





In [133]: #Handling Categorical Values  
df[['Total\_Stops']]=df[['Total\_Stops']].map({'non-stop':0, '2 stops':2, '1 stop':1, '3 stops':3, '4 stops':4})  
air\_dummy=pd.get\_dummies(df[['Airline']],drop\_first=True)  
# print(df['Source'].value\_counts())  
# print(df['Destination'].value\_counts())  
source\_dest\_dummy=pd.get\_dummies(df[['Source','Destination']],drop\_first=True)  
df=pd.concat([air\_dummy,source\_dest\_dummy,df],axis=1)

In [134]: df.drop(['Airline','Source','Destination'],inplace=True,axis=1)

In [135]: df.head()

Out[135]:

	Air India	GoAir	IndiGo	Jet Airways	Jet Airways Business	Multiple carriers Premium economy	SpiceJet	Trujet	Vistara	...	Additional Info	Price	Day_of_Journey	Mor
0	0	0	1	0	0	0	0	0	0	...	No info	3887	24	
1	1	0	0	0	0	0	0	0	0	...	No info	7662	5	
2	0	0	0	1	0	0	0	0	0	...	No info	13862	6	
3	0	0	1	0	0	0	0	0	0	...	No info	6218	5	
4	0	0	1	0	0	0	0	0	0	...	No info	13302	3	
5 rows x 15 columns														

In [136]: df.shape

Out[136]: (10682, 15)

## Test Data

In [137]: df\_test=pd.read\_excel("Test\_set.xlsx")

In [138]: df\_test.head()

Out[138]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional Info
0	Jet Airways	6/06/2019	Delhi	Cochin	DEL --> BOM --> COK	17:30	04:25 07 Jun	10h 55m	1 stop	No info
1	IndiGo	12/05/2019	Kolkata	Banglore	CCU --> MAA --> BLR	06:20	10:20	4h	1 stop	No info
2	Jet Airways	21/05/2019	Delhi	Cochin	DEL --> BOM --> COK	19:15	19:00 23 May	23h 45m	1 stop	In-flight meal not included
3	Multiple carriers	21/06/2019	Delhi	Cochin	DEL --> BOM --> COK	08:00	21:00	13h	1 stop	No info
4	Air Asia	24/06/2019	Banglore	Delhi	BLR --> DEL	23:55	00:45 25 Jun	2h 50m	non-stop	No info

## Let's Perform that same steps as we did for train data earlier

In [139]: df\_test['Date\_of\_Journey']=pd.to\_datetime(df\_test['Date\_of\_Journey'])  
df\_test['Day\_of\_Journey']=(df\_test['Date\_of\_Journey']).dt.day  
df\_test['Month\_of\_Journey']=(df\_test['Date\_of\_Journey']).dt.month  
  
#Dep time  
df\_test['Dep\_hr']=pd.to\_datetime(df\_test['Dep\_Time']).dt.hour  
df\_test['Dep\_min']=pd.to\_datetime(df\_test['Dep\_Time']).dt.minute  
  
#Arrival time  
df\_test['Arrival\_hr']=pd.to\_datetime(df\_test['Arrival\_Time']).dt.hour  
df\_test['Arrival\_min']=pd.to\_datetime(df\_test['Arrival\_Time']).dt.minute  
  
#Splitting duration time  
  
a=df\_test['Duration'].str.split(' ',expand=True)  
a[1].fillna('00m',inplace=True)  
df\_test['dur\_hr']=a[0].apply(lambda x: x[:-1])  
df\_test['dur\_min']=a[1].apply(lambda x: x[:-1])  
  
#dropping the data  
df\_test.drop(['Date\_of\_Journey','Duration','Arrival\_Time','Dep\_Time'],inplace=True,axis=1)  
  
#Handling Categorical Values  
df\_test['Total\_Stops']=df\_test['Total\_Stops'].map({'non-stop':0, '2 stops':2, '1 stop':1, '3 stops':3, '4 stops':4})  
air\_dummy=pd.get\_dummies(df\_test[['Airline']],drop\_first=True)  
source\_dest\_dummy=pd.get\_dummies(df\_test[['Source','Destination']],drop\_first=True)  
df\_test=pd.concat([air\_dummy,source\_dest\_dummy,df\_test],axis=1)

In [140]: df\_test.drop(['Airline','Source','Destination','Additional\_Info','Route'],inplace=True,axis=1)

In [141]: print('train\_shape',df.shape)  
print('test\_shape',df\_test.shape)  
train\_shape (10682, 15)  
test\_shape (2671, 10)

In [142]: df.head()

Out[142]:

	Air India	GoAir	IndiGo	Jet Airways	Jet Airways Business	Multiple carriers Premium economy	SpiceJet	Trujet	Vistara	...	Additional Info	Price	Day_of_Journey	Mor
0	0	0	1	0	0	0	0	0	0	...	No info	3887	24	
1	1	0	0	0	0	0	0	0	0	...	No info	7662	5	
2	0	0	0	1	0	0	0	0	0	...	No info	13862	6	
3	0	0	1	0	0	0	0	0	0	...	No info	6218	5	
4	0	0	1	0	0	0	0	0	0	...	No info	13302	3	
5 rows x 15 columns														

