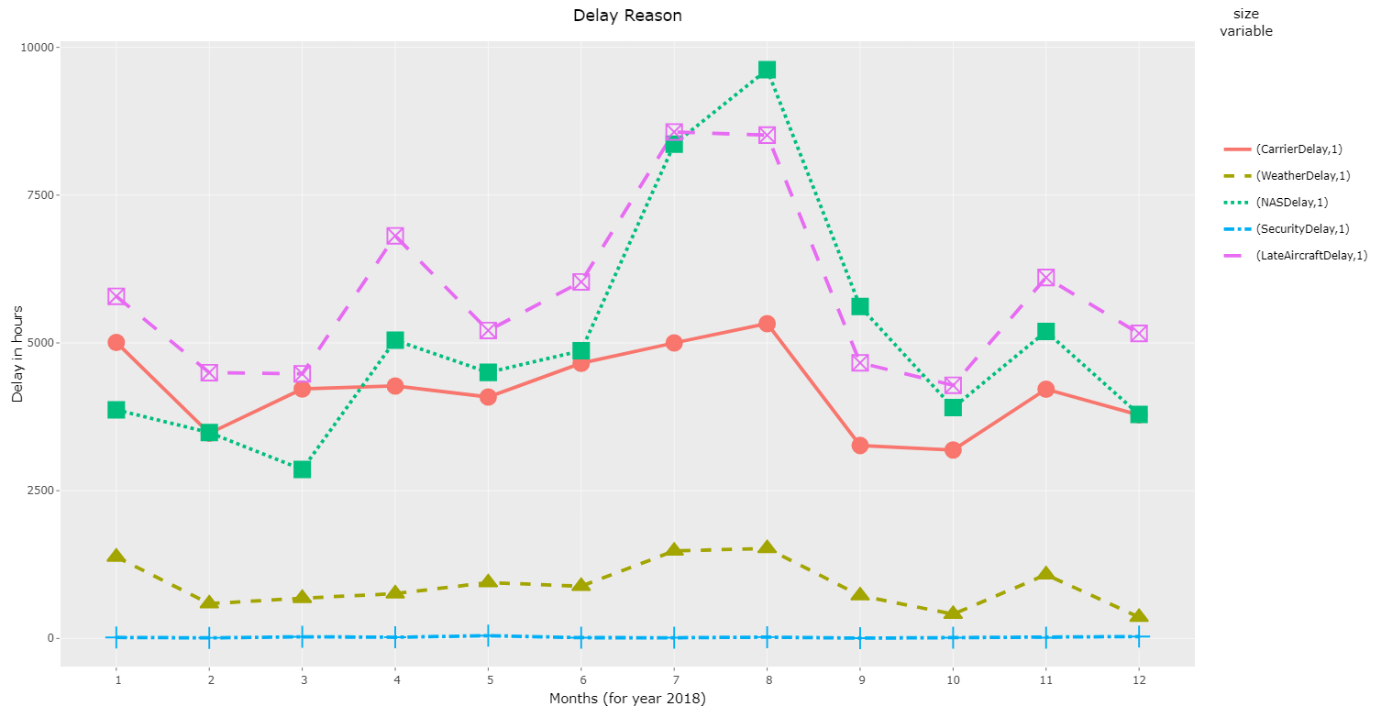
**Fig 3.6 Destination Airport Dimension ETL Diagram**

For Destination Airport data, we take data from the SQL Server as well as the Excel Source then merge them and then into the Destination Database.

# **4. VISUALIZATIONS AND REPORTS**

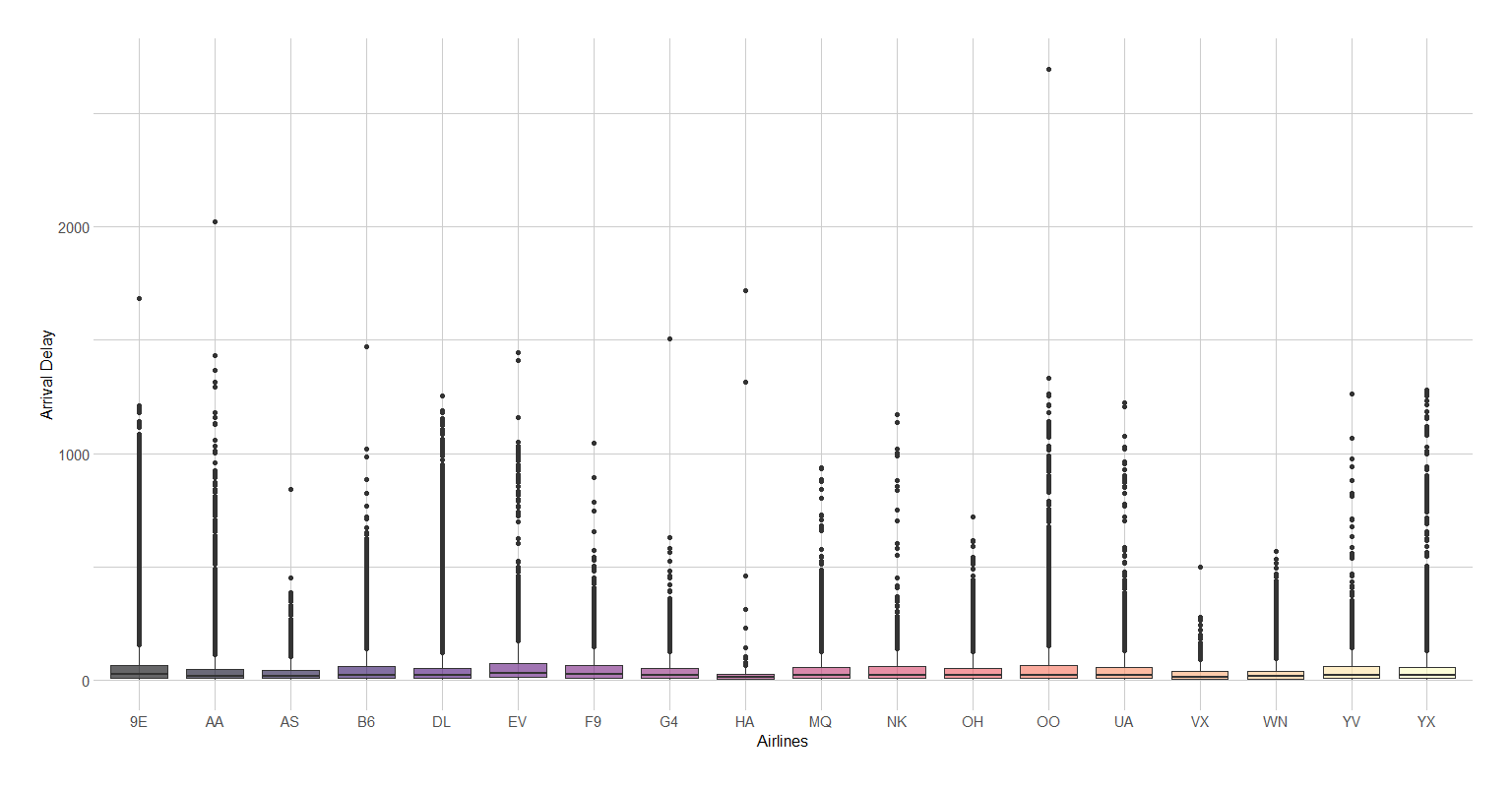
## **4.1 VISUALIZATIONS**

As the volume of data is increasing day by day, it is impossible to tell stories without visualizations. Data visualization is an art of how to turn the given numbers to useful facts. We used R as it offers inbuilt functions and libraries to build visualizations and present data. For our visualization we have used Bar Graph, Box Plot, Jitter Plot and Line Chart.

****

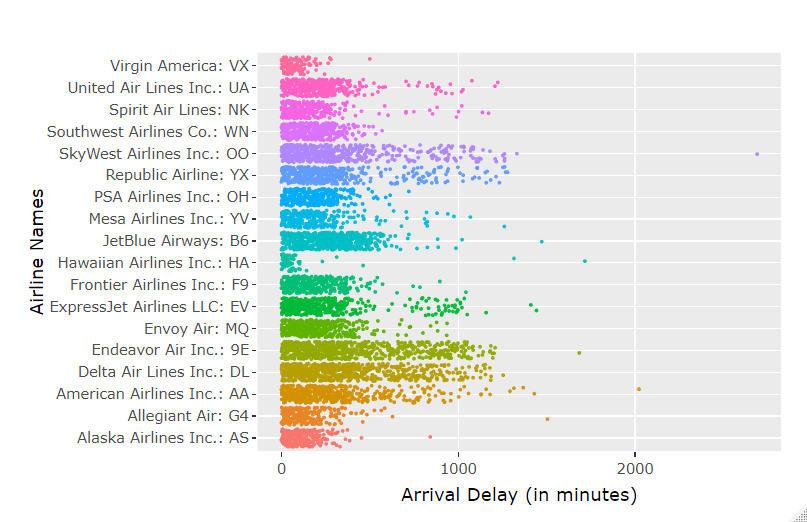
**Fig 4.1.1 Line Chart of Delay Reason**

We used the line charts to know which type of delay is highest during particular period. This would help to find out who is responsible for the delay. We can see that delay caused by NAS and Security tend to be higher during month of August.

