

Title: Rental Housing Price in New York City

Name: Winsten Coellins

New York, one of the most popular cities in the United States has attracted many tourists from many countries in the world for decades because of its places of interest. Tourists have been spending their day and night wandering around New York to fully experience the atmosphere where they would never regret in their lives. Besides enjoying the atmosphere, tourists often decide regarding how much spending per day they will need to get the estimation of spending they need to. Thus, my motivation for this project is to help tourists to decide for their spending so that tourists can estimate their spending while enjoying their vacation.

The source of data is taken from New York City's leading real estate marketplace called StreetEasy - <https://github.com/Codecademy/datasets/tree/master/streeteasy>. Taking three datasets which are manhattan.csv, queens.csv, and brooklyn.csv. I then concatenate the data and merge them into one file called data.csv. Therefore, I have three borough of New York City which are Manhattan, Queens, and Brooklyn for analysis.

For the analysis, I selected 12 of the 18 features from the data. I selected the number of bedrooms, number of bathrooms, size square feet of the rental property, minutes to subway, number of floors, building age, fee, roof deck, washer dryer, doorman, elevator, dishwasher, patio, and gym.

Figure One shows the average rent of each borough in New York City. From the visualization, it shows that the most expensive borough in New York City is Manhattan averaging US\$5138.94, follow up by Brooklyn averaging US\$3327.40, and Queens averaging US\$2516.14. As an addition, from the analysis I have, the total rental housing properties provided in each borough are different with the statistics: Manhattan (70.78%), Brooklyn (20.26%), Queens (8.96%). Thus, the most rental housing properties provided are in Manhattan with high price, second is Brooklyn with a middle-class price, and Queens with the cheapest price in New York City.

Figure Two shows the Principal Components 1 and Principal Components 2. The method I used here is the Principal Components Analysis (PCA) to do dimensionality reduction from 15 features into two dimensions to enable the visualization. From the clusters, I have to say that Manhattan has a variety type of properties that has features that are almost similar to Queens and Brooklyn since the clusters distribution almost has no specific cluster or area at all for Manhattan. As an addition, the explained variance ratio from the Principal Components Analysis is 99.22% and 0.752%.

Figure Three shows the correlation of the actual price versus the predicted price from the model by using Multiple Linear Regression. The accuracy of the training model compared to testing set was 72.81% with the coefficients: -447.67 (number of bedroom), 1324.27 (number of bathrooms), 4.67 (size of the property in square feet), -18.57 (minutes to subway), 39.19 (number of floors), -4.3 (building age), -83.34 (fee), 66.9 (roof deck), 225.44 (washer dryer), -95.92 (doorman), 217.16 (elevator), 42.17 (dishwasher), -38.8 (patio), and 22.47 (gym). I then try to increase the accuracy of the model by using Standard Scaler and Polynomial Features, and the accuracy of the model goes up by 2%.

In conclusion, we have discovered that to visit New York City, one does not need to spend a huge amount of money to stay since New York is famous of its luxury. Moreover, from the analysis, we could use the Linear Regression model to estimate the spending that one will need to help manage the expenses. Overall, these models may be helpful for tourists to predict their expenses to enjoy the beauty of New York City.

Figure1: Average Rent of Each Borough in New York City

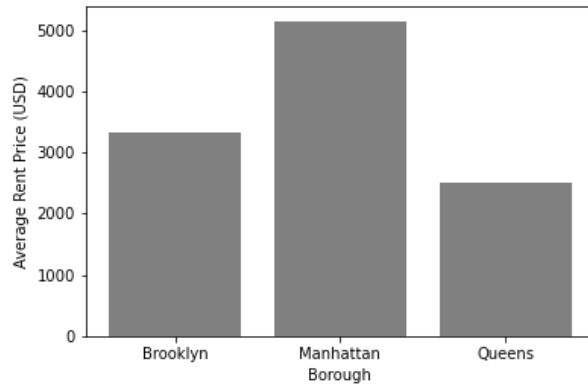


Figure2: Principal Components 1 and Principal Components 2

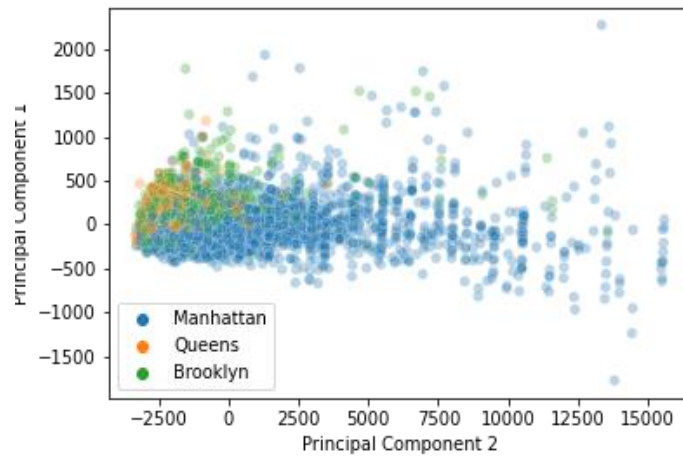


Figure3: Correlation of Actual Price vs Predicted Price (MLR)

