BAN 602 Quantitative Fundamentals

Homework Assignment 1 (**GROUP 2**)

# Flight Delays at RegionEx


**GROUP MEMBERS**

Gyati Mittal

Maitreyee Das

Rajamani Chidambaram

Rashme Gunasekaran

Vine Merline Ananthi Baskaran

## Case Details:

## Metrics taken into consideration during the assessment of Airline Performance

- Percentage of scheduled flights that were delayed
- Percentage of scheduled flights that arrived on time
- Average arrival delay in minutes of an airline's flights

Other factors related to flight delays affect customers' perceptions of an airline:

- Two different airlines might schedule two different durations for the same route, which could affect their apparent on-time performance
- a delay on the first leg of a two-leg flight is often a more serious problem than a comparable delay on the second leg, because it could result in a missed connection, forcing the customer to wait until the next available flight

## RegionEx's routes:

Route 1: Dallas-Fort Worth (DFW) to New Orleans (MSY)

Route 2: New Orleans (MSY) to Dallas-Fort Worth (DFW)

Route 3: New Orleans (MSY) to Pensacola (PNS)

Route 4: Pensacola (PNS) to New Orleans (MSY)

# ANSWERS

## Question 1: Descriptive Analysis on the Delay in Arrival of RegionEx and MDA Airlines

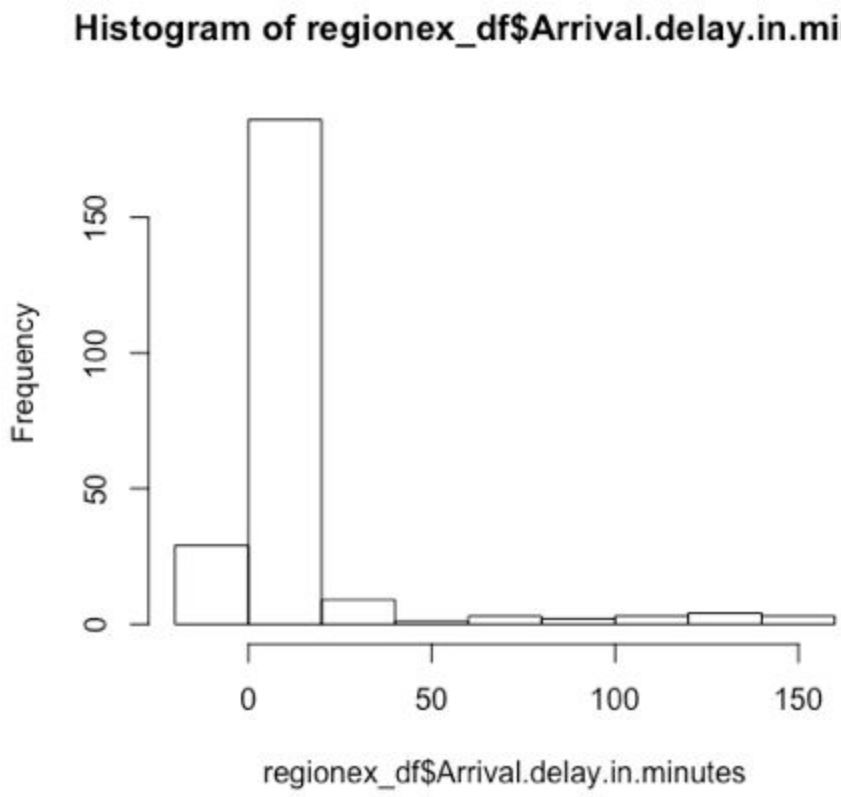
Tabular Summary of results:

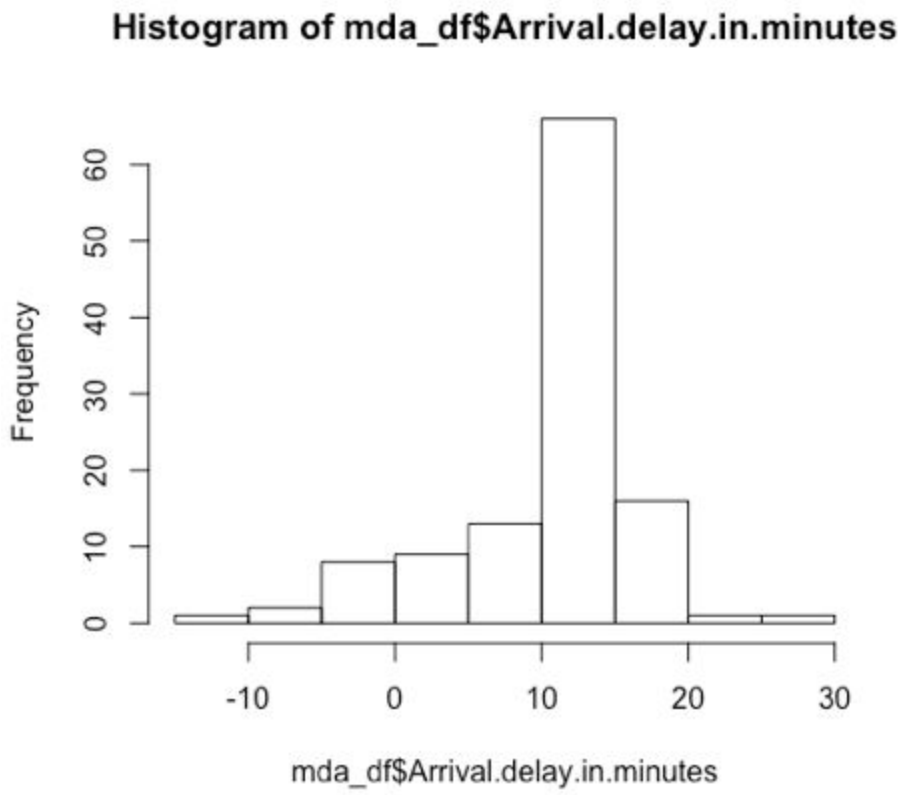
| AIRLINE | Mean | Median | 90th percentile | Standard Deviation |
|---|---|---|---|---|
| **RegionEx** | 15.6625 | 9 | 21 | 27.65036 |
| **MDA** | 10.92308 | 13 | 16.4 | 6.418123 |

MDA airlines appears to perform better than RegionEx since Arrival Delay mean time i.e the average time of arrival delay in minutes is higher for RegionEx compared to MDA airlines. However, the Median is more accurate as data points are positioned in the middle of the dataset and are ordered from smallest to largest. Therefore, RegionEx appears to perform better, as datasets with extremely large data points can skew the actual result. However, looking at 90% percentile, MDA appears to be performing well compared to RegionEx and Standard Deviation arrival delay in minutes for RegionEx airlines data points are more spread out than MDA airlines. Standard Deviation of RegionEx is more than 5 times the standard deviation of MDA. Therefore, it is considered that MDA flights are performing well.

Z score values for RegionEx airlines lies between -0.96427 to 4.966934. Z score greater than +3 means the skewness is positive and it means that mean is usually more than Median.Z score values for MDA airlines lies between -3.72743 to 2.504926. Z score less than -3, means the skewness is negative and it means that mean is usually less than the median.

Question 2: Comparative distribution of RegionEx and MDA based on Histogram






As per above histogram of RegionEx, approximately 24 flights are delayed by more than 20 minutes so in terms of percentage there are 10% flights which are delayed more than 20 minutes. On the other hand in histogram of MDA, Only 2 flights are delayed by more than 20 minutes so around 1.75% flights are delayed more than 20 minutes.
Additionally, 9.5% RegionEx and 12% MDA flights have no delay so marginally RegionEx is better but overall MDA has good results as it has very low delay in comparison of RegionEx flight.
Also, RegionEx is highly skewed within 10 to 20 minutes delay whereas MDA is skewed within 10 to 15 minutes.
While comparing the histograms for both airlines, we analyze that the delay in RegionEx flights seemed to be distributed over a longer time range (-10 to - 150 minutes), whereas the MDX flight delays seemed to be in a shorter range (-20 to 30 mins). This is in line with our conclusion from the previous analysis.

