

Continue with Visualizations

# VISUAL ANALYSIS

```
In [45]: df.head()
```

Out[45]:

	Airline	Source	Destination	Total_Stops	Additional_Info	Price	Date	Month	Year	Arrival_Hour	Arrival_Minute	Dep_Hour	Dep_Min	Duration_hr	Duration_min
0	IndiGo	Banglore	New Delhi	0	No info	3897	24	3	2019	1	10	22	20	2	50
1	Air India	Kolkata	Banglore	2	No info	7662	1	5	2019	13	15	5	50	7	25
2	Jet Airways	New Delhi	Cochin	2	No info	13882	9	6	2019	4	25	9	25	19	0
3	IndiGo	Kolkata	Banglore	1	No info	6218	12	5	2019	23	30	18	5	5	25
4	IndiGo	Banglore	New Delhi	1	No info	13302	1	3	2019	21	35	16	50	4	45

## 1) Does the number of stops (Total\_Stops) impact the flight price?

```
In [46]: plt.figure(figsize=(5,3))

sns.barplot(x = 'Total_Stops', y = 'Price', data = df, color='purple', linewidth = 1.3, edgecolor='black', errorbar=None)
plt.xlabel("Total Stops", fontsize=12, fontweight='bold')
plt.ylabel("Price", fontsize=12, fontweight='bold')
plt.title("Stops vs Price", fontsize=12, fontweight='bold')

plt.show()
```



**Observation** - As the number of stops increases, the price of flight increases.

## 2) Which airlines have the highest and lowest prices on average?

```
In [47]: avg_price = df.groupby('Airline')['Price'].mean().reset_index()
filtered_avg_price = avg_price[avg_price['Airline'] != 'Jet Airways Business']
display(filtered_avg_price)
```

	Airline	Price
0	Air Asia	5590.260188
1	Air India	9606.804112
2	GoAir	5861.056701
3	IndiGo	5673.682903
4	Jet Airways	11643.923357
6	Multiple carriers	10902.678094
7	Multiple carriers Premium economy	11418.846154
8	SpiceJet	4338.284841
9	Trujet	4140.000000
10	Vistara	7796.348643
11	Vistara Premium economy	8962.333333

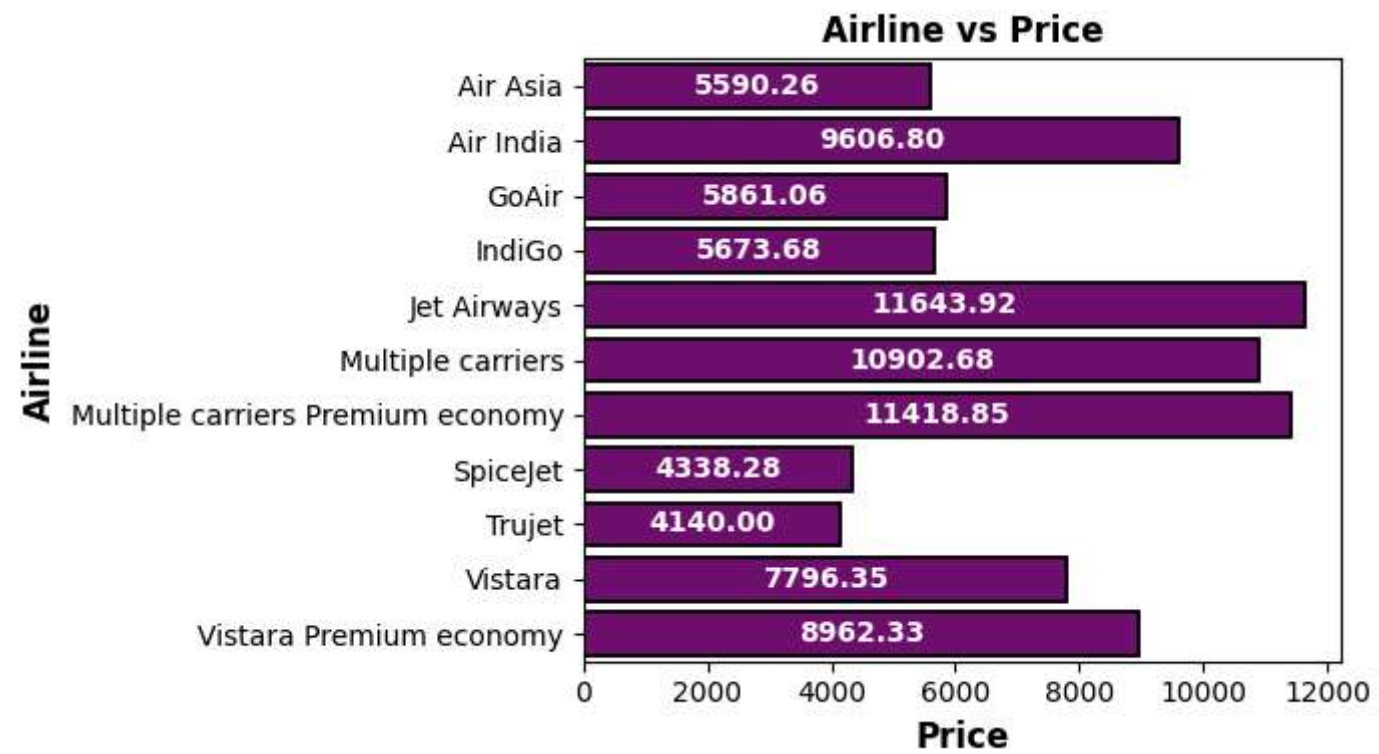
```
In [48]: plt.figure(figsize=(7,4))

ax= sns.barplot(x = 'Price', y = 'Airline', data = filtered_avg_price, color='purple', linewidth = 1.3, edgecolor='black', errorbar=None)

# For values inside the bars
for x in ax.patches:
    width = x.get_width()          # Gives actual width of the bars (Price)
    ax.text(width / 2,            # x - coordinate (center of bar)
            x.get_y() + x.get_height() / 2,    # y - coordinate (center of the bar)
            f'{width:.2f}',        # Number in 2 decimal places
            ha = 'center',        # Horizontal allignment
            va = 'center',        # Vertical allignment
            color = 'white',
            fontweight = 'bold')

plt.xlabel("Price", fontsize=12, fontweight='bold')
plt.ylabel("Airline", fontsize=12, fontweight='bold')
plt.title("Airline vs Price", fontsize=12, fontweight='bold')

plt.tight_layout()
plt.show()
```



