# OBJECTIVE

To clean, normalise and analyse the data to derive relevant insights from a given dataset. The dataset includes various details about flights, such as flight numbers, dates, times, airlines, and delays.

Actions performed on the dataset:

1. First I import necessary libraries to jupyter notebook.
2. Load the dataset, aviation_data.csv.
3. Then checked for any null values, I found 2 NaN values in the DelayMinutes column.
4. For the NaN values I chose to use the median instead of the mean because it's less affected by outliers, giving us a more reliable and typical value. Given the small sample size, the median might be more robust.
5. Checked for the duplicate entries, there were none.
6. Created another copy of the dataset to perform some experimental operations while maintaining the integrity of the main dataset.
7. I merged the DepartureDate with the DepartureTime column and ArrivalDate with the ArrivalTime and dropped the unnecessary columns so that I could simplify the data analysis, allowed me to perform operations like calculating the total duration of flights, determining delays, and analysing trends over time more efficiently.
8. Rearrange the columns for better readability.
9. Converted date and time in appropriate format.
10. I have created a function to find the inconsistency in the time entries and correct them.
11. Created a new column for FlightDuration by calculating the difference betweenDepartureTime and ArrivalTime on the same day considering feature engineering.


Data Analysis:

Distribution of Flight Delays
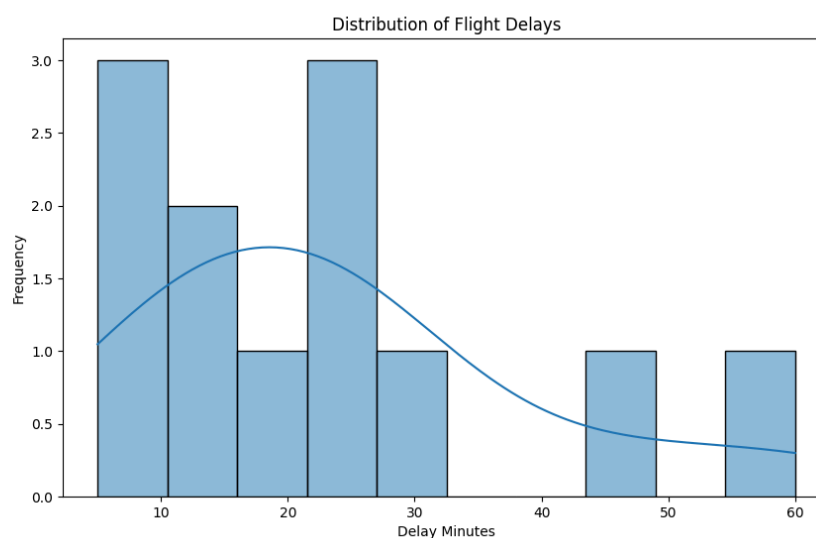
1. These statistics suggest that while most delays are around 23.75 minutes.
2. There is a significant variation, with some delays being much shorter and others much longer.
3. median being close to the mean indicates a relatively symmetric distribution but the presence of a maximum value of 60 minutes suggests there are some outliers and a right-skewed distribution.

Average Delay by Airline
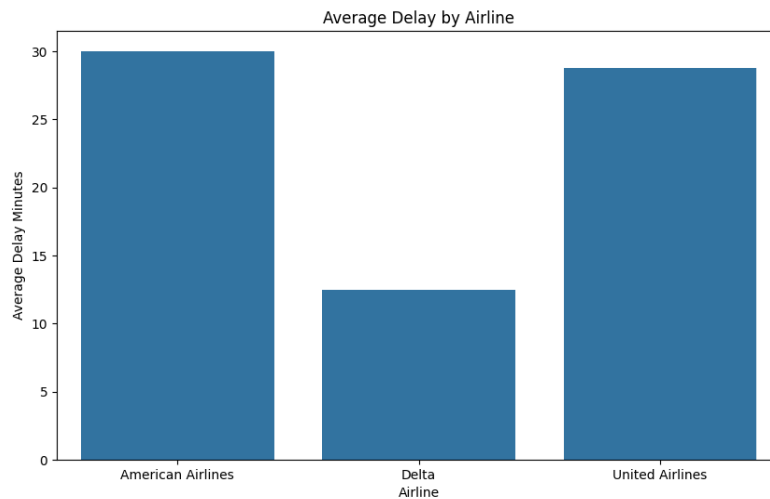
1. This suggests that Delta has the shortest average delay time while American Airlines has the longest.
2. United Airlines falls in between but is closer to American Airlines in terms of delay duration.

