Name - Abhishek Jani

Roll No - 16

**DSPL PROJECT REPORT**

**Topic - House Rent Prediction System**

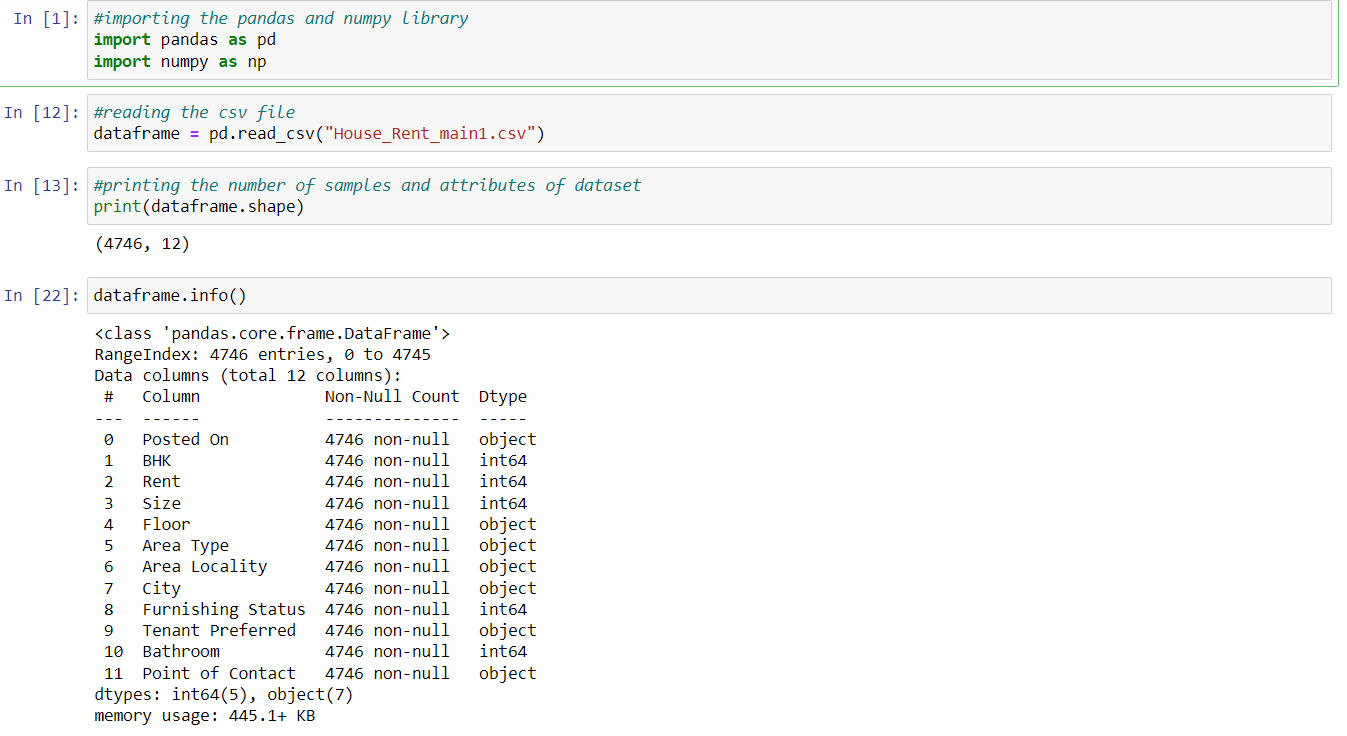
**Problem Statement -**

The House Rent price Prediction System strives to locate the best properties in your neighborhood with the most affordable rental rates.

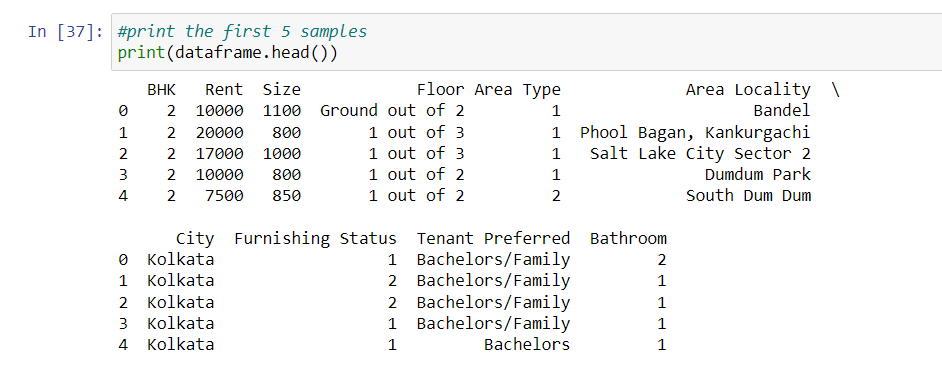
The House Rent Prediction System aims to address these issues by providing accurate and transparent rent predictions based on relevant factors. This will help landlords set rents that are in line with market rates, and tenants make more informed decisions about their housing options.

**Data Preprocessing-**

1. **Cleaning:**



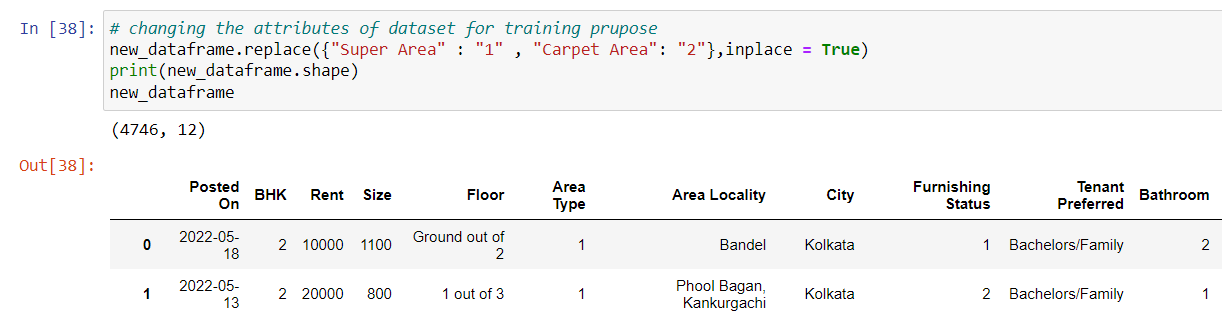
1.a) Head of the dataset:

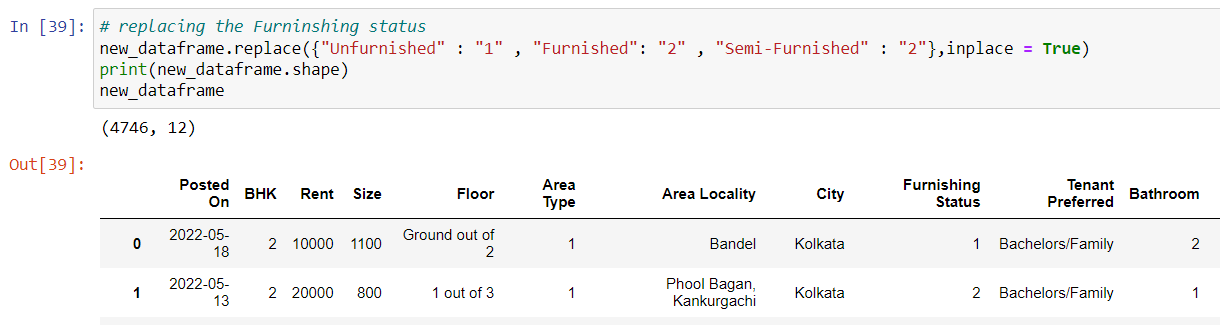


1.b) Missing Values:

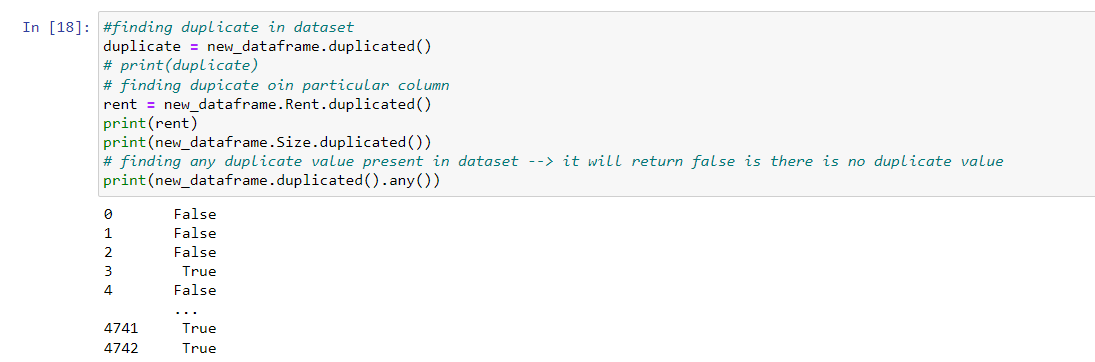


1.c) Changing the string values to integers:



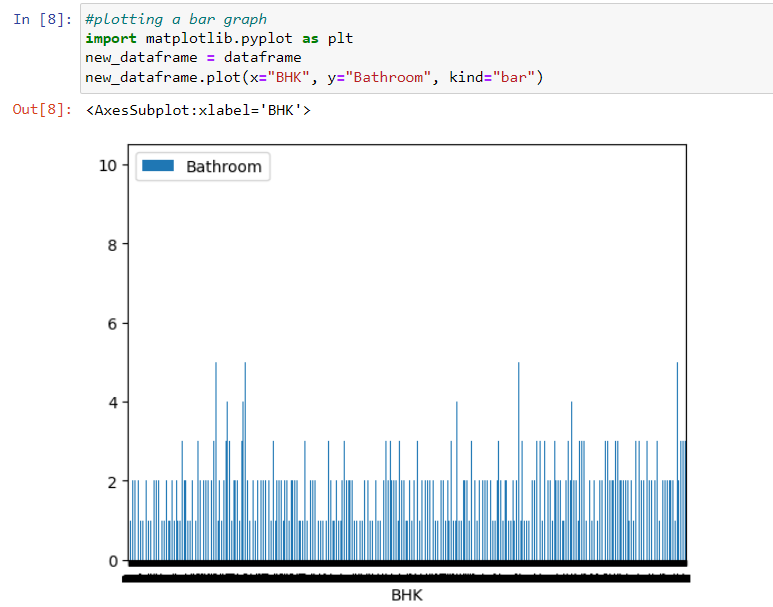


1.d) Finding Duplicates:

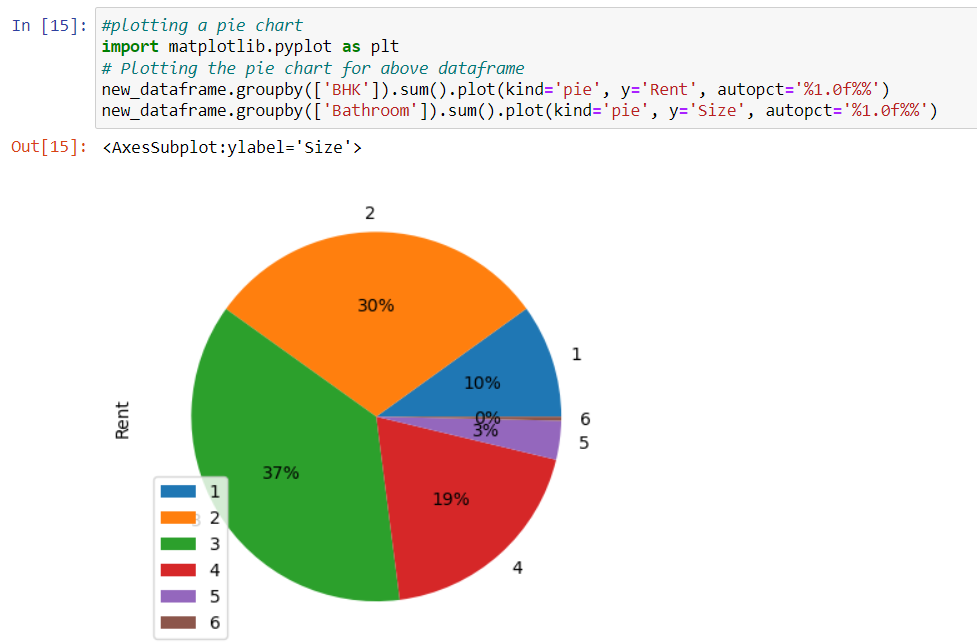


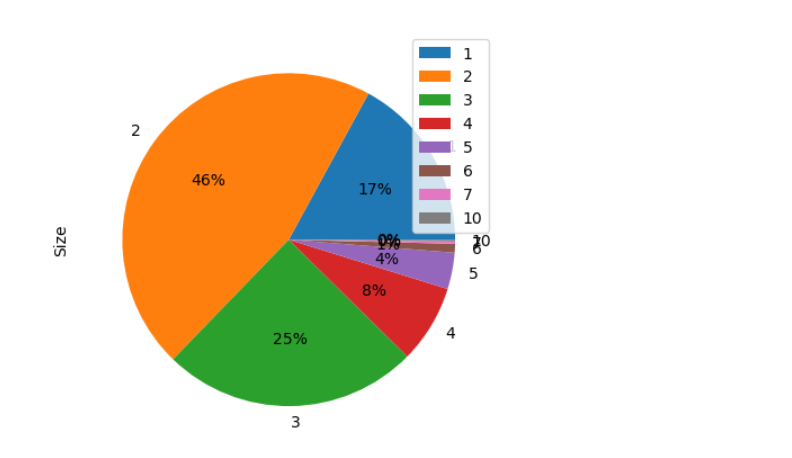
1. **Data Visualization:**

2.a) Bar Graph

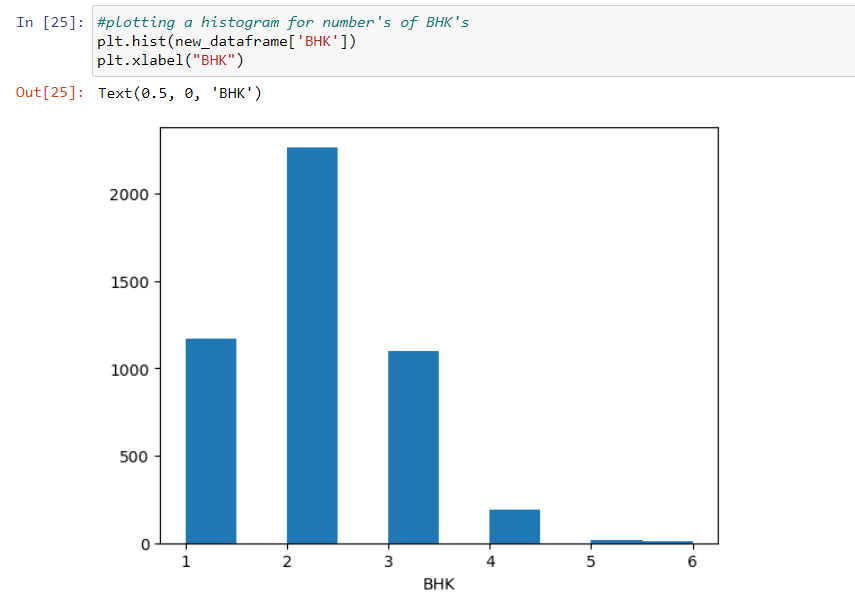


2.b) Pie Chart

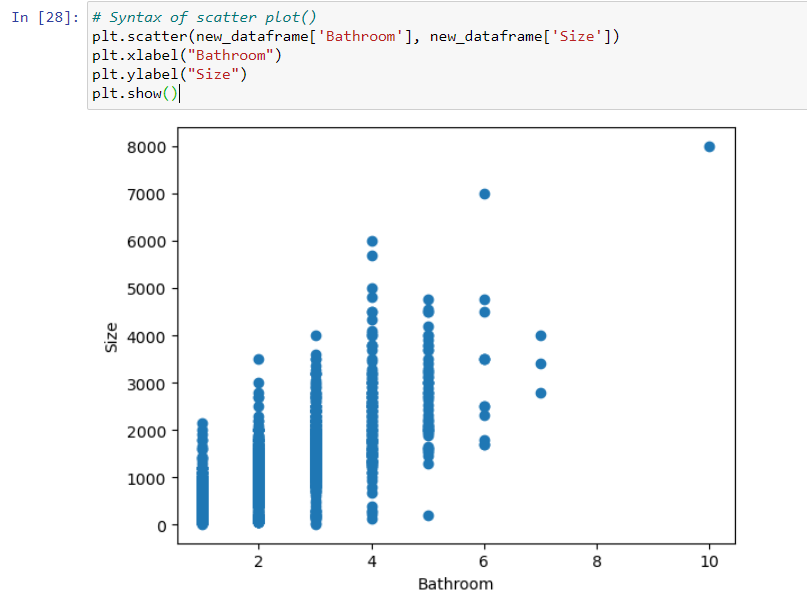




2.c) Univariate Histogram -

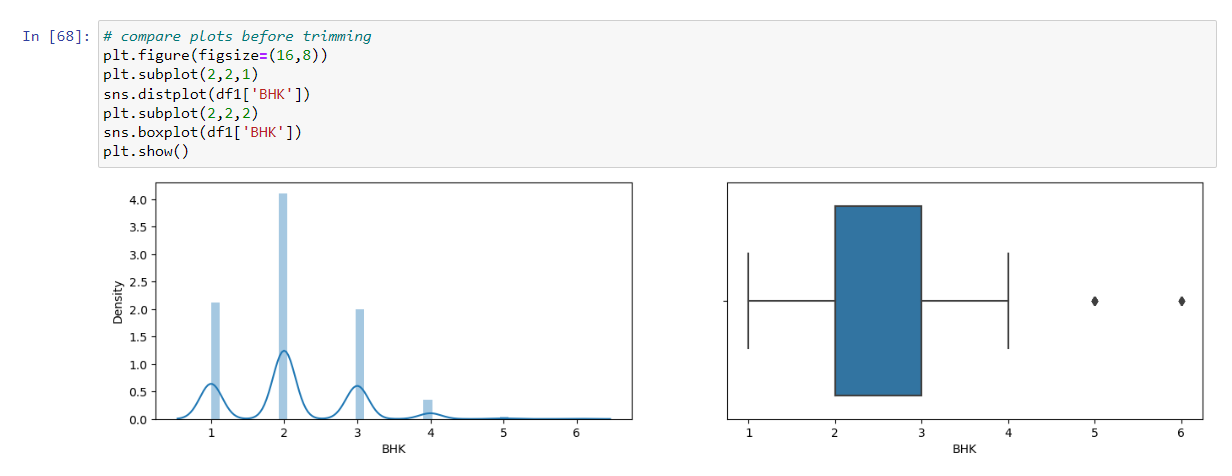


2.d) Scatter Plot

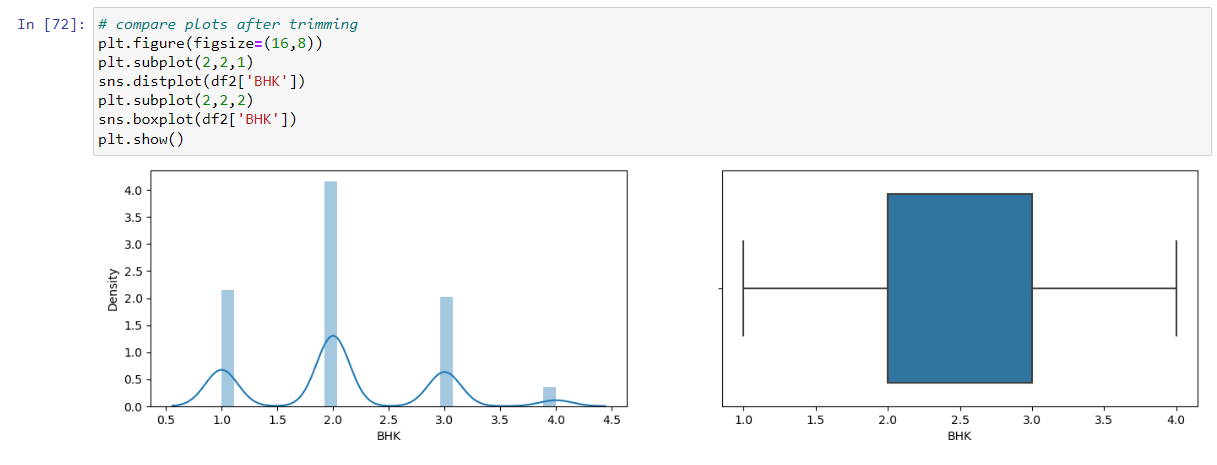


**3.Removing the outliers**

3.a) Before Trimming:



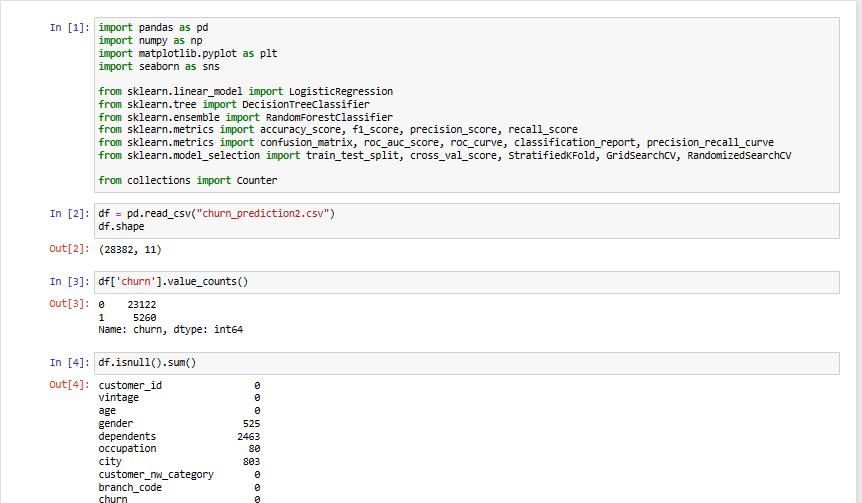
3.b) After Trimming:



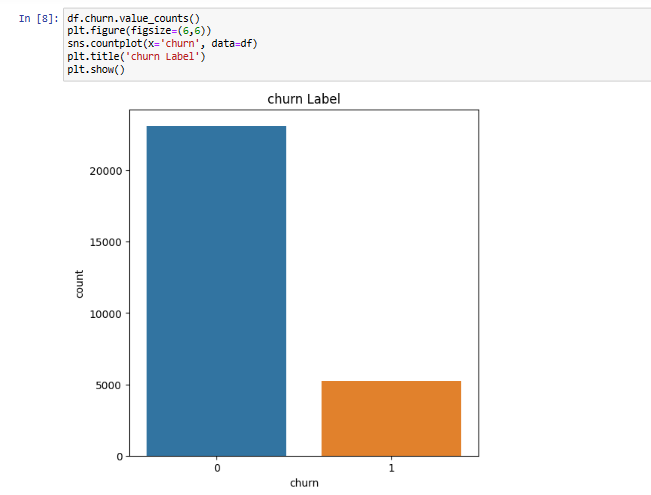
