**NYC Realty Estate Analysis: Baychester District**

by

**Hsin Fang Hu**

5th Oct 2022

**Table of contents**

Summary........................................................................................................................2

Object.............................................................................................................................2

Data Preparation.............................................................................................................3

Exploratory Data Analysis.............................................................................................5

Predictive Analytics.......................................................................................................8

Data Modeling and Optimization.................................................................................11

Conclusion....................................................................................................................15

**Summary**

This analysis report aims to determine the feasibility of establishing a real estate firm in the Baychester District of New York. To accomplish this, I utilized NYC Real Estate SQL data, along with tools such as R Studio, Power BI, and Advanced Excel. The report follows a comprehensive approach, beginning with descriptive analytics for data exploration, followed by predictive analytics using Time Series and Regression forecast models to predict future sales performance. Finally, prescriptive analytics utilizing Excel Solver is employed to assess the profitability of opening an office.

Historical data shows residential sales in Baychester significantly surpass commercial sales: $120.7 million versus $29.76 million. Using a regression forecasting model, we projected residential sales for the next eight quarters (2022-2023), expecting approximately $1,093,024 in revenue. However, considering costs (initial investment, office space, labor, marketing), our estimated net present value of profit is around $430,000, with a 65% return on investment. Opening a Baychester office, focusing on residential sales, appears favorable. Practical aspects include limited manpower with only three employees and the need for in-depth risk analysis before making a final decision.

**Objectives**

* Determine what type of realty to open

Choose between residential, commercial, mixed-use, or a combination of realty services based on the assigned neighborhood.

* Assess the market

Justify the business's capital requirements to lenders by highlighting the demand for services, expected growth, and the potential for profitable operations with satisfactory revenue levels.

* Strategic location

Minimize expenses and maximize efficiency by avoiding heavy competition and reducing travel time for agents. This ensures higher productivity, minimizes opportunity costs, prevents employee fatigue, saves on travel expenses, increases the frequency of closed sales, and establishes expertise in the chosen locale.