**Observation** - Jet Airways has the highest price whereas Trujet has the lowest price.

### 3) Are certain routes (Source to Destination) more expensive than others?

```
In [49]: route_price = df.groupby(['Source', 'Destination'])['Price'].mean().reset_index()
route_price['Price'] = route_price['Price'].round(2)
display(route_price)
```

	Source	Destination	Price
0	Banglore	New Delhi	8017.46
1	Chennai	Kolkata	4789.89
2	Kolkata	Banglore	9158.39
3	Mumbai	Hyderabad	5042.08
4	New Delhi	Cochin	10539.44

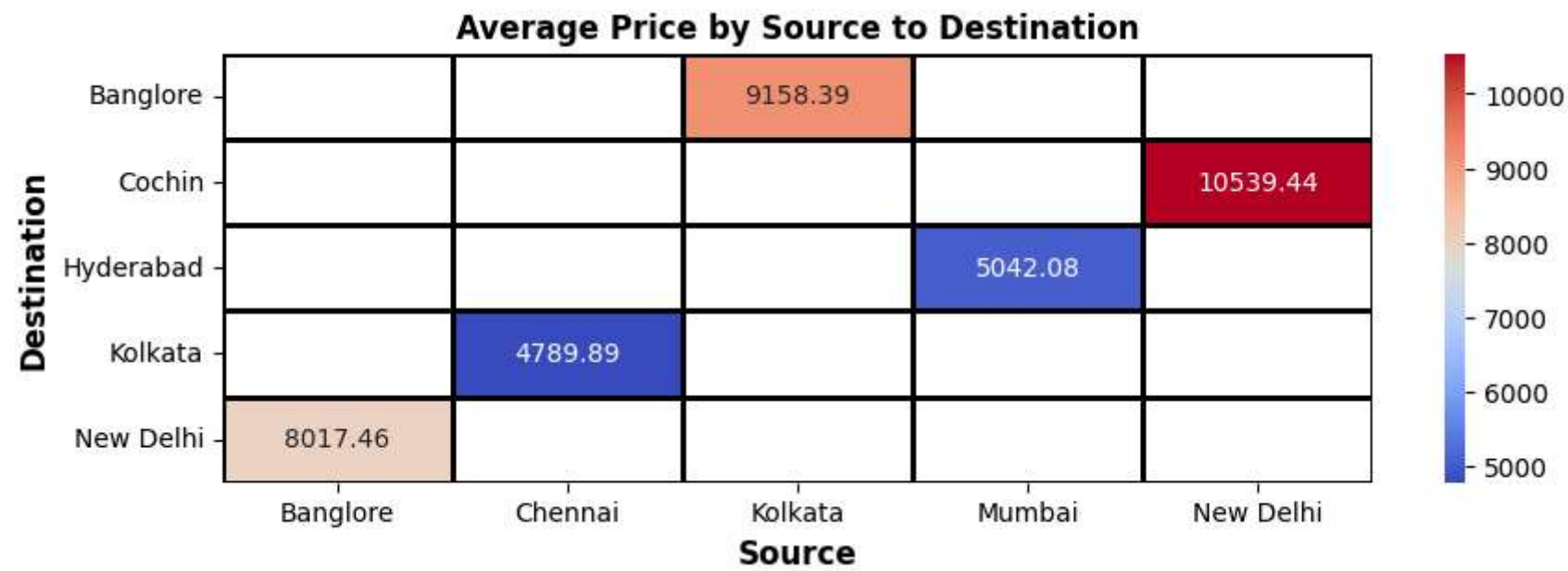
```
In [50]: route_pivot = route_price.pivot_table(index='Destination', columns='Source', values='Price', aggfunc='mean')
```

```
In [57]: plt.figure(figsize=(10,3))

ax = sns.heatmap(route_pivot, annot=True, cmap='coolwarm', fmt='.2f', linewidths=1, linecolor='black')

ax.set_title('Average Price by Source to Destination', fontweight='bold')
ax.set_xlabel('Source', fontsize=12, fontweight='bold')
