IBM – Coursera

Data Science Specialization

Capstone project - Final report

Choosing the best residence based on neighboring amenities.

Md Ibtehajul Islam – 2019

Table of content

Table of Figures:

1. Introduction

1.1 Background

Toronto is the most populous city of Canada and the provincial capital of Ontario. For being welcoming to new immigrants, Toronto is full of diversity and rising opportunities. These opportunities attract many individuals from all over the world to move into this city. It is tough for a person to move from one city and rent a suitable place for living in such a big city according to the individual's convenience. Every person has their preferences in terms of their choice of residence. These preferences vary from person to person based on the surrounding amenities of the apartments or the neighborhoods. For example – one person may prefer a grocery store near his/her living place. Similarly, a specific restaurant type or bars might be the priority to another individual. So, the location of the residence is one of the most important decisions that will determine whether the nearby amenities are convenient or not according to the individual need.

1.2 Problem

The objective of this capstone project is to find a suitable residence for any person moving to the city of Toronto that will cover most of his/her nearby amenity preferences. This project will suggest suitable residential buildings or a neighborhood to narrow down any individual's list while finding a suitable place for living. Using data science methodology and Foursquare API this project aims to provide the solutions to answer the business question: If a person is moving to Toronto, Ontario, Canada, with a new job, where he/she should look for a residence near the workplace that has most of the amenities in his/her list of preference?

1.3 Interest

This project is particularly useful to:

- Any student, graduate or any individual starting a job in Toronto,
- Families migrating from a different city, province or country moving to Toronto.

2 Data acquisition and cleaning

2.1 Data sources

- **Location of interest:** It is mainly the address of the workplace or the center of interest based on what the whole analysis will be conducted. For this project we are considering **'301 Front St W, Toronto, ON' –** address of CN tower, as our workplace location.
- **Apartment data:** This the data that contains all the nearby Residential Building (Apartment / Condo) with in 5 Km distance from the location of interest. This data mainly contains the latitude and longitude coordinates of those apartments. This is required in order to plot the map and to get the nearby venue data.
- **Postal codes of Canada:** (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) This data contains the Postal codes, Borough and the Neighborhoods of Toronto city. This data was merged with the '**Apartment data**' on the postal codes to check the Boroughs of each apartments within the radius of the location of interest.
- **Geospatial data:** (http://cocl.us/Geospatial_data) This data contains the latitude and the longitude coordinates of all the Postal codes of Toronto city. This data was merged with the

'**Postal codes of Canada**' data to get the neighborhoods with corresponding coordinates. This data is useful to find a residence based on the neighborhood amenities instead of the apartments.

2.2   Reasons

Toronto city neighborhoods were chosen as the observation target due to the following reasons:

- The availability of higher job opportunity,
- The availability of geo data which can be used to visualize the dataset onto a map.

2.3   Data acquisition

- Location of interest or the workplace data was chosen randomly at the heart of Toronto downtown by searching on google.
- Apartment data was found using foursquare API by searching for the nearby Residential Building (Apartment / Condo) with in 5 km distance.
- Scraped Wikipedia page and transformed **Postal codes of Canada data** into a structured data frame.
- Geospatial data of Toronto neighborhoods were found from week 3 module of the IBM capstone project course and merged with the 'Postal codes of Canada' data.

2.4   Data cleaning

Initially apartment data was retrieved from the location data. Then, the data was filtered by keeping required features for the analysis, dealt with missing values and cleaned by extracting information from unnecessary elements. Additionally, the feature names were changed in a meaningful way.

categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet
[{'id': '4d954b06a243a5684965b473', 'name': 'R...	False	4e040bb0b61c7592ae6e588b	16 willison sq	CA	Toronto	Canada	spandina AVe
[]	False	5364e2f6498e2cafa29810aa	73 Richmond St. West	CA	Toronto	Canada	NaN
[{'id': '4bf58dd8d48988d1fa931735', 'name': 'H...	False	4ad4c05bf964a5209df520e3	77 Ryerson Ave	CA	Toronto	Canada	NaN
[{'id': '4d954b06a243a5684965b473', 'name': 'R...	False	542cc75a498e68715daf6f88	822-65 Scadding Avenue	CA	Toronto	Canada	NaN
[{'id': '4bf58dd8d48988d12a941735', 'name': 'C...	False	4fca7852e4b0c116823a6344	NaN	CA	NaN	Canada	NaN

