FLIGHT TICKET Dataset Project EDA Report

1. UNDERSTANDING THE GIVEN TASK

Task Given by instructor Atharva Kulkarni

Performed By PRASHANT B. AGHAO

- Solve in ipynb:

1. Import the dataset. [flight.ipynb]\flight_price_1.csv
2. Do proper EDA(analysis) of your dataset and create a report
3. Then perform necessary preprocessing steps

+ Code + Markdown

2. DATA COLLECTION AND UNDERSTANDING

- I used a dataset on **FLIGHT TICKET PRICE** from UCI.

2.1 ObjectivesClearly state the goals of the exploratory data analysis, such as identifying patterns, trends, and anomalies in the flight ticket data.

Empty markdown cell, double-click or press enter to edit.

3. DATA EXPLORATION

- In this step, we will apply Exploratory Data Analysis (EDA) to extract insights from the data set to know which features have contributed more in predicting Forest fire by performing Data Analysis using Pandas and Data visualization using Matplotlib & Seaborn. It is always a good practice to understand the data first and try to gather as many insights from it.

Below are tasks to be performed in EDA:

1. Importing Libraries

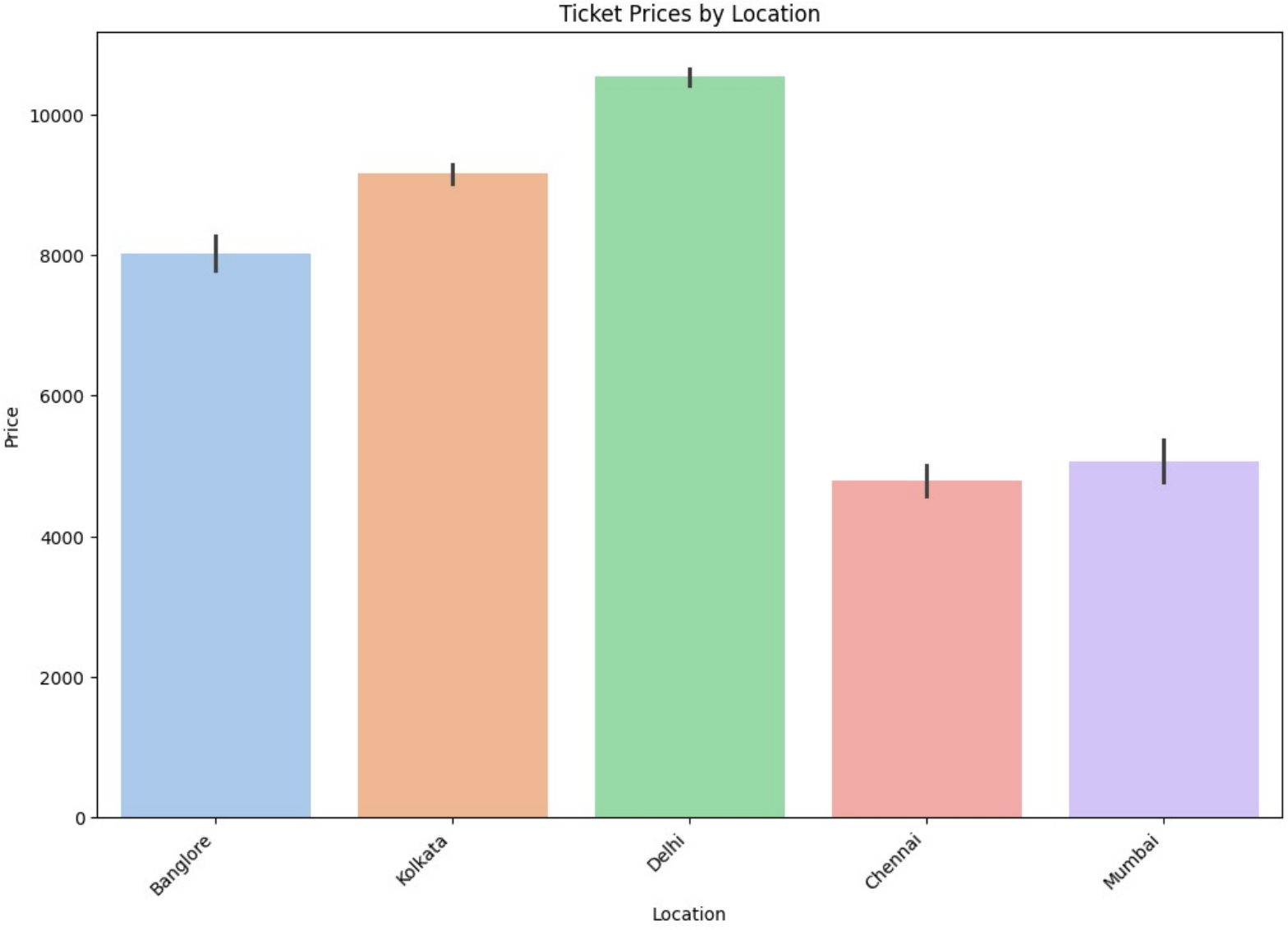
2. Data Cleaning for EDA Report

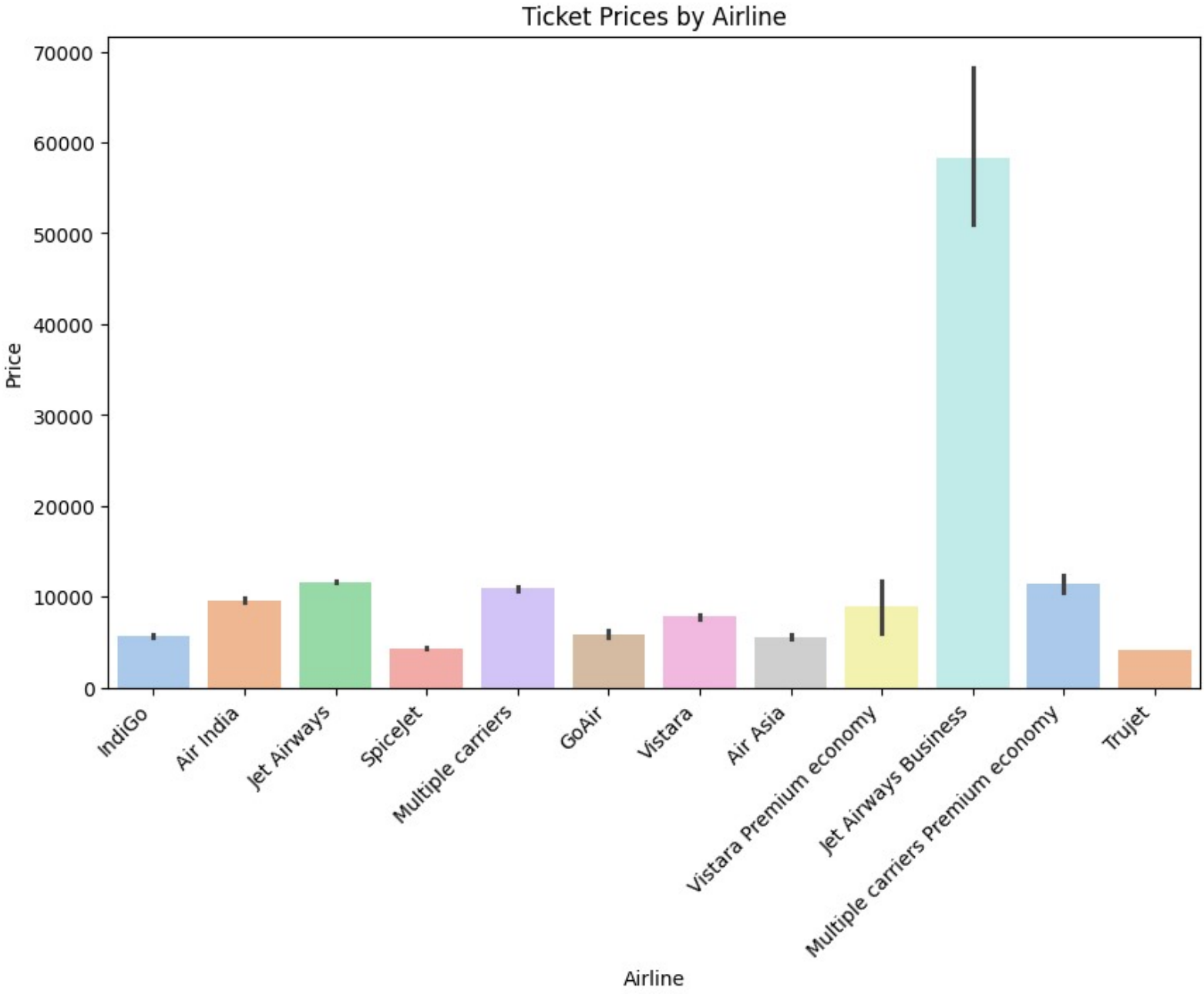
3. Exploratory Data Analysis (EDA) on all Features

