Pattern Recognition and Machine Learning
Bonus Project : Predict The Flight Ticket Price
Project Report

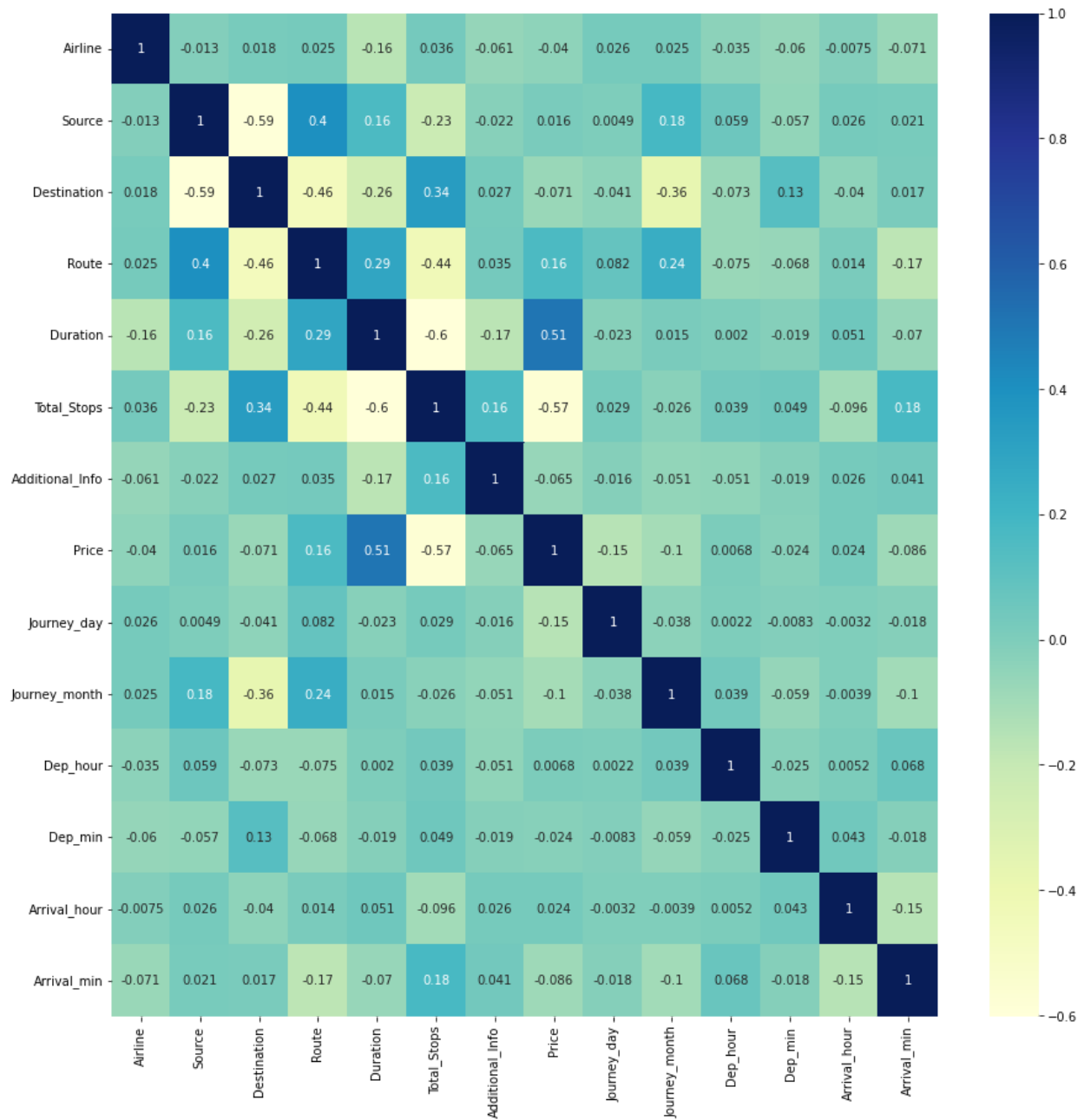## Tammireddi Sai Unnathi - B20AI045

Data Pre-Processing

Airline, Date_of_Journey, Source, Destination, Route, Dep_Time, Arrival_Time,
Duration, Total_Stops, Additional_Info, Price are the columns of the data set
after pre-processing and completing label encoding the dataset is ready.

Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
Airline	10683.000000	3.965927	2.352155	0.000000	3.000000	4.000000	4.000000	11.000000
Source	10683.000000	1.952261	1.177221	0.000000	2.000000	2.000000	3.000000	4.000000
Destination	10683.000000	1.436113	1.474782	0.000000	0.000000	1.000000	2.000000	5.000000
Route	10683.000000	74.825143	36.730924	0.000000	48.000000	73.000000	104.000000	128.000000
Duration	10683.000000	643.093232	507.862001	5.000000	170.000000	520.000000	930.000000	2860.000000
Total_Stops	10683.000000	1.458579	1.806560	0.000000	0.000000	0.000000	4.000000	5.000000
Additional_Info	10683.000000	7.393054	1.214211	0.000000	8.000000	8.000000	8.000000	9.000000
Price	10683.000000	9087.064121	4611.359167	1759.000000	5277.000000	8372.000000	12373.000000	79512.000000
Journey_day	10683.000000	13.508378	8.479277	1.000000	6.000000	12.000000	21.000000	27.000000
Journey_month	10683.000000	4.708602	1.164357	3.000000	3.000000	5.000000	6.000000	6.000000
Dep_hour	10683.000000	12.490686	5.748650	0.000000	8.000000	11.000000	18.000000	23.000000
Dep_min	10683.000000	24.411214	18.767980	0.000000	5.000000	25.000000	40.000000	55.000000
Arrival_hour	10683.000000	13.348778	6.859125	0.000000	8.000000	14.000000	19.000000	23.000000
Arrival_min	10683.000000	24.690630	16.506036	0.000000	10.000000	25.000000	35.000000	55.000000

Data Analysis

	Airline	Source	Destination	Route	Duration	Total_Stops	Additional_Info	Price	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min
Airline	1	-0.013	0.018	0.025	-0.16	0.036	-0.061	-0.04	0.026	0.025	-0.035	-0.06	-0.0075	-0.071
Source	-0.013	1	-0.59	0.4	0.16	-0.23	-0.022	0.016	0.0049	0.18	0.059	-0.057	0.026	0.021
Destination	0.018	-0.59	1	-0.46	-0.26	0.34	0.027	-0.071	-0.041	-0.36	-0.073	0.13	-0.04	0.017
Route	0.025	0.4	-0.46	1	0.29	-0.44	0.035	0.16	0.082	0.24	-0.075	-0.068	0.014	-0.17
Duration	-0.16	0.16	-0.26	0.29	1	-0.6	-0.17	0.51	-0.023	0.015	0.002	-0.019	0.051	-0.07
Total_Stops	0.036	-0.23	0.34	-0.44	-0.6	1	0.16	-0.57	0.029	-0.026	0.039	0.049	-0.096	0.18
Additional_Info	-0.061	-0.022	0.027	0.035	-0.17	0.16	1	-0.065	-0.016	-0.051	-0.051	-0.019	0.026	0.041
Price	-0.04	0.016	-0.071	0.16	0.51	-0.57	-0.065	1	-0.15	-0.1	0.0068	-0.024	0.024	-0.086
Journey_day	0.026	0.0049	-0.041	0.082	-0.023	0.029	-0.016	-0.15	1	-0.038	0.0022	-0.0083	-0.0032	-0.018
Journey_month	0.025	0.18	-0.36	0.24	0.015	-0.026	-0.051	-0.1	-0.038	1	0.039	-0.059	-0.0039	-0.1
Dep_hour	-0.035	0.059	-0.073	-0.075	0.002	0.039	-0.051	0.0068	0.0022	0.039	1	-0.025	0.0052	0.068
Dep_min	-0.06	-0.057	0.13	-0.068	-0.019	0.049	-0.019	-0.024	-0.0083	-0.059	-0.025	1	0.043	-0.018
Arrival_hour	-0.0075	0.026	-0.04	0.014	0.051	-0.096	0.026	0.024	-0.0032	-0.0039	0.0052	0.043	1	-0.15
Arrival_min	-0.071	0.021	0.017	-0.17	-0.07	0.18	0.041	-0.086	-0.018	-0.1	0.068	-0.018	-0.15	1

Scores before applying Grid Search CV

	Model Name	Accuracy Scores
0	Random Forest Regressor	0.883526
1	Gradient Boosting Regressor	0.822685
2	XGBRegressor	0.818542
3	Decision Tree Regressor	0.787071

Scores after applying Grid Search CV

	Model Name	Accuracy Scores
1	Gradient Boosting Regressor	0.903748
2	XGBRegressor	0.894173
0	Random Forest Regressor	0.792177
3	Decision Tree Regressor	0.496854

Before applying grid search cv random forest regressor has highest score, but after applying grid search cv gradient boosting regressor has highest accuracy because of the hypertuning the model by changing the parameters.