

Project Proposal Summary

This is just a summary of Project Proposal. For the Project proposal Markdown, please refer to the datafile “Rental_Freq_Summary_Modelling “ file with coding, output and findings.

Introduction

It is quite interesting to use analytic models to price the property. In this project. We focus on the renting properties in GTA. The dataset is scraped from trebhome.com using BeautifulSoup and Requests with Python. The objective of this project is to find which properties are the best for investments. The result of this study can give some information to the property investors to price the renting properties.

Dataset

The dataset is scraped from trebhome.com posted on March 20 to March 24, 2019 using BeautifulSoup and Requests with Python. The filters used are area” Toronto” , min_price”\$300” , and max_price “\$50,000” (Excluded locker and parking spot lease priced below \$300) .The dataset contains 3164 records and 6 variables with duplicates. The variable names are ID, Address, Bedrooms, Bathrooms, Type, Price. The final dataset is written as rental1.csv file.

Project Analysis Plan

- 1) Load data
- 2) Explore data
- 3) Clean data, remove duplicates, remove missing values
- 4) Change datatype as necessary
- 5) Statistics about Price, Bedrooms , Bathrooms (min, max, mean , median, Q1,Q3)
- 5) Look into counts(frequency) by Type and visualization
- 6) Look into Price distribution related to Type , Bedrooms, Bathrooms
- 7) Find Top 10 Type by the Number of Properties
- 8) Find Top 10 Type by the Average Price of Properties
- 9) Summarize the Price with Type, Bedrooms, Bathrooms and Look at Price Trend
(see Summary_Type_Beds_Bath.csv)
Look at one specific example: # Look the price trend in Type “Detached 2-storey”
- 10) Summarize the Price with Bedrooms, Bathrooms and Look at Price Trend
(See Summary_Beds_Baths.csv)

11) Summarize the Price with Bedrooms and Look at Price Trend (High level summary)
(See Summary_Beds.csv)

Project Modelling Analysis Plan

- 1) train dataset 80% , test dataset 20%
- 2) decision tree model
- 3) Random Forest model with 500 trees
- 4) Evaluate models: compare decision tree vs Random Forest model (rmse or mae)
- 5) Regression Model

Model 1: Look at Price with Bedrooms and Bathrooms

Model 2: Look at log(Price) with Bedrooms and Bathrooms

To overcome heteroskedasticity with building log(Price).

- 6) Evaluate models: compare Model1 vs Model 2, look at Adjusted R squared and p-value.

compare rmse or mae

- 7) write the formula for Regression Model

Discussion about Models

Check if the objective of this project is achieved. what other factors can affect the price etc..?

Note: This is just a summary of Project Proposal. For the Project proposal Markdown, please refer to the datafile “Rental_Freq_Summary_Modelling “ file with coding, output and findings.