

# Apartment Prices in Queens

Daniel Varghese

May 25 2022

## 1 Introduction

The task of modeling apartment prices is very difficult considering the fact that many observations are made. The internet is filled with data that contains the number of bedrooms, number of bathrooms, model type, etc. that may be analyzed with the purpose of figuring out the particular price of an apartment. Considering that for some features data values are given while for others data values are missing, the task of modeling becomes even more difficult.

## 2 Data

The type of data that was used in this project was data on a .csv file from online. It was numerical data as well as data consisting of words. Observations were made and for some values there was an "NA" indicating that data values were missing or not available. The given data set was representative of Queens because it was from several neighborhoods and not just one area or part of the county. The problem with extrapolation is that the modeler would then be working beyond the data set and into the unknown which can lead to several errors.

## 3 Errors and Misgivings

A misgiving that I noticed while working on this assignment is that there are a lot of data values where there are NAs for the half bathrooms. It seems to me that there would either be a half bathroom or there would not be and that this would be pretty straight forward for a real estate agent. Maybe real estate agents intentionally leave this information out to leave the interested party wondering.

## 4 Regression Tree Modeling

The top ten layers that seem most important for predicting the price of an apartment are the number of full bathrooms, number of bedrooms, number of

floors, model type, sale price, garage, pets allowed or not, zip code, maintenance cost and a dining room.