Figure 1: Initial apartment dataset

	Apt_Name	Address	Distance	Formatted_Address	Latitude	Longitude	Postalcode
0	University Apartment	16 willison sq	1427	[16 willison sq (spandina AVe), Toronto ON M5T...	43.652168	-79.398821	M5T
1	Elm Place Apartments	222 Elm St.	1576	[222 Elm St. (Elm & Mccaul), Toronto ON M5T 1K...	43.656243	-79.392139	M5T
2	Epitome Apartments	160 Huron Street	1836	[160 Huron Street, Toronto ON M5T 2B6, Canada]	43.657119	-79.397814	M5T
3	1 Homewood Ave Apartments	1 Homewood Ave	2514	[1 Homewood Ave (Carlton St), Toronto ON M4Y 2...	43.663222	-79.374450	M4Y
4	Allan Plaza Apartments	166 Carlton Street	2568	[166 Carlton Street, Toronto ON M5A 2K5, Canada]	43.663463	-79.373573	M5A

Figure 2: Cleaned apartment dataset

Later, the postal codes and the boroughs were extracted from **Postal codes of Canada** dataset and merged with the apartment dataset to identify the boroughs of the apartments. Final dataset for analysis is given below:

	Apt_Name	Address	Distance	Formatted_Address	Latitude	Longitude	Postalcode	Borough
0	University Apartment	16 willison sq	1427	[16 willison sq (spandina AVe), Toronto ON M5T...	43.652168	-79.398821	M5T	Downtown Toronto
1	Elm Place Apartments	222 Elm St.	1576	[222 Elm St. (Elm & Mccaul), Toronto ON M5T 1K...	43.656243	-79.392139	M5T	Downtown Toronto
2	Epitome Apartments	160 Huron Street	1836	[160 Huron Street, Toronto ON M5T 2B6, Canada]	43.657119	-79.397814	M5T	Downtown Toronto
3	1 Homewood Ave Apartments	1 Homewood Ave	2514	[1 Homewood Ave (Carlton St), Toronto ON M4Y 2...	43.663222	-79.374450	M4Y	Downtown Toronto
4	Cromwell Apartments	55 Isabella Street	2869	[55 Isabella Street, Toronto ON M4Y 1M8, Canada]	43.668187	-79.383245	M4Y	Downtown Toronto

Figure 3: Final apartment dataset for analysis

2.5   Solution process

- Initially a location is required to look for nearby neighborhoods or residences. **Location of interest** data will provide the nearby residential building (Apartment / Condo) data within 5 Km distance.
- **Apartment data** and the **Postal codes of Canada** data will be useful to get the postal codes, neighborhoods, boroughs, distances from the location of interest and nearby amenities from the apartments to analyze the residential buildings.
- **Geospatial data** will provide the coordinates of the Toronto neighborhoods those will be used to analyze the neighborhood amenities for choosing the best one among others if analyzed by the neighborhoods instead of residences.

3   Methodology:

Scenarios for choosing a residence can vary from person to person. In this project, the idea of finding the perfect place for living is cantered by a workplace or location of interest. One method is to identify the residential building (apartment / condo) with in a certain distance and selecting the best among those by comparing the individual need. And, the other method can be identifying the nearby neighborhoods and choosing the convenient one based on the priority.

3.1   Analysis by apartments

In this project, foursquare API was used to find nearby apartments/condos by centering the workplace location within 5 kilometers. But the taste and the priority of every individual is different. For example- a person may prefer a place close to his/her workplace to save travelling time whereas others may consider surrounding amenities on top of short distance.