In [144]: x=df.drop(['Route','Price','Additional\_Info'],axis=1)  
y=df['Price']

In [154]: x.head(3)

Out[154]:

	Air India	GoAir	IndiGo	Jet Airways	Jet Airways Business	Multiple carriers Premium economy	SpiceJet	Trujet	Vistara	...	Destination	Kolkata	Total_Stops	Day_of_Jou
0	0	0	1	0	0	0	0	0	0	...			0	0
1	1	0	0	0	0	0	0	0	0	...			0	2
2	0	0	0	1	0	0	0	0	0	...			0	2
3 rows x 15 columns														

## Model Building and Hyperparameter Tuning

### ExtraTreesRegressor

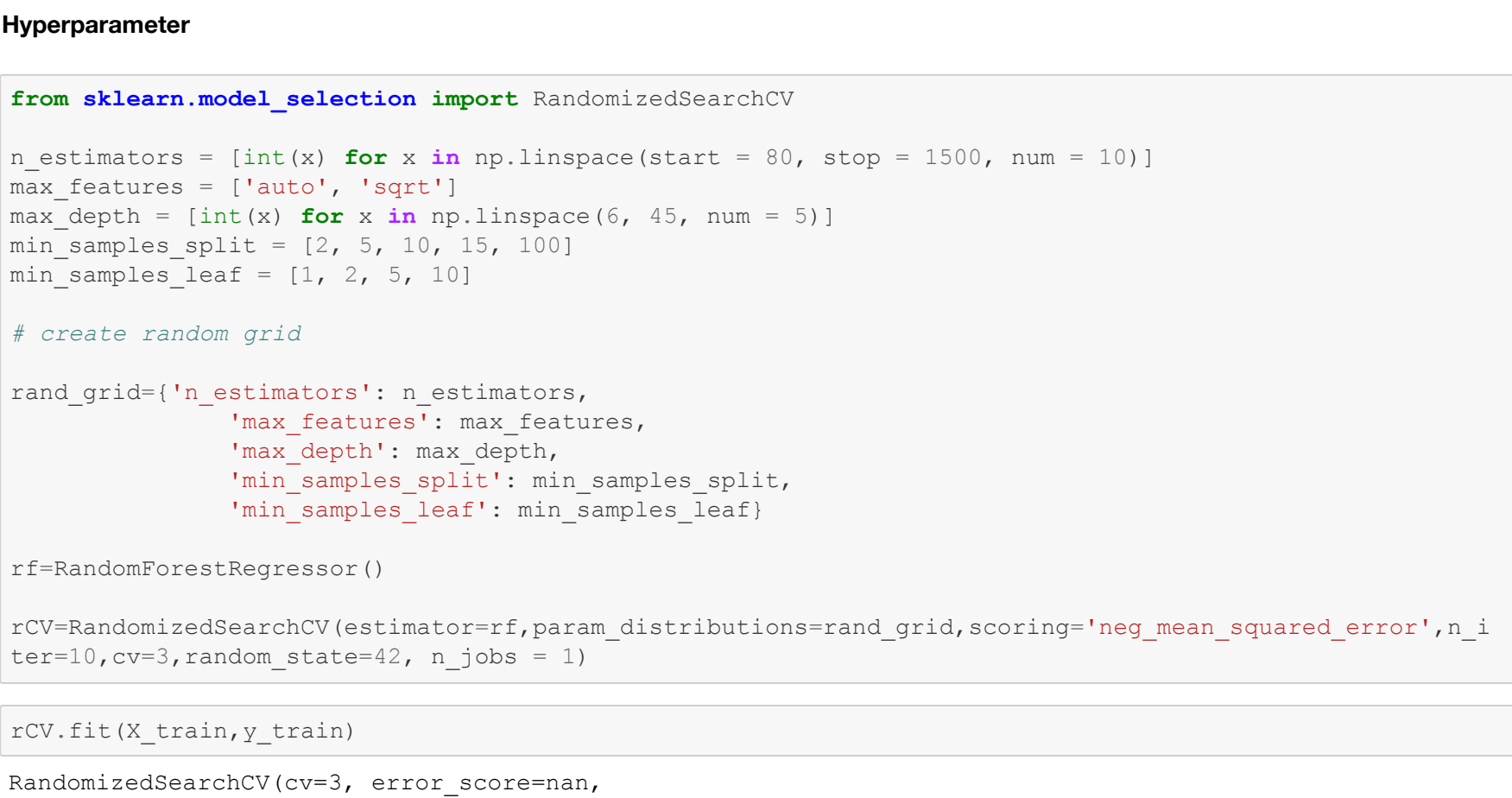
In [150]: from sklearn.ensemble import ExtraTreesRegressor  
ET\_Model=ExtraTreesRegressor()

Out[150]: ExtraTreesRegressor(bootstrap=False, ccp\_alpha=0.0, criterion='mse', max\_depth=None, max\_features='auto', max\_leaf\_nodes=None, max\_samples=None, min\_impurity\_decrease=0.0, min\_impurity\_split=None, min\_samples\_leaf=1, min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0, n\_estimators=100, n\_jobs=None, oob\_score=False, random\_state=None, verbose=0, warm\_start=False)

