EXPERIMENT 2

AIM:

The goal of this project is to analyze and visualize **flight-related data** from the `flights.csv` dataset using **matplotlib and seaborn**. The dataset contains details such as **Airline, Journey Date, Source, Destination, Departure Time, Duration, Stops, Additional Info, and Price**.
PROCEDURE:

Step 1: Importing Required Libraries

- Open **Google Colab** and create a new notebook.
- Import essential libraries such as **pandas, matplotlib, seaborn**, and **google.colab.files** to handle CSV uploads and data visualization.

Step 2: Uploading the Dataset

- Use `files.upload()` to manually upload the `flights.csv` dataset to Google Colab.
- Read the dataset into a **pandas DataFrame** using `pd.read_csv()`.
- Display the **first few rows** using `df.head()` to understand the structure of the dataset.

Step 3: Data Preprocessing

- Convert the `Date_of_Journey` column to **datetime format** using `pd.to_datetime()`.
- Extract **day, month, and year** as separate columns for further analysis.
- Handle missing values or incorrect data formats if necessary.

Step 4: Generating Visualizations

1 **Number of Flights Per Airline**

- Use **seaborn's countplot** to create a **bar chart** representing the number of flights for each airline.
- Helps in identifying which airline operates the most/least flights.

2 **Flight Prices by Airline (Box Plot)**

- Use a **box plot** to compare ticket prices across different airlines.
- Helps in identifying **price variations** among airlines and detecting **outliers**.

3 Number of Flights Per Month

- Use a **bar chart** to visualize the **flight distribution by month**.
- Helps in analyzing seasonal trends in flight operations.

4 Flight Price Distribution (Histogram)

- Use a **histogram** with **KDE (Kernel Density Estimation)** to visualize how flight ticket prices are distributed.
- Helps in understanding the **most common ticket price range**.

5 Flight Duration vs. Ticket Price (Scatter Plot)

- Convert flight duration from hours/minutes into a **numeric format** (in minutes).
- Use a **scatter plot** to compare flight duration with ticket prices.
- Helps in identifying whether **longer flights cost more**.

Step 5: Displaying the Results

- Show all five visualizations using `plt.show()`.
- Analyze insights derived from the visualizations.

Step 6: Conclusion

- Interpret the results based on visual trends.
- Provide recommendations for airlines, passengers, or travel agencies.
- Suggest future enhancements, such as **predicting flight prices using machine learning models**.

CODE:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files

# 📌 Upload and Load Dataset
uploaded = files.upload()
file_name = list(uploaded.keys())[0]
df = pd.read_csv(file_name)

# Convert Date_of_Journey to datetime format
```

```python
df['Date_of_Journey'] = pd.to_datetime(df['Date_of_Journey'])

# Extract 'Day', 'Month', 'Year' from 'Date_of_Journey'
df['Day'] = df['Date_of_Journey'].dt.day
df['Month'] = df['Date_of_Journey'].dt.month
df['Year'] = df['Date_of_Journey'].dt.year

# Display dataset info
print(df.head())

# 📌 1️⃣ Visualization: Number of Flights Per Airline
plt.figure(figsize=(12, 6))
sns.countplot(x='Airline', data=df, palette="coolwarm",
order=df['Airline'].value_counts().index)
plt.xlabel("Airline")
plt.ylabel("Number of Flights")
plt.title("✈️ Number of Flights Per Airline")
plt.xticks(rotation=45)
plt.show()

# 📌 2️⃣ Visualization: Flight Prices by Airline
plt.figure(figsize=(12, 6))
sns.boxplot(x='Airline', y='Price', data=df, palette="viridis")
plt.xlabel("Airline")
plt.ylabel("Ticket Price (INR)")
plt.title("💰 Flight Prices by Airline")
plt.xticks(rotation=45)
plt.show()

# 📌 3️⃣ Visualization: Number of Flights Per Month
plt.figure(figsize=(10, 5))
sns.countplot(x='Month', data=df, palette="magma")
plt.xlabel("Month")
plt.ylabel("Number of Flights")
plt.title("📅 Monthly Flight Distribution")
plt.xticks(rotation=45)
plt.show()