3.1.1   Exploratory Data Analysis

3.1.1.1   Distance

At first, the distances between the workplace and the apartments were analyzed for those individuals who prefer short distances. Below diagram illustrates the geographical location of the apartments that were retrieved from the foursquare API and roughly provides the distance comparison:
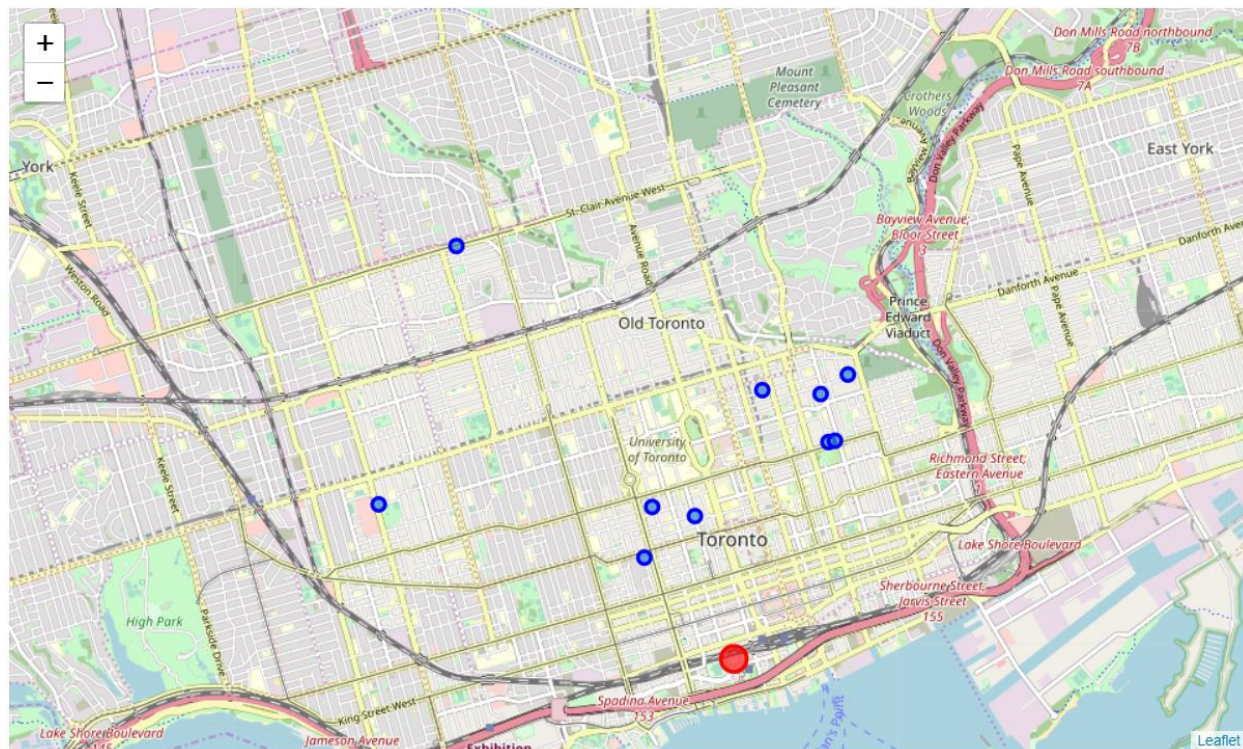
Figure 4: Geographical locations of the apartments

3.1.1.2   Amenities

The nearby amenities were analyzed by exploring each apartment within 500 meters, for those persons prioritize specific amenities over distances. Below data frame shows the top ten amenities around the apartments:

Apt. Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1 Homewood Ave Apartments	Japanese Restaurant	Coffee Shop	Gay Bar	Hotel	Grocery Store	Fast Food Restaurant	Diner	Ramen Restaurant	Gastropub	Falafel Restaurant
Allan Plaza Apartments	Coffee Shop	Gastropub	Restaurant	Grocery Store	Fast Food Restaurant	Japanese Restaurant	Diner	Hotel	Pub	Deli / Bodega
Bunker Apartment	Coffee Shop	Ice Cream Shop	Bakery	Italian Restaurant	Restaurant	Café	Fast Food Restaurant	Sushi Restaurant	Pizza Place	Thai Restaurant
CAPREIT Toronto Apartments - 100 Wellesley St East	Coffee Shop	Grocery Store	Pizza Place	Breakfast Spot	Dance Studio	Burger Joint	Caribbean Restaurant	Shopping Mall	Rock Club	Restaurant
Cromwell Apartments	Coffee Shop	Sushi Restaurant	Café	Japanese Restaurant	Restaurant	Burger Joint	Yoga Studio	Thai Restaurant	Gay Bar	Gym
Dufferin Park Apartment	Bakery	Bar	Coffee Shop	Park	Café	Clothing Store	Cocktail Bar	Restaurant	Beer Store	Skating Rink
Elm Place Apartments	Coffee Shop	Sandwich Place	Café	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant	Ice Cream Shop	Bar	Ramen Restaurant	Arts & Crafts Store
Epitome Apartments	Café	Vegetarian / Vegan Restaurant	Chinese Restaurant	Bar	Vietnamese Restaurant	Mexican Restaurant	Dumpling Restaurant	Bakery	Coffee Shop	Ice Cream Shop
RodXtein Apartment	Coffee Shop	Café	Pizza Place	Grocery Store	Food & Drink Shop	Library	Market	Filipino Restaurant	Park	Breakfast Spot
University Apartment	Bar	Café	Vegetarian / Vegan Restaurant	Chinese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Burger Joint	Dumpling Restaurant	Dessert Shop	Record Shop

Figure 5: Top ten amenities around the apartments

This data frame will help a person to choose the nearby apartment surrounded by his/her choice of amenity.

3.2   Analysis by neighborhoods

Similarly, a neighborhood can be chosen instead of apartments. This project can help to find the boroughs located within 5 kilometers of the workplace and get all the neighborhoods in those boroughs. Then all the neighborhoods can be compared by checking corresponding distances from the workplace and exploring the nearby amenities. The goal here is to choose the convenient neighborhood and find residence in it.
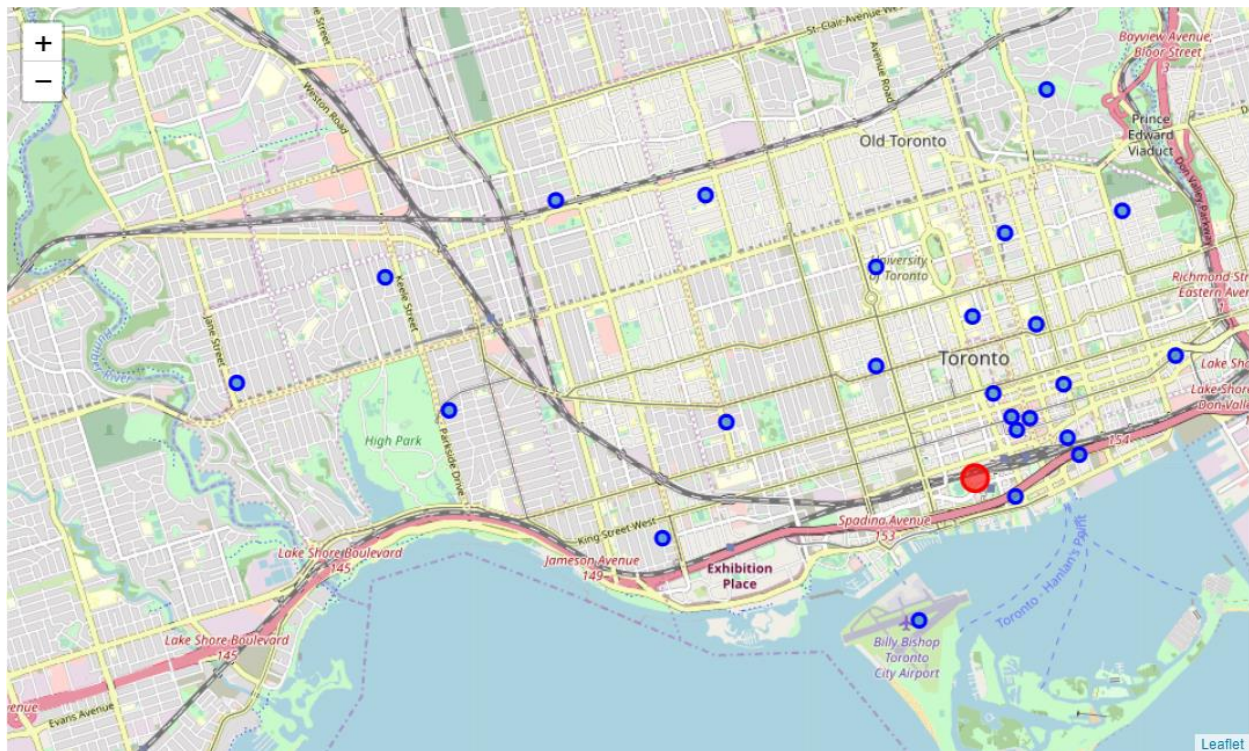