### Future Importance

In [153]: pd.Series(ET\_Model.feature\_importances\_,index=x.columns).sort\_values(ascending=False).plot(kind='bar',figsize=(18,10))

Out[153]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1a2e18050>



In [156]: #splitting the dataset  
from sklearn.model\_selection import train\_test\_split  
X\_train,X\_test,y\_train,y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 50)

In [157]: #Preparing Extra Tree Regression  
from sklearn.ensemble import ExtraTreesRegressor  
ET\_Model=ExtraTreesRegressor(n\_estimators = 120)  
ET\_Model.fit(X\_train,y\_train)  
y\_pred=ET\_Model.predict(X\_test)

Out[157]:

In [160]: from sklearn.metrics import r2\_score  
r2\_score(y\_test,y\_pred)

Out[160]: 0.7682238280133812

### Hyperparameter

In [178]: from sklearn.model\_selection import RandomizedSearchCV

```
n_estimators = [int(x) for x in np.linspace(start = 80, stop = 1500, num = 10)]  
max_features = ['auto', 'sqrt']  
max_depth = [int(x) for x in np.linspace(6, 45, num = 5)]  
min_samples_split = [5, 10, 15, 100]  
min_samples_leaf = [1, 2, 5, 10]  
  
# create random grid  
rand_grid={'n_estimators': n_estimators,  
           'max_features': max_features,  
           'max_depth': max_depth,  
           'min_samples_split': min_samples_split,  
           'min_samples_leaf': min_samples_leaf}  
  
rf=RandomForestRegressor()  
rCV=RandomizedSearchCV(estimator=rf,param_distributions=rand_grid,scoring='neg_mean_squared_error',n_iter=10,cv=3,random_state=42, n_jobs = 1)
```

In [179]: rCV.fit(X\_train,y\_train)

Out[179]: RandomizedSearchCV(cv=3, error\_score='nan', estimator=RandomForestRegressor(bootstrap=True, ccp\_alpha=0.0, criterion='mse', max\_depth=None, max\_features='auto', max\_leaf\_nodes=None, max\_samples=None, min\_impurity\_decrease=0.0, min\_impurity\_split=None, min\_samples\_leaf=1, min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0, n\_estimators=100, n\_jobs=None, oob\_score=False, random\_state=None, verbose=0, warm\_start=False), scoring='neg\_mean\_squared\_error', verbose=0)

In [180]: rf\_pred=rCV.predict(X\_test)  
rf\_pred

Out[180]: array([[12500.2967405, 4286.76073494, 6136.50519908, ..., 11657.36838275, 9659.32444484, 8655.77083065]])

In [181]: from sklearn.metrics import mean\_absolute\_error,mean\_squared\_error  
print('MSE',mean\_squared\_error(y\_test,rf\_pred))  
print('MSE',mean\_absolute\_error(y\_test,rf\_pred))

MSE 1156.8110505706231  
MSE 3974736.1005828027

In [182]: r2\_score(y\_test,y\_pred)

Out[182]: 0.8064284149245494

In [189]: !pip install catboost

Collecting catboost  
Downloading catboost-0.23.2-cp37-none-macosx\_10\_6\_intel.macosx\_10\_9\_intel.macosx\_10\_9\_x86\_64.whl (10.9 MB)  
Requirement already satisfied: scipy in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (feo m catboost) (1.4.1)  
Requirement already satisfied: plotly in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (from catboost) (4.8.1)  
Requirement already satisfied: six in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (from catboost) (1.14.0)  
Collecting graphviz  
Downloading graphviz-0.14.1-py2.py3-none-any.whl (18 kB)  
Requirement already satisfied: matplotlib in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (from catboost) (3.2.2)  
Requirement already satisfied: pandas>=2.0.4 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-pack ages (from catboost) (1.0.5)  
Requirement already satisfied: numpy>=1.16.0 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packa ges (from catboost) (1.18.1)  
Requirement already satisfied: retrying>=1.3.3 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-pac kages (from plotly->catboost) (1.3.3)  
Requirement already satisfied: kiwisolver>=1.0.1 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-p ackages (from matplotlib->catboost) (1.1.0)  
Requirement already satisfied: pyparsing>=2.0.4 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packa ges (from pandas>=2.0.0->catboost) (2.0.9)  
Requirement already satisfied: python-dateutil>=2.1 in /Users/mybeast/opt/anaconda3/lib/python3.7/sit e-packages (from matplotlib->catboost) (2.8.1)  
Requirement already satisfied: cycler>=0.10 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packa ges (from matplotlib->catboost) (0.10.0)  
Requirement already satisfied: setuptools>=2017.2 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packa ges (from pandas>=2.0.0->catboost) (2019.9)  
Requirement already satisfied: cycle>=0.10 in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packa ges (from kiwisolver>=1.0.1->matplotlib->catboost) (46.0.0.post20200309)  
Installing collected packages: graphviz, catboost  
Successfully installed catboost-0.23.2 graphviz-0.14.1

