Final Report for NYC Database

Jing-Chian Teng

Boston University

AD 571Business Analytics Foundation

Ajit Appari

12.12.22

# Report summary ----------------------------------------------------------------------- 3

# Location -------------------- ------------------------------------------------------------- 4

# Price from the past----------------------------------------------------------------------4

# Forecasting -------------------------------------------------------------------------------5

**Optimize the profit ---------------------------------------------------------------------6**

**Table ---------------------------------------------------------------------------------------8**

Report Summary

In the project, I have to find the best place to open the office in 5 neighborhoods: Clinton Hill, Cobble Hill, Cobble Hill-West, Coney Island, and Crown Heights. To find the best place to open, I have to know the price from the past to understand the market price because there is some useful information that can let me judge if the location is suitable to open the office. After receiving the price from the past, I made the forecasting to predict the future price because I try to figure out the future price to open the office. If we can predict the future price, we can make sure we can have profit at the location in the future. In the forecasting part, I list some additional information to explain the forecast. In the end, I want to optimize the profit. There are a lot of factors to affect our profit, such as cost, commission, penetration rate, and the number of employees. I have to do my best to lower the cost and the number of employees. Also, I have to increase the commission and the penetration rate.

Location

In this project, I chose 5 neighborhoods in Brooklyn: Clinton Hill, Cobble Hill, Cobble Hill-West, Coney Island, and Crown Heights. In these five neighborhoods, I have to choose which neighborhood is the best place to open the office. To determine the best location to open, I have to obtain numerous data such as the sales number, the mean price, the forecast, the average price per square foot, and so on.

Price from the past

First, I have to know the price from the past because we can learn some information from the past to prepare for the future.



In the ggplot, we can see the movement of each neighborhood. From starting point to the ending point, the price in each neighborhood is going up. We can analyze the detail from the ggplot. From 2004 to about 2007, the price in each neighborhood goes up, but the price from about 2007 to 2010, the price goes down a lot, and I think there is a huge change in the market during that period. The price of each sqft in each neighborhood started to leap from 2010 until about 2016. However, the price drops from about 2016 to about 2018. There is a cycle in these 5 neighborhoods. In conclusion, the price will go up for a few years, and then goes down for another few years, but the trend is still going up.

Forecasting

After we have the price from the past, we want to predict the future price because we want to make sure we can have profit in the future. I chose 2 neighborhoods from the 5 neighborhoods based on the mean price and the sold numbers, and the 2 neighborhoods are Coney Island and Crown Heights, the low-volume, and high-volume neighborhoods. In these 2 neighborhoods, I made a model to forecast the future price to see the difference.

Crown Heights:



Coney Island:



From the prediction graphics, I try to predict the sales for the next 8 quarters. We can notice that the prediction sales for the next 8 quarters are going up and down for Crown Heights during the period, but the result may be the same. And the predicted sales in the next 8 quarters for Coney Island might be the same.

During the predicting part, the most useful predictor is gross square feet for both neighborhoods. It has the highest estimate coefficients and p-value less than 0.05. The least useful predictors are building type for both neighborhoods because most of the building type predictor has the p-value greater than 0.05, which means we need to consider remove it from our regression.

For both neighborhoods, I think the year built and building type seem like redundant. Because both independent variables has high p-value. To figure out, I remove both variables from my model and comparing the result with previous. The bellowed figures show after removing, all the rest independent variable are statistically significant and high correlated. Therefore, the year built and building type are redundant independent variables.

Properties with more units will be cheaper, and properties with higher gross square feet will be overpriced. I think the result makes sense, because with more units in a property, each person will be allocated relatively fewer resources and the usual living environment may be worse, such as noise impact and so on

Optimize the profit

In the final part, I made an optimization model to determine the best course of action for your real estate company. To find out the maximum profit for the next 8 quarters, I have to follow the constraints of the budget, the monthly budget is $15000, the value of the employees, the maximum is 3, the value of space, the minimum is 250, and the penetration, it should be between 4% and 6%. My office is opened in Crown Heights, Brooklyn because from the previous result, Crown Heights has the highest sales and the mean price, and it means the profit is the highest. The average rent per square foot in commercial space is $66.

To increase the profit, we have to optimize the penetration, commission, rent, and number of employees. There are a lot of ways to increase the penetration rate such as decreasing the commission rate and increasing the number of employees. However, when we decrease the commission rate, we will decrease the revenue because the commission rate means the money we can get from the sale. If we want to increase the penetration by increasing the number of employees, we will increase the salary and the rent because we will need more space for the employees and hire more employees to increase the penetration rate, but the profit will decrease. To find out how to maximize the profit, we use Excel, and we input the forecast of the next 8 quarters to predict the profit of the next 8 quarters and follow the constraints. In Excel, we are focusing on changing the number of employees and the commission rate because they can affect the penetration rate. In the output, the commission rate should be 5.23%, and the number of employees should be 2.69 or around about 3 so we can maximize the net profit to $ 3,971,513.69 in the neighborhood.

Table

The forecast of Crown Heights:

Text

Description automatically generated

Crown Heights:

Text

Description automatically generated

Text

Description automatically generated

The forecast of Coney Island:

Text

Description automatically generated

Coney Island:

Text

Description automatically generated

Text

Description automatically generated with medium confidence