Problem statement:- To find Which Model is suitable for Flight Price Prediction

importing packages

In [53]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

To read train DataSet

In [54]:

```
traindf=pd.read_csv(r"C:\Users\lenovo\Downloads\Data_Train.csv")
traindf
```

Out[54]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Dura
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h
...	...	...	...	...	...	...	...	
10678	Air Asia	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h
10679	Air India	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h
10680	Jet Airways	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	
10681	Vistara	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h
10682	Air India	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h

10683 rows × 11 columns

In [55]:

```
testdf=pd.read_csv(r"C:\Users\lenovo\Downloads\Data_Train.csv")
testdf
```

Out[55]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Dura
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h
...	...	...	...	...	...	...	...	
10678	Air Asia	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h
10679	Air India	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h
10680	Jet Airways	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	
10681	Vistara	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h
10682	Air India	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h

10683 rows × 11 columns

Data Collection and Preprocessing

In [56]:

```
traindf.head()
```

Out[56]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m

In [57]:

```
testdf.head()
```

Out[57]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m

In [58]:

```
traindf.tail()
```

Out[58]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Dura
10678	Air Asia	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h
10679	Air India	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h
10680	Jet Airways	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	
10681	Vistara	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h
10682	Air India	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h

In [59]:

```
testdf.tail()
```

Out[59]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Dura
10678	Air Asia	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h
10679	Air India	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h
10680	Jet Airways	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	
10681	Vistara	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h
10682	Air India	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h

In [60]:

```
traindf.describe()
```

Out[60]:

	Price
count	10683.000000
mean	9087.064121
std	4611.359167
min	1759.000000
25%	5277.000000
50%	8372.000000
75%	12373.000000
max	79512.000000

In [61]:

```
testdf.describe()
```

Out[61]:

	Price
count	10683.000000
mean	9087.064121
std	4611.359167
min	1759.000000
25%	5277.000000
50%	8372.000000
75%	12373.000000
max	79512.000000

In [62]:

```
traindf.shape
```

Out[62]:

```
(10683, 11)
```

In [63]:

```
testdf.shape
```

Out[63]:

```
(10683, 11)
```

In [64]:

```
traindf.columns
```

Out[64]:

```
Index(['Airline', 'Date_of_Journey', 'Source', 'Destination', 'Route',
       'Dep_Time', 'Arrival_Time', 'Duration', 'Total_Stops',
       'Additional_Info', 'Price'],
      dtype='object')
```

In [65]:

```
testdf.columns
```

Out[65]:

```
Index(['Airline', 'Date_of_Journey', 'Source', 'Destination', 'Route',
       'Dep_Time', 'Arrival_Time', 'Duration', 'Total_Stops',
       'Additional_Info', 'Price'],
      dtype='object')
```

In [66]:

```
traindf.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 [67]:

```
testdf.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
```

To Find out any null or Duplicate values in DataSet

In [68]:

```
traindf.isnull().sum()
```

Out[68]:

```
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 [69]:

```
testdf.isnull().sum()
```

Out[69]:

```
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
```

To Remove Null values in DataSet

In [70]:

```
traindf.dropna(inplace=True)
```

In [71]:

```
traindf.isnull().sum()
```

Out[71]:

```
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 [72]:

```
traindf.shape
```

Out[72]:

```
(10682, 11)
```

Replacing the String values to Numerical values in given DataSet

In [73]:

```
traindf['Airline'].value_counts()
```

Out[73]:

```
Airline
Jet Airways                          3849
IndiGo                               2053
Air India                            1751
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: count, dtype: int64
```

In [74]:

```
traindf['Source'].value_counts()
```

Out[74]:

```
Source
Delhi       4536
Kolkata     2871
Banglore    2197
Mumbai       697
Chennai      381
Name: count, dtype: int64
```

In [75]:

```
traindf['Destination'].value_counts()
```

Out[75]:

```
Destination
Cochin       4536
Banglore     2871
Delhi        1265
New Delhi     932
Hyderabad     697
Kolkata       381
Name: count, dtype: int64
```

In [76]:

```
traindf['Total_Stops'].value_counts()
```

Out[76]:

```
Total_Stops
1 stop      5625
non-stop    3491
2 stops     1520
3 stops       45
4 stops        1
Name: count, dtype: int64
```

In [77]:

```
airline={"Airline":{"Jet Airways":0,"IndiGo":1,"Air India":2,"Multiple carriers":3,
 "SpiceJet":4,"Vistara":5,"Air Asia":6,"GoAir":7,
 "Multiple carriers Premium economy":8,
"Jet Airways Business":9,"Vistara Premium economy":10,"Trujet":11}}
traindf=traindf.replace(airline)
traindf
```

Out[77]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Durat
0	1	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 5
1	2	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 2
2	0	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	
3	1	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 2
4	1	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 4
...	...	...	...	...	...	...	...	
10678	6	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h 3
10679	2	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h 3
10680	0	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	
10681	5	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h 4
10682	2	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h 2

10682 rows × 11 columns

In [78]:

```
city={"Source":{"Delhi":0,"Kolkata":1,"Banglore":2,
 "Mumbai":3,"Chennai":4}}
traindf=traindf.replace(city)
traindf
```

Out[78]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Durati
0	1	24/03/2019	2	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 5C
1	2	1/05/2019	1	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25
2	0	9/06/2019	0	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	1'
3	1	12/05/2019	1	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25
4	1	01/03/2019	2	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45
...	...	...	...	...	...	...	...	
10678	6	9/04/2019	1	Banglore	CCU → BLR	19:55	22:25	2h 3C
10679	2	27/04/2019	1	Banglore	CCU → BLR	20:45	23:20	2h 35
10680	0	27/04/2019	2	Delhi	BLR → DEL	08:20	11:20	
10681	5	01/03/2019	2	New Delhi	BLR → DEL	11:30	14:10	2h 4C
10682	2	9/05/2019	0	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h 2C

10682 rows × 11 columns

In [79]:

```
dest={"Destination":{"Cochin":0,"Banglore":1,"Delhi":2,
 "New Delhi":3,"Hyderabad":4,"Kolkata":5}}
traindf=traindf.replace(dest)
traindf
```

Out[79]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Durati
0	1	24/03/2019	2	3	BLR → DEL	22:20	01:10 22 Mar	2h 50
1	2	1/05/2019	1	1	CCU → IXR → BBI → BLR	05:50	13:15	7h 25
2	0	9/06/2019	0	0	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	1
3	1	12/05/2019	1	1	CCU → NAG → BLR	18:05	23:30	5h 25
4	1	01/03/2019	2	3	BLR → NAG → DEL	16:50	21:35	4h 45
...	...	...	...	...	...	...	...	
10678	6	9/04/2019	1	1	CCU → BLR	19:55	22:25	2h 30
10679	2	27/04/2019	1	1	CCU → BLR	20:45	23:20	2h 35
10680	0	27/04/2019	2	2	BLR → DEL	08:20	11:20	
10681	5	01/03/2019	2	3	BLR → DEL	11:30	14:10	2h 40
10682	2	9/05/2019	0	0	DEL → GOI → BOM → COK	10:55	19:15	8h 20

10682 rows × 11 columns

In [80]:

```
stops={"Total_Stops":{"non-stop":0,"1 stop":1,"2 stops":2,
 "3 stops":3,"4 stops":4}}
traindf=traindf.replace(stops)
traindf
```

Out[80]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Durati
0	1	24/03/2019	2	3	BLR → DEL	22:20	01:10 22 Mar	2h 50
1	2	1/05/2019	1	1	CCU → IXR → BBI → BLR	05:50	13:15	7h 25
2	0	9/06/2019	0	0	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	1
3	1	12/05/2019	1	1	CCU → NAG → BLR	18:05	23:30	5h 25
4	1	01/03/2019	2	3	BLR → NAG → DEL	16:50	21:35	4h 45
...	...	...	...	...	...	...	...	
10678	6	9/04/2019	1	1	CCU → BLR	19:55	22:25	2h 30
10679	2	27/04/2019	1	1	CCU → BLR	20:45	23:20	2h 35
10680	0	27/04/2019	2	2	BLR → DEL	08:20	11:20	
10681	5	01/03/2019	2	3	BLR → DEL	11:30	14:10	2h 40
10682	2	9/05/2019	0	0	DEL → GOI → BOM → COK	10:55	19:15	8h 20

10682 rows × 11 columns

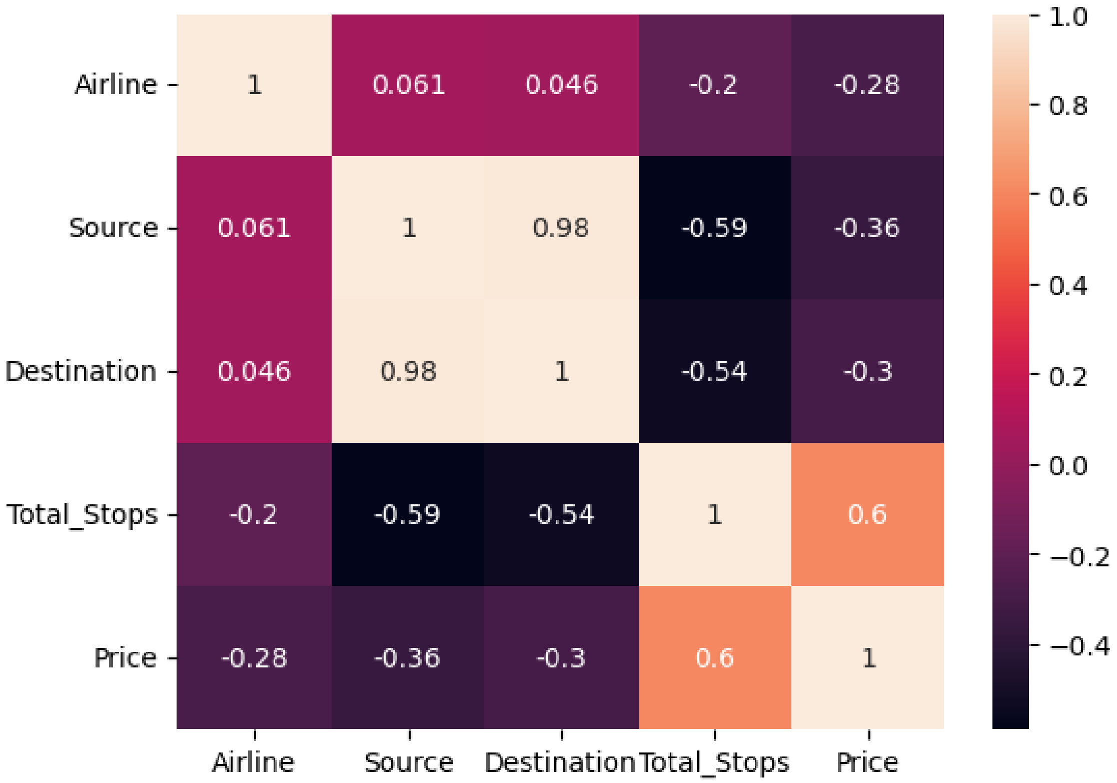
Data visualization:-

In [81]:

```
fdf=traindf[['Airline','Source','Destination','Total_Stops','Price']]
sns.heatmap(fdf.corr(),annot=True)
```

Out[81]:

```
<Axes: >
```



Feature Scaling :- To Split the data into train data and test data

In [82]:

```
x=fdf[['Airline','Source','Destination','Total_Stops']]
y=fdf['Price']
```

In [83]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=100)
```

Linear Regression

In [84]:

```
from sklearn.linear_model import LinearRegression
regr=LinearRegression()
regr.fit(X_train,y_train)
print(regr.intercept_)
coeff_df=pd.DataFrame(regr.coef_,x.columns,columns=['coefficient'])
coeff_df
```

```
7211.098088897486
```

Out[84]:

	coefficient
Airline	-418.483922
Source	-3275.073380
Destination	2505.480291
Total_Stops	3541.798053

In [85]:

```
score=regr.score(X_test,y_test)
print(score)
```

```
0.41083048909283504
```

In [86]:

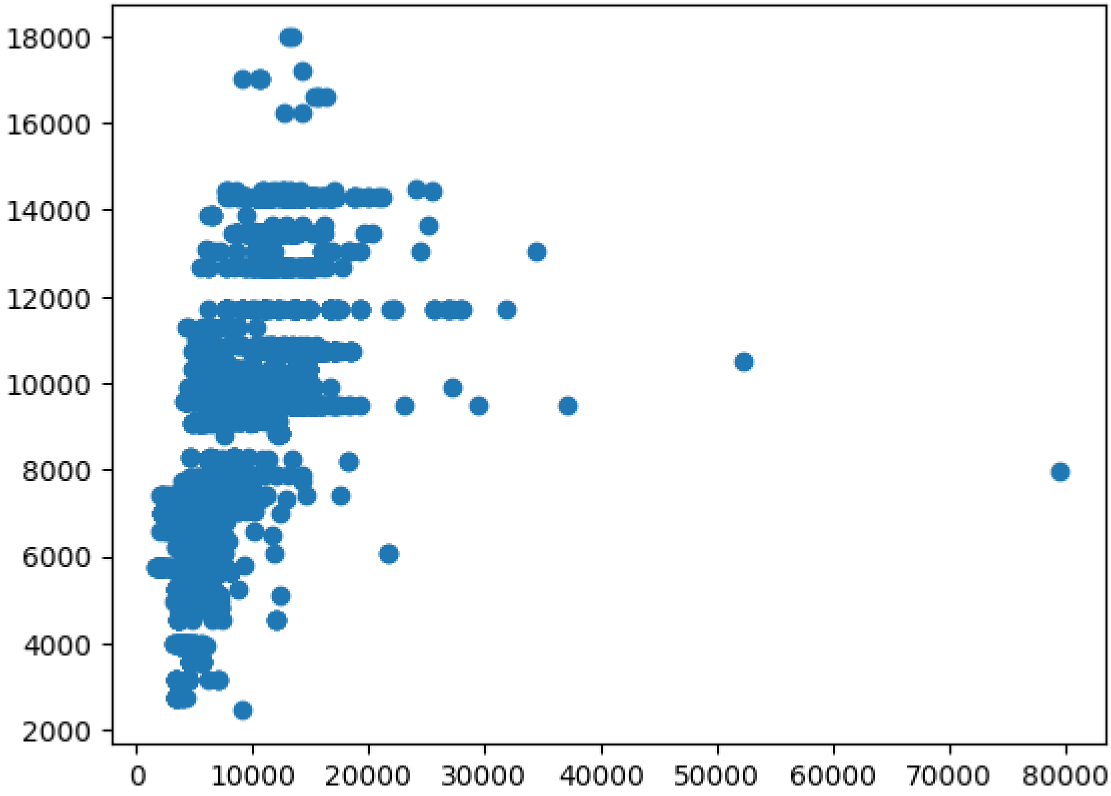
```
predictions=regr.predict(X_test)
```

In [87]:

```
plt.scatter(y_test,predictions)
```

Out[87]:

```
<matplotlib.collections.PathCollection at 0x2360f883750>
```



In [88]:

```
x=np.array(fdf['Price']).reshape(-1,1)
y=np.array(fdf['Total_Stops']).reshape(-1,1)
fdf.dropna(inplace=True)
```

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_29572\521034954.py:3: Setting
WithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  fdf.dropna(inplace=True)
```

In [89]:

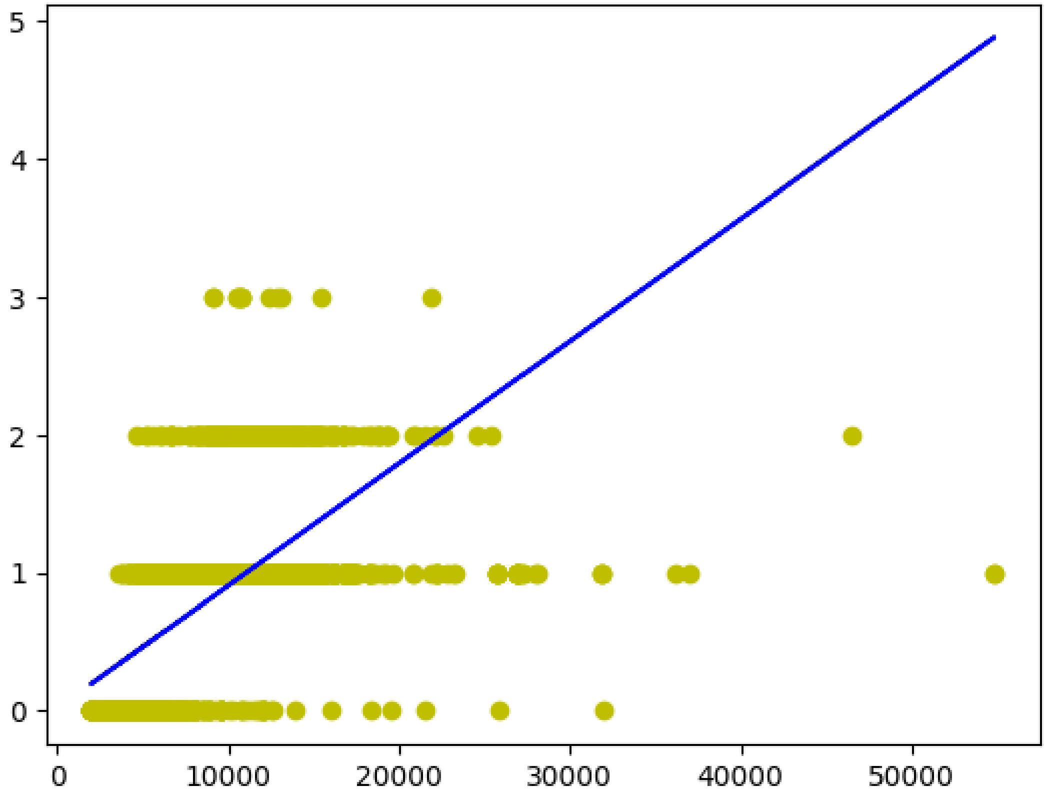
```
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
regr.fit(X_train,y_train)
regr.fit(X_train,y_train)
```

Out[89]:

```
▾ LinearRegression
LinearRegression()
```

In [90]:

```
y_pred=regr.predict(X_test)
plt.scatter(X_test,y_test,color='y')
plt.plot(X_test,y_pred,color='b')
plt.show()
```



Since in the above Linear regression we could not get accuracy so we can check for Logistic regression model.

Logistic Regression

In [91]:

```
x=np.array(fdf['Price']).reshape(-1,1)
y=np.array(fdf['Total_Stops']).reshape(-1,1)
fdf.dropna(inplace=True)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=1)
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression(max_iter=10000)
```

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_29572\497261869.py:3: Setting
WithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  fdf.dropna(inplace=True)
```

In [92]:

```
lr.fit(x_train,y_train)
```

```
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\utils\validation.py:1143: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[92]:

```
▾        LogisticRegression
LogisticRegression(max_iter=10000)
```

In [93]:

```
score=lr.score(x_test,y_test)
print(score)
```

```
0.7160686427457098
```

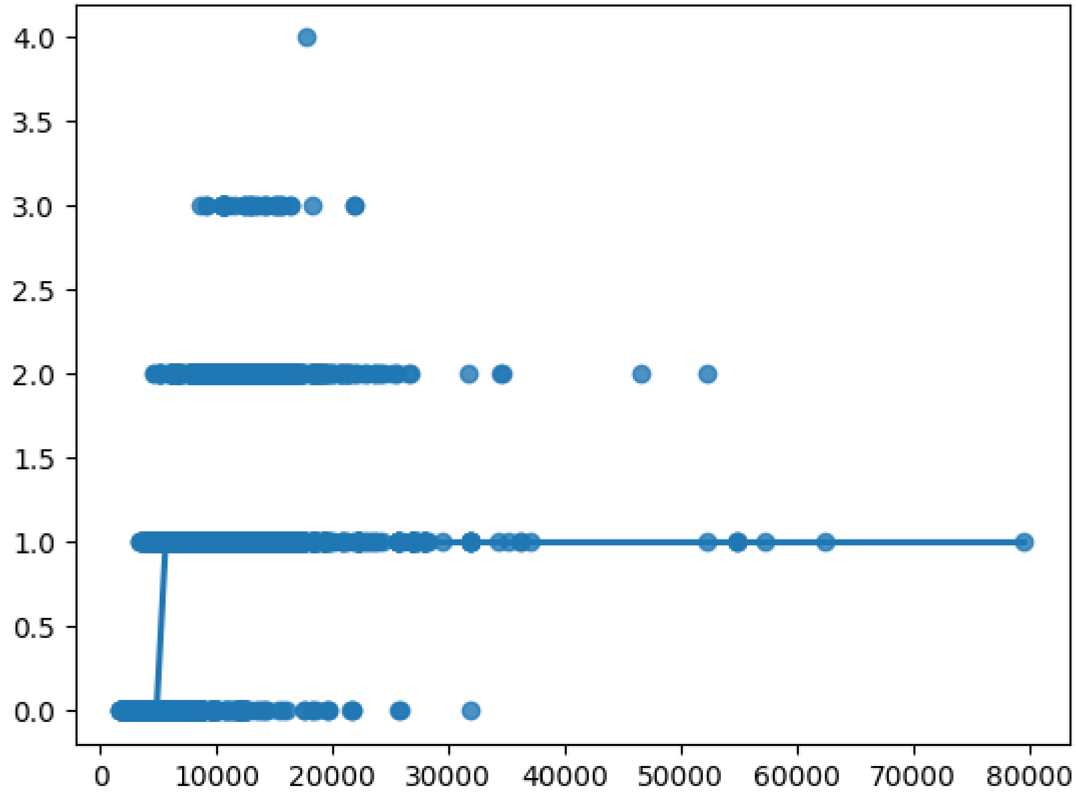
In [94]:

```
sns.regplot(x=x,y=y,data=fdf,logistic=True,ci=None)
```

```
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\statsmodels\genmod\families\links.py:198: RuntimeWarning: overflow encoun
tered in exp
  t = np.exp(-z)
```

Out[94]:

```
<Axes: >
```



In Logistic Regression model we could not get accuracy.we can use other models like Decision Tree and Random Forest to check the accuracy.

Decision Tree

In [95]:

```
from sklearn.tree import DecisionTreeClassifier
clf=DecisionTreeClassifier(random_state=0)
clf.fit(x_train,y_train)
```

Out[95]:

```
▾        DecisionTreeClassifier
DecisionTreeClassifier(random_state=0)
```

In [96]:

```
score=clf.score(x_test,y_test)
print(score)
```

```
0.9369734789391576
```

Random Forest

In [97]:

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(X_train,y_train)
```

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_29572\4104924521.py:3: DataCo
nversionWarning: A column-vector y was passed when a 1d array was expecte
d. Please change the shape of y to (n_samples,), for example using ravel
().
  rfc.fit(X_train,y_train)
```

Out[97]:

```
▾ RandomForestClassifier

RandomForestClassifier()
```

In [98]:

```
params={'max_depth':[2,3,5,10,20],
 'min_samples_leaf':[5,10,20,50,100,200],'n_estimators':[10,25,30,50,100,200]}
```

In [99]:

```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=params,cv=2,scoring="accuracy")
```

In [100]:

```
grid_search.fit(X_train,y_train)
```

```
the shape of y to (n_samples,), for example using ravel().
  estimator.fit(X_train, y_train, **fit_params)

C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packag
es\sklearn\model_selection\_validation.py:686: DataConversionWarning: A
column-vector y was passed when a 1d array was expected. Please change
the shape of y to (n_samples,), for example using ravel().
  estimator.fit(X_train, y_train, **fit_params)
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packag
es\sklearn\model_selection\_validation.py:686: DataConversionWarning: A
column-vector y was passed when a 1d array was expected. Please change
the shape of y to (n_samples,), for example using ravel().
  estimator.fit(X_train, y_train, **fit_params)
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packag
es\sklearn\model_selection\_validation.py:686: DataConversionWarning: A
column-vector y was passed when a 1d array was expected. Please change
the shape of y to (n_samples,), for example using ravel().
  estimator.fit(X_train, y_train, **fit_params)
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packag
es\sklearn\model_selection\_validation.py:686: DataConversionWarning: A
```

In [101]:

```
grid_search.best_score_
```

Out[101]:

```
0.5240069998086772
```

In [102]:

```
rf_best=grid_search.best_estimator_
rf_best
```

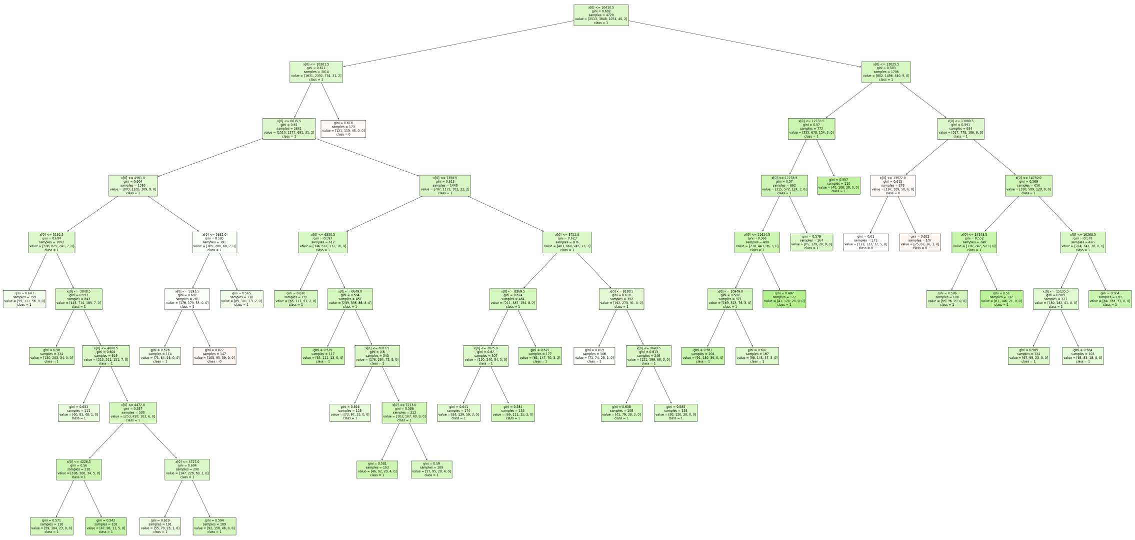
Out[102]:

```
▾ RandomForestClassifier

RandomForestClassifier(max_depth=20, min_samples_leaf=100, n_estimators=2
5)
```

In [103]:

```
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rf_best.estimators_[4],class_names=['0','1','2','3','4'],filled=True);
```



In [104]:

```
score=rfc.score(x_test,y_test)
print(score)
```

```
0.46770670826833307
```

CONCLUSION : Based on the accuracy of all models that are implemented above we can conclude that "Decision Tree is best model for given DataSet" ¶

In [ ]: