Dubai Estates Prices

Hasan Soufan

10/1/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Dubai's Estates Prices Prediction.
- We want to reduce the time and the effort of searching about the Meter Price of the estates in Dubai by predict the Meter Price and figuring out what features effect it.
- Starting by collecting the data from Dubai Government Land Department, then doing some wrangling to find patterns in the data.
- Using exploratory data analysis (EDA) find information like distribution between the standards and the Meter Price.
- Using Statistical Experiments and significance testing find standards that have effect on the Meter Price.
- Performing a predictive analysis and building models to predict the Meter Price depending on the previous standards and chose the best models based on some metrics like R-Square.
- The results that we can figure out which standards effect the Meter Price and predict its value with 67% of certainty.

Introduction

- Dubai Estates are being bought and sold many times and when someone need to buy/sell an estate, it need another one to approximate the meter price, instead of searching for its prices and features that effect that price, this project is to figure out the way to make this process easy and fast based on studding Dubai Real Estates data for 2022.
- Depending on the previous we will study Dubai Real Estates Data to figure out the following:
  - What are the features that effect the meter price of the estates.
  - Build a prediction model trying to predict the meter price based on the different situation of the previous standards.

Methodology

Section 1

Methodology

Executive Summary

- **Data Collection**, data was collected using Dubai Government Land Department Website.
- **Data Wrangling**, solved any problem in the data like extreme values .. , found patterns and object types of each feature like rooms and Areas.
- **Exploratory Data Analysis (EDA)**, using visualization to figure out which features could effect or related to the Meter Price like Near Landmark, Near Metro .. .
- **Statistical Experiments and Significance Testing**, using some testing methods found features that effect the Meter Price.
- **Predictive Analysis**, built regression models like K Nearest Neighbors and XGBoost, evaluated by R Square, MAE* and MSE*

- *MAE : Mean Absolute Error
- *MSE: Mean Squared Error

Data Collection

- the data was collected from Dubai Government Land Department Website Here, as a csv file, then extracted, here some of it

Transaction Number	Transaction Date	Property ID	Transaction Type	Transaction sub type	Registration type	Is Free Hold?	Usage	Area	Property Type	Property Sub Type	Property Size (sq.m)	Room(s)	Parking	Nearest Metro	Nearest Mall	Nearest Landmark	No. of Buyer	No. of Seller	Master Project	Project
102-1-2022	2022-01-03 07:30:34	1131371499	Sales	Sell - Pre registration	Off-Plan	Free Hold	Residential	DUBAI HARBOUR	Unit	Flat	119.25	2 B/R	1	NaN	NaN	NaN	1	1	NaN	Beach Mansion
102-10-2022	2022-01-03 07:32:56	100227279	Sales	Sell - Pre registration	Off-Plan	Free Hold	Residential	Wadi Al Safa 7	Unit	Flat	33.44	Studio	1	NaN	NaN	NaN	1	1	NaN	RUKAN
102-100-2022	2022-01-03 12:30:23	1147007457	Sales	Sell - Pre registration	Off-Plan	Free Hold	Residential	ARJAN	Unit	Flat	51.05	1 B/R	1	Sharaf Dg Metro Station	Mall of the Emirates	Motor City	1	1	NaN	SKYZ By Danube
102-1000-2022	2022-01-10 13:21:58	1133339823	Sales	Sell - Pre registration	Off-Plan	Free Hold	Residential	BUSINESS BAY	Unit	Flat	36.24	Studio	1	Business Bay Metro Station	Dubai Mall	Downtown Dubai	2	1	NaN	Peninsula One
102-10000-2022	2022-04-13 11:21:32	1171598130	Sales	Sell - Pre registration	Off-Plan	Free Hold	Residential	Jumeirah First	Unit	Flat	118.34	2 B/R	1	NaN	NaN	NaN	1	1	NaN	Port De La Mer - Le Ciel

Data Wrangling

- Dealing with Null values.
- Figuring out features types and patterns, found that there are features with non balanced data like (Usage) and features with many unique values like (Parking).
- created label column as Meter Price by dividing (Amount on Property Size sq.m).

Exploratory Data Analysis

- **Histogram plot:** the distribution of the (Meter Price) feature showed that it skewed to the left.
- **Box Plots**: show the distribution of the features and figuring out if there is any overlapping between the categories in each of the following features:
  - Transaction Type vs Meter Price.
  - Area vs Meter Price.
  - Rooms vs Meter Price.
  - Parking vs Meter Price.
  - Nearest Metro vs Meter Price.
  - Nearest Mall vs Meter Price.
  - Nearest Landmark vs Meter Price.

Statistical Experiment

- **ANOVA** testing: this method was used because all of the features have categorical type data and the label is continuous:
  - features that we selected have an effect on the label (Meter Price) with P-value near to zero.

Predictive Analysis

- Apply one hot encoding on the features, then because that all of the features are binary we did Corresponding Analysis (CA) as a feature reduction method to reduce the dimensions from 43 to 28.
- We split the dataset into train and test data to test the models if they accurate on samples from out of the range.
- Using Grid search and cross validation for model selection, and hyper parameter tuning.
- Train regression models (K Nearest Neighbors – XGBoost – Random Forest Regressor – Neural Network)
- Using metrics like R-Square, MAE, MSE, and Sum of Residuals to evaluate the models and selecting the best model.

Results

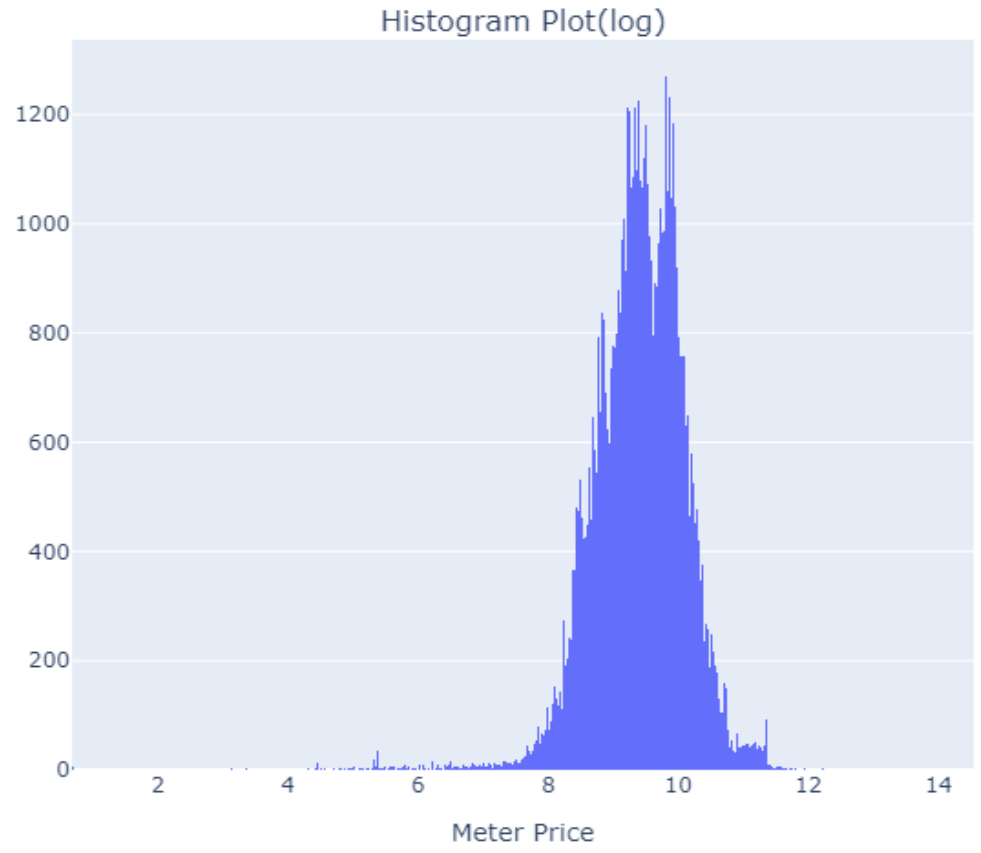
- **Exploratory Data Analysis:**
  - There are too many outliers and the label skewed to the left so we encode it using log.
  - There is overlap between (Rooms) categories.
  - There is overlap between (Parking) categories.
  - There is a little overlap between (Nearest Metro) categories.
  - There is a little overlap between (Nearest Mall) categories.
- **Statistical Experiment:**
  - Transaction Type effect the label.
  - Area effect the label.
  - Rooms effect the label.
  - Parking effect the label.
  - Nearest Metro effect the label.
  - Nearest Mall effect the label.
  - Nearest Landmark effect the label.
- **Predictive Analysis:**
  - More than one model can predict the Meter Price with 67% of certainty.