# 📌 4️⃣ Visualization: Flight Price Distribution
plt.figure(figsize=(10, 5))
```

```python
sns.histplot(df['Price'], bins=50, kde=True, color="blue")
plt.xlabel("Ticket Price (INR)")
plt.ylabel("Frequency")
plt.title("📊 Flight Price Distribution")
plt.show()

# 📌 5️⃣ Visualization: Duration of Flights
plt.figure(figsize=(10, 5))
df['Duration_in_Minutes'] = df['Duration'].str.replace('h',
'*60').str.replace('m', '').str.replace(' ', '+').apply(eval)
sns.scatterplot(x=df['Duration_in_Minutes'], y=df['Price'],
hue=df['Airline'], palette="coolwarm")
plt.xlabel("Flight Duration (Minutes)")
plt.ylabel("Ticket Price (INR)")
plt.title("⏳ Flight Duration vs. Ticket Price")
plt.show()
```
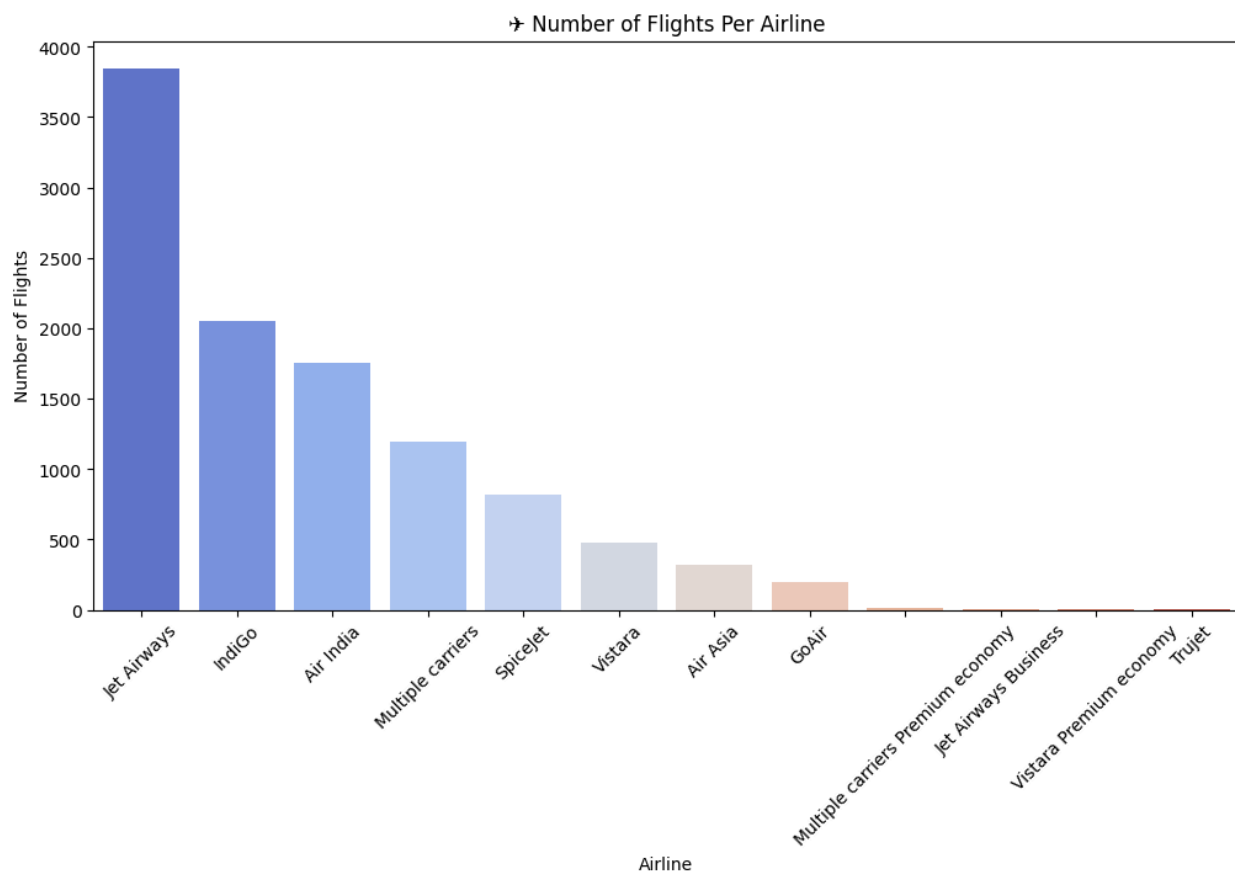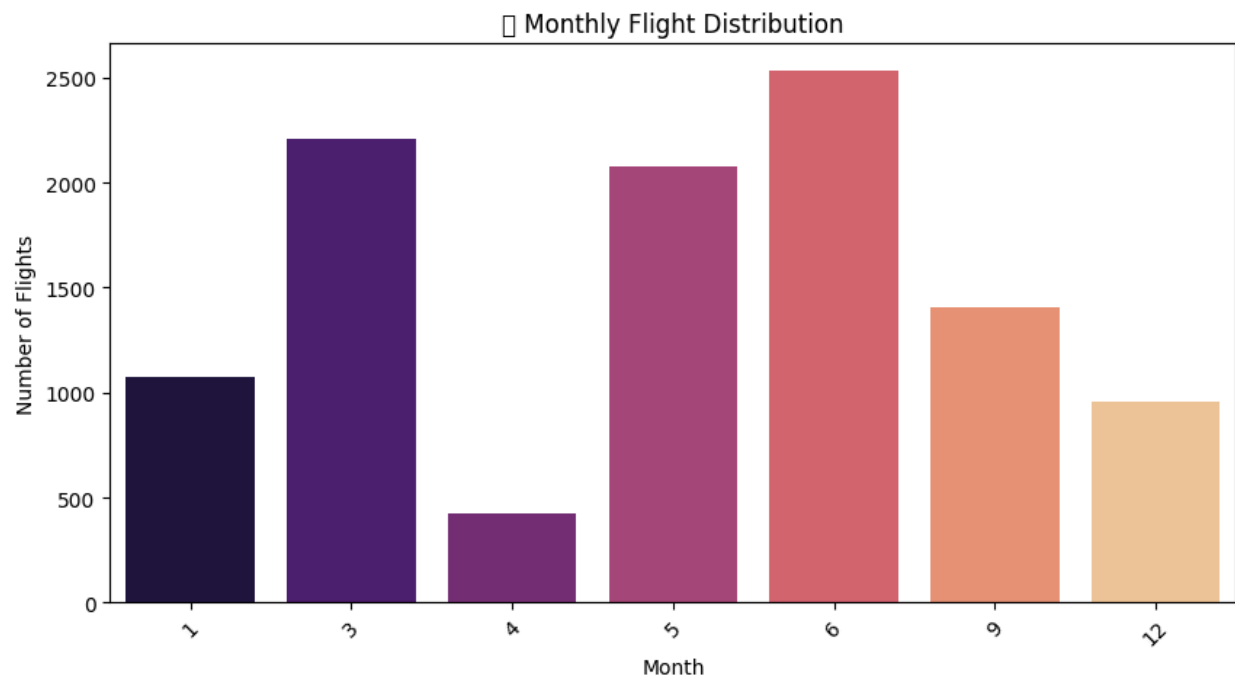
OUTPUT:

🛫 Flight Prices by Airline

Flight Price Distribution

Flight Duration vs. Ticket Price

⬚ Monthly Flight Distribution

