New York Real Estate Case Study

By Jiaxing Lin

AD571 A1: Business Analytics Foundations

Prof. Hyunuk Kim

Table of contents

01 Business Problem overview

Problem statement

Given the dynamic nature of the real estate industry in New York City, the company faced with the ongoing decision of whether to establish a new office in NYC, particularly in the **HARLEM-CENTRAL** neighborhood. This decision arises from the changing landscape and conditions based on *transactions, sales unit, building types, regions and sqft* within the real estate market from 2000 to 2021.

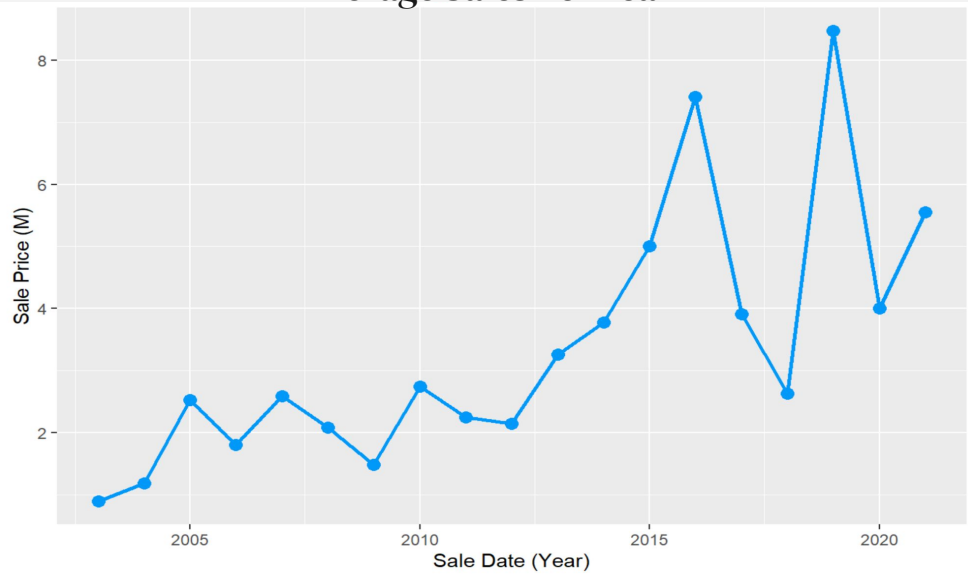
Neighborhood Overview

02

The Average Sales in HARLEM-CENTRAL

The general pattern of sales from 2000 to 2021 shows an **upward trend,** despite experiencing a few variations.

