

House Sigma

Web Scrapping Using Selenium

Sapna Maheshwari

Agenda



Company information & Motivation



Data



Data Analysis



Conclusion



Challenges



Next Steps

House Sigma

- Industry: Computer Software
 - Founder : Joseph Zeng & Reed Zhao
 - Founded: 2016
 - Type: Private
 - Headquarter: Markham Ontario
 - Company Size: 11-50 employees
 - Website: <http://www.housesigma.com>
-

Motivation

- New to the country.
- Looking for Property to buy.
- Real Estate getting hotter every month.
- Bidding war looks crazy.
- Comprehend house pricing pattern in the different areas.



Website

HouseSigma
Ontario
Search listing #, address
Map Search Market Trend Reports Contact Us
Watched List (0)

View Listing In Full Map

14 WESTLAKE Boulevard
Brantford Twp - Brant
Single Family

Listed for: ~~\$1,399,900~~
Sold for: \$1,505,000
Sold 1 day ago
Listed by Redline Properties Inc.

☐ Watch this listing. Receive notification when it is sold
☐ Watch this community. Receive updates on Single Family homes in Brantford Twp - Brant

Sold For: ~~\$800,000~~ **Sold** Single Family
3 Bedroom, 2 Bathroom, 2 Garage
200325 10 Sidersad, Georgian Bluffs - Rural Georgian Bluffs
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$1,505,000~~ **Sold** Single Family
3 Bedroom, 3 Bathroom, 2 Garage
14 WESTLAKE Boulevard, Brant - Brantford Twp
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$785,500~~ **Sold** Detached
3+2 Bedroom, 3 Bathroom, 2 Garage
57 Walwins Way, Barrie - Innis-Shore
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$1,300,000~~ **Sold** Freehold Townhouse
3 Bedroom, 3 Bathroom, 2 Garage
78 Carlaw Ave, Toronto - South Riverdale
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$1,060,000~~ **Sold** Semi-Detached
3 Bedroom, 2 Bathroom, 1 Garage
155 Marlmar Ave Toronto - North York
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$980,500~~ **Sold** Semi-Detached
3 Bedroom, 3 Bathroom, 1 Garage
1111 Brimley Ave, Mississauga - Mississauga
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$803,000~~ **Sold** Condo Townhouse
3 Bedroom, 3 Bathroom, 2 Garage
111 - 78 Broadbent Dr, Scarborough - Midland
List: 2021-02-22 End: 2021-02-23

Sold For: ~~\$880,000~~ **Sold** Single Family
3 Bedroom, 3 Bathroom, 1 Garage
448 1001 ET Street Waterloo - Other
List: 2021-02-22 End: 2021-02-23

Code

Web Scraping using Selenium

```
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
driver = webdriver.Chrome(chromedriver_win32\\chromedriver.exe')

# using seleniu function to login onto the website
driver.get("https://housesigma.com/web/en") #home
driver.implicitly_wait(6)
#sign in click button
driver.find_element_by_xpath('//*[@id="nav"]/div[1]/div[2]/div[3]/a[1]').click()
from time import sleep
sleep(2)
driver.find_element_by_xpath('//*[@id="pane-email"]/form/div[1]/div/div/input').send_keys("*****")
driver.find_element_by_xpath('//*[@id="pane-email"]/form/div[2]/div/div/input').send_keys("*****")
driver.find_element_by_xpath('//*[@id="login"]/div[2]/div/div[3]/button/span').click()

# Using Customize button to select GTA West
customize_button = driver.find_element_by_xpath('//*[@id="index"]/div[3]/div[2]/div/span/button').click()
Brampton = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[1]').click()
Mississauga = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[2]').click()
Oakville = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[3]').click()
Burlington = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[4]').click()
Milton = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[5]').click()
Halton Hills = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[6]').click()
Caledon = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[2]/div[3]/div[2]/div[4]/div[2]/span[7]').click()
save = driver.find_element_by_xpath('//*[@id="index"]/div[2]/div/div[3]/span/button[3]').click()

# opening webpage of just sold by clicking "see more"
banners = driver.find_element_by_xpath('//*[@id="index"]/div[3]/div[4]/div[1]/div[1]/a').click()

# Extracting the price for which the listings were listed for
listed_price(xxxx,xxxx) = []
for i in tqdm(range(2000,2373)):
    driver.get('https://housesigma.com'+prop_pagelink[i])
    sleep(5)
    soup = BeautifulSoup(driver.page_source)
    try:
        priceb = soup.find('div', class_='price_listing price_grey').get_text().split(" ")[-1]
    except:
        priceb = np.nan
    listed_pricexxxx,xxxx.append(priceb)

# Extracting list of information
address = tag.find_all('span',class_="address").get_text()
house = tag.find_all('span',class_="house_type").get_text()
room = tag.find_all('span',class_="rooms").get_text()
soldfor = tag.find_all('span',class_="price").get_text()
date = tag.find_all('div',class_="bottom").get_text()
```