**Data Preparation**

**Data Overview**

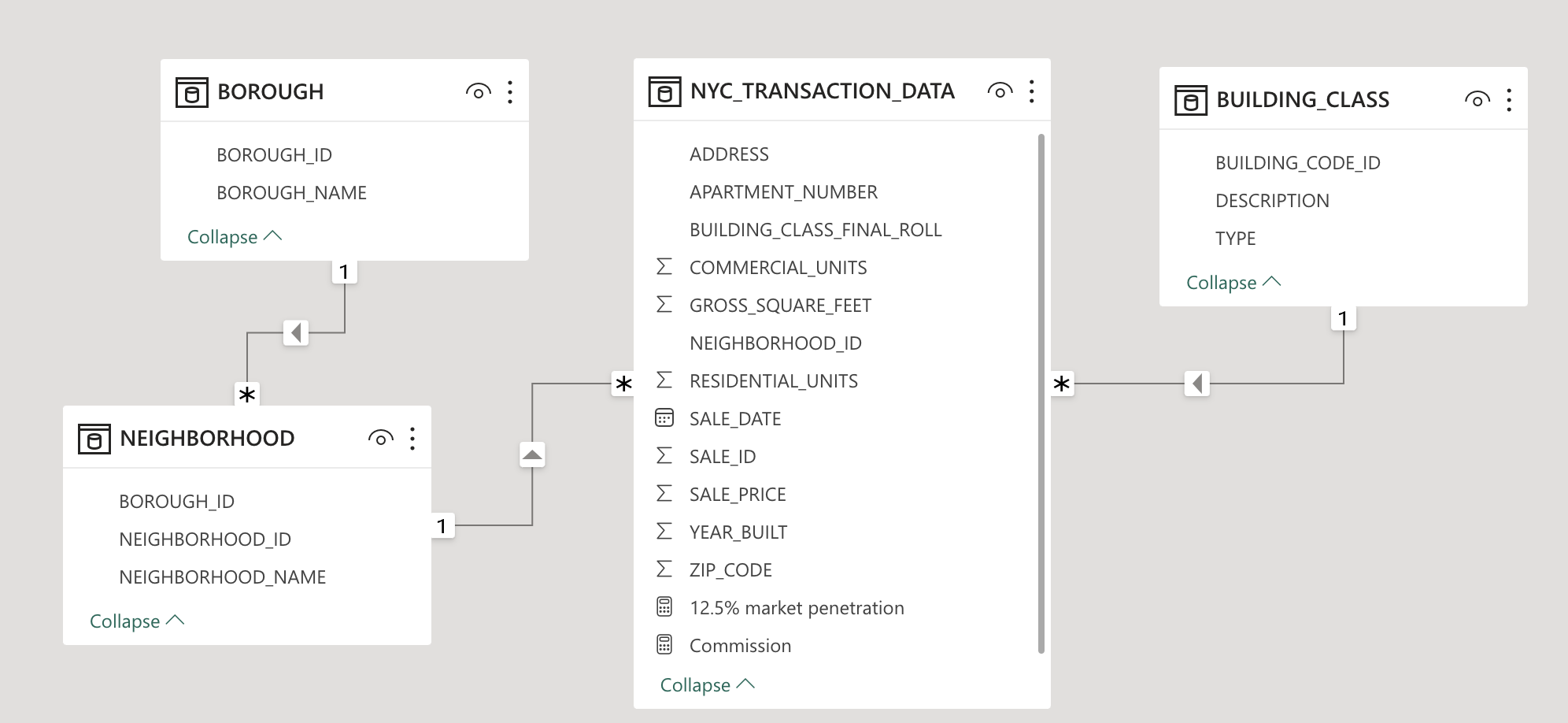
The NYC Real Estate Dataset was from the Boston University MET Lab SQL Server database. We will use 4 entities for analysis: Borough, Neighborhood, NYC transactions, Building classification. It contains information such as building number, neighborhood area, year of construction, address, date of sale (2009 to 2021), gross profit per square meter, etc.

**ETL**

ETL stands for Extract, Transform, and Load. It involves extracting data from various sources, such as databases, applications, and external systems. The extracted data is then transformed and standardized to conform to a common data model or schema, ensuring consistency and compatibility across different data sets. Finally, the transformed data is loaded into a target database or data warehouse for analysis, reporting, and decision-making purposes.

The datasets are in the Boston University MET Lab SQL Server, so I first establish a connection to this "NYC Real Estate" database, hosted at "met-sql19.bu.edu". Then I extract data from 4 tables. After that, the transformed data is loaded into a new data frame by using left join functions.

