

Web scraping Project example – Royal LePage

Tianyuan Xu
Xinzhen Fan



Contents



PART ONE
Motivation



PART TWO
Data



PART THREE
Visualizations



PART FOUR
Conclusions



PART FIVE
Challenges



Background



- **Industry:** Real estate
- **Founder:** Albert E. LePage
- **Founded:** Toronto, Ontario, Canada (July 2, 1913)
- **Products:** Real estate brokerage services
- **Headquarter:** North York, Ontario, Canada
- **Number of employees:** 20000+
- **Number of locations:** 650+





Motivation

Properties in greater Vancouver

Property Type	Bedrooms	Bathrooms	Price	Address	Location	Agent
House	6	2	\$2,780,000	300 Balmoral Road	North Vancouver, BC	Courtesy of: Vanak Realty
Condo	2	3	\$2,697,000	3202 1111 Alberni Street	Vancouver, BC	Courtesy of: eXp Realty
House	4	4	\$2,620,000	5625 Daffodil Lane	West Vancouver, BC	Courtesy of: RE/MAX Masters Realty
House	6	3	\$2,499,999	57 Glenmore Drive	West Vancouver, BC	Courtesy of: Royal LePage Sussex
Condo	3	3	\$2,498,000	203 3101 Burfield Place	West Vancouver, BC	Courtesy of: Sutton Group-West Coast Realty
Condo	2	3	\$2,449,900	1603 1365 Davie Street	Vancouver, BC	Courtesy of: Engel & Volkers Vancouver

- Aspect of data scraping
 - Clear and enough data
 - Data scraping is not blocked
- Aspect of interest
 - Famous real estate company
 - Necessary and important for people
 - Housing price increases fast



Data scraping

```
## Define function to go to every page
def get_page(page):
    url = f'https://www.royallepage.ca/en/search/homes/bc/vancouver/{page}/?property_type=&house_t:

    result = requests.get(url)
    soup = BeautifulSoup(result.content)
    return soup

## set up for loop to collect data in page in set up range
for page_num in range(1,31):
    soup = get_page(page_num)

    ## types
    for tag in soup.find_all('div', class_='listing-meta listing-meta--small'):
        try:
            details['types'].append(tag.span.get_text(strip=True))
        except:
            details['types'].append('missing')

    ## bedrooms number
    for tag in soup.find_all('div', class_='listing-meta listing-meta--small'):
        try:
            details['bds'].append(int(list(tag.children)[3].get_text(strip = True)[0]))
        except:
            details['bds'].append(0) ## some property doesnt have bedroom like comercial
```

```
for tag in soup.find_all('div', class_='listing-meta listing-meta--small'):
    try:
        details['bths'].append(int(list(tag.children)[3].get_text(strip = True).replace(' ', '')[-6]))
    except:
        details['bths'].append(0) ## some property doesnt have bathrooms like comercial

    ## price
    for tag in soup.find_all('span', class_='title--h3 price'):
        try:
            details['price'].append(tag.span.get_text(strip=True))
        except:
            details['price'].append(0)

    ## address
    for tag in soup.find_all('address', class_='address-1'):
        try:
            details['address'].append(tag.get_text(strip=True))
        except:
            details['address'].append('missing')

    ## region
    for tag in soup.find_all('address', class_='card_address-2'):
        try:
            details['region'].append(tag.get_text(strip = True).split(',')[0])
        except:
            details['region'].append('missing')

    ## belong to which realty company
    for tag in soup.find_all('div', class_='card_body'):
        try:
            details['realty'].append(list(tag.children)[8].get_text(strip = True).split(':')[1])
        except:
            details['realty'].append('missing')
```



Dataframe – raw data

	types	bds	bths	price	address	region	realty
0	House	2	3	\$1,799,000	4089 Ash Street	Vancouver	Macdonald Realty
1	House	5	3	\$2,758,000	3115 Benbow Road	West Vancouver	Sutton Group-West Coast Realty, Sotheby's Int...
2	House	3	3	\$1,349,900	1215 E 27th Avenue	Vancouver	Macdonald Realty
3	Commercial	0	0	\$418,000	10506 Confidential	Vancouver	SaveEditRemoveView
4	Commercial	0	0	\$178,000	3364 Cambie Street	Vancouver	SaveEditRemoveView
...
1375	Condo	1	1	\$575,000	1003 550 Taylor Street	Vancouver	Faith Wilson Group
1376	Condo	1	1	\$499,000	304 1330 Marine Drive	North Vancouver	Coldwell Banker Prestige Realty
1377	Commercial	0	0	\$299,000	6459 Victoria Drive	Vancouver	SaveEditRemoveView
1378	House	6	8	\$14,980,000	1337 Cambridge Road	West Vancouver	Sutton Group-West Coast Realty
1379	House	7	7	\$9,850,000	1466 Bramwell Road	West Vancouver	VIRANI REAL ESTATE ADVISORS

1380 rows × 7 columns





Dataframe – cleaned data

	types	bds	bths	price(million)	address	region	realty
0	House	2	3	1.7990	4089 Ash Street	Vancouver	Macdonald Realty
1	House	5	3	2.7580	3115 Benbow Road	West Vancouver	Sutton Group-West Coast Realty
2	House	5	3	2.7580	3115 Benbow Road	West Vancouver	Sotheby's International Realty Canada
3	House	3	3	1.3499	1215 E 27th Avenue	Vancouver	Macdonald Realty
4	Commercial	0	0	0.4180	10506 Confidential	Vancouver	NaN
...
1058	Condo	1	1	0.5750	1003 550 Taylor Street	Vancouver	Faith Wilson Group
1059	Condo	1	1	0.4990	304 1330 Marine Drive	North Vancouver	Coldwell Banker Prestige Realty
1060	Commercial	0	0	0.2990	6459 Victoria Drive	Vancouver	NaN
1061	House	6	8	14.9800	1337 Cambridge Road	West Vancouver	Sutton Group-West Coast Realty
1062	House	7	7	9.8500	1466 Bramwell Road	West Vancouver	VIRANI REAL ESTATE ADVISORS

