

# Airline Delays

---

Vipul Munot, Disha Wagle, Avruti Srivastava

INDIANA UNIVERSITY | BIG DATA ANALYTICS AND APPLICATIONS | FALL 2015

## Table of Contents

Abstract .....	2
Software .....	3
Data Sets .....	4
Results.....	5
Arrival Time Trend Analysis .....	5
Departure Delay Month wise Analysis .....	5
Delays by day of Week .....	6
Flight Cancellations Analysis .....	6
Busiest Airports .....	8
Worst Performing Carriers.....	9
Best Performing Carriers .....	10
Recommendations.....	11
References.....	12

## Abstract

The past decade has seen air travel become one of the most preferred mode of transport over long distances. Apart from passenger carriers, air shipping also facilitates transportation of perishables and premium express deliveries. This unprecedented growth over time coupled with other factors has led to increased traffics at airports and disturbed schedules of flights. In this project, we analyze the departure and arrival delay trends in passenger carriers, over a series of years. We have inspected the various reasons behind Flight Cancellations. Analysis of crowd at various airports have been performed determining the busiest airports. Also, performance of various Airlines have been examined to provide the carriers with worst and best performance.

## Software

In order to achieve our objectives of analyzing the reasons behind flight delays we utilized dataset consisting of common separated values and analyzed it using python. The python code was run over Enthought Canopy. The various libraries included are:

1. Matplotlib: This is a plotting library for Python in an object oriented way <sup>[1]</sup>.
2. Numpy: This is a fundamental package in Python which supports large, multidimensional arrays and matrices along with a library of mathematical functions to work on these arrays <sup>[2]</sup>.
3. Seaborn: It is a visualization library based on Matplotlib. It provides an interface for generating visually appealing statistical graphics <sup>[3]</sup>.
4. Pandas: This is a software library in Python for data manipulation and analysis <sup>[4]</sup>.

Link to the code: <https://github.com/vipmunot/bdaa>

## Data Sets

We have made use of data sets <sup>[5]</sup> pertaining to the flight details and includes the following attributes:

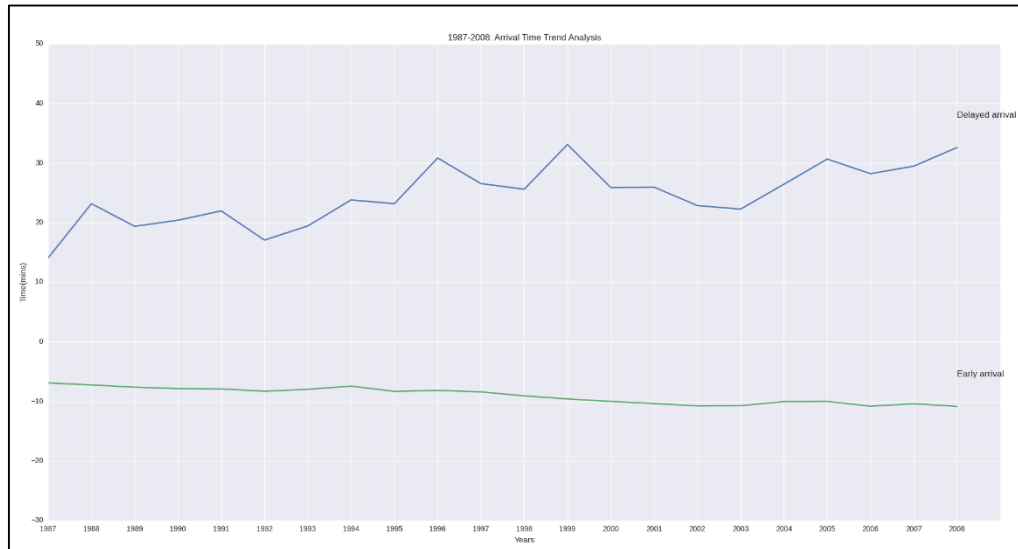
No.	Name	Description
1	Year	1987-2008
2	Month	1-12
3	DayofMonth	1-31
4	DayOfWeek	1 (Monday) - 7 (Sunday)
5	DepTime	actual departure time (local, hhmm)
6	CRSDepTime	scheduled departure time (local, hhmm)
7	ArrTime	actual arrival time (local, hhmm)
8	CRSArrTime	scheduled arrival time (local, hhmm)
9	UniqueCarrier	<u>unique carrier code</u>
10	FlightNum	flight number
11	TailNum	plane tail number
12	ActualElapsedTime	in minutes
13	CRSElapsedTime	in minutes
14	AirTime	in minutes
15	ArrDelay	arrival delay, in minutes
16	DepDelay	departure delay, in minutes
17	Origin	origin <u>IATA airport code</u>
18	Dest	destination <u>IATA airport code</u>
19	Distance	in miles
20	TaxiIn	taxi in time, in minutes
21	TaxiOut	taxi out time in minutes
22	Cancelled	was the flight cancelled?
23	CancellationCode	reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
24	Diverted	1 = yes, 0 = no
25	CarrierDelay	in minutes
26	WeatherDelay	in minutes
27	NASDelay	in minutes
28	SecurityDelay	in minutes
29	LateAircraftDelay	in minutes