**ERD**



BOROUGH

Primary Key: BOROUGH\_ID

NEIGHBORHOOD

Primary Key: NEIGHBORHOOD\_ID

Foreign key: BOROUGH\_ID

NYC\_TRANSACTION\_DATA

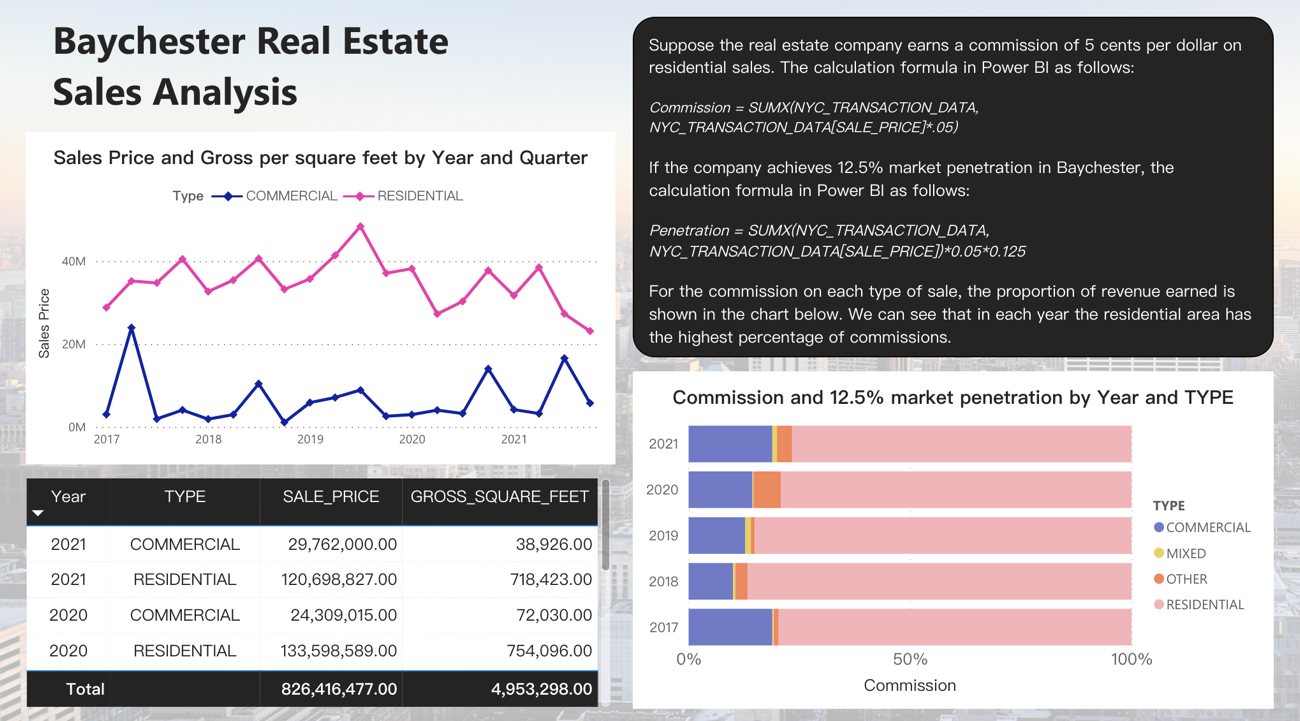
Foreign key: NEIGHBORHOOD\_ID, BUILDING\_CODE\_ID

BUILDING\_CLASS

Primary Key: BUILDING\_CODE\_ID

**Exploratory Data Analysis**

**Data Exploration: Power BI**



*Residential and Commercial Sales of Baychester*

In the last year 2021, Baychester's residential sales totaled a whopping $120.7 million. Commercial sales were only $29.76 million compared to residential sales. Therefore, residential type should be targeted to open in this area.

*Commission of Baychester*

Suppose the real estate company earns a commission of 5 cents per dollar on residential sales; then the total revenue earned in Baychester over the last year is $6.03 million for our company. The calculation formula in Power BI as follows:

*Commission = SUMX(NYC\_TRANSACTION\_DATA, NYC\_TRANSACTION\_DATA[SALE\_PRICE]\*.05)*

If the company achieves 12.5% market penetration in Baychester, the revenue would be $754.37 thousand. And the calculation formula in Power BI as follows:

*Penetration = SUMX(NYC\_TRANSACTION\_DATA, NYC\_TRANSACTION\_DATA[SALE\_PRICE])\*0.05\*0.125*

**Data Exploration: R**

*Cleaning Data - Remove 0*

In this step, I added the description that the sale price and sum of all areas must be greater than zero in the filter:

*filter(NEIGHBORHOOD\_NAME=="BAYCHESTER", TYPE=="RESIDENTIAL", GROSS\_SQUARE\_FEET>0, SALE\_PRICE>0)*

|  |  |
| --- | --- |
|  |  |

From the table above, it can be seen that I accurately calculated the average price of one square foot of real estate and the total sale price.

