Student Name: Pendurthi hemanandini

Student ID: 23106112

Git Hub: https://github.com/hemanandinipendurthi/Student-Name-Pendurthi-hemanandini-.git

Introduction

This paper aims at discovering how K-Means clustering and linear regression can be used in the analysis of a housing dataset. To analyze the abstract data the K-Means clustering algorithm is used to group properties based on their similarity, and linear regression checks the parameters of property area and price. In this report, calculations involve Mean Squared Error (MSE) and the coefficient of determination, R-squared (R^2).

Visualizations

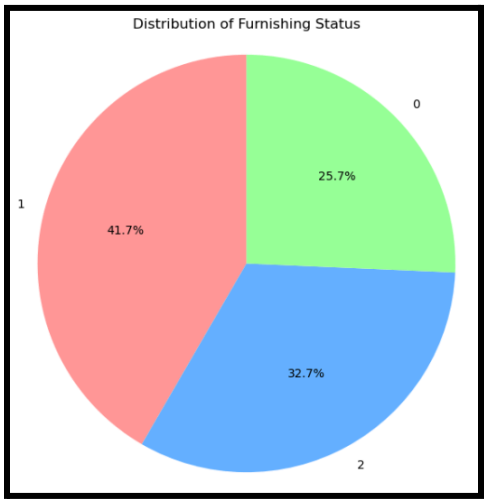

Figure 1: Distribution of Furnishing Status

This pie chart shows three categories of furnishing status to the data set namely; furnished, semi-furnished and unfurnished. The largest percentage goes to semi furnished (41.7%) then furnished (32.7%) while the least is the unfurnished ones (25.7%). This visualization depicts the primary furnishing state, to help analysis of the buying trends among customers.

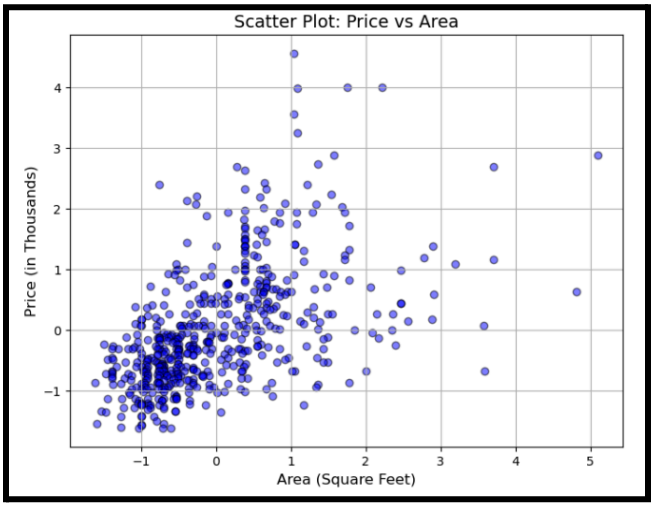

Figure 2: Price vs Area Scatterplot

The scatterplot above represents the property area and the price. The general trend is that larger properties cost more; that is, the correlation is positive. However, much variation with fluctuations ranging from 2.5 % to 23.5 % is observed, especially in mid-sized regions, implying other drivers for pricing. The plot is helpful in interpreting the outcome in terms of derivatives market of price volatilities and distributions.

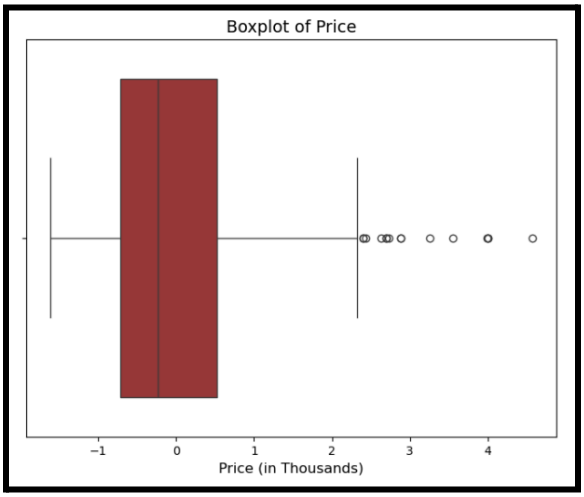

Figure 3: Boxplot of Price

In boxplot below, it exhibits the median, the interquartile range and any outliner of property prices. As with most distributions, the great majority of the prices fall within a given bandwidth while a few are high-value properties which are considered as outliners. This visualization highlights issues with pricing, allowing for subsequent investigation of the high-end property segments.