ax.set_ylabel('Destination', fontsize=12, fontweight='bold')

plt.show()
```



**Observation** - New Delhi to Cochin is the most expensive route whereas Chennai to Kolkata is the cheapest.

#### 4) Is there a relationship between flight duration and price?

```
In [52]: total_duration = df['Duration_hr'] * 60 + df['Duration_min']
duration_price_avg = df.groupby(total_duration)['Price'].mean().reset_index()
display(duration_price_avg.head(2))
```

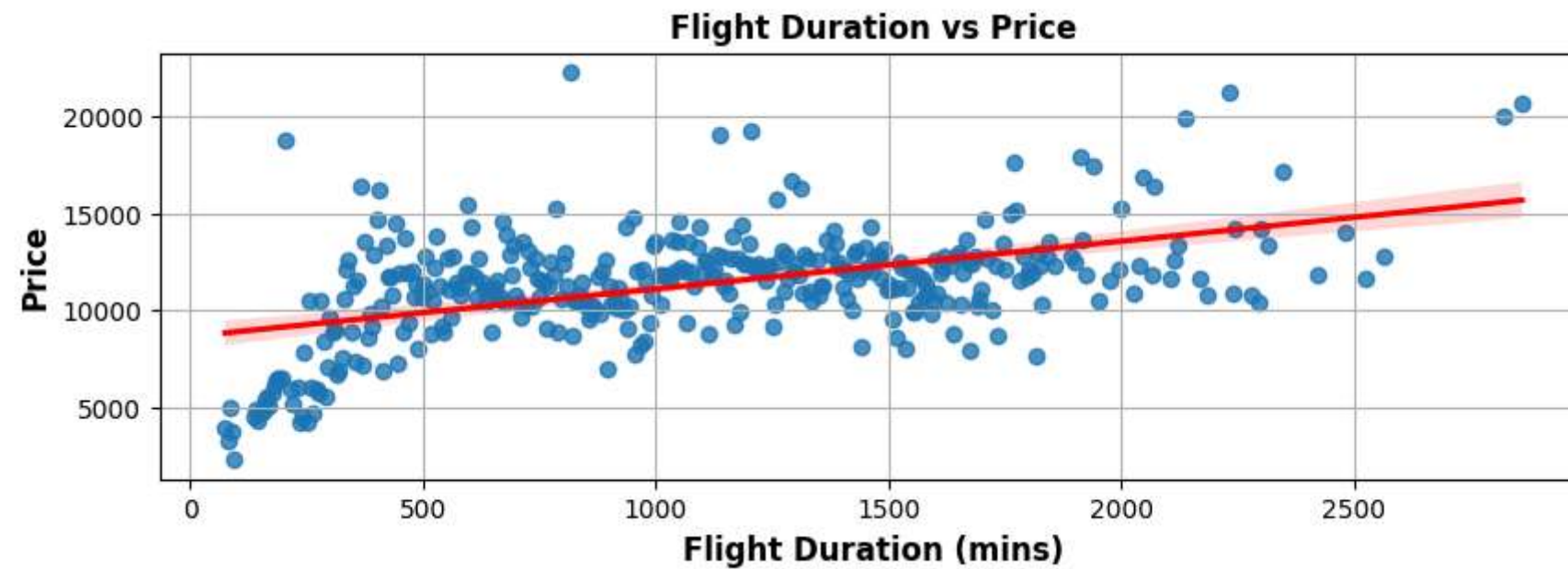
	index	Price
0	75	3944.333333
1	80	3286.377049

```
In [60]: plt.figure(figsize=(10,3))

sns.regplot(x = 'index', y = 'Price', data = duration_price_avg, line_kws={'color': 'red'})

plt.xlabel("Flight Duration (mins)", fontsize=12, fontweight='bold')
plt.ylabel("Price", fontsize=12, fontweight='bold')
plt.title("Flight Duration vs Price", fontsize=12, fontweight='bold')
plt.grid(True)

plt.show()
```