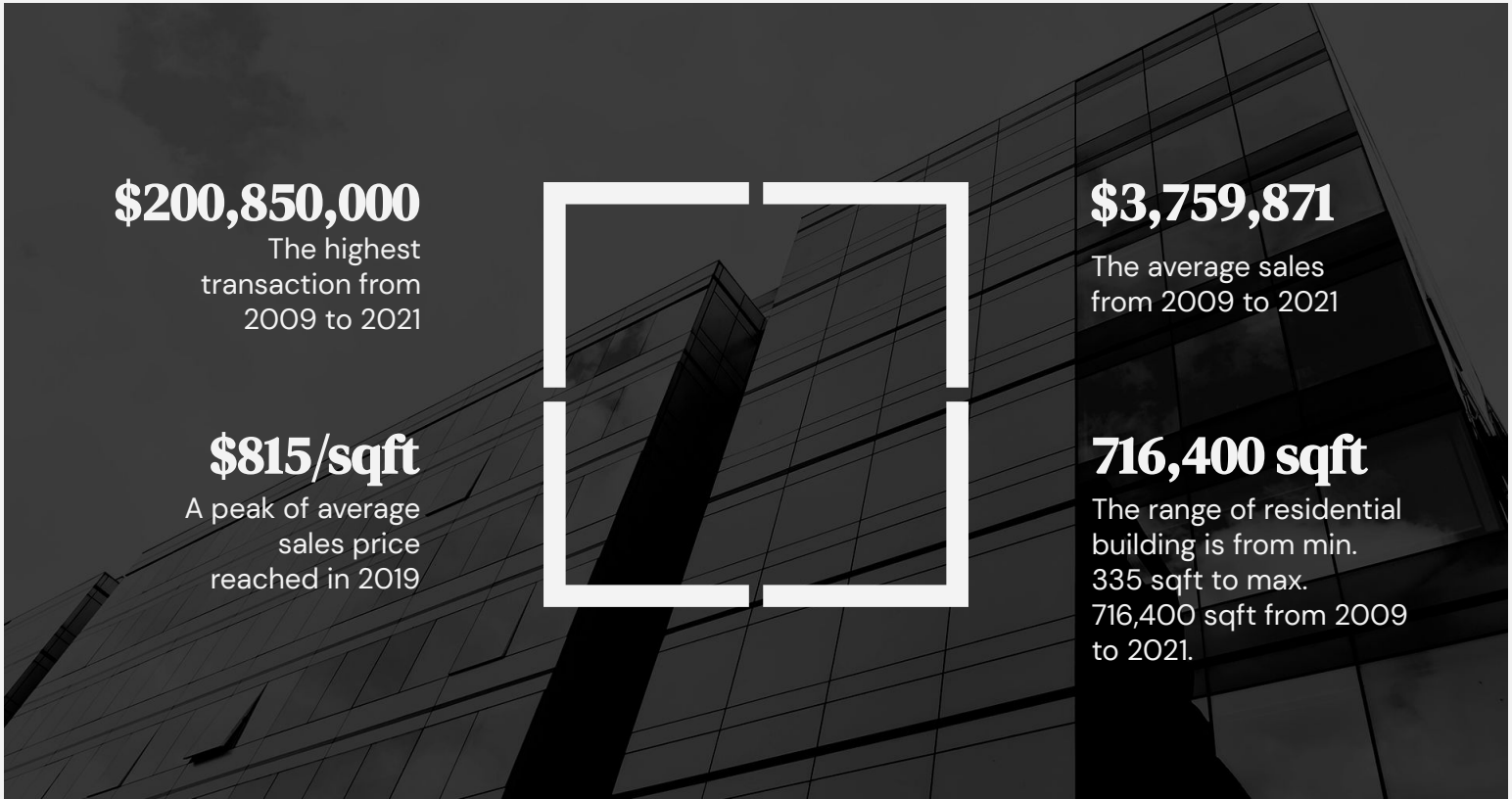
Average Sales Per Year



Top 3 Building Class in Average Sales

1. Elevator Apt: Miscellaneous
2. Hotel/Boatel
3. Elevator Apt: Converted

Key Indicators of Residential Buildings in HARLEM-CENTRAL



$200,850,000
The highest transaction from 2009 to 2021

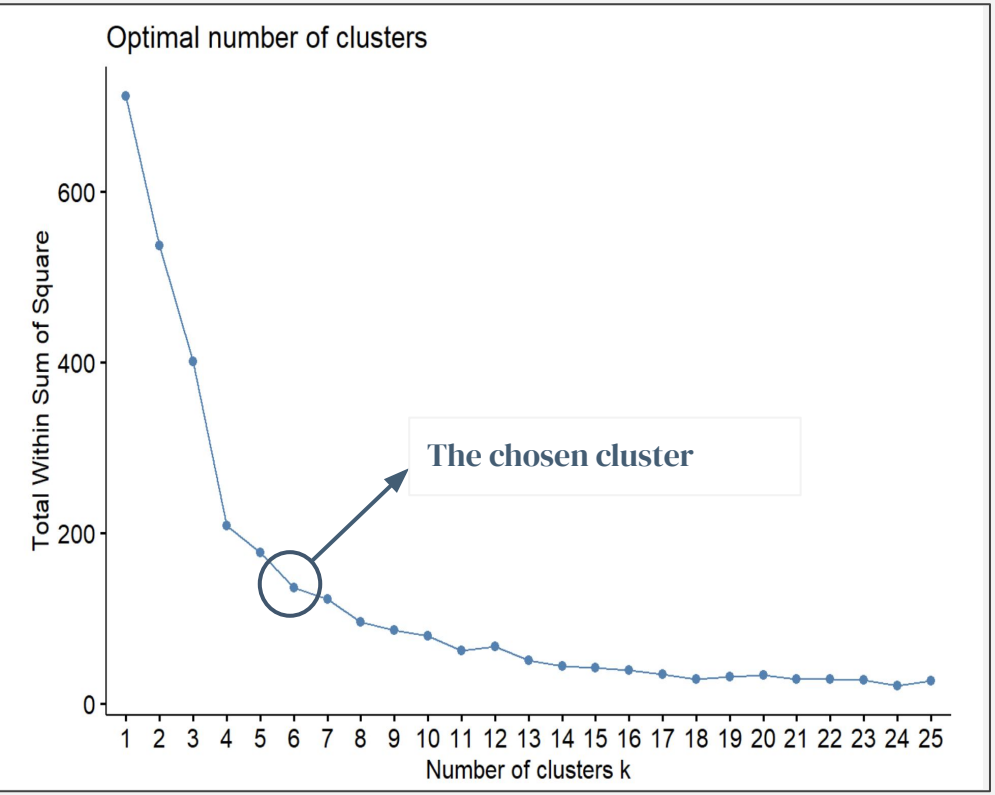
$815/sqft
A peak of average sales price reached in 2019