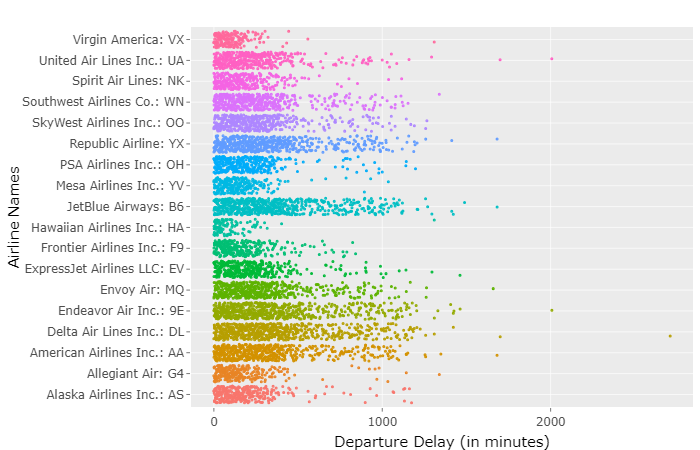
****

**Fig 4.1.2 Box Plot of Arrival Delay**

With the use of Box Plot, we tried to show the arrival delay for various flights in the US. We observed that Hawaiian Airlines (HA) is the best performing airlines in the comparison for Arrival

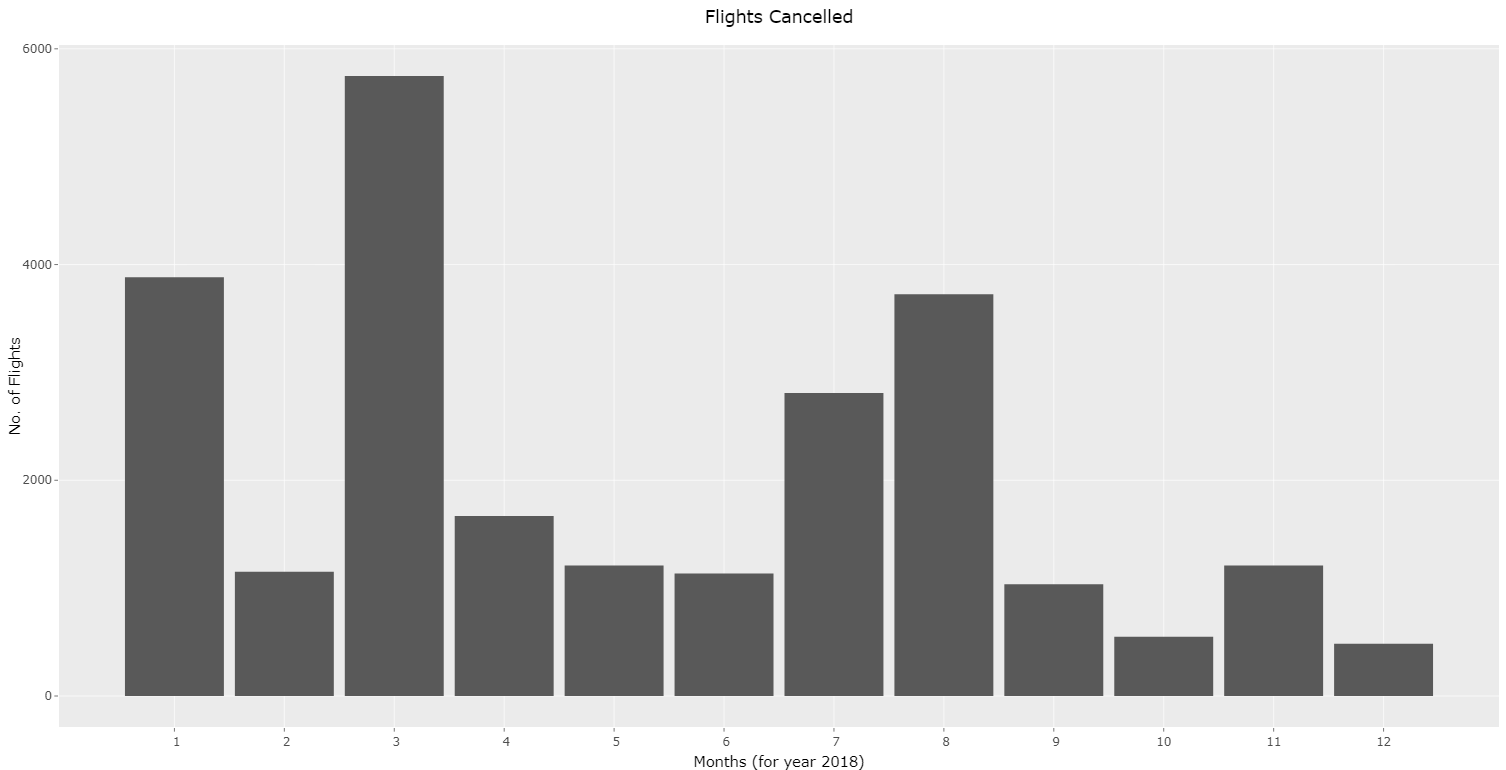
****

**Fig 4.1.3 Jitter Plot of Arrival Delay**

****

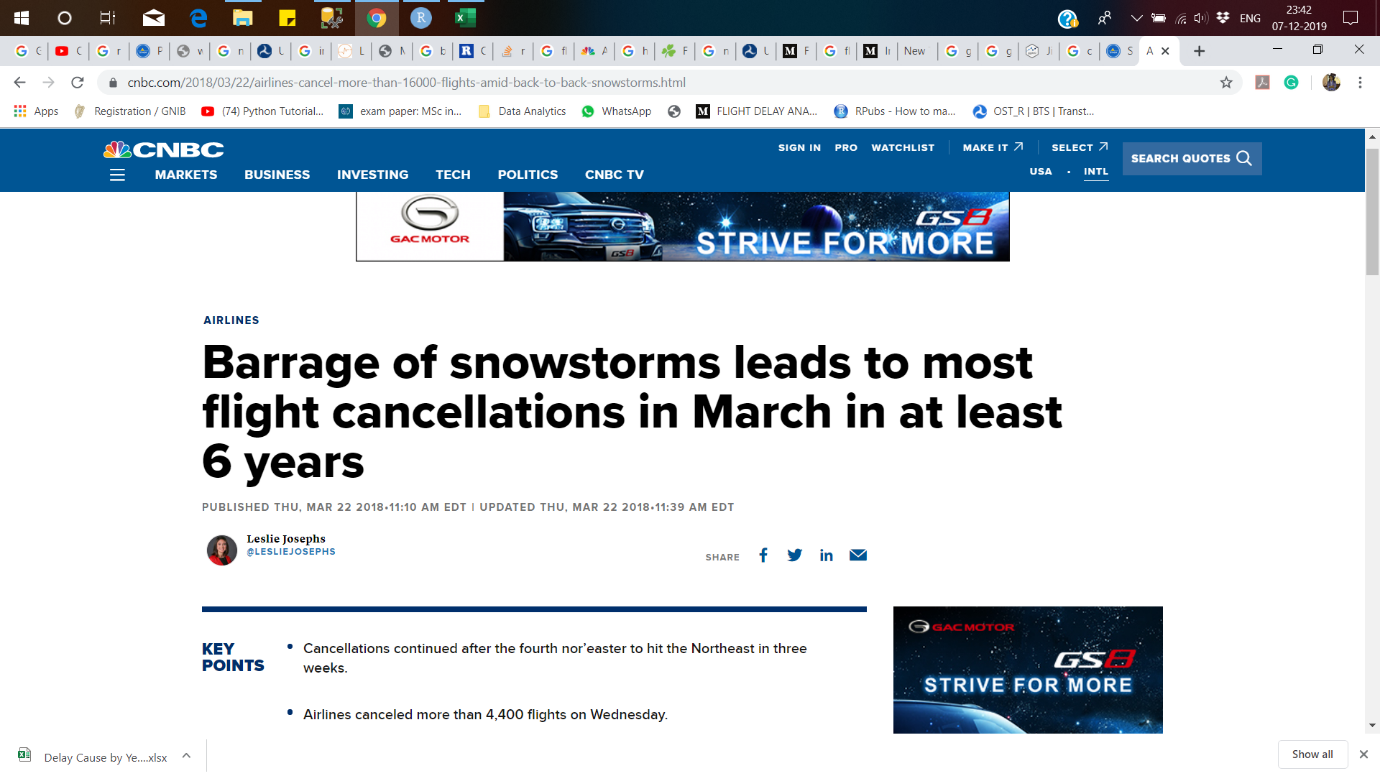
**Fig 4.1.4 Jitter Plot of Departure Delay**

Both above jitter plots show the census of all the delays that were measured in January 2018.This plot gives us the dispersion of data as compared to the box plot details. We can see that occasionally we can face large delays that can reach a few tens of hours.



**Fig 4.1.5 Bar Graph of Flights Cancelled**

We used the Bar Graph to find out no of Flights that got cancelled during the year 2018. We can see that during the month of March the No of flights that got cancelled is highest. The reason for this was Snowstorm that hit during that time period.