1063 rows × 7 columns





Data visualization

□ Number distribution

- Different types
- Different regions
- Realty companies

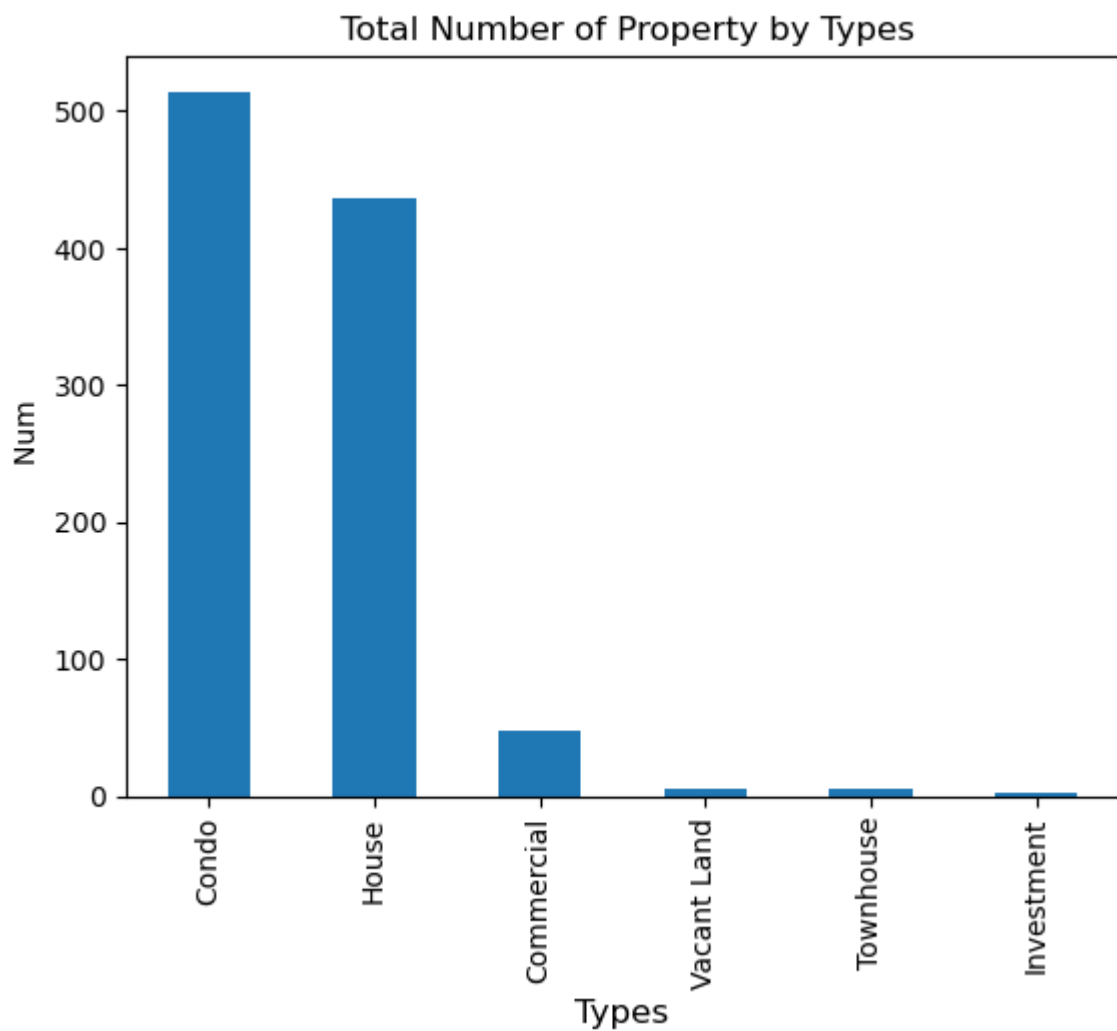
□ Price

- House
- condo

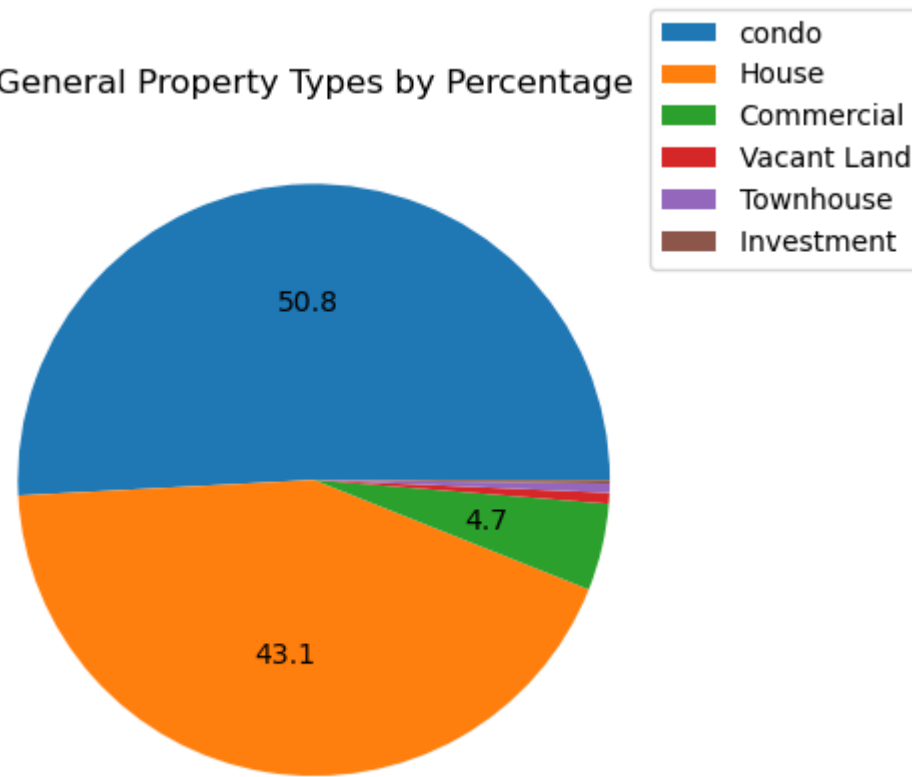




Different types

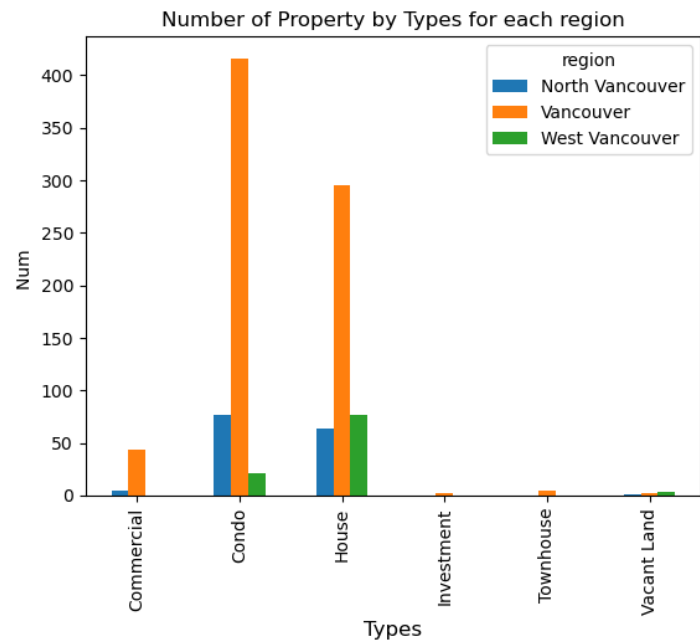


General Property Types by Percentage

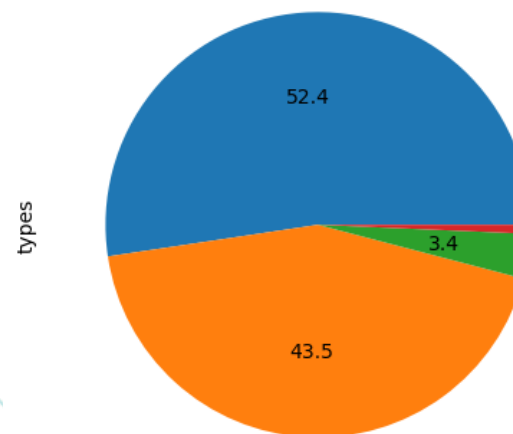




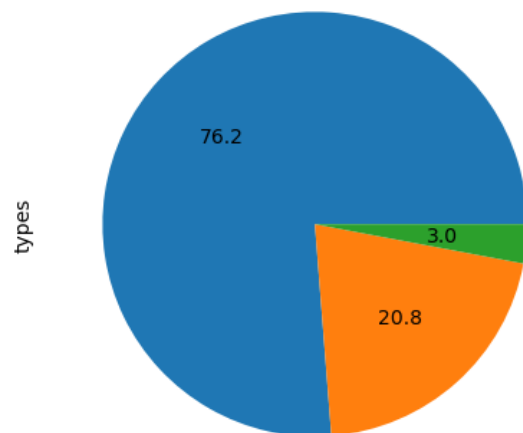
Different regions



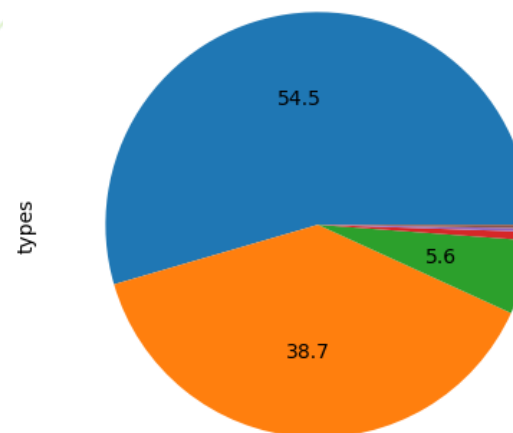
North Vancouver Property Types by Percentage



