3) Narendra Chigili

We have chosen to classify mode price as a feature for milestone 3 based on the neighborhood group. Performing arbitrary analysis

As you can see from the picture's highlighted rows, those Airbnbs were all from the same neighborhood and three of them had the same kind of room.

How can we defend the optimal price for the rooms with higher prices if we classify the mode price? Except for the latitude and longitude, which had a minimal impact, our dataset did not contain any features that could explain it. Learned that we need to include more variables to justify the price predictions and helped to include that in PCA.

Used GridSearch to attempt to boost the efficiency of the K cluster.

Classification based on price was performed by dividing the price into two groups, less than $250 and greater than $250. two classes were classified using a new variable. When we used $500 as the marginal value for both classes, there was an unequal distribution of wealth between the two. Compared to numbers below 500, those over 500 are significantly fewer. There were significantly fewer values outside of 500 in the above 500 class than there were within 500. Values above 500 were eliminated because they are fewer in number to address the class imbalance.

The results are entirely accurate, according to classification metrics. which suggests the model is not a good fit.

Envisioned the test's accuracy with more complex models, training accuracy improves. Models that are too complex or not complex enough suffer from testing accuracy penalties. In KNN models, the value of K (lower value = more complex) determines the complexity.

Milestone 3 reflection -Ronald Adomako

From this project, using the AirBnB dataset, I learned a lot about using statistical behavior to understand and bend a dataset for understanding. I saw how the mode can behave as a stable statistical when the mean and median shift and do not match in a skewed or log-normal distribution. I saw how the statistics of a subset (room\_type: Entire home/ apt) of a dataset was representative of the aggregate statistics.  I saw how the scree plot of parsimonious-like model could have a constant slope. I saw how that was consistent with summary statistics of seeing nearly equally weighted variables for logistic regression. I learned how to simplify a data set into binary behavior and build off incremental insights and develop skills for exploratory, descriptive, prescriptive, and predictive analyses for data science.