**\*\*Note:** We have made use of 400,000 records for each year in the graphs for ‘Arrival Time trend Analysis’, ‘Delays by Day of Week’, ‘Busiest Airports’, ‘Worst Performing Carriers’ and ‘Worst Performing Carriers’ due to hardware constraints.

## Results

### Arrival Time Trend Analysis

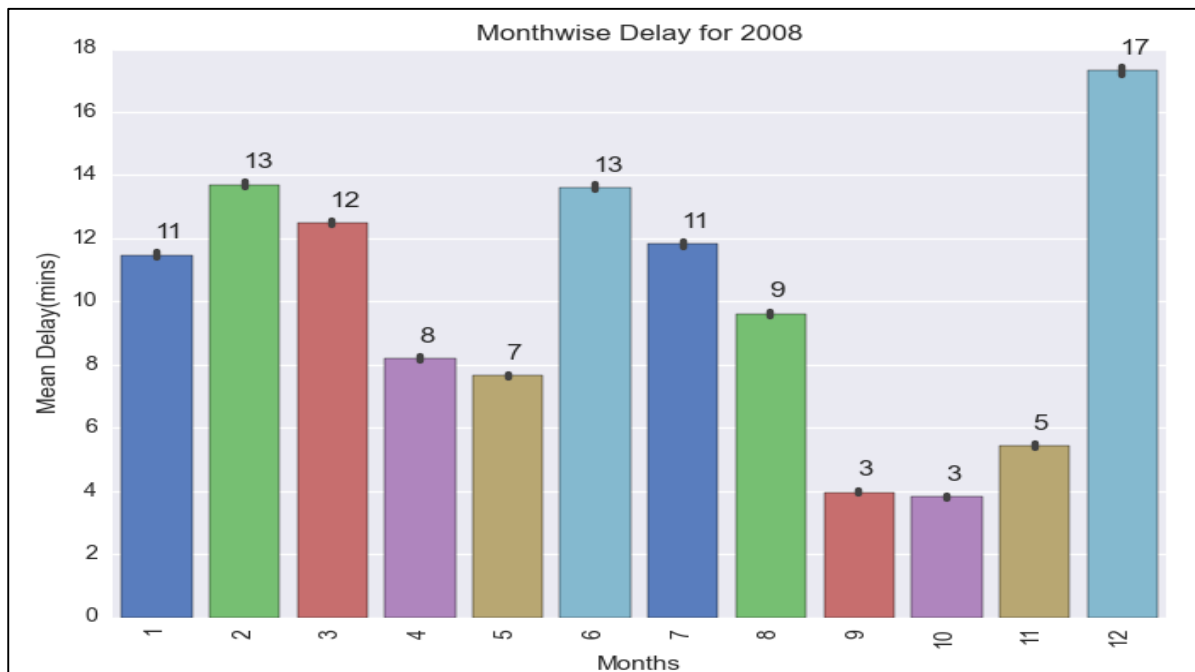
The following graph shows the **variation in average delays** that occurred for a period of 21 years from 1987 to 2008. It also gives the average early arrivals for the same duration.



We can infer from the graph that year 2005 and 2008 has maximum instances of delays whereas year 2006 has the maximum instances early of arrivals.