**4.Handling Class Imbalance :**

Handling class imbalance refers to the techniques and strategies used to address the problem of imbalanced class distributions in a dataset, where one class (the minority class) has significantly fewer instances than another class (the majority class).

Imported the Churns Dataset



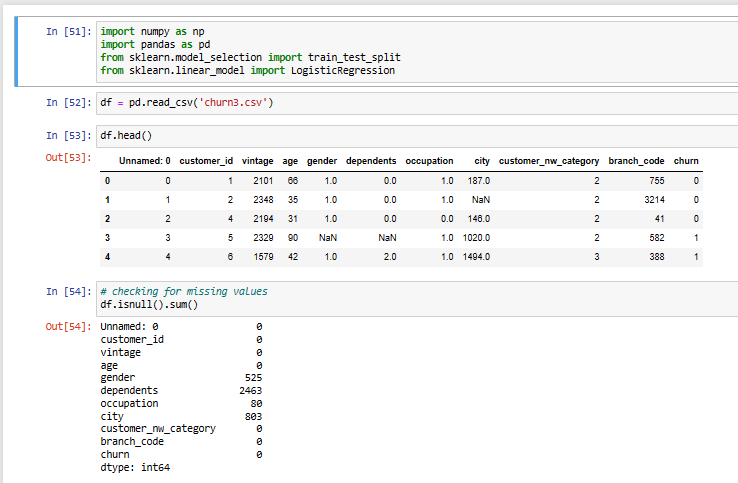
We have used **SMOTE** (Synthetic Minority Oversampling Technique) in our dataset. This technique will duplicate the tuples of minority class and it will balance our Class Label.