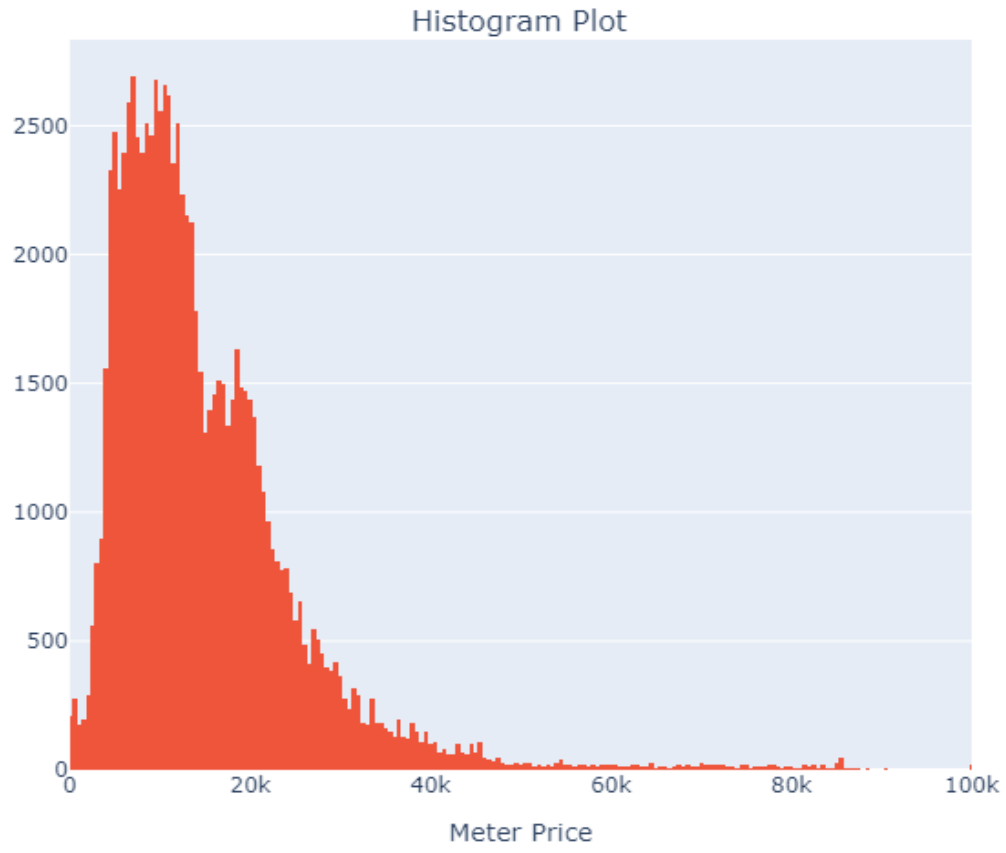
EDA Insights

Section 2

Meter Price Distribution

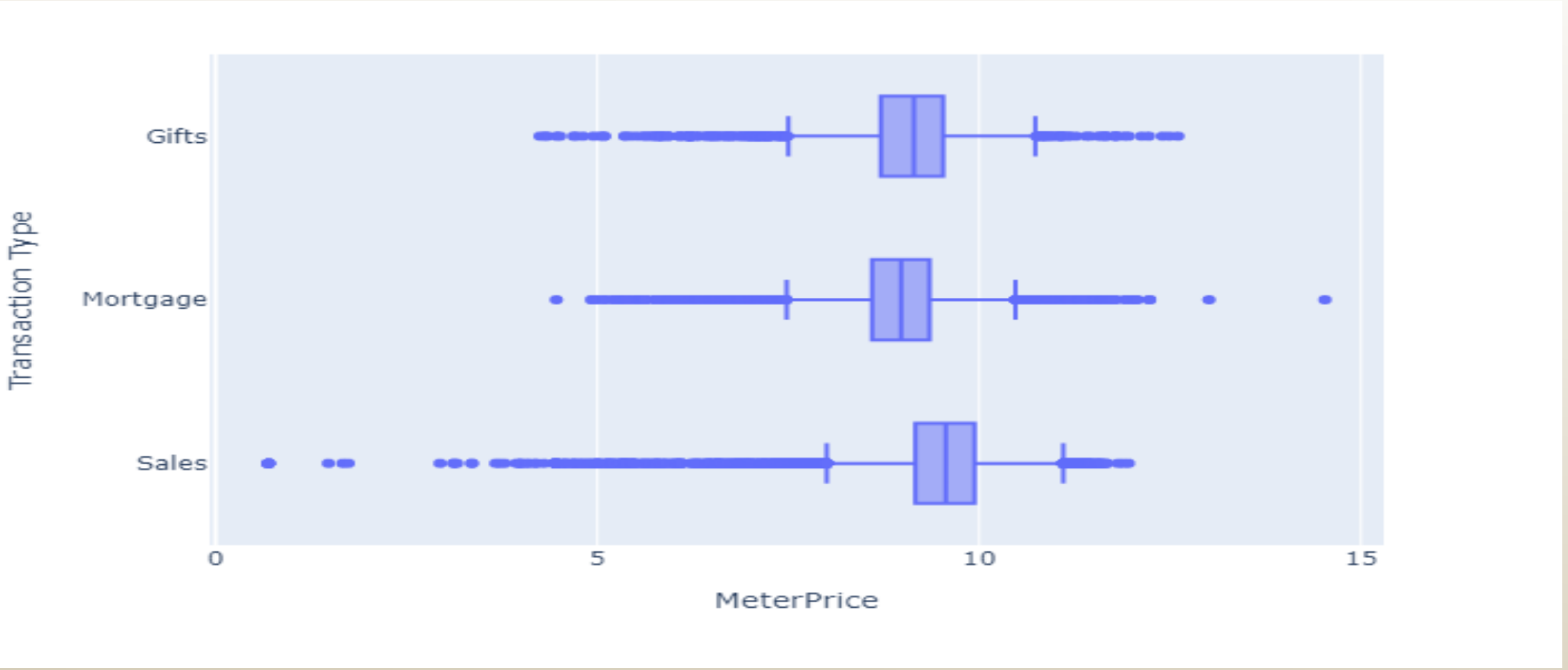
- We found that (Meter Price) skewed to the left, so we take the (log) of it as in the blue plot.





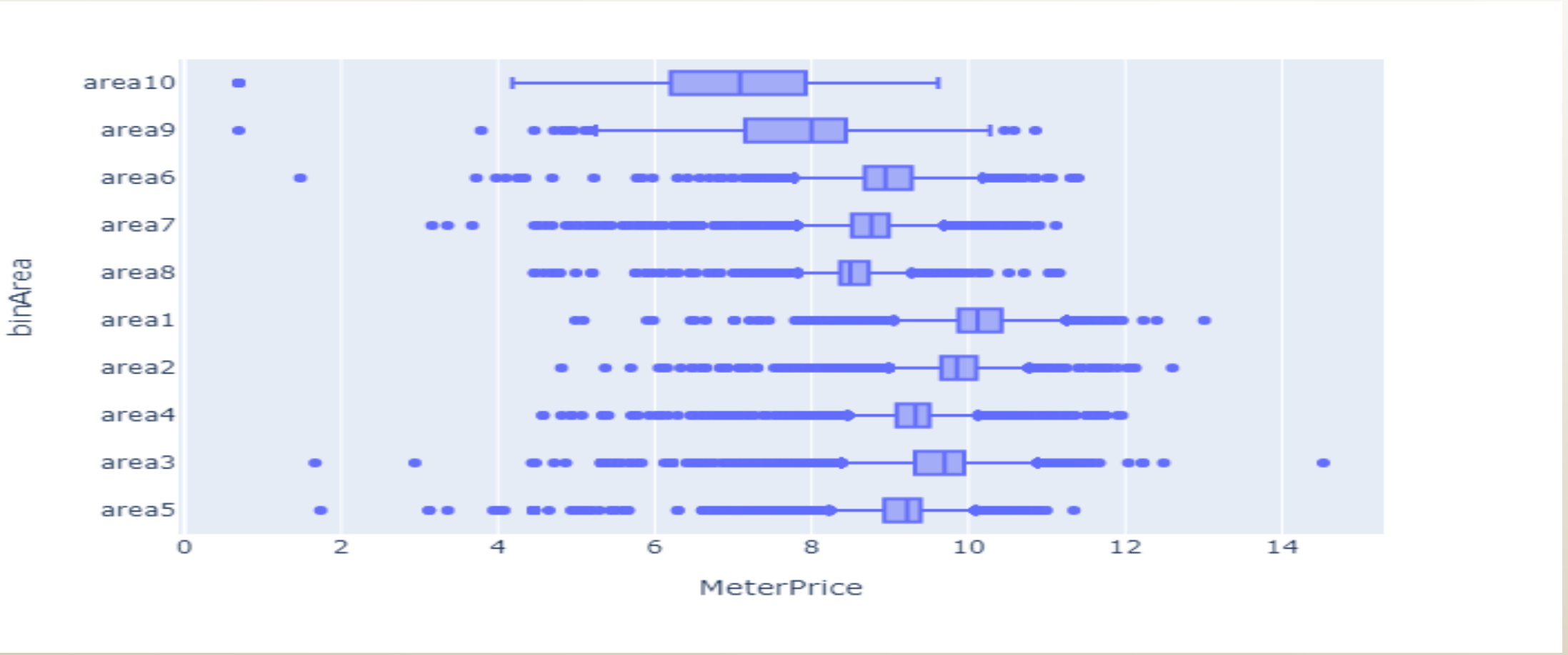
Meter Price differentiated by Transactions Type

- We found that (Transaction Type) categories have difference in the median, also in the spread specifically in the outliers.



