Analysis of Property Prices in Toronto Greater Area, Canada

Project overview

- In this project, the house prices listed in the various regions of Greater Toronto Area (GTA) on the website www.realtor.ca are analyzed.
- The prices were scraped from the website on March 16, 2022.
- The data was cleaned bringing it into a suitable format facilitating the data analysis.
- Exploratory data analysis was performed on the scraped data set.
- Classification and regression machine learning models were develop to predict the property prices.

Skills demonstrated:

- Web-scraping (using Selenium package available with Python programming language.
- Data cleaning (using regex and pandas package in Python).
- Exploratory data analysis (pandas, numpy, seaborn and matplotlib packages in Python).
- Classification and regression model development and their hyper-parameter tuning using scikit-learn package in Python.

- Complete code for the project can be found in the Github repository at: Mangaljit/Toronto_Property_Prices_Analysis (github.com)

- A similar project analyzing the property prices in the city of Montreal, Canada was also developed. The details can be found at Mangaljit/Montreal_Property_Prices_Analysis (github.com)

Web scraping

- A snapshot of the python code for the web scraping is shown below.
- Complete code can be found in the GitHub repository at: [Mangaljit/Toronto_Property_Prices_Analysis (github.com)]

```
import os
import time
import AppKit
import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.common.exceptions import TimeoutException
from selenium.webdriver.support.wait import WebDriverWait

# Minimum number of clicks to scroll through different pages.
clicks = 51

# Enter the video url below.
url = "https://www.realtor.ca/"

# Regions in the Greater Toronto area to be scraped.
areas = [
        'Ajax, ON', 'Clarington, ON', 'Brock, ON', 'Oshawa, ON',
        'Pickering, ON', 'Scugog, ON', 'Uxbridge, ON', 'Whitby, ON',
        'Burlington, ON', 'Halton Hills, ON', 'Milton, ON',
        'Oakville, ON', 'Brampton, ON', 'Caledon, ON',
        'Mississauga, ON', 'Aurora, ON', 'East Gwillimbury, ON',
        'Georgina, ON', 'King, ON', 'Markham, ON', 'Newmarket, ON',
        'Richmond Hill, ON', 'Vaughan, ON', 'Whitchurch-Stouffville, ON',
        'Old Toronto, Toronto, ON', 'Hamilton, ON', 'Guelph, ON',
        'Kitchener, ON', 'Cambridge, ON', 'Brantford, ON', 'Scarborough, ON'
    ]


# Loop through different regions to scrape property listing prices and other relevant details.
for area in areas:
    driver = webdriver.Chrome()
    print('Opening the Browser.')
    time.sleep(4)
    driver.get(url)
    print('Browser opened the requested url.')
    AppKit.NSBeep()
    print('Waiting for the manual captcha entering by the user.')
```

A quick look at the dataset

price	region	address	bedrooms	bathrooms	pricem
799000	Ajax, ON	2 ROLLO DR, Ajax, Ontario	3	3	0.799
989000	Ajax, ON	717 OLD HARWOOD AVE, Ajax, Ontario	2	1	0.989
999900	Ajax, ON	52 ADDLEY CRES, Ajax, Ontario	3	4	0.9999
799900	Ajax, ON	249 MONARCH AVE, Ajax, Ontario	3	3	0.7999
899999	Ajax, ON	18 MONK CRES, Ajax, Ontario	3	3	0.899999
...	...	...	...	...	...
1899000	Scarborough, Toronto, ON	#MAIN UN -2977 LAWRENCE AVE E AVE, Toronto, On...	8	4	1.899
752880	Scarborough, Toronto, ON	#406 -3220 SHEPPARD AVE E, Toronto, Ontario	2	2	0.75288
1150000	Scarborough, Toronto, ON	#902 -2799 KINGSTON RD, Toronto, Ontario	3	3	1.15
550000	Scarborough, Toronto, ON	#506 -2201 KINGSTON RD, Toronto, Ontario	1	1	0.55
1820000	Scarborough, Toronto, ON	103 SLAN AVE, Toronto, Ontario	7	4	1.82

The dataset has a total of 7324 rows and 6 columns.