## Question 3: Percentage of Late Flights and On-Time flights

Below is a tabular display showing the percentage of "late" and "on-time" flights of RegionEx and MDA September flights according to the FAA definition

| | **RegionEx** | **MDA** |
|---|---|---|
| Late flights (in %) | 73.75 | 26.49573 |
| On-time flights (in %) | 26.25 | 73.50427 |

## Question 4: Descriptive statistics of delay minutes for each of the four routes

**ROUTE 1**: If we look at Table 4.2, the percentage of late flights of MDA Airlines on Route 1 is higher than that of RegionEx Airlines. But, when we compare the median and standard deviation of both the airlines in Table 4.1, *RegionEx flights has had many flights with larger delays which affects their on-time performance.*

**ROUTE 2**: Same as the analysis we did for Route 1, the delay percentage of Route 2 MDA Airlines is higher than that of RegionEx Airlines. But *as we look at their standard deviations, RegionEx has had many flights with larger delays.*

**ROUTE 3**: Comparing the percentage of delayed flights for RegionEx and MDA Airlines on Route 2 doesn't help in the assessment of on-time performance as their values are equal. But *the higher standard deviation of RegionEx flights denotes that it had flight legs with larger delays compared to MDA flights.*

**ROUTE 4**: *The analysis on Route 4 looks similar to that of all the other routes which shows that MDA Airlines is still performing better in all aspects.*

Table 4.1 Descriptive Statistics of RegionEx and MDA airlines on each route

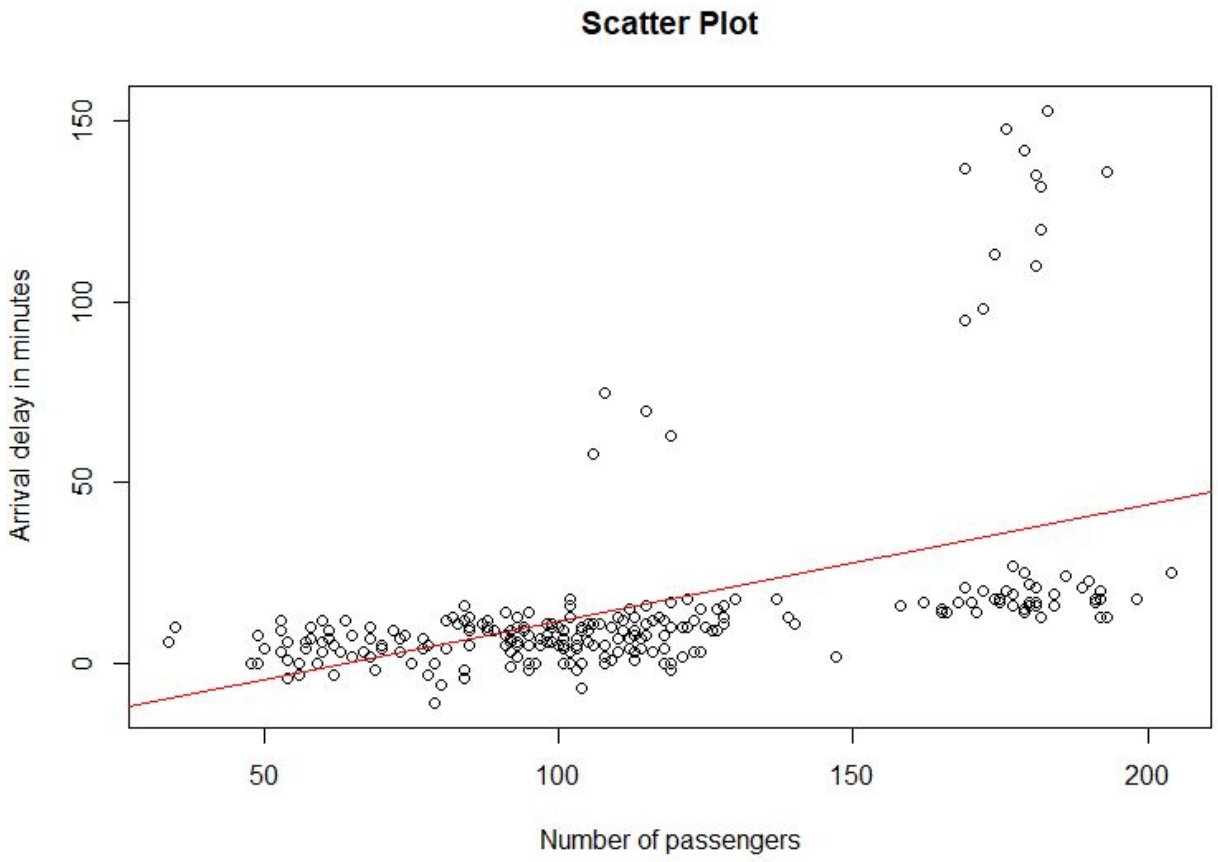
| | **Descriptive Statistics** | **RegionEx** | **MDA** |
|---|---|---|---|
| **Route 1** | Mean | 16.42222 | 13.71429 |
| | Median | 10 | 14 |
| | 90th percentile | 20 | 15.3 |
| | Standard Deviation | 28.91288 | 1.329359 |
| **Route 2** | Mean | 18.21111 | 13.62069 |
| | Median | 9 | 13 |
| | 90th percentile | 24.1 | 15.2 |
| | Standard Deviation | 31.84742 | 1.497946 |
| **Route 3** | Mean | 11 | 10.2 |
| | Median | 6.5 | 11 |
| | 90th percentile | 18 | 18 |
| | Standard Deviation | 16.1074 | 7.16986 |
| **Route 4** | Mean | 10.4 | 6.433333 |
| | Median | 7.5 | 4.5 |
| | 90th percentile | 18 | 17.1 |
| | Standard Deviation | 17.03263 | 8.564898 |

