Exploring Neighborhoods In Toronto

# Introduction

Almost a year has passed since the World Health Organization (WHO) declared Covid-19 a pandemic. Hundreds of millions of people have lived through lockdowns and city of Toronto is no different. Many of people are working from home and there is no need for them to commute. Due to this reason, there is no need for employees to be close to their office or be located close to hubs and city cores such as downtown. Covid-19 is reshaping our lives.

In this report I will be looking at the issue of housing and how the pandemic has standardized remote work, and what that might mean. Will we go to the office again – and, if so, how often? Since there is no need for people to be located close to their office and most people especially located in small apartments in downtown are deciding to move further away. Some of these employees are parents and have kids or pets. Prior to moving it will be good for these employees to look at neighborhoods that is either similar or has the necessary things they need such as school, university, trails and park, library, restaurants.

In this report I will go through step-by-step process to decide which neighborhood is a good area to move to. I will analyze the neighborhoods to identify the most or at least suitable areas to move to.

## Target Audience

Employees who are working in the downtown area which is expensive are remotely working from home. This could be:

* Parents that want to live in a neighborhood other than downtown and has almost most all the things they currently have.
* Young employees that want to live in a neighborhood other than downtown which is cheaper but at the same time has good bars and restaurants or in general night life.
* Realtors that want general information to sell a house to their clients.

# Data

## Data Sources

In this report, I will explore, segment, and cluster the neighborhoods in the city of Toronto. However, the data is not readily available on the internet. There is a Wikipedia page that has all the information we need to explore and cluster all the neighborhoods in Toronto. The link is provided “<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>”

The above link is only for postal codes that starts with “M”. However, I will try to expand the analysis to more neighborhoods based on information from other sites such as “<http://www.geonames.org/postal-codes/CA/ON/ontario.html>”

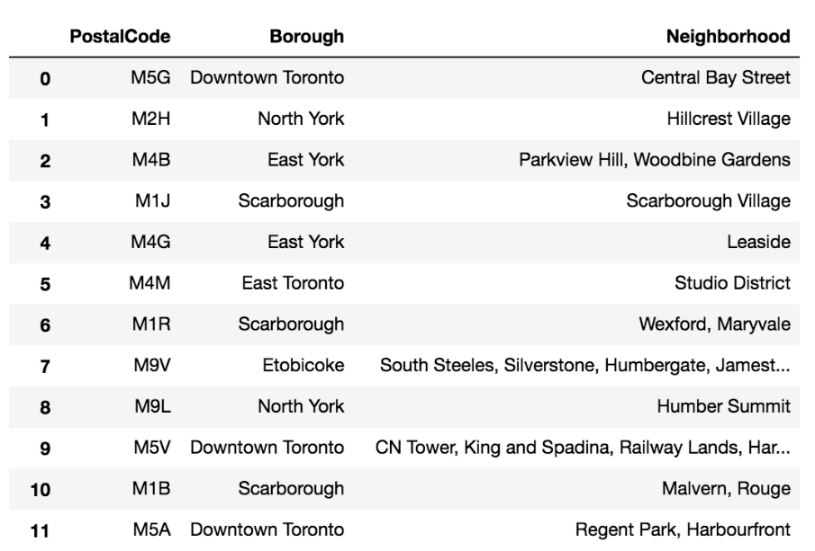
These might not be the most updated information, but it includes most of the neighborhoods needed for this report. I will be scraping these pages and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format.

I’ll be using “[https://cocl.us/Geospatial\_data”](https://cocl.us/Geospatial_data%E2%80%9D) csv file to get all the geographical coordinates of the neighborhoods for the postal codes that start with “M” and extract the postal codes from the rest of the neighborhoods from “<http://www.geonames.org/postal-codes/CA/ON/ontario.html>”.

To get information about venues in different neighborhoods Foursquare’s explore API will be used. This API will retrieve information and details such as name, category, latitude, and longitude for different neighborhoods.

## Data Wrangle & Cleaning

Scraped the following Wikipedia page, “*List of Postal code of Canada: M*” that was mentioned in the previous section in order to obtain the data about the Toronto and the neighborhoods in it. The Dataframe will consist of three columns: PostalCode, Borough, and Neighborhood. Only the cells that have an assigned borough will be processed and Boroughs that is not assigned are ignored. More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma. as shown in row 11 in the above table. If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.