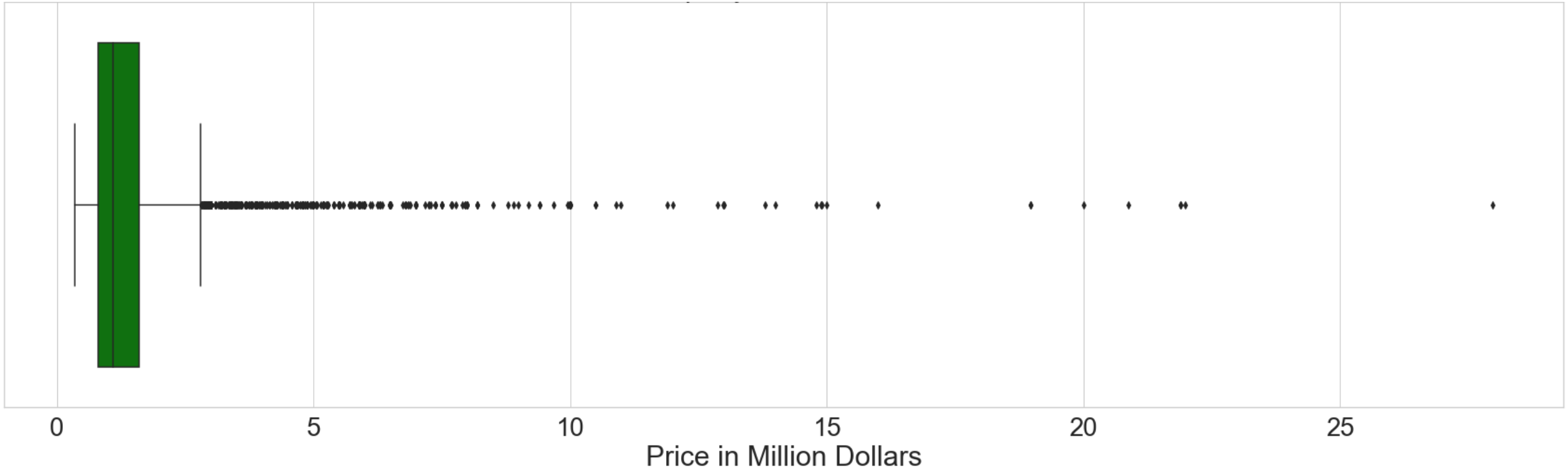
There are 31 regions in the scraped data set.

Region
King
Caledon
Oakville
Richmond Hill
East Gwillimbury
Aurora
Whitchurch-Stouffville
Markham
Vaughan
Halton Hills
Uxbridge
Milton
Newmarket
Brampton
Scugog
Burlington
Whitby
Pickering
Ajax
Mississauga
Georgina
Scarborough Toronto
Clarington
Oshawa
Brock
Guelph
Old Toronto Toronto
Brantford
Hamilton
Cambridge
Kitchener

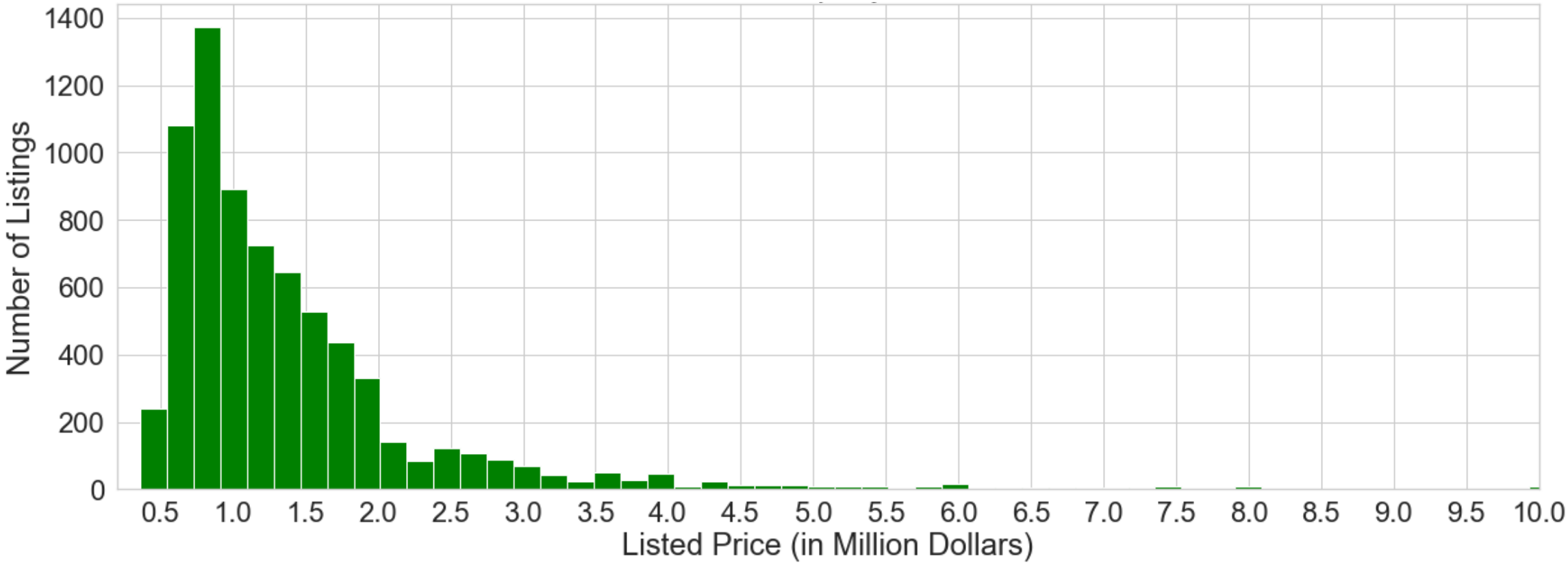
Median property price in different regions