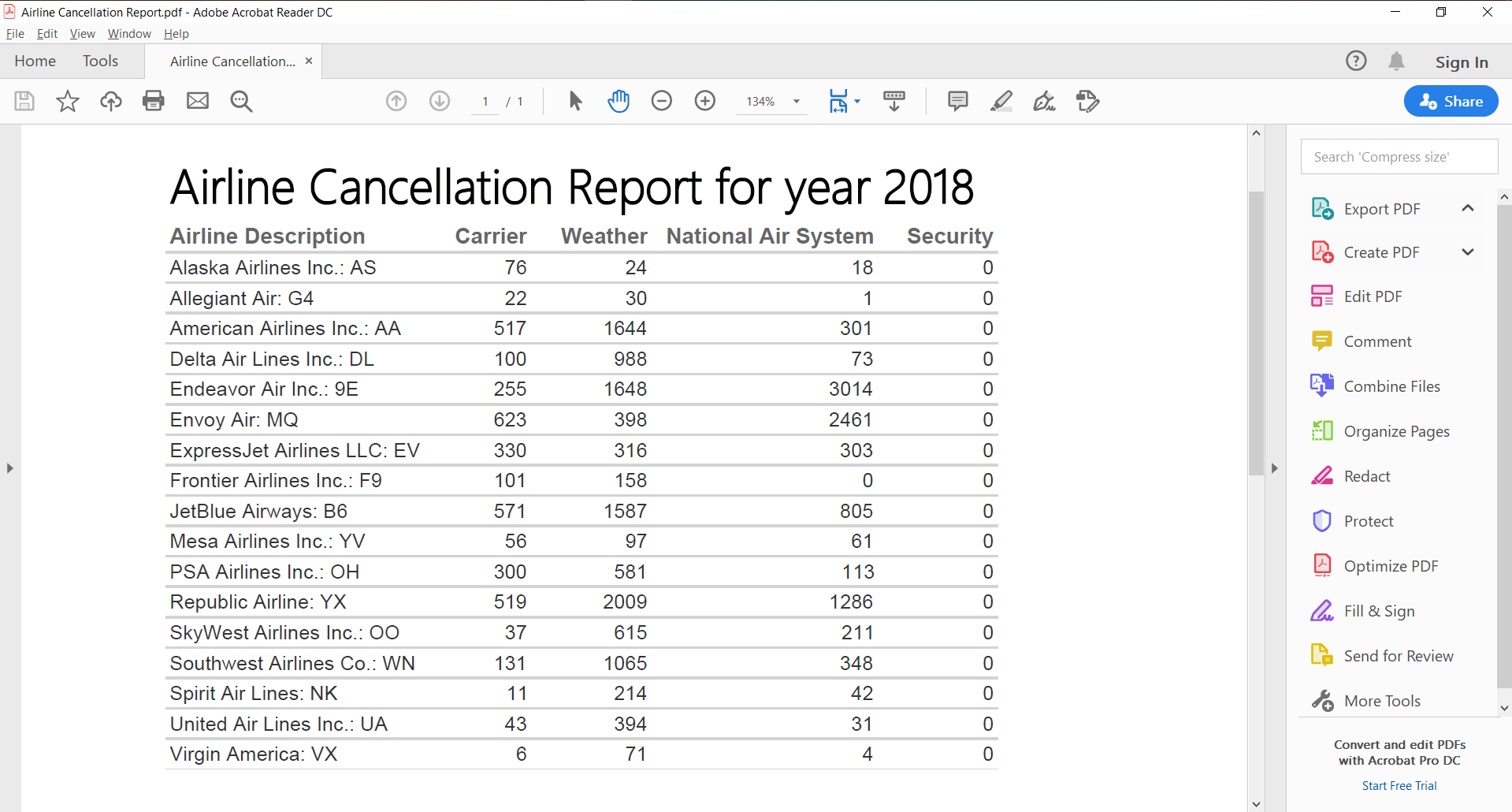


**Fig 4.1.6 Reason for huge number of flights cancelled**

# **4.2 Reports**

## **4.2.1 Airline Cancellation Report:**

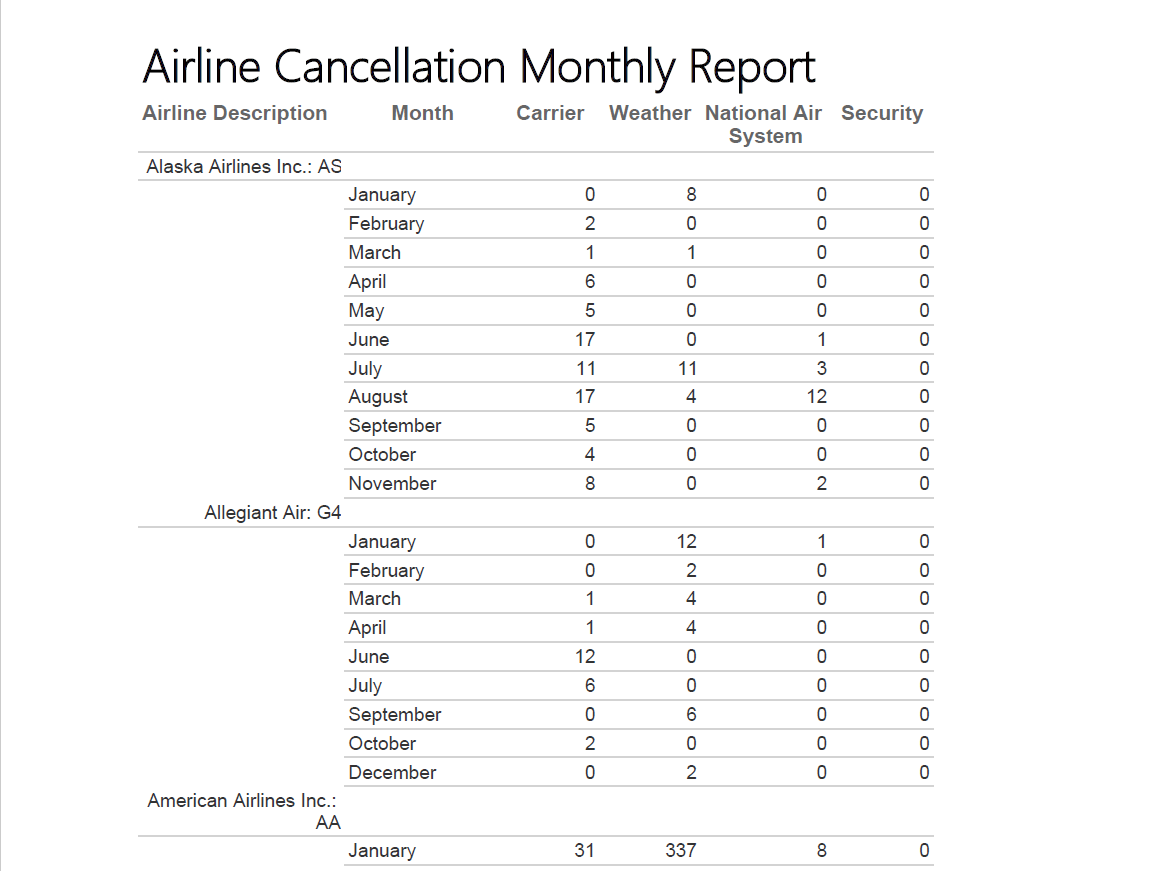
This Report tells us about the Cancellation data for the year 2018.It contains data about various flights and factors leading to the cancellation.



**Fig 4.2.1 Airline Cancellation Report**

## **4.2.2 Airline Cancellation Monthly Report**

This is a drilldown report for the above cancellation report. This report displays Airline’s monthly cancellation.



**Fig 4.2.2 Airline Cancellation Monthly Report**

## **4.2.3 Airline Diversion Report**

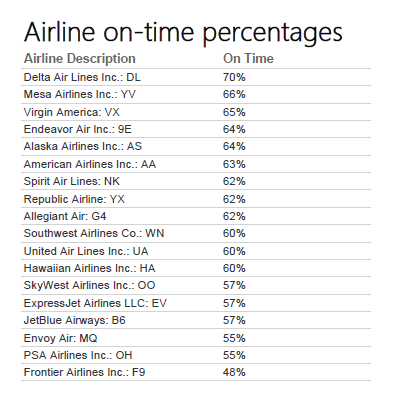
This Report tells us about the Airlines which were diverted during the year 2018.It contains data about count of various diverted flights.

****

**Fig 4.2.3 Report on Airline Diversion 2018**

## **4.2.4 Airline on-time percentages**

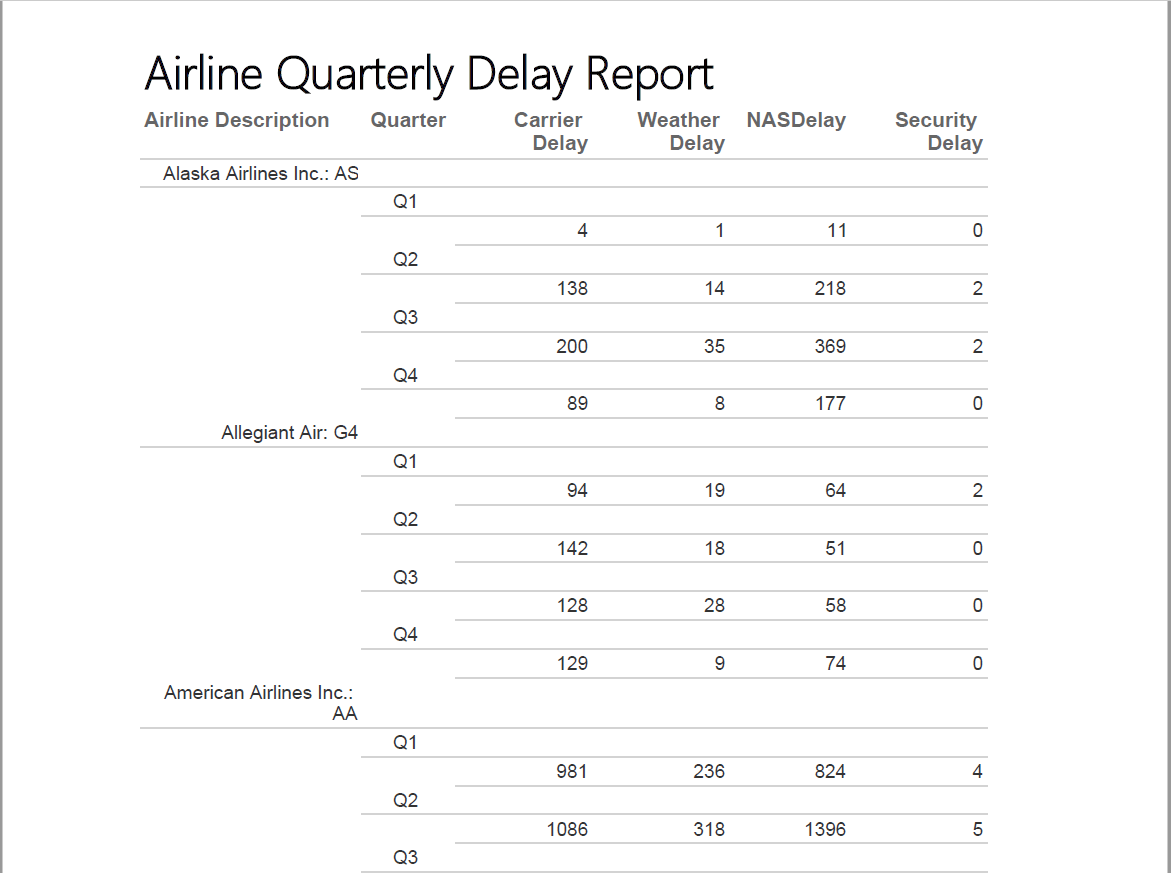
This Report tells us about the on-time arrival percentage of Flights coming and going from New York during the year 2018.It contains data about various flights and On-Time Arrival Percentage.