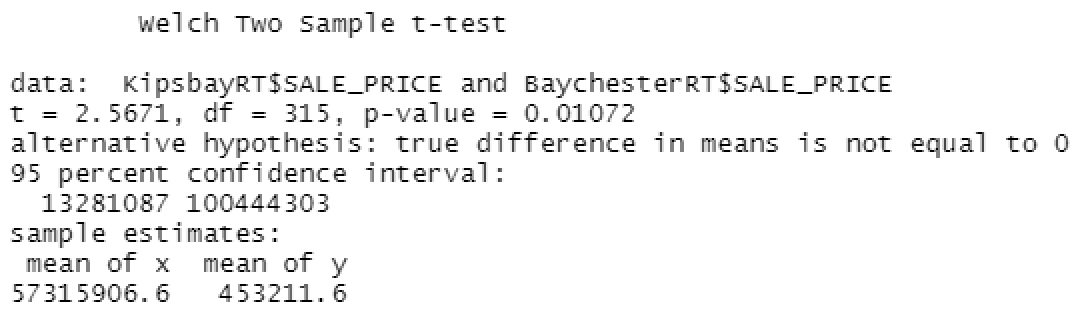
*Compare Baychester with Williams Bridge and Kips Bay*

I picked two nearby neighborhoods. In terms of location, Williams Bridge is right next to Baychester. In terms of total sales price, Kips Bay was the top five sales hotspot. I'd like to see what sales are near Baychester; also compare the top sales areas, which is why I chose these two locations in particular.



It can be seen from the chart that the average price of one square foot in Baychester and Williams Bridge is similar. It can be explained that the price level in this area may be the case. And Kips Bay's average price of one square foot can be clearly seen as much higher than Baychester and Williams Bridge's. This comparison chart can give supervisors a good basis for decision-making.

*Compare Baychester with Kips Bay (T-test)*

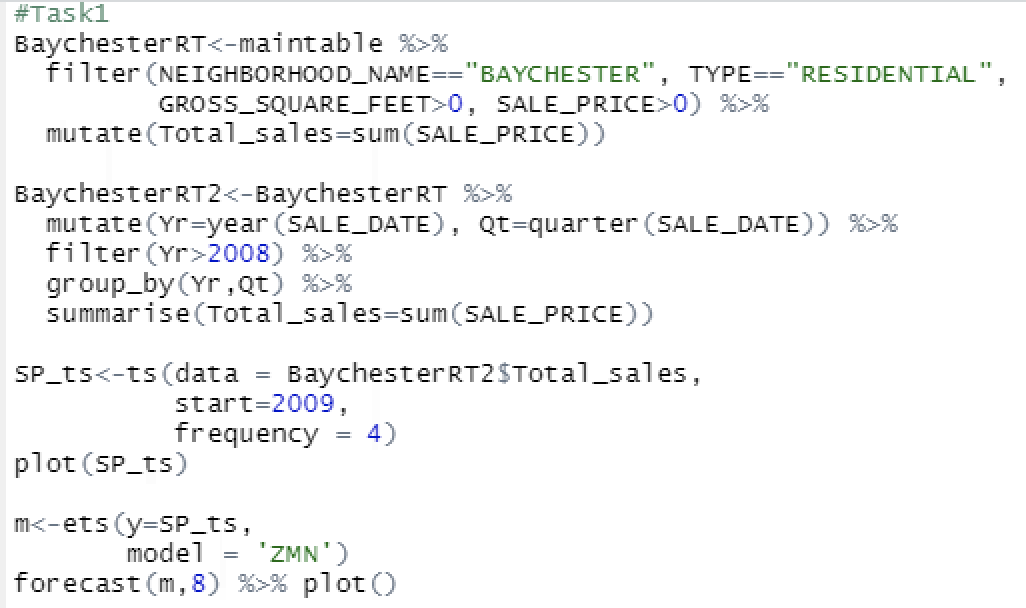


In T-test, I use "t" to run the code. We can see that Kips Bay's sales price and Bay Chester's sales price have different level. From the mean of these two area, we know that Kips Bay's sales price is much higher than Bay Chester. We understand that still have another neighborhood's sale price is higher, so it seems that Bay Chester is not the best place to open a real estate office.