region	median price in million dollars	no. of listings
King	2.969	56
Caledon	1.899	132
Oakville	1.79945	422
Richmond Hill	1.6985	382
East Gwillimbury	1.6885	92
Aurora	1.688	130
Whitchurch-Stouffville	1.499	98
Markham	1.39945	482
Vaughan	1.399	478
Halton Hills	1.397	80
Uxbridge	1.314	18
Milton	1.2999	215
Newmarket	1.299	139
Brampton	1.199	597
Scugog	1.1749	16
Burlington	1.1499	306
Whitby	1.099	130
Pickering	0.99995	84
Ajax	0.999	114
Mississauga	0.999	579
Georgina	0.999	118
Scarborough Toronto	0.988	414
Clarington	0.899999	127
Oshawa	0.899	234
Brock	0.899	26
Guelph	0.898988	186
Old Toronto Toronto	0.839	444
Brantford	0.799999	131
Hamilton	0.7999	579
Cambridge	0.7999	163
Kitchener	0.7499	297

Spread of property prices in GTA

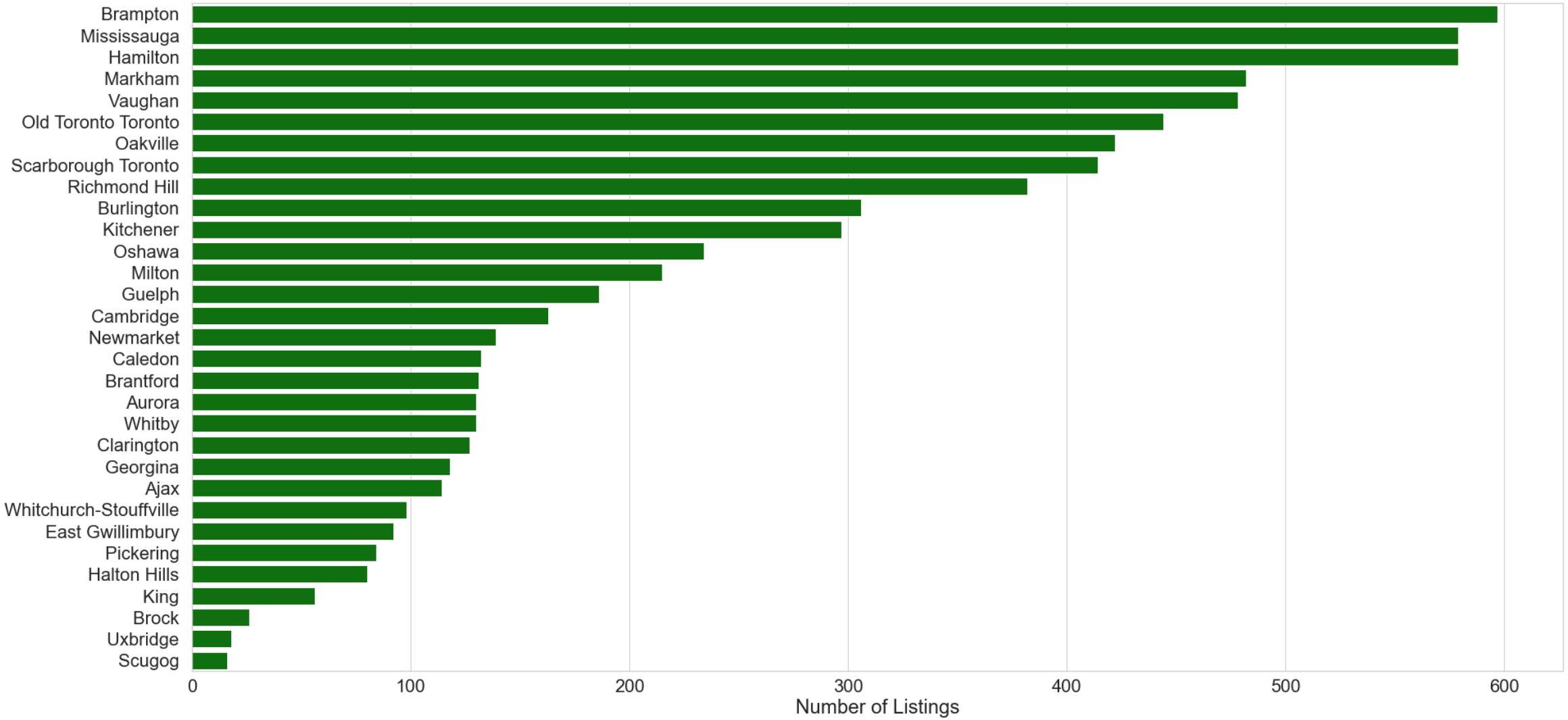


Distribution of Property prices in GTA



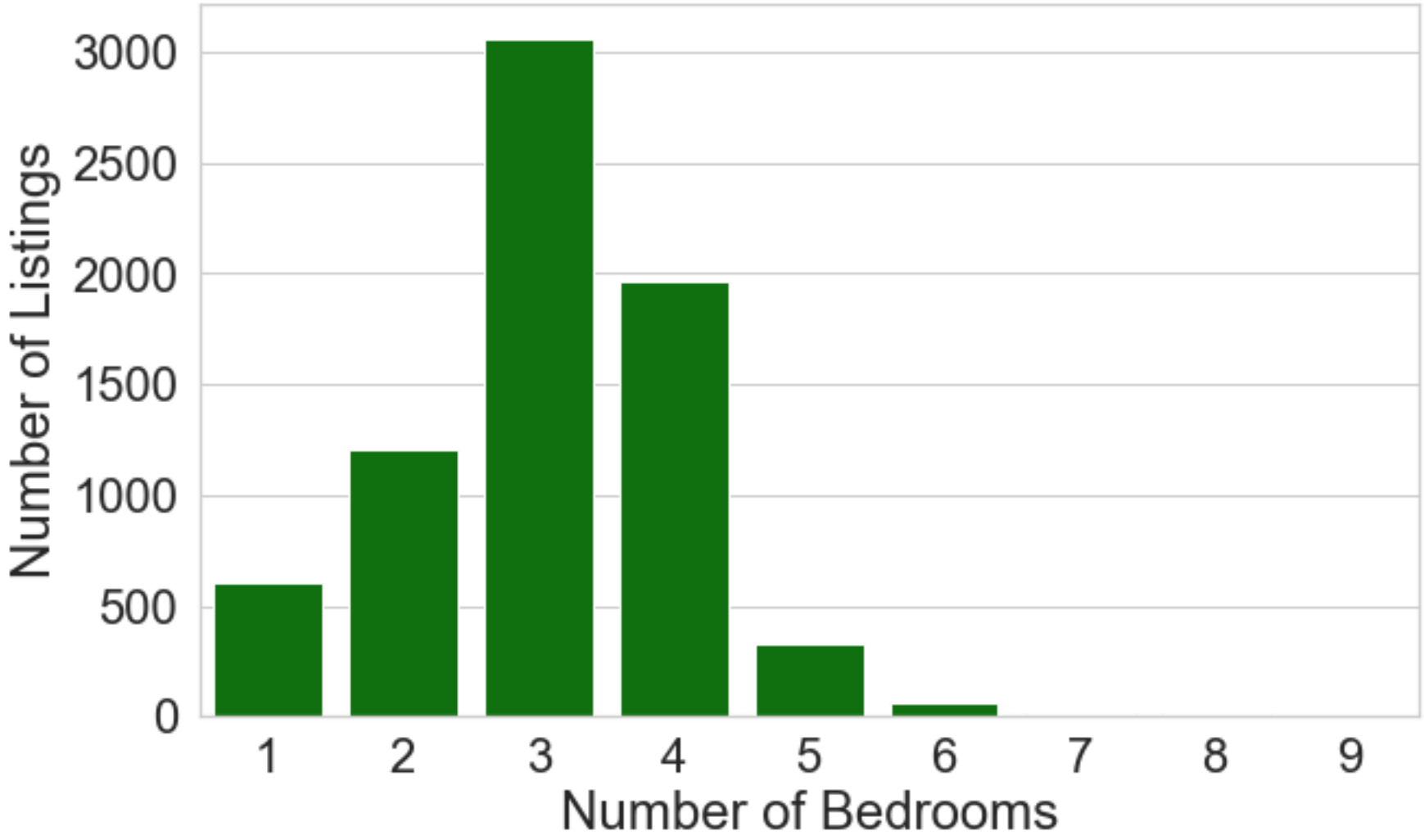
Above figure shows that majority of the property prices in GTA lie in a range between 0.5-2.0 million dollars.