****

**Fig 4.2.4 Report on Airline On-time Percentage**

## **4.2.5 Airline Quarterly Delay Report**

This is a drill down report which displays Airline delays for each Quarter. All the different type of delays is displayed in minutes for each Quarter.



**Fig 4.2.5 Report on Airline On-time Percentage**

**5. INCLUDE XML AND SCHEMA**

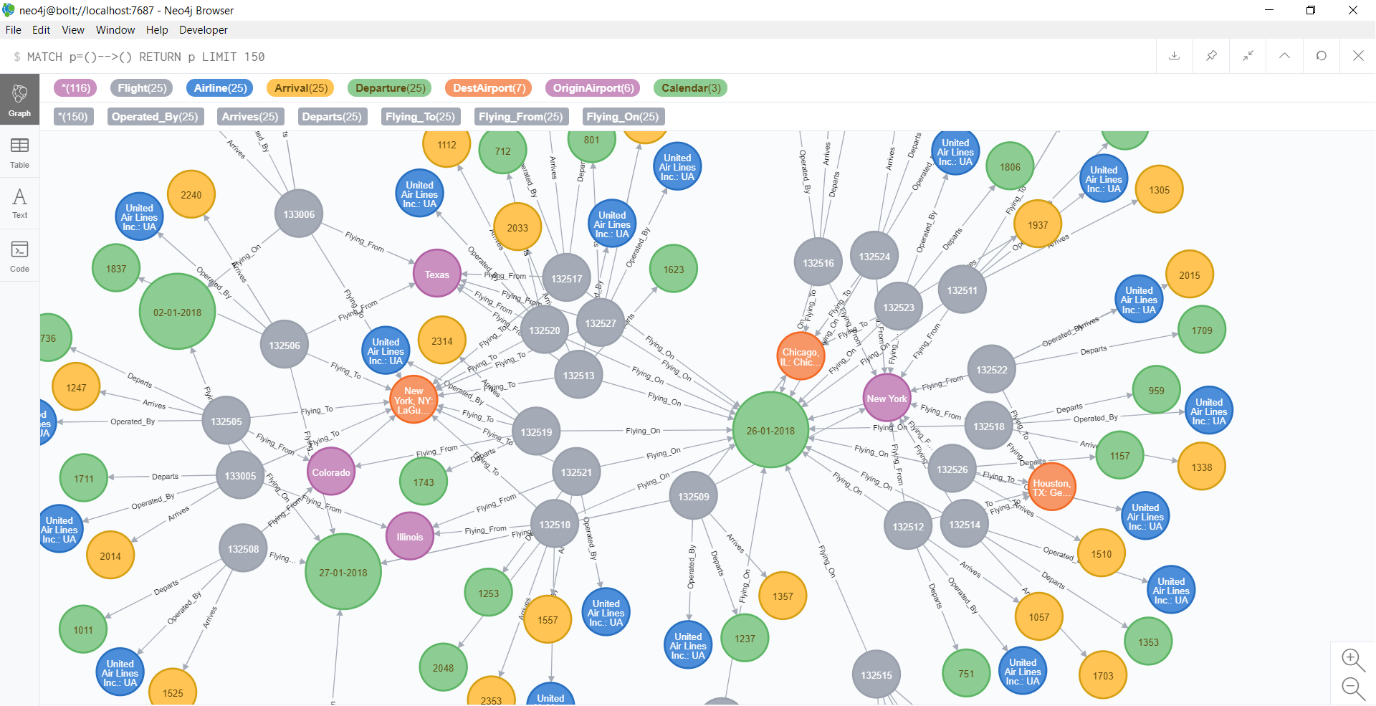
XML (Extensible Markup Language) that helps us to create our own customized tags to provide functionality not available with HTML. The XML for the Flight Delay Analysis is attached below.



**6. GRAPH DATABASES**

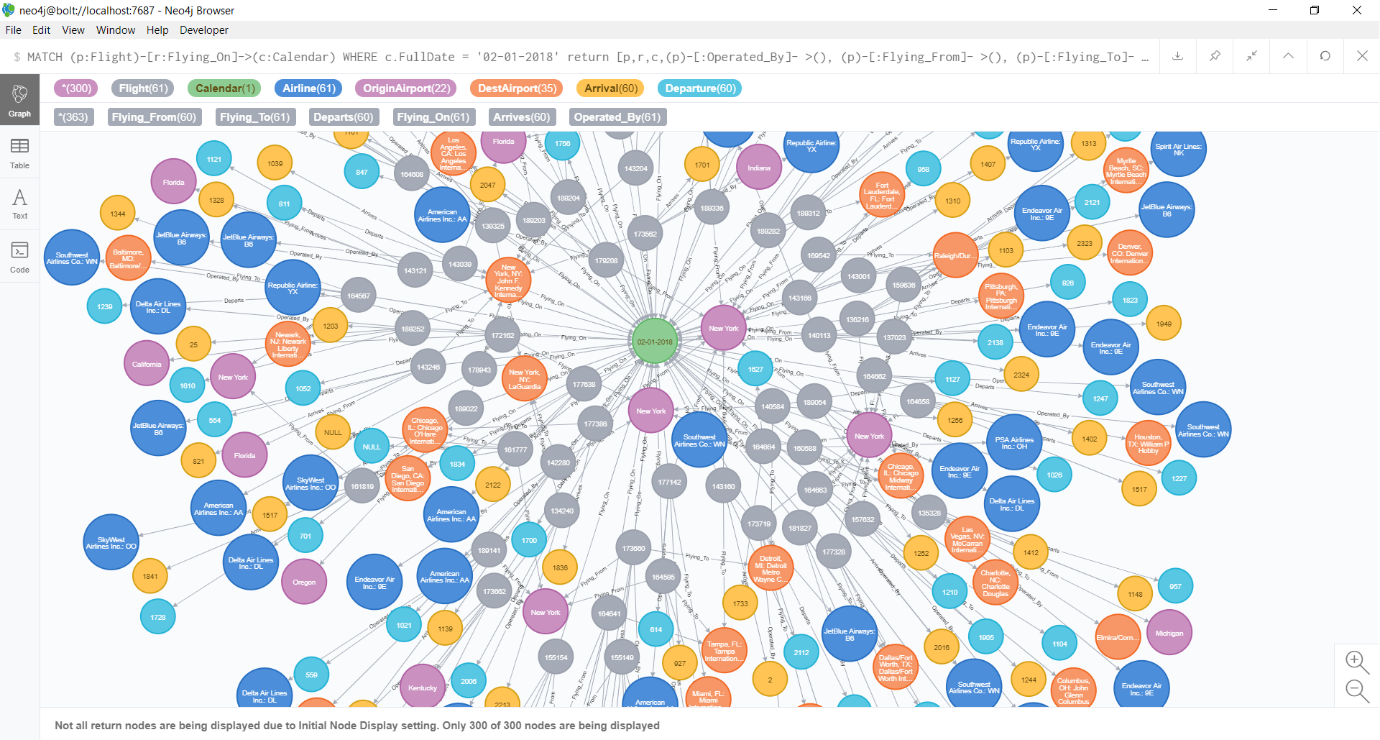
Using Neo4j we were able to represent data in a more graphical and interactive way. Neo4j Graph is a collection of objects related to each other. These objects are called vertices and the relationship between them is called edge. The other advantage of Graph database is that it handles scattered data much more user friendly than the relational database. As compared to relational databases graph databases consider relationships as separate objects which help in achieving good performance and this increases the readability of the model. In Graph database we use Cypher as a language as it is a declarative.

