Week 6: Seasonal and Cancellation Analysis

Objective

This week's analysis focuses on understanding how flight cancellations vary across months and seasons, identifying common causes of delays, and exploring how external factors like winter months or holidays may influence flight disruptions.

---

1. Monthly Cancellation Trends

Code Section:

```
monthly_cancellations = (
  df[df['Cancelled'] == 1]
  .groupby('Month')
  .size()
  .reindex(month_order, fill_value=0)
)
```

Visualization:
A bar chart showing the total number of cancelled flights for each month.

Insight:

- Identifies peak cancellation months (for example, December or January).
- Seasonal spikes may indicate weather-related issues or holiday congestion.
- The pattern helps assess how stable flight operations are throughout the year.

---

2. Cancellation Causes (Carrier, NAS, Security)

Code Section:

```
cancelled_flights = df[df['Cancelled'] == 1][['CarrierDelay', 'NASDelay', 'SecurityDelay']].sum()

cancelled_flights.plot(kind='bar', color='#817282')
```

Visualization:
A bar chart representing the total delay (in minutes) caused by different factors.

Insight:

- CarrierDelay represents airline-specific issues such as crew scheduling or maintenance.
- NASDelay reflects air traffic control or airport capacity problems.
- SecurityDelay may increase during stricter travel periods or national events.
- This graph shows which factor contributes the most to overall cancellations.

---

3. Heatmap of Cancellation Rates by Month and Airline

Code Section:

```
pivot_cancel = df.groupby(['Month', 'Airline'])['Cancelled'].mean().unstack()

sns.heatmap(pivot_cancel, cmap='YlGnBu', annot=True)
```

Visualization:
A heatmap comparing cancellation rates across airlines and months.

Insight:

- Enables visual comparison of airline reliability over time.
- Darker cells indicate higher cancellation rates.
- Useful for benchmarking airline performance and identifying operational consistency.

---

4. Winter vs Non-Winter Cancellations

Code Section:

```
df['Season'] = df['Month'].apply(lambda m: 'Winter' if m in [11, 12, 1] else 'Non-Winter')
```

Visualization:
A bar chart comparing cancellation rates between winter (November–January) and non-winter months.

Insight:

- Highlights how cold-weather conditions impact flight reliability.
- Winter months generally show higher cancellations due to snow, fog, and operational delays.
- The difference between the two bars clearly shows the seasonal effect.

---

5. Cancellations by Season (Pie Chart)

Code Section:

```
df['Season'] = df['Month'].apply(season_from_month)

seasonal = df.groupby('Season')['Cancelled'].mean() * 100

plt.pie(seasonal.values, labels=seasonal.index, autopct='%1.1f%%')
```

Visualization:
A pie chart showing the percentage of cancellations across all four seasons (Winter, Spring, Summer, Autumn).

Insight:

- Shows the seasonal proportion of cancellations.
- Winter may dominate the chart, confirming the environmental impact.
- Provides a complete year-wide overview of cancellation trends.

---

6. Pairwise Delay Relationships (Pairplot)

Code Section:

```
sns.pairplot(df[delay_cols])
```

Visualization:
A pairplot showing scatter and distribution plots between delay types (Carrier, NAS, Security).

Insight:

- Reveals whether delays are correlated, for example, NASDelay and CarrierDelay rising together.
- Identifies clusters or shared causes for delays.
- Helps in understanding interdependencies among delay factors.

---

7. Carrier Delay Distribution by Season (Violin Plot)

Code Section:

```
sns.violinplot(x='Season', y='CarrierDelay', data=df)
```

Visualization:
A violin plot showing how carrier delays vary across different seasons.

Insight:

- Displays the spread and concentration of delay values.
- Wider sections indicate a higher frequency of delays at those levels.
- Helps identify if certain seasons have extreme or consistent delays.

Overall Summary

Aspect	Insight
Monthly Trends	Certain months (especially winter) show higher cancellation counts.
Delay Causes	Carrier-related delays are often the major contributor.
Airline Reliability	Varies significantly month-to-month.
Seasonal Impact	Winter months exhibit higher cancellation rates.
Delay Correlation	Some delay types are interlinked, suggesting systemic issues.

Conclusion

Through this week's analysis, we identified that:

- Cancellations spike during winter, likely due to weather and operational challenges.
- Carrier delays remain the leading cause, followed by NAS delays.
- Airline performance is not uniform across the year, reflecting differences in capacity or management.