Data

Total Listings: 2373

Date range for Listing house: 13th July,2020 to 19th February, 2021

Price range: Upto \$1.2M

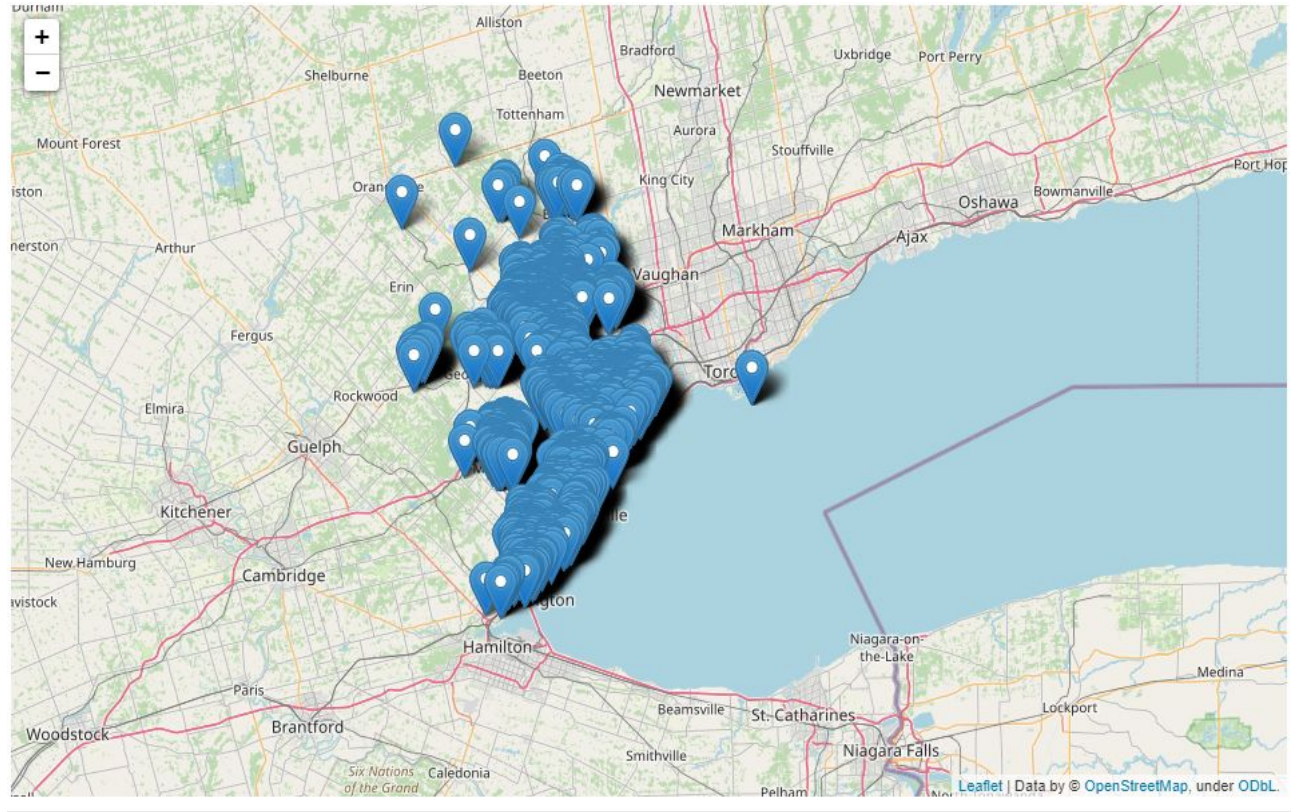
Focused Area: Greater Toronto Area - West

	address	housetype	sold	listed_price	city	bedroom	bathroom	garage	listing_date	acquisition_date	days_on_market	sold_vs_askprice
0	1653 Glenburnie Rd, Mississauga - Mineola	Detached	1056000	1049000	Mississauga	3	2	1	2021-02-12	2021-02-21	9	7000
1	15 - 2970 Berwick Dr, Burlington - Rose	Condo Townhouse	1100000	849900	Burlington	3	3	1	2021-02-19	2021-02-20	1	250100
2	36 Solway Ave, Brampton - Heart Lake East	Detached	910000	839900	Brampton	3+1	3	1	2021-02-19	2021-02-20	1	70100
3	38 Ridgefield Crt, Brampton - Vales of Castlemore	Semi- Detached	880000	849000	Brampton	3+1	4	1	2021-02-18	2021-02-20	2	31000

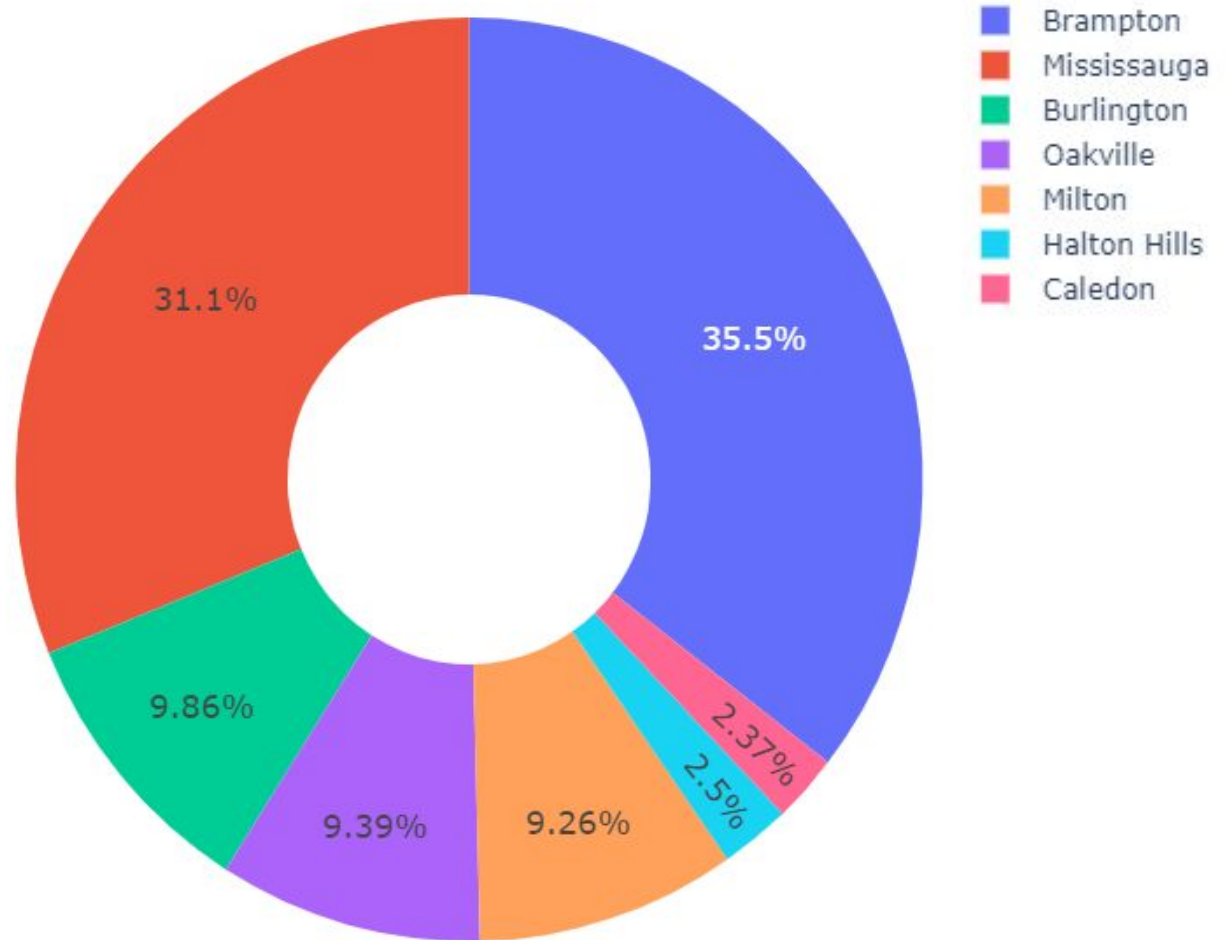
Data Analysis

- Distribution of properties sold in the GTA- West Area
- Total listings per property type
- Average days property stayed in the market
- Listed Price Vs Selling Price
- Selling Price Vs Listed Price: Difference

Distribution of properties sold in the GTA- West Area

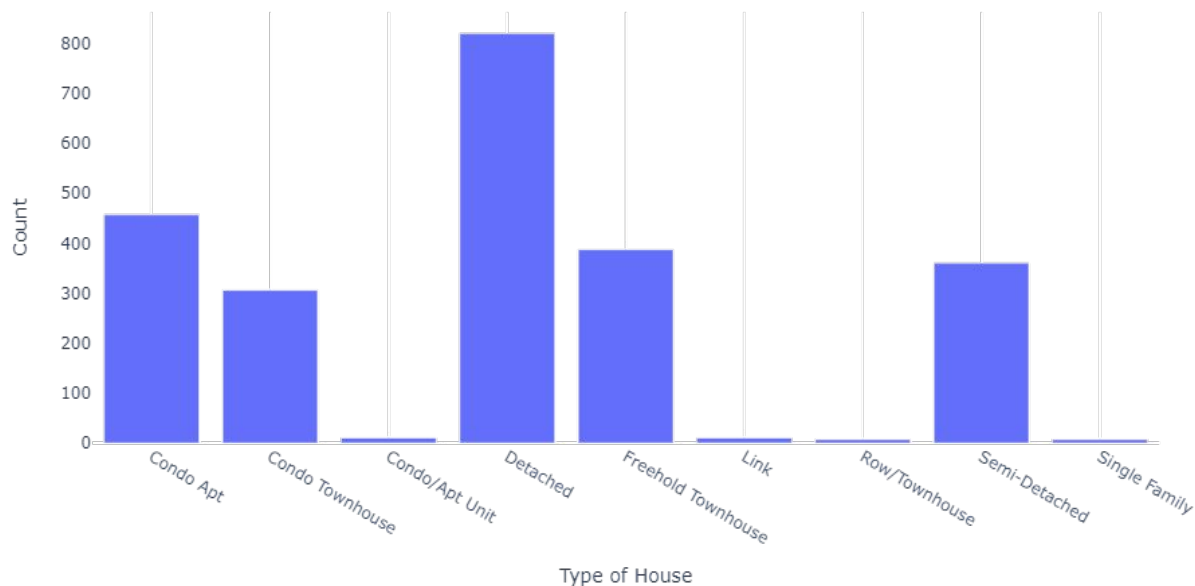


Distribution of properties sold in the GTA- West Area



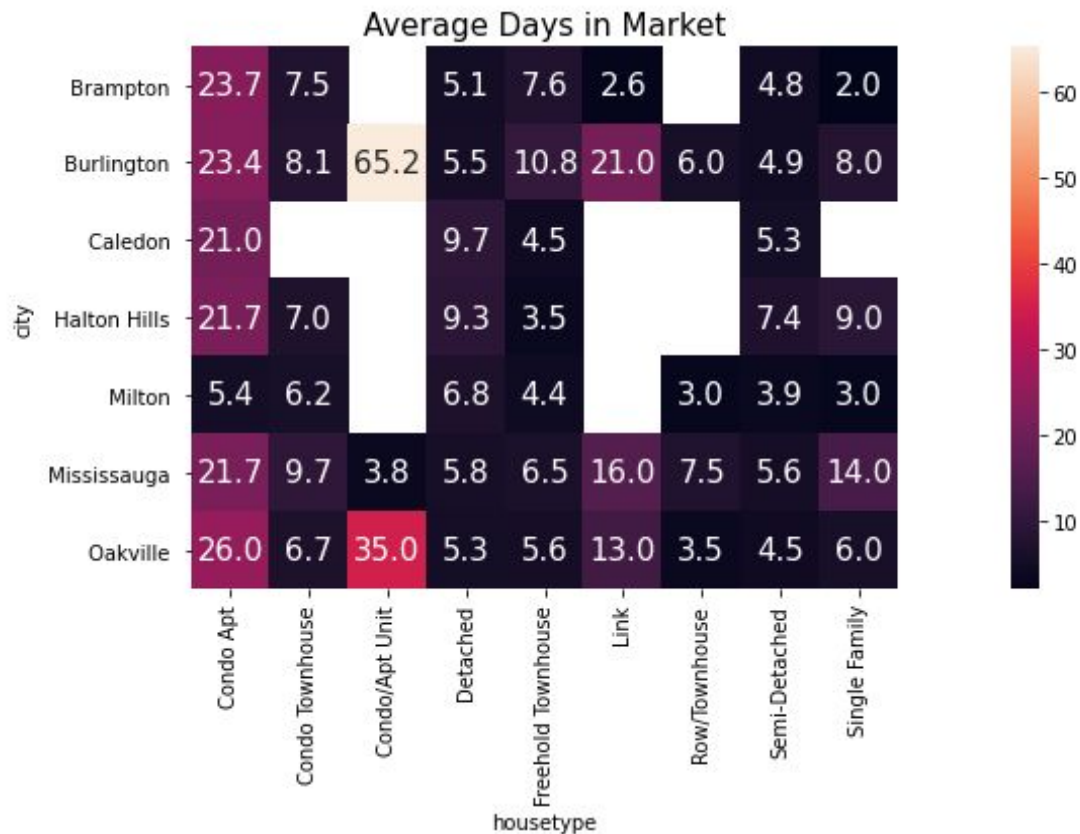
Types of Properties Sold

Total Properties Sold

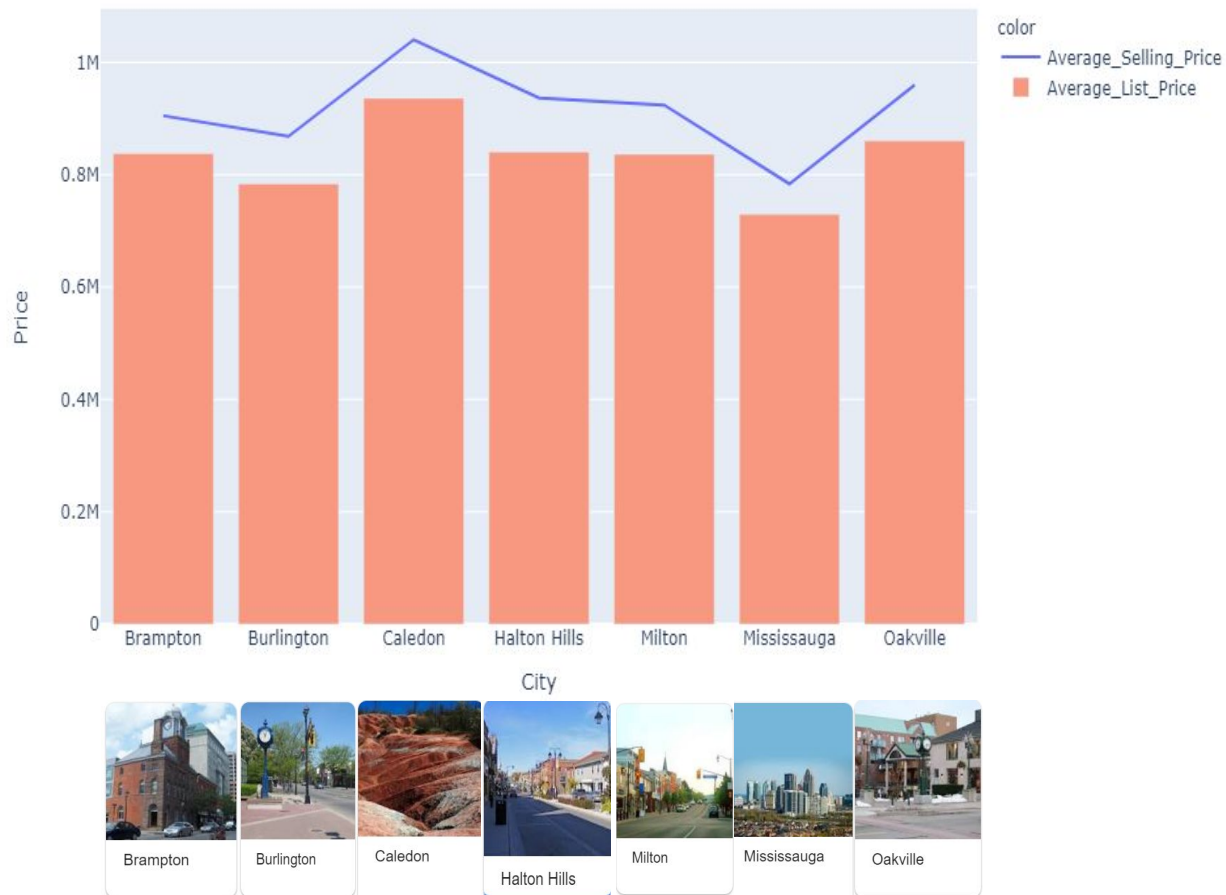


housetype	Condo Apt	Condo Townhouse	Condo/Apt Unit	Detached	Freehold Townhouse	Link	Row/Townhouse	Semi-Detached	Single Family
city									
Brampton	59	64	0	383	143	5	0	184	1
Burlington	56	33	4	92	25	3	2	17	1
Caledon	1	0	0	33	12	0	0	10	0
Halton Hills	3	2	0	43	4	0	0	5	2
Milton	14	13	0	69	94	0	1	27	1
Mississauga	283	159	5	145	34	1	2	106	1
Oakville	41	35	1	55	75	1	2	11	1