We have visualised our data warehouse as in the below diagram:

****

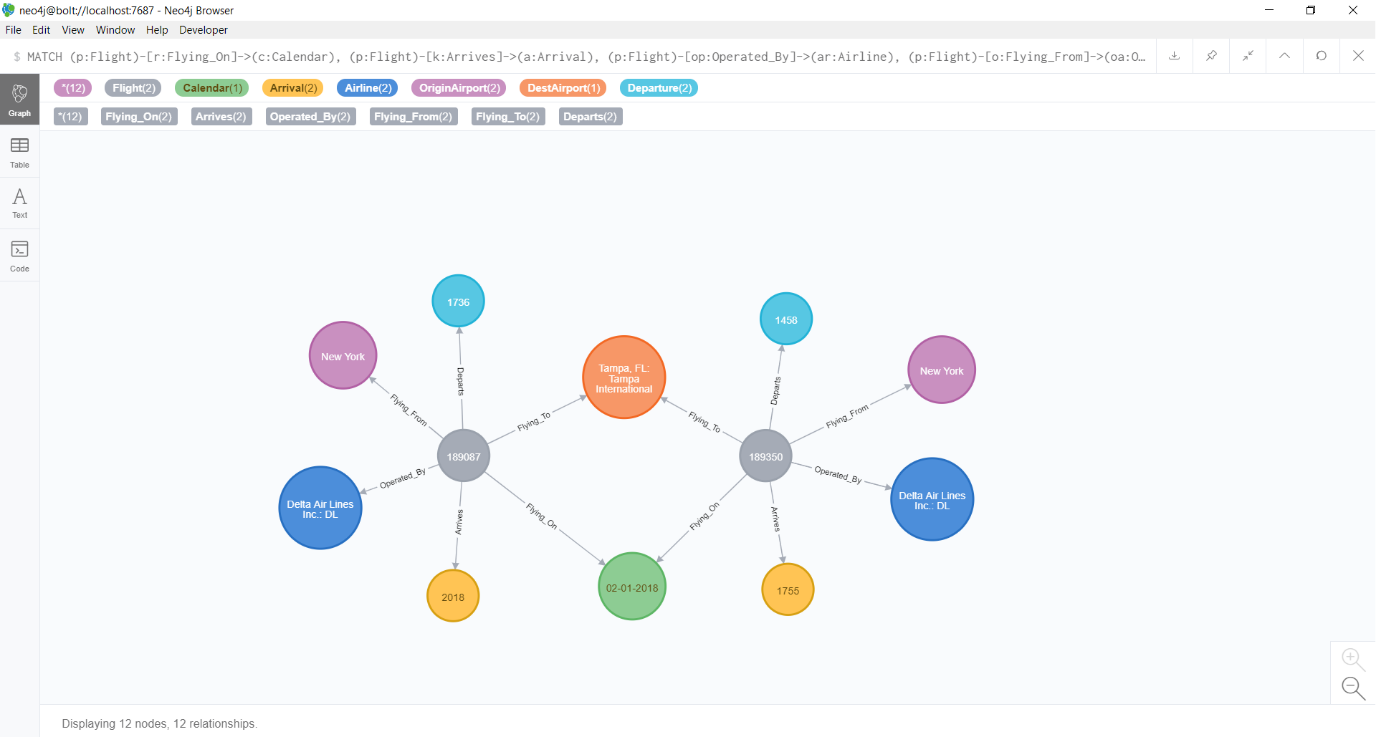
**Fig 6.1 Warehouse Graph (Visualization using Neo4j)**

The above Graph describes the warehouse design of the dataset. We tried to explain various flights taking off from New York to different locations across the US as well as arriving to New York in the month of January. The graphical form of the data makes it easy to know the relationship between various data points easier.



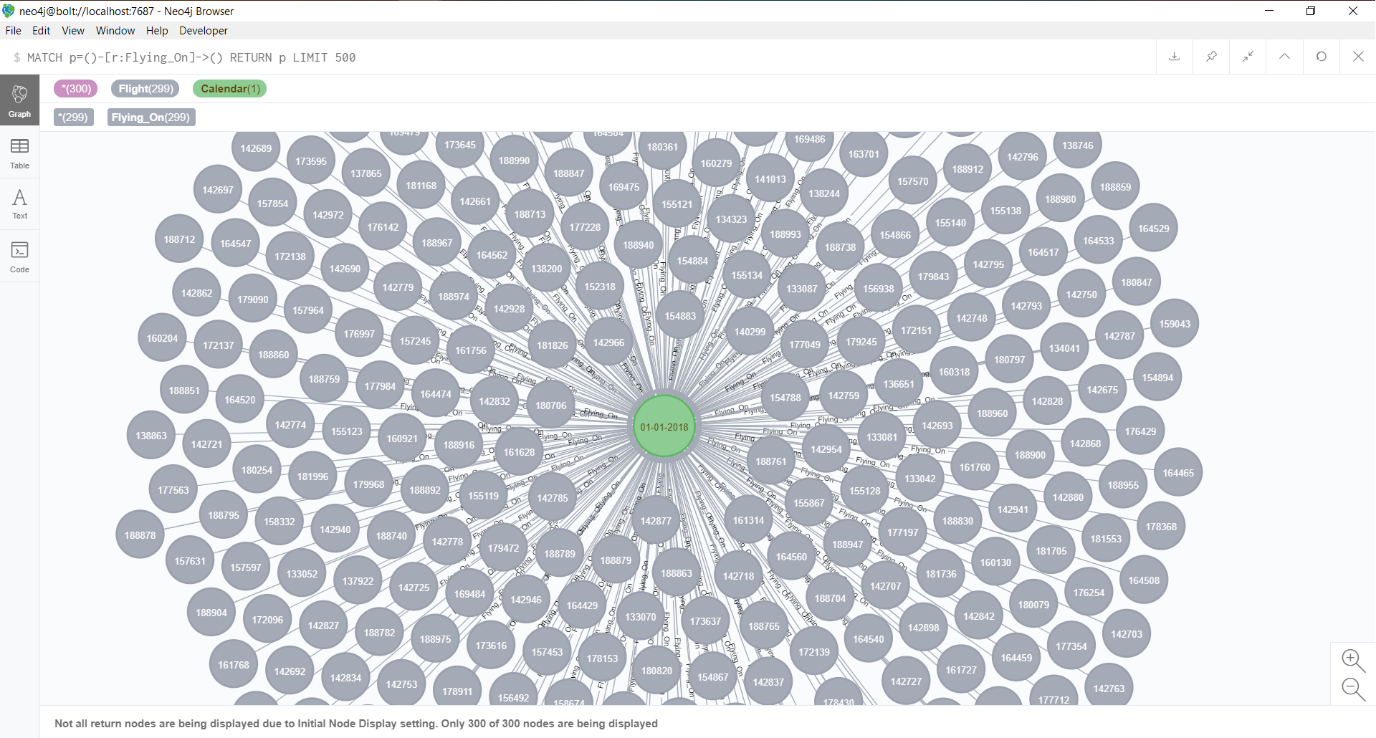
**Fig 6.2 Warehouse Design for flights for a single day**

In the above diagram we tried to show the Airlines travelling in USA on 02/01/2018.In a single diagram we were able to show various Airlines that are having services to and from New York with all the details like Arrival time as well as Departure Time. The colour code of various nodes helps to differentiate the purpose of these nodes. The edges connecting the objects are written with the relationship they have each other.



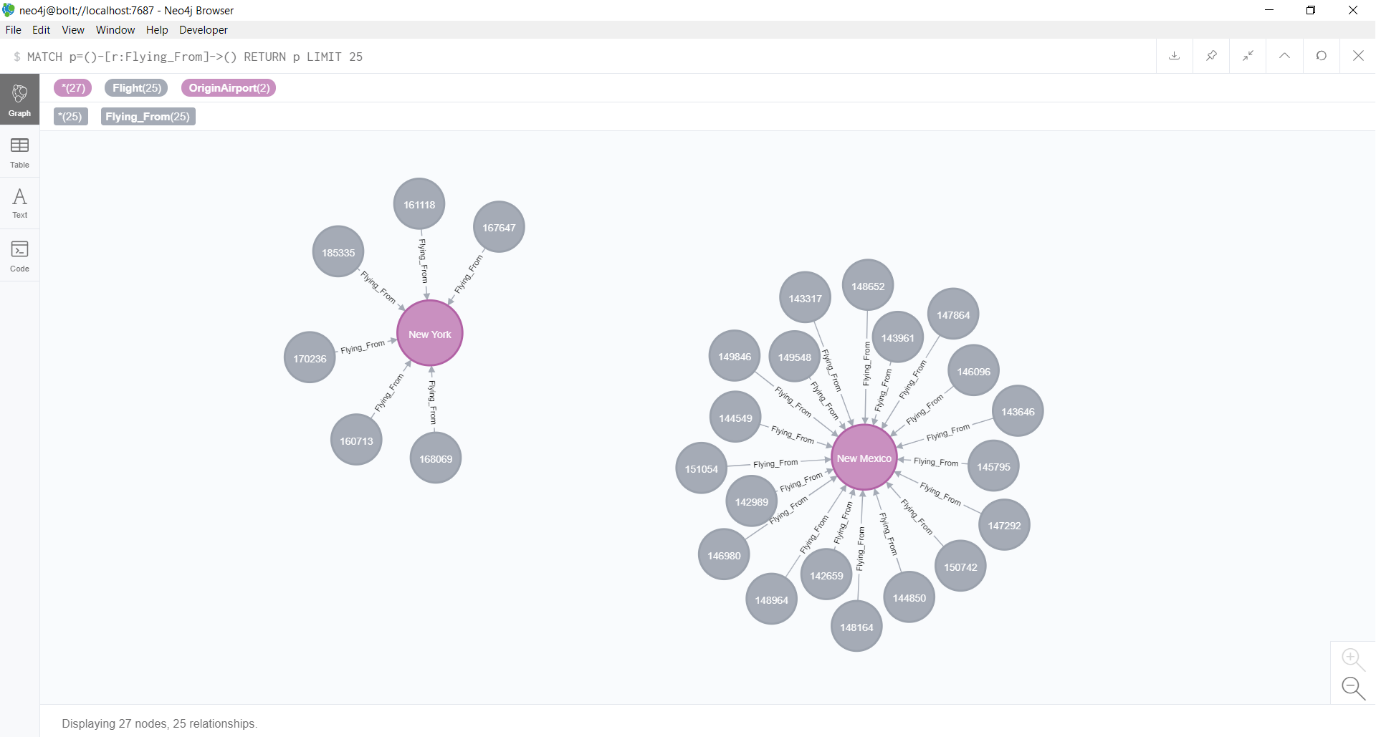
**Fig 6.3 Individual Node Graph**

The above diagram helps us elaborate the warehouse. In this graph there are two flights of Delta Airlines travelling from New York to Tampa International Airport at 17:36 and 14:58 respectively.