Figure 6: Geographical locations of the neighborhoods

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Thai Restaurant	Steakhouse	Cosmetics Shop	Asian Restaurant	Gym	Burger Joint	Hotel
1	Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Steakhouse	Cheese Shop	Café	Farmers Market	Beer Bar	Seafood Restaurant	Breakfast Spot
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Coffee Shop	Grocery Store	Intersection	Falafel Restaurant	Convenience Store	Burrito Place	Restaurant	Stadium
3	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Lounge	Airport Terminal	Boutique	Airport	Airport Food Court	Airport Gate	Sculpture Garden	Bar	Boat or Ferry
4	Cabbagetown, St. James Town	Coffee Shop	Restaurant	Café	Pub	Italian Restaurant	Pizza Place	Bakery	Market	Outdoor Sculpture	Butcher

Figure 7: Top ten amenities around the neighborhoods

3.2.1 Reasons

There are several reasons of analyzing neighborhoods over residential buildings for finding a residence. In real life scenario, townhouses, independent houses, basements and single rooms can also be rented. But those do not appear in APIs while looking for residences. Moreover, there are 2 boroughs and 24 neighborhoods within 5 kilometers of the workplace location used in this project. It will consume a long time to check every residence from every neighborhood and find a suitable one. In this case, this project can help to find a suitable neighborhood among others by exploring surrounding amenities. Then it will be easier to find a convenient place for living in that neighborhood in big city like Toronto.

4 Results

4.1 Observations

The goal of this project was to suggest suitable residential buildings or a neighborhood to narrow down his/her list while finding a place for living. Findings are given below with examples –

Scenario 1: If a person wants to minimize his travelling time by reducing distances between his home and workplace which apartment should be considered?
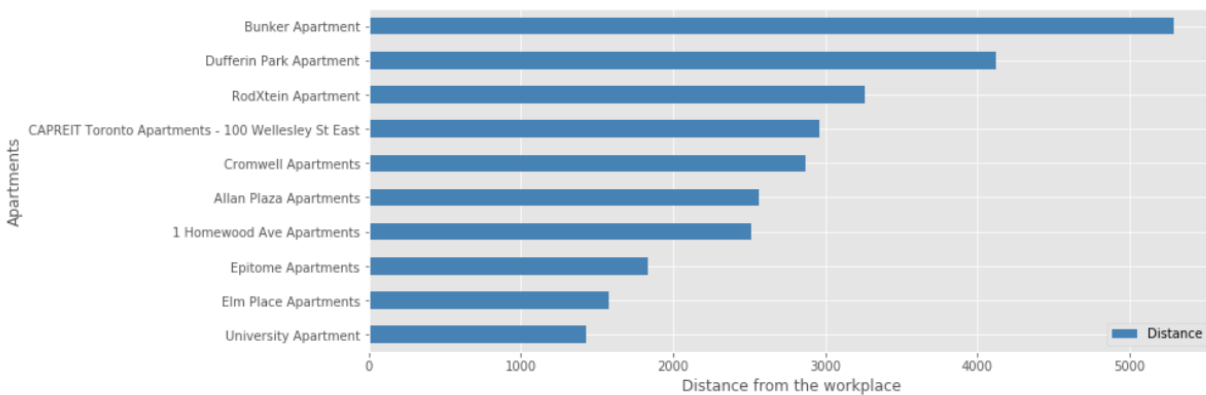

Figure 8: Distances from the workplace

From the above figure the distance comparison is clearly visible, and a person can easily choose his preferred home from the list.

Scenario 2: If a person's priority is to look for an apartment near grocery stores, which one he/she should choose?

