Importing necessary libraries

In [13]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [14]:
```
#Reading CSV File which was saved earlier and gathered data using API calls.
df=pd.read_csv("datawithPrices.csv")
pd.set_option('display.max_columns', None)
df.head(10)
```

Out[14]:

	Unnamed: 0	Place Name	Category	PostalCode	Borough	Neighborhood	Latitude	Longitude	HousingPrice
0	2	Elgin & Winter Garden Theatre Ctr	Music Venue	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	2200
1	11	Carville Mill Park	Park	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	2200
2	23	Hina's Beauty Care	Hair Salon	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	2200
3	34	Richmond Station	American Restaurant	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	2200
4	44	Cactus Club Cafe	American Restaurant	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	2200
5	46	Markham Corners	Shopping Mall	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	2200
6	54	Cherry beach sports fields	Soccer Field	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	2200
7	63	Metro Grill	Fast Food Restaurant	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	2200
8	82	Toca	Bar	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420	2200
9	87	Akira Back Toronto	Japanese Restaurant	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420	2200

In [15]:
```
df.shape
```

Out[15]:
```
(1242, 9)
```

In [16]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1242 entries, 0 to 1241
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Unnamed: 0    1242 non-null   int64
 1   Place Name    1242 non-null   object
 2   Category      1242 non-null   object
 3   PostalCode    1242 non-null   object
 4   Borough       1242 non-null   object
 5   Neighborhood  1242 non-null   object
 6   Latitude      1242 non-null   float64
 7   Longitude     1242 non-null   float64
 8   HousingPrice  1242 non-null   int64
dtypes: float64(2), int64(2), object(5)
memory usage: 87.5+ KB
```

Most common place type in the data is Restaurant.

In [17]:
```
df.Category.value_counts()
```

Out[17]:
```
Restaurant               53
Park                     52
Café                     50
Bakery                   44
Diner                    39
                         ..
Laser Tag Center          1
Aquarium                  1
Pest Control Service      1
Karaoke Bar               1
Dumpling Restaurant       1
Name: Category, Length: 274, dtype: int64
```

No null values in the dataset.

In [18]:
```
df.isnull().sum()
```

Out[18]:
```
Unnamed: 0      0
Place Name      0
Category        0
PostalCode      0
Borough         0
Neighborhood    0
Latitude        0
Longitude       0
HousingPrice    0
dtype: int64
```

In [19]:
```
#Unique Values
unique_values = pd.DataFrame(
 columns=['Unique Values']
)
for row in list(df.columns.values):
 unique_values.loc[row] = [df[row].nunique()]
unique_values
```

Out[19]:

	Unique Values
Unnamed: 0	1242
Place Name	1242
Category	274
PostalCode	101
Borough	13
Neighborhood	101
Latitude	101
Longitude	73
HousingPrice	13

No missing data in the dataset.

In [20]:
```
missing_data = pd.DataFrame(
 df.isnull().sum(),
 columns=['Missing Values']
)
missing_data
```

Out[20]:

	Missing Values
Unnamed: 0	0
Place Name	0
Category	0
PostalCode	0
Borough	0
Neighborhood	0
Latitude	0
Longitude	0
HousingPrice	0

In [21]:
```
data_types = pd.DataFrame(
 df.dtypes,
 columns=['Data Type']
)
data_types
```

Out[21]:

	Data Type
Unnamed: 0	int64
Place Name	object
Category	object
PostalCode	object
Borough	object
Neighborhood	object
Latitude	float64
Longitude	float64
HousingPrice	int64

Data Quality Report

In [22]:
```
dq_report = data_types.join(missing_data).join(unique_values)
dq_report
```

Out[22]:

	Data Type	Missing Values	Unique Values
Unnamed: 0	int64	0	1242
Place Name	object	0	1242
Category	object	0	274
PostalCode	object	0	101
Borough	object	0	13
Neighborhood	object	0	101
Latitude	float64	0	101
Longitude	float64	0	73
HousingPrice	int64	0	13

Analysis of the data - Overall data quality is good. It has more than 1200 columns and no missing or null values in the dataset. Target column HousingPrice is there to predict its values later on using machine learning models.