**Predictive Analytics**

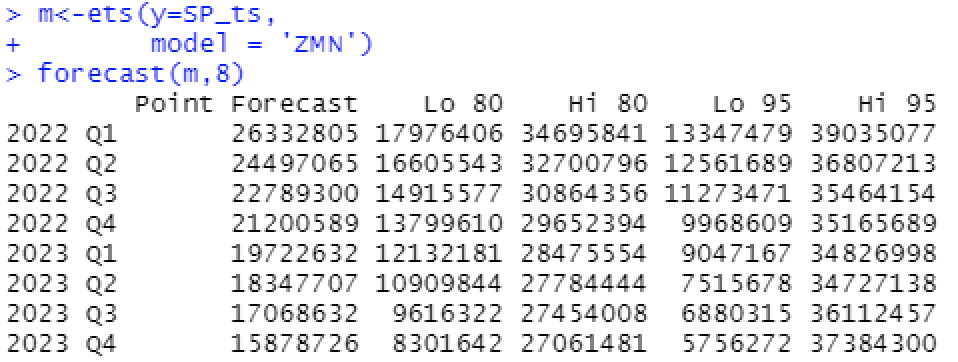
**Time Series**

At the beginning, I made a filter according to the requirements of the title, and chose to start from 2009. After making sales line chart, I find that the sales volume presents multiplicative trend, which is slightly curved, and none seasonality. Therefore, I put "M" for trend and "N" for seasonality to analyze. For the error type, I use "Z" to let the R choose.



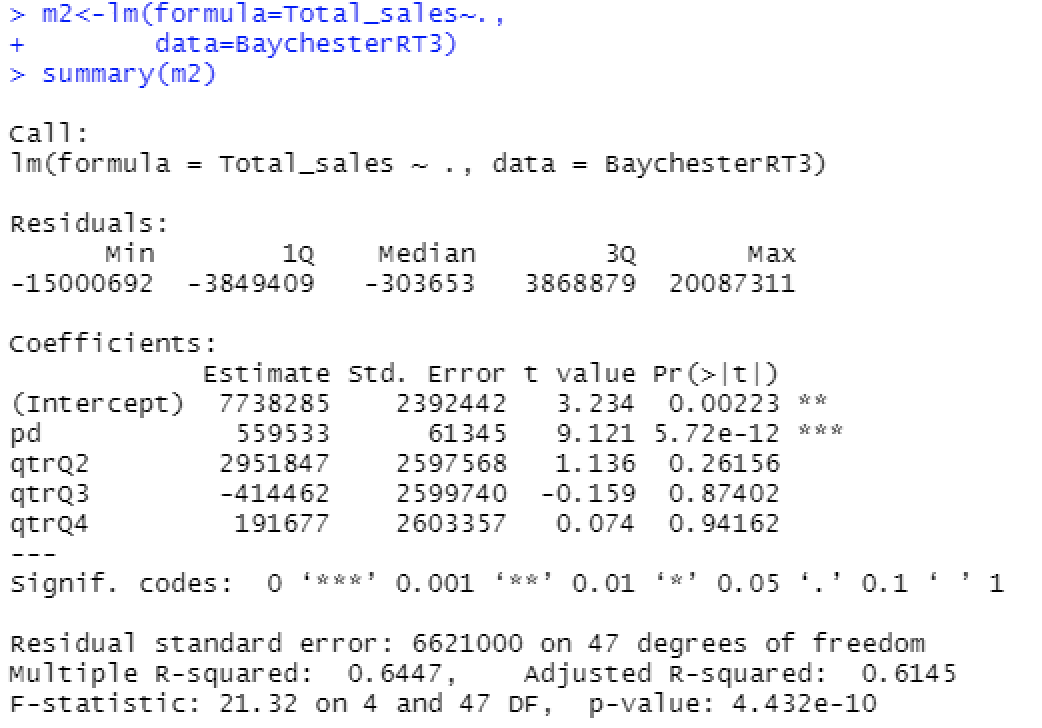
The numbers predicted by the time series can be referred to the following table, which shows the forecast numbers, and confidence bands for the next 8 quarters. It can be seen that the sales number predicted by the time series is downward. This may be because historically, residential sales in 2020 began to decline due to the impact of the epidemic and it reflected in the forecast.