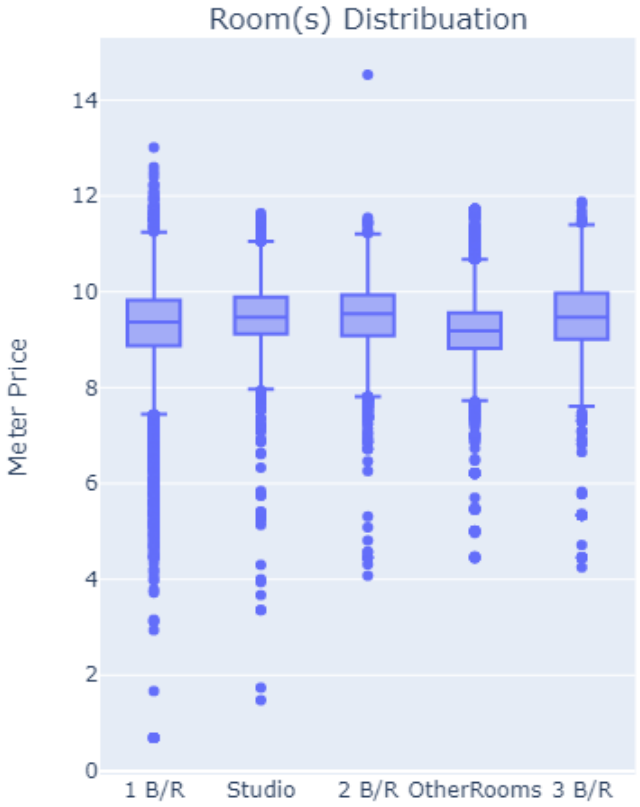
Meter Price differentiated by Area

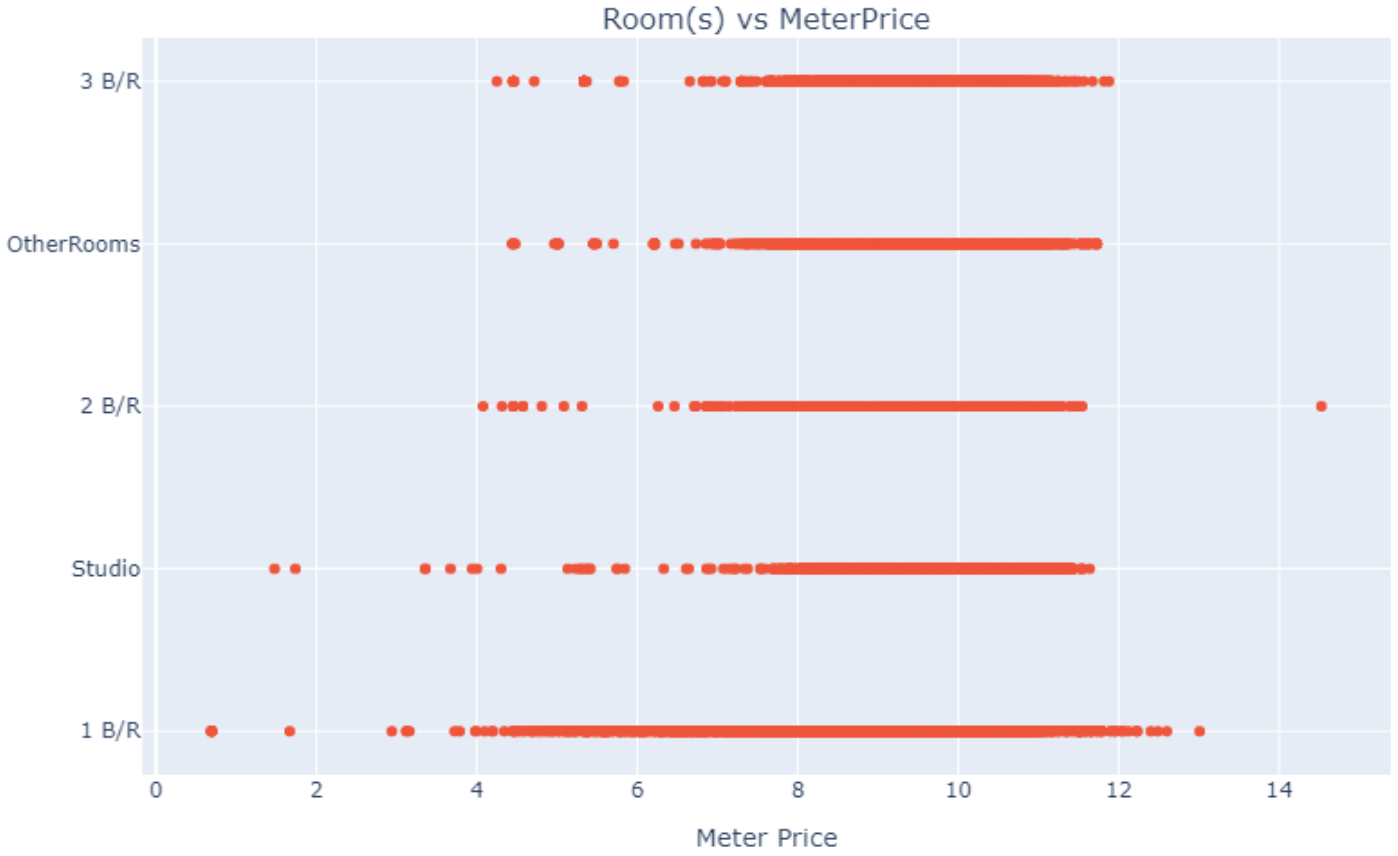
- After we binned the feature (Area) due to its many value, We found that its categories have difference in the median, also in the spread



Meter Price differentiated by Rooms

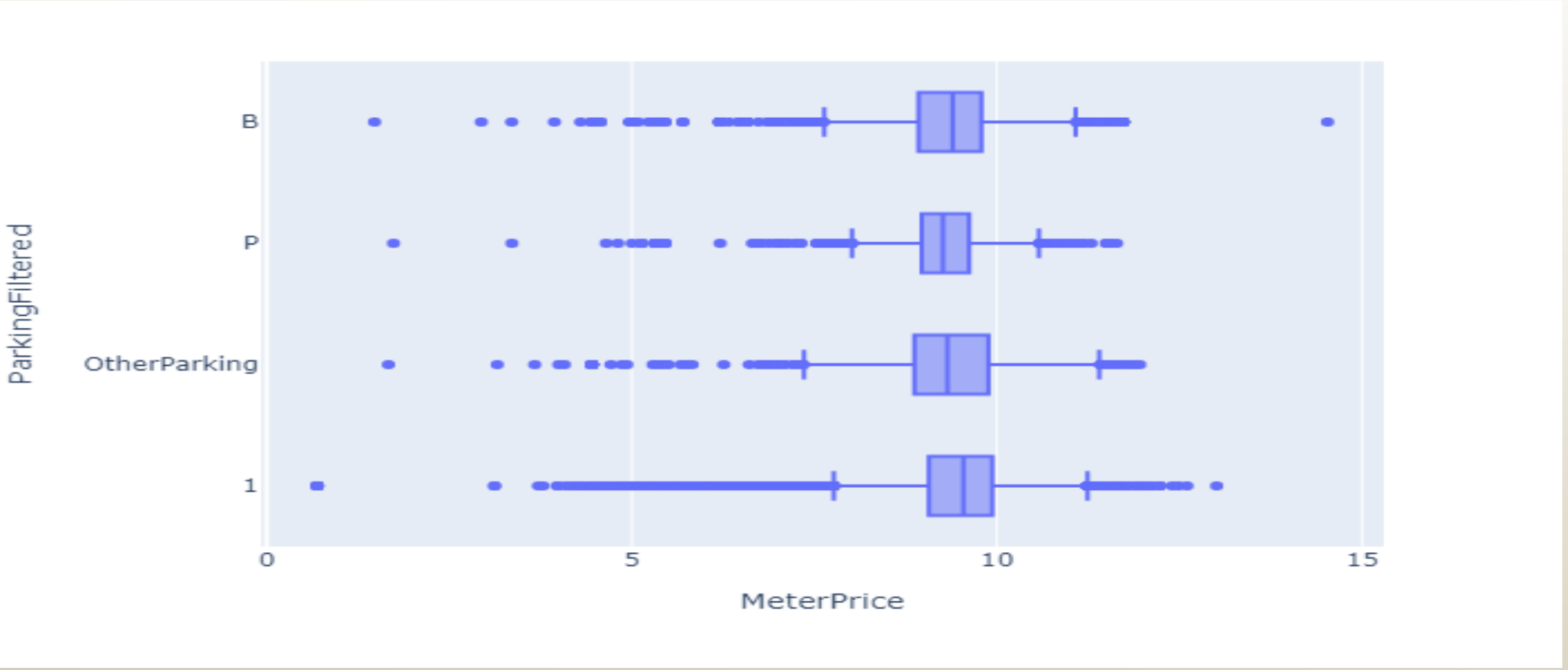
- We found that (Rooms) categories medians are closed and there is overlap.