Before Applying the Smote technique the churn attributes was Unbalanced

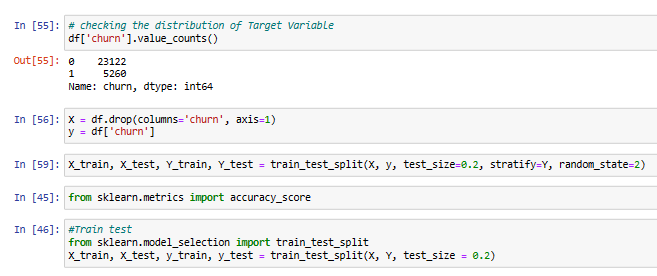
After using SMOTE churn class label get balance :

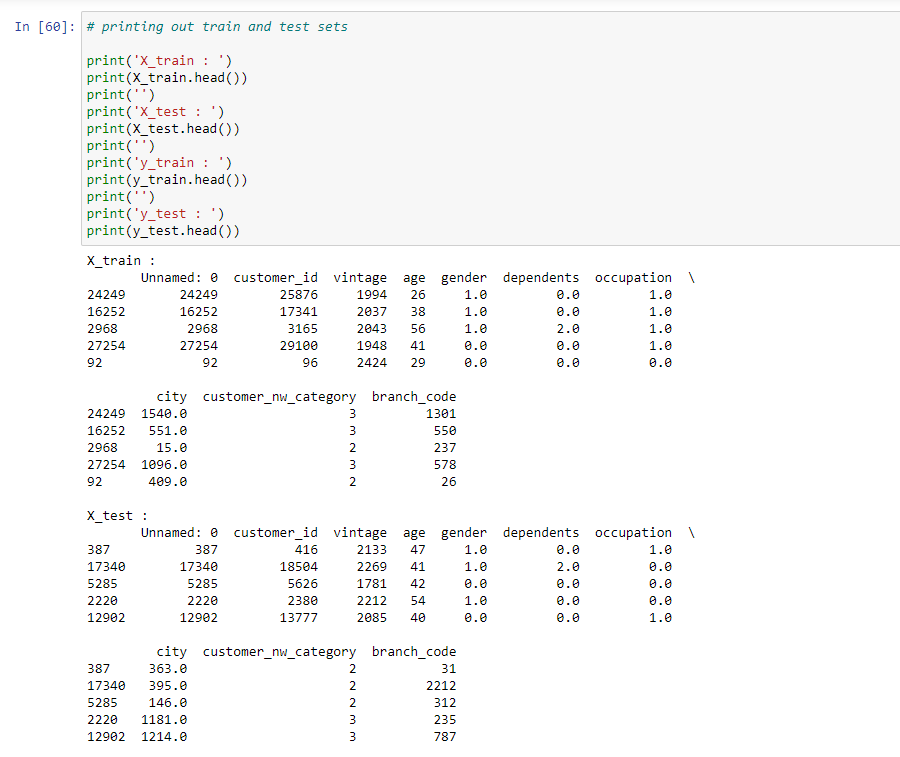


**5. Partition the dataset in training and testing:**

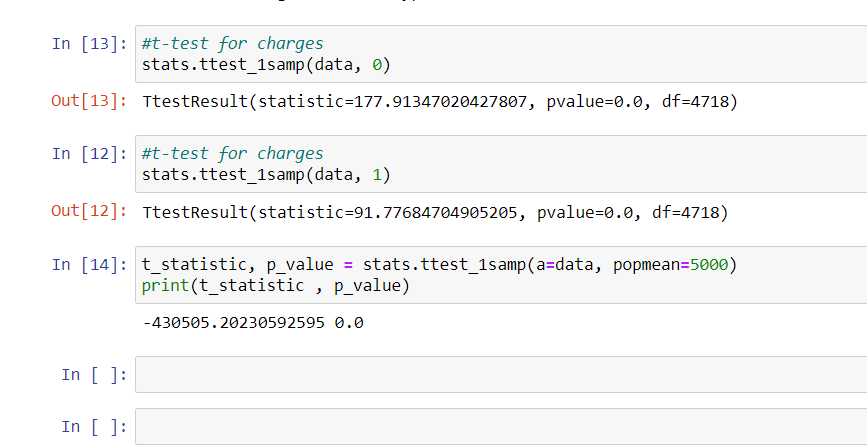
****

Dividing the dataset into training and testing

****

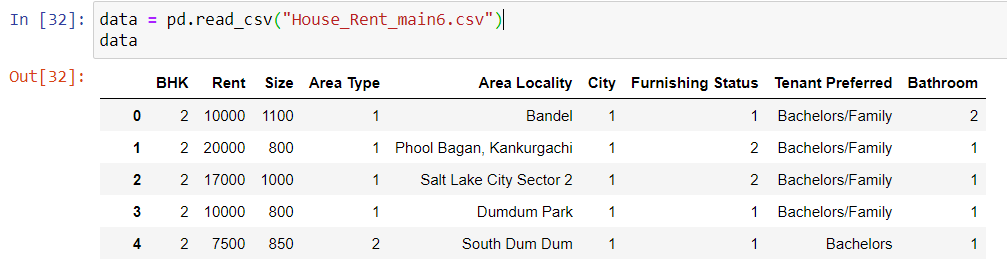
****

**6. T-Test :**

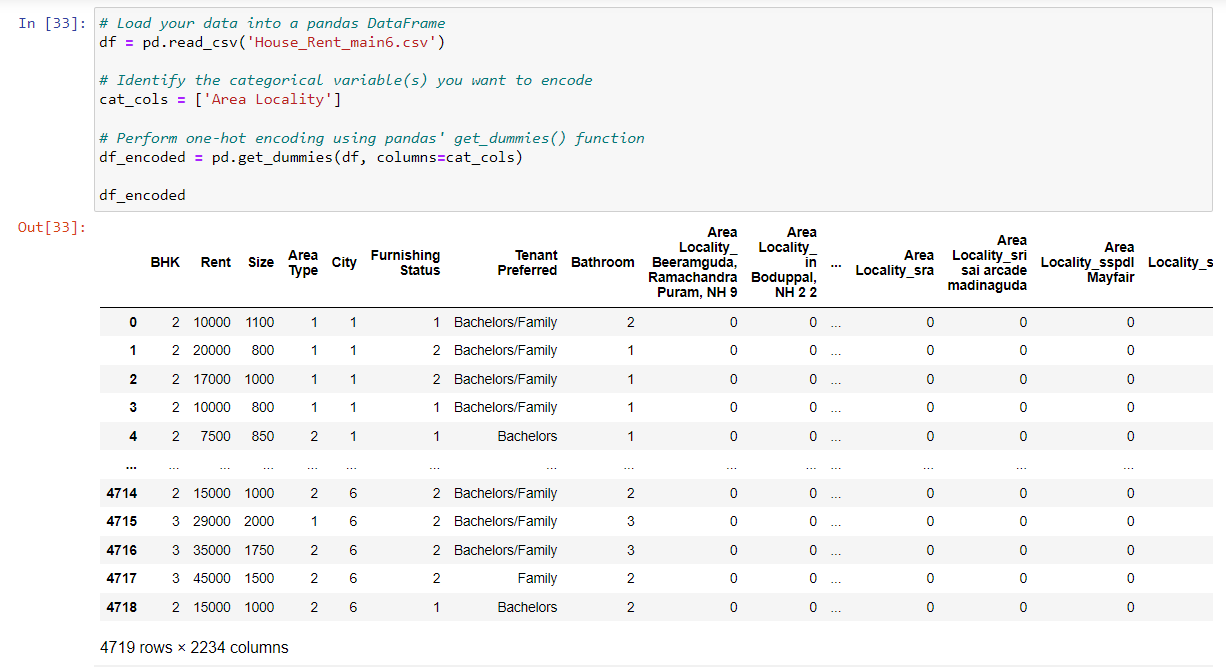