Number of listings in different regions

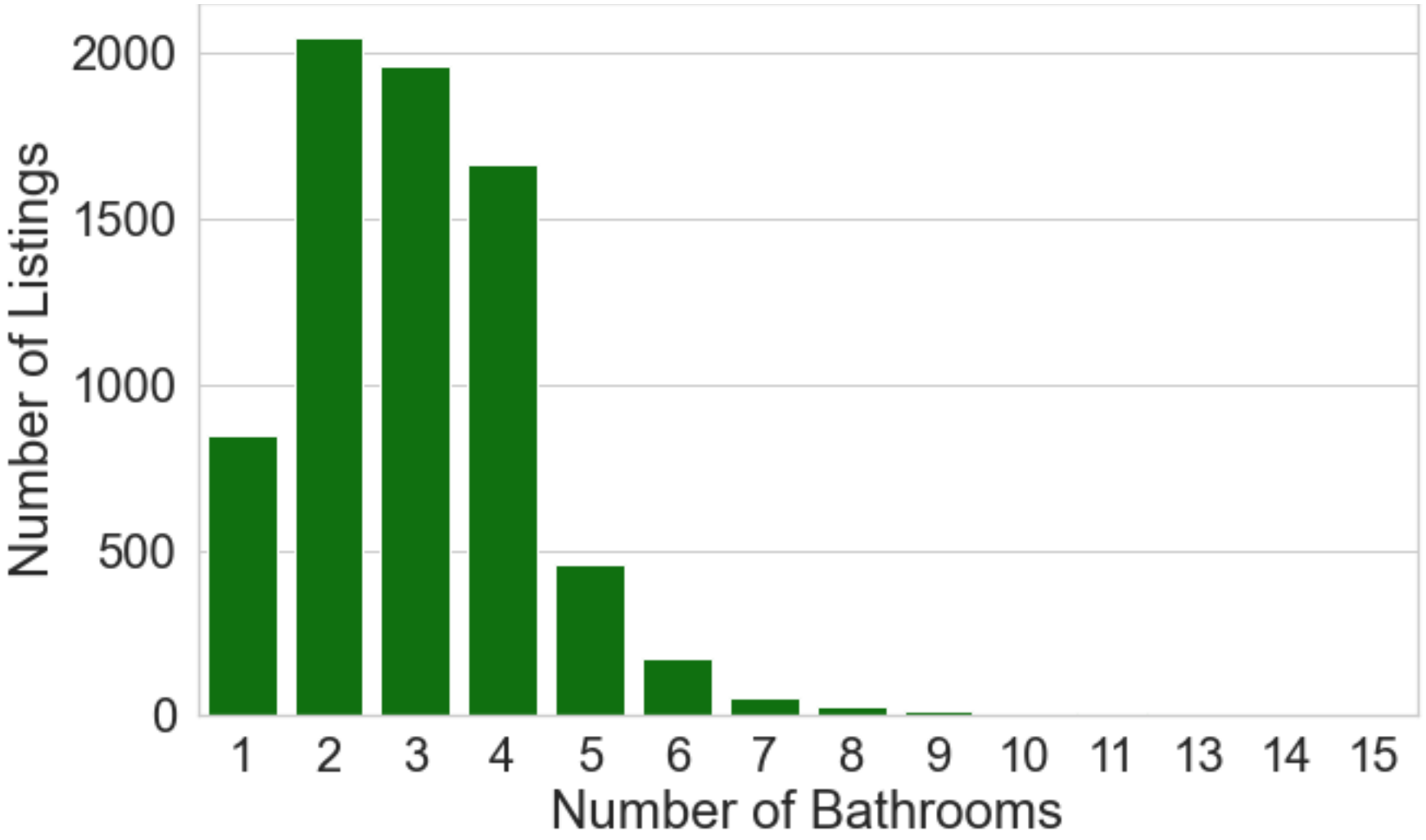


Brampton , Mississauga , Hamilton , Markham and Vaughan are regions with one of the highest number of listings.

