Week 3: Univariate and Bivariate Visual Analysis Report

Objective

To perform visual exploration of the flight dataset by identifying trends, distributions, and relationships using bar charts, histograms, boxplots, scatterplots, and line plots. The main focus was to understand top airlines, routes, busiest months, flight delays, and airport traffic.

---

1. Top Airlines by Number of Flights

Code Summary: Used sns.countplot() to visualize the number of flights per airline.

Insight:
This graph shows which airlines operated the most flights. Major airlines such as Delta, United, and Southwest dominated the flight volume, while smaller airlines had fewer flights.
It helps identify the key players in the dataset.

---

2. Top 10 Routes by Number of Flights

Code Summary: Created a new column combining Origin and Destination airports, then plotted the top 10 routes using a bar chart.

Insight:
Certain city pairs like *ATL → LAX* or *ORD → DFW* appeared most frequently.
These represent high-demand routes, likely due to their economic or geographic importance.

---

3. Departure Delay Distribution (Histogram & Boxplot)

Code Summary: Used sns.histplot() for the delay distribution and sns.boxplot() to detect outliers.

Insight:
Most flights departed on time or with short delays (< 20 minutes), but a few outliers showed long delays.
The boxplot highlighted the skewed nature of delays and the presence of extreme late departures.

---

4. Flight Count by Month

Code Summary: Used sns.countplot() with months sorted from 1 to 12.

Insight:
The distribution showed seasonal trends — some months (for example, July and December) had higher flight frequencies, possibly due to holiday travel peaks.

---

5. Flight Count by Day of Week

Code Summary: Used sns.countplot() on the DayOfWeek column (1 = Monday ... 7 = Sunday).

Insight:
Weekdays had a higher number of flights, indicating business travel dominance, while weekends saw slightly fewer flights.

---

6. Flight Distribution by Departure Time

Code Summary: Converted DepTime into hourly bins and plotted a countplot.

Insight:
Peak flight activity occurred in the morning (6 – 10 AM) and evening (5 – 9 PM).
These are the most preferred travel times for passengers.

---

7. Flight Distribution by Airport

Code Summary: Mapped airport codes to full names and plotted the top 15 busiest airports using a horizontal bar chart.

Insight:
Airports such as Atlanta (ATL), Chicago O'Hare (ORD), and Dallas/Fort Worth (DFW) emerged as the busiest hubs.
The visualization clearly highlighted the major national and international flight centers.

---

8. Average Departure Delay by Airline

Code Summary: Grouped flights by airline, calculated mean DepDelay, and plotted the results.

Insight:
Some airlines consistently showed higher average delays, indicating potential operational inefficiencies or congestion at specific hubs.

---

9. Arrival Delay Distribution (Histogram & Boxplot)

Code Summary: Similar approach to departure delays, but used the ArrDelay column.

Insight:
Arrival delays followed a similar pattern but had slightly higher variance, suggesting cumulative effects from earlier delays.

---

10. Departure Delay vs Flight Distance (Scatter Plot)

Code Summary: Used sns.scatterplot() to compare flight distance with departure delay.

Insight:
No strong correlation was found — both short and long routes experienced delays.
This suggests weather, congestion, or airline management are stronger factors than distance.

---

11. Average Monthly Delay (Line Plot)

Code Summary: Grouped data by month and plotted average DepDelay using a line chart.

Insight:
Certain months showed higher average delays, likely during peak travel or adverse weather seasons.
This trend helps identify time-based operational bottlenecks.

---

12. Pairwise Delay and Distance Relationships

Code Summary: Used sns.pairplot() on DepDelay, ArrDelay, Distance, and CRSElapsedTime.

Insight:
Pairwise plots revealed mild relationships between departure and arrival delays, and expected correlations between flight distance and scheduled time.

Summary of Findings

Category	Visualization	Key Insight
Top Airlines	Bar Chart	Major carriers dominate the dataset
Top Routes	Bar Chart	High-demand routes between major hubs
Delay Distributions	Histogram & Boxplot	Delays are skewed, with outliers
Busiest Months	Bar Chart	Seasonal peaks in summer and winter
Day of Week	Bar Chart	Weekday dominance for business travel
Departure Time	Countplot	Morning and evening flight peaks
Busiest Airports	Horizontal Bar	ATL, ORD, DFW are top airports
Airline Delays	Bar Chart	Some airlines have consistently higher delays
Distance vs Delay	Scatter Plot	No clear correlation
Monthly Delay	Line Plot	Delay spikes in certain months
Pairwise Analysis	Pairplot	Relationship between Dep & Arr delays

---

Conclusion

Through univariate and bivariate visual analysis, the dataset revealed valuable insights into:

- Airline performance and route popularity
- Seasonal and daily flight patterns
- Delay behavior and operational efficiency
- The dominance of major airports

These analyses provide a solid foundation for building predictive models and optimizing airline operations.