

Insights-2347107

Aryan Majhi 2347107

AKASA - Task 1 (Python + SQL)

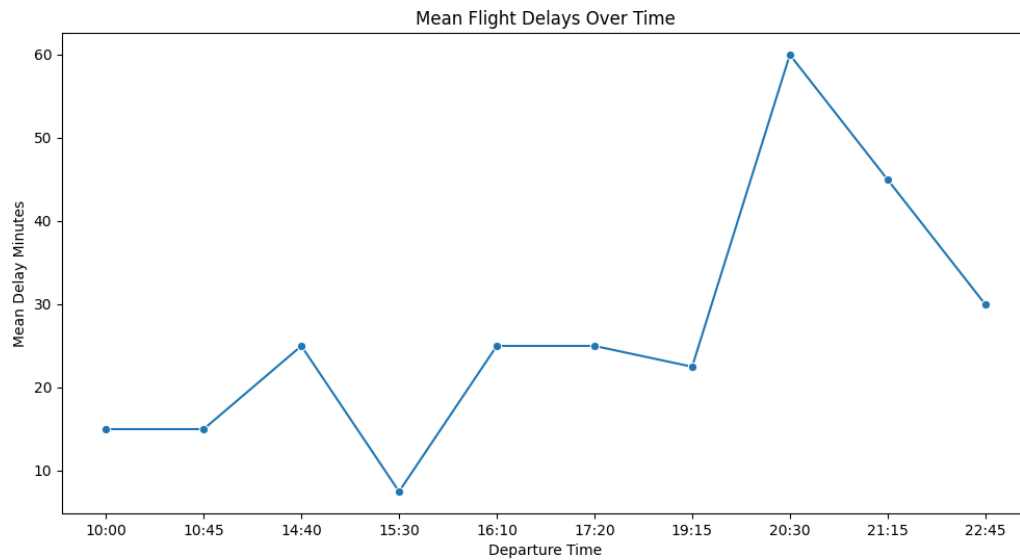
Provide a summary of the key findings from the data

- There are total of 12 observations (rows) and 8 variables (columns).
- There are no duplicate entries in the dataset.
- There were missing values which were handled.
- The statistical summary of Delay Minutes **(after preprocessing)** :
 - Number of non null values are 12
 - The average delays in minutes is 25
 - The amount of dispersion(spread) of data in the dataset is 15.07
 - The minimum delay is 5 minutes
 - The maximum delay is 60 minutes (1 hour)
 - First Quartile(25%) is 15.0, means that 25% of the data points are less than 15.0.
 - Second Quartile(50%) is 25.0, means that 50% of the data points are less than or equal to 25.
 - Third Quartile(75%) is 26.25, means that 75% of the data points are less than or equal to 26.25
- The missing values are handled by using the mean method, because if the columns of NaN would have been replaced with 0 it would be unrealistic as in real-world scenario there are always delays. On an average the delay is 25 mins.
- For determining the significant difference between times between the airlines, I have used the approach of t-tests. T-tests are ideal while dealing with groups and since there are three

different airlines, hence 3 different groups.

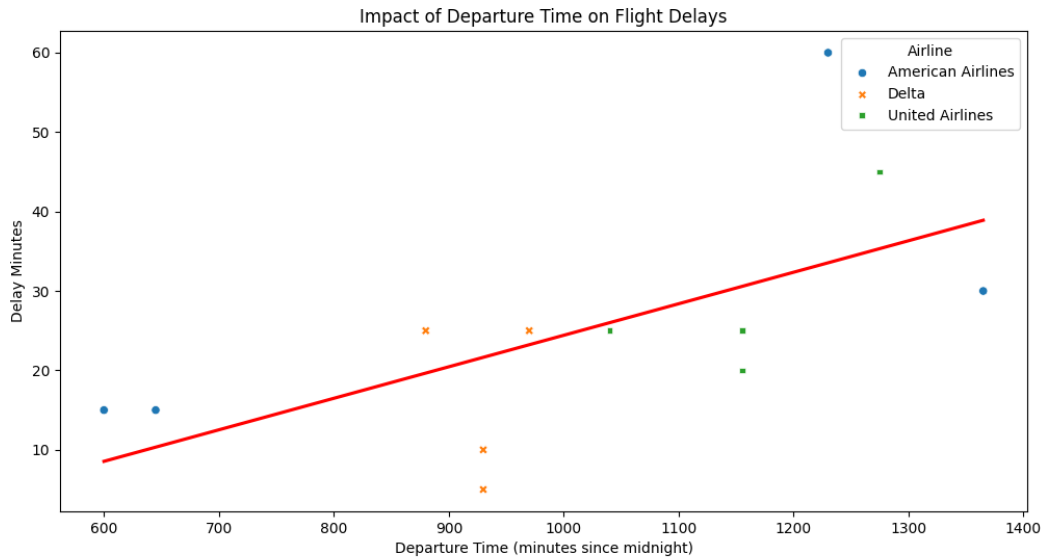
- **Null hypothesis (H0)** is there is no significant difference in the mean delay times between the airlines.
- **Alternative Hypothesis (H1)** is there is a significant difference in the mean delay times between the airlines.
- The alpha value is taken as standard 0.05
- We are performing paired t-test as we can find the difference between delays for any two airlines. The groups are:
 - American Airlines, Delta
 - Delta, United Airlines
 - United Airlines, American Airlines
- After **comparing American Airlines and Delta**:
 - T-statistic: 1.165997668006996, P-value: 0.2878631950549536
 - Conclusion is **fail to reject the null hypothesis**.
- After **comparing Delta and United Airlines**:
 - T-statistic: -1.651445647689541, P-value: 0.14973733521583907
 - Conclusion is **fail to reject the null hypothesis**.
- After **comparing United Airlines and American Airlines**:
 - T-statistic: -0.1044465935734187, P-value: 0.9202189075644182
 - Conclusion is **fail to reject the null hypothesis**.
- We accept the Null Hypothesis for all the airlines. So, there is no significant difference in delays between the airlines.