Delay Distribution by Airline
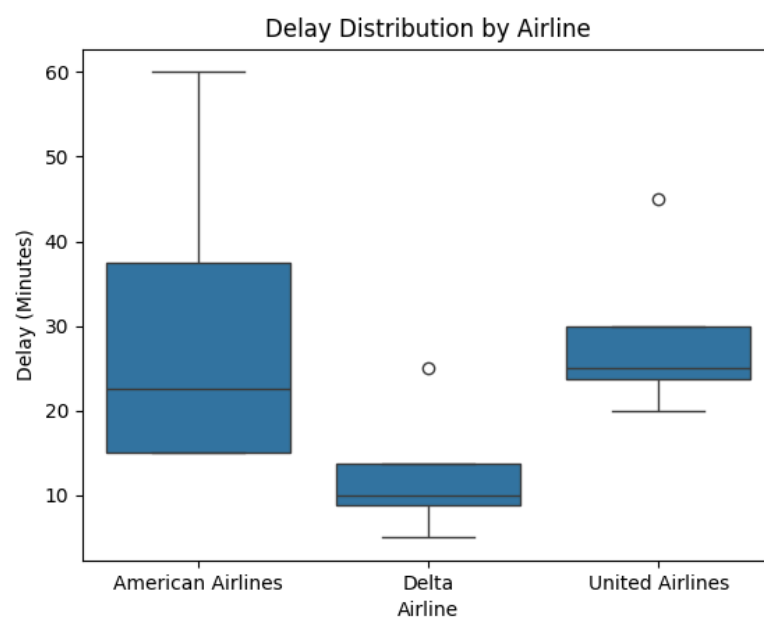
American Airlines:
1. Median delay is around 25 minutes (shown by the line in the middle of the box).
2. The interquartile range (IQR) is large, suggesting a wider variability in delays, with delays ranging between approximately 15 to 40 minutes for the central 50% of the data.
3. There are no extreme outliers, but the maximum delay reaches around 60 minutes.

Delta Airlines:
1. Median delay is very low, around 10 minutes.
2. The IQR is much smaller compared to the other airlines, indicating more consistent and shorter delays.
3. One outlier delay is much higher than the usual, around 30 minutes.

United Airlines:
1. Median delay is around 25 minutes.
2. The IQR is smaller than American Airlines, suggesting less variability in delays, though still more than Delta.
3. There is one outlier delay around 45 minutes.

Flight Delays by Departure Hour
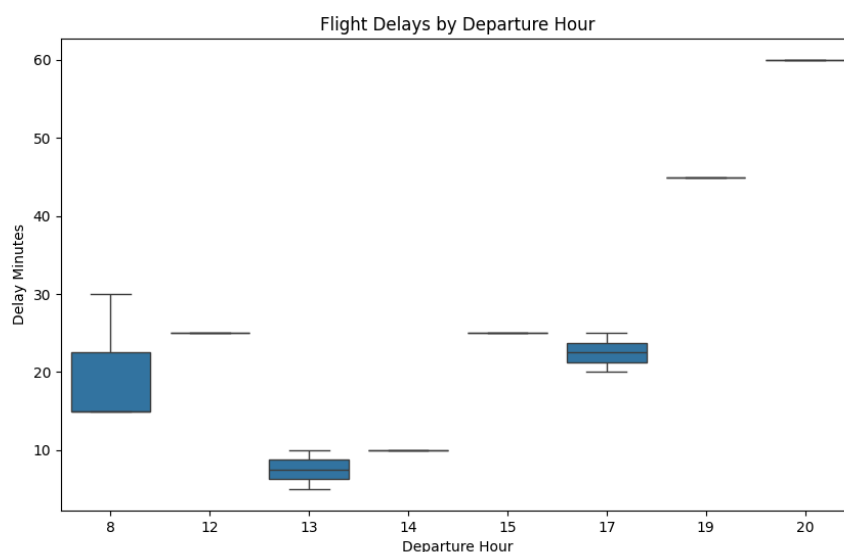
1. The highest delay occurs at 20:00 with 60 minutes.
2. Significant delays are also observed at 19:00 (45.0 minutes) and 12:00 and 15:00 (both 25.0 minutes)
3. The lowest delay is at 13:00 with 7.5 minutes
4. Delays tend to increase significantly in the evening hours, particularly after 17:00

Insights:

- **Time of Day Effect:** The first and most important insight is that delays grow as the day progresses, with evening flights facing the longest wait times. Early afternoon flights experience fewer delays, possibly due to less congestion or more streamlined operations. This pattern could be the result of a build-up of earlier delays, increased air traffic, operational hiccups, or even crew scheduling issues.

- **Airline-Specific Factors:** Different airlines have varying average delay times. This could be tied to factors like operational efficiency, the age of their fleet, the complexity of their routes, or how they handle scheduling.

- **Airport or Air Traffic Control (ATC) Congestion:** The rising delay trend throughout the day might point to increasing congestion at airports or within airspace. It's worth noting that weather could also be a factor, especially for the longer delays, though more context would help clarify this.

Recommendations:

- **Adjust Evening Flight Schedules:** Airlines might consider padding evening flight schedules with extra buffer time to account for the increased likelihood of delays as the day goes on.
- **Boost Turnaround Efficiency:** Improving ground operations can help, especially for airlines that tend to have longer delays (like American Airlines and United Airlines in this case).
- **Better Delay Recovery:** Developing strategies to make up for lost time throughout the day could prevent delays from piling up.
- **Optimise Crew Scheduling:** Ensure that crew rotations are planned with the understanding that delays are more common later in the day.
- **Invest in Predictive Analytics:** Using data to forecast potential delays and tackle them proactively can make a significant difference.
- **Collaborate with ATC and Airports:** Work on reducing congestion during peak hours. This might involve tweaking flight schedules or enhancing air traffic management systems.
- **Fleet Management:** Airlines experiencing higher delays should review their fleet and maintenance schedules to ensure they have the right aircraft available at the right times.
- **Customer Communication:** Enhance systems to keep passengers in the loop about delays and their reasons. Better communication can help manage expectations and reduce frustration.