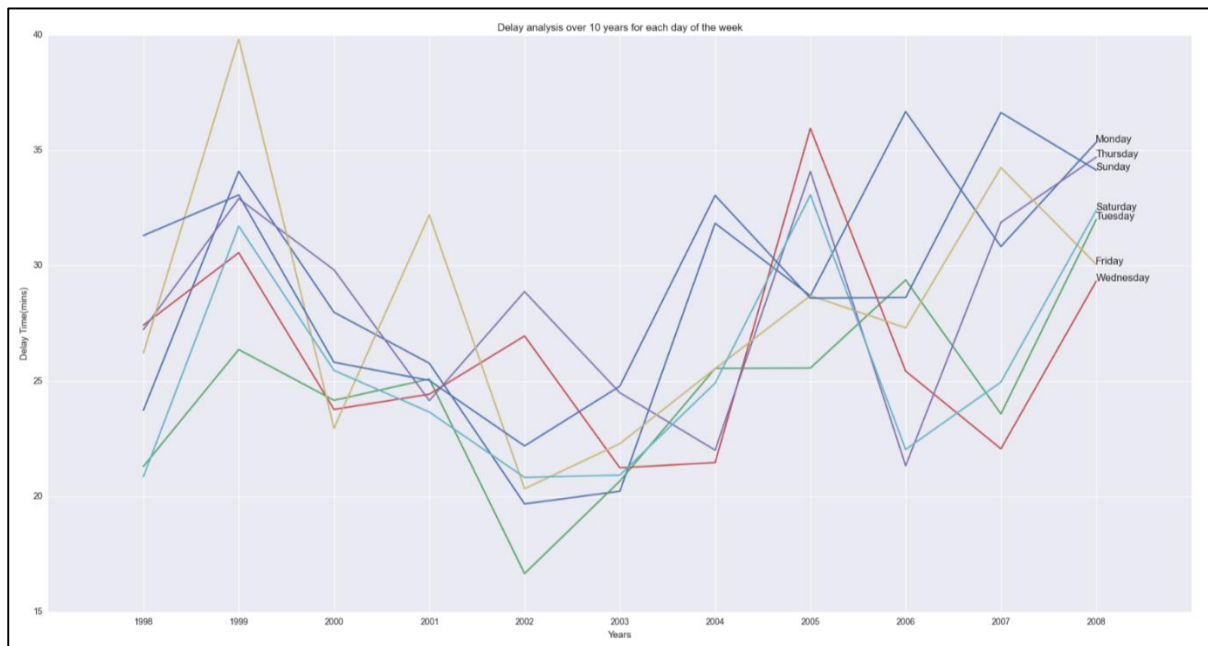
### Departure Delay Month wise Analysis

This graph for the year 2008, shows December as the month with maximum delays with a mean of 17 minutes. Marking the start of winter, a harsh climate is expected to cause delays. Also, being the festive vacation season, more people fly to various places, increasing air traffic.



(Note: We have considered the month wise delay for the year 2008, as data for other years was not available.)

