**Graphs :**

**Accuracy**

**Feature Selection :**

In mile stone 1: selected features are:

Average\_Score","Review\_Total\_Negative\_Word\_Counts","Review\_Total\_Positive\_Word\_Counts",'Reviewer\_Score

In mile stone 2: selected features are:

Reviewer\_Score','Hotel\_Address','Review\_Date','Hotel\_Name','Reviewer\_Nationality',

'Tags','Review\_Total\_Positive\_Word\_Counts','Review\_Total\_Negative\_Word\_Counts

1-We removed the Average\_Score feature as it didn’t affect the accuracy and got low correlation

2-Added the Hotel\_Address ,Review\_Date,Hotel\_Name,Reviewer\_Nationality,Tags As it got Higher correlation and it made the accuracy higher.

**Hyperparameters :**

**K in KNN:** K=35 ,Accuracy =65.5 , train time = 0.90, test\_time= 26.5

K=20 ,Accuracy=65.1,train time=1.1 ,test\_time=21.6

K=10,Accuracy=64.1,train time=0.8, test\_time=16.2

**Max\_Depth in Decision tree**: depth=4,Accuracy=85.9,train\_time=0.7,test time=0.04

depth=10,Accuracy=86.8,train\_time=1.3,test time=0.02

depth=15,Accuracy=86.3,train\_time=1.9,test time=0.03

**Dimension Reduction:**

**Decision Tree(max Depth=10):Setting Components to:**

N\_components=1,Accuracy=82.9,train time =0.9,test time =0.01

N\_Components=5,Accuracy=85.6,train time=3.5,test time=0.02

N\_components=10,Accuracy=86.7,train time=7.4,test time=0.03

**Naïve Bayes:Setting Components to:**

N\_components=1,Accuracy=81.3,train time =0.1,test time =0.05

N\_Components=5,Accuracy=84.8,train time=0.3,test time=0.1

N\_components=10,Accuracy=85.6,train time=0.4,test time=0.1

**Random Forest (n=1):Setting Components to:**

N\_components=1,Accuracy=78.4,train time =3.5,test time =0.1

N\_Components=5,Accuracy=81.8,train time=3.5,test time=0.04

N\_components=10,Accuracy=83,train time=5.9,test time=0.4

**KNN (n=1):Setting Components to:**

N\_components=1,Accuracy=56,train time =1,test time =2.7

N\_Components=5,Accuracy=59.5,train time=0.4,test time=10.3

N\_components=10,Accuracy=65,train time=0.6,test time=95

Conclusion: The problem in our mile stone was to handle the missing values without dropping and we decided to replace the missing value in numerical data types with the mean of the data in that column ,while the string data types columns was replaced with not found .

Using Classifiers with the proper preprocessing techniques should increases the accuracy, as the accuracy increased and the mean squared error decreased more than mile stone one while using Regression models .