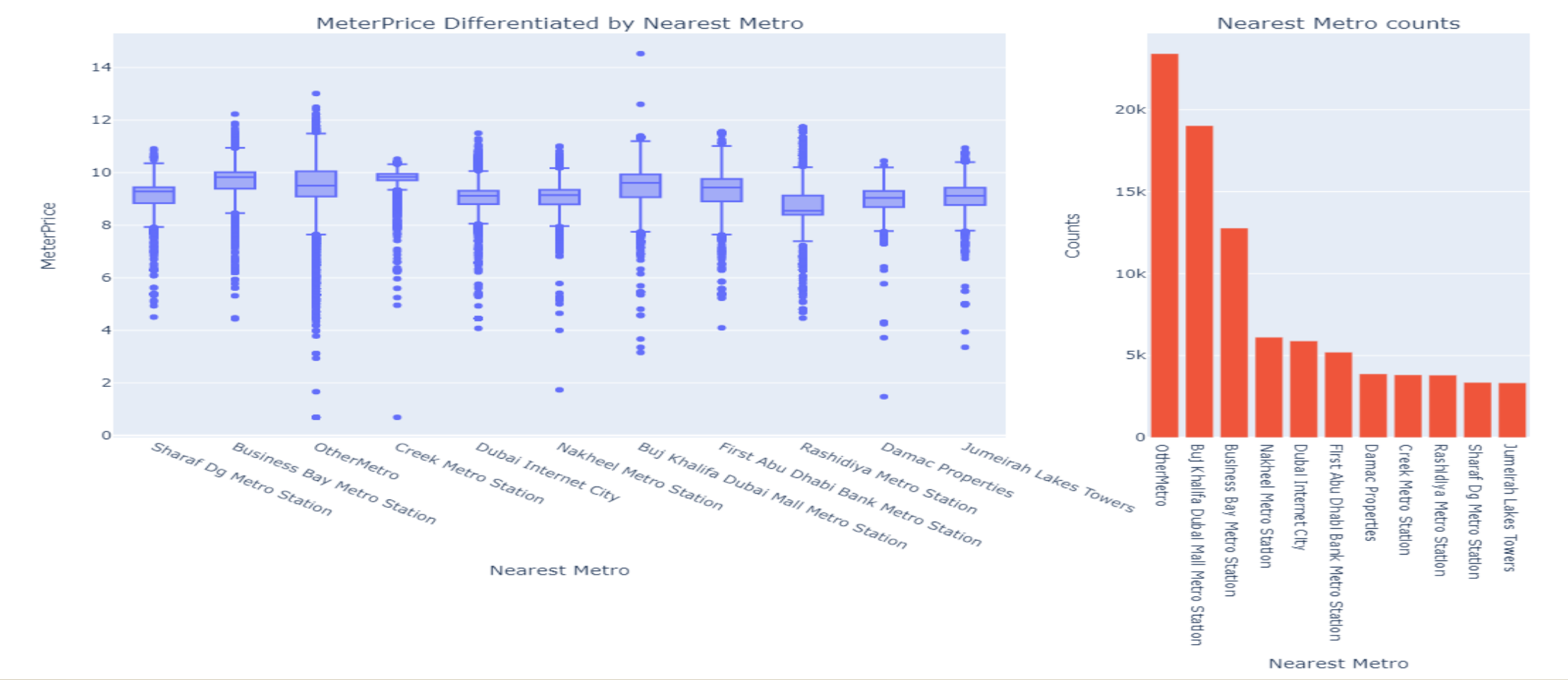
Meter Price differentiated by Parking

- After we filtered the feature (Parking) due to its many values We found that its categories medians are closed and there is overlap.



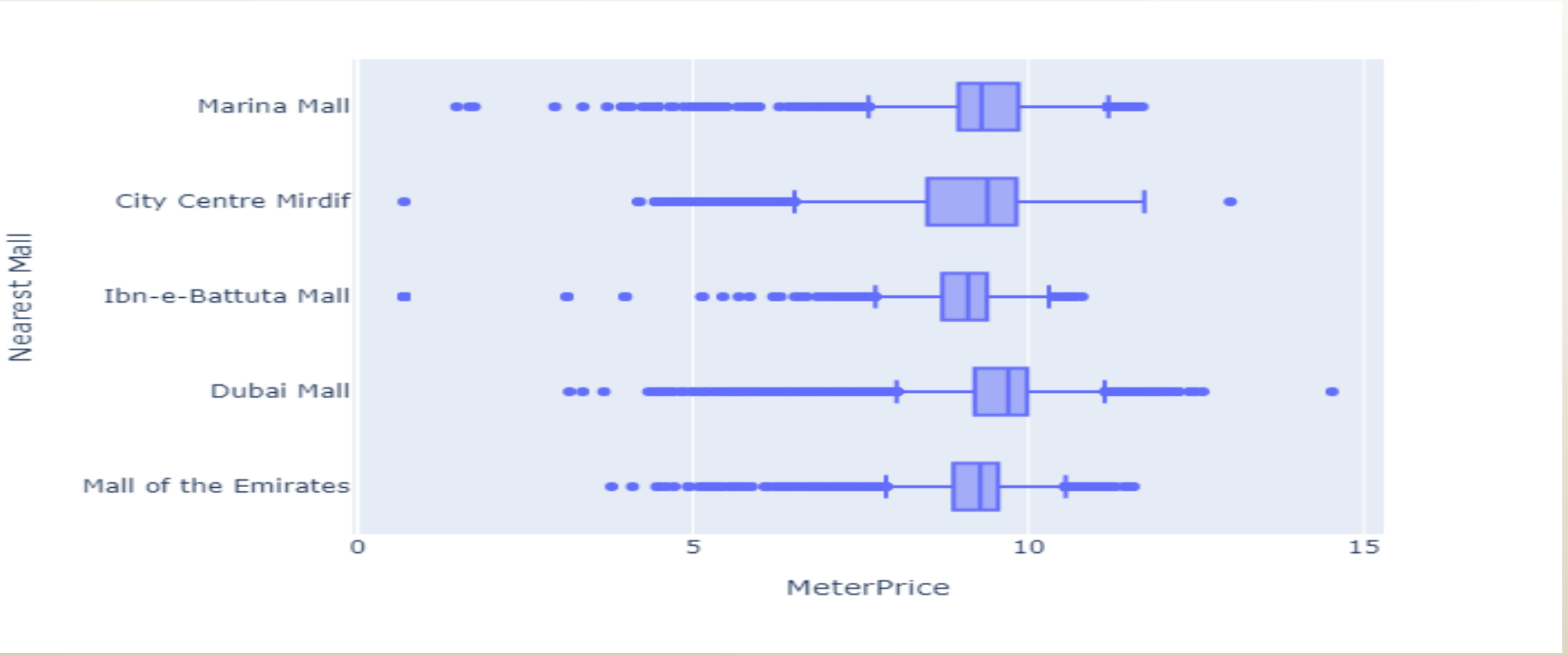
Meter Price differentiated by Nearest Metro

- We found that (Nearest Metro) categories have overlap but have difference in the medians and counts.