## Delays by day of Week



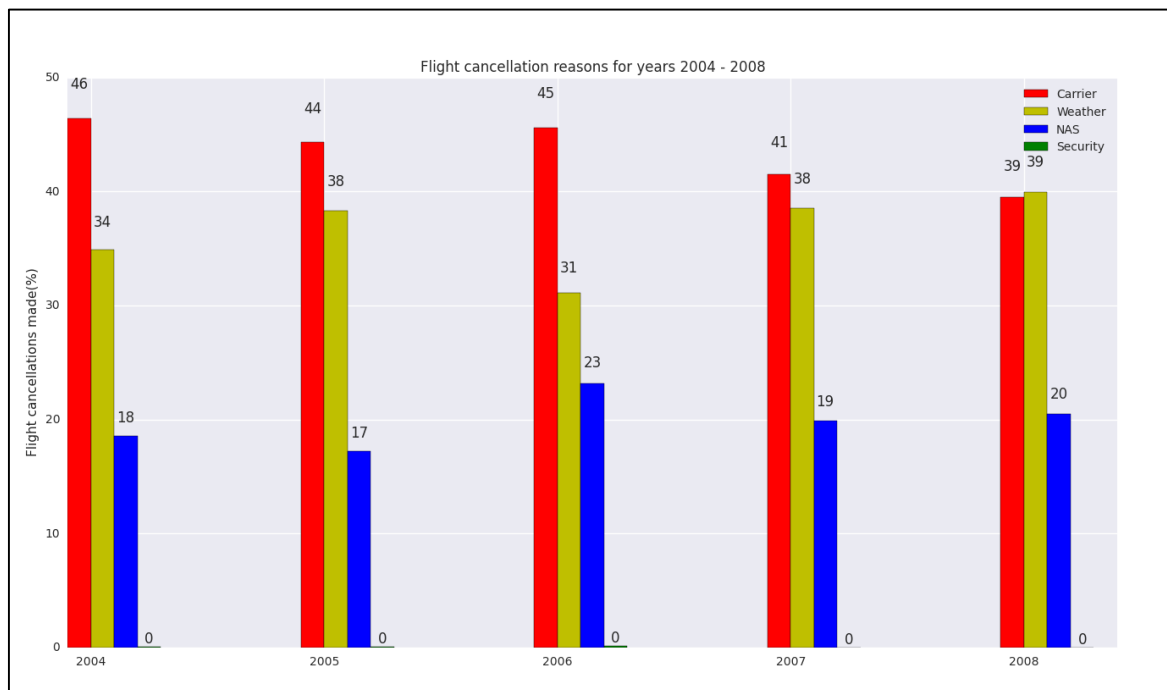
The above graph depicts the plotting of **delay occurrences based on the days of the week** for a series of years.

As seen above, for most of the years, the weekend: Friday, Saturday, and Sunday appear to be the days with maximum delays. The weekdays see comparatively fewer delays with the exception of 2005, where Wednesday witnesses maximum delays. Hence, we can conclude that best days to travel are weekdays and travelling on weekends will increase the chances of a passenger running behind schedule, due to delays.

## Flight Cancellations Analysis

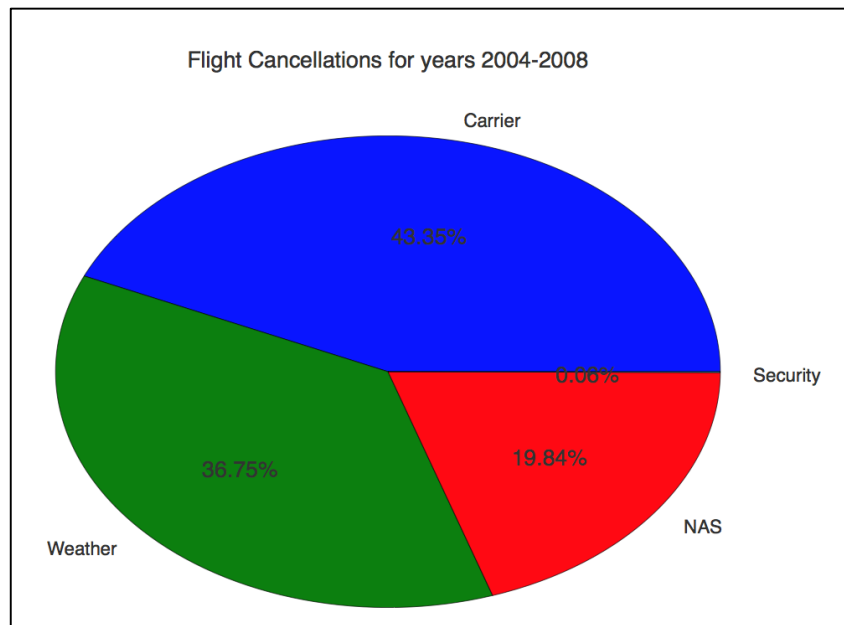
The next graph shows the various reasons of flight cancellations over a period of 5 years. There exist primarily four reasons for cancellations:

- Carrier Issues
- Weather Issues
- NAS
- Security Reasons



The major reason behind cancellations remained carrier issues, however, year 2007 was an exception for which maximum flight cancellation was encountered due to weather issues.