**Regression Forecast**

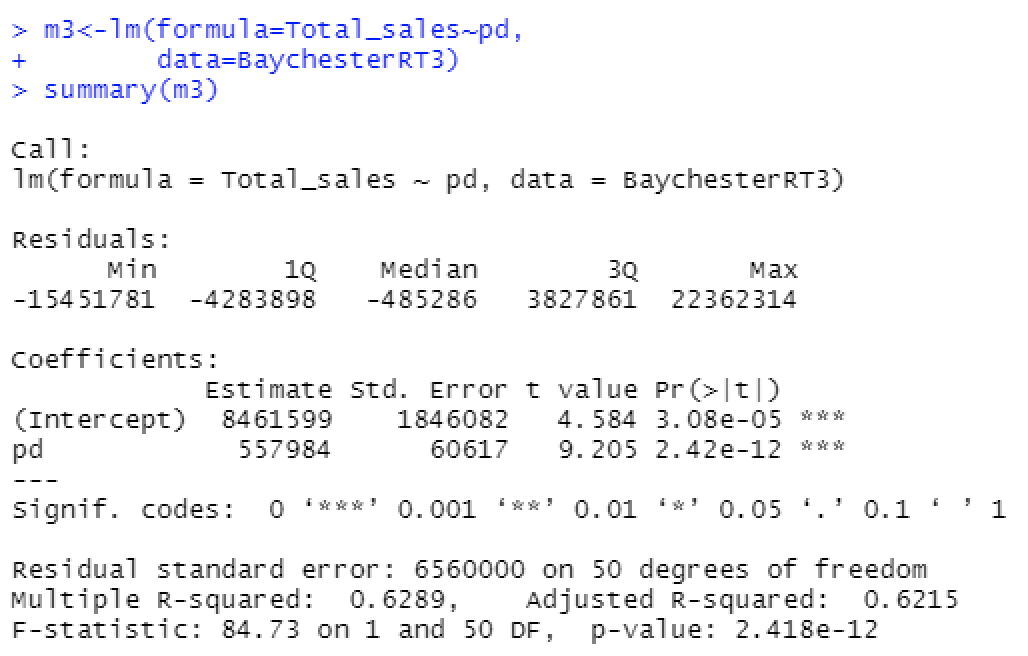
*Including Both Time and Seasonality*



The model equation is:

As we can see at the significance row of "pd", time variable is very important, which is really small (5.72e-12). However, the quarter might not be so important. Because non of its significance row is smaller than 0.05. The multiple R squared is 0.64 and adjusted R squared is 0.61. Adjusted R squared tells us whether the model does a good job of explaining the variance in comparison to the total variance. And a R squared value of 0.64 means that 64% of the variance can be explained or predicted using the predictors we see as coefficients.