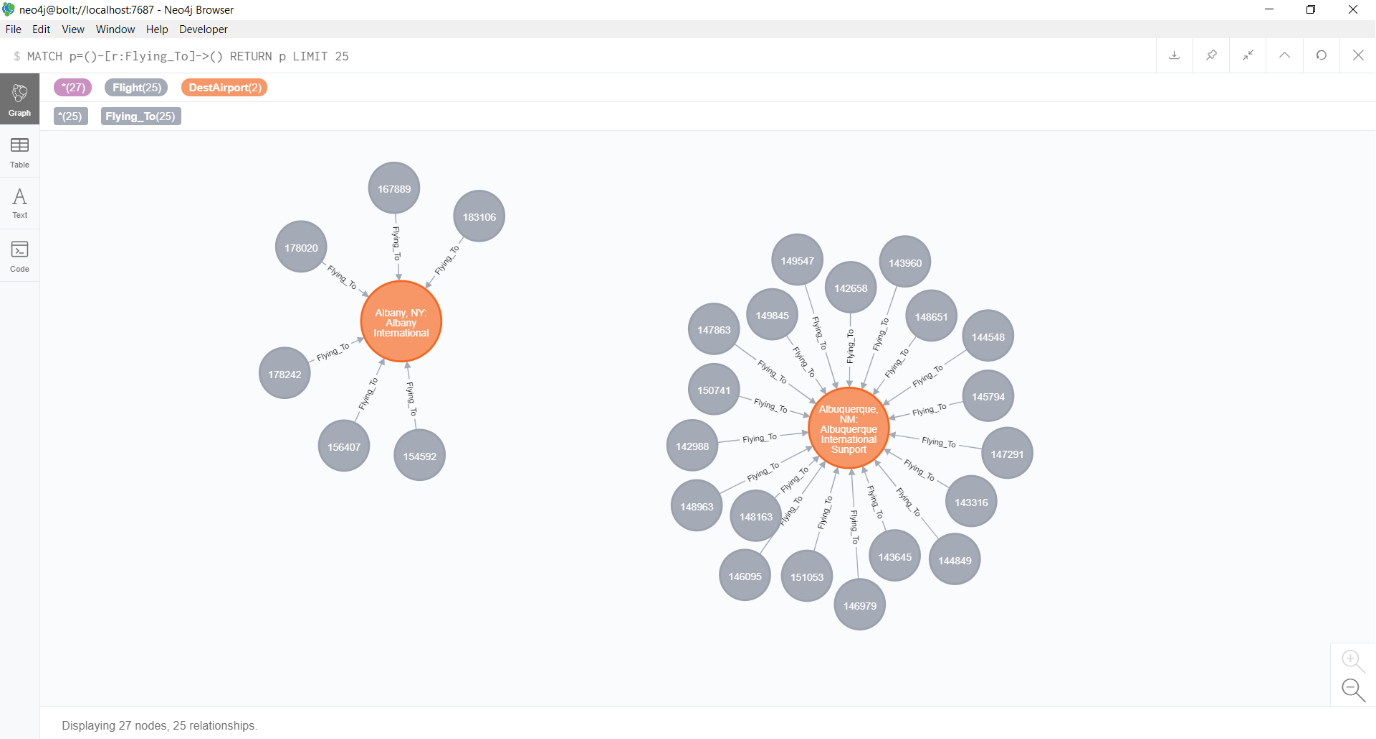
**Fig 6.4 Graph for flights on a Single day**

The above graph helps us to know the flights that are flying on 01/01/2018. This is just to understand one to one relationship that exist. The green node tells us the date and the pink one tells us about the flight number.



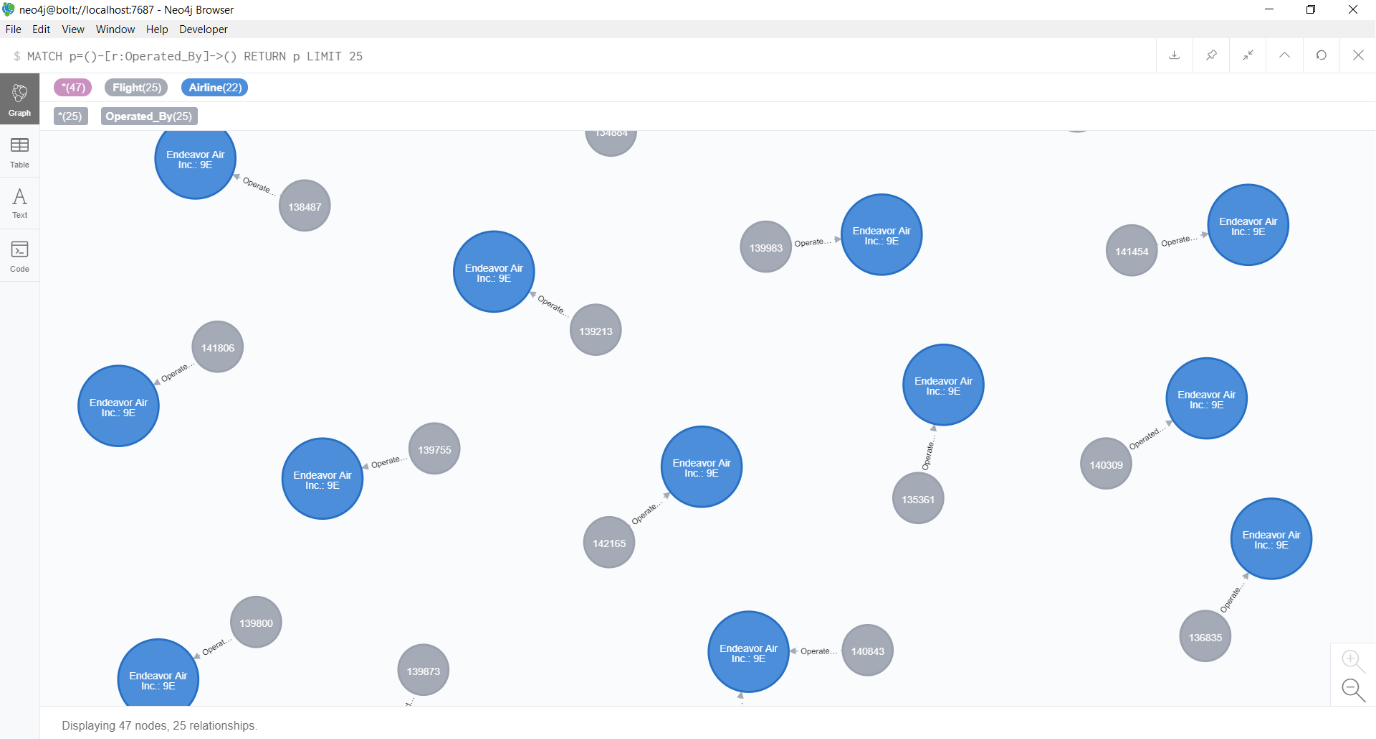
**Fig 6.5 Graph for Flights flying from Airports**

Above diagram shows the no of flights that are taking off from two Cities that are as follows New York and New Mexico. So, we can say that 5 flights take off from New York and 19 flights take off from New Mexico.



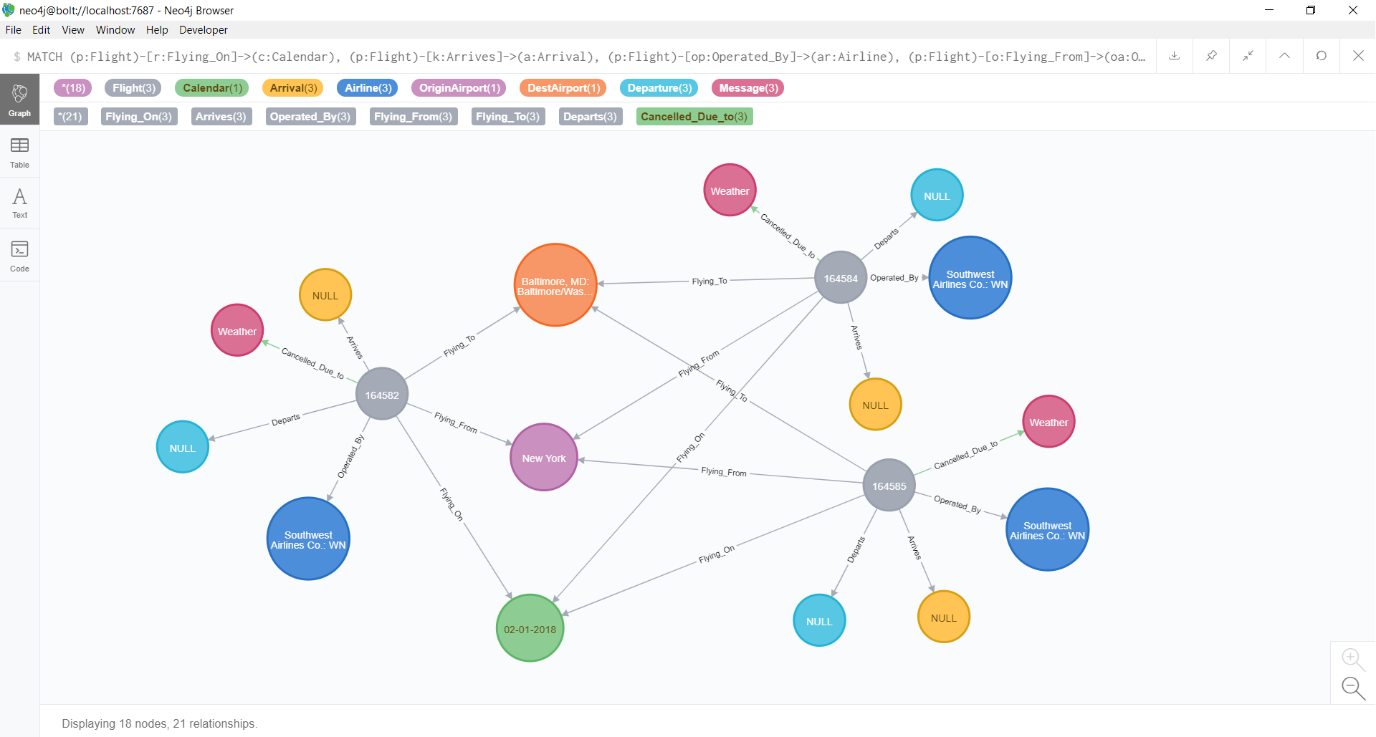
**Fig 6.5 Graph for Flights flying to Airports**