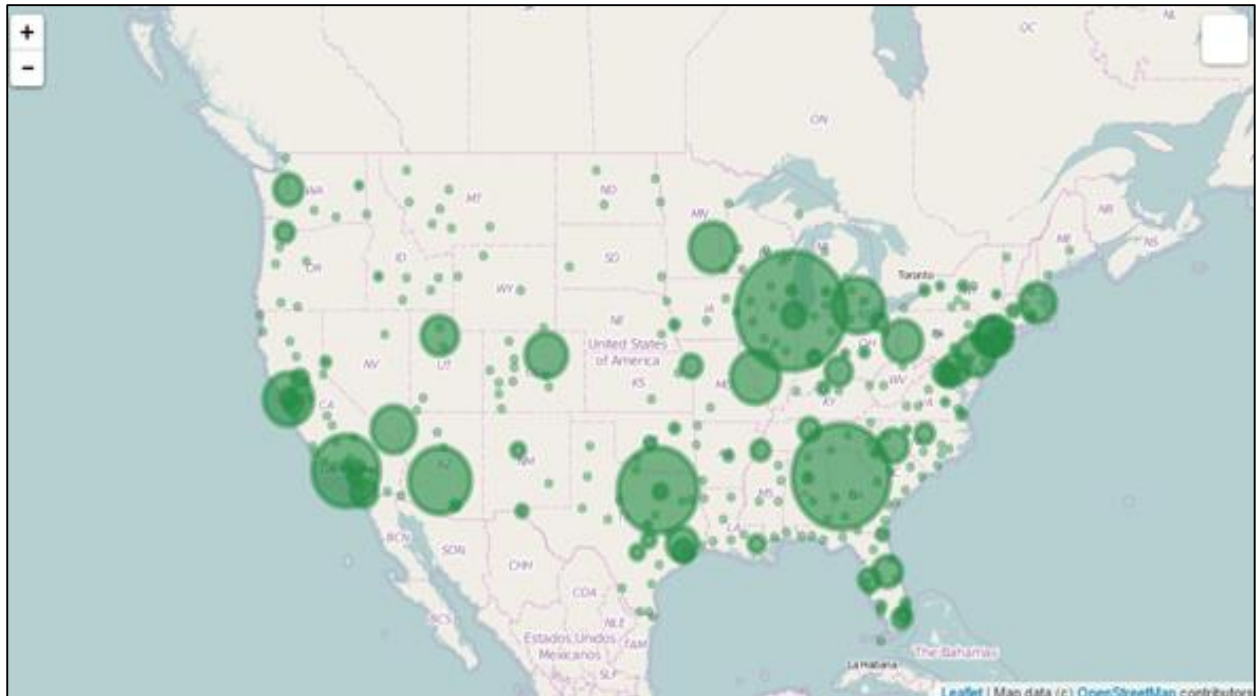
Following is a cumulative graph, for the various cancellation reasons from the year 2004 to 2008.



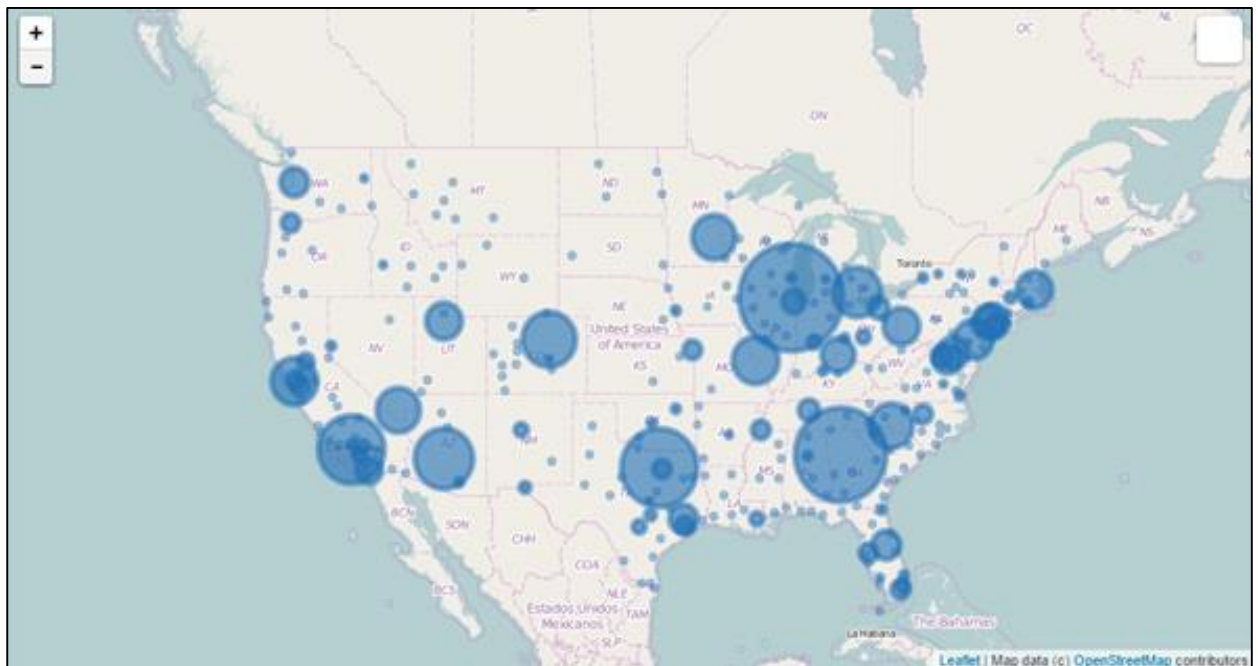
It can be observed that over 5 years, Carrier issues have been the reason behind maximum cancellations accounting for 43.35% of the total cancellations.