**Observation** - The red Trendline shows an upward trend meaning that as the flight duration increases, the price of flight increases.

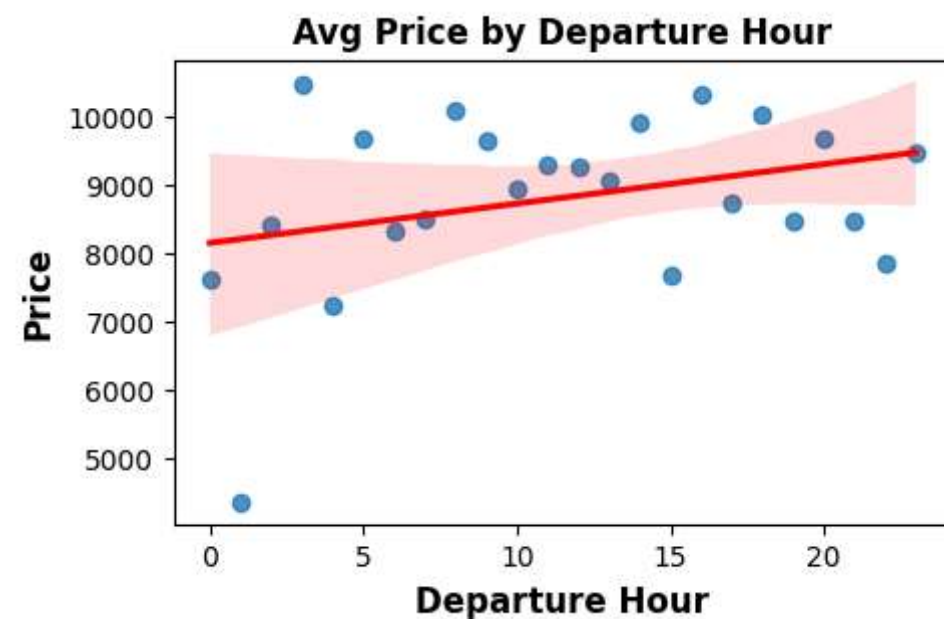
### 5) Does the time of day (e.g., early morning, afternoon, evening) affect the price of the flight?

```
In [86]: avg_price_by_dep = df.groupby('Dep_Hour')['Price'].mean().reset_index()
```

```
In [110]: plt.figure(figsize=(5,3))

sns.regplot(x='Dep_Hour', y='Price', data=avg_price_by_dep, line_kws={'color': 'red'})
plt.title('Avg Price by Departure Hour', fontsize = 12, fontweight = 'bold')
plt.xlabel('Departure Hour', fontsize = 12, fontweight = 'bold')
plt.ylabel('Price', fontsize = 12, fontweight = 'bold')

plt.show()
```



**Observation** - The red Trendline suggests that, on average, flight prices tend to increase slightly as the departure hour progresses through the day. However, the pink shaded area around the red line represents a confidence interval, which gives an estimate of the uncertainty in the trend line. A wider shaded region indicates more variability or less confidence in the trend at certain points.

### 6) How does the average price vary over the months of the year?

```
In [120... avg_price_by_date = df.groupby(['Month'])['Price'].mean().reset_index()
avg_price_by_date
```

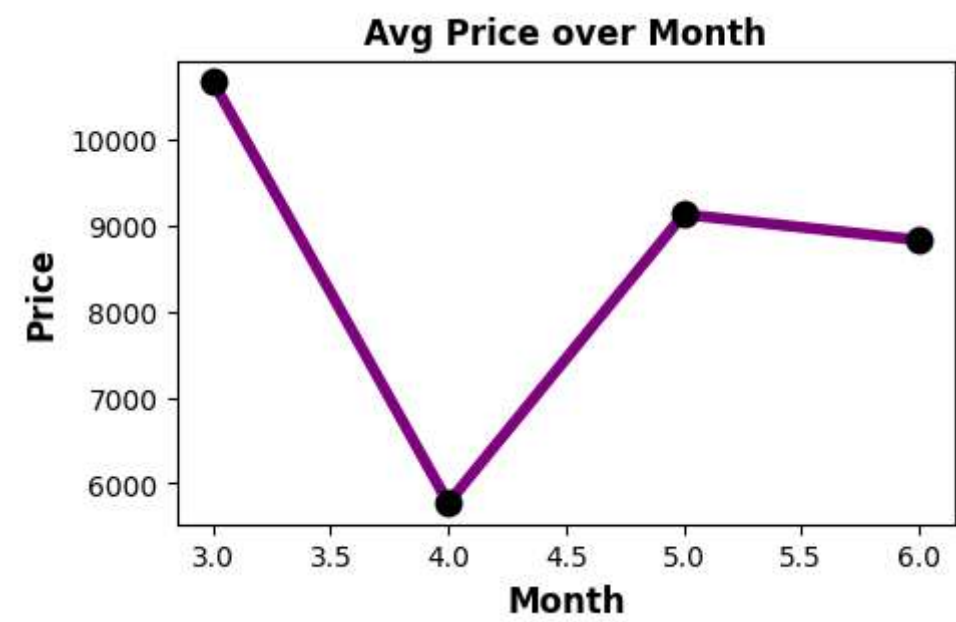
Out[120...

	Month	Price
0	3	10670.762027
1	4	5770.847081
2	5	9127.247548
3	6	8828.796134

```
In [122... plt.figure(figsize=(5,3))

plt.plot('Month', 'Price', data = avg_price_by_date, color='purple', linestyle='-', linewidth=4, marker='o', markerfacecolor='black', markeredgecolor='black', markersize=9)
plt.title('Avg Price over Month', fontsize = 12, fontweight = 'bold')
plt.xlabel('Month', fontsize = 12, fontweight = 'bold')
plt.ylabel('Price', fontsize = 12, fontweight = 'bold')

plt.show()
```



**Observation** - The month of March has most expensive fares, whereas April is the cheapest.

```
In [126... df.head()
```

Out[126...

	Airline	Source	Destination	Total_Stops	Additional_Info	Price	Date	Month	Year	Arrival_Hour	Arrival_Minute	Dep_Hour	Dep_Min	Duration_hr	Duration_min
0	IndiGo	Banglore	New Delhi	0	No info	3897	24	3	2019	1	10	22	20	2	50
1	Air India	Kolkata	Banglore	2	No info	7662	1	5	2019	13	15	5	50	7	25
2	Jet Airways	New Delhi	Cochin	2	No info	13882	9	6	2019	4	25	9	25	19	0
3	IndiGo	Kolkata	Banglore	1	No info	6218	12	5	2019	23	30	18	5	5	25
4	IndiGo	Banglore	New Delhi	1	No info	13302	1	3	2019	21	35	16	50	4	45

7) Find the busiest airport

```
In [136... total_dep = df.groupby('Source').size().reset_index(name='Total_Departures') # Best for counting rows directly
total_dep
```

Out[136...

	Source	Total_Departures
0	Banglore	2197
1	Chennai	381
2	Kolkata	2871
3	Mumbai	696
4	New Delhi	4537

```
In [158... total_arrivals = df.groupby('Destination').size().reset_index(name='Total_Arrivals')
total_arrivals
```

Out[158...

	Destination	Total_Arrivals
0	Banglore	2871
1	Cochin	4537
2	Hyderabad	696
3	Kolkata	381
4	New Delhi	2197

```
In [184... plt.figure(figsize=(13,4))

plt.subplot(1,2,1)

plt.suptitle('BUSIEST AIRPORTS', fontsize=14, color = 'purple' ,fontweight='bold', y=0.98)

bx = plt.bar('Source', 'Total_Departures', data = total_dep, color = 'purple', linewidth = 1.3, edgecolor='black')

for x in bx.patches:
    height = x.get_height()
    plt.text(
        x.get_x() + x.get_width() / 2,
        height / 2,
        f'{height:.0f}',
        ha = 'center',
        va = 'center',
        color = 'white',
        fontweight = 'bold')

plt.title('Most Departures', fontsize = 12, fontweight = 'bold')
plt.xlabel('Airport', fontsize = 10, fontweight = 'bold')
plt.ylabel('Total_Departures', fontsize = 10, fontweight = 'bold')

plt.subplot(1,2,2)

cx = plt.bar('Destination', 'Total_Arrivals', data = total_arrivals, color = 'purple', linewidth = 1.3, edgecolor='black')
```



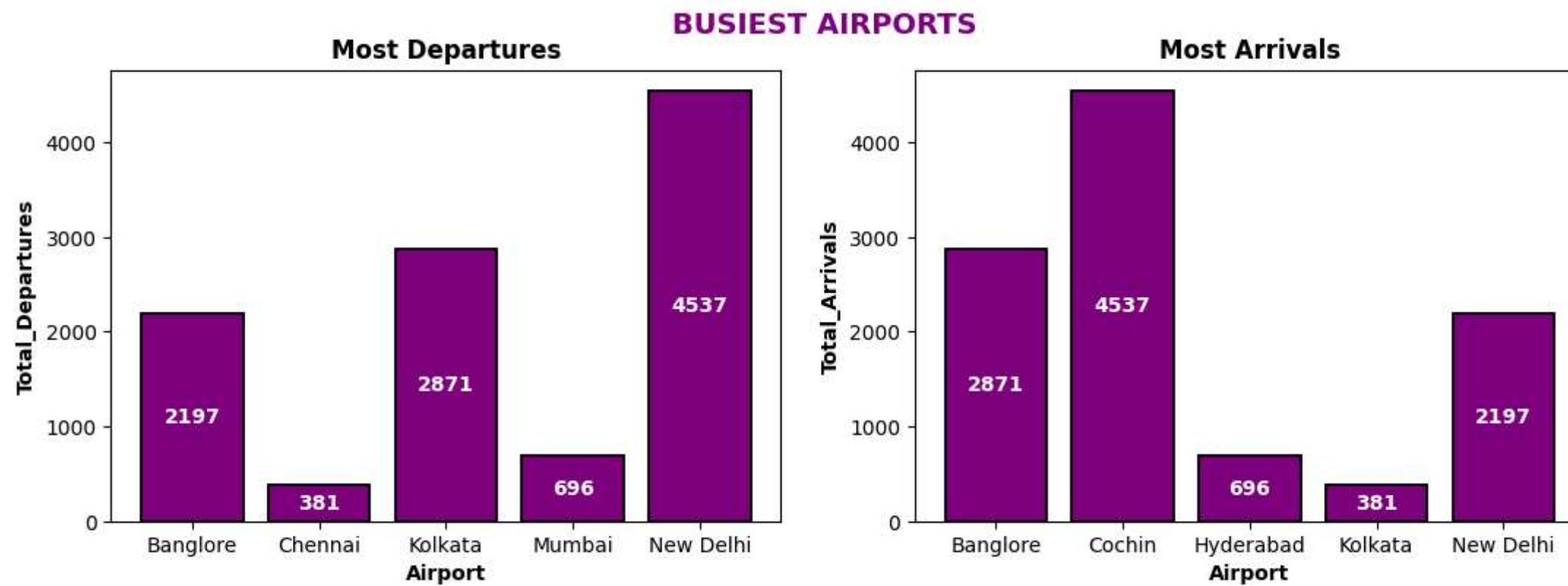
```

for x in cx.patches:
    height = x.get_height()
    plt.text(
        x.get_x() + x.get_width() / 2,
        height / 2,
        f'{height:.0f}',
        ha = 'center',
        va = 'center',
        color = 'white',
        fontweight = 'bold')

plt.title('Most Arrivals', fontsize = 12, fontweight = 'bold')
plt.xlabel('Airport', fontsize = 10, fontweight = 'bold')
plt.ylabel('Total_Arrivals', fontsize = 10, fontweight = 'bold')

plt.show()

```



**Observation** - New Delhi and Cochin were the busiest airports.