Analyze the impact of departure times on delays.



This figure shows the delays for time of the day and we can deduce the following points from it:

- The delay pattern is non-linear
- Flights departing after 19:15 have high chances of delay of 25 min or more
- The morning flights have less delays
- The evening flights have more delays
- From this graph, the unusual amount of delays are:
 - At 15:30 with less than 10 mins delay
 - At 20:30 with almost an hour delay



We can infer the following from this graph:

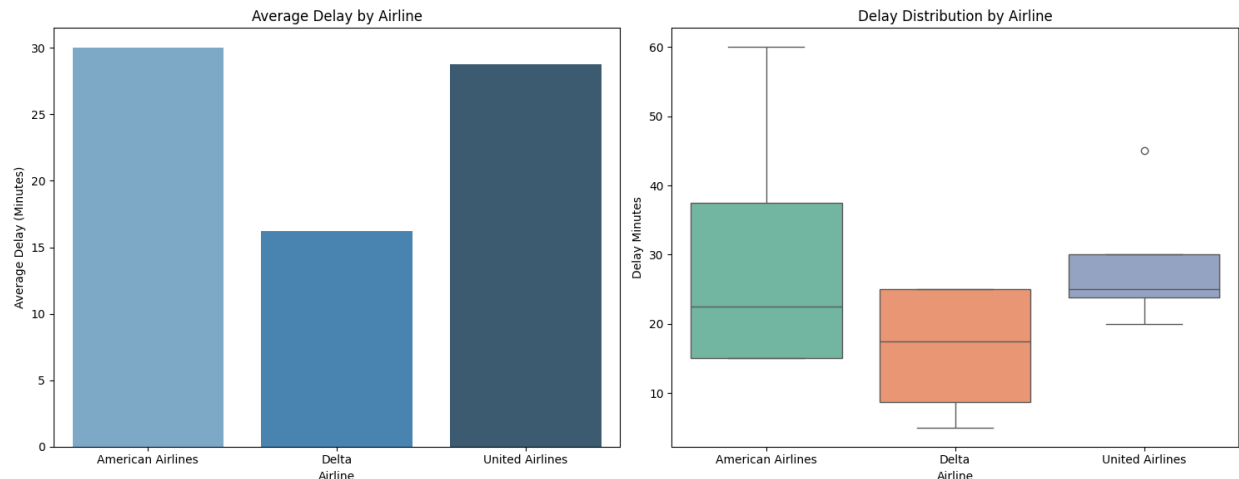
- There is a positive correlation i.e., for increase in x-axis, there is increase in y-axis.
- The delay increases with time i.e., the later in the day it is, the more is the chances of having long delays.
- For American airlines, there is an outlier with around 60 mins delay.
- For United airlines, the delay is within the range of 20-30 mins mostly.
- For Delta, the delay is very less for any given time within the range of 10-30 mins, the highest being below 30 mins

Compare delay distributions between airlines.

- Delay and average delay per airline are as follows:

Airlines	Delay (in mins)	Average (in mins)
American Airlines	120	30.0
Delta	65	16.25
United Airlines	115	28.75
Total	300	75

Visualize the average delay by airline and the delay distribution using appropriate charts.



- Average delay by Airline
 - **American Airlines** has the highest average delay, approximately **30 minutes**.
 - **United Airlines** has a slightly lower average delay, just under **30 minutes**.
 - **Delta** has the lowest average delay, roughly **15 minutes**.
- Delay Distribution by Airline
 - **American Airlines:**
 - The median delay is around **20 minutes**.
 - The **box** spans from about **15 to 40 minutes**, representing the **interquartile range (IQR)**, meaning 50% of the delay times are within this range.
 - The **whiskers** extend up to **60 minutes**, indicating the general spread of delays, without any significant outliers.
 - **Delta:**
 - The **median delay** is around **15 minutes**.
 - The **interquartile range** IQR for Delta is much narrower, ranging from **5 to 25 minutes**, indicating that Delta's delay times are more consistent and less variable.
 - There are no major outliers in Delta's delay distribution.

- **United Airlines:**

- The **median delay** is about **25 minutes**.
 - The **interquartile range** IQR ranges from **20 to 30 minutes**, showing a somewhat consistent delay pattern, but the box is narrower than American Airlines.
 - There is one **outlier** beyond the whiskers, indicating that at least one flight had an unusually long delay.
-

Offer recommendations based on the analysis, such as suggestions for airlines to improve their punctuality.

- The American and United Airlines should increase staff for the later part of the day when the delay is more, since the delay for these airlines are highest.
- The airlines should review the time schedule of their flights and also change the timings so that delays can be minimized.
- The departure of one flight to another should have a significant time gap so that if any flight is late then the other flights after it are not affected.
- The airlines should first check their workforce and capacity and then schedule flights and bookings, as with lesser workforce it is difficult to manage and deliver on time.