$3,759,871
The average sales from 2009 to 2021

716,400 sqft
The range of residential building is from min. 335 sqft to max. 716,400 sqft from 2009 to 2021.

03 Analytics Models and Processes

K-means Clustering



Three KPIs:

- Median sale prices
- Sum of sales units
- Standard deviation of sales

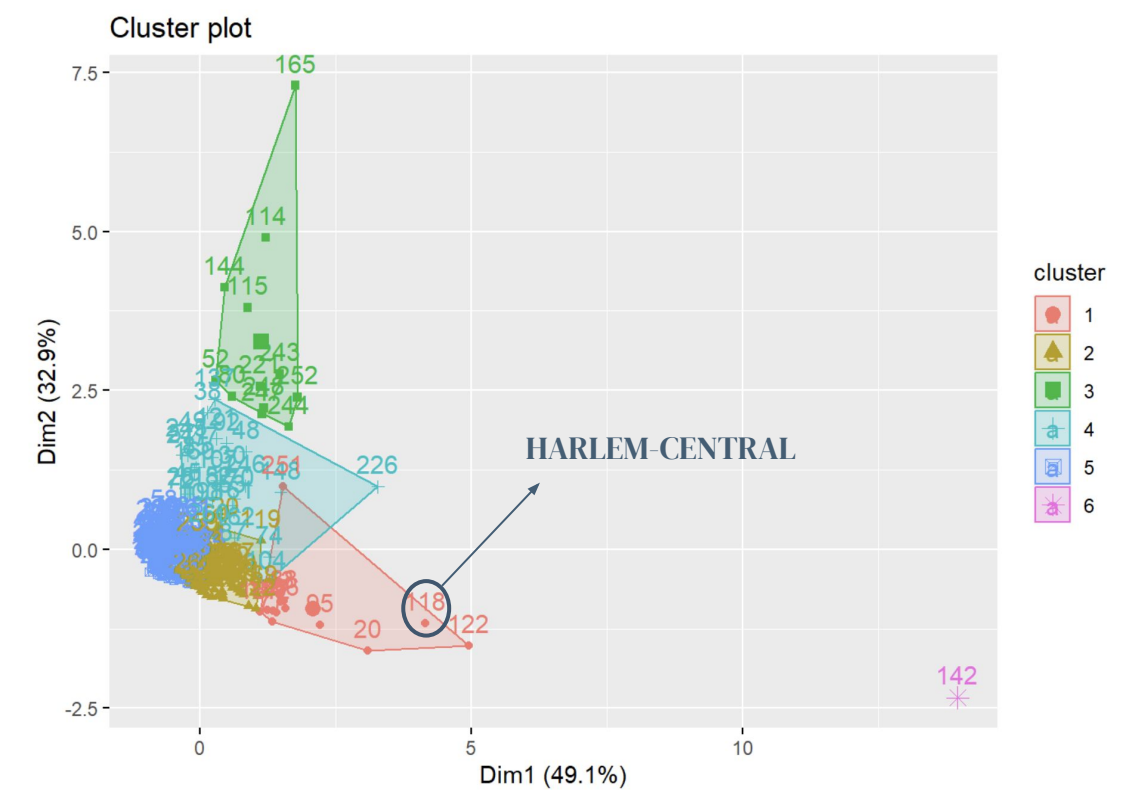
Goals:

- Find the ideal cluster.
- Identify the neighborhoods share the same characteristics with HARLEM-CENTRAL in terms of three KPIs from 2009 to 2021.

Process:

- Standardized all KPIs into same scale.
- Adjust unique identifiers.
- Remove missing value.
- Choose the optimal cluster and pinpoint the neighborhood.

K-means Clustering



13 neighborhoods belonging to the same cluster exhibit common attributes, suggesting similarities in median sale prices, total sales unit volumes, and sales variability.