In [190]:



```
learn_rate: 0.00518
01: learn: 4339.395229 total: 87ms remaining: 11.26s
02: learn: 4339.395229 total: 90.7ms remaining: 45.35s
03: learn: 4339.395229 total: 92.9ms remaining: 45.35s
31: learn: 3909.533292 total: 99.4ms remaining: 20.3s
32: learn: 3909.533292 total: 102ms remaining: 20.3s
33: learn: 3909.533292 total: 103ms remaining: 20.3s
41: learn: 3686.976593 total: 103ms remaining: 17.7s
42: learn: 3686.976593 total: 105ms remaining: 17.7s
71: learn: 3685.971553 total: 110ms remaining: 13.7s
81: learn: 3498.942693 total: 113ms remaining: 12.4s
91: learn: 3498.942693 total: 115ms remaining: 12.4s
101: learn: 3498.942693 total: 118ms remaining: 10.6s
111: learn: 3258.936377 total: 120ms remaining: 9.9s
121: learn: 3258.936377 total: 122ms remaining: 9.9s
131: learn: 3258.936377 total: 125ms remaining: 8.2s
141: learn: 3064.930273 total: 128ms remaining: 8.12s
151: learn: 3064.930273 total: 132ms remaining: 8.22s
161: learn: 2965.955295 total: 135ms remaining: 8.25s
171: learn: 2974.932184 total: 140ms remaining: 7.62s
181: learn: 2874.498490 total: 143ms remaining: 7.42s
191: learn: 2874.498490 total: 145ms remaining: 7.42s
201: learn: 2874.498490 total: 156ms remaining: 7.29s
211: learn: 2760.792210 total: 160ms remaining: 7.11s
221: learn: 2760.792210 total: 163ms remaining: 6.92s
231: learn: 2698.979443 total: 165ms remaining: 6.92s
241: learn: 2670.860702 total: 175ms remaining: 6.28s
251: learn: 2647.450113 total: 184ms remaining: 6.03s
261: learn: 2645.951760 total: 187ms remaining: 6.03s
271: learn: 2602.312844 total: 202ms remaining: 7.0s
281: learn: 2577.739452 total: 208ms remaining: 6.96s
291: learn: 2577.739452 total: 210ms remaining: 6.96s
301: learn: 2577.739452 total: 219ms remaining: 6.85s
311: learn: 2521.293464 total: 224ms remaining: 6.77s
321: learn: 2506.926271 total: 230ms remaining: 6.74s
331: learn: 2493.981963 total: 233ms remaining: 6.72s
341: learn: 2481.947619 total: 241ms remaining: 6.64s
351: learn: 2468.991084 total: 246ms remaining: 6.65s
361: learn: 2468.991084 total: 252ms remaining: 6.58s
371: learn: 2443.986731 total: 258ms remaining: 6.52s
381: learn: 2442.913643 total: 261ms remaining: 6.61s
391: learn: 2429.234794 total: 268ms remaining: 6.61s
401: learn: 2421.979808 total: 271ms remaining: 6.51s
411: learn: 2399.983018 total: 278ms remaining: 6.63s
421: learn: 2389.506627 total: 282ms remaining: 6.27s
431: learn: 2378.951343 total: 285ms remaining: 6.27s
441: learn: 2359.934712 total: 289ms remaining: 6.14s
451: learn: 2361.299493 total: 293ms remaining: 6.08s
461: learn: 2340.950293 total: 315ms remaining: 6.93s
471: learn: 2334.962574 total: 324ms remaining: 6.34s
481: learn: 2317.956440 total: 329ms remaining: 6.24s
491: learn: 2313.938150 total: 337ms remaining: 6.13s
501: learn: 2305.983481 total: 341ms remaining: 6.09s
511: learn: 2277.945310 total: 342ms remaining: 6.08s
521: learn: 2289.914364 total: 352ms remaining: 5.99s
531: learn: 2284.760473 total: 359ms remaining: 5.98s
541: learn: 2277.945310 total: 363ms remaining: 5.98s
551: learn: 2252.967693 total: 378ms remaining: 5.89s
561: learn: 2248.054611 total: 392ms remaining: 5.83s
571: learn: 2236.950360 total: 407ms remaining: 5.78s
581: learn: 2231.092268 total: 423ms remaining: 5.99s