****

**7. One Hot Encoding for all String values in our Area Attribute:**

Before:

****

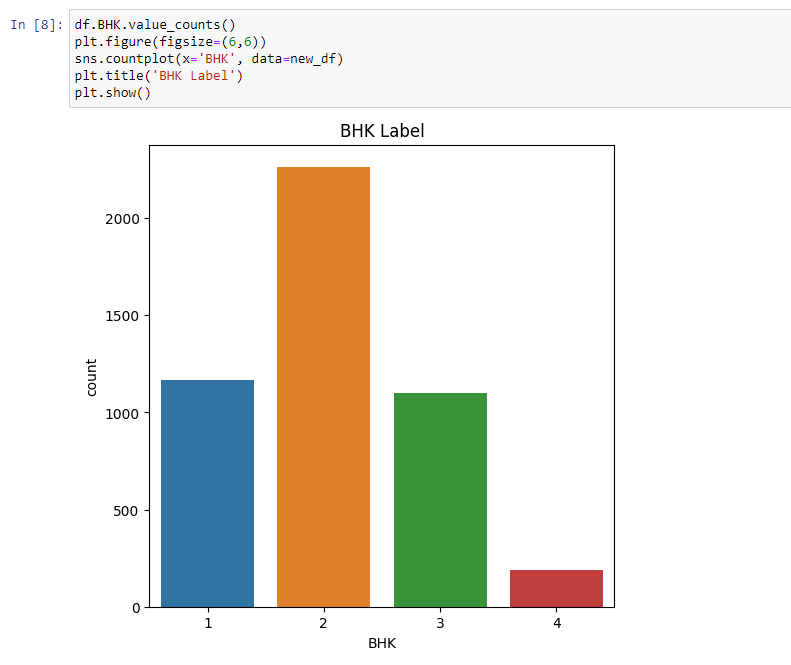
After:



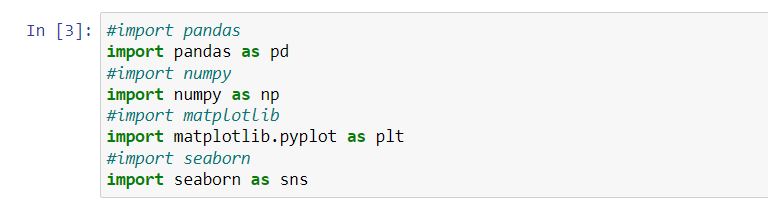
**8. Applying Classification Modelling:**

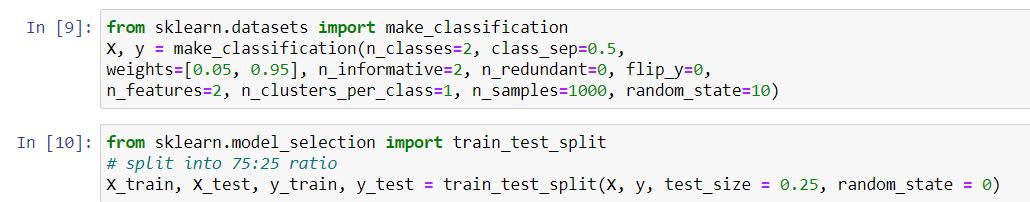
We will use a variety of classification methods to determine which algorithms are most accurate and suitable for our dataset.

There are 4 class labels in our dataset.