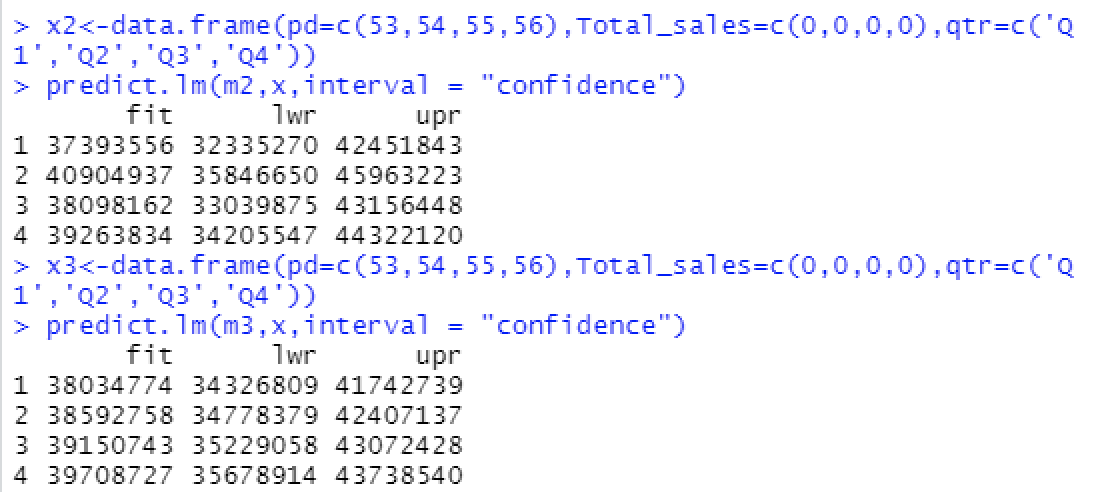
*Including Time Only*



The model equation is:

After deleting the quarter, now we can see the significance row of "pd" become smaller than before, and we also have a slightly better adjusted R squared which is 0.62.

*Regression Forecast*



As we can see in the table, both of this two regression models think that sales are going up, while when we go back to time series analysis, it thinks that sales were going down. So the reason for that is because the sales started by going up, and then ended by going down. The regression model doesn't weigh newer data entries more heavily than older, whereas the time series model does.

**Data Modeling and Optimization**

**Market Conditions and Background in Baychester**

Before setting up the optimization model, we should first predict the future sales of residential properties in Baychester. According to the previous regression forecast, I developed two models. One is only including time as a predictor. The other includes both time and seasonality. The adjusted R squared of former is higher than the latter, 0.62 and 0.61 respectively. Therefore, we should choose the forecast outcome from the one only including time.

With the prediction of 8 quarters' residential sales, we can start to set our optimization model. First, we wish to maximize the net present value of profit from 2022 to 2023 and we assume the annual discount rate is 6%. After researching the market in the Baychester area, we assumed that we would reach 5.5% market penetration and earn a 5% commission on sales after opening. This is Baychester's market conditions and background for the next two years.

**Total Cost of Opening**

If we open in Baychester, the total cost can be divided into four parts: initial investment, space cost, staff salary and marketing cost. After we calculate the total cost, we can subtract it from the total revenue forecast above, and then we can know the profit from 2022 to 2023. In order to control our expenses, our monthly operating budget is $25,000.

1. Initial Investment

Before opening, we must first maintain a fund. These funds can be used to pay licenses, insurance, and professional fees, about 2% of total revenue. We’ll also need to buy some furniture, computers, Wi-Fi, and other utilities, which might be 2%. The most important part is that we have to build a CRM system to make our future sales more comprehensive, and the cost will be 4% of our total revenue.

1. Space Cost

In terms of rent, we use the average cost per gross square foot of commercial real estate in Baychester for the last year to estimate. The average price in 2021 is 373.73 per gross square foot. Each month we will pay 1.5% of this for rent and electricity for each square foot of office space we rent. We must rent at least 250 square feet of office space, plus 125 for each employee we hire.

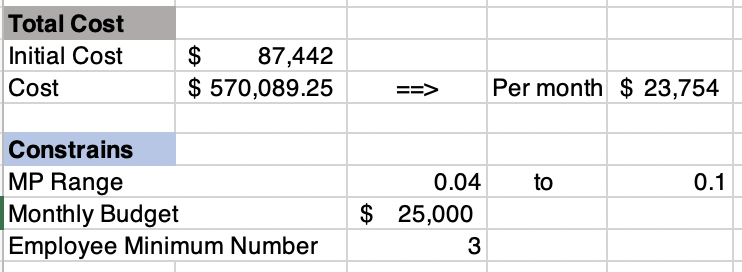
1. Labor Cost

In order for the office to operate normally, we must employ at least 3 full-time employees. Each employee's salary is $70,000 yearly and each will increase our market penetration by 0.5%.

1. Marketing Cost

We may host a grand opening event to attract customer attention, which costs $6,000. After that, we are going to have advertisements on Facebook, Instagram, or Google. At the same time, we also should create longer-lasting digital and printed marketing materials like postcards, brochures, or flyers and which will cost $30,000 yearly. For every $10,000 spent on marketing, we will increase our market penetration rate by 0.1%.

|  |  |
| --- | --- |
|  |  |



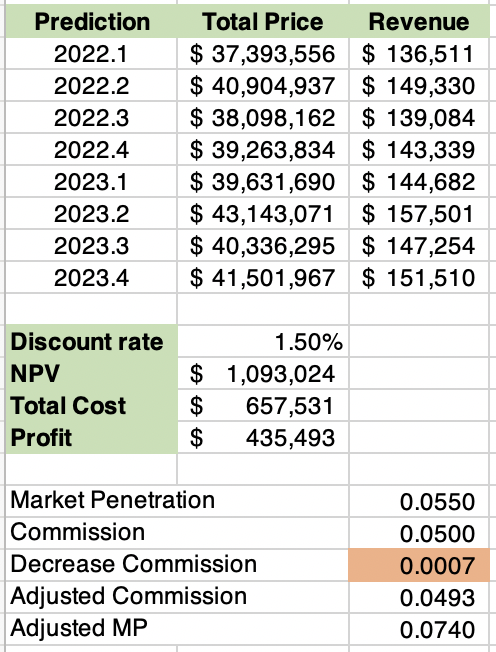
**Optimization Model**

With the previous data and cost calculations, we can build an optimization model now. First, for every 0.1% the commission we charge decreases, our market penetration will increase by 0.15% (down to a minimum penetration of 4% or a maximum of 10%). Second, hiring one employee can increase our market penetration by 0.5%. Lastly, every $10,000 spent also increases our market penetration by 0.1%. According to the information, we can come up with the formula:

*Adjusted Market Penetration (MP) = 5.5% + Increased MP% due to reduced commission + Increased MP% by hiring employees + Increased MP% through marketing spending*

With this formula, we can use the Solver in Excel to figure out the commission ratio we should reduce and the number of employees we should hire to maximize our profits. In addition, we also need to input the constraints of real conditions.

In the end, we figured out that we would generate the greatest profit when reducing the commission by 0.07% and employing 3 employees. The adjusted market penetration rate will be 7.4% and adjusted commission rate will be low to 4.93%. According to this, we use the net present value function to calculate our revenue is $1,093,024 from 2022 to 2023, and the total cost is $657,531. Subtracting the two values yields a profit of $435,493.



**Model Monitoring and Improvement**

We can compare the actual sales figures in 2022 with the forecasted values, and first examine the accuracy of the forecasted values. Then check the rationality of the market penetration rate and commission ratio, and fine-tune the ratio. With the previous monitoring data, we can improve the prediction model and optimization model. Making it feasible and profitable to open an office in Baychester.

**Conclusion**

According to the data, we conduct optimization model analysis and calculate that we would have earned a total of approximately $1,093,024 in revenue over the 8 quarters. Taking into account the various factors such as the initial investment, office space, labor, and marketing expenses, the estimated total cost sums up to $657,531. In this case, our net present value of profit will be around $430,000. The return on investment rate is 65%. Based on regression predictors and optimization model analysis, opening an office in Baychester and targeting residential sales seems to be a good investment.

**Choose Baychester?**

However, we must also carefully consider the practical aspects of our decision. With a limited workforce consisting of only three employees, there is a potential risk of experiencing manpower shortages and operational challenges. Additionally, it is essential to conduct a detailed analysis that encompasses potential risks and factors like the ever-changing market environment, including the influence of events like the Covid-19 pandemic. Only by carefully evaluating these aspects can we determine whether opening a real estate office in Baychester is the right move.

**Other Potential Neighborhoods**

After conducting a thorough exploratory data analysis, we have discovered that Baychester may not be the most lucrative neighborhood for opening our real estate offices. In comparison to Baychester, our analysis has revealed that neighborhoods like Kips Bay exhibit significantly higher average prices per square foot. These findings suggest that there are other areas with greater potential where we should consider establishing our new office. By focusing our efforts on these promising neighborhoods, we might position ourselves for greater success in the real estate market.