Multi-linear Regression

Decipher the linear intricate interactions among the variables and offer a more holistic view of the data tendencies.

Goals:

- Understand how the independent variables of property's construction year, building code ID, gross square footage, and the number of residential units could be utilized to explain the sale price of the residential properties in HARLEM-CENTRAL from 2009 to 2021.

- Find the predictors that are most and least valuable in determining the sales amount.

Statistical Results:

Most:

Gross Sqft

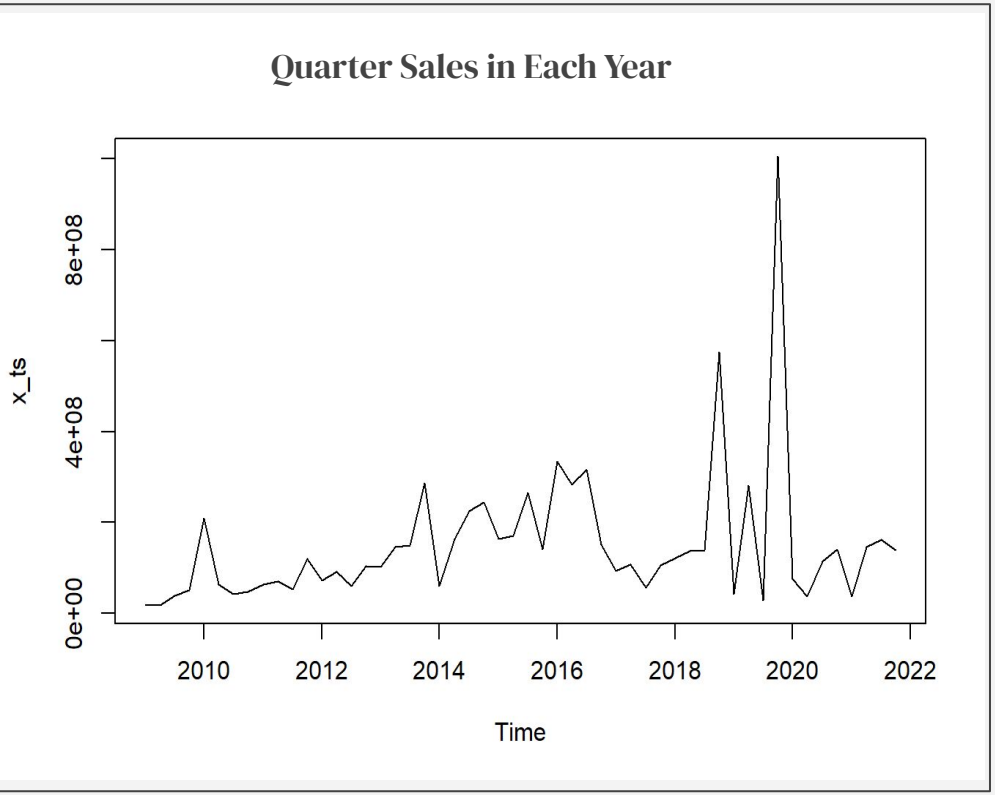
The p-value is way smaller than 0.05, indicating statistical significance and a strong relationship with building sales

LEAST:

Year of Build

The p-value is 0.97, outside the 95% confidence interval, showing no enough evidence proving that there is a relationship between building sales and the year of build.

Time Series Forecasting



Goals:

- Identify the seasonality and trends to forecast future sales.
- Compute confidence band with predicted sales for optimization model.

Process:

- Formulate the sales of residential building in each quarter from 2009 to 2021 in HARLEM-CENTRAL.
- Then predict the sales of following 8 quarters utilized a more appropriate model after comparing two prediction models.

Time-series Forecasting

AAA Model

The AAA (additive error, additive trend, and additive seasonality) model is suitable for forecasting sales when there are consistent and predictable trends or seasonality patterns.

ANA Model

The ANA (additive error, non-additive trend, and additive seasonality) model is preferable for capturing non-linear trends in sales data, offering a comprehensive representation of complex patterns that surpasses the limitations of a linear trend.

Statistical Results:

The ANA model presents lower values for AICc and BIC and a mean percentage error closer to zero, suggesting a smaller average difference between its predicted and actual values compared to the AAA model. This implies that the ANA model could offer more accurate forecasts.