Pandas Module was used to import the table. Sample of the result can be seem in the table below:

# Webpage url

url = 'https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M'

# Extract tables

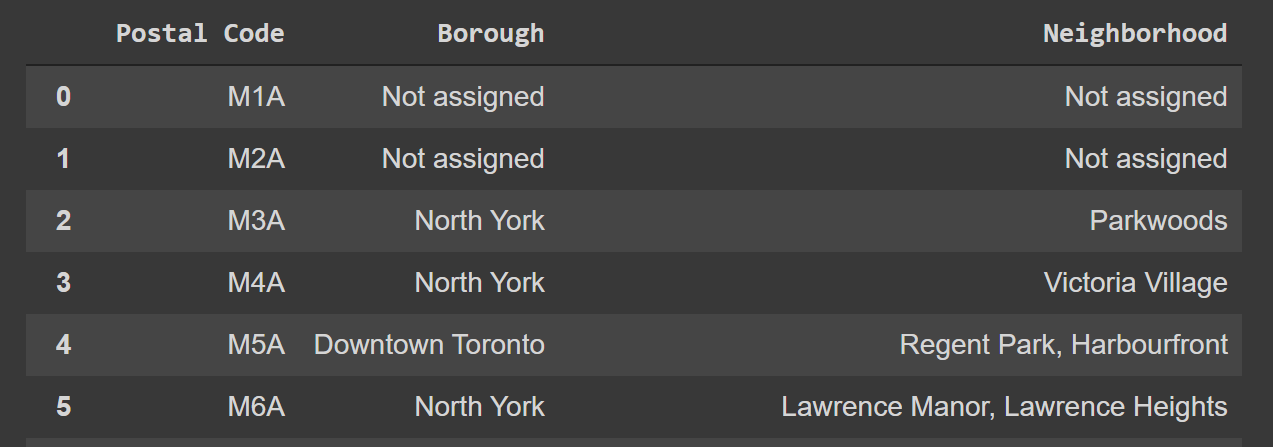
dfs = pd.read\_html(url)

# Get first table

df = dfs[0]

print(df)

print(df.columns)



The dataframe was wrangled in the order of Postal code, Borough and Neighborhood.

#Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned

df2=df[df["Borough"]!="Not assigned"]

df2.reset\_index(drop=True,inplace=True)

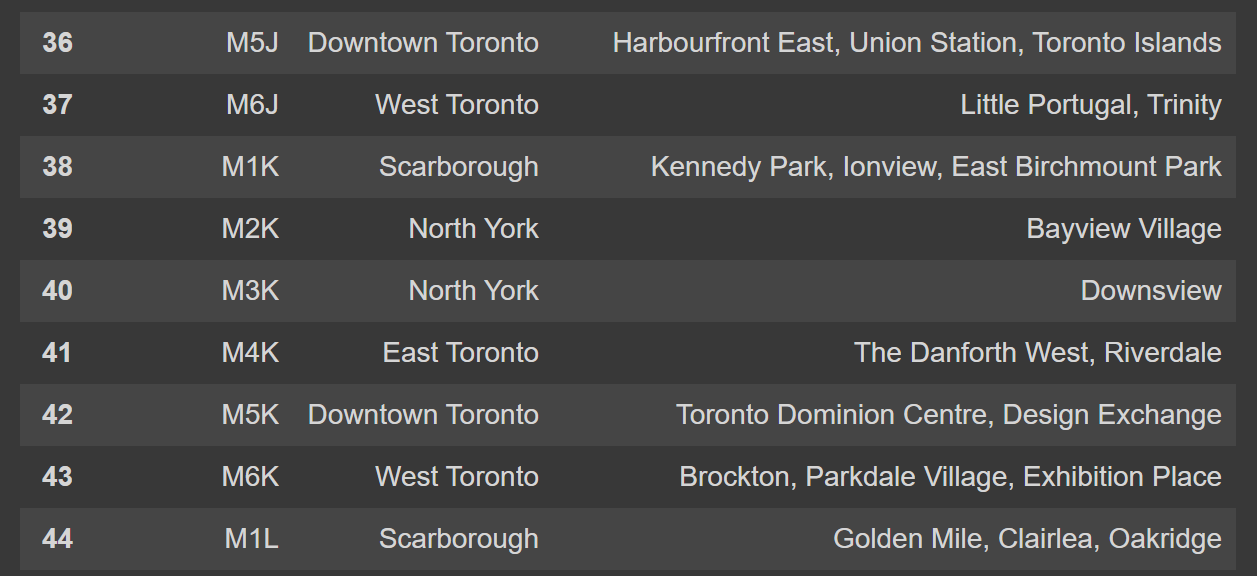
print(df2)

df=df.rename(columns={"Neighbourhood": "Neighborhood"})

for i,row in enumerate(df2["Neighborhood"]):

  if row=="Not assigned":

    df2[i,"Neighborhood"]=df2[i,"Borough"]