Average Days Properties Stayed in Market

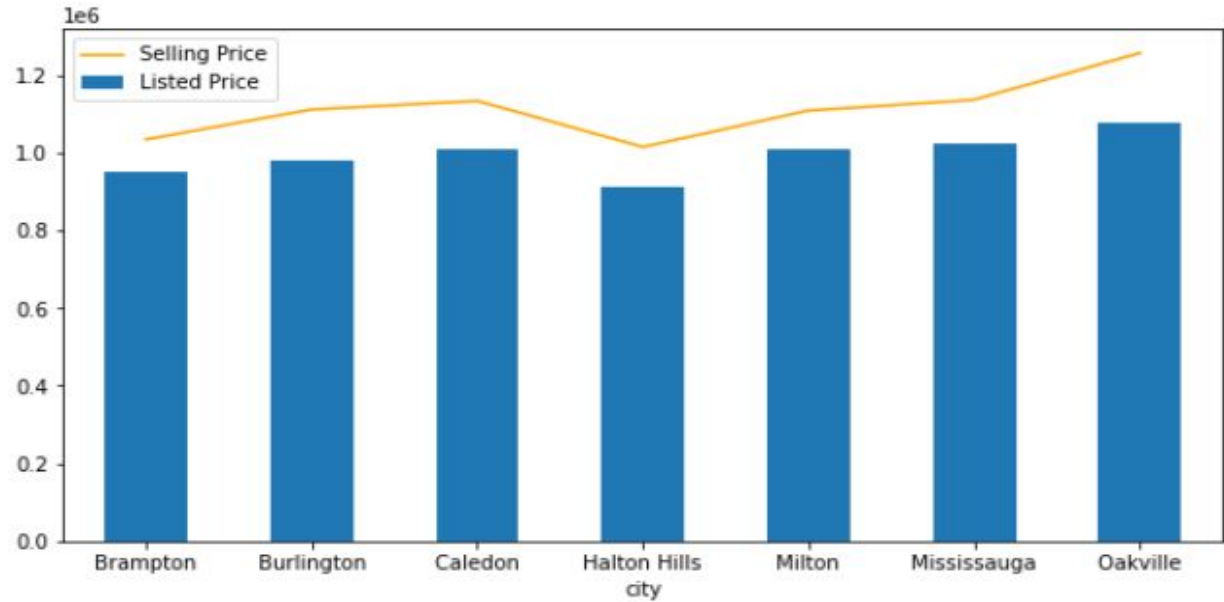


Listed Price Vs Selling Price



Listed Price Vs Selling Price

For Detached Properties Only



Brampton



Burlington



Caledon



Halton Hills



Milton

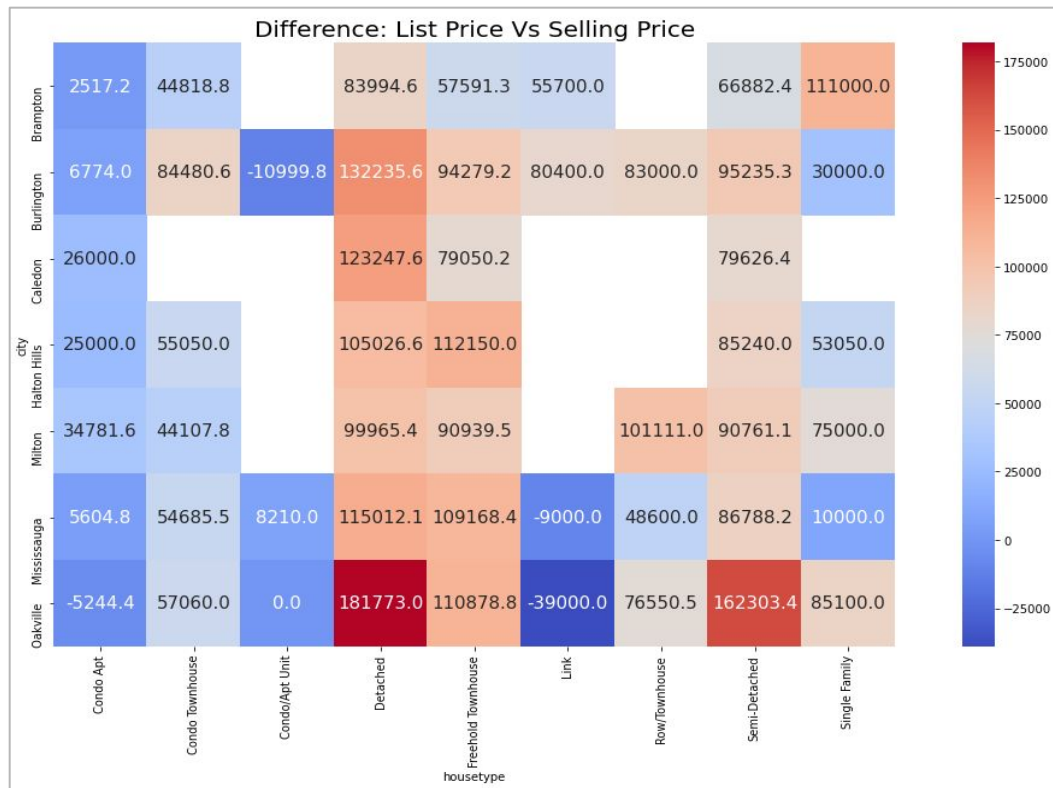


Mississauga



Oakville

Average Difference: List Price Vs Selling Price



Conclusion

- Detached , Semi-Detached and Townhouse were sold in less than 2 weeks of their listings.
- Detached properties were sold approximately 80k - 190k above of their asking price.
- Condos took 3-4 weeks to get sold.
- Buyers prefer to buy Detached and Townhouses than condos/Apt Units.
- Pandemic and Immigrants does impact the prices and demand and supply of Real Estate.
- Sharp increase in the housing market gave old buyers an additional disposal income to invest in new properties.



Challenges

- Loading data: time consuming task.'
- Understand working of Selenium.
- Dealing with semi -unstructured data.
- Extra Cautious: Analyzing with smaller dataset may leads to skewed results.



Next Steps

- Widening the scope: Extract and analyze data of other parts of GTA.

Thank You!