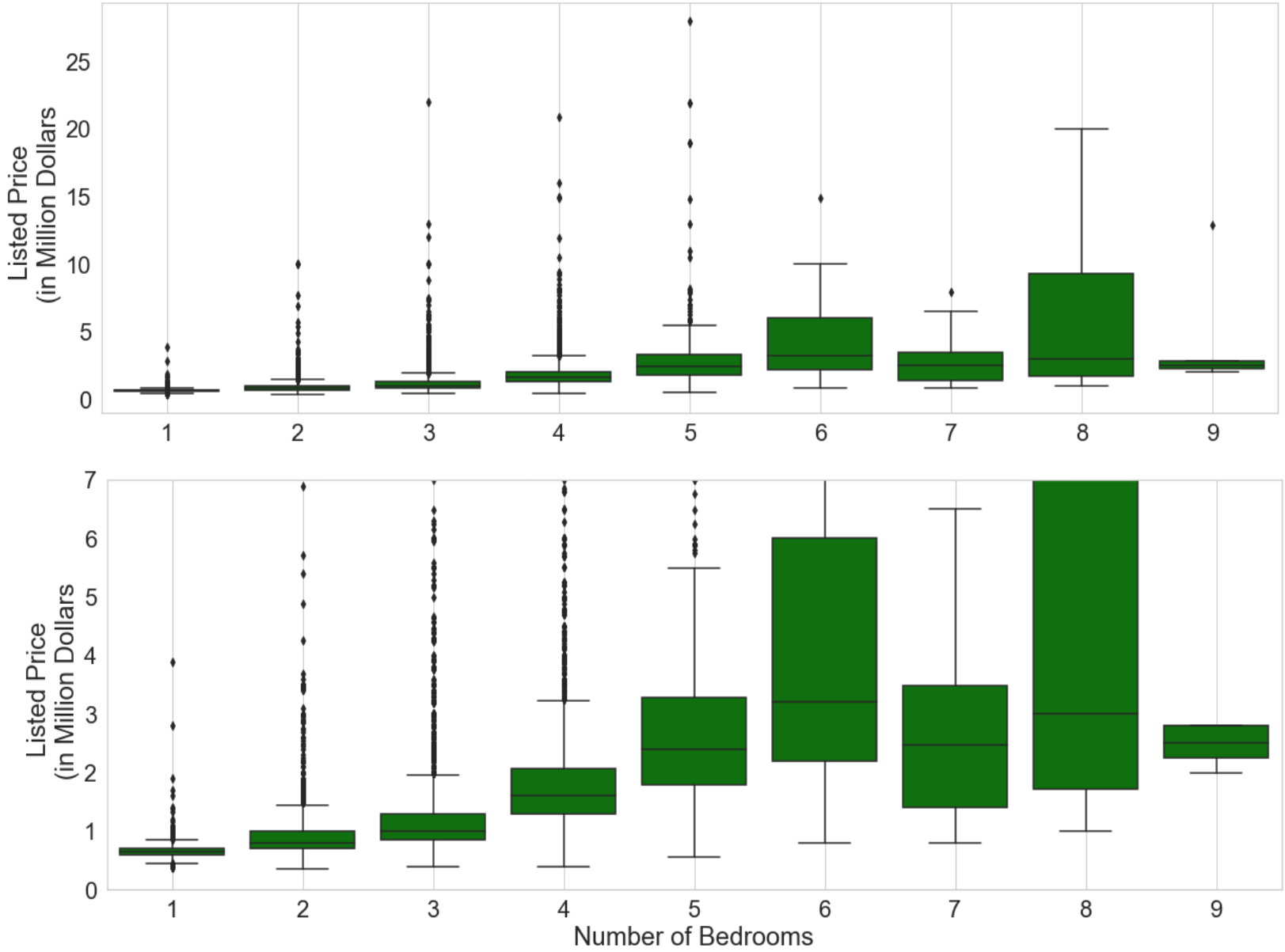
Number of property listings with a given number of bedrooms