Table 4.2 Percentage of late flights in RegionEx and MDA airlines on each route

| | **RegionEx late flights (in %)** | **MDA late flights (in %)** |
|---|---|---|
| Route 1 | 25.5555555555556 | 28.5714285714286 |
| Route 2 | 28.8888888888889 | 31.0344827586207 |
| Route 3 | 20 | 20 |
| Route 4 | 26.6666666666667 | 26.6666666666667 |

## Question 5: Scatter Plot and the Correlation Coefficient

The correlation coefficient between arrival delay minutes and number of passengers for RegionEx's flights is 0.48. Even though the value is positive it is close to the value 0, and hence the relationship between arrival delay minutes and number of passengers is considered a weaker relationship.




## Question 6: Analysis on Scheduled and Actual Flight Durations

Comparing scheduled flight duration for the two airlines by route

| | Route 1 | Route 2 | Route 3 | Route 4 |
|---|---|---|---|---|
| Regionex | 90 | 90 | 70 | 70 |
| MDA | 100 | 100 | 75 | 75 |

Comparing actual flight duration for the two airlines by route

| | Route 1 | Route 2 | Route 3 | Route 4 |
|---|---|---|---|---|
| Regionex | 106.422 | 108.21 | 81 | 80.4 |
| MDA | 113.71 | 113.62 | 85.2 | 81.43 |

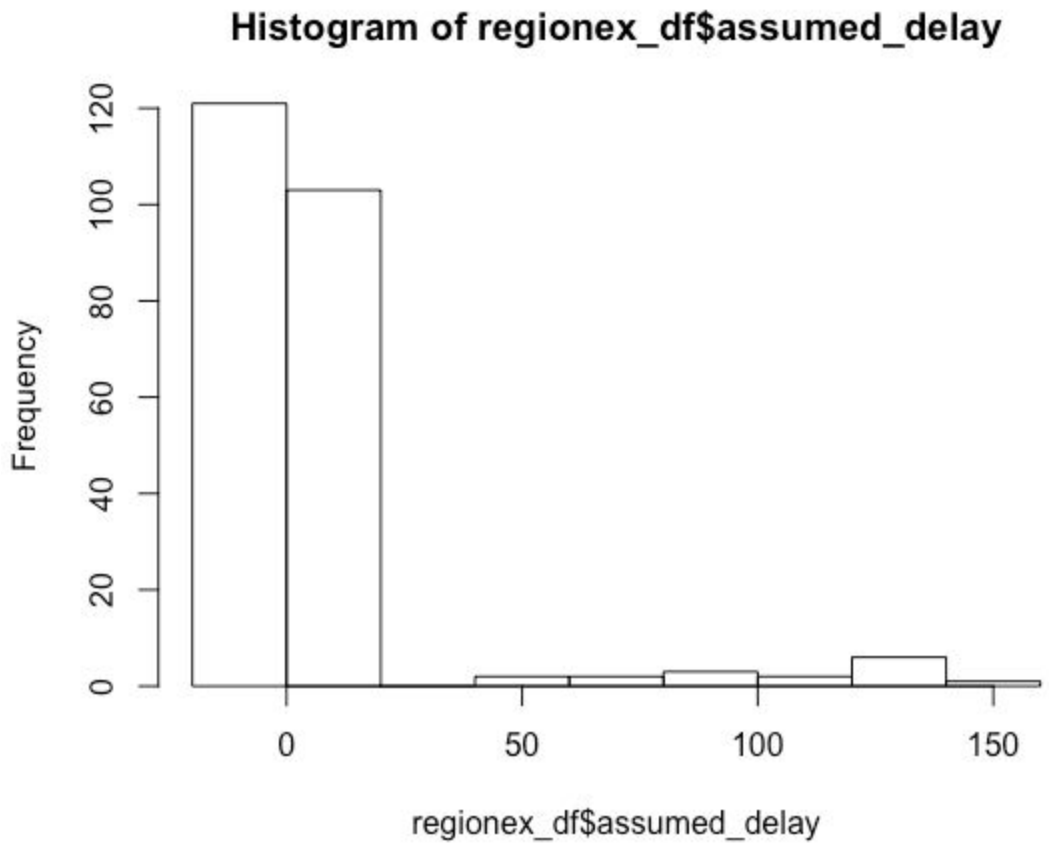
From the table, we can observe the following:

- MDA always schedules for a longer duration compared to Regionex.
- Regionex's actual duration is always better than MDA , but their delays are higher than MDA because they schedule for a shorter duration.

Now, lets compare with the same scheduled duration for both airlines,

| | Route 1 | Route 2 | Route 3 | Route 4 |
|---|---|---|---|---|
| Regionex | 100 | 100 | 75 | 75 |
| MDA | 100 | 100 | 75 | 75 |

Comparing the assumed (same) scheduled duration by route, we find that the delays for regionex are much lower compared to MDA. So with better planning of the scheduled duration, regionex performs much better than MDA.



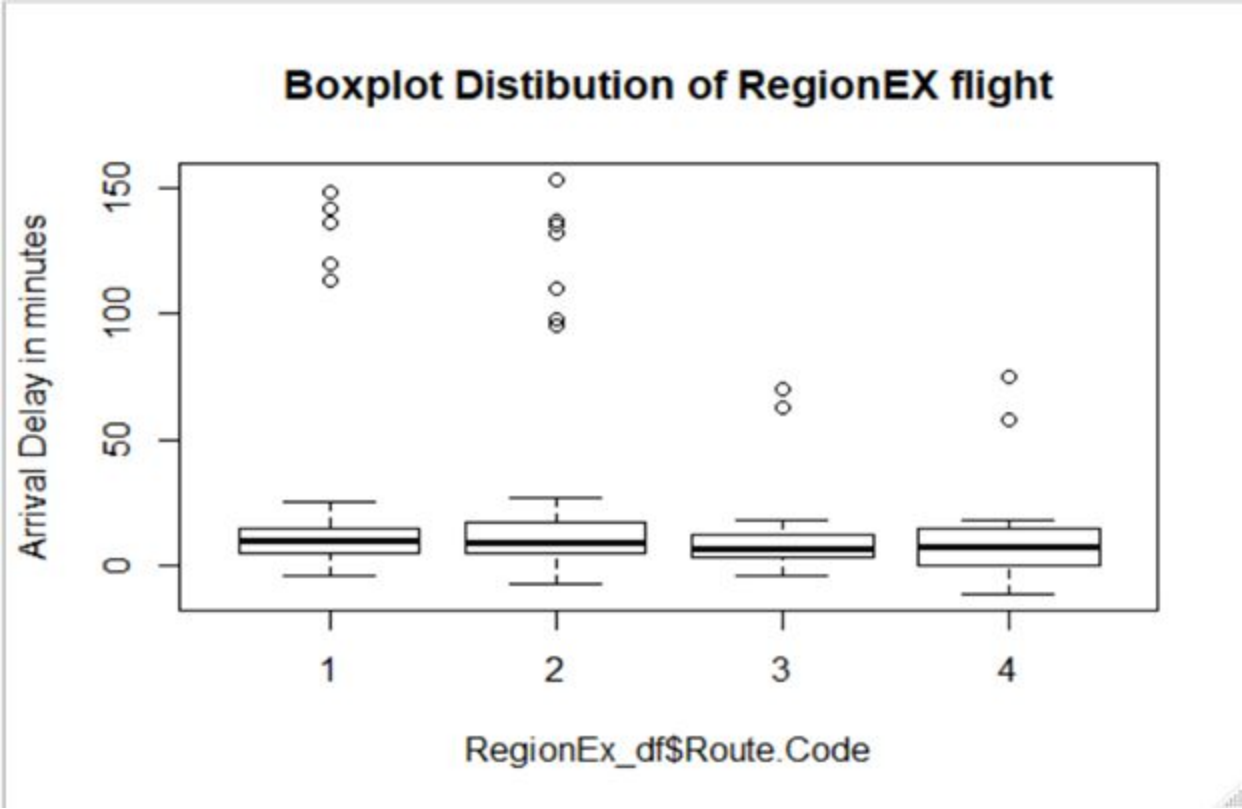

Tabular Display: Descriptive Statistics of RegionEx Airlines with assumed scheduled duration