Optimization Model

Forecasted Sales in following 8 quarters

Year	Quarter	Forecasted Sales $
2022	Q1	126,874,067
2022	Q2	151,434,168
2022	Q3	140,043,799
2022	Q4	262,359,096
2023	Q5	126,874,067
2023	Q6	151,434,168
2023	Q7	140,043,799
2023	Q8	262,359,096

Goals:

- The model should discovers the optimal commission rate and number of employee to hire that can maximize the NPV of profit from selling residential buildings in the next 8 quarters,
- The projected sales of residential buildings should rely on the time series model, while the rental cost estimation for commercial office space can be derived from the average sales per square foot of commercial buildings in 2021.

Optimization Model

Assumptions:

- IRR(8 %) Compound quarterly = 2.00%
- Baseline commission = 5%
- Commission vs Mrkt Pen = -0.10%
- Baseline Mrkt Pen = 5.50%
- Mrkt pen vs Commission = 0.15%
- Cost per sqft (Commercial) = $514.83
- Rent & Utilities = 2.50%
- Cost per Employee (year) = $70,000
- Operating Budget (Month) = $20,000
- The presence of employees is assumed to contribute to an additional increase in market penetration beyond the commission's influence.

Constrains:

- Min. Penetration = 4%
- Max Penetration = 6%
- Min Employee = 1
- Max Employee = 4
- Mrkt Pen increase by employee = 0.50%
- Space per Employee = 150 sqft
- Office Space = 250 sqft
- Rent & Utilities (Quarterly) = 7.50%
- Employee Salary (Quarterly) = $17,500
- Budget (Quarterly) = $60,000

Optimization Model

- Market Penetration = 5.50%+((commission rate-5%)/(-0.1%)*5.5%)
- Market Penetration added by employee = Market penetration + (number of employee*0.5%)
- Rent & Utilities = office space *($514.83*7.5%)
- Employee Salary = number of employee* $17,500

1. **If the budget cannot be carry over to next quarter :**

   The estimated NPV from 2022Q1 to 2023Q4 is approximately $363.7 million, with 2 employees per quarter, a commission rate of 4.7%, and a market penetration of 7%.

2. **If the budget can be carry over to next quarter :**

   The total NPV could reach around $376.3 million. Yet, 4 employees required in 2022Q4, 2023Q2, and 2023Q4 due to higher sales and compounded budgets, while the rest quarters only require one employee.

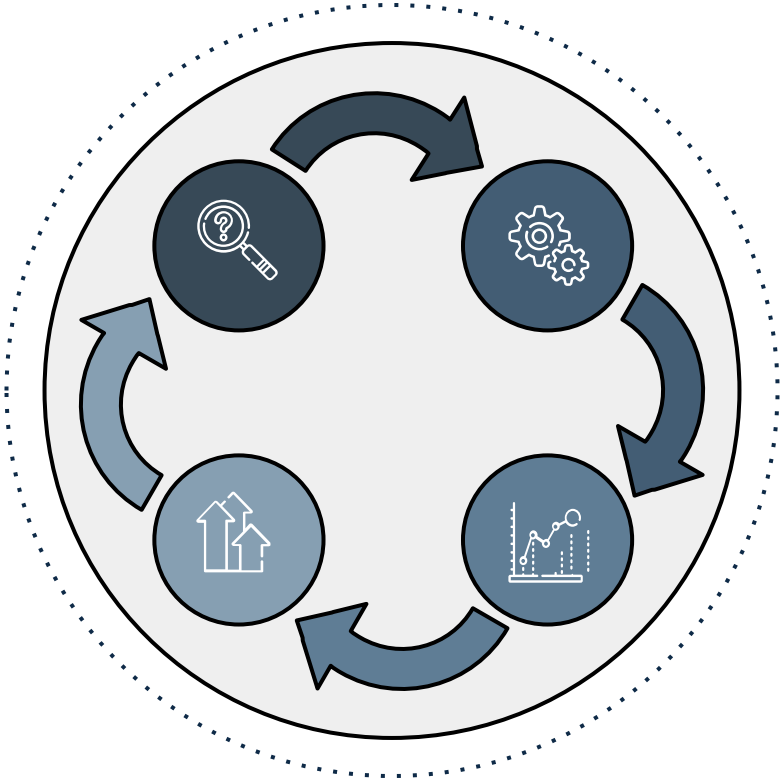
Life Cycle Management

04

The Life Cycle of Analytics Model

Define the Business Goal of the Model

Manage the Necessary Data



Model Development

Model Improvement

Data Cleaning (Manage Necessary Data)

Exclude non-useful data including sale prices and square feet of buildings equal to zero in order to provide more statistically accurate results.

