Name: Senthil Kumar S
Reg No: 212221230091

In [15]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv("FlightInformation.csv")
df
```

Out[2]:

	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
...	...	...	...	...	...	...	...	...	...	...	...
10678	Air Asia	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h 30m	non-stop	No info	4107
10679	Air India	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h 35m	non-stop	No info	4145
10680	Jet Airways	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	3h	non-stop	No info	7229
10681	Vistara	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h 40m	non-stop	No info	12648
10682	Air India	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h 20m	2 stops	No info	11753

10683 rows × 11 columns

In [3]:
```
df.head()
```

Out[3]:

	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

In [4]:
```
df.info()
```

```
<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)
memory usage: 918.2+ KB
```

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
Airline           0
Date_of_Journey   0
Source            0
Destination       0
Route             1
Dep_Time          0
Arrival_Time      0
Duration          0
Total_Stops       1
Additional_Info   0
Price             0
dtype: int64
```

In [7]:
```
df['Route'] = df['Route'].fillna(df['Route'].mode()[0])
df['Total_Stops'] = df['Total_Stops'].fillna(df['Total_Stops'].mode()[0])
df.isnull().sum()
```

Out[7]:
```
Airline           0
Date_of_Journey   0
Source            0
Destination       0
Route             0
Dep_Time          0
Arrival_Time      0
Duration          0
Total_Stops       0
Additional_Info   0
Price             0
dtype: int64
```

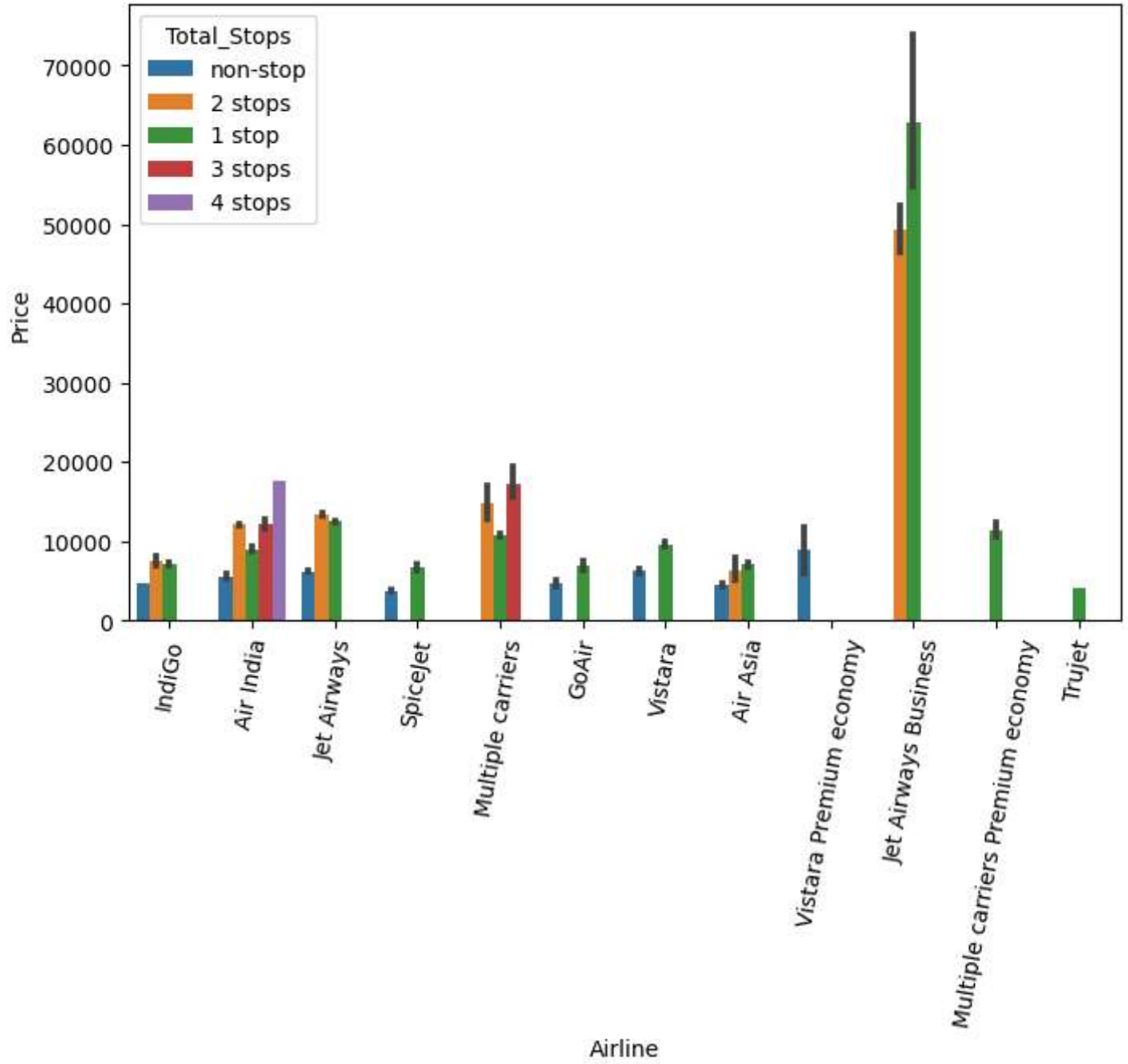
In [8]:
```
df.shape
```

Out[8]:
```
(10683, 11)
```

In [23]:
```
plt.figure(figsize=(8,5))
plt.xticks(rotation = 80)
sns.barplot(df["Airline"],df["Price"],hue=df["Total_Stops"])
```

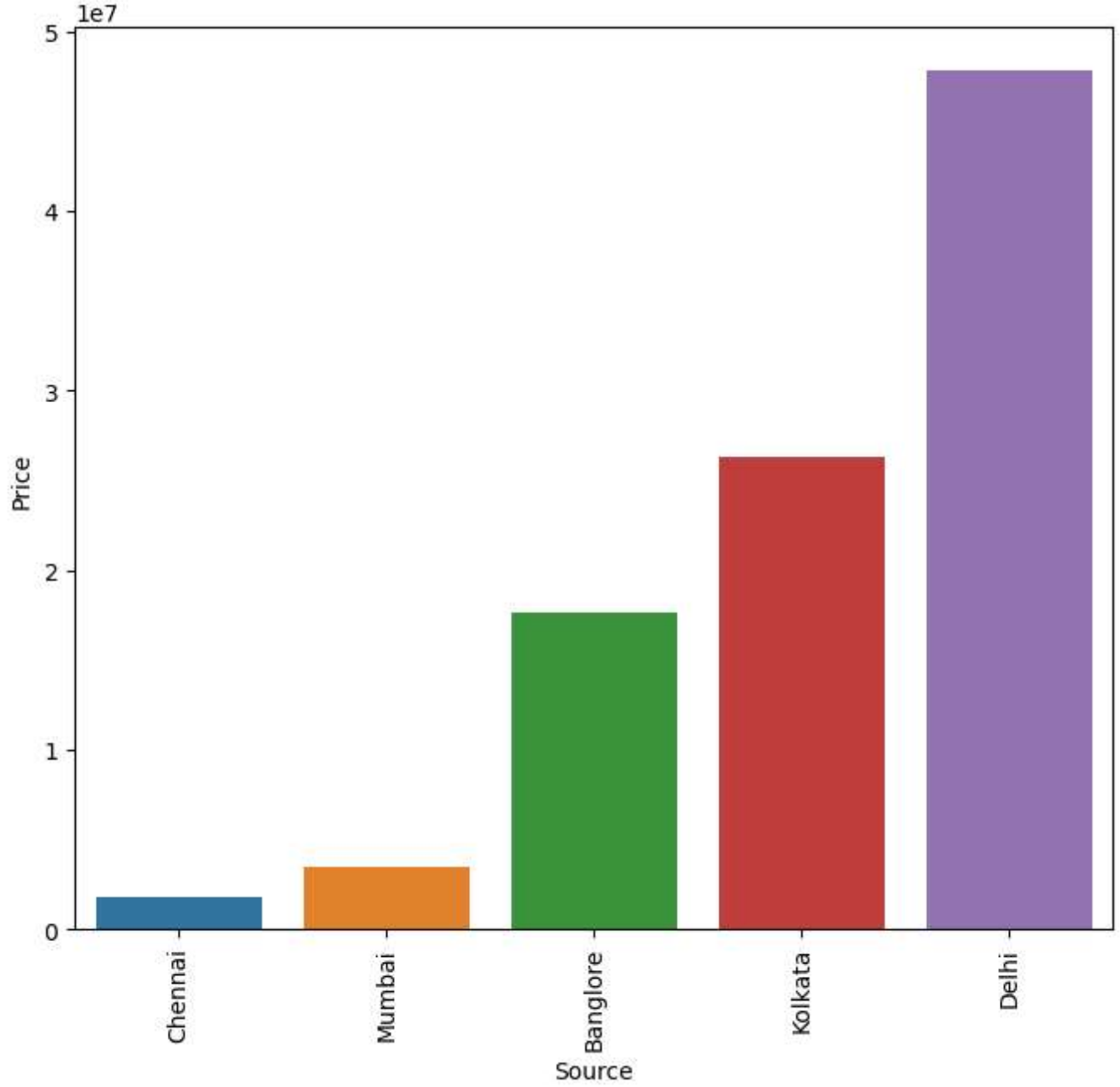
```
C:\Users\shafe\anaconda3\envs\AiMath\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[23]:
```
<AxesSubplot:xlabel='Airline', ylabel='Price'>
```



Costliest Ticket is from Jet Airways and is a 1 stop airplane

In [58]:
```
states=df.loc[:,["Source","Price"]]
states=states.groupby(by=["Source"]).sum().sort_values(by="Price")
plt.figure(figsize=(8,7))
sns.barplot(x=states.index,y="Price",data=states)
plt.xticks(rotation = 90)
plt.xlabel=("SOURCE")
plt.ylabel=("PRICE")
plt.show()
```

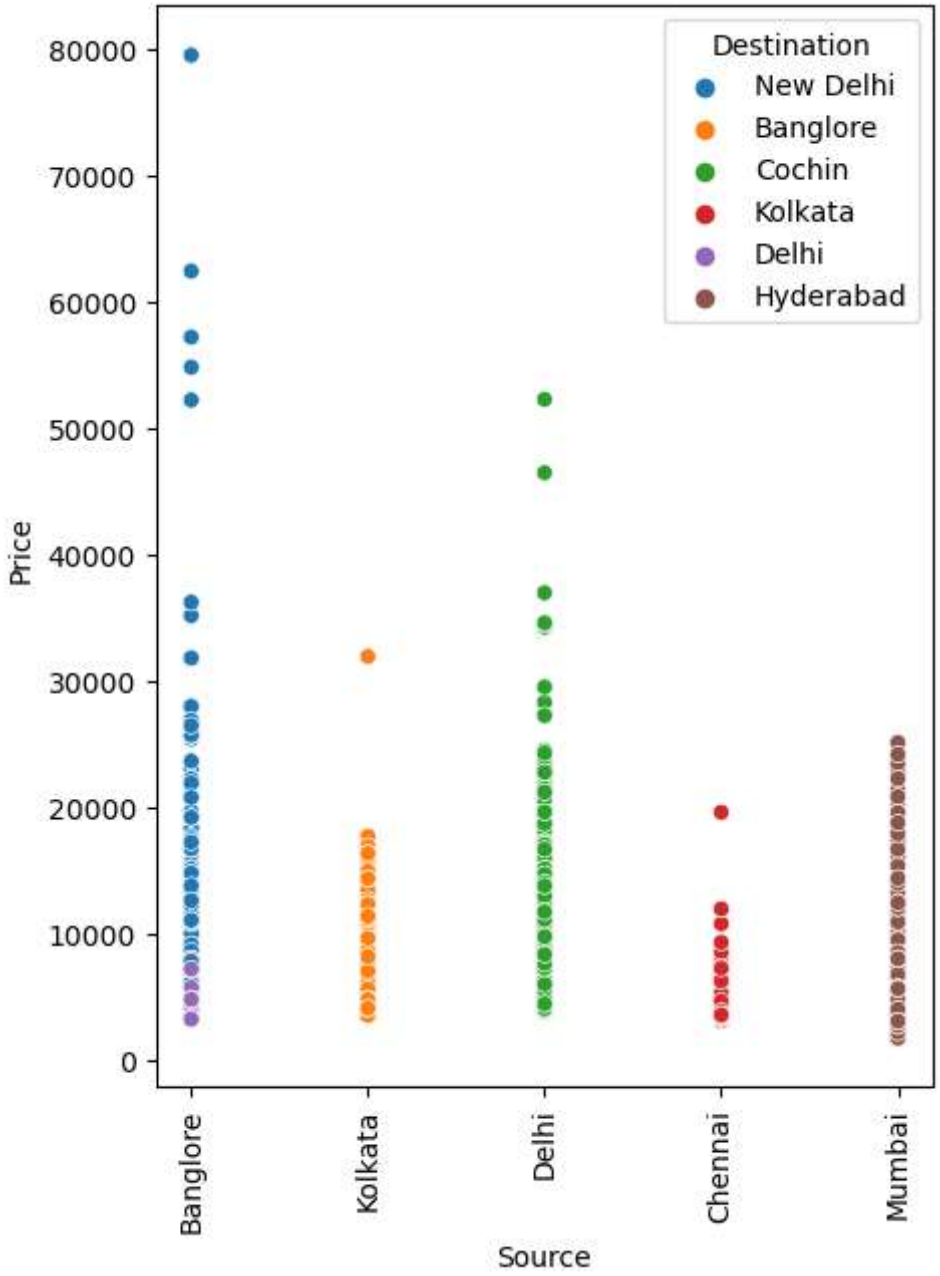


The Costliest flights depart from Delhi

In [33]:
```
plt.figure(figsize=(5,7))
sns.scatterplot(df['Source'], df['Price'], hue=df['Destination'])
plt.xticks(rotation = 90)
```

```
C:\Users\shafe\anaconda3\envs\AiMath\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

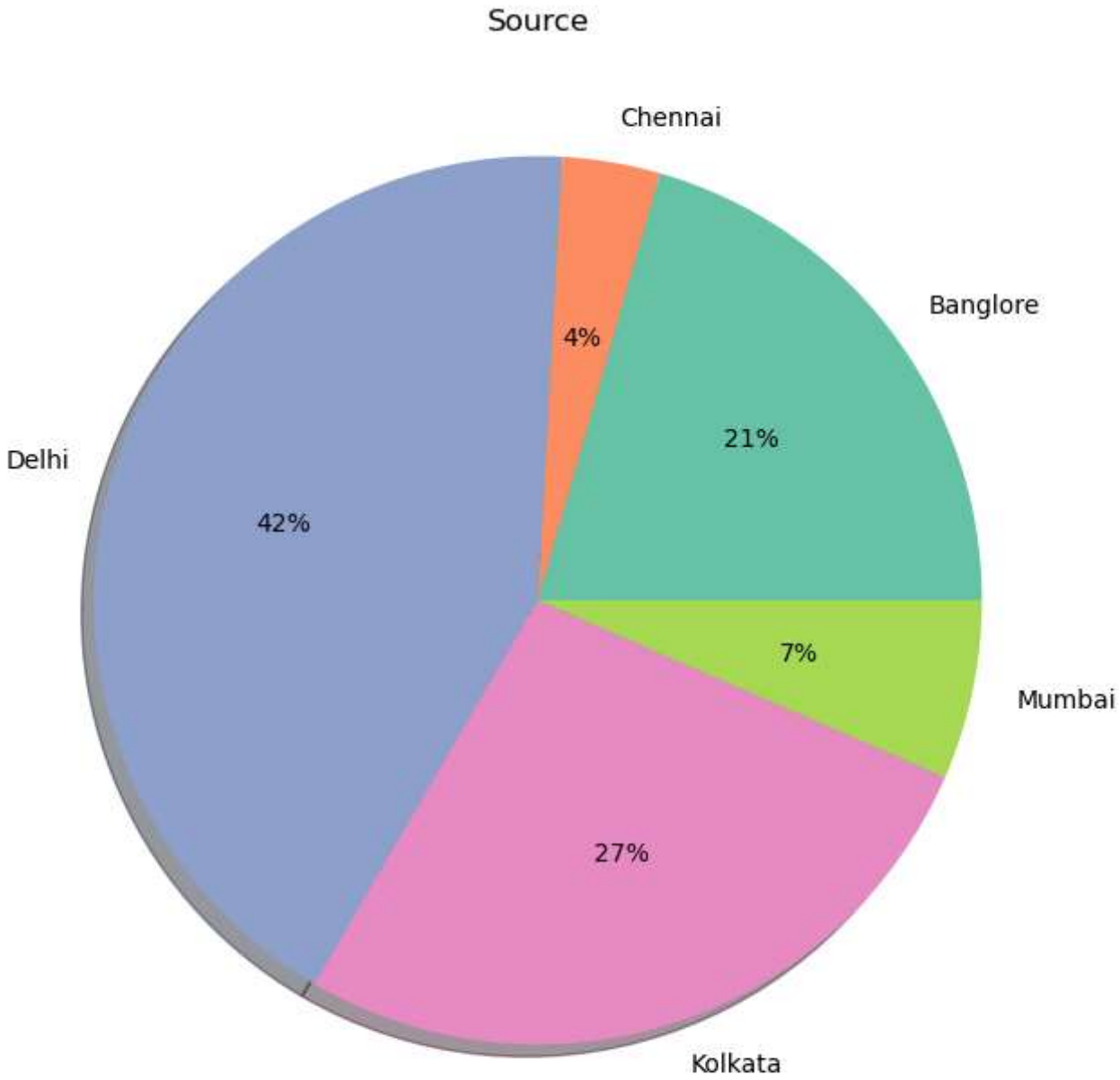
Out[33]:
```
([0, 1, 2, 3, 4],
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```



Most of flights departing from Chennai arrives at Kolkata

In [55]:

```
df_count = df.groupby(by=["Source"]).count()
labels=[]
for i in df_count.index:
    labels.append(i)
plt.figure(figsize=(8,8))
colors = sns.color_palette("Set2")
plt.pie(df_count["Price"], colors = colors, labels=labels, autopct = "%0.0f%%",shadow = True)
plt.title("Source")
plt.show()
```



In [56]:
```
df['Airline'].value_counts()
```

Out[56]:
```
Jet Airways                          3849
IndiGo                               2053
Air India                            1752
Multiple carriers                    1196
SpiceJet                              818
Vistara                               479
Air Asia                              319
GoAir                                 194
Multiple carriers Premium economy      13
Jet Airways Business                    6
Vistara Premium economy                 3
Trujet                                  1
Name: Airline, dtype: int64
```

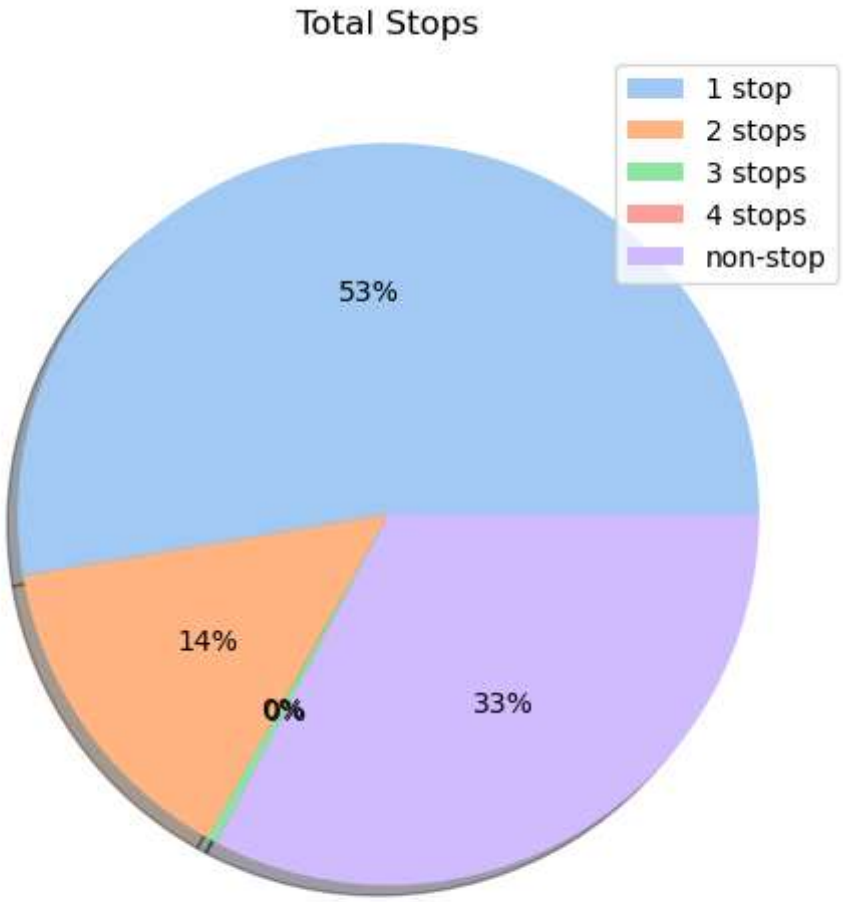
Jet Airways has the most number of flights

In [40]:
```
df['Duration'].value_counts()
```

Out[40]:
```
2h 50m     550
1h 30m     386
2h 45m     337
2h 55m     337
2h 35m     329
          ...
31h 30m      1
30h 25m      1
42h 5m       1
4h 10m       1
47h 40m      1
Name: Duration, Length: 368, dtype: int64
```

In [54]:
```
df_segment = df.groupby(by=["Total_Stops"]).count()
labels = []
for i in df_segment.index:
    labels.append(i)

plt.figure(figsize=(6,6))
colors = sns.color_palette('pastel')
pie = plt.pie(df_segment["Price"], colors = colors, autopct = "%0.0f%%",shadow = True)
plt.title("Total Stops")
plt.legend(pie[0], labels, loc="upper right")
plt.show()
```



Most of flights are 1 stops

In [52]:
```
df_region = df.groupby(by=["Destination"]).count()
labels = []
for i in df_region.index:
    labels.append(i)

plt.figure(figsize=(8,8))
colors = sns.color_palette('pastel')
pie = plt.pie(df_region["Price"], colors = colors, autopct = "%0.0f%%",shadow = True)
plt.title("Destination")
```

```
plt.legend(pie[0], labels, loc="upper right")
plt.show()
```