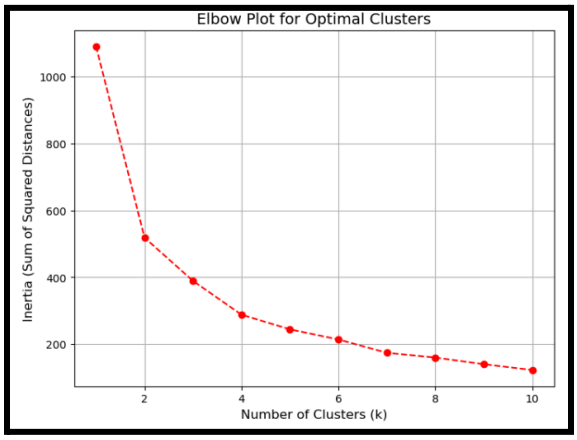

Figure 4: Optimal Customers

The elbow plot is used to assess the number of clusters from which the K-Means clustering algorithm should be used, based on inertia, or the sum of distances squared. “The ‘elbow’ is found at k = 3, suggesting that three clusters provide an adequate trade-off between both compactness and simplicity.

K-Means Clustering

```
Mean Squared Error (MSE): 2488861398180.66
R-squared (R²): 0.29
```

Figure 5: Clustering Results

This figure shows the results of clustering with K-Means; MSE values for applying regression analysis are also given. From equations (7) and (8), the increasing order of R^2 (0.29) and MSE (2488861398180.66) show the fitness of clustering and regression. These measures help in the evaluation of the grasped model precision, as well as the appropriateness of the segmentation for property price prediction.

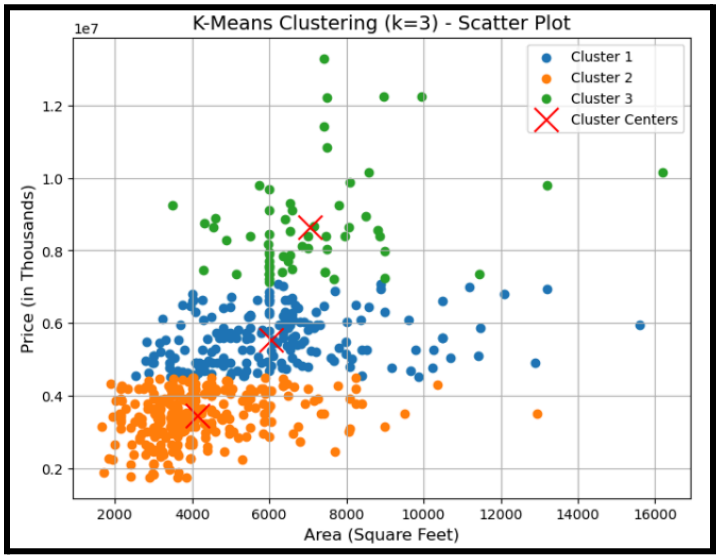

Figure 6: K-Means Clustering

This scatterplot shows K-Means clustering model using three clusters. Each cluster has its color and the centroids are marked with a red "X", with data collected by price and area. Clustering entails matching homeowner groups because it helps in customized selling or setting of fees and other charges.

Conclusion

Having established this, the study shows how K-Means clustering and linear regression work in segmenting out properties and in molding and portraying price associations. Through clustering, there are three types of property distinguished, additionally, regression offered a positive linear regression between area and price. According to the assessment of the simple and multiple regression models, modified MSE ad R^2 enabled to determine the models' accuracy in making decisions.