	Apt. Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1 Homewood Ave Apartments	Japanese Restaurant	Coffee Shop	Gay Bar	Diner	Gastropub	Grocery Store	Hotel	Fast Food Restaurant	Ramen Restaurant	Pet Store
1	Allan Plaza Apartments	Coffee Shop	Gastropub	Restaurant	Diner	Grocery Store	Japanese Restaurant	Fast Food Restaurant	Hotel	Pub	Garden
3	CAPREIT Toronto Apartments - 100 Wellesley St ...	Coffee Shop	Grocery Store	Dance Studio	Pizza Place	Breakfast Spot	Caribbean Restaurant	Park	Smoothie Shop	Steakhouse	Wine Shop
7	RodXtein Apartment	Coffee Shop	Pizza Place	Café	Grocery Store	Hotel	Bistro	Playground	Diner	Dive Bar	Park
9	University Apartment	Bar	Café	Vegetarian / Vegan Restaurant	Chinese Restaurant	Dessert Shop	Vietnamese Restaurant	Mexican Restaurant	Dumpling Restaurant	Burger Joint	Grocery Store

Figure 9: Inquiry for a specific venue of choice for comparing the residential places.

According to the result found in figure 9, **1 Homewood Ave Apartments, Allan Plaza Apartments, CAPREIT Toronto Apartments, RodXtein Apartment and University Apartments** are the best choice for him/her from all 10 residential buildings.

Similarly, if a person's priority is to look for a neighborhood near grocery stores, which one he/she should choose?

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Coffee Shop	Grocery Store	Intersection	Falafel Restaurant	Convenience Store	Burrito Place	Restaurant	Stadium
7	Christie	Grocery Store	Café	Park	Nightclub	Athletics & Sports	Diner	Restaurant	Baby Store	Italian Restaurant	Coffee Shop
16	High Park, The Junction South	Mexican Restaurant	Café	Bar	Grocery Store	Furniture / Home Store	Fast Food Restaurant	Bookstore	Flea Market	Cajun / Creole Restaurant	Speakeasy

Figure 10: Inquiry for a specific venue of choice for comparing the neighborhoods.

According to the result found in figure 10, **Brockton, Exhibition Place, Parkdale Village, Christie, High Park and The Junction South** are the best choice for him/her from all 49 neighborhoods in Central, downtown and west Toronto.

4.2   Limitations

Even though this project is performing well to narrow down the list of choice but due to some limitation's, retrieval of all the residential buildings was poor and that caused the Apartment dataset to become smaller. Thus, it cannot be used to precisely find all the residential places.

Explanations for the poor findings can be –

- Limitations of retrieval of residential buildings in foursquare API,
- The dataset and the information were not enough (small dataset)

But again, on the bright side, the insight, gotten from observing the analysis results, seems consistent and logical. And the insight can serve the needs of most normal people finding the preferred neighborhoods.

5   Discussion

I was able to locate residential buildings and neighborhoods within 5 kilometers of distance from the location of interest, and their corresponding surrounding facilities within 500 meters of the radius. However, foursquare API was unable to return all the apartment buildings, independent houses, townhomes. As foursquare technology crowd-sources their data, the locations like different houses are not available on the API. Apart from that, rent plays a vital role in terms of choosing a place for most of the people. But usually, the average rent data is not publicly available for different neighborhoods that lead the project without considering the major part like rent. Moreover, there were some real challenges in constructing the data set those are given below:

- For dataset constructed through API calls, had a lot of missing and unnecessary information.

- Similarly, scrapping data from the Wikipedia page and merging multiple sources took lots of efforts to check, research and change to decide which data should be kept or transformed before the merge.

It can be considered the most important process in the whole data science pipeline. Which can affect the most on the result.

6   Conclusion

In this study, I analyzed the nearby surrounding amenities both for residential buildings and neighborhoods, centering a location of interest or workplace location. I identified residences and neighborhoods adjacent to a workplace and their corresponding facilities according to individual convenience. I used geolocator to get the coordinates, and foursquare API to collect the distances from the work location and the information of neighboring amenities close to the residences. This project can help people finding a new place for living in Toronto in several ways. For example, it can help a person identify the nearest home or neighborhood from the workplace as well as locate a perfect residence surrounded by an individual's preferred amenities.