Narrowed down the data to include only residential building types to find specific information, except for the model that aim to find the optimize parameters for commercial buildings.

Interpretation of results

05

Interpretation of Statistical results

k-means clustering

the study of residential real estate transactions throughout all boroughs since 2009 states that HARLEM-CENTRAL stands as the second highest-selling neighborhood group in NYC.

Regression Model

Gross square feet of the residential building has significant relationship with building sales.

Time Series Model

The model successfully predicted sales prices for the next 8 quarters, serving as a reliable reference for decision-making and future modeling.

Optimization Model

4.7% of commission rate and 7% of market penetration can maximize the NPV of profit to $363.7m with 2 employees, if there is no budget carry-over.

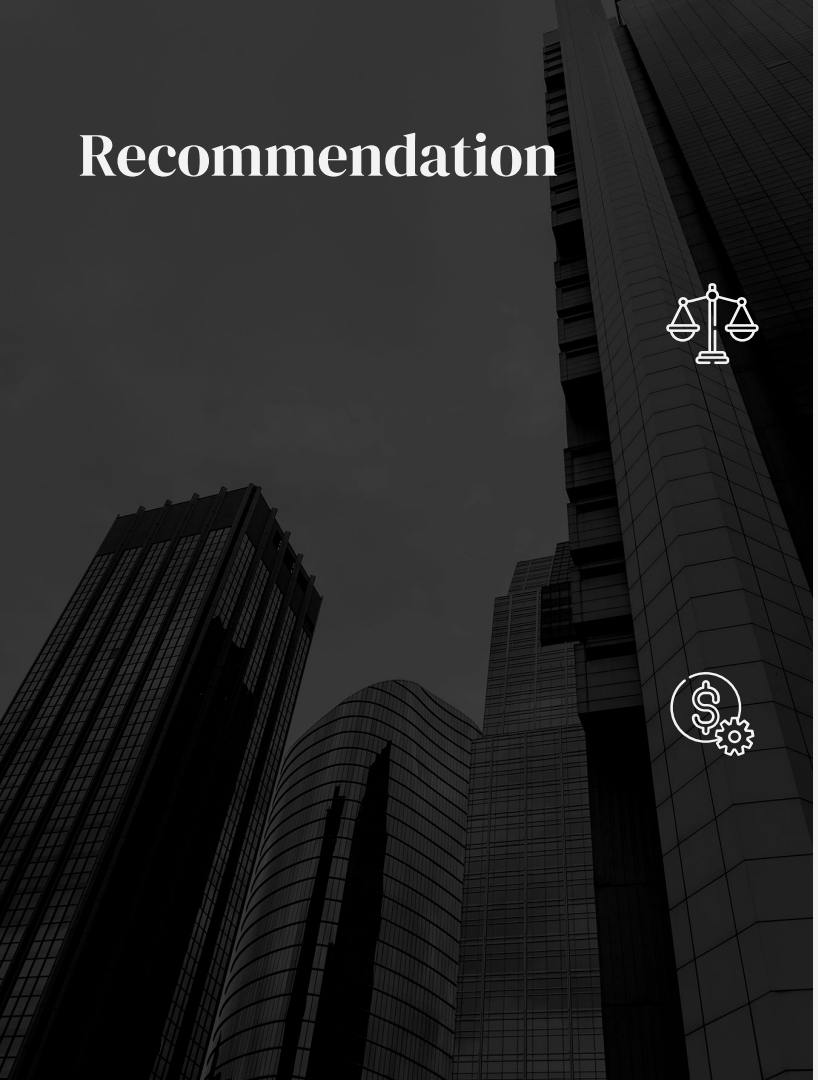
Interpretation of Statistical results

- Most KPIs analyzed indicate the neighborhood is among the top choices.
- The company can start by sourcing properties from the other 12 neighborhoods that share common characteristics, with the focus being on the square footage of the properties.
- Overall, there is an increase in sale price per square foot from 2000 to 2021.

06

Recommendation

Recommendation



Open an office in HARLEM-CENTRAL?

Based on models, statistical results and predictions, this neighborhood has presented a comparable outstanding performance in multiple dimensions. **Therefore, it is recommended to open an office in this neighborhood.**

Next Step:

Based on the 2-years foresees, it is reasonable to open a real estate office. However, to ensure a more precise prediction of long-term sales, the company may consider build further business analysis to cover more influential variables and longer period of predictions.


Thank You for Listening.
Any Question?