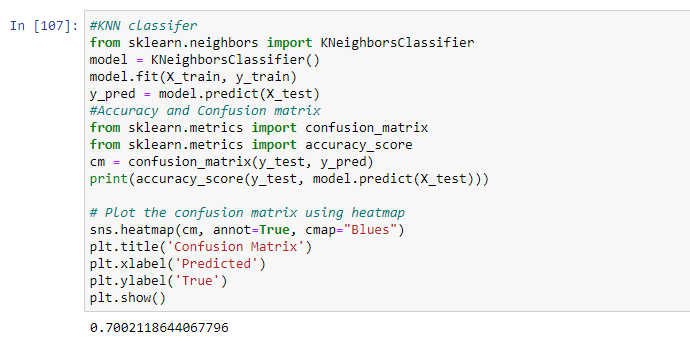


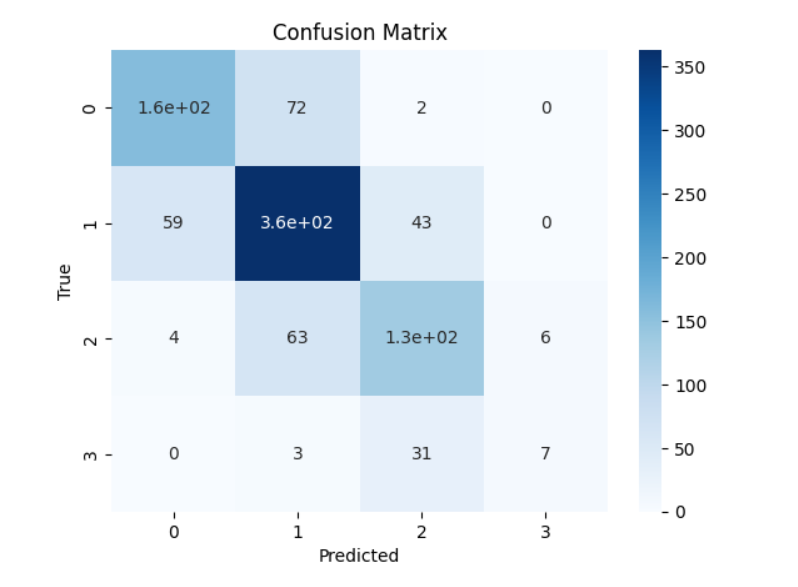
Splitting into training and testing dataset



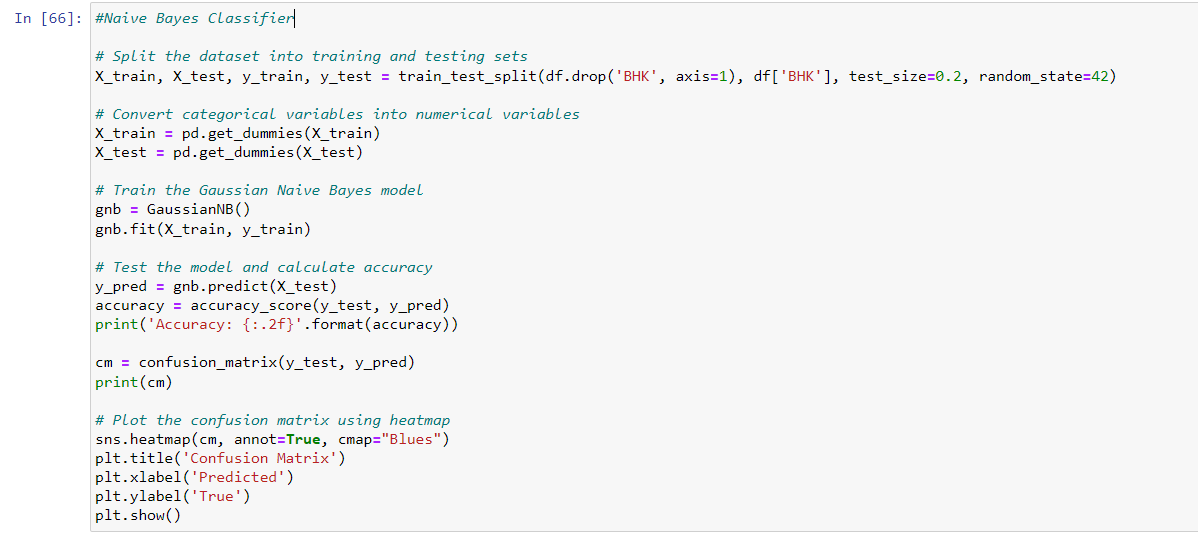


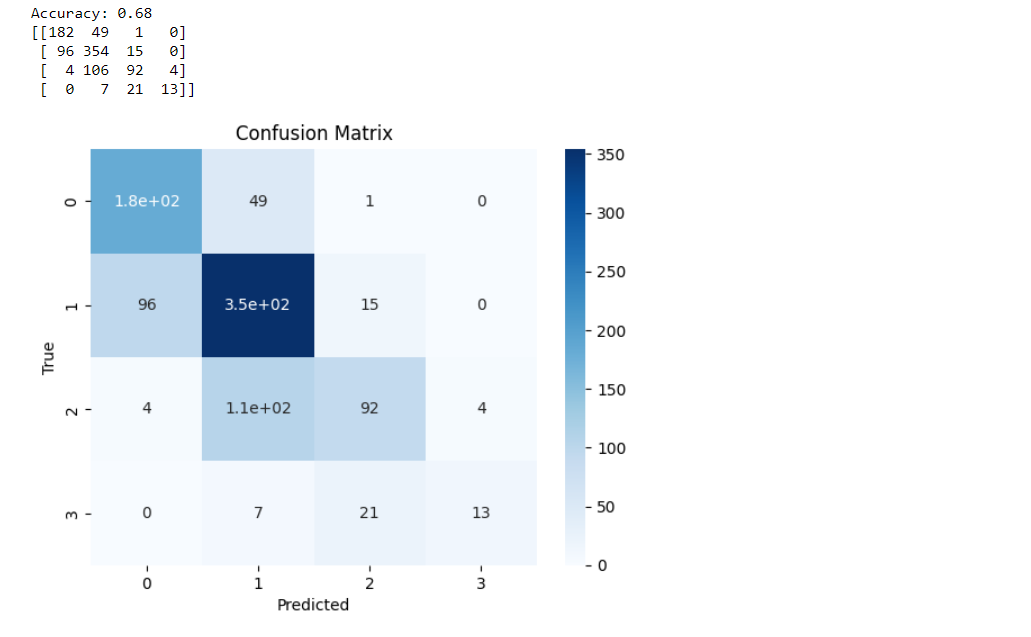
**8.a) Applying KNN Classifier Algorithm:**