Number of property listings with a given number of bathrooms

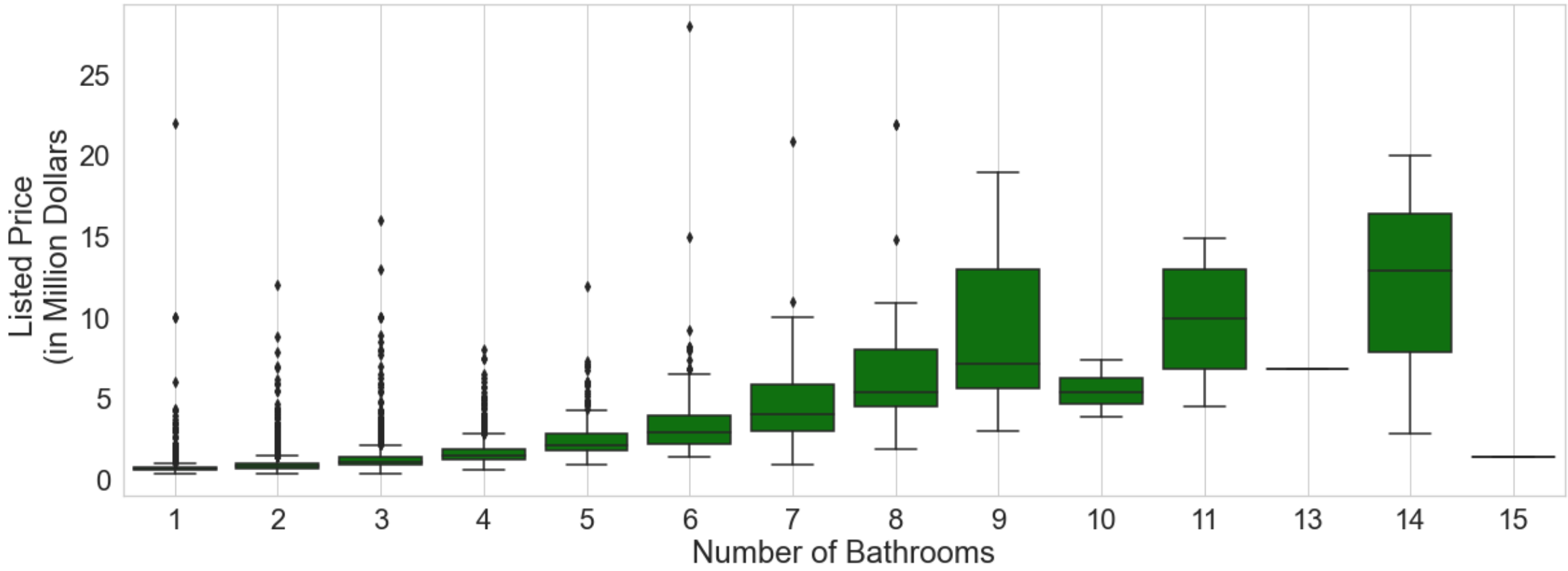


Distribution of property prices for a given number of bedrooms



The median price, indicated by the horizontal line in the box increases with the increase in the number of bedrooms a property contains.

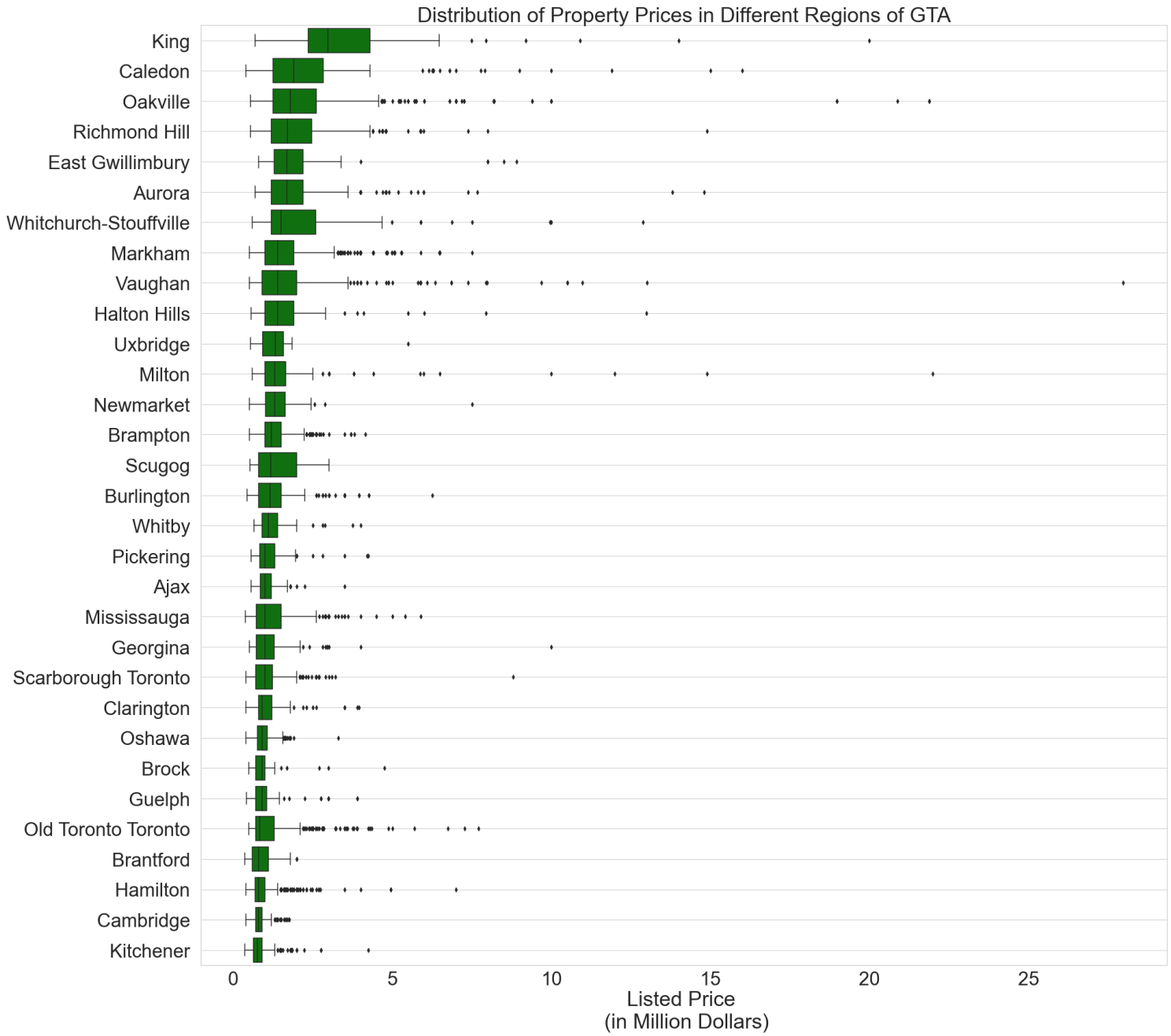
Distribution of property prices for a given number of bathrooms



The median price also increases with the increase in the number of bathrooms a property contains.