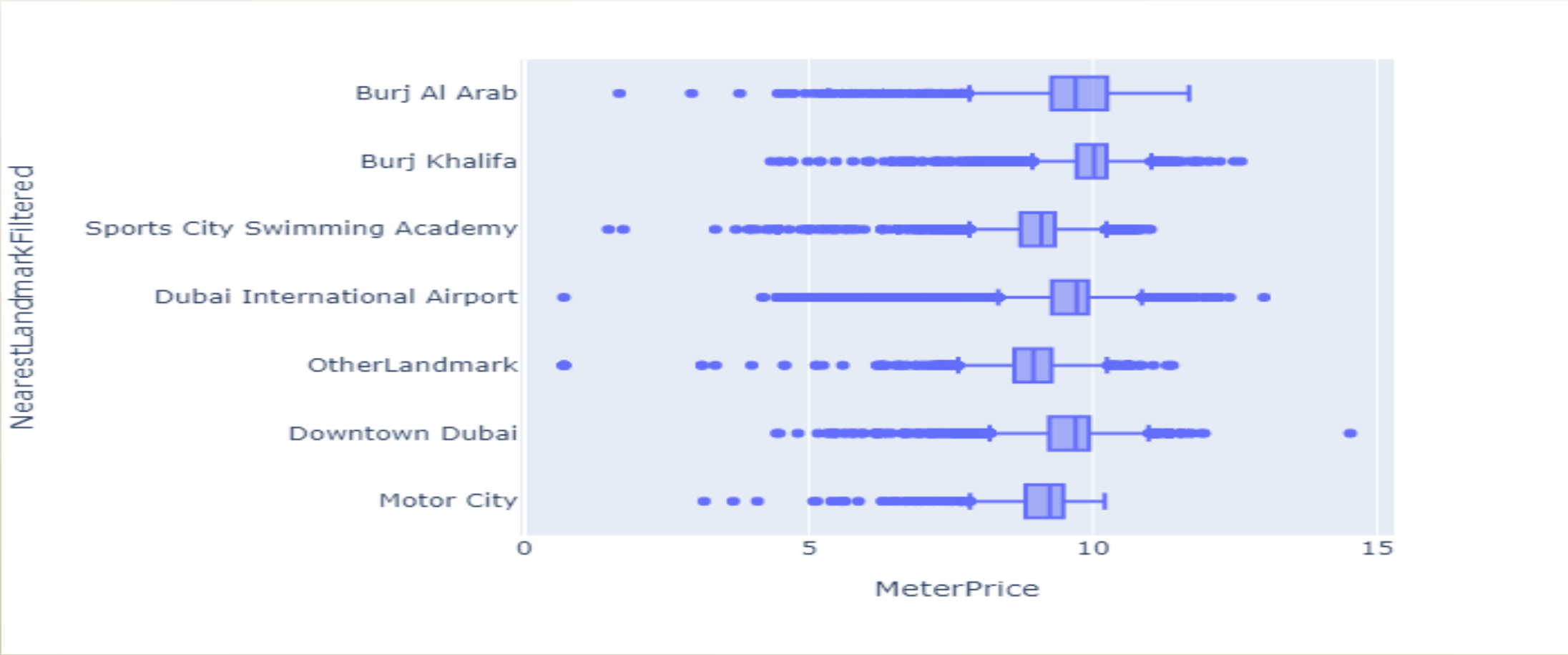
Meter Price differentiated by Nearest Mall

- We found that (Nearest Mall) categories have a little overlap but have difference in the median and the spread.



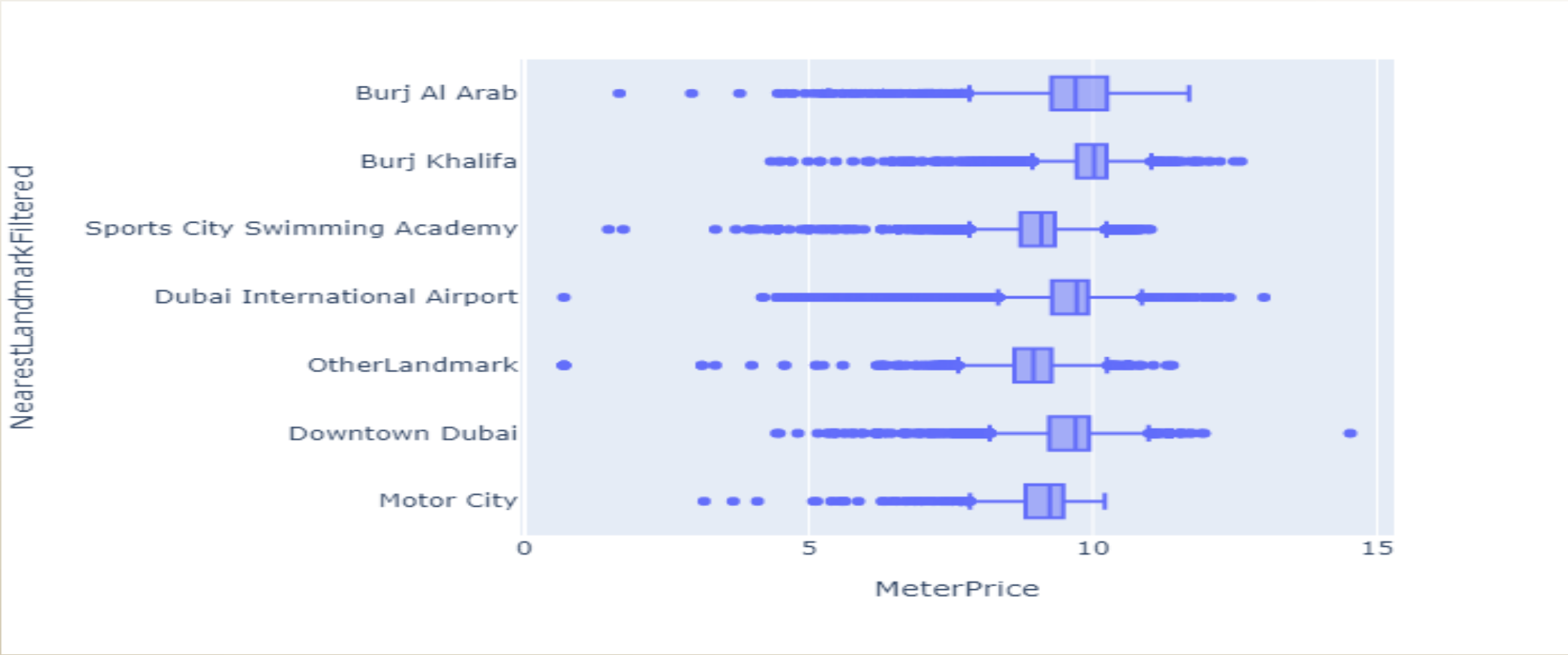
Meter Price differentiated by Nearest Landmark

- We found that (Nearest Landmark) categories have difference in the median and the spread.



Meter Price differentiated by Nearest Landmark

- We found that (Nearest Landmark) categories have difference in the median and the spread.



Real estates Transactions

Transaction Type	Area	Room(s)	Parking	Nearest Metro	Nearest Mall	Nearest Landmark	MeterPrice
Sales	ARJAN	1 B/R	1	Sharaf Dg Metro Station	Mall of the Emirates	Motor City	12732.615083
Sales	BUSINESS BAY	Studio	1	Business Bay Metro Station	Dubai Mall	Downtown Dubai	20719.536424
Sales	DOWN TOWN JABAL ALI	Studio	1	UAE Exchange Metro Station	Ibn-e-Battuta Mall	Expo 2020 Site	12825.112108
Sales	DUBAI CREEK HARBOUR	2 B/R	1	Creek Metro Station	City Centre Mirdif	Dubai International Airport	16312.785263
Sales	JUMEIRAH VILLAGE CIRCLE	1 B/R	1	Dubai Internet City	Mall of the Emirates	Sports City Swimming Academy	9317.785349

Number Transaction Type

Transaction Type	Counts
Sales	67723
Mortgage	18812
Gifts	4124

Top 5 Areas based on the number of Transactions

Area	Counts
BUSINESS BAY	12674
JUMEIRAH VILLAGE CIRCLE	7082
DUBAI MARINA	6193
BURJ KHALIFA	6156
PALM JUMEIRAH	4257

Top 5 Rooms based on the number of Transactions

Room(s)	Counts
1 B/R	38032
2 B/R	21940
Studio	14639
3 B/R	9207
Office	3141

Top 5 Nearest Metro based on the number of Transactions

Nearest Metro	Counts
Buj Khalifa Dubai Mall Metro Station	19026
Business Bay Metro Station	12781
Nakheel Metro Station	6125
Dubai Internet City	5889
First Abu Dhabi Bank Metro Station	5212

Number of Nearest Malls

Nearest Mall	Counts
Dubai Mall	36331
Marina Mall	27432
Mall of the Emirates	13423
City Centre Mirdif	7884
Ibn-e-Battuta Mall	5589

Number of Nearest Landmarks

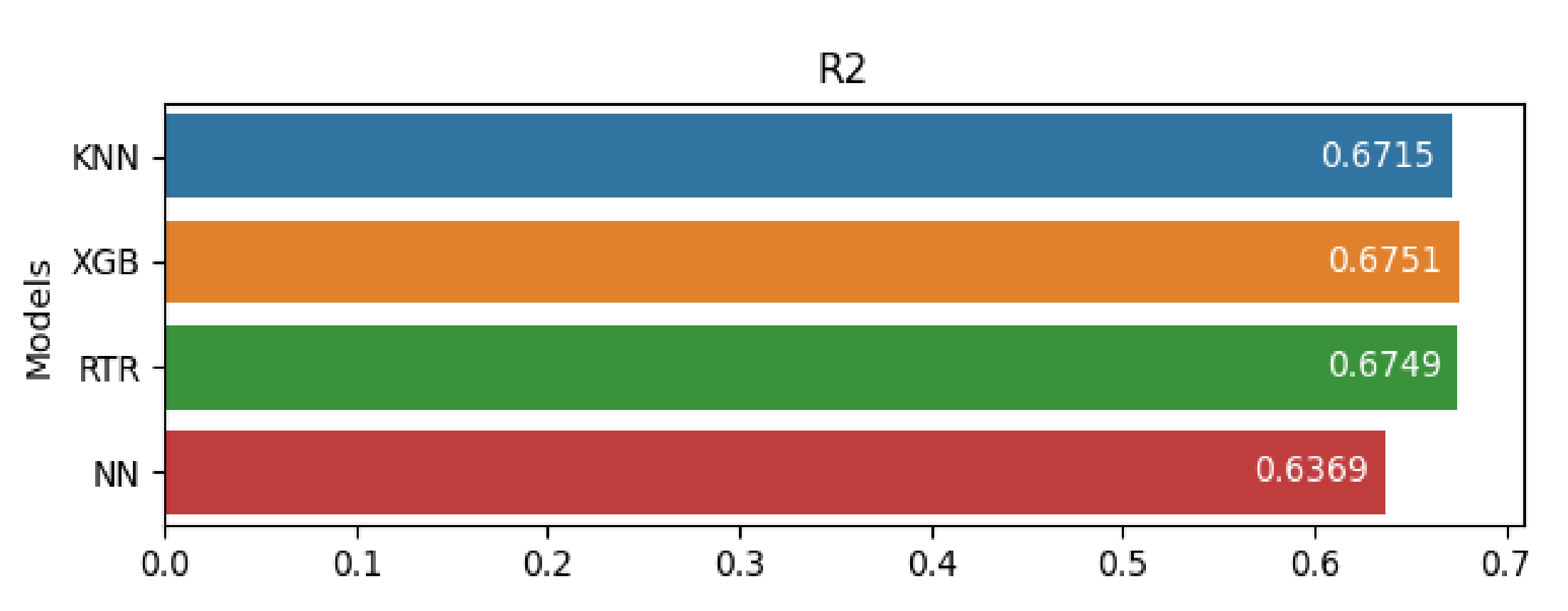
Nearest Landmark	Counts
Downtown Dubai	23966
Sports City Swimming Academy	19520
Burj Al Arab	16447
Motor City	8419
Burj Khalifa	7428
Dubai International Airport	6536
IMG World Adventures	3725
Expo 2020 Site	3437
Dubai Cycling Course	582
Global Village	290
Al Makhtoum International Airport	225
Dubai Parks and Resorts	82
Jabel Ali	2

Predictive Analysis (Regression)

Section 3

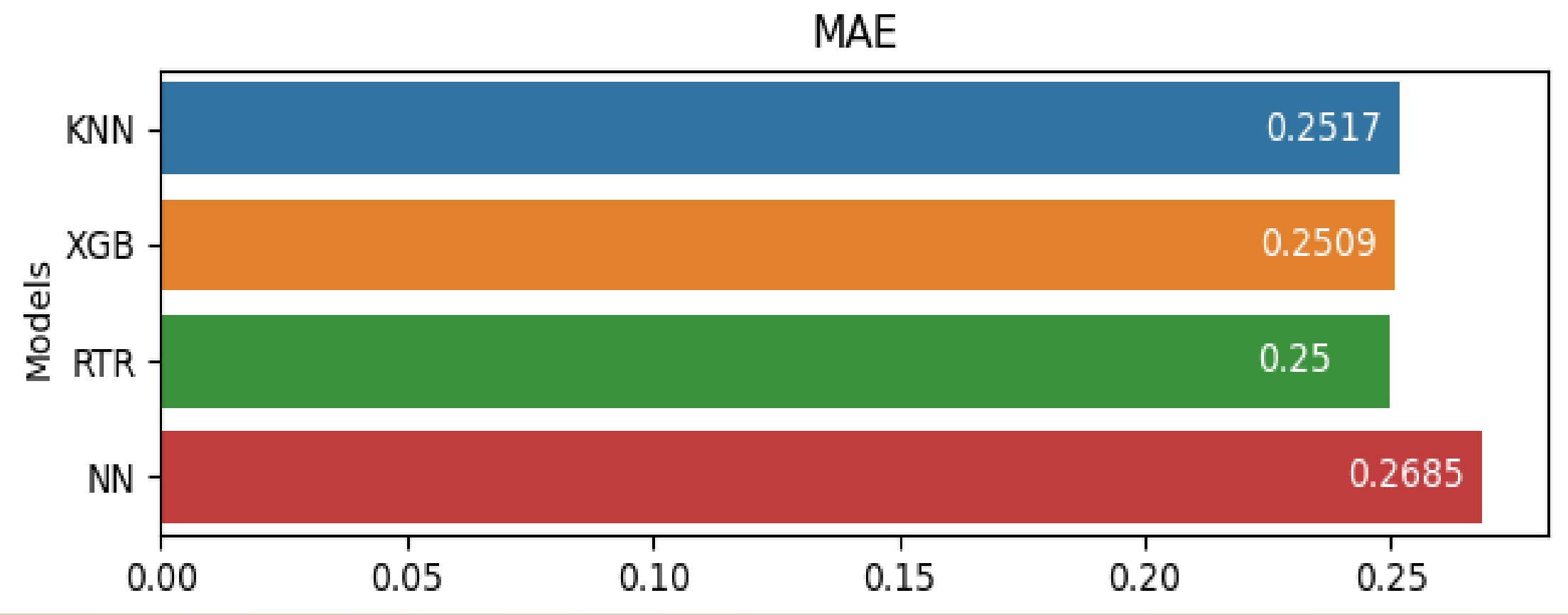
Regression R-Square

- The R-Square of (KNN, XGB and RTR) are very close, but neural network is less than the others.