## Busiest Airports



This graph represents the busiest airports in the country on the basis of the origin of the flight. It can be deduced that maximum flights take off from William B Hartsfield-Atlanta International Airport followed by Chicago O'Hare International Airport and Dallas-Fort Worth International Airport.



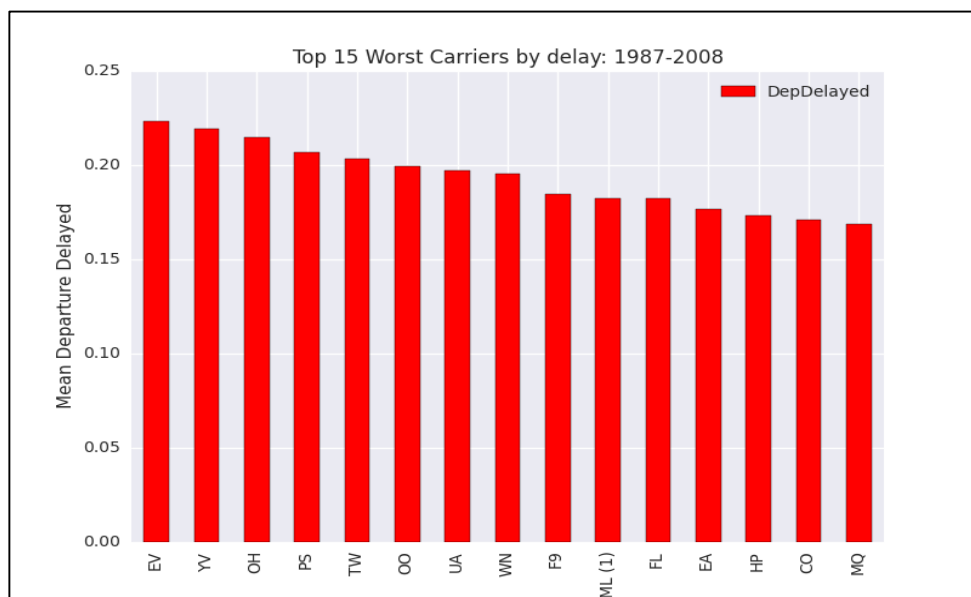
The above graph is the pictorial representation of the busiest airports in the country on the basis of the destination of the flight. We can see that ORD(Chicago O'Hare International Airport) proves to be the Destination for the maximum number of flights. The second in line is William

B Hartsfield-Atlanta International Airport, followed by Dallas-Fort Worth International Airport.

As we can see the busiest airports above, there are high chances that the traffic might affect the schedule of flights and may lead to unforeseen delays. Also, when taking a flight which has a halt at those airports may cause inconvenience, as it could lead to messing up of baggage, extended waiting time and other issues. So, unless that particular airport is not the final destination, flights with those airports as layovers can be avoided for a smoother journey.

### Worst Performing Carriers

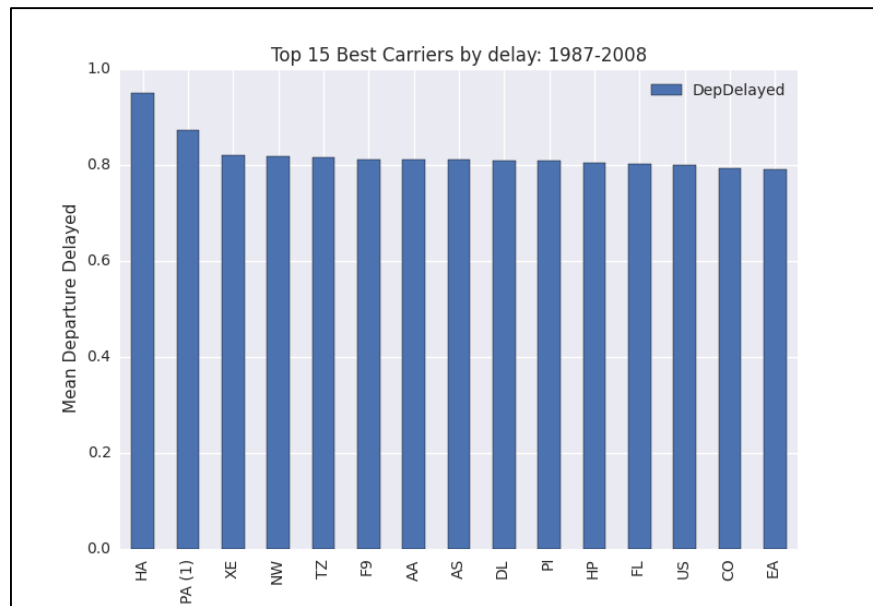
We calculated top fifteen worst performing carriers with respect to departure delays having departure delays more than 15 minutes.



The rankings obtained are as follows: ( 1 representing the worst carrier)

1. Atlantic Southeast Airlines
2. Mesa Airlines Inc.
3. Comair Inc.
4. Pacific Southwest Airlines
5. Trans World Airways LLC
6. Skywest Airlines Inc.
7. United Air Lines Inc.
8. Southwest Airlines Co.
9. Frontier Airlines Inc.
10. Midway Airlines Inc. (1)
11. AirTran Airways Corporation
12. Eastern Air Lines Inc.
13. America West Airlines Inc.
14. Continental Air Lines Inc.
15. American Eagle Airlines Inc.

## Best Performing Carriers



The top fifteen best performing carriers with least rate of departure delays are as follows:

1. Hawaiian Airlines Inc.
2. Pan American World Airways (1)
3. Expressjet Airlines Inc.
4. Northwest Airlines Inc.
5. ATA Airlines d/b/a ATA
6. Frontier Airlines Inc.
7. American Airlines Inc.
8. Alaska Airlines Inc.
9. Delta Air Lines Inc.
10. Piedmont Aviation Inc.
11. America West Airlines Inc.
12. AirTran Airways Corporation
13. US Airways Inc.
14. Continental Air Lines Inc.
15. Eastern Air Lines Inc.

## Recommendations

Based on the graphs obtained and analysis performed, we can make the following recommendations:

- On the basis of ‘Month Wise Delay Analysis Graph’, we see that December witnesses maximum delays. So when travelling around this time, passengers should schedule their flights well ahead of time, in order to avoid any inconvenience being caused due to the same, as these delays could affect their planned activities or important meetings.
- The graph obtained for ‘Delays by day of week’ concludes that it is better to travel on weekdays if the schedule permits.
- Analysis of ‘Flight cancellation reasons’, shows ‘Carrier Issues’ as the primary reason behind cancellations. This calls for the carriers to be more vigilant with their technicalities in order to avoid the inconvenience caused to the passengers.
- Airports like Chicago O’Hare International Airport, William B Hartsfield-Atlanta International Airport and Dallas-Fort Worth International Airport have proved to be the busiest airports. As the traffic dealt with, is high, the chances of baggage mess up, extended delays and other problems are higher. Therefore, unless these airports are final destinations, flights with them as layovers could be avoided.

## References

- [1] <https://pypi.python.org/pypi/matplotlib>
- [2] <https://pypi.python.org/pypi/numpy>
- [3] <https://pypi.python.org/pypi/seaborn/>
- [4] <http://pandas.pydata.org/>
- [5] Source: <http://stat-computing.org/dataexpo/2009/the-data.html>
- [6] <http://stat-computing.org/dataexpo/2009/posters>