✈ Number of Flights Per Airline
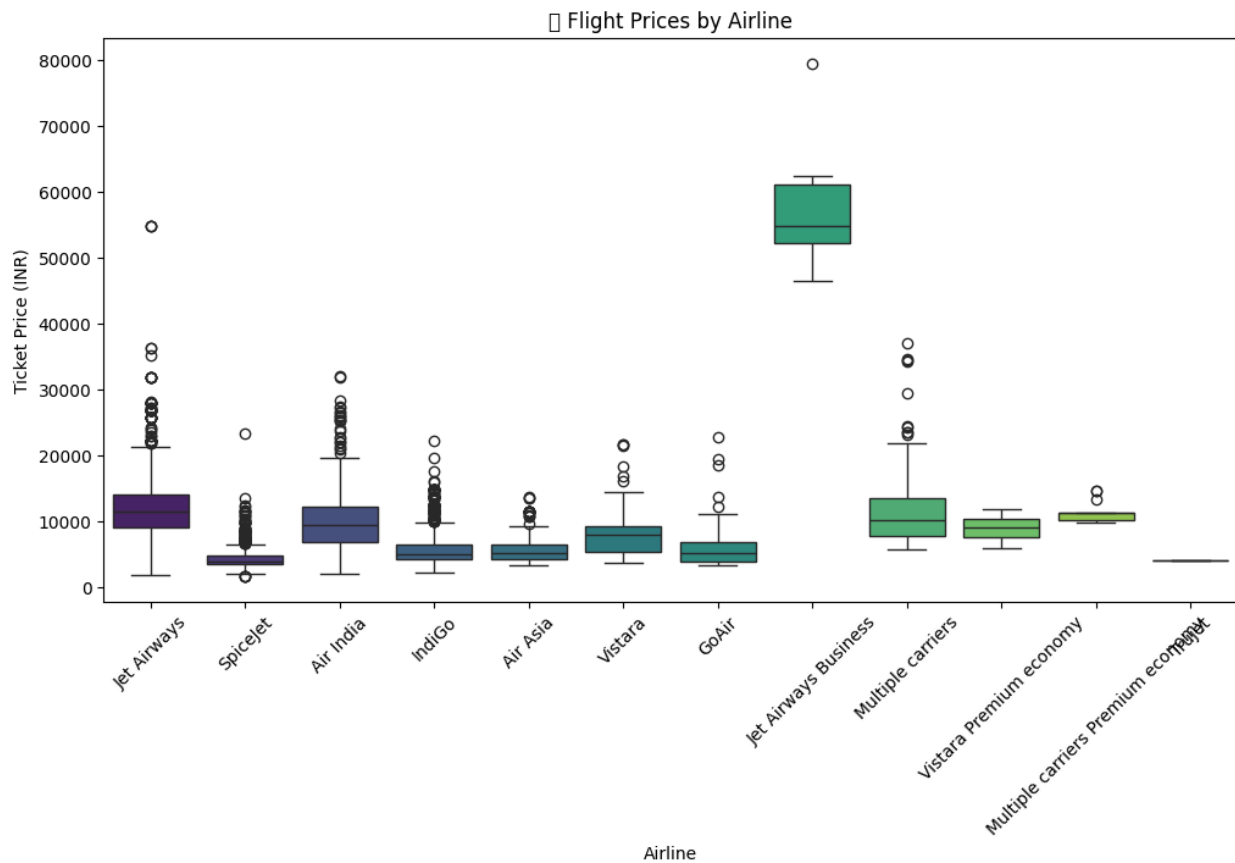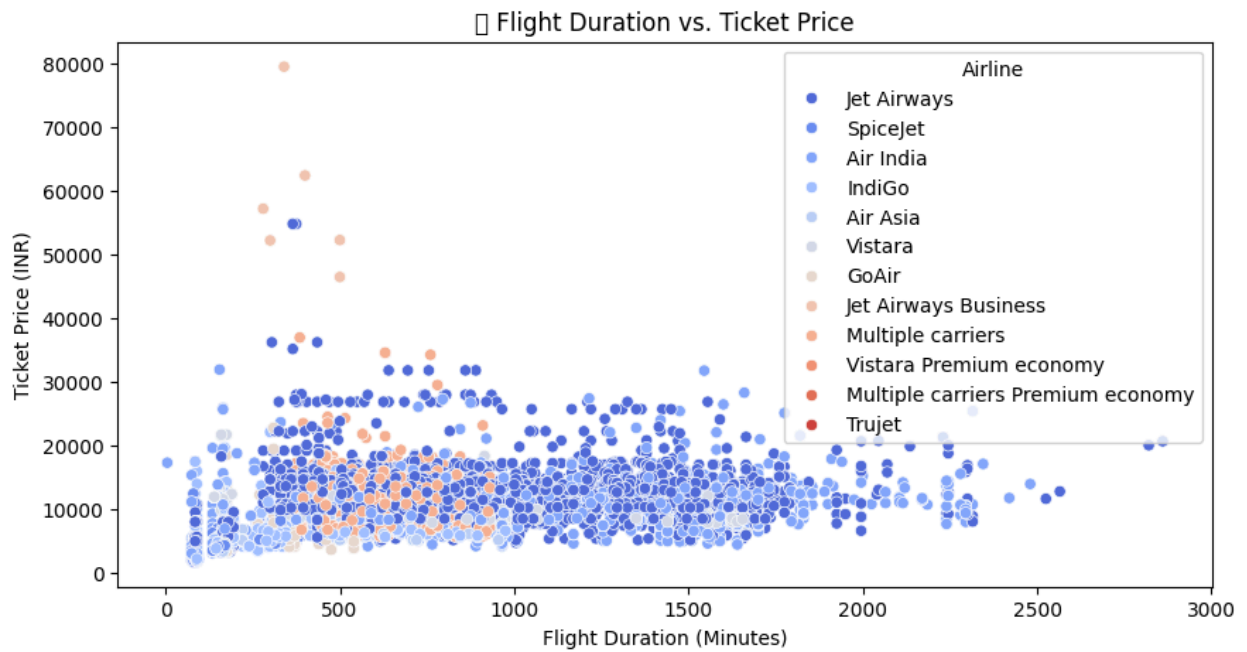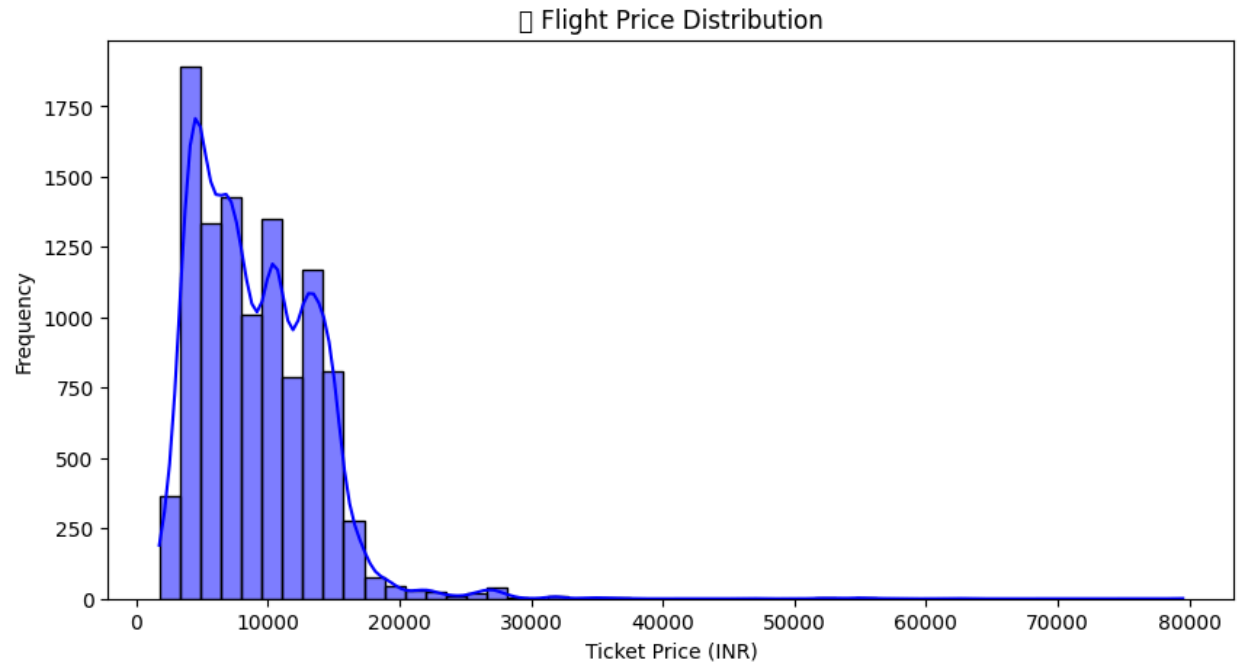
RESULT:

Display all visualizations using `plt.show()` and analyze trends (e.g., airline operations, price distribution, seasonal trends).