Once the above format was achieved the latitude and longitude was added to these neighborhoods. These information was added by merging with the “[http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data)” and I’m combining it with the existing neighborhood dataframe by merging them both based on the postal code.

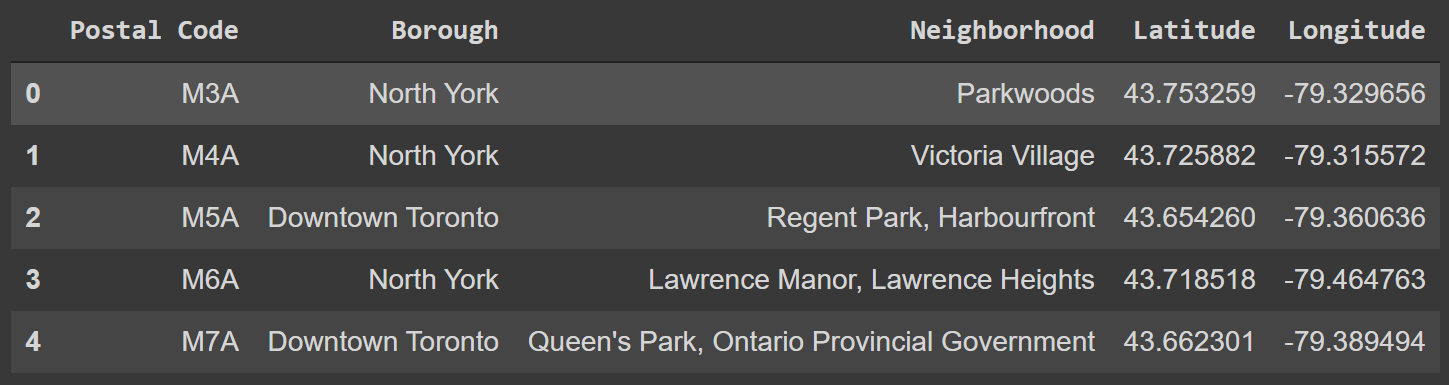
import geocoder # import geocoder

url\_coords='https://cocl.us/Geospatial\_data'

df\_coords=pd.read\_csv(url\_coords)

df\_with\_coord=pd.merge(df2,df\_coords,how='inner',on='Postal Code')

df\_with\_coord.head()



The information of this website <http://www.geonames.org/postal-codes/CA/ON/ontario.html> was also used and formatted into the same format as the above dataframe. The initial table from the website looks like below:



After cleaning up the above data first in Excel by converting it into the proper table form by removing the coordinates to from below the Place column to their own column Latitude and Longitude and removing the Country column and then doing the same process as we did to the Wikipedia information. This tables were appended to the Wikipedia information that has the coordinates. Our final dataframe is now ready for next procedure.

Foursquare was used by creating a function to explore all the neighborhoods by going over the latitudes and longitudes and sending a GET Request.

# function that extracts the category of the venue

def get\_category\_type(row):

    try:

        categories\_list = row['categories']

    except:

        categories\_list = row['venue.categories']

    if len(categories\_list) == 0:

        return None

    else:

        return categories\_list[0]['name']

The JSON information is now ready to be cleaned and structured into a pandas dataframe.

venues = results['response']['groups'][0]['items']

nearby\_venues = json\_normalize(venues) # flatten JSON

# filter columns

filtered\_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']

nearby\_venues =nearby\_venues.loc[:, filtered\_columns]

# filter the category for each row

nearby\_venues['venue.categories'] = nearby\_venues.apply(get\_category\_type, axis=1)

# clean columns

nearby\_venues.columns = [col.split(".")[-1] for col in nearby\_venues.columns]

nearby\_venues.head()

The information will look similar to the table below:

