DS8007-Advanced Data Visualization Project

Toronto Housing Market

Instructor: Dr Sanaz Jafari
Prestended By :Vaidehi Atodaria (April,2023)
Master of Data Science and Analytic
(Toronto Metropolitan University)

Table Of Contents:

INTRODUCTION

In many cities around the world,including Toronto,housing affordability and availability are major issues.Over the past decade,the city of Toronto has experienced rapid growth in real estate prices,making it increasingly difficult for many resident to find affordable housing.In this data visualization project,I aim to provide insights and understanding into the trends and patterns in the Toronto housing market over the past decade,using publicly available dataset.

PROBLEM DEFINITION

The rising housing prices in Toronto pose a significant challenge for businesses and individuals looking to invest in the real estate market,as well as for those looking to find affordable housing.This project aims to explore and visualize the trends and patterns in the Toronto Housing Market,with the goal of providing valuable insights to stakeholders and the public.

RESEARCH QUESTION

To achieve this goal,we will explore the following questions:

1. How has the average price of a home in Toronto Changed over the past decade?
2. Are there any patterns or trends in the housing market that can be observed over time?

3. How do different neighbourhoods in Toronto compare in terms of housing prices and trends?
4. What factors are driving changes in the Toronto housing market,such as economy,population growth?
5. How has pandemic affected the Toronto housing market,and what are the long term implications?

MOTIVATION

The motivation for this project is to contribute to ongoing conversation around housing affordability and availability in Toronto by visualising the data I aim to provide insights and understanding to public.

PUBLICLY AVAILABLE DATASET

For this project,I will be using the Toronto Home Price Index dataset from Kaggle(https://www.kaggle.com/ahnafalfarisi/toronto-home-price-index). This dataset includes monthly home price indices for the city of Toronto,broken down by neighbourhood and dwelling type,from January 2005 to august 2021.By using a publicly available dataset,we aim to provide transparency and accessibility to our analysis and findings.

DATA DESCRIPTION

VARIABLE	DATA TYPE	DESCRIPTION
Neighborhood	Categorical	The name of the neighborhood where the property is located
Dwelling_Type	Categorical	The type of dwelling, such as single-family home, condo, or townhouse
Date	Date	The date of the home price index observation
Home_Price_Index	Numeric	Numeric The home price index for the given observation
Year	Numeric	The year of the observation
Month	Numeric	The month of the observation
Quarter	Numeric	The quarter of the observation
Avg_Home_Price_Index	Numeric	The average home price index for the given neighborhood
Above_Median	Binary	Whether the home price index is above or below the median for the dataset
Price_Change_Percent	Numeric	The percentage change in the average home price index for the given neighborhood over the past decade

Data Cleaning and Preprocessing of Data

Clean and preprocess the data, such as handling missing values converting data types.After completing these cleaning and preprocessing steps,the datasets was ready for exploratory data analysis.

Python code for checking null,missing values and data set description

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Read the CSV file and store it in a DataFrame
file_path = 'mls.csv'
df = pd.read_csv(file_path)

def check_df(dataframe, head=5):
    print(" SHAPE ".center(70, '#'))
    print('Rows: {}'.format(dataframe.shape[0]))
    print('Columns: {}'.format(dataframe.shape[1]))
    print(" INFO ".center(70, '#'))
    print(dataframe.info())
    print(" MISSING VALUES ".center(70, '#'))
    print(dataframe.isnull().sum())
    print(" DUPLICATED VALUES ".center(70, '#'))
    print(dataframe.duplicated().sum())
    print(" DESCRIBE ".center(70, '#'))
    print(dataframe.describe().T)


check_df(df)
```

```
################################ SHAPE ################################
Rows: 5091
Columns: 17
################################ INFO #################################
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5091 entries, 0 to 5090
Data columns (total 17 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Location           5091 non-null   object
 1   CompIndex          5076 non-null   float64
 2   CompBenchmark      5076 non-null   float64
 3   CompYoYChange      5076 non-null   float64
 4   SFDetachIndex      5076 non-null   float64
 5   SFDetachBenchmark  5076 non-null   float64
 6   SFDetachYoYChange  5075 non-null   float64
 7   SFAttachIndex      4949 non-null   float64
 8   SFAttachBenchmark  4949 non-null   float64
 9   SFAttachYoYChange  4951 non-null   float64
 10  THouseIndex        3803 non-null   float64
 11  THouseBenchmark    3803 non-null   float64
 12  THouseYoYChange    3804 non-null   float64
 13  ApartIndex         4010 non-null   float64
 14  ApartBenchmark     4010 non-null   float64
 15  ApartYoYChange     4008 non-null   float64
 16  Date               5091 non-null   object
dtypes: float64(15), object(2)
memory usage: 676.3+ KB
```

```
############################ MISSING VALUES #############################
Location                 0
CompIndex               15
CompBenchmark           15
CompYoYChange           15
SFDetachIndex           15
SFDetachBenchmark       15
SFDetachYoYChange       16
SFAttachIndex          142
SFAttachBenchmark      142
SFAttachYoYChange      140
THouseIndex           1288
THouseBenchmark       1288
THouseYoYChange       1287
ApartIndex            1081
ApartBenchmark        1081
ApartYoYChange        1083
Date                     0
dtype: int64
```

```
########################### DUPLICATED VALUES ###########################
62
############################### DESCRIBE ################################
                     count           mean            std        min  \
CompIndex           5076.0     244.871592      41.374120     135.00
CompBenchmark       5076.0  772049.881797  272588.169295  257000.00
CompYoYChange       5076.0      10.222695       9.186082     -19.58
SFDetachIndex       5076.0     246.021277      41.829704     126.00
SFDetachBenchmark   5076.0  948100.059102  376586.547562  259100.00
SFDetachYoYChange   5075.0       9.329878      10.411804     -21.32
SFAttachIndex       4949.0     247.290139      43.124667     137.20
SFAttachBenchmark   4949.0  718284.198828  243377.690928  273300.00
SFAttachYoYChange   4951.0       9.311238      10.105563    -100.00
THouseIndex         3803.0     245.168393      55.364090       0.00
THouseBenchmark     3803.0  582943.965291  221250.438808       0.00
THouseYoYChange     3804.0       9.466312      14.473397    -100.00
ApartIndex          4010.0     236.428279      52.340839     106.00
ApartBenchmark      4010.0  483821.970075  145759.334129  171400.00
ApartYoYChange      4008.0      11.521093       9.079515     -10.60

                         25%        50%           75%         max
CompIndex            217.600     247.10  2.705000e+02      409.90
CompBenchmark     589000.000  726500.00  9.033000e+05  2162900.00
CompYoYChange          4.380       9.39  1.538250e+01       46.78
SFDetachIndex        220.300     246.70  2.720000e+02      415.60
SFDetachBenchmark  682200.000  873450.00  1.137150e+06  2536900.00
SFDetachYoYChange      2.245       8.54  1.571500e+01       49.10
SFAttachIndex        220.700     246.70  2.699000e+02      709.00
SFAttachBenchmark  542900.000  671600.00  8.399000e+05  1677200.00
SFAttachYoYChange      2.715       8.76  1.537000e+01       43.71
THouseIndex          210.500     246.40  2.783000e+02      466.80
THouseBenchmark    439450.000  555000.00  6.731000e+05  1750500.00
THouseYoYChange        4.080       9.63  1.560250e+01       56.49
ApartIndex           196.625     239.60  2.722000e+02      395.30
ApartBenchmark     381925.000  477600.00  5.692500e+05  1005500.00
ApartYoYChange         5.640       9.74  1.529000e+01       64.32
```

EXPLORATORY DATA ANALYSIS

The goal of EDA is to explore the data,identify patterns and relationship,and generate insights that can inform further analysis and modeling.

```
# 1. Handle missing values
# Drop missing value fill with mean, median, or mode)
df.dropna(inplace=True)

# 2. Convert data types if necessary
# In this case, we will convert the 'Date' column to a datetime object
df['Date'] = pd.to_datetime(df['Date'])

# 3. Handle outliers (example: IQR method for 'CompIndex')
Q1 = df['CompIndex'].quantile(0.25)
Q3 = df['CompIndex'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Remove outliers from the dataset
df = df[(df['CompIndex'] >= lower_bound) & (df['CompIndex'] <= upper_bound)]
print(df)
# 4. Normalize or scale features (if necessary)
#  we will normalize the 'CompIndex' column using Min-Max normalization
from sklearn.preprocessing import MinMaxScaler

# Select the feature(s) to normalize
feature_to_normalize = ['CompIndex']

# Instantiate the scaler object
scaler = MinMaxScaler()

# Fit and transform the data
df[feature_to_normalize] = scaler.fit_transform(df[feature_to_normalize])
```

```
         Location  CompIndex  CompBenchmark  CompYoYChange  SFDetachIndex  \
69           Ajax   0.175214       444300.0          16.00          178.8
70           Ajax   0.178063       445800.0          16.31          179.2
71           Ajax   0.189459       451700.0          15.55          181.4
72           Ajax   0.194207       454200.0          15.74          182.3
73           Ajax   0.186135       450000.0          12.69          180.3
...           ...        ...            ...            ...            ...
5086  York Region   0.684236       970300.0          11.76          292.0
5087  York Region   0.696106       978700.0          12.56          295.2
5088  York Region   0.721747       997000.0          13.34          301.7
5089  York Region   0.792972      1047700.0          17.16          318.6
5090  York Region   0.860399      1095700.0          19.82          334.4

      SFDetachBenchmark  SFDetachYoYChange  SFAttachIndex  SFAttachBenchmark  \
69             478900.0              17.86          187.1           402400.0
70             480000.0              15.76          187.6           403500.0
71             485800.0              15.10          190.1           408800.0
72             488300.0              14.51          191.2           411200.0
73             482900.0              11.99          190.3           409300.0
...                 ...                ...            ...                ...
5086          1112000.0              12.65          293.2           847300.0
5087          1124200.0              13.89          296.3           856200.0
5088          1148900.0              15.20          302.9           875300.0
5089          1213300.0              20.23          322.6           932200.0
5090          1273400.0              23.49          337.2           974400.0

      SFAttachYoYChange  THouseIndex  THouseBenchmark  THouseYoYChange  \
69                15.64        170.5         311600.0            18.16
70                16.81        170.4         311400.0            15.76
71                15.99        172.3         314900.0            14.33
72                15.95        177.1         323700.0            17.13
73                13.14        173.1         316300.0            12.99
...                 ...          ...              ...              ...
5086              12.73        271.3         704100.0            11.28
5087              13.66        270.4         701800.0            10.28
5088              14.78        272.6         707500.0            10.14
5089              20.15        280.7         728500.0            11.04
5090              22.48        294.0         763000.0            14.75

      ApartIndex  ApartBenchmark  ApartYoYChange       Date
69         148.7        252500.0            5.61 2015-07-01
70         151.5        257200.0            8.99 2015-08-01
71         154.3        262000.0           10.53 2015-09-01
72         152.4        258800.0            8.70 2015-10-01
73         151.2        256700.0            5.00 2015-11-01
...          ...             ...             ...        ...
5086       252.2        562700.0            6.91 2020-11-01
5087       253.2        564900.0            6.84 2020-12-01
5088       254.7        568200.0            5.86 2021-01-01
5089       260.2        580500.0            5.39 2021-02-01
5090       268.7        599500.0            6.00 2021-03-01

[3641 rows x 17 columns]
```

```
1 HousingData.Location.unique()
```

```
array(['Adjala-Tosorontio', 'Ajax', 'Aurora', 'Barrie', 'Bradford West',
       'Bradford West Gwillimbury', 'Brampton', 'Brock', 'Burlington',
       'Caledon', 'City of Toronto', 'Clarington', 'Dufferin County',
       'Durham Region', 'EGswsiallimbury', 'East Gwillimbury', 'Essa',
       'GEswsiallimbury', 'Georgina', 'Halton Hills', 'Halton Region',
       'Innisfil', 'King', 'Markham', 'Milton', 'Mississauga',
       'New Tecumseth', 'Newmarket', 'Oakville', 'Orangeville', 'Oshawa',
       'Peel Region', 'Pickering', 'Richmond Hill', 'Scugog',
       'Simcoe County', 'TREB Total', 'Toronto C01', 'Toronto C02',
       'Toronto C03', 'Toronto C04', 'Toronto C06', 'Toronto C07',
       'Toronto C08', 'Toronto C09', 'Toronto C10', 'Toronto C11',
       'Toronto C12', 'Toronto C13', 'Toronto C14', 'Toronto C15',
       'Toronto E01', 'Toronto E02', 'Toronto E03', 'Toronto E04',
       'Toronto E05', 'Toronto E06', 'Toronto E07', 'Toronto E08',
       'Toronto E09', 'Toronto E10', 'Toronto E11', 'Toronto W01',
       'Toronto W02', 'Toronto W03', 'Toronto W04', 'Toronto W05',
       'Toronto W06', 'Toronto W07', 'Toronto W08', 'Toronto W09',
       'Toronto W10', 'Uxbridge', 'Vaughan', 'Whitby',
       'Whitchurch-Stouffville', 'York Region'], dtype=object)
```

GENERATE SUMMARY STATISTICS OF DATA(MEAN,STD,MAX)

```
# Generate summary statistics
summary_stats = df.describe()
print(summary_stats)
```

```
         CompIndex  CompBenchmark  CompYoYChange  SFDetachIndex  \
count  3641.000000   3.641000e+03    3641.000000    3641.000000
mean      0.500849   7.972443e+05       9.806894     249.818923
std       0.186450   2.618999e+05       8.431952      39.304269
min       0.000000   3.099000e+05     -19.580000     126.300000
25%       0.374159   6.220000e+05       4.630000     228.000000
50%       0.514890   7.517000e+05       9.160000     250.800000
75%       0.624400   9.154000e+05      14.250000     275.500000
max       1.000000   2.162900e+06      38.990000     380.500000

       SFDetachBenchmark  SFDetachYoYChange  SFAttachIndex  SFAttachBenchmark  \
count       3.641000e+03        3641.000000    3641.000000       3.641000e+03
mean        1.023623e+06           8.573197     248.359324       7.653170e+05
std         3.633570e+05           9.996615      38.818005       2.318607e+05
min         3.455000e+05         -21.320000     137.200000       2.868000e+05
25%         7.679000e+05           2.090000     223.000000       6.042000e+05
50%         9.470000e+05           8.020000     248.800000       7.206000e+05
75%         1.205400e+06          14.100000     271.800000       8.781000e+05
max         2.536900e+06          45.690000     385.400000       1.677200e+06

       SFAttachYoYChange  THouseIndex  THouseBenchmark  THouseYoYChange  \
count        3641.000000  3641.000000     3.641000e+03      3641.000000
mean            8.563227   245.432601     5.859174e+05         9.383057
std             9.449906    53.588073     2.220566e+05        14.404316
min           -21.470000     0.000000     0.000000e+00      -100.000000
25%             2.450000   212.800000     4.455000e+05         4.010000
50%             8.150000   247.300000     5.582000e+05         9.640000
75%            14.210000   278.900000     6.727000e+05        15.470000
max            38.700000   407.600000     1.750500e+06        56.490000

        ApartIndex  ApartBenchmark  ApartYoYChange
count  3641.000000     3641.000000     3641.000000
mean    235.368223   473605.657786       11.776416
std      51.128247   137680.059889        9.185079
min     106.000000   171400.000000       -7.450000
25%     195.400000   376100.000000        5.810000
50%     239.500000   469100.000000        9.950000
75%     271.700000   555000.000000       15.710000
max     395.300000   975100.000000       64.320000
```

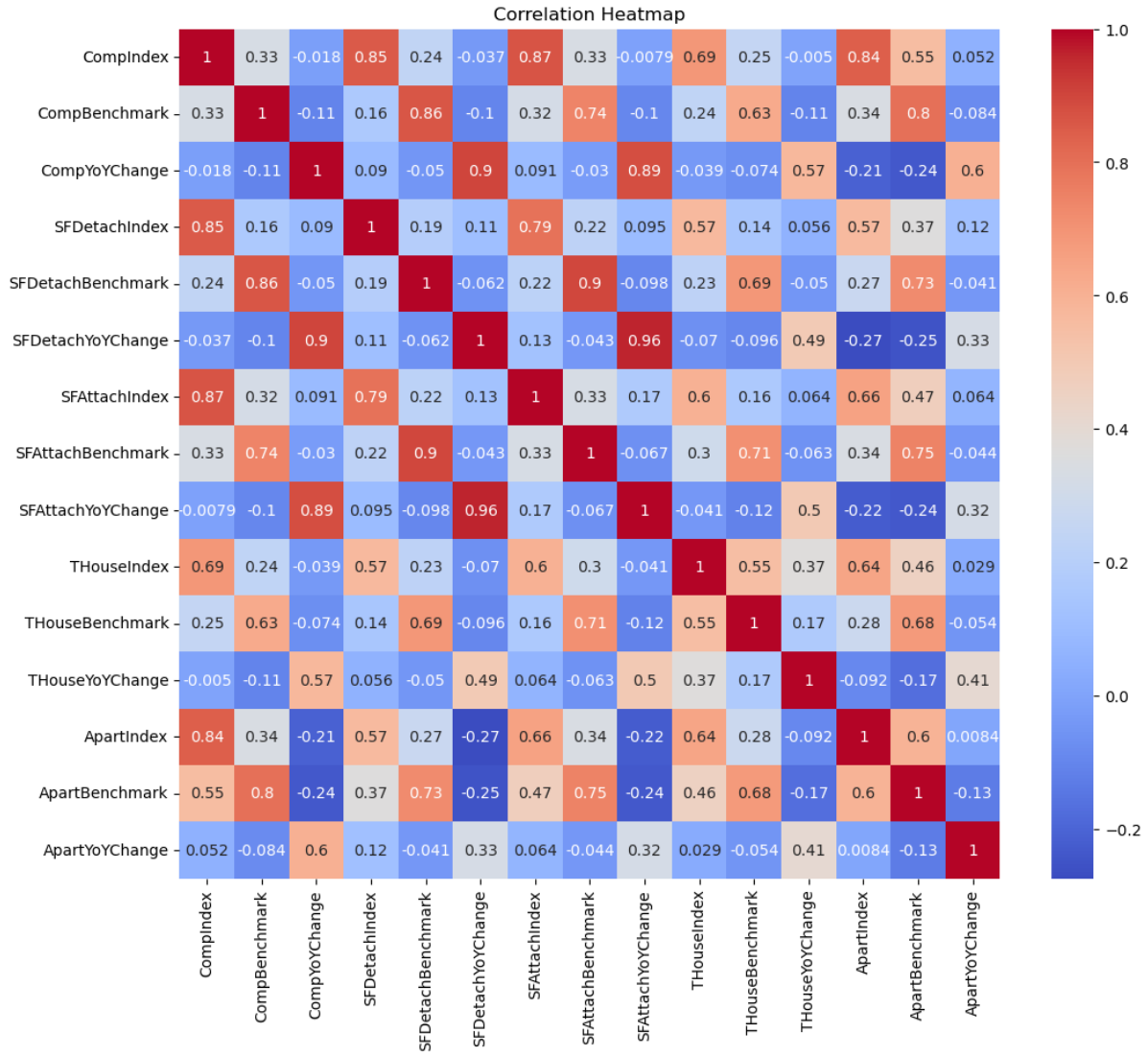
Correlation matrix is useful for identifying pattens and relationships between variables.

```
# Compute the correlation matrix
correlation_matrix = df.corr()

# Display the correlation matrix
print(correlation_matrix)

```

```
                  CompIndex  CompBenchmark  CompYoYChange  SFDetachIndex  \
CompIndex          1.000000       0.328066      -0.018261       0.852718
CompBenchmark      0.328066       1.000000      -0.107641       0.159185
CompYoYChange     -0.018261      -0.107641       1.000000       0.090013
SFDetachIndex      0.852718       0.159185       0.090013       1.000000
SFDetachBenchmark  0.238686       0.864591      -0.049742       0.190061
SFDetachYoYChange -0.036574      -0.104853       0.903644       0.110729
SFAttachIndex      0.867531       0.320635       0.091051       0.793545
SFAttachBenchmark  0.334586       0.744551      -0.030471       0.224755
SFAttachYoYChange -0.007899      -0.099701       0.885183       0.095199
THouseIndex        0.691185       0.237763      -0.039156       0.570696
THouseBenchmark    0.248319       0.625719      -0.073597       0.144254
THouseYoYChange   -0.005035      -0.108694       0.574262       0.056310
ApartIndex         0.843132       0.339658      -0.214379       0.566165
ApartBenchmark     0.552933       0.799227      -0.235054       0.366750
ApartYoYChange     0.051782      -0.084088       0.603070       0.117378

                  SFDetachBenchmark  SFDetachYoYChange  SFAttachIndex  \
CompIndex                  0.238686          -0.036574       0.867531
CompBenchmark              0.864591          -0.104853       0.320635
CompYoYChange             -0.049742           0.903644       0.091051
SFDetachIndex              0.190061           0.110729       0.793545
SFDetachBenchmark          1.000000          -0.062388       0.218774
SFDetachYoYChange         -0.062388           1.000000       0.125918
SFAttachIndex              0.218774           0.125918       1.000000
SFAttachBenchmark          0.895961          -0.042564       0.328836
SFAttachYoYChange         -0.097731           0.961786       0.170163
THouseIndex                0.227162          -0.070166       0.597840
THouseBenchmark            0.688486          -0.096282       0.161061
THouseYoYChange           -0.049854           0.494568       0.063663
ApartIndex                 0.272749          -0.274423       0.656046
ApartBenchmark             0.725649          -0.252473       0.470998
ApartYoYChange            -0.041279           0.327214       0.064424
```

CORRELATION HEATMAP

This is a correlation matrix show's correlation coefficients between each pair of variables in the dataset. The correlation coefficient ranges from -1 to 1, with -1 indicating a perfect negative correlation, 0 indicating no correlation, and 1 indicating a perfect positive correlation. The darker colors in the heatmap indicate a stronger correlation between the variables. We can see that the highest positive correlation (dark red color) exists between the CompIndex and SFDetachIndex (0.85) and SFAttachIndex (0.87). On the other hand, the highest negative correlation (dark blue color) exists between CompYoYChange and SFDetachYoYChange (-0.903) and SFAttachYoYChange (-0.885).

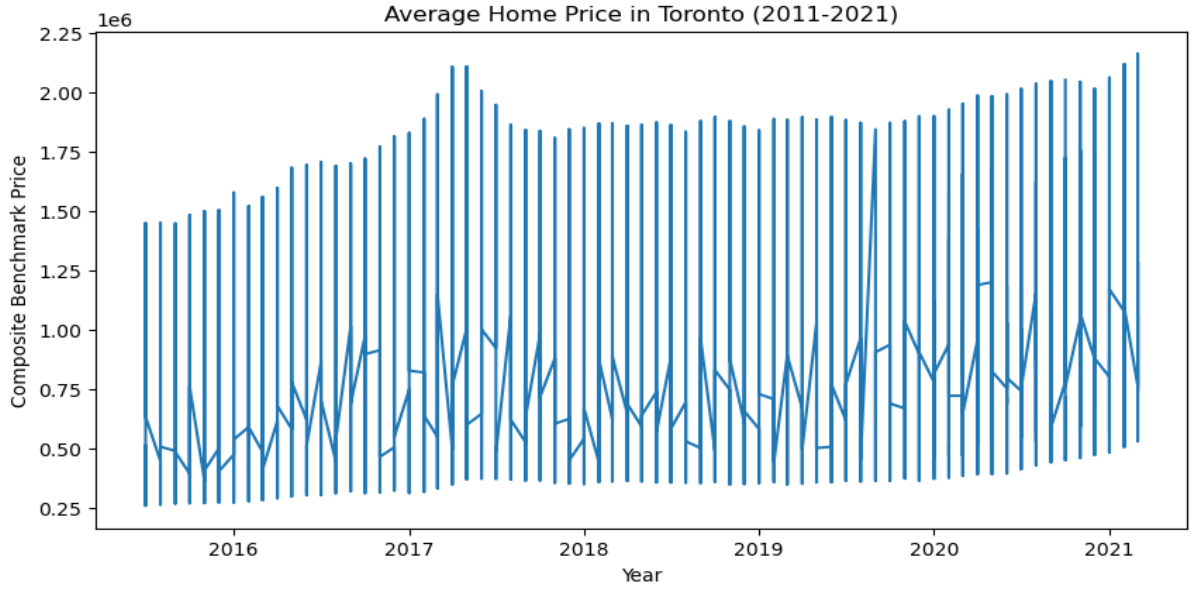


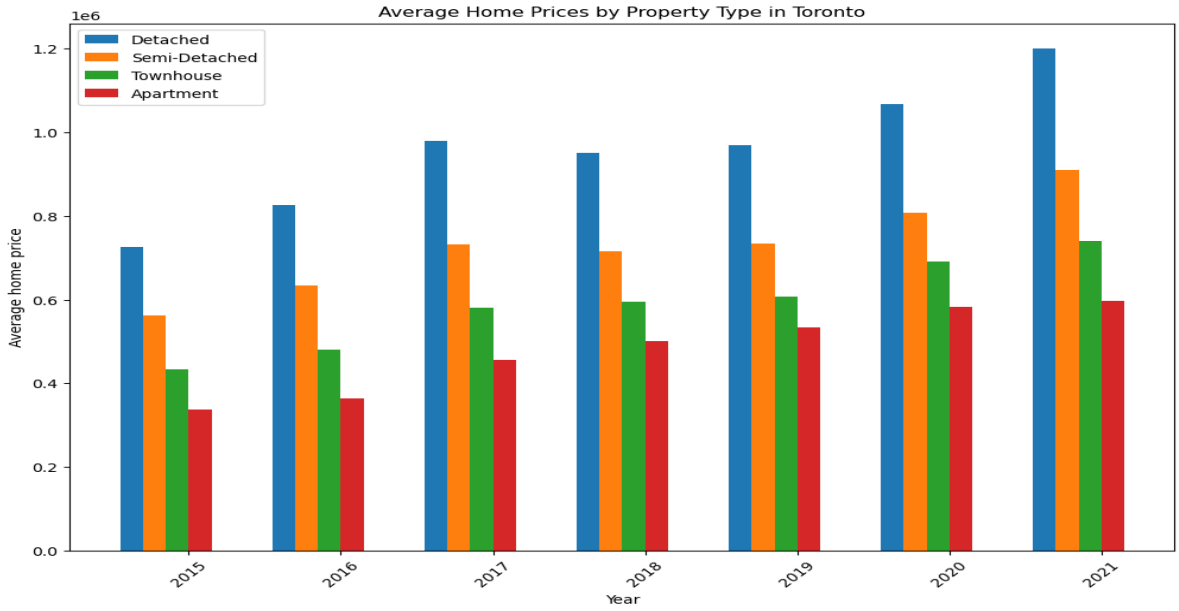
FINDING AND RESULTS

1.How has the average price of a home in Toronto changed over the past decade?

Visualization: Time series plot, For all property types comparison-bar graph

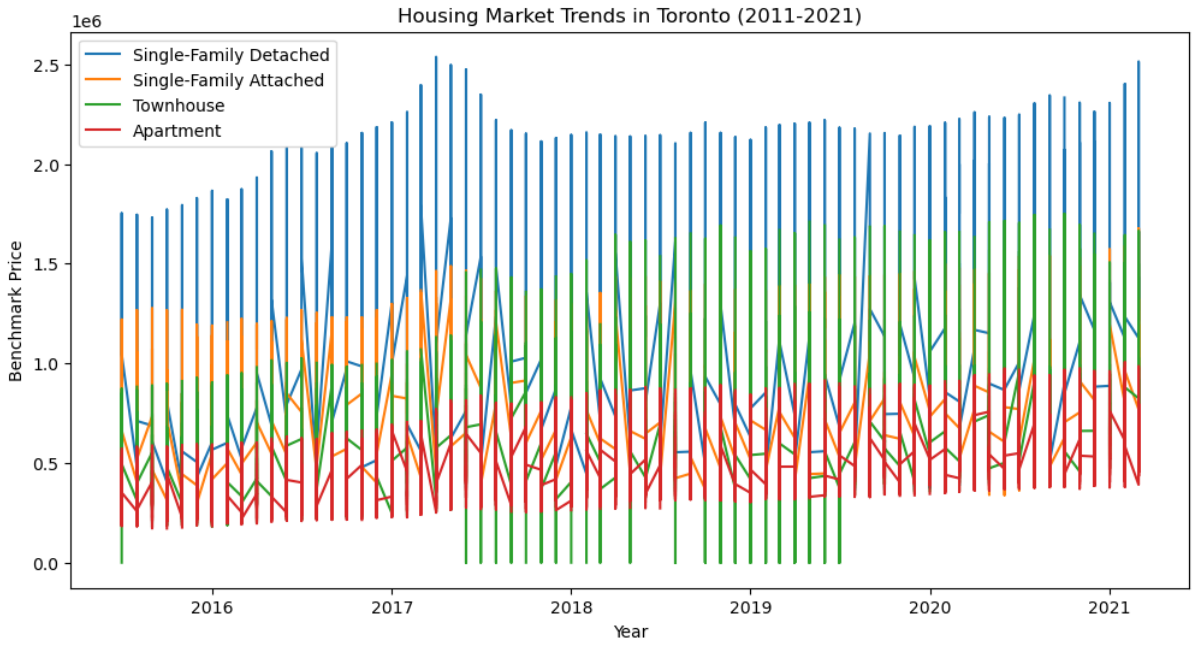
Variables: Date (x-axis) and CompBenchmark or a specific property type benchmark (y-axis)or all property types in the dataset (Single-Family Detached, Single-Family Attached, Townhouse, and Apartment).The composite benchmark price provides a comprehensive view of the overall housing market by taking into account The price trends across different property types.



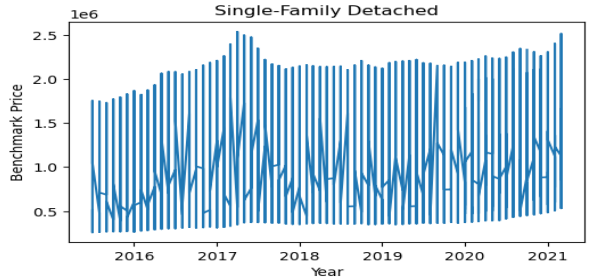


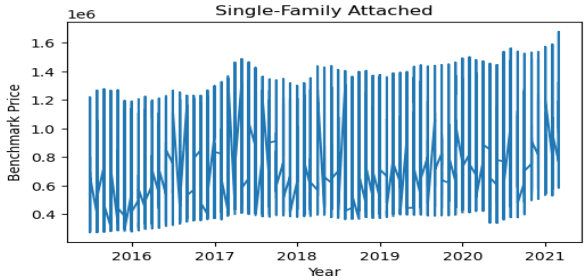
2.Are there any patterns or trends in the housing market that can be observed over time?

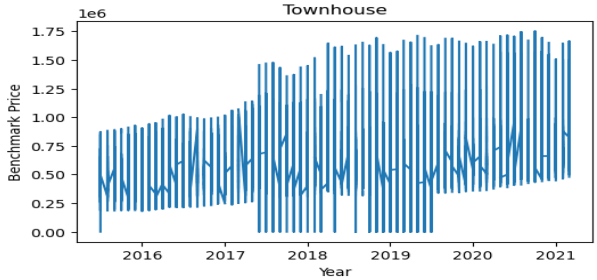
Visualization: Multiple time series plots on the same graph for different property .**Variables:** Date (x-axis), SFDetachBenchmark, SFAttachBenchmark, THouseBenchmark, and ApartBenchmark (all on the y-axis)Some potential factors that can cause patterns or trends in the housing market include:

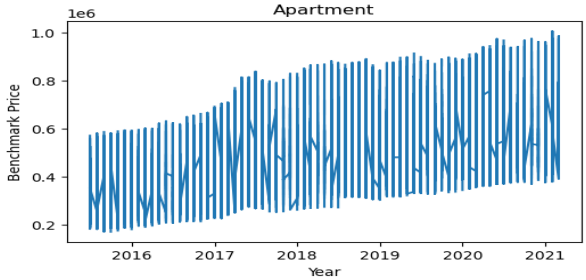


Over the past decade, the Toronto housing market has experienced significant fluctuations in 2017, the market saw a steady increase in home prices, with some years experiencing double-digit growth. This was fueled by low interest rates, a growing economy, and high demand for housing.







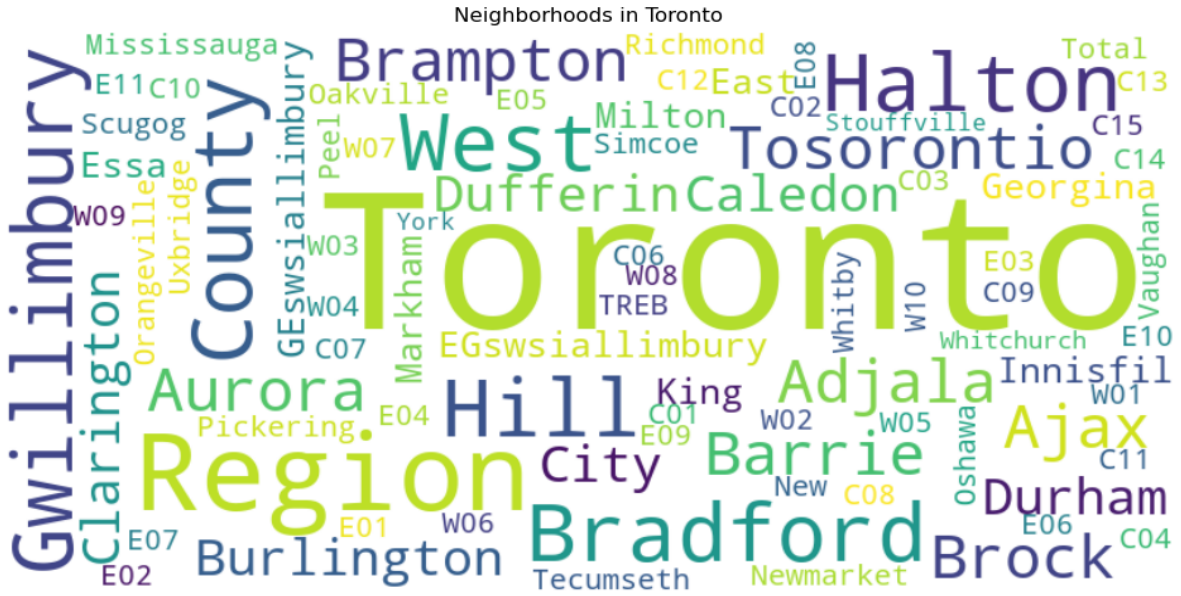


3.How do different neighbourhoods in Toronto compare in terms of housing prices and trends?

Variables:COMPEBENCHMARK,NEIGHBORHOOD

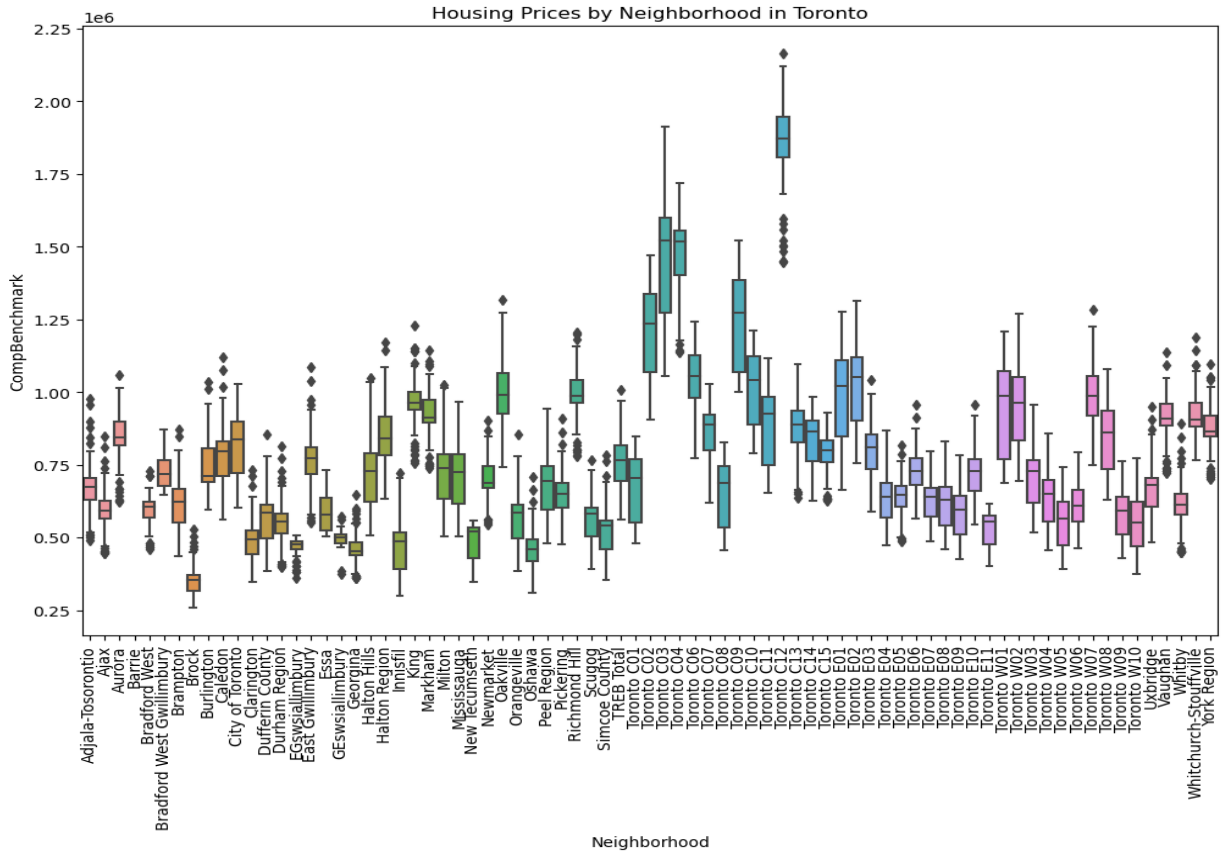
VISULIZATION: **world cloud for popular neighbourhood** around Toronto area.By looking at the **box plot,** we can gain several insights about the Toronto housing market:There is a wide range of housing prices across different neighborhoods in Toronto. Some neighborhoods have higher housing prices than others, which suggests that the demand for housing varies depending on the location.

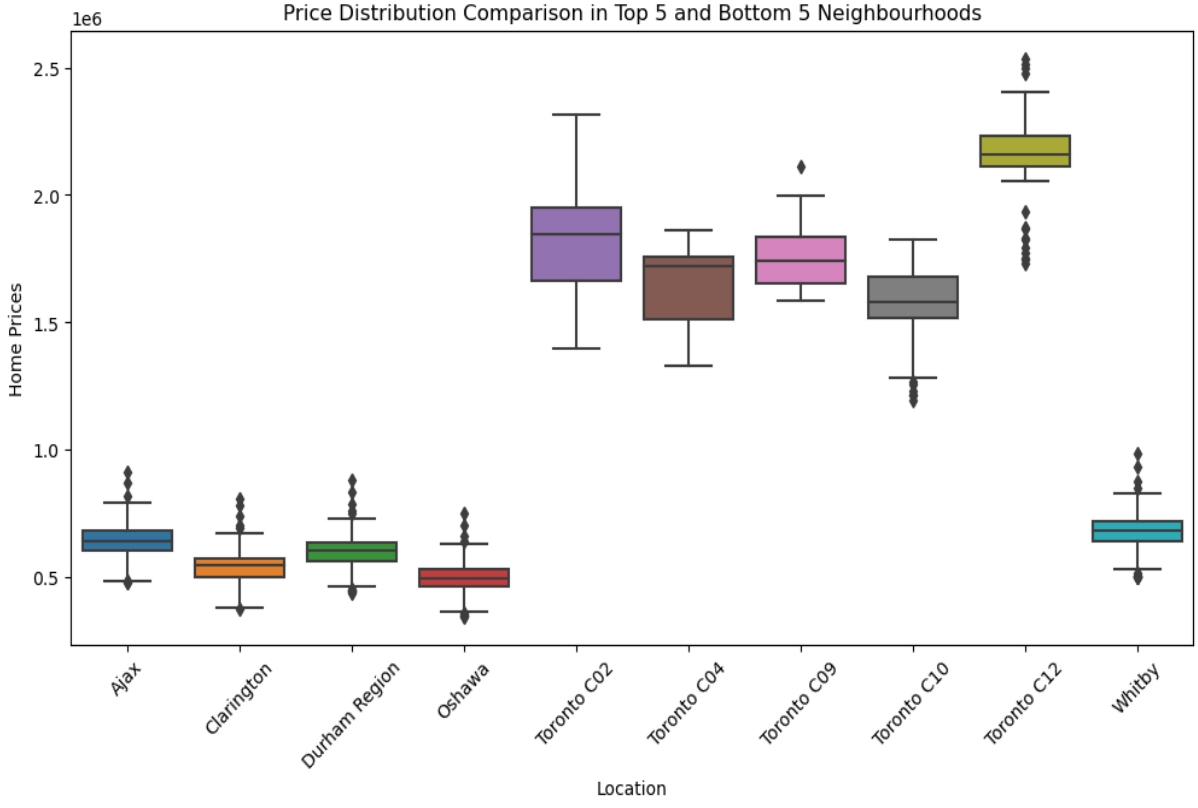
There are several outliers in the box plot, which represent individual data points that fall outside of the range of typical values for that neighborhood. These outliers could be due to a variety of factors, such as unique features or amenities of a property, or other market conditions that are not representative of the overall trend in that neighborhood.

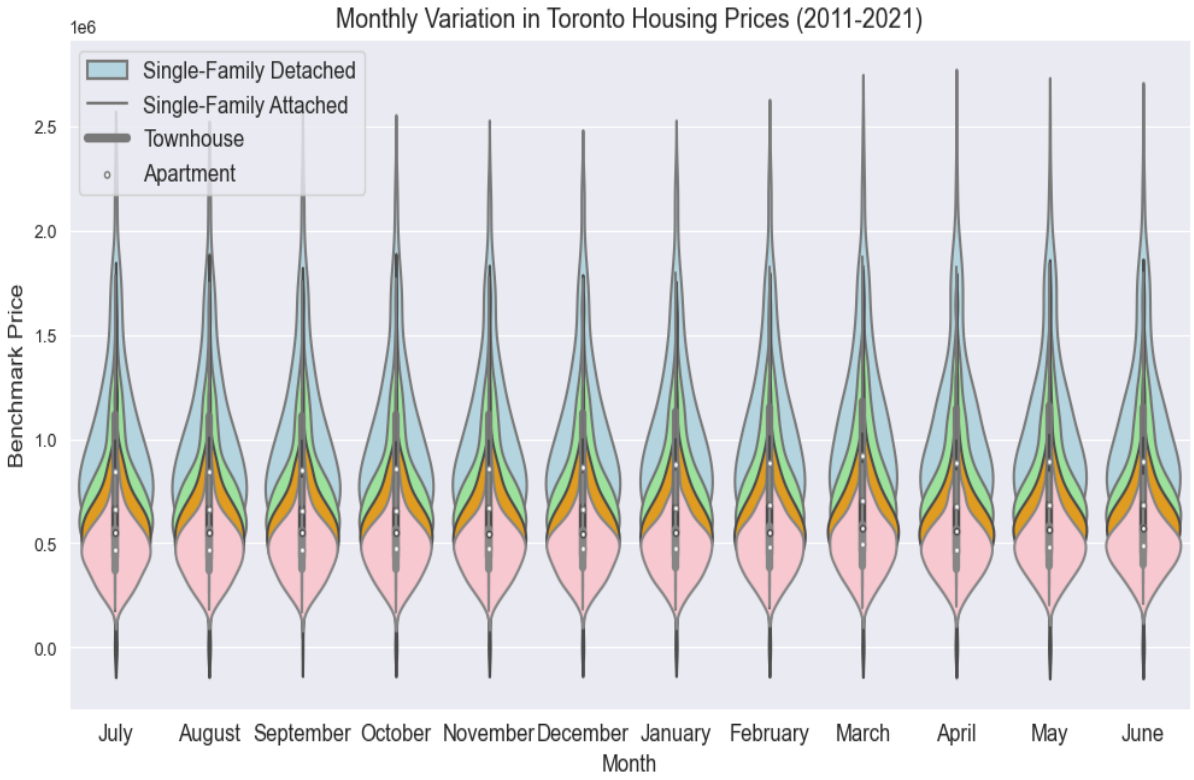


The first letter in the code indicates the district (e.g., "C" for the old city of Toronto, "E" for the former city of East York)C12 includes neighborhoods such as Rosedale, Summerhill, and Moore Park, which are known for their luxury homes, exclusive clubs, and high-end shopping. The average home prices in C12 are among the highest in Toronto, with single detached homes selling for over $3 million and luxury condos selling for over $1 million.

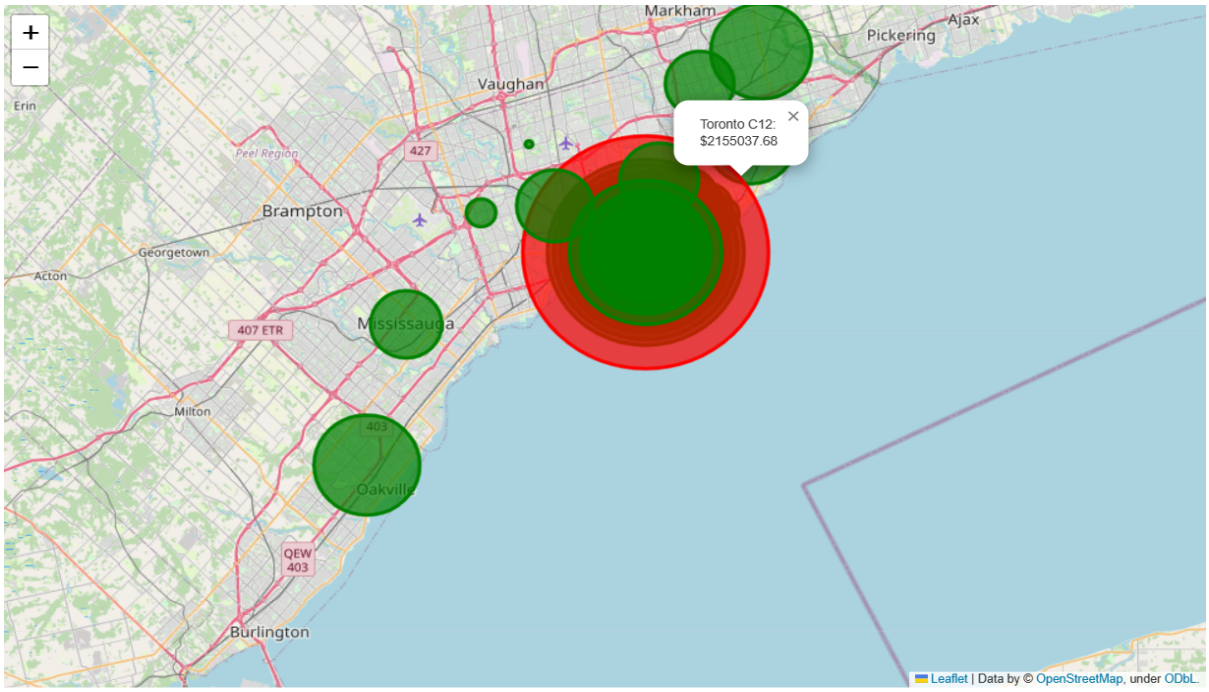
it has a mix of older and newer housing options, ranging from historic homes to newer condos and townhouses. Prices tend to be more affordable in Oshawa compared to the more expensive neighborhoods in Toronto.


Housing Prices by Neighborhood in Toronto
1e6
CompBenchmark
Neighborhood


Price Distribution Comparison in Top 5 and Bottom 5 Neighbourhoods
1e6
Home Prices
Ajax
Clarington
Durham Region
Oshawa
Toronto C02
Toronto C04
Toronto C09
Toronto C10
Toronto C12
Whitby
Location



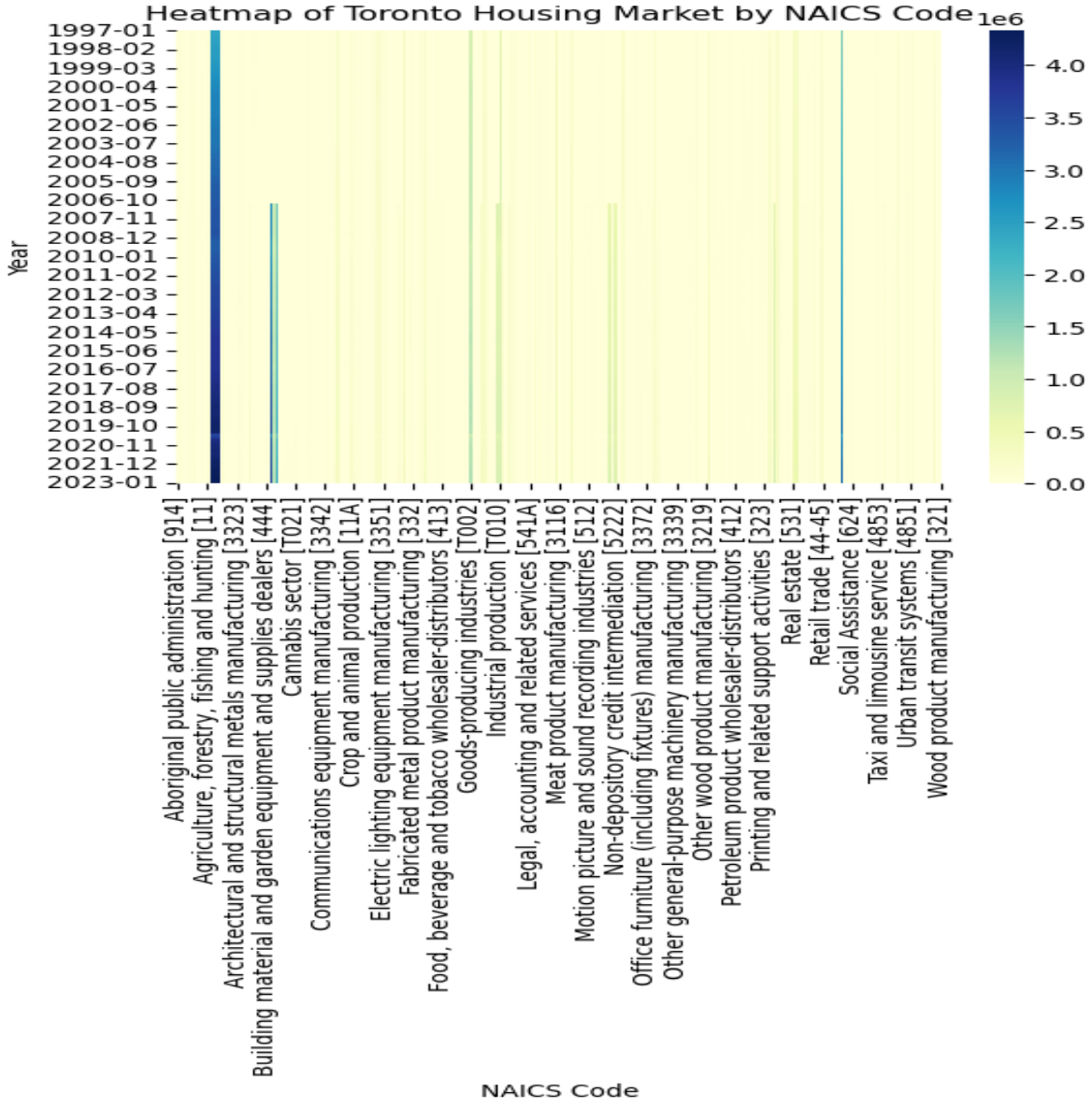
Neighborhood Map



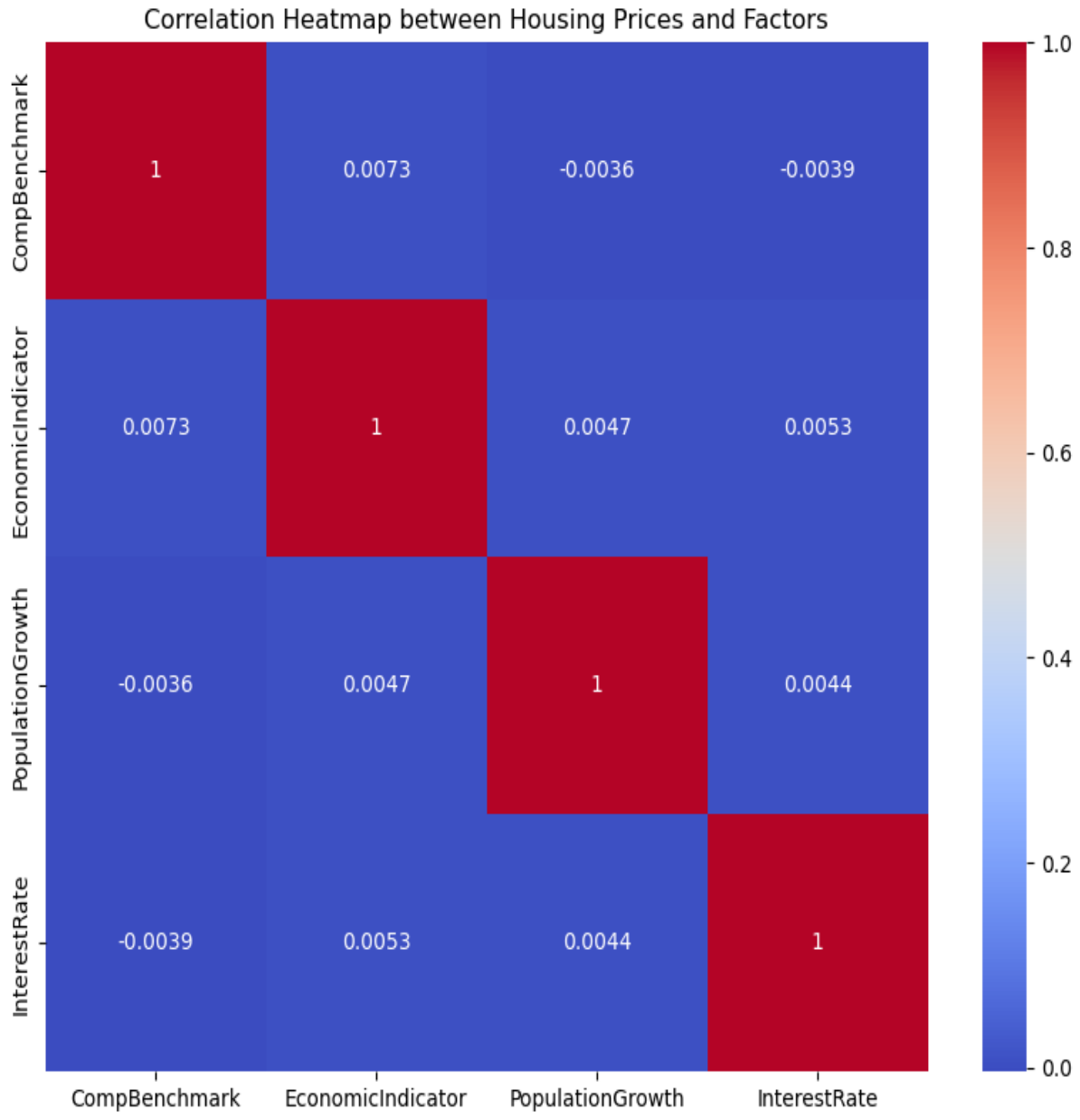
4.What factors are driving changes in the Toronto housing market,such as economy,population growth?

Agriculture, forestry, and fishing industries often require large amounts of land for their operations.If the cost of land in Toronto is high, then the value of properties used for these industries could also be high, which would contribute to the overall value of the Toronto housing market.

Social assistance has high contribution of Toronto housing market because it is related to affordable housing market. I have taken data from **North AMERICAN INDUSTRY CLASSIFICATION SYSYEM.**



Correlation Heatmap between Housing prices with factors



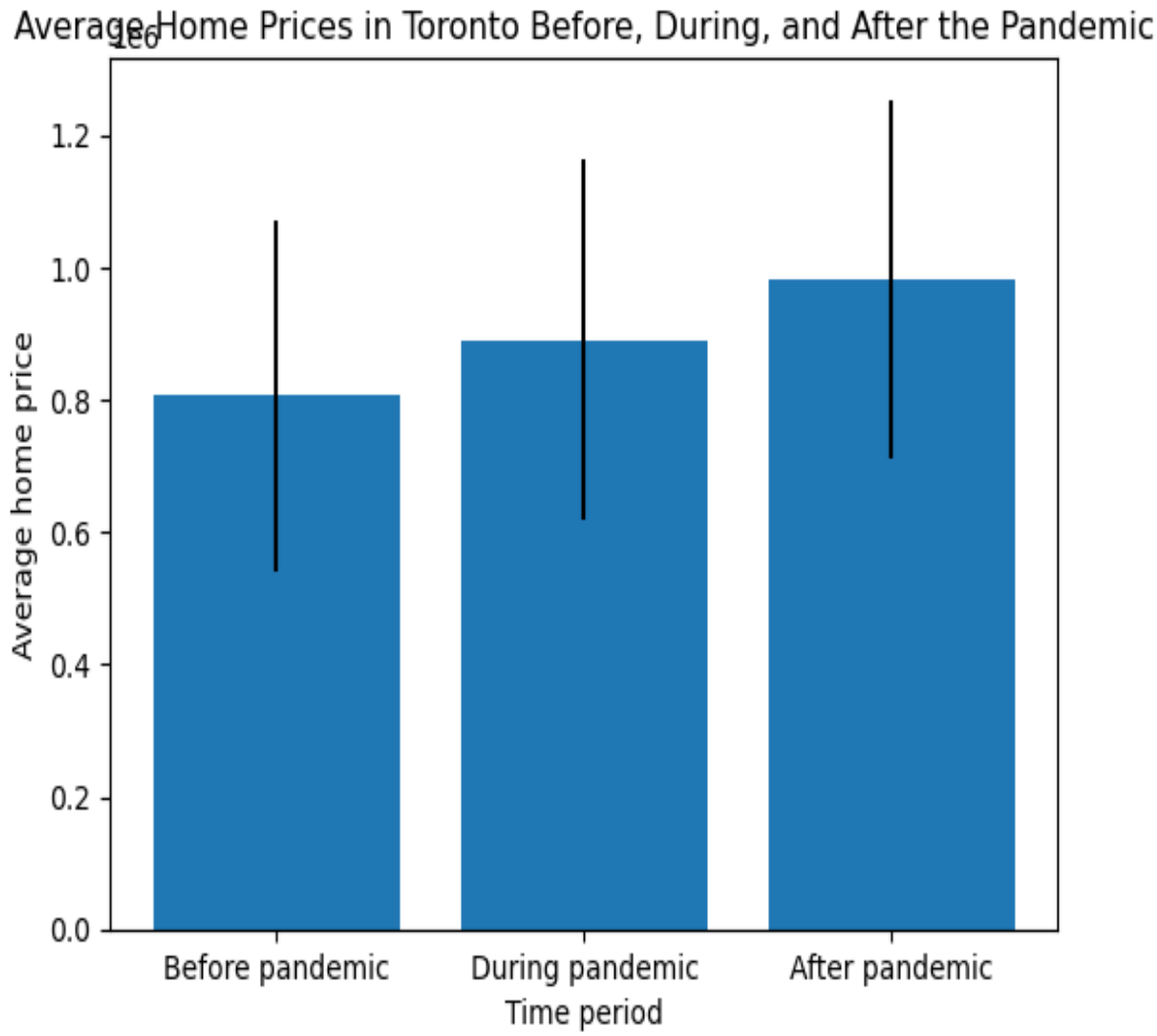
5.How has pandemic affected the Toronto housing market,and what are the long term implications?

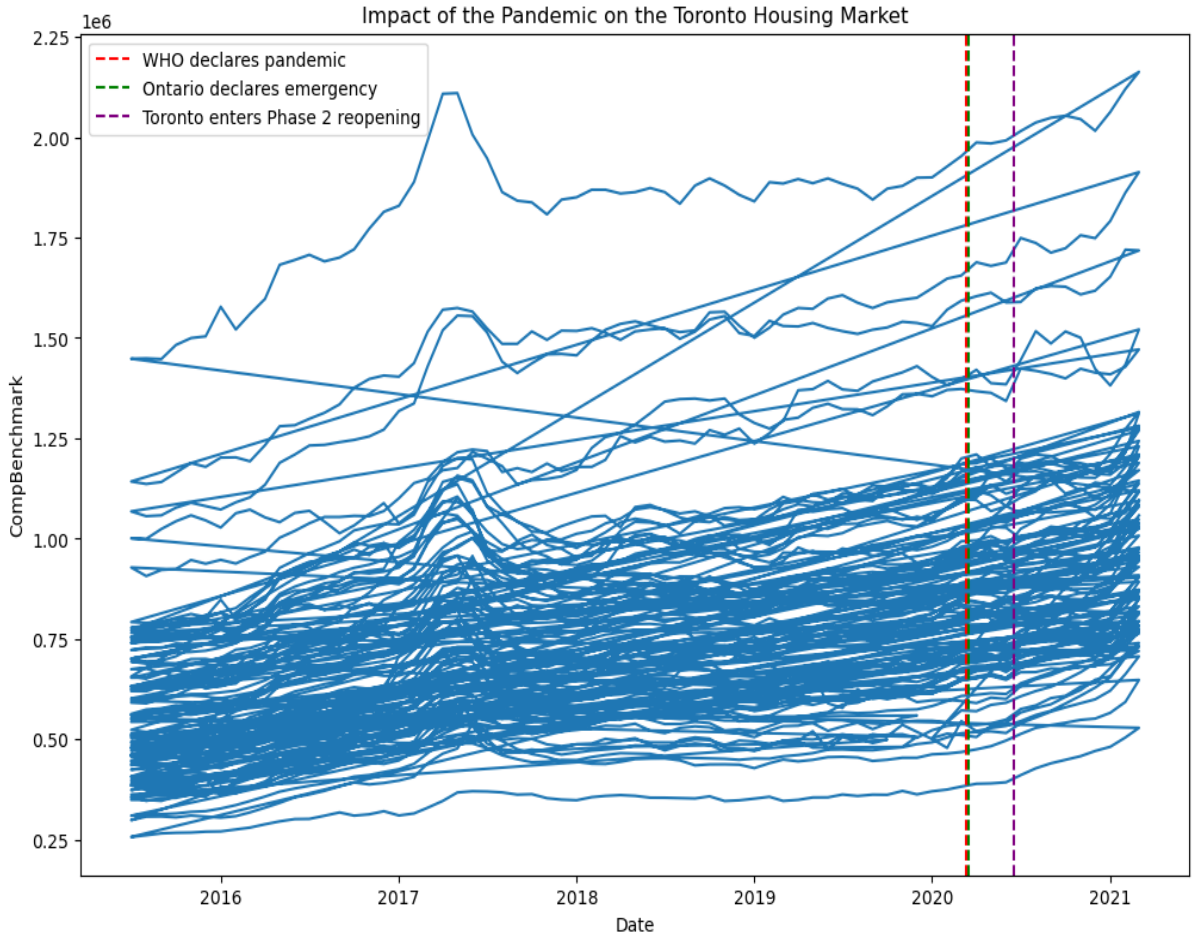
According to the Toronto Regional Real Estate Board (TRREB), the average selling price for all home types in the Greater Toronto Area (GTA) reached $1,097,565 in 2020, an increase of 13.5% compared to the previous year. The price increase was particularly pronounced in the suburban areas surrounding Toronto, where buyers were looking for larger properties with more outdoor space.

Pandemic

0 The pandemic has caused many people to reassess their housing needs, with some looking for larger homes or more outdoor space due to spending more time at home. This has led to increased demand for certain types of properties, such as single-family homes with yards or outdoor space.

1





CONCLUSION AND FUTURE WORK

Based on the analysis of the Toronto housing market,it can be concluded that the average price of a home in Toronto has increased significantly over the past decade based on the clear patterns and trends of housing market.There are factors such as changes in the economy,population growth,and policy changes have played a significant role in driving changes in the Toronto housing market.

The market has also experienced periods of instability and uncertainty, such as during the 2008 financial crisis and the current COVID-19 pandemic. These events have had a significant impact on the Toronto housing market, causing fluctuations in prices and demand.

FUTURE WORK

- Toronto market is complex and multifaceted,and there is still much to be learned and understood about it patterns and trends.
- By continuing to analyze and study the market,we can better understand its dynamics and make more informed decisions about future of housing in Toronto.

References

Toronto Home Price Index | Kaggle

TRREB - MLS Home Price Index Archive

Plotly graphing library for matlab® in MATLAB

North American Industry Classification System (NAICS) Canada 2022 Version 1.0 (statcan.gc.ca)

https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/find-your-neighbourhood/