AXE CAPITAL

DEVELOPMENT GROUP PROPOSAL

Valentina Rizzati (AKA Taylor Mason)

May 2021


VINTAGE
EDITION
NO.3


Process & Tools

Data Gathering

Output: Listings Dataframe (Zillow) / Zip code Dataframe (NYC Gov)

Tool: Selenium, Beautiful Soup, Pandas

Data Cleaning & EDA

Output: Listings Dataframe (clean)

Tool: Pandas, Matplotlib, Seaborn

Linear Regression Modeling

Output: Linear Regression Equation

Tool: scikit-learn

Data Requirements

Data Source: Zillow

Scraped with Selenium and BeautifulSoup

NYC Zip code/ Area Mapping from NYC Gov

Feature	Feature Name	Data Type	Transformation in Final Model
y	Actual Property Price	Numerical Continuous	
y_p	Estimated Property Price	Numerical Continuous	
x_1	Number of bedrooms	Numerical Continuous	Included rooms = beds + baths / dropped number of bedrooms
x_2	Number of bathrooms	Numerical Continuous	Included rooms = beds + baths / dropped number of bathrooms
x_3	Square footage	Numerical Continuous	
x_4	Zip code	Categorical Nominal	Transformed in Borough / zip code dropped
x_5	Parking	Categorical Nominal	parking_binary
x_6	Year Built	Numerical Discrete	
x_7	Property Type	Categorical Nominal	Recategorized and transformed in dummy
x_8	Number of Schools	Numerical Discrete	Dropped
x_9	Average School Rating | 0-10	Numerical Continuous	
x_{10}	Walk Score | 0-100	Numerical Discrete	Included walk_transit_binary / walk_score dropped
x_{11}	Transit Score | 0-100	Numerical Discrete	Included walk_transit_binary / transit_score dropped
x_{12}	Pool	Categorical Binary	
x_{13}	City View	Categorical Binary	Recategorized and dropped
x_{14}	Water View	Categorical Binary	Recategorized and dropped
x_{15}	Park View	Categorical Binary	Recategorized and dropped
x_{16}	Mountain View	Categorical Binary	Recategorized and dropped

Data Modeling

Interpretation

Top Coefficients

01. Positive | Condominium
02. Negative | Small Size

Property Size

03. The smaller the the property type the larger the negative impact on price

Property Type

04. Condos are high in demand in NYC
05. Other Types include costly types

Borough

06. Lower demand for properties in Bronx

To investigate further

07. Negative coefficient of Year Built
08. Negative coefficient of Parking Binary

Feature Name	Coefficients
Intercept	1,911,670
Rooms	71,410
Year Built	-885
Walk Transit Binary	228,105
Abg School Rating	30,930
Med Size	-47,043
Small Size	-272,664
Parking Binary	-57,624
Pool Binary	55,046
Condominium	389,066
Cooperative	24,937
Multi Family	107,737
Single Family	47,716
Stock Cooperative	52,819
Two Family	66,521
Other Property Types	214,839
Bronx	-111,245
Brooklyn	195,023
Manhattan	249,371
Queens	207,041

Scale	Impact on Price
	Intercept
	High positive
	Moderate positive
	Low positive
	Low negative
	High negative

Data Modeling

Prediction

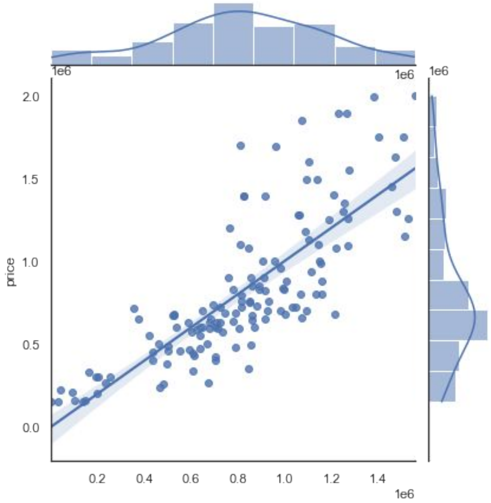
Main Considerations

01. Jointplot presents a fairly good fit
02. Residuals can be assumed to be normally distributed

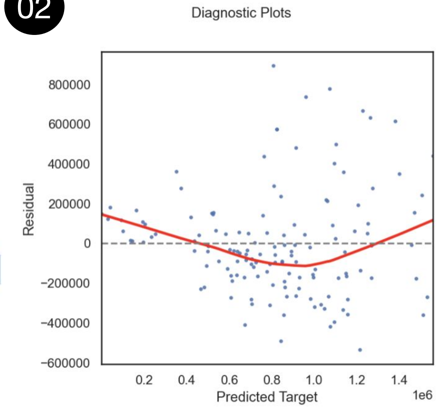
    Residuals’ variance can be assumed to be constant

    No heteroskedasticity
03. No multicollinearity since the features that were highly correlated have been transformed

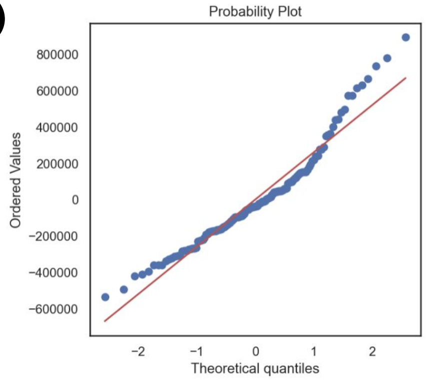
01



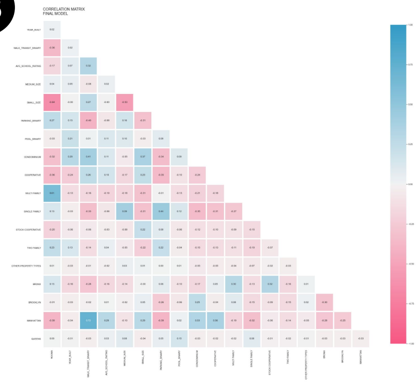
02



02



03



Summary Metrics

Linear Regression Model

0.64

R^2

$263,208

RMSE

0.58

Adjusted R^2

$196,466

MAE

Investment Opportunity

- **Location**: Manhattan
- **Property Type**: Condominiums
- **Transportation & Walkability**: Focus on areas that are well connected to public transports
- **Size**: Large condos, especially due to the working from home trend we have observed with the pandemic
- **Amenities**: Pool in the building, Parking less of a priority (but to be further investigated)


Questions?