| **AIRLINE** | **Mean** | **Median** | **90th percentile** | **Standard Deviation** |
|---|---|---|---|---|
| **RegionEx** | 15.6625 | 9 | 21 | 27.65036 |
| **MDA** | 10.92308 | 13 | 16.4 | 6.418123 |
| **RegionEx with Assumed Scheduled Duration** | 6.9125 | 0 | 13 | 27.5 |

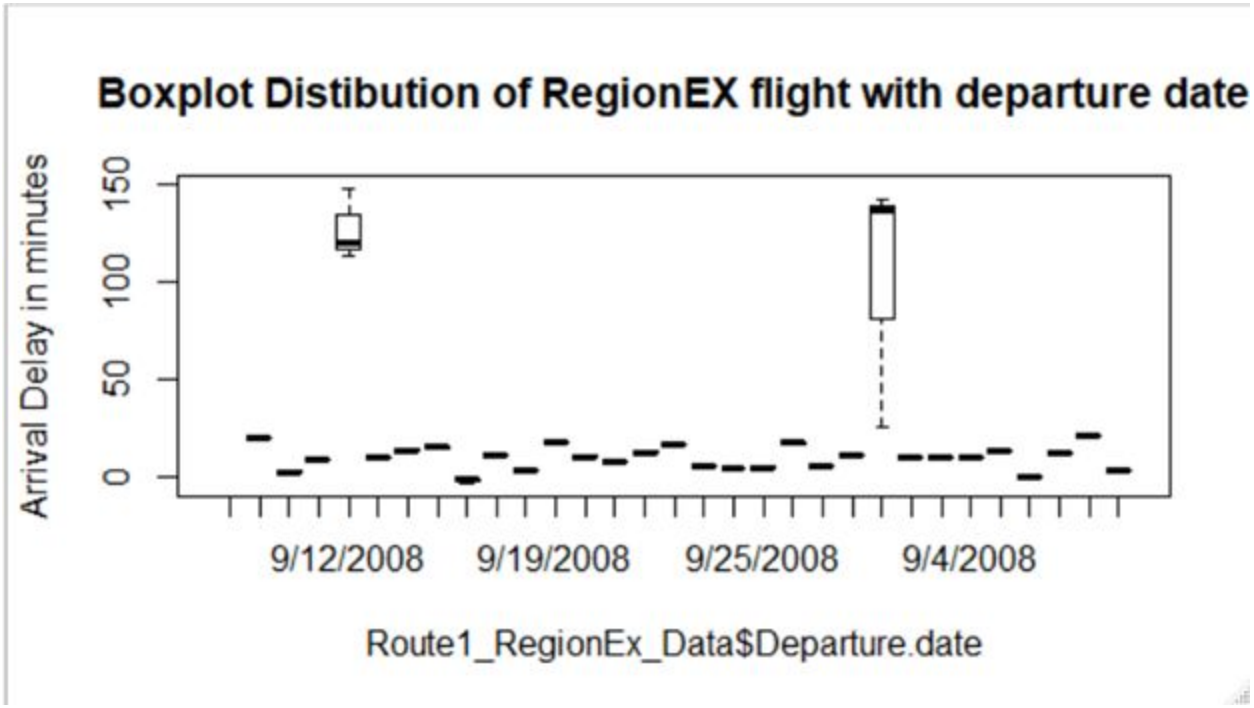
The above data is consistent with our conclusion that regionex performs better than MDA by all metrics with the same scheduled duration.