West Vancouver Property Types by Percentage



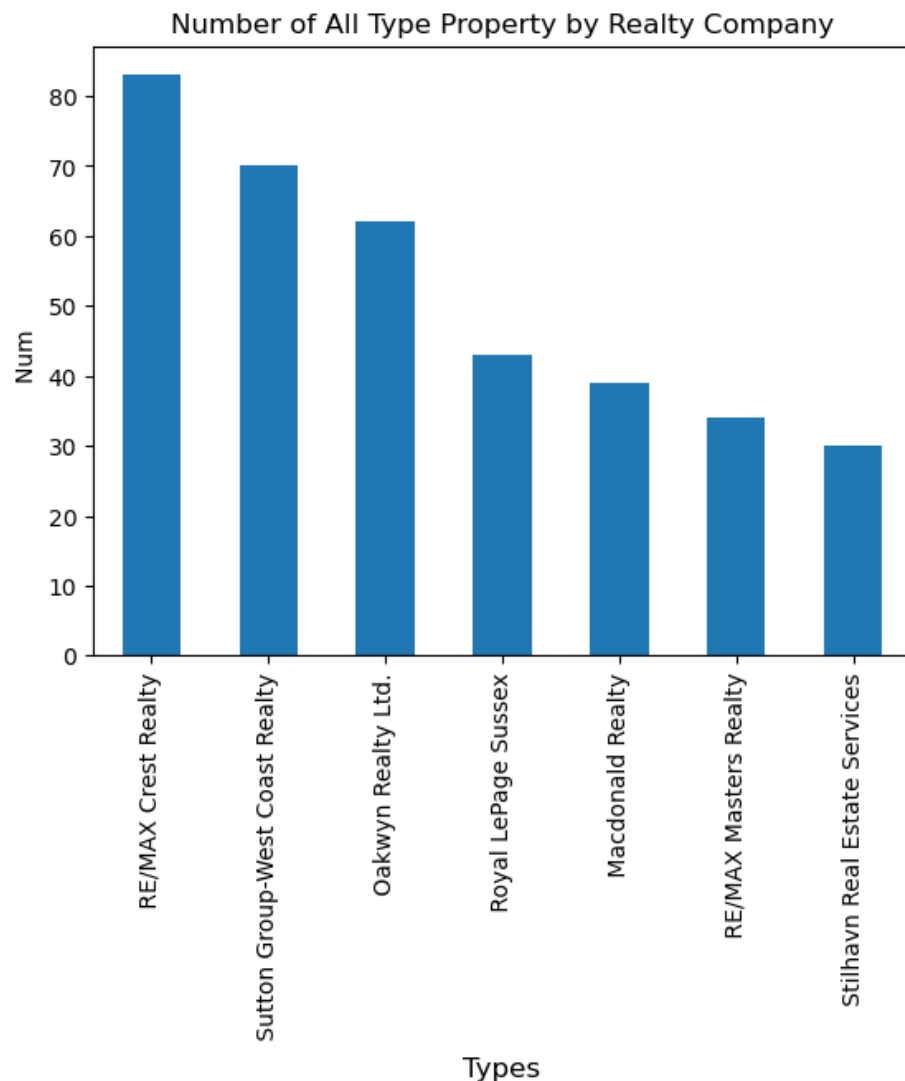
Vancouver Property Types by Percentage





Realty companies

Realty companies have more than 30 properties

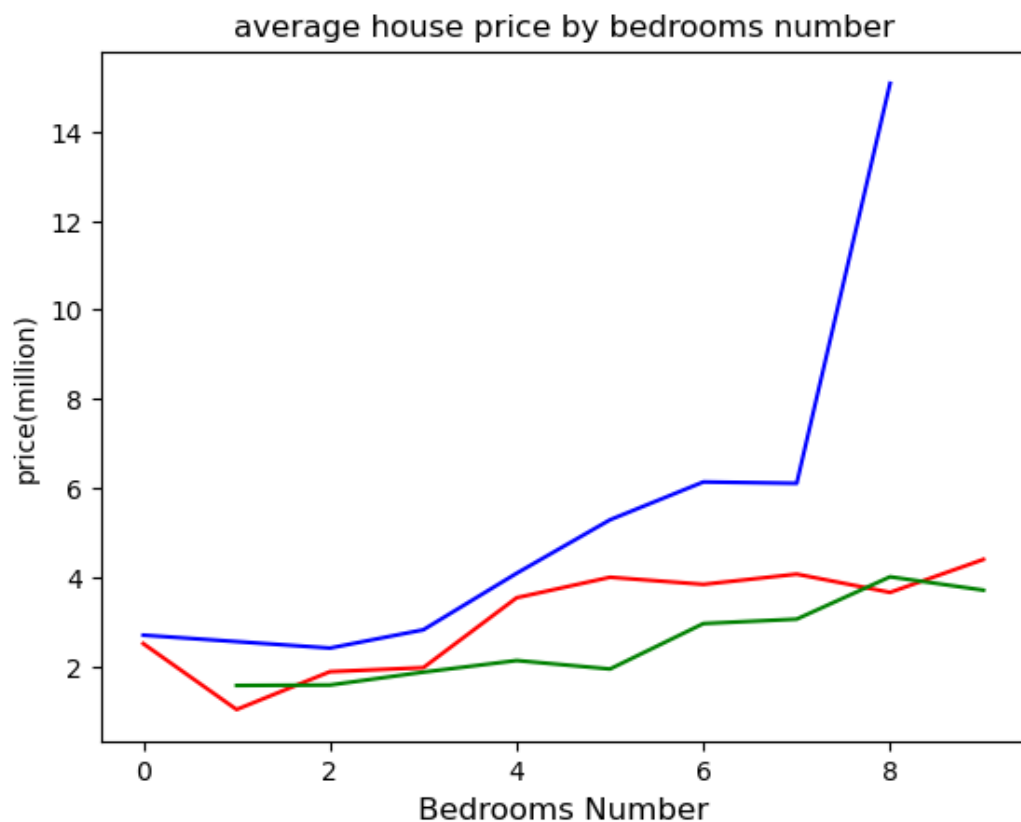


```
RE/MAX Crest Realty      83
Sutton Group-West Coast Realty  70
Oakwyn Realty Ltd.      62
Royal LePage Sussex     43
Macdonald Realty        39
..
Planet Group Realty Inc.  1
Swift Realty Ltd.        1
Oneflatfee.ca           1
Macdonald Realty (Surrey/152)  1
Birds Nest Properties    1
Name: realty, Length: 133, dtype: int64
```

- House and condo are the majority in all regions (V, WV, NV)
- The region Vancouver has all property types
- 133 realty companies but only 7 of them have more than 30 properties



House price – bedrooms number



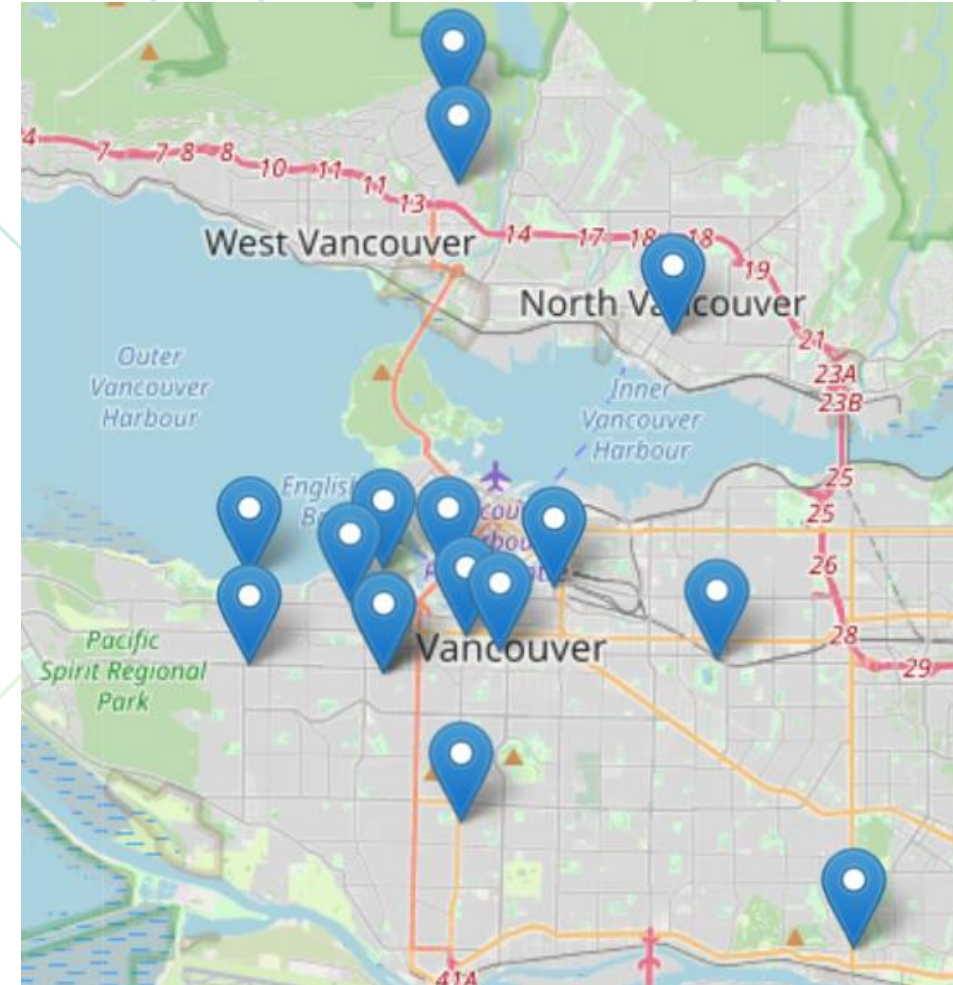
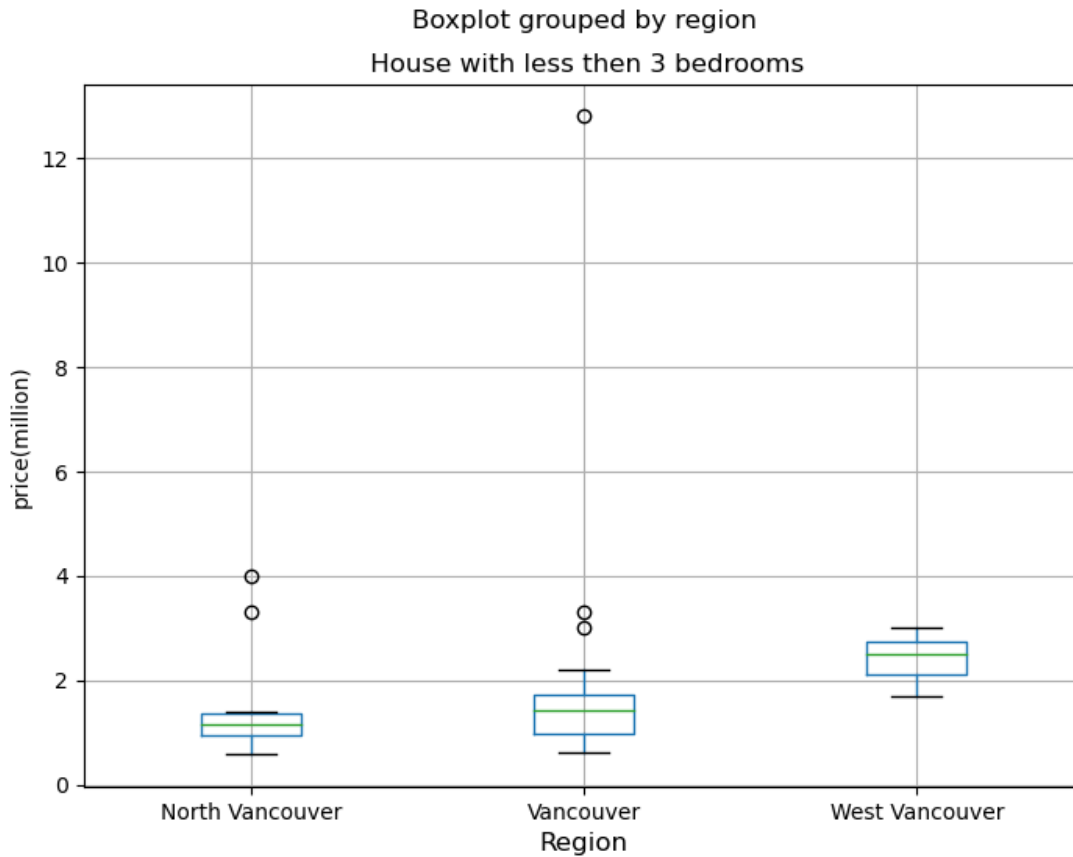
— Vancouver
— North Vancouver
— West Vancouver

- House price increases with number of bedrooms
- House in West Vancouver is more expensive than other regions
- 0 means the number of bedrooms is not indicated



House price – less than 3 bedrooms

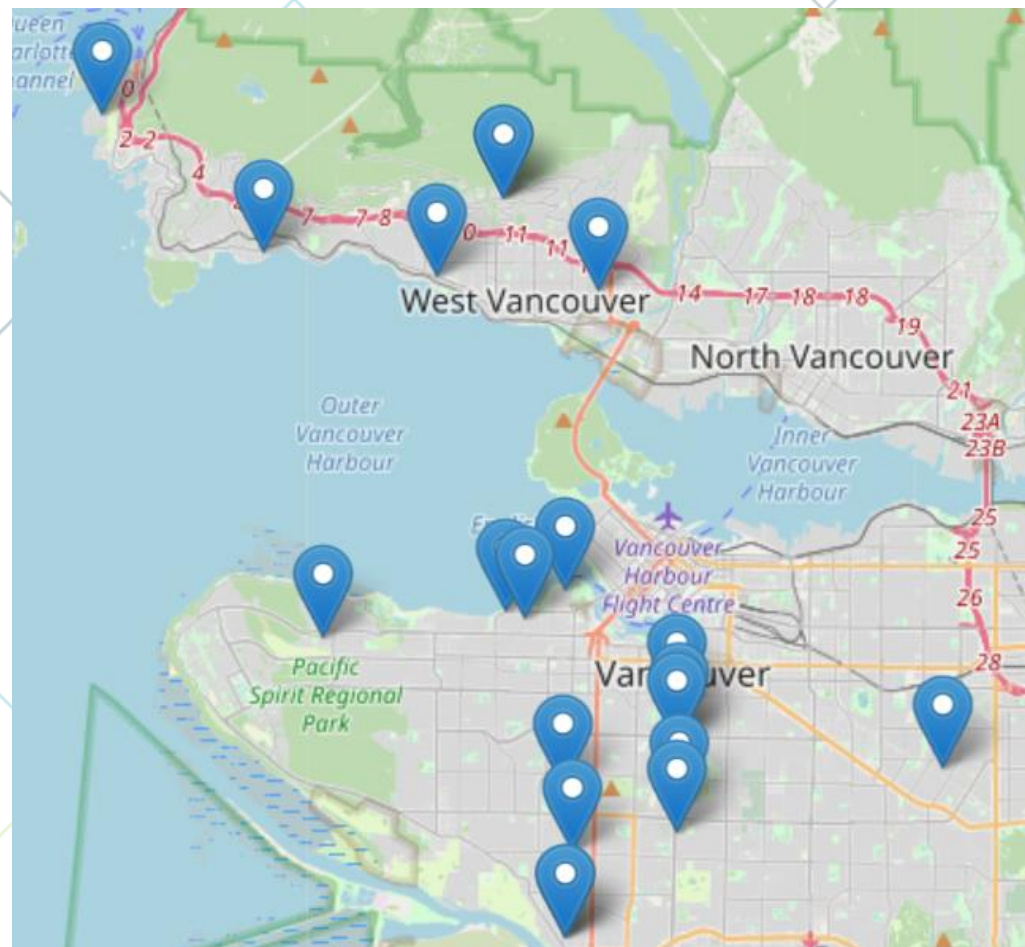
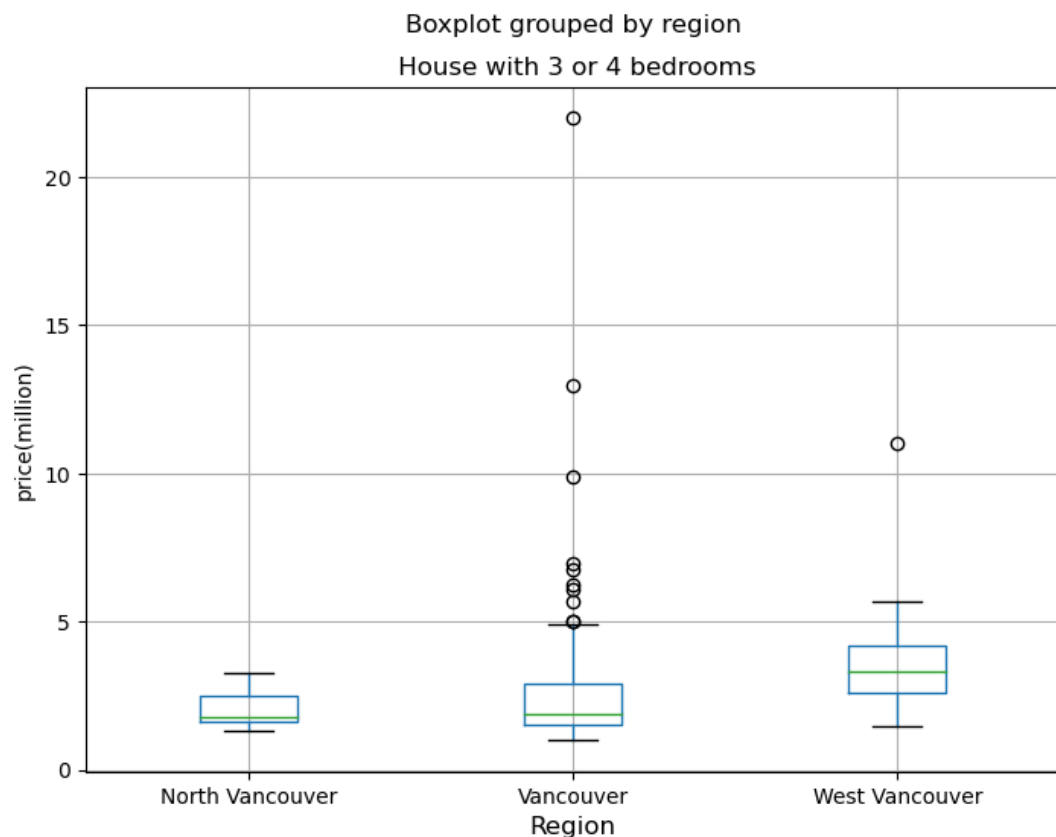
Top 20 price of less than 3 bedrooms house in all region





House price – 3 or 4 bedrooms

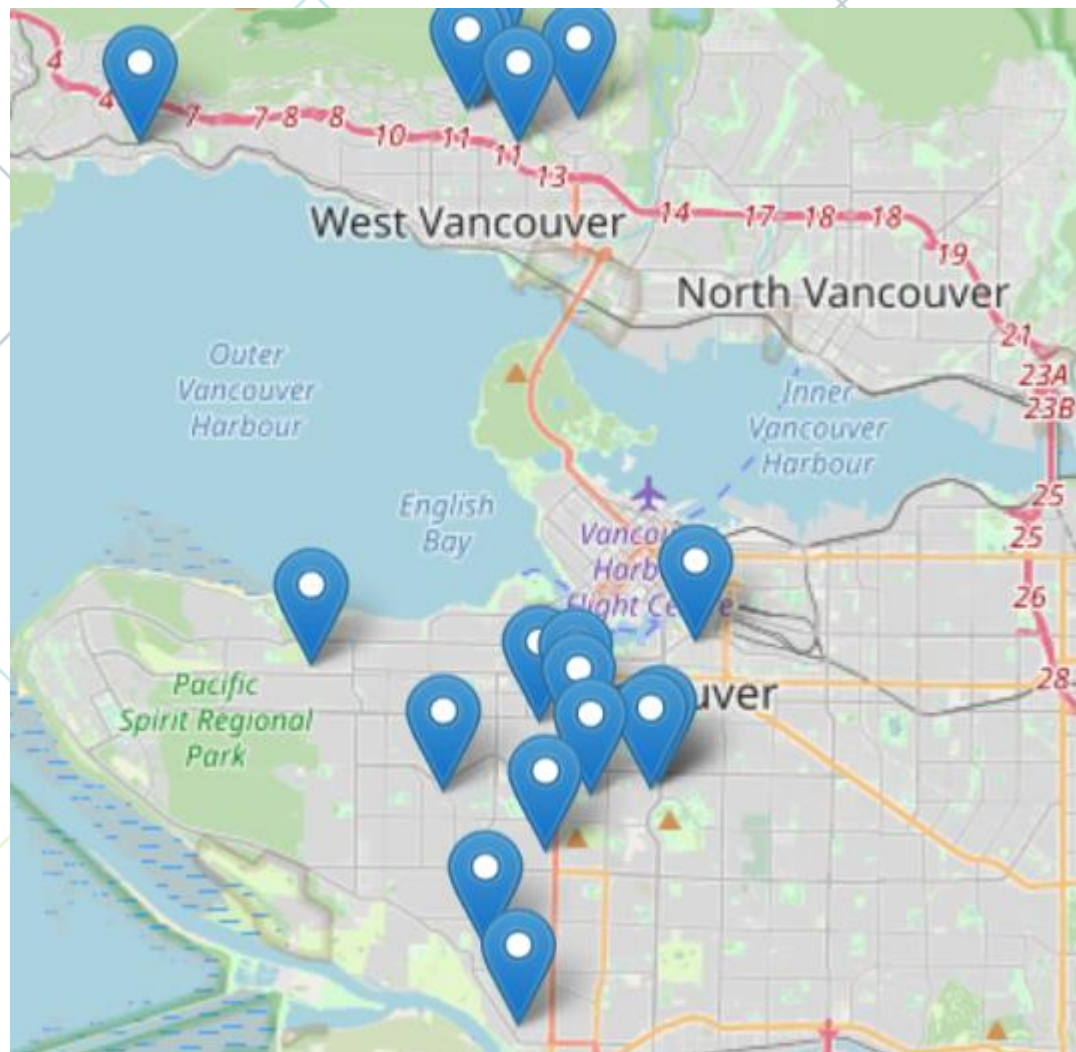
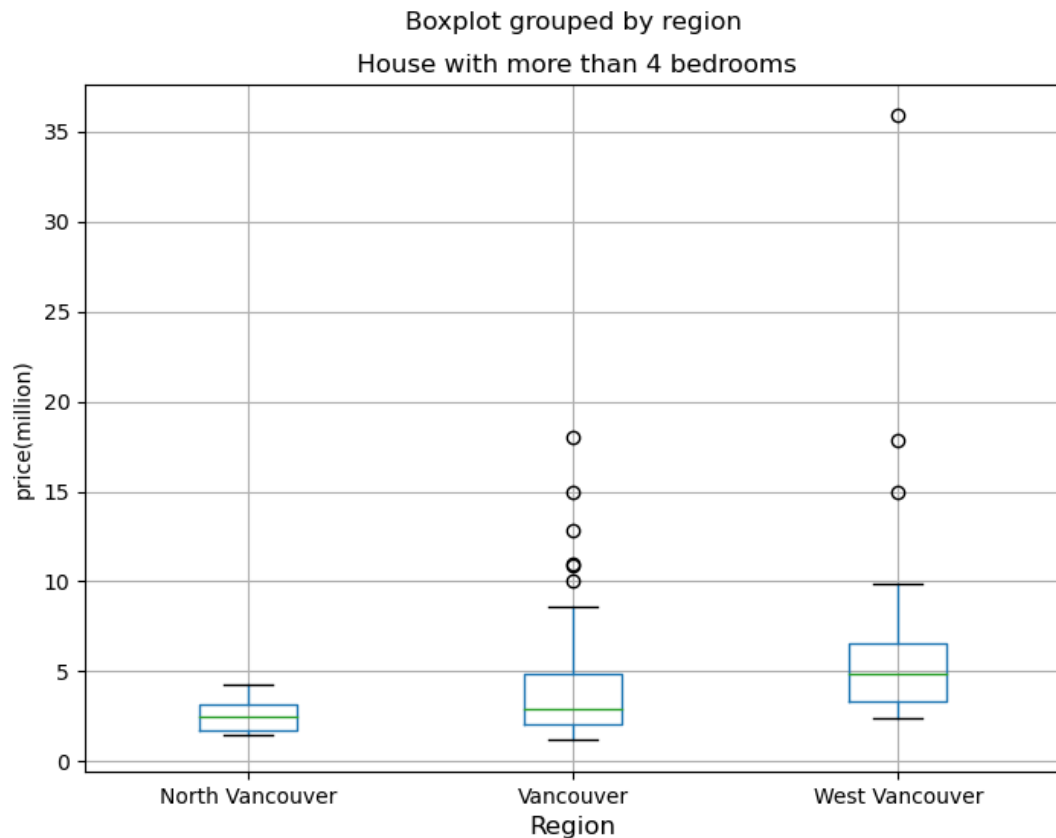
Top 20 price of 3 or 4 bedrooms house in all region





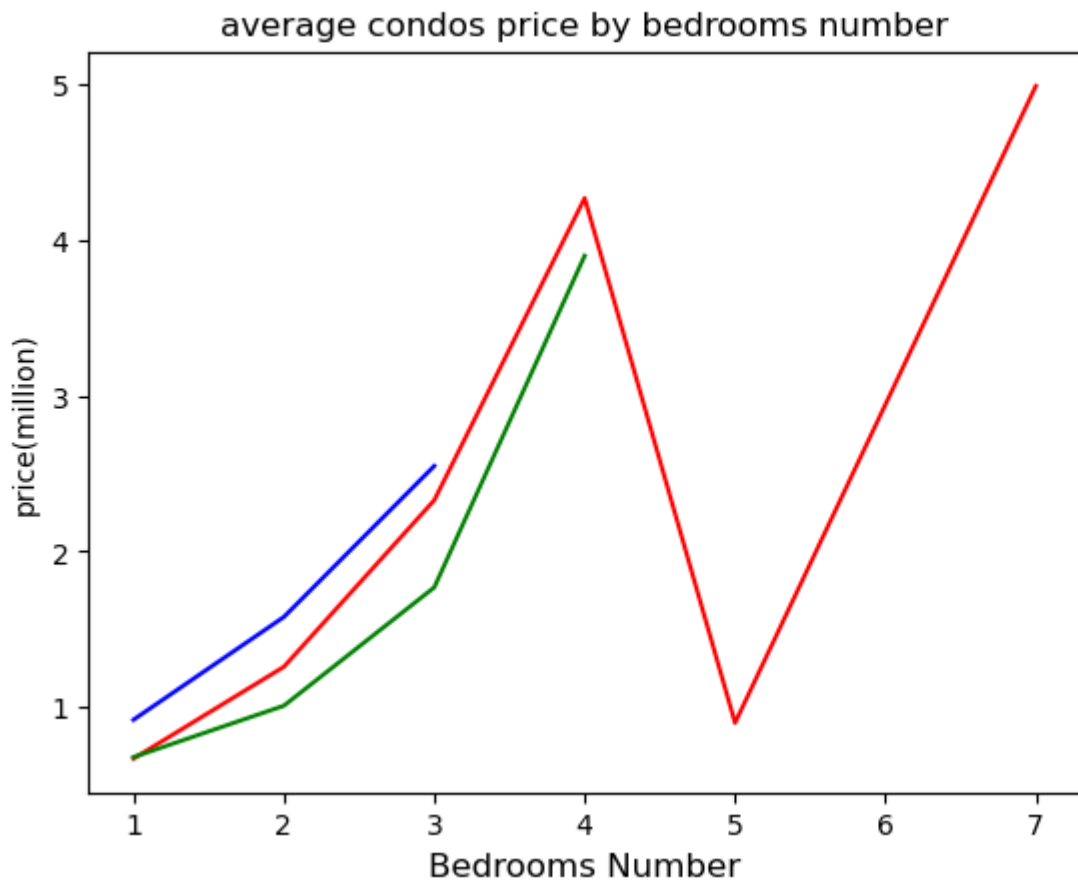
House price – more than 4 bedrooms

Top 20 price of more than 4 bedrooms house in all region





Condo price – bedrooms number

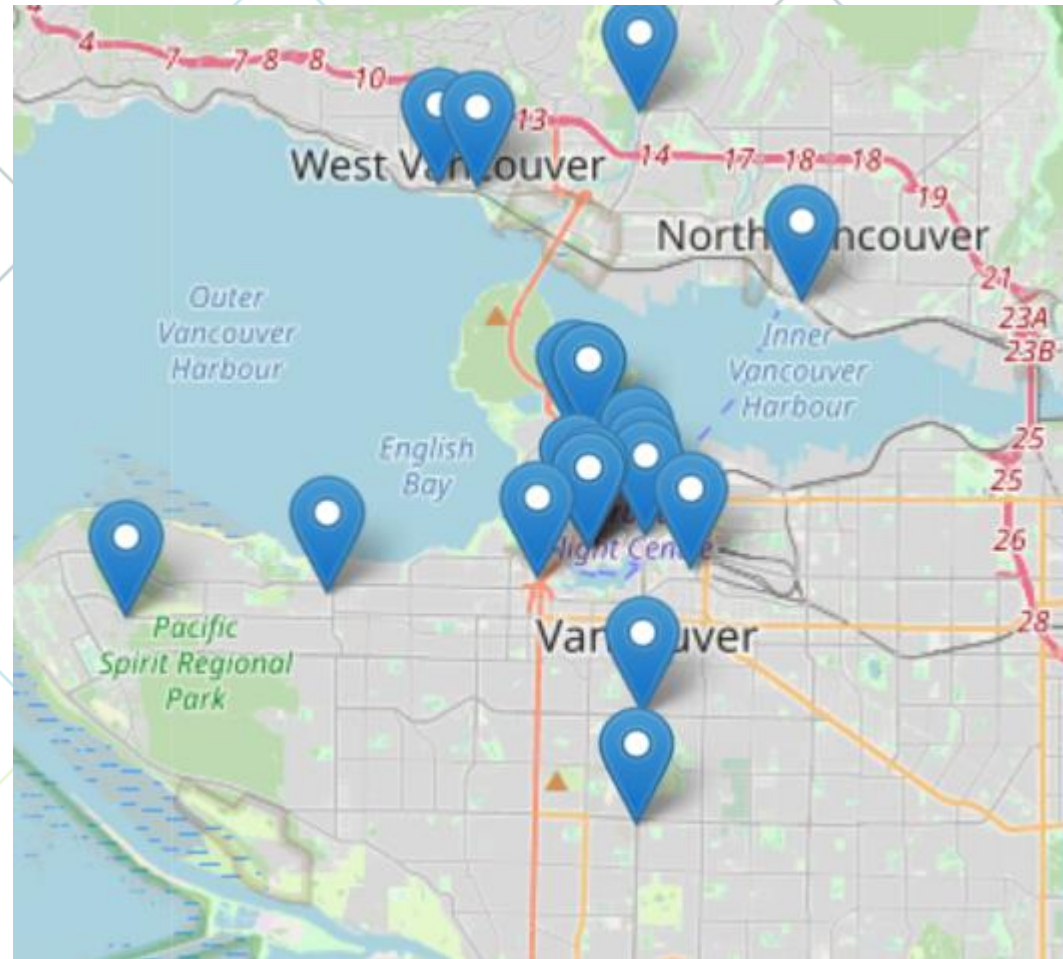
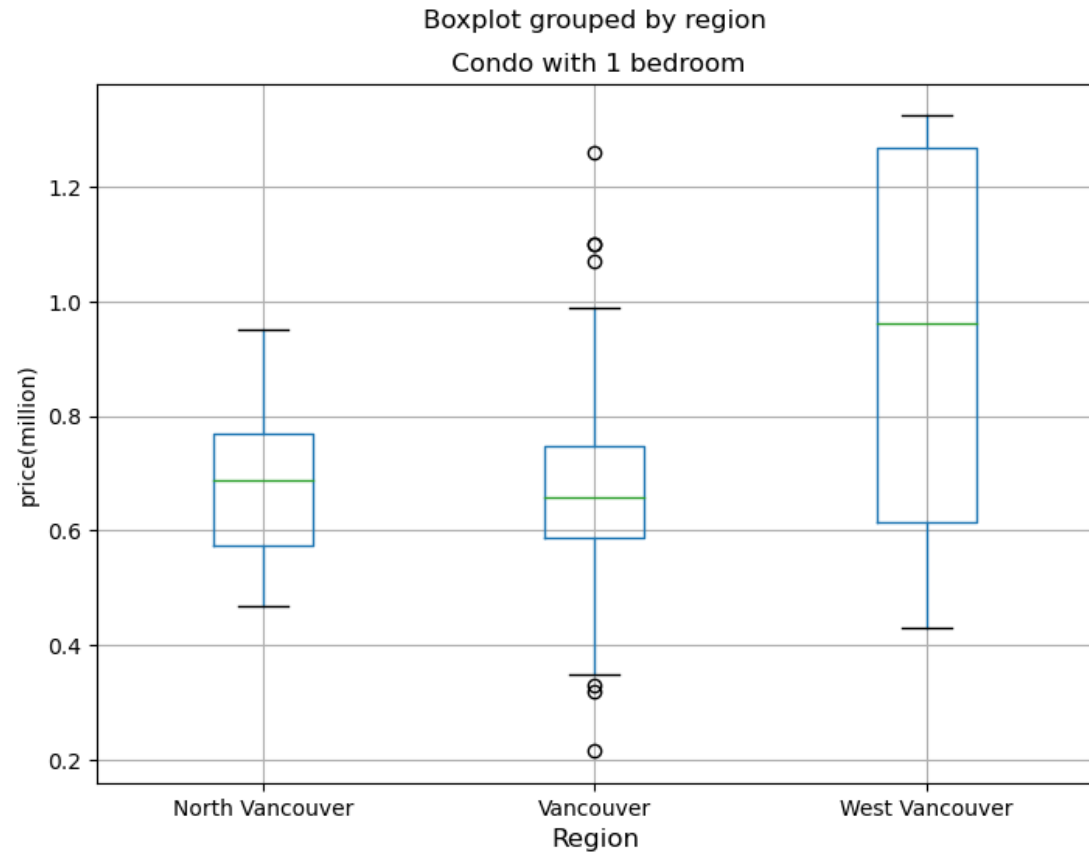


- Only the region Vancouver has condo with more than 4 bedrooms
- Condo price increases with number of bedrooms (except 5 bedrooms in Vancouver)
- Condo in West Vancouver is more expensive than other regions



Condo price – 1 bedroom

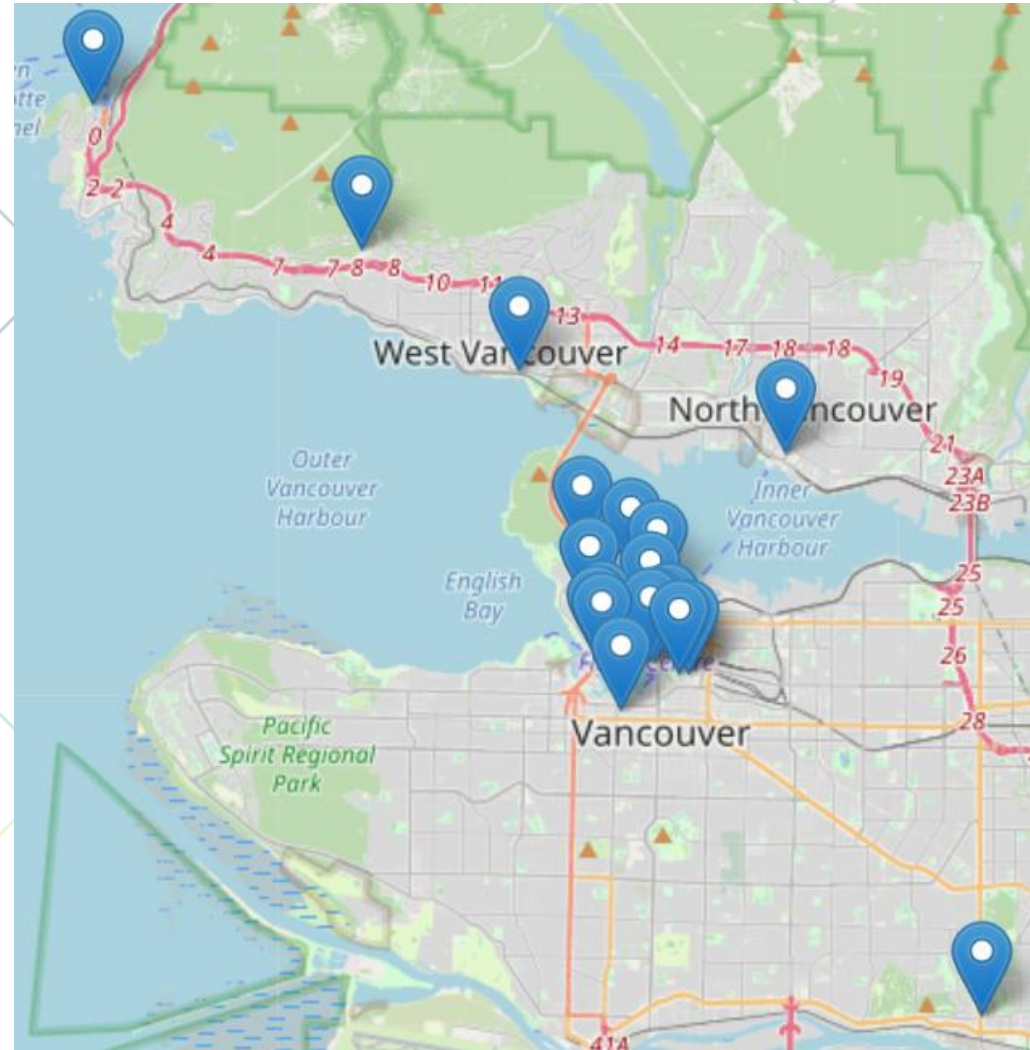
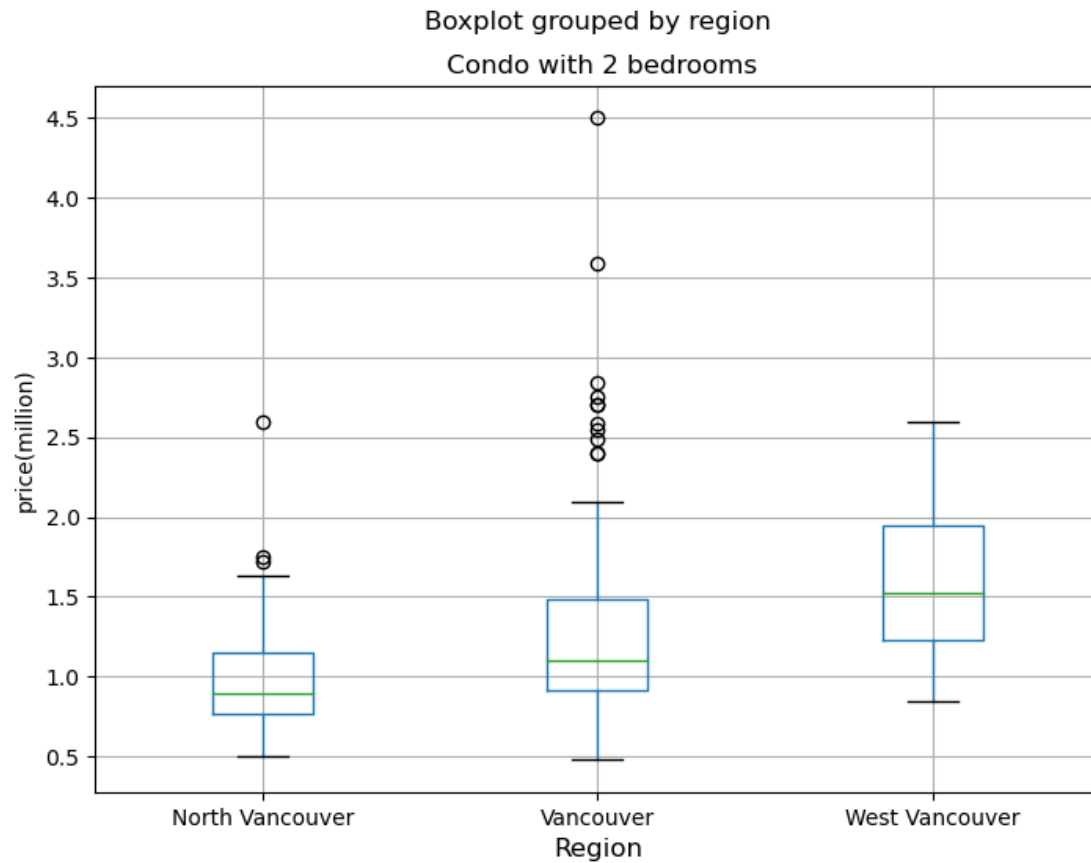
Top 20 price of 1 bedroom condo in all region





Condo price – 2 bedrooms

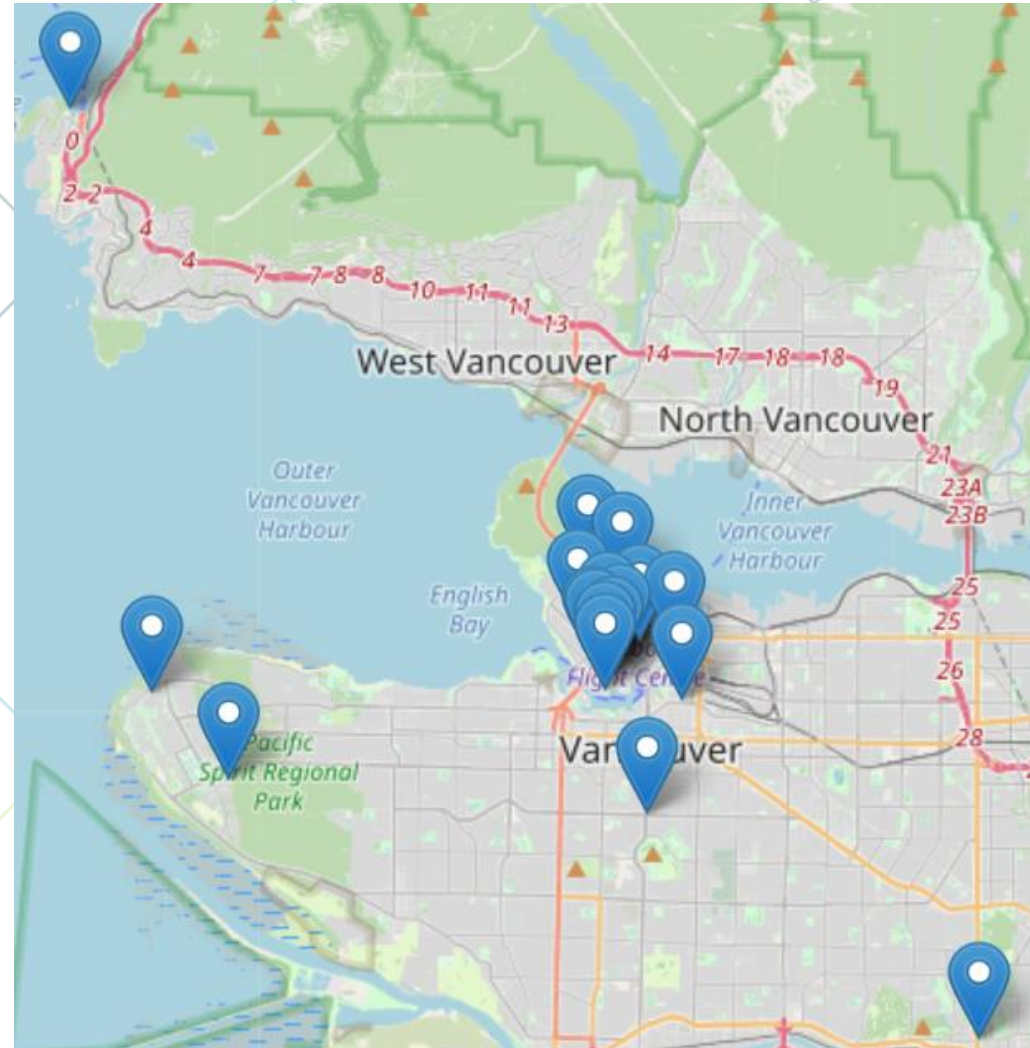
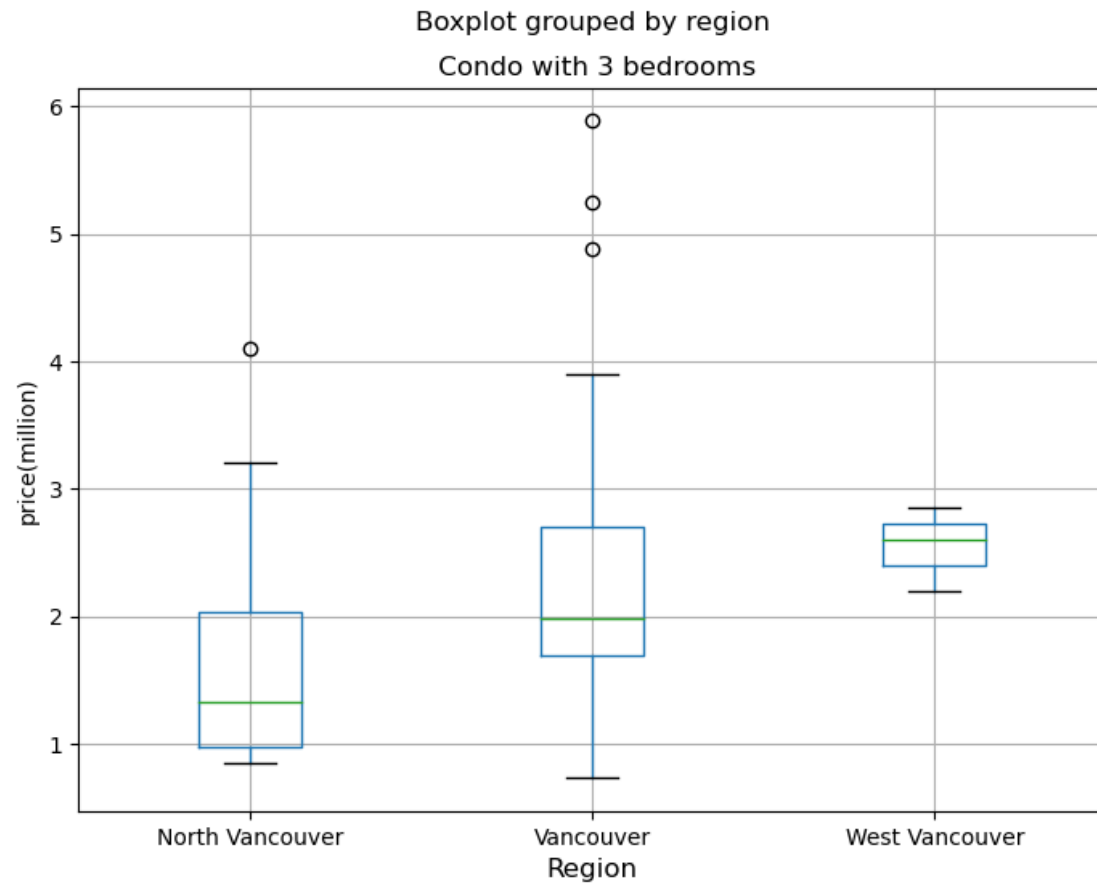
Top 20 price of 2 bedrooms condo in all region





Condo price – 3 bedrooms

Top 20 price of 3 bedrooms condo in all region

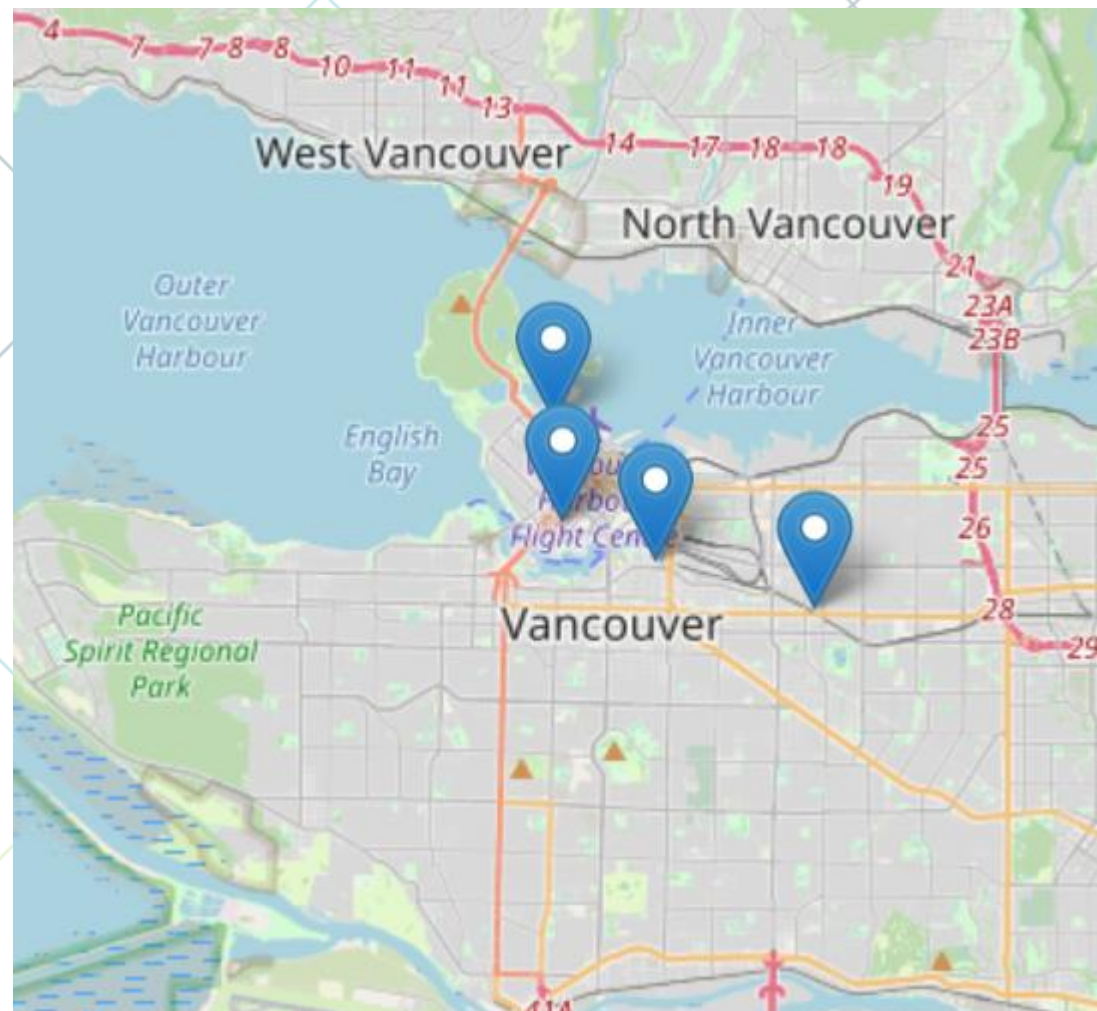
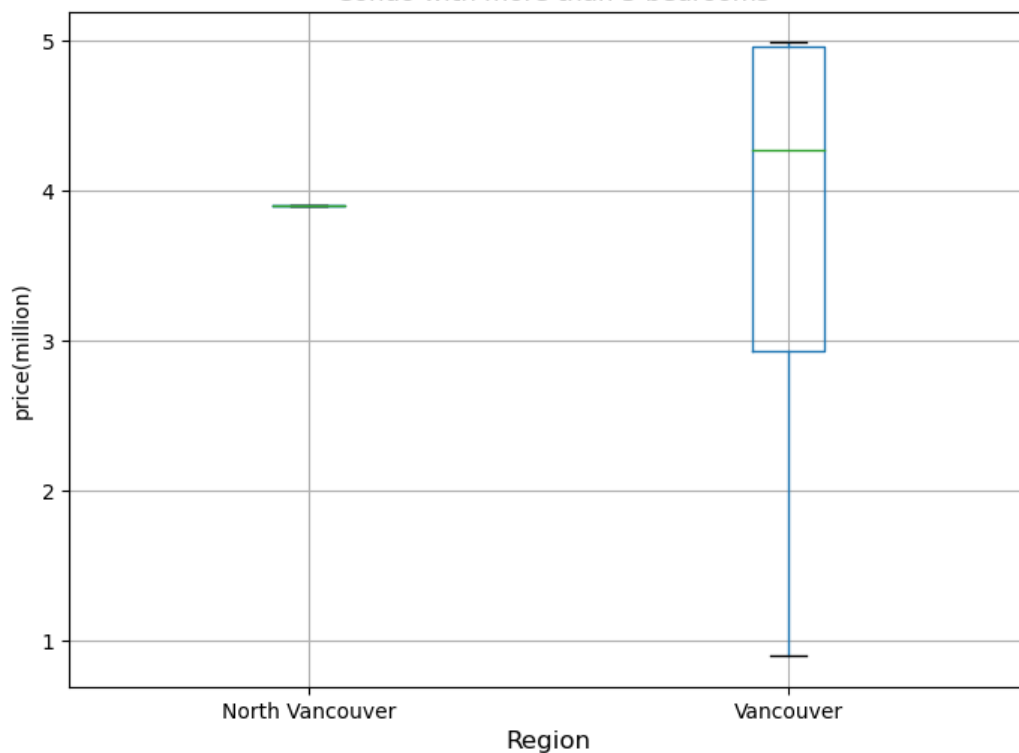




Condo price – more than 3 bedrooms

Top 20 price of more than 3 bedrooms condo in all region

Boxplot grouped by region
Condo with more than 3 bedrooms



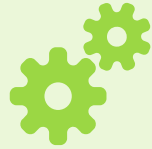


Conclusions

- 01 Most of the realty properties in greater Vancouver are house and condo
- 02 Total number of condo is more than number of house
- 03 Luxury houses (top 20 price) are concentrated in Vancouver region and West Vancouver
- 04 Luxury condos (top 20 price) are concentrated in Vancouver region

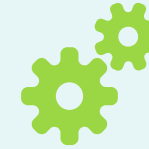


Challenges & Next steps



Xu

- Start Project
 - Data Cleaning
 - Better visualization
-
- ✓ Continue work hard
 - ✓ Review and Practice more



Fan

- Selenium
 - Complicated coding and visualization is always the challenge to me, since chemistry is far away from coding
-
- ✓ Review recorded videos
 - ✓ Work efficiently

The background is a complex geometric design. On the left, there are large, overlapping squares in shades of lime green and dark teal, some with drop shadows. A series of smaller, outlined squares in these colors and white extend from the left towards the center. On the right, a large grid of thin, light blue and green lines forms a diamond pattern. The text "THANK YOU" is centered in the middle of the image.

THANK YOU