591: learn: 2221.938866 total: 434ms remaining: 5.95s
601: learn: 2214.930351 total: 437ms remaining: 5.95s
611: learn: 2212.285270 total: 451ms remaining: 5.93s
621: learn: 2175.920340 total: 456ms remaining: 5.92s
711: learn: 2159.918039 total: 479ms remaining: 6.04s
721: learn: 2150.940870 total: 482ms remaining: 6.04s
731: learn: 2139.993260 total: 489ms remaining: 6.03s
741: learn: 2189.950393 total: 497ms remaining: 6.04s
751: learn: 2183.879244 total: 504ms remaining: 6.04s
761: learn: 2183.879244 total: 512ms remaining: 6.04s
771: learn: 2175.915266 total: 525ms remaining: 6.12s
781: learn: 2173.935233 total: 550ms remaining: 6.03s
791: learn: 2167.962700 total: 582ms remaining: 6.32s
801: learn: 2164.921245 total: 599ms remaining: 6.62s
811: learn: 2159.945410 total: 609ms remaining: 6.62s
821: learn: 2156.920486 total: 622ms remaining: 6.62s
831: learn: 2150.846411 total: 632ms remaining: 6.63s
841: learn: 2145.084703 total: 645ms remaining: 6.64s
851: learn: 2141.061610 total: 653ms remaining: 6.64s
861: learn: 2136.920360 total: 657ms remaining: 6.57s
871: learn: 2136.920360 total: 668ms remaining: 6.55s
881: learn: 2126.979947 total: 674ms remaining: 6.52s
891: learn: 2126.979947 total: 682ms remaining: 6.52s
901: learn: 2123.953150 total: 684ms remaining: 6.43s
911: learn: 2120.916832 total: 687ms remaining: 6.39s
921: learn: 2117.920360 total: 693ms remaining: 6.39s
931: learn: 2107.931464 total: 702ms remaining: 6.32s
941: learn: 2102.915273 total: 707ms remaining: 6.29s
951: learn: 2100.862410 total: 724ms remaining: 6.58s
961: learn: 2097.938478 total: 751ms remaining: 6.22s
971: learn: 2094.982700 total: 759ms remaining: 6.22s
981: learn: 2094.982700 total: 762ms remaining: 6.22s
991: learn: 2084.971197 total: 773ms remaining: 6.12s
1001: learn: 2084.971197 total: 776ms remaining: 6.09s
1011: learn: 2078.976560 total: 782ms remaining: 6.07s
1021: learn: 2075.912699 total: 786ms remaining: 6.07s
1031: learn: 2073.969822 total: 791ms remaining: 6.01s
1041: learn: 2073.969822 total: 792ms remaining: 6.01s
1051: learn: 2069.934910 total: 794ms remaining: 5.98s
1061: learn: 2066.940195 total: 796ms remaining: 5.96s
1071: learn: 2062.928393 total: 797ms remaining: 5.95s
1081: learn: 2059.993802 total: 798ms remaining: 5.93s
1091: learn: 2055.943741 total: 799ms remaining: 5.92s
1101: learn: 2047.975081 total: 803ms remaining: 5.89s
1111: learn: 2044.971574 total: 810ms remaining: 5.88s
1121: learn: 2044.971574 total: 812ms remaining: 5.88s
1131: learn: 2040.919263 total: 822ms remaining: 5.85s
1141: learn: 2036.913273 total: 828ms remaining: 5.85s
1151: learn: 2035.906761 total: 834ms remaining: 5.83s
1161: learn: 2032.951896 total: 838ms remaining: 5.83s
1171: learn: 2031.940401 total: 845ms remaining: 5.81s
1181: learn: 2029.980578 total: 851ms remaining: 5.81s
1191: learn: 2029.980578 total: 857ms remaining: 5.78s
1201: learn: 2022.958214 total: 869ms remaining: 5.76s
1211: learn: 2016.083145 total: 877ms remaining: 5.72s
1221: learn: 2014.974735 total: 882ms remaining: 5.67s
1231: learn: 2009.979393 total: 886ms remaining: 5.68s
1241: learn: 2007.973400 total: 891ms remaining: 5.67s
1251: learn: 2006.966194 total: 895ms remaining: 5.
```