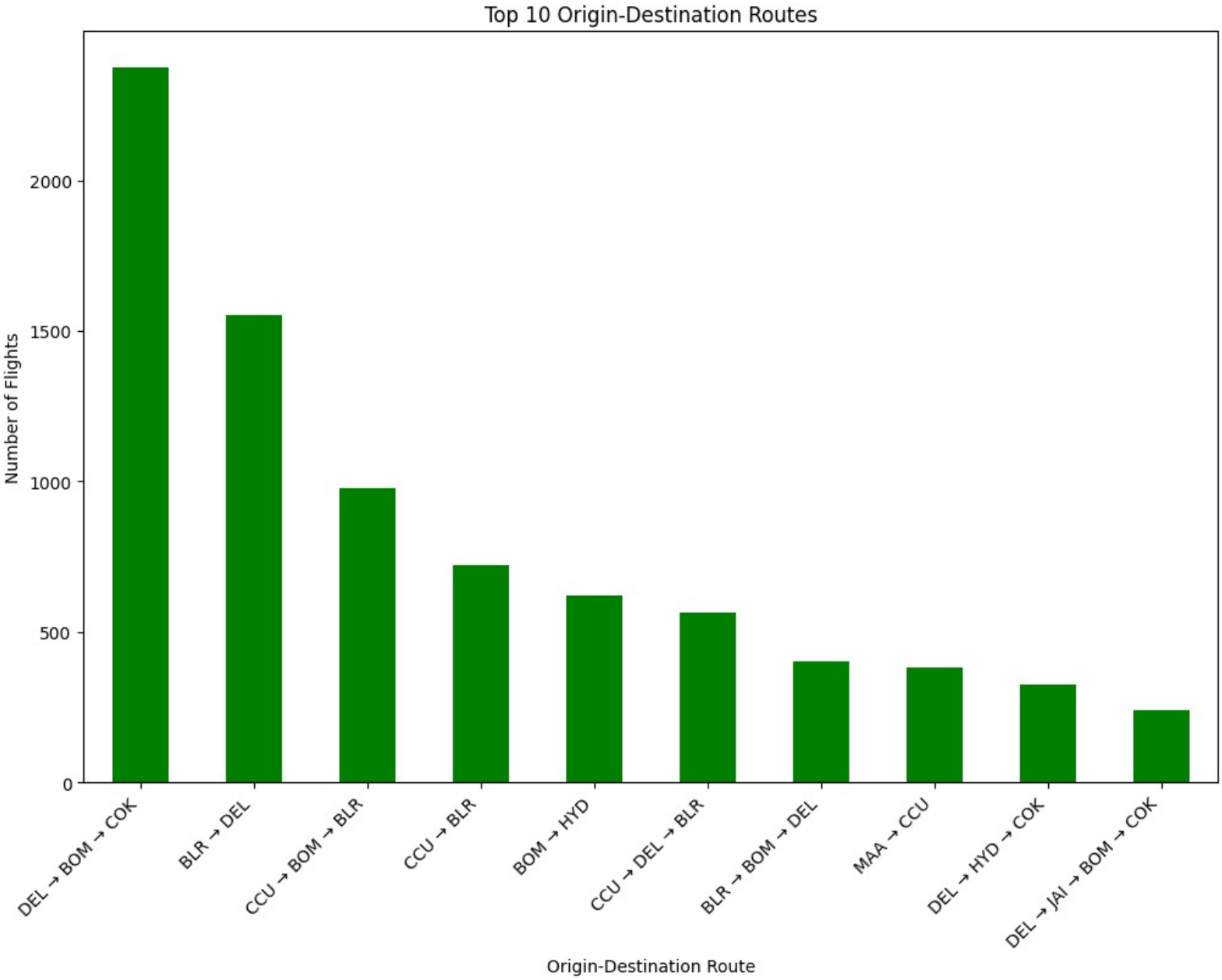
	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302

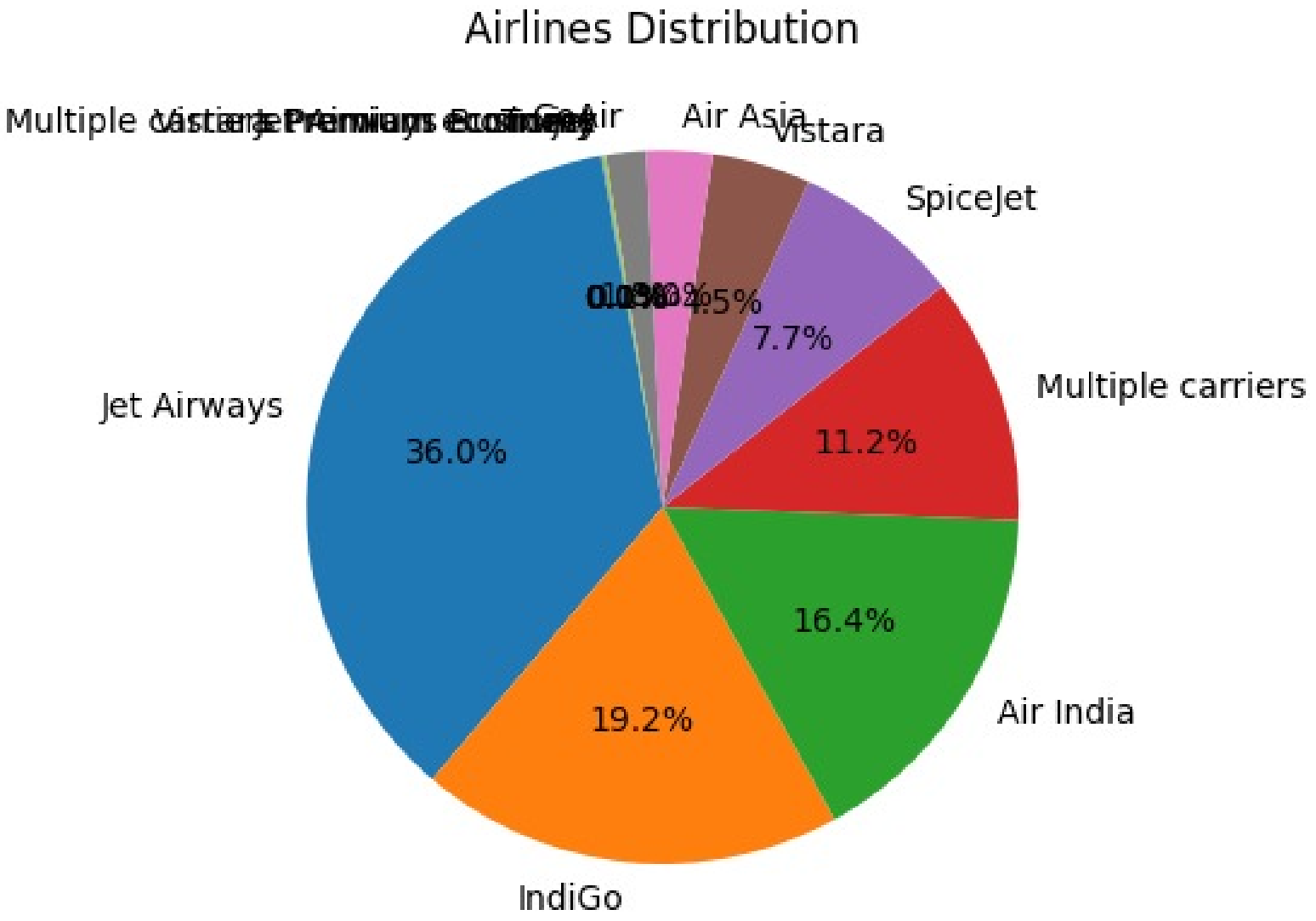
```python
# Overview of the dataset
print("\nDataset Overview:")
print(df.info())
```

```
Dataset Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   Airline          10683 non-null  object
 1   Date_of_Journey  10683 non-null  object
 2   Source           10683 non-null  object
 3   Destination      10683 non-null  object
 4   Route            10682 non-null  object
 5   Dep_Time         10683 non-null  object
 6   Arrival_Time     10683 non-null  object
 7   Duration         10683 non-null  object
 8   Total_Stops      10682 non-null  object
 9   Additional_Info  10683 non-null  object
 10  Price            10683 non-null  int64 
dtypes: int64(1), object(10)
```


Ticket Prices by Location
Price
10000
8000
6000
4000
2000
0
Banglore
Kolkata
Delhi
Chennai
Mumbai
Location


Ticket Prices by Airline
Price
70000
60000
50000
40000
30000
20000
10000
0
IndiGo
Air India
Jet Airways
SpiceJet
Multiple carriers
GoAir
Vistara
Air Asia
Vistara Premium economy
Jet Airways Business
Multiple carriers Premium economy
Trujet
Airline


Top 10 Origin-Destination Routes
Number of Flights
0
500
1000
1500
2000
DEL → BOM → COK
BLR → DEL
CCU → BOM → BLR
CCU → BLR
BOM → HYD
CCU → DEL → BLR
BLR → BOM → DEL
MAA → CCU
DEL → HYD → COK
DEL → JAI → BOM → COK
Origin-Destination Route


Airlines Distribution
Air Asia
Vistara
SpiceJet
7.7%
Multiple carriers
11.2%
Jet Airways
36.0%
16.4%
Air India
19.2%
IndiGo

+ Code + Markdown

REPORT

- JET AIRWAYS BUSINESS have highest ticket prizes.
- DELHI have most no of departure flight.
- JET AIRWAYS have most airlines distribution'
- Most flight fly between DELHI-COCHIN.