Distribution of property prices in different regions

The exploratory data analysis shows that the number of bedrooms and number of bathrooms a property contains, as well as the region where a property is located are important factors in determining the listed price of the property.

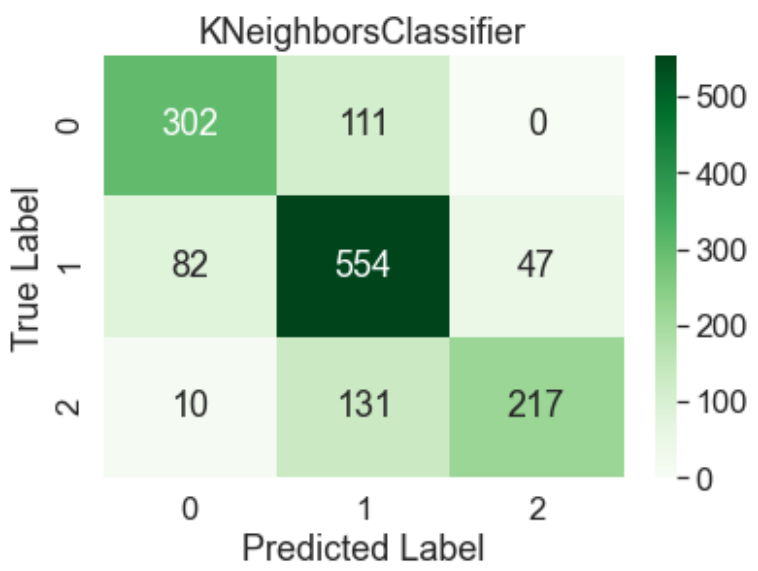


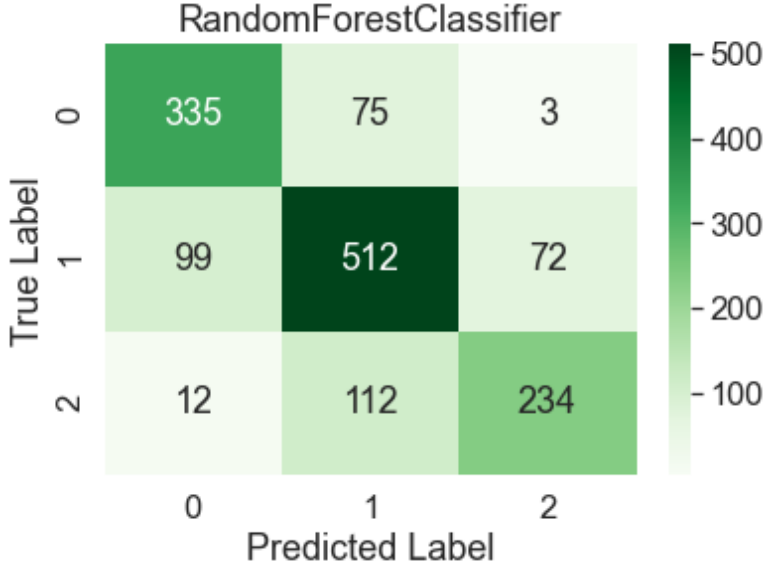
Regression model predicting the property prices in GTA

- Both linear as well as the second order polynomial regression models were developed.
  - Linear regression model delivered a mean absolute accuracy of 0.244 million dollars on the testing data set.
  - Polynomial regression model delivered a mean absolute accuracy of 0.233 million dollars on the testing data set.

Classification model predicting the property prices in GTA

- For the classification model, the entire dataset was categorized into three levels. The first level, assigned with a value of 0 indicates a property price less that 0.8 million dollars. The second level, assigned with a value of 1 will indicates a property price in the range of 0.8-1.6 million dollars. And finally, the third level, assigned with a value of 2 will indicate a property price greater than 1.6 million dollars.
- K-nearest neighbors and Random forest model was developed, delivering an overall accuracy of 0.7379 and 0.7434 respectively.





- Complete code and the detailed methodologies can be found in the GitHub repository at: [Mangaljit/Toronto_Property_Prices_Analysis (github.com)]

End of presentation