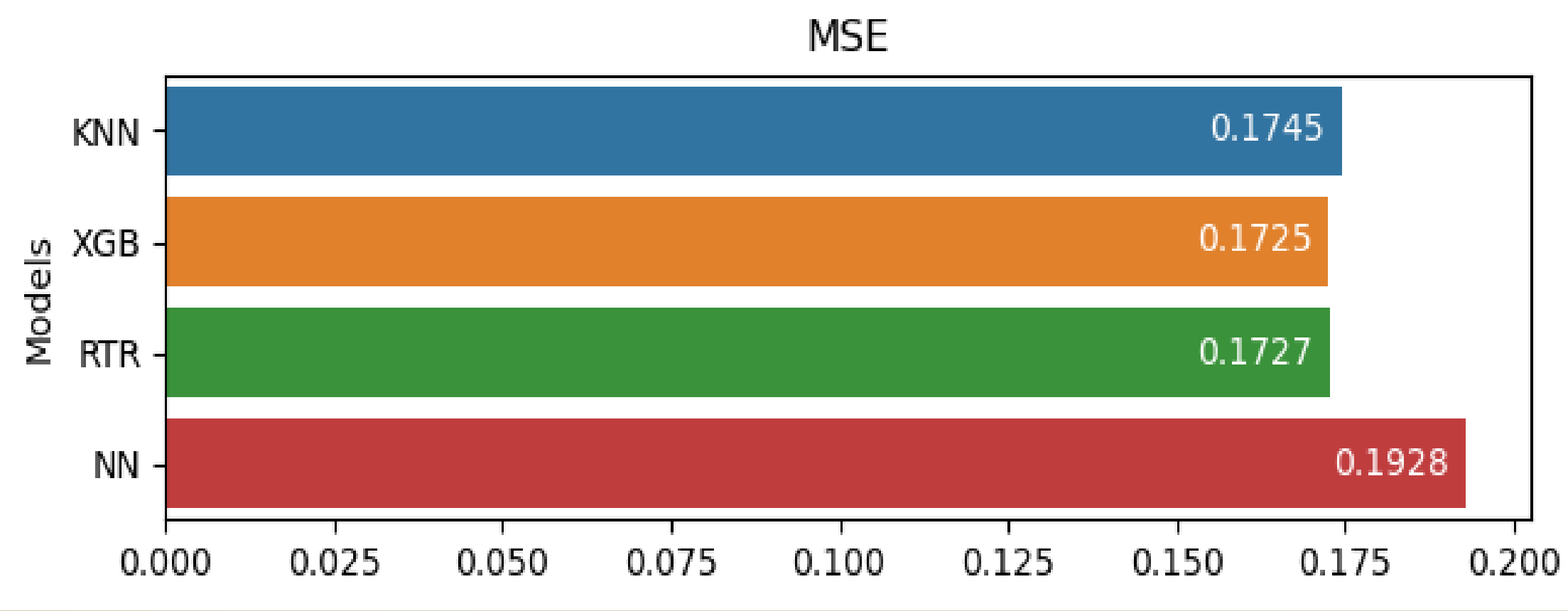
Regression Mean Absolute Error

- The MAE of (KNN, XGB and RTR) are very close, but neural network is higher than the others.



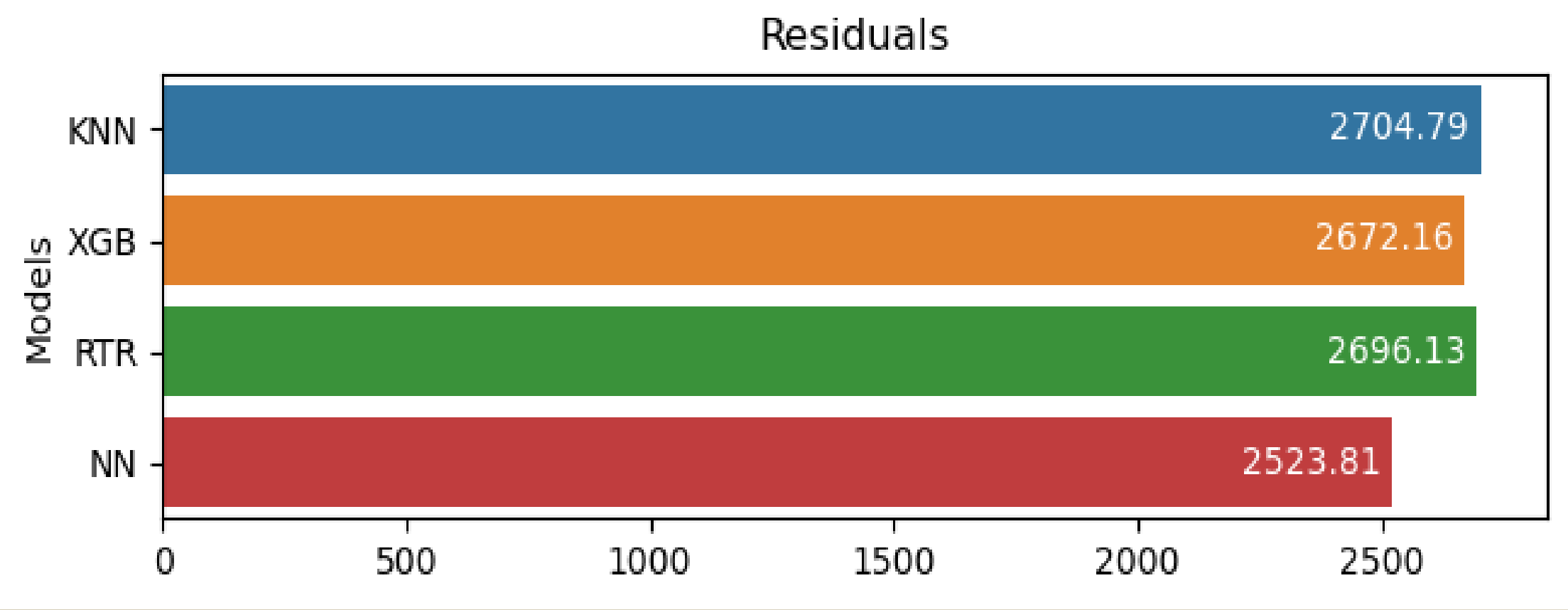
Regression Mean Squared Error

- The MSE of (KNN, XGB and RTR) are very close, but neural network is higher than the others.



Regression Residuals

- The MSE of (KNN, XGB and RTR) are very close, but neural network is less than the others.



Conclusions

- Studied the real estates data from Dubai Government Land Department to figure out what features that have an effect on the Meter Price of the estates in Dubai and built a prediction model to predict the Meter Price.
- Found that there is a relation between (Transaction Type, Area, Rooms, Parking, Nearest Metro, Nearest Mall and Nearest Landmark) and the (Meter Price).
- Performed a predictive analysis and built a regression models to predict the Meter Price based on the previous standards, evaluated by R-Square, MAE, MSE and Residuals metrics, afford in the best situation 67% of certainty, using the XGBoost model because it is fast more than the others.
- The results are we can figure out what the standards that effect the Meter Price in Dubai and predict it with 67% of certainty.

Appendix

Thank You!