Above diagram shows the no of flights that are arriving in two Cities that are as follows Albany and Albuquerque. So, we can say that 5 flights arrive at Albany and 19 flights arrive at Albuquerque.



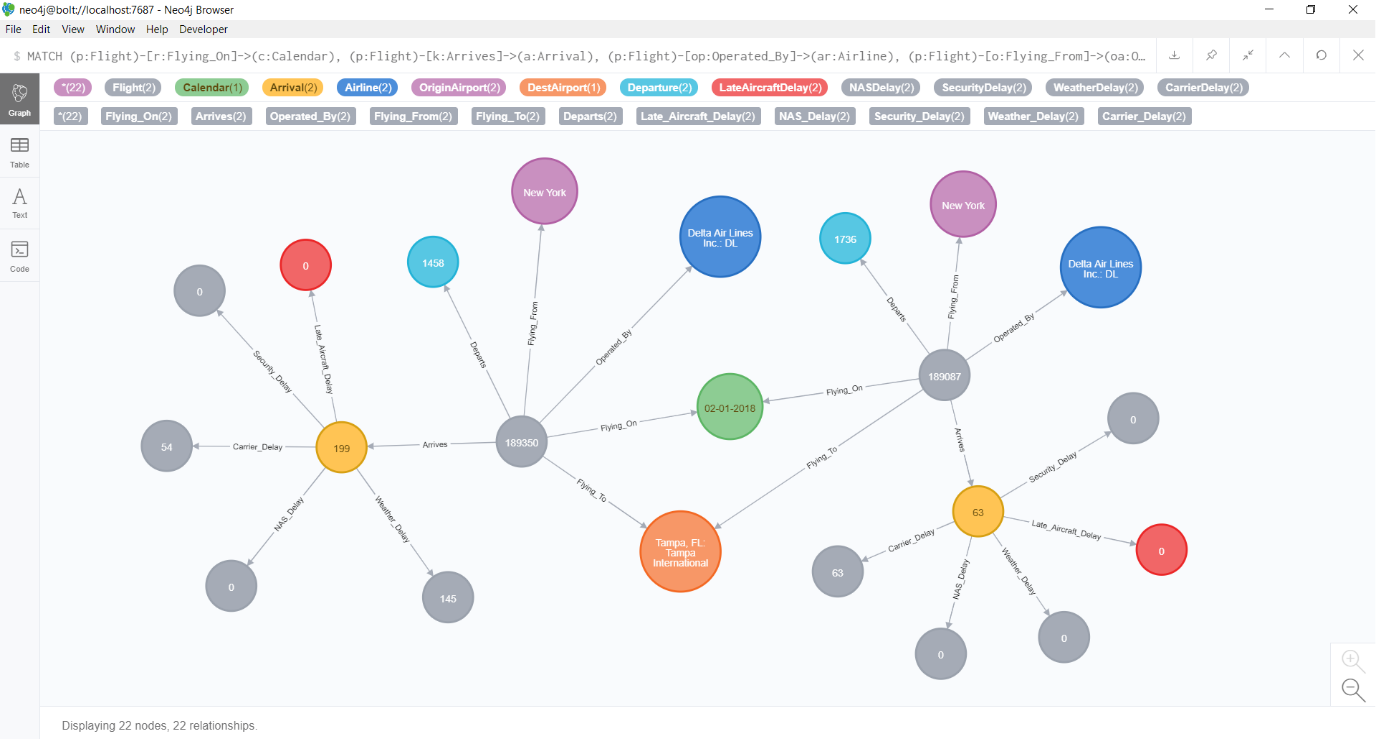
**Fig 6.6 Graph for Flights operated by**

Above graph tells us about the Name of various flights operating across the US. The Unique Airline number is given for each Airline operating to a specific City. That means a single Airline can have multiple Airline Number.



**Fig 6.7 Graph for Flights Cancelled Cause**

The above graph tells us about 3 Southwest Airlines flights that got cancelled on 02/01/2018 due to weather condition which was to fly from New York to Baltimore. We can see that departure time and arrival time for all the Flights are null as these flights got cancelled.



**Fig 6.8 Graph for Flights Delayed Causes**

In the graph we tried to visualize 2 Delta Airlines flights that were travelling from New York to Tampa which got delayed by 199 and 63 minutes respectively .The delay in the time can be divided amongst weather, Late Aircraft, Carrier, NAS and Security Delay. The Cypher Query for the above graph are attached:



# **7. CONCLUSIONS**

The Passengers are always stuck in the airports longer than we have planned due to the flight delays. For avoiding these delays, we acquired the flight data and used Visualization techniques to understand data. We tried to explore the dataset with the aim of understanding some properties of delay which occur on flights. We were helped by many visualization tools for exploring the data.

# **BIBLIOGRAPHY**

OST\_R | BTS | Transtats [WWW Document], n.d. URL https://www.transtats.bts.gov/DL\_SelectFields.asp (accessed 12.16.19).

Understanding the Reporting of Causes of Flight Delays and Cancellations | Bureau of Transportation Statistics [WWW Document], n.d. URL https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations (accessed 12.16.19).

The first GraphGist Challenge completed, 2013. . Neo4j Graph Database Platf. URL https://neo4j.com/blog/the-first-graphgist-challenge-completed/ (accessed 12.16.19).

The State of the Aviation Industry - KPMG Ireland [WWW Document], 2019. . KPMG. URL https://home.kpmg/ie/en/home/insights/2019/01/aviation-industry-leaders-report-2019-state-of-aviation-industry.html (accessed 12.16.19).

Holtz, Y., n.d. The R Graph Gallery – Help and inspiration for R charts [WWW Document]. R Graph Gallery. URL https://www.r-graph-gallery.com/index.html (accessed 12.16.19).

Why Do Flights Get Delayed - 15 Reasons Why Your Next Flight May Be Delayed, 2018. . Travel Tips. URL https://www.cruisinaltitude.com/why-do-flights-get-delayed-15-reasons-next-flight-may-delayed/ (accessed 12.16.19).

# **APPENDIX A –VISUALIZATIONS CODE**

library(RODBC)

library(tidyr)

library(dplyr)

library(ggplot2)

library(tidyr)

library(reshape2)

library(plotly)

library(viridis)

library(hrbrthemes)

conn = odbcDriverConnect("Driver={SQL Server};server=LAPTOP-3LA7V46O; database=Airline\_DW; trusted\_connection=yes")

DelayReason = as.data.frame(sqlQuery(conn, "SELECT CONVERT(int, Month)hiash AS Month, SUM(CONVERT(int,CarrierDelay))/60 CarrierDelay,SUM(CONVERT(int,WeatherDelay))/60 WeatherDelay, SUM(CONVERT(int,NASDelay))/60 NASDelay, SUM(CONVERT(int,SecurityDelay))/60 SecurityDelay, SUM(CONVERT(int,LateAircraftDelay))/60 LateAircraftDelay

FROM Flight\_Summaries\_Fact FS

INNER JOIN Calendar\_Dim C ON FS.CalendarKey = C.CalendarKey

Group By Month

order by Month ASC"))

d <- melt(DelayReason, id="Month")

lChart <- ggplot(d, aes(x=Month, y=value, color=variable)) + geom\_line(aes(linetype=variable), size=1) +

geom\_point(aes(shape=variable, size=5)) +

scale\_x\_discrete(limits=1:12) +labs(title = "Delay Reason") +

theme(plot.title = element\_text(hjust = 0.5))+ylab("Delay in hours") +

xlab("Months (for year 2018)") ggplotly(lChart)

ArrivalDelay = as.data.frame(sqlQuery(conn, "SELECT DISTINCT Airline\_Description, CONVERT(int,AR.ArrDelayMinutes) ArrivalDelay

FROM Airline\_Dim AD

INNER JOIN Flight\_Summaries\_Fact FS ON AD.AirlineKey = FS.AirlineKey

INNER JOIN Arrival\_Dim AR ON FS.ArrivalKey = AR.ArrivalKey

ORDER BY Airline\_Description ASC"))

graph\_1 <- ArrivalDelay %>% ggplot(aes(x= Airline\_Description, y = ArrivalDelay,

col = Airline\_Description)) +geom\_jitter(alpha = 1, size = 0.3) +coord\_flip() +

theme(legend.position = "none") +xlab("Airline Names") +ylab("Arrival Delay (in minutes)")

ggplotly(graph\_1)