840: Learn: 1444.6836141 total: 5.664 remaining: 1.07s  
841: Learn: 1443.963568 total: 5.678 remaining: 1.06s  
842: Learn: 1443.963568 total: 5.668 remaining: 1.05s  
843: Learn: 1444.5793488 total: 5.668 remaining: 1.05s  
844: Learn: 1443.2765572 total: 5.678 remaining: 1.04s  
845: Learn: 1442.502917 total: 5.686 remaining: 1.03s  
846: Learn: 1442.5857041 total: 5.698 remaining: 1.03s  
847: Learn: 1441.7708073 total: 5.718 remaining: 1.02s  
848: Learn: 1441.5324245 total: 5.728 remaining: 1.02s  
849: Learn: 1441.1731914 total: 5.728 remaining: 1.01s  
850: Learn: 1440.7375024 total: 5.738 remaining: 1s  
851: Learn: 1440.2537645 total: 5.748 remaining: 999ms  
852: Learn: 1440.3015257 total: 5.758 remaining: 998ms  
853: Learn: 1440.2978302 total: 5.758 remaining: 998ms  
854: Learn: 1439.9885696 total: 5.768 remaining: 976ms  
855: Learn: 1439.8306255 total: 5.768 remaining: 976ms  
856: Learn: 1439.1773033 total: 5.778 remaining: 963ms  
857: Learn: 1438.9957066 total: 5.798 remaining: 958ms  
858: Learn: 1438.4603273 total: 5.798 remaining: 958ms  
859: Learn: 1437.5399741 total: 5.88s remaining: 945ms  
860: Learn: 1437.2149320 total: 5.818 remaining: 938ms  
861: Learn: 1436.9406554 total: 5.828 remaining: 931ms  
862: Learn: 1436.4516603 total: 5.838 remaining: 929ms  
863: Learn: 1436.2854041 total: 5.838 remaining: 918ms  
864: Learn: 1435.9279608 total: 5.848 remaining: 912ms  
865: Learn: 1435.7649645 total: 5.928 remaining: 910ms  
866: Learn: 1435.2703883 total: 5.868 remaining: 899ms  
867: Learn: 1434.7197487 total: 5.878 remaining: 892ms  
868: Learn: 1434.2556845 total: 5.878 remaining: 892ms  
869: Learn: 1434.2560496 total: 5.888 remaining: 879ms  
870: Learn: 1433.9749701 total: 5.898 remaining: 872ms  
871: Learn: 1433.2291566 total: 5.898 remaining: 869ms  
872: Learn: 1432.8211329 total: 5.98s remaining: 859ms  
873: Learn: 1432.4084235 total: 5.91s remaining: 852ms  
874: Learn: 1432.0185610 total: 5.92s remaining: 845ms  
875: Learn: 1432.0164106 total: 5.92s remaining: 836ms  
876: Learn: 1431.4370466 total: 5.938 remaining: 832ms  
877: Learn: 1431.1682666 total: 5.948 remaining: 825ms  
878: Learn: 1430.9511382 total: 5.978 remaining: 819ms  
879: Learn: 1430.9477950 total: 5.998 remaining: 812ms  
880: Learn: 1430.4077459 total: 5.958 remaining: 804ms  
881: Learn: 1430.0129827 total: 5.968 remaining: 797ms  
882: Learn: 1429.4996655 total: 5.978 remaining: 790ms  
883: Learn: 1429.1418463 total: 5.978 remaining: 784ms  
884: Learn: 1428.9533784 total: 5.988 remaining: 777ms  
885: Learn: 1428.9534875 total: 5.998 remaining: 770ms  
886: Learn: 1428.6645775 total: 5.998 remaining: 763ms  
887: Learn: 1428.1646398 total: 6s remaining: 757ms  
888: Learn: 1427.9468079 total: 6.018 remaining: 750ms  
889: Learn: 1427.4871135 total: 6.018 remaining: 743ms  
890: Learn: 1427.6841885 total: 6.028 remaining: 736ms  
891: Learn: 1427.4921877 total: 6.028 remaining: 729ms  
892: Learn: 1427.2639447 total: 6.028 remaining: 722ms  
893: Learn: 1426.7911065 total: 6.048 remaining: 716ms  
894: Learn: 1426.4899662 total: 6.058 remaining: 709ms  
895: Learn: 1426.3107111 total: 6.058 remaining: 702ms  
896: Learn: 1425.8850561 total: 6.068 remaining: 696ms  
897: Learn: 1425.5080254 total: 6.078 remaining: 689ms  
898: Learn: 1425.0586399 total: 6.078 remaining: 682ms  
899: Learn: 1424.5001044 total: 6.098 remaining: 676ms  
900: Learn: 1424.2669264 total: 6.098 remaining: 669ms  
901: Learn: 1423.8482508 total: 6.1s remaining: 663ms  
902: Learn: 1423.6232016 total: 6.128 remaining: 656ms  
903: Learn: 1423.2905391 total: 6.118 remaining: 649ms  
904: Learn: 1423.1176866 total: 6.128 remaining: 642ms  
905: Learn: 1423.1143925 total: 6.128 remaining: 635ms  
906: Learn: 1422.5288967 total: 6.138 remaining: 628ms  
907: Learn: 1422.2110598 total: 6.148 remaining: 622ms  
908: Learn: 1421.7953912 total: 6.148 remaining: 615ms  
909: Learn: 1421.4991174 total: 6.178 remaining: 608ms  
910: Learn: 1421.5456320 total: 6.168 remaining: 602ms  
911: Learn: 1421.3602840 total: 6.178 remaining: 595ms  
912: Learn: 1421.0583839 total: 6.178 remaining: 588ms  
913: Learn: 1420.5044573 total: 6.188 remaining: 581ms  
914: Learn: 1420.0678475 total: 6.198 remaining: 575ms  
915: Learn: 1419.5380979 total: 6.198 remaining: 568ms  
916: Learn: 1419.2384519 total: 6.198 remaining: 561ms  
917: Learn: 1419.5650873 total: 6.218 remaining: 554ms  
918: Learn: 1419.3920275 total: 6.218 remaining: 548ms  
919: Learn: 1419.1755008 total: 6.228 remaining: 541ms  