## Question 7: Other factors that Marion Volero should consider regarding the data analysis

When a boxplot is created for RegionEx flight considering Arrival Delay in minutes over route , we can see that there are more outliers for Route1 and Route2 .(Refer to figure1)




For doing further Analysis for Route 1 of RegionEx flight,a boxplot is created considering arrival delay in minutes over the date of it departure (Refer figure 2)



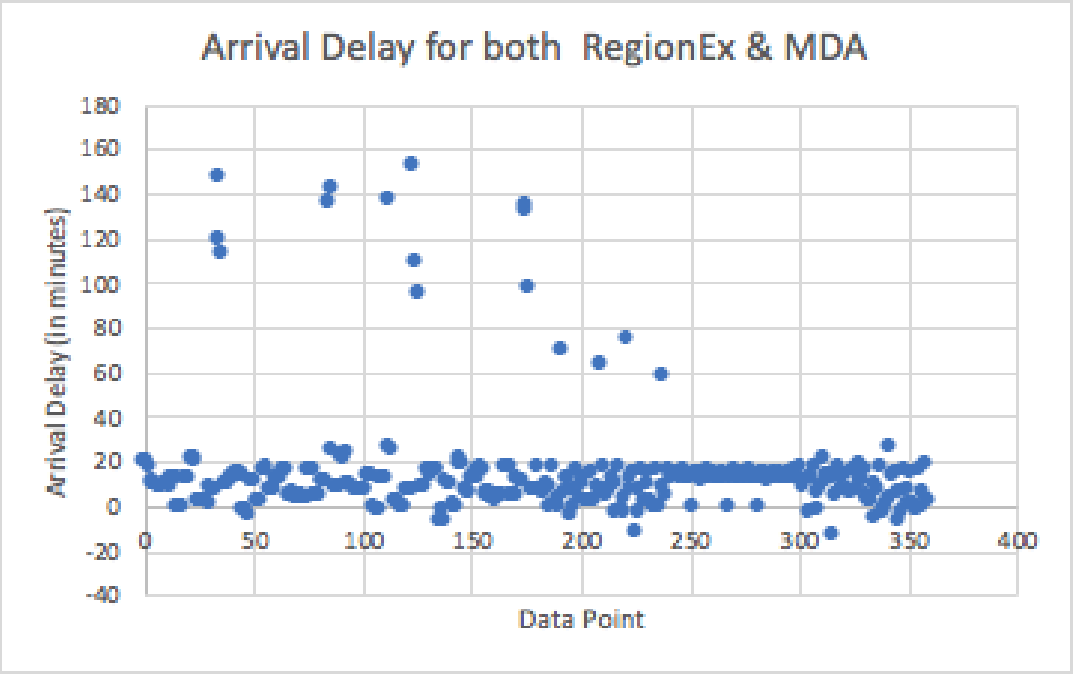

It shows there are a lot of outliers for date 9/12/2008 and 9/29/2008 .RegionEx has most delays in time for route 1 is on 9/12/2008. Similarly on both of the same days i.e on 9/12/2008 and 9/29/2008 RegionEx flight has most delays for route 2 .

On the other hand, MDA flight data for route 1 and roue2 for those two days has an interesting pattern. While all the flights Of MDA for route 1 and route2 were cancelled on 9/12/2008,On 9/12/2008 it had one flight cancellation for Route 1.

The outliers on those two dates for RegionEx flight has remarkably high delays on arrival time that has impacted RegionEx flight's overall performance .MDA has cancelled flight on those days .

So while performing analysis, Marion Volero has to consider there might be some unprecedented circumstances on 9/12/2008 and 9/29/2008 that had caused cancellation of MDA flight and has caused delay for Region Ex flight and has affected its overall performance.

One more factor that Marion Valero could consider is to plan their flights better, as explained below:




As per above scatter plot, there are lots of flights which are delayed between 15 to 20 minutes so as per FAA guidelines they can avoid being considered delayed if they plan their flights better (for eg., they can do so by slightly increasing their schedule flight duration to avoid being marked as delayed)

These are the bulk of the flights which are impacting the ranking of MDA and RegionEx because they are considered delayed by FAA (delay indicator is a major factor that FAA uses while ranking the Airlines).

## Conclusion

Based on the given scheduled and actual durations, we found that MDA was performing better than RegionEx in terms of delay. But it turned out that the performance in terms of actual duration was much better for Regionex.

Once we changed the scheduled duration for Regionex to be the same as MDA, Regionex performed much better than MDA.

We also found that RegionEx's higher delays were on certain days where MDA had cancelled flights on those days.This implies that MDA's delay record could be worse if they had actually operated on those days.