DepatureDelay = as.data.frame(sqlQuery(conn, "SELECT DISTINCT AD.Airline\_Description, CONVERT(int,DP.DepDelayMinutes) DepartureDelay

FROM Airline\_Dim AD

INNER JOIN Flight\_Summaries\_Fact FS ON AD.AirlineKey = FS.AirlineKey INNER JOIN Departure\_Dim DP ON FS.ArrivalKey = DP.DepatureKey

ORDER BY Airline\_Description ASC"))

graph\_2 <- DepatureDelay %>% ggplot(aes(x= Airline\_Description, y = DepartureDelay,

col = Airline\_Description)) + geom\_jitter(alpha = 1, size = 0.3) + coord\_flip() +

theme(legend.position = "none") + xlab("Airline Names") +

ylab("Depature Delay (in minutes)")

ggplotly(graph\_2)

CancelledData = as.data.frame(sqlQuery(conn, "SELECT CONVERT(int, Month) AS Month,

COUNT(CONVERT(int,Cancelled)) CancelledCount

FROM Flight\_Summaries\_Fact FS

INNER JOIN Calendar\_Dim C ON FS.CalendarKey = C.CalendarKey

WHERE Cancelled is not null and Cancelled > 0

Group By Month

order by Month ASC"))

Cancelled\_Graph <- ggplot(CancelledData, aes(x=CancelledData$Month, y=CancelledData$CancelledCount)) + geom\_bar(stat = "identity") +scale\_x\_discrete(limits=1:12) +ylab("No. of Flights") +xlab("Months (for year 2018)")

ggplotly(Cancelled\_Graph)

FlightData = as.data.frame(sqlQuery(conn, "SELECT AD.AirlineKey, AD.Reporting\_Airline, AR.ArrDelayMinutes

FROM Flight\_Summaries\_Fact FS

INNER JOIN Airline\_Dim AD ON FS.AirlineKey = AD.AirlineKey

INNER JOIN Arrival\_Dim AR ON FS.ArrivalKey = AR.ArrivalKey

WHERE AR.ArrDelayMinutes > 0"))

FlightData %>% ggplot( aes(x=Reporting\_Airline, y=ArrDelayMinutes, fill=Reporting\_Airline)) + geom\_boxplot() + scale\_fill\_viridis(discrete = TRUE, alpha=0.6, option="A") + theme\_ipsum() + theme( legend.position="none", plot.title = element\_text(size=11, hjust = 0.5)) + ylab("Arrival Delay") + xlab("Airlines")

# **APPENDIX B – NEO4J CODE**

**Determining Relationships**

Airline\_Dim (Operated\_By)

Arrival\_Dim (Arrives)

Calendar\_Dim (Flying\_On)

Departure\_Dim (Departs)

Destination\_Airport\_Dim (Flying\_To)

Origin\_Airport\_Dim (Flying\_From)

Flight\_Summaries\_Fact (Flight)

**Create Relationship**

MATCH (c:Flight), (s:Airline) where c.AirlineKey = s.AirlineKey create (c)-[r:Operated\_By]- >(s) return c,s, r LIMIT 25

MATCH (c:Flight), (s:Arrival) where c.ArrivalKey = s.ArrivalKey create (c)-[r:Arrives]- >(s) return c,s, r LIMIT 25

MATCH (c:Flight), (s:Calendar) where c.CalendarKey = s.CalendarKey create (c)-[r:Flying\_On]- >(s) return c,s, r LIMIT 25

MATCH (c:Flight), (s:Departure) where c.DepatureKey = s.DepatureKey create (c)-[r:Departs]- >(s) return c,s, r LIMIT 25

MATCH (c:Flight), (s:DestAirport) where c.DestAirportKey = s.DestAirportKey create (c)-[r:Flying\_To]- >(s) return c,s, r LIMIT 25

MATCH (c:Flight), (s:OriginAirport) where c.OriginAirportKey = s.OriginAirportKey create (c)-[r:Flying\_From]- >(s) return c,s, r LIMIT 25

**See all relationship**

MATCH p=()-->() RETURN p LIMIT 150

**Get Flights based Flight date**

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar)

WHERE c.FullDate = '02-01-2018'

return p,r,c limit 150

**Get All Nodes based Flight date**

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar)

WHERE c.FullDate = '02-01-2018'

return [p,r,c,(p)-[:Operated\_By]- >(), (p)-[:Flying\_From]- >(), (p)-[:Flying\_To]- >(),

(p)-[:Arrives]- >(), (p)-[:Departs]- >()]

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar), (p:Flight)-[k:Arrives]->(a:Arrival)

WHERE c.FullDate = '02-01-2018' AND a.ArrDelay > 0

return [p,r,k,c,a,(p)-[:Operated\_By]- >(), (p)-[:Flying\_From]- >(), (p)-[:Flying\_To]- >(),

(p)-[:Departs]- >()]

limit 150

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar), (p:Flight)-[k:Arrives]->(a:Arrival),

(p:Flight)-[op:Operated\_By]->(ar:Airline), (p:Flight)-[o:Flying\_From]->(oa:OriginAirport)

WHERE c.FullDate = '02-01-2018' AND a.ArrDelay > '60'

AND ar.Airline\_Description='Delta Air Lines Inc.: DL'

AND oa.OriginState = 'NY'

return [p,r,k,c,a,(p)-[:Operated\_By]- >(), (p)-[:Flying\_From]- >(), (p)-[:Flying\_To]- >(),

(p)-[:Departs]- >()]

**Individual Node graph query**

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar), (p:Flight)-[k:Arrives]->(a:Arrival),

(p:Flight)-[op:Operated\_By]->(ar:Airline), (p:Flight)-[o:Flying\_From]->(oa:OriginAirport),

(p:Flight)-[ft:Flying\_To]->(d:DestAirport)

WHERE c.FullDate = '02-01-2018' AND a.ArrDelay > '15'

AND ar.Airline\_Description='Delta Air Lines Inc.: DL'

AND oa.OriginState = 'NY' AND d.Dest = 'TPA'

return [p,r,k,c,a,(p)-[:Operated\_By]- >(), (p)-[:Flying\_From]- >(), (p)-[:Flying\_To]- >(),

(p)-[:Departs]- >()]

**Creating Cancelled Relationship**

MATCH (n:Flight) WHERE n.Cancelled = '1'

CREATE (m:Message {text : n.CancellationCause} )<-[:Cancelled\_Due\_to]-(n)

**Cancelled Cause Query**

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar), (p:Flight)-[k:Arrives]->(a:Arrival),

(p:Flight)-[op:Operated\_By]->(ar:Airline), (p:Flight)-[o:Flying\_From]->(oa:OriginAirport),

(p:Flight)-[ft:Flying\_To]->(d:DestAirport)

WHERE c.FullDate = '02-01-2018' AND p.Cancelled > '0'

AND oa.OriginState = 'NY' AND d.Dest = 'BWI'

return [p,r,k,c,a,(p)-[:Operated\_By]- >(), (p)-[:Flying\_From]- >(), (p)-[:Flying\_To]- >(),

(p)-[:Departs]- >(),(p)-[:Cancelled\_Due\_to]->(:Message)]

**Create Delay Relationship**

MATCH (n:Flight), (a:Arrival) WHERE n.ArrivalKey = a.ArrivalKey AND a.ArrDelayMinutes > '15'

CREATE (m:CarrierDelay {text : n.CarrierDelay})<-[:Carrier\_Delay]-(a)

MATCH (n:Flight), (a:Arrival) WHERE n.ArrivalKey = a.ArrivalKey AND a.ArrDelayMinutes > '15'

CREATE (o:WeatherDelay {text : n.WeatherDelay})<-[:Weather\_Delay]-(a)

MATCH (n:Flight), (a:Arrival) WHERE n.ArrivalKey = a.ArrivalKey AND a.ArrDelayMinutes > '15'

CREATE (p:NASDelay {text : n.NASDelay})<-[:NAS\_Delay]-(a)

MATCH (n:Flight), (a:Arrival) WHERE n.ArrivalKey = a.ArrivalKey AND a.ArrDelayMinutes > '15'

CREATE (q:SecurityDelay {text : n.SecurityDelay})<-[:Security\_Delay]-(a)

MATCH (n:Flight), (a:Arrival) WHERE n.ArrivalKey = a.ArrivalKey AND a.ArrDelayMinutes > '15'

CREATE (r:LateAircraftDelay {text : n.LateAircraftDelay})<-[:Late\_Aircraft\_Delay]-(a)

MATCH (p:Flight)-[r:Flying\_On]->(c:Calendar), (p:Flight)-[k:Arrives]->(a:Arrival),

(p:Flight)-[op:Operated\_By]->(ar:Airline), (p:Flight)-[o:Flying\_From]->(oa:OriginAirport),

(p:Flight)-[ft:Flying\_To]->(d:DestAirport)

WHERE c.FullDate = '02-01-2018' AND a.ArrDelay > '15'

AND ar.Airline\_Description='Delta Air Lines Inc.: DL'

AND oa.OriginState = 'NY' AND d.Dest = 'TPA'

return [p,r,k,c,a,(p)-[:Operated\_By]- >(), (p)-[:Flying\_From]- >(), (p)-[:Flying\_To]- >(),

(p)-[:Departs]- >(),(a)-[:Late\_Aircraft\_Delay]->(),(a)-[:NAS\_Delay]->(),(a)-[:Operated\_By]->(),

(a)-[:Security\_Delay]->(),(a)-[:Weather\_Delay]->(),(a)-[:Carrier\_Delay]->()]