****

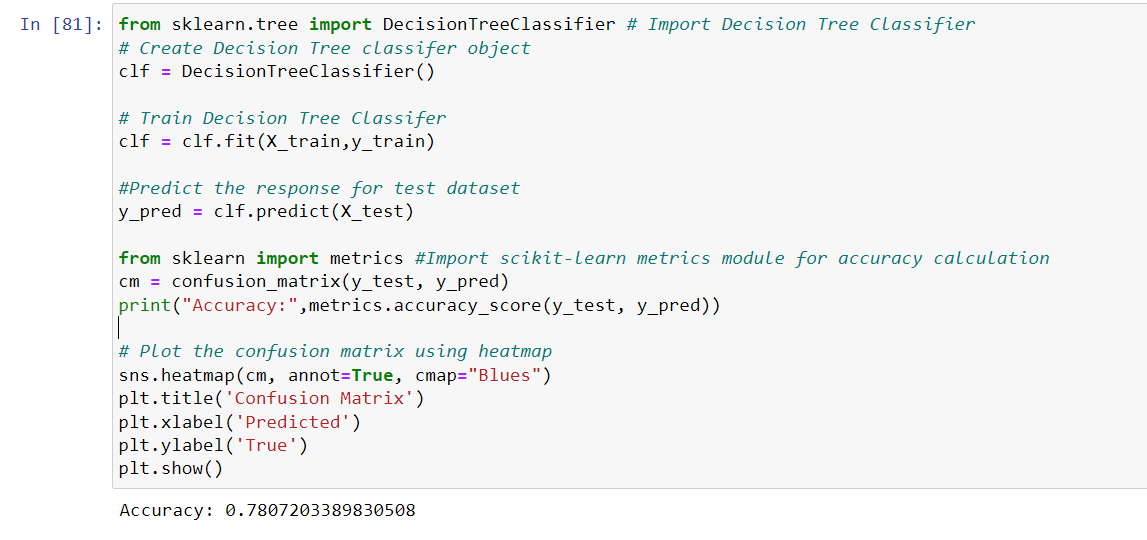
****

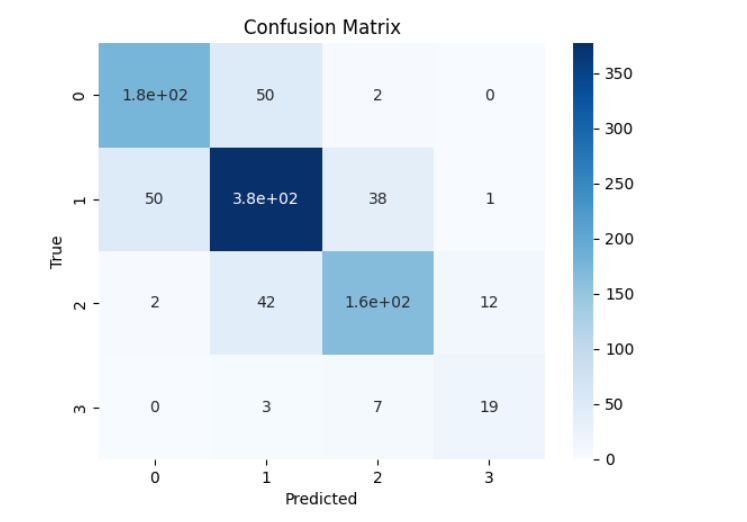
**8.b) Applying Naive Bayes Classifier Algorithm:**



****

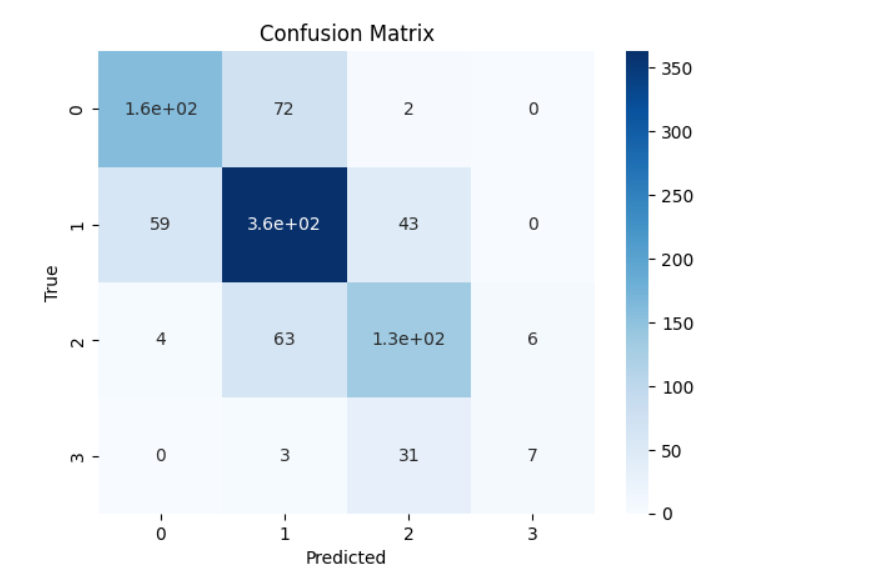
**8.c) Applying Decision Tree Classifier Algorithm:**



****

**8.d) Applying Random Forest Classifier Algorithm:**





**Conclusion:**

With the above experiments we can conclude that Random Forest Classifier gives us the most accuracy.

Confusion matrix:

True Positive (TP) = 160

True Negative (TN) = 7

False Positive (FP) = 74

False Negative (FN) = 63

Accuracy Rate: 0.83 or 83%

In conclusion, we have performed classification operations on the house rent dataset using various algorithms such as KNN Classifier, Decision Tree, Random Forest and Naive Bayes Classifier. We also evaluated the performance of these algorithms using metrics such as accuracy and confusion Matrix.

Based on our experiments, we found that the **Random Forest Classifier** performed the best with an accuracy of around 83%, which is a reasonably good performance given the complexity and variability of the data. The other algorithms also performed reasonably well, with accuracy ranging from 68% to 78%. Overall, the results suggest that machine learning algorithms can be useful in predicting the house rent prices based on various features such as location, area, number of rooms, etc. However, there is still room for improvement, and more sophisticated techniques can be explored to improve the accuracy and performance of the models.