920: Learn: 1418.5597912 total: 6.238 remaining: 534ms  
921: Learn: 1418.2654170 total: 6.238 remaining: 527ms  
922: Learn: 1418.0593003 total: 6.248 remaining: 520ms  
923: Learn: 1417.6692082 total: 6.258 remaining: 514ms  
924: Learn: 1417.3704203 total: 6.268 remaining: 507ms  
925: Learn: 1416.9973968 total: 6.268 remaining: 500ms  
926: Learn: 1416.4516603 total: 6.278 remaining: 493ms  
927: Learn: 1416.3590845 total: 6.278 remaining: 487ms  
928: Learn: 1416.0422781 total: 6.288 remaining: 480ms  
929: Learn: 1415.6560350 total: 6.298 remaining: 473ms  
930: Learn: 1415.2973622 total: 6.298 remaining: 467ms  
931: Learn: 1414.7907054 total: 6.3s remaining: 460ms  
932: Learn: 1414.4843311 total: 6.318 remaining: 453ms  
933: Learn: 1413.9827997 total: 6.328 remaining: 447ms  
934: Learn: 1413.5659494 total: 6.338 remaining: 440ms  
935: Learn: 1413.3011488 total: 6.338 remaining: 433ms  
936: Learn: 1412.9303088 total: 6.348 remaining: 426ms  
937: Learn: 1412.9093580 total: 6.348 remaining: 419ms  
938: Learn: 1412.4984716 total: 6.358 remaining: 413ms  
939: Learn: 1412.3107111 total: 6.368 remaining: 406ms  
940: Learn: 1412.1012221 total: 6.378 remaining: 399ms  
941: Learn: 1411.8666136 total: 6.378 remaining: 392ms  
942: Learn: 1411.3555451 total: 6.388 remaining: 386ms  
943: Learn: 1411.2384519 total: 6.398 remaining: 379ms  
944: Learn: 1410.758195 total: 6.48s remaining: 373ms  
945: Learn: 1410.2869816 total: 6.41s remaining: 366ms  
946: Learn: 1410.1686109 total: 6.428 remaining: 359ms  
947: Learn: 1409.9576388 total: 6.428 remaining: 352ms  
948: Learn: 1409.7019315 total: 6.438 remaining: 346ms  
949: Learn: 1409.4523003 total: 6.448 remaining: 339ms  
950: Learn: 1409.4242917 total: 6.458 remaining: 332ms  
951: Learn: 1409.1164346 total: 6.468 remaining: 326ms  
952: Learn: 1408.9998123 total: 6.468 remaining: 319ms  
953: Learn: 1408.9456486 total: 6.478 remaining: 312ms  
954: Learn: 1408.3353653 total: 6.488 remaining: 305ms  
955: Learn: 1407.9314986 total: 6.488 remaining: 298ms  
956: Learn: 1407.5987632 total: 6.498 remaining: 292ms  
957: Learn: 1407.4323027 total: 6.5s remaining: 285ms  
958: Learn: 1406.9693403 total: 6.518 remaining: 278ms  
959: Learn: 1406.4843112 total: 6.518 remaining: 271ms  
960: Learn: 1406.4428982 total: 6.528 remaining: 265ms  
961: Learn: 1405.9767107 total: 6.538 remaining: 258ms  
962: Learn: 1405.7042827 total: 6.548 remaining: 251ms  
963: Learn: 1405.5023008 total: 6.548 remaining: 244ms  
964: Learn: 1405.1841130 total: 6.558 remaining: 238ms  
965: Learn: 1404.840563 total: 6.568 remaining: 231ms  
966: Learn: 1404.6805300 total: 6.578 remaining: 224ms  
967: Learn: 1404.4669353 total: 6.64s remaining: 220ms  
968: Learn: 1404.3237897 total: 6.678 remaining: 213ms  
969: Learn: 1403.8947241 total: 6.688 remaining: 207ms  
970: Learn: 1403.633009 total: 6.718 remaining: 200ms  
971: Learn: 1403.3692929 total: 6.728 remaining: 193ms  
972: Learn: 1403.1661247 total: 6.728 remaining: 187ms  
973: Learn: 1402.8613265 total: 6.738 remaining: 180ms  
974: Learn: 1402.5869729 total: 6.748 remaining: 173ms  
975: Learn: 1402.3533067 total: 6.748 remaining: 166ms  
976: Learn: 1402.0986185 total: 6.748 remaining: 159ms  
977: Learn: 1401.9177844 total: 6.758 remaining: 152ms  
978: Learn: 1401.7392305 total: 6.758 remaining: 145ms  
979: Learn: 1401.4846041 total: 6.758 remaining: 138ms  
980: Learn: 1401.2180449 total: 6.778 remaining: 131ms  
981: Learn: 1400.6664373 total: 6.768 remaining: 124ms  
982: Learn: 1400.3446627 total: 6.778 remaining: 117ms  
983: Learn: 1400.2177865 total: 6.788 remaining: 110ms  
984: Learn: 1399.8623806 total: 6.788 remaining: 103ms  
985: Learn: 1399.5501711 total: 6.788 remaining: 96.3ms  
986: Learn: 1399.0022085 total: 6.798 remaining: 89.4ms  
987: Learn: 1398.6201150 total: 6.798 remaining: 82.5ms  
988: Learn: 1398.5153012 total: 6.798 remaining: 75.6ms  
989: Learn: 1398.3524555 total: 6.8s remaining: 68.7ms  
990: Learn: 1397.9041039 total: 6.818 remaining: 61.8ms  
991: Learn: 1397.6403814 total: 6.818 remaining: 54.9ms  
992: Learn: 1397.2668980 total: 6.828 remaining: 48ms  
993: Learn: 1397.1235980 total: 6.828 remaining: 41.2ms  
994: Learn: 1396.9517827 total: 6.828 remaining: 34.3ms  
995: Learn: 1396.6956847 total: 6.838 remaining: 27.4ms  
996: Learn: 1396.4003923 total: 6.838 remaining: 20.6ms  
997: Learn: 1395.9341021 total: 6.848 remaining: 13.7ms  
998: Learn: 1395.8518997 total: 6.848 remaining: 6.85ms  
999: Learn: 1395.7331331 total: 6.858 remaining: 0s

Out[192]: <catboost.core.CatBoostRegressor at 0x1a53ab8680>

In [ ]:

In [193]: cat\_pred=cat.predict(X\_test)

In [195]: r2\_score(y\_test,cat\_pred)

Out[195]: 0.8301928561140887

In [196]: !pip install lightgbm

Collecting lightgbm  
 Downloading lightgbm-2.3.1-py2.py3-none-macosx\_10\_9\_x86\_64.macosx\_10\_10\_x86\_64.macosx\_10\_11\_x86\_64.macosx\_10\_12\_x86\_64.macosx\_10\_13\_x86\_64.macosx\_10\_14\_x86\_64.macosx\_10\_15\_x86\_64.whl (679 kB)  
Requirement already satisfied: scipy in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (fco m lightgbm) (1.4.1)  
Requirement already satisfied: numpy in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (fco m lightgbm) (1.18.1)  
Requirement already satisfied: scikit-learn in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packag es (from lightgbm) (0.22.1)  
876: Learn: 1429.4996655 total: 5.978 remaining: 790ms  
es (from scikit-learn>lightgbm) (0.14.1)  
Installing collected packages: lightgbm  
Successfully installed lightgbm-2.3.1

In [201]: X\_train[['duration\_hr','duration\_min']]>X\_train[['duration\_hr','duration\_min']].astype(int)  
X\_test[['duration\_hr','duration\_min']]>X\_test[['duration\_hr','duration\_min']].astype(int)

In [202]: from lightgbm import LGBMRegressor  
lgb\_model = LGBMRegressor()  
lgb\_model.fit(X\_train,y\_train)

Out[202]: LGBMRegressor(boosting\_type='gbdt', class\_weight=None, colsample\_bytree=1.0, importance\_type='split', learning\_rate=0.1, max\_depth=1, min\_child\_samples=20, min\_child\_weight=0.01, min\_split\_gain=0.0, n\_estimators=100, n\_jobs=-1, num\_leaves=31, objective=None, random\_state=None, reg\_alpha=0.0, reg\_lambda=0.0, silent=True, subsample=1.0, subsample\_for\_bin=200000, subsample\_freq=0)

In [204]: lgb\_pred=lgb\_model.predict(X\_test)  
r2\_score(y\_test,lgb\_pred)

Out[204]: 0.8030064936005559

In [ ]:

In [206]: !pip install xgboost

Collecting xgboost  
 Downloading xgboost-1.1.1-py3-none-macosx\_10\_13\_x86\_64.macosx\_10\_14\_x86\_64.macosx\_10\_15\_x86\_64.whl (1.1 MB)  
Requirement already satisfied: numpy in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (fco m xgboost) (1.18.1)  
Requirement already satisfied: scipy in /Users/mybeast/opt/anaconda3/lib/python3.7/site-packages (fco m xgboost) (1.4.1)  
Installing collected packages: xgboost  
Successfully installed xgboost-1.1.1

In [209]: import xgboost as xgb  
xgb\_model=xgb.XGBRegressor()  
xgb\_model.fit(X\_train,y\_train)  
xgb\_pred=xgb\_model.predict(X\_test)  
r2\_score(y\_test,xgb\_pred)

Out[209]: 0.8212496826037233

In [210]: df.head()

Out[210]:

	Air India	GoAir	Indigo	Jet Airways	Jet Airways Business	Multiple carriers	Multiple carriers Premium economy	SpiceJet	Trujet	Vistara	...	Additional Info	Price	Day_of_Journey	Mor
0	0	0	1	0	0	0	0	0	0	0	...	No info	3867	24	
1	1	0	0	0	0	0	0	0	0	0	...	No info	7662	5	
2	0	0	0	1	0	0	0	0	0	0	...	No info	13882	6	
3	0	0	1	0	0	0	0	0	0	0	...	No info	6218	8	
4	0	0	1	0	0	0	0	0	0	0	...	No info	13302	3	
5 rows x 31 columns															

In [219]: # Use pickle to save our model so that we can use it later

! import pickle  
# Saving model to disk  
pickle.dump(lgb\_model, open('model.pkl','wb'))  
# model=pickle.load(open('model.pkl','rb'))

In [213]: df.columns

Out[213]: Index(['Air India', 'GoAir', 'Indigo', 'Jet Airways', 'Jet Airways Business', 'Multiple carriers', 'Multiple carriers Premium economy', 'SpiceJet', 'Trujet', 'Vistara', 'Vistara Premium economy', 'Source\_Chennai', 'Source\_Delhi', 'Source\_Hyderabad', 'Source\_Kolkata', 'Source\_Mumbai', 'Destination\_Cochin', 'Destination\_Delhi', 'Destination\_Hyderabad', 'Destination\_Kolkata', 'Route', 'Total\_Stops', 'Additional\_Info', 'Price', 'Day\_of\_Journey', 'Month\_of\_Journey', 'Dep\_hr', 'Dep\_min', 'Arrival\_hr', 'Arrival\_min', 'duration\_hr', 'duration\_min'], dtype=object)

In [216]: deploy\_df=df.drop(['Route','Additional\_Info'],axis=1)

In [217]: deploy\_df.to\_csv('deploy\_df')

In [ ]: