Coursera Capstone Project

The Battle of Neighborhoods: United States
Final Report

Composed by: Ekaterina Deriabina

Problem Definition

1. Introduction

The United States today are not only one of the cultural and economic centers of the world, but also a major scientific and technological center. This in turn leads to a constant growth and development of the cities, and also to an increasing demand on housing. Accordingly, not only the cost, place and availability of public transport, but also access to public places such as shops and kindergartens are becoming the criteria for choosing the place to live.
Customers, making a decision on the acquisition of real estate in different parts of the USA, are guided today by a combination of factors, the analysing of which is not possible without the use of machine learning due to extremely large amounts of data. New technologies come to the help to analyze large amounts of data, as well as present the results in a form convenient for buyers.
The goal of this project is to develop a machine learning algorithm that will simplify the search for real estate for purchase.
As an example the city of Seattle is shown, as the list of the postal codes of Seattle is easier to work with, downloaded from the government web-site.

2. Intended users

Real estate agents, that need the tools to analyse properties of housing and customers demands to develop the best offer of the property in the UK.

3. Stakeholders

Government of the USA, Real estate agents, Clients

Data

In this project the following data were used:

Data	Name	Type	Condition	Content	Source
Zip Codes	List of zip codes	Wiki	N/A	List of zip codes of Seattle	
Geocoding	Geocoding API	JSON	N/A	Google Maps API calls	Google Maps
Forsquare	Foursquare location data	JSON	N/A	Foursquare API calls	https://foursquare.com/

In my project the API Forsquare were used as data source of all the required information about the schools, shops, public places and their ratings. To get the information about the location of the required objects (latitude and longtitute) the zip codes of the required city were used.
To avoid the problems with the requests limitation in the Foursquare API the radius was selected to be 1000, and the places number is to be 500.

To make the choice of the housing easier, the five neighborhoods were compared with the following parameter: the schools ratings and the pricing of housing. Moreover the amount of the public places to visit were compared on the basis of the most common venues.

The comparation of the communities were performed with the using of API Foursquare, whereas the data was worked through with such python's libraries as pandas, numpy, scilearn kit. The detailed description of the used libraries are given in the following table:

Pandas	Library for Data Analysis
NumPy	Library to handle data in a vectorized manner
JSON	Library to handle JSON files
Geopy	To retrieve Location Data
Requests	Library to handle http requests
Matplotlib	Python Plotting Module
Sklearn	Python machine learning Library
Folium	Map rendering Library

Methodology

The used Dataset includes information about the property prices and the schools rating for different communities in Seattle, USA.

The goal of the project is to generate the automated algorithm, that will help the real estate managers to find the best choice of the property for the client on the basis of the demands. As an example the schools ratings and the house pricing data base were used.

The generated script fulfils the following functions:

- The postal codes of the USA are cleansed and prepared for using
- The coordinates of the streets are obtained from Google Maps to build the data frage to analyse
- The information about the venues is obtained from the Foursquare API calls to develop the working data base
- The gathered data frame is used to find out, the best five communities to further comparation on the base of budget of the client and his/her field of interest in public places
- The found five communities are compared with more details to find the best choice for the client

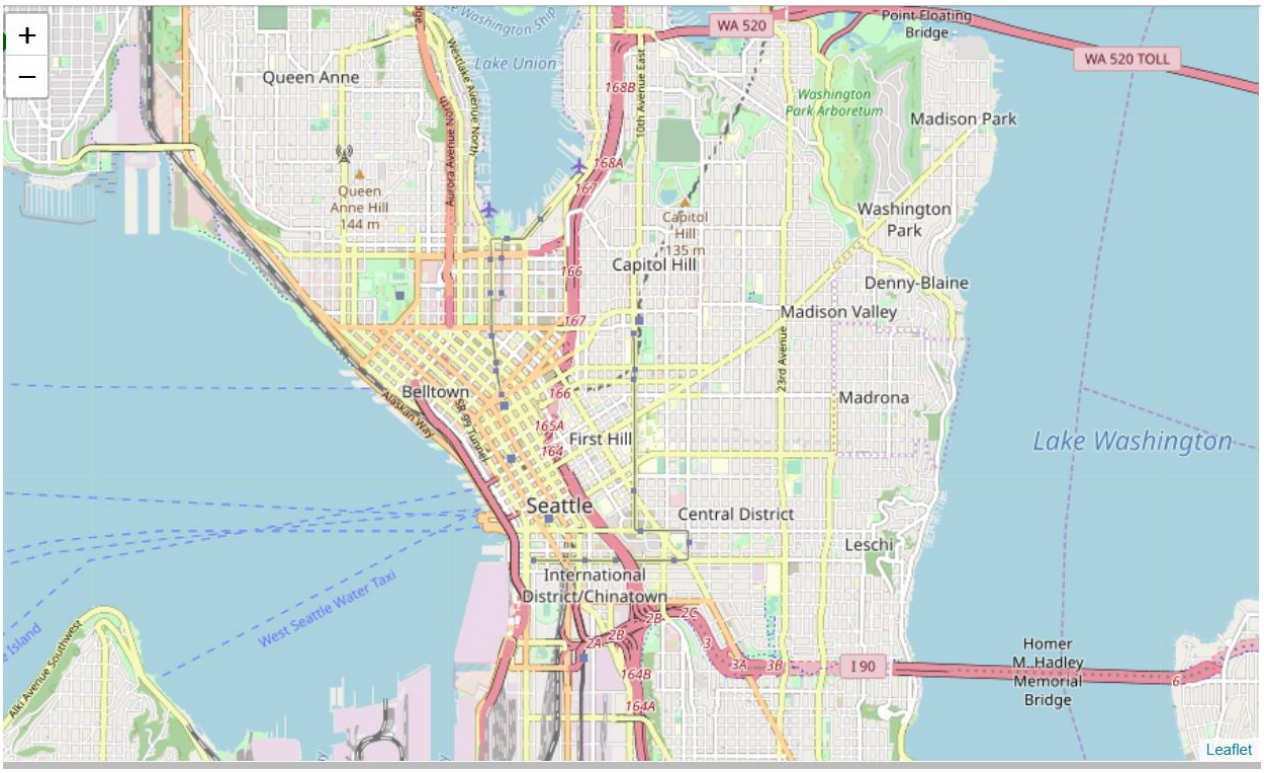
Results

The machine learning algorithm uses the list of the postal codes from the government web-site and gives the list of communities of the city Seattle, in which the client can search for the property.

	Post_Code	Neighborhood_Name
0	98003	Federal Way
1	98005	Bellevue
2	98033	Kirkland
3	98037	Lynnwood
4	98040	Mercer Island
5	98052	Redmond
6	98055	Renton
7	98101	Seattle
8	98101	Downtown
9	98102	Capital Hill
10	98103	Greenwood
11	98103	Freemont
12	98103	Greenlake
13	98104	International District
14	98104	Pioneer Square
15	98105	University District
16	98105	Laurelhurst
17	98107	Ballard

The data with the location coordinates is used to generate the map of Seattle, on which the communities are shown with the marks.

	Post_Code	Neighborhood_Name	Latitude	Longitude
0	98003	Federal Way	47.642242	-122.399689
1	98005	Bellevue	47.642242	-122.399689
2	98033	Kirkland	47.642242	-122.399689
3	98037	Lynnwood	47.642242	-122.399689
4	98040	Mercer Island	47.642242	-122.399689
5	98052	Redmond	47.642242	-122.399689
6	98055	Renton	47.642242	-122.399689
7	98101	Seattle	47.642242	-122.399689
8	98101	Downtown	47.642242	-122.399689
9	98102	Capital Hill	47.642242	-122.399689
10	98103	Greenwood	47.642242	-122.399689
11	98103	Freemont	47.642242	-122.399689
12	98103	Greenlake	47.642242	-122.399689
13	98104	International District	47.642242	-122.399689
14	98104	Pioneer Square	47.642242	-122.399689
15	98105	University District	47.642242	-122.399689
16	98105	Laurelhurst	47.642242	-122.399689
17	98107	Ballard	47.642242	-122.399689
18	98109	South	47.642242	-122.399689



With the using of Foursquare the list of public places in different categories found in the given city can be shown.

	name	categories	lat	lng
0	Il Corvo	Italian Restaurant	47.602522	-122.331952
1	Columbia Tower Club	Social Club	47.604507	-122.330484
2	Biscuit B*tch	Breakfast Spot	47.603237	-122.332010
3	Juicy Cafe	Café	47.604329	-122.330958
4	Tat's Delicatessen	Sandwich Place	47.601901	-122.332423
5	Smith Tower	Building	47.601858	-122.332152
6	Top Pot Doughnuts	Donut Shop	47.604023	-122.332499
7	Metropolitan Grill	Steakhouse	47.604617	-122.334280
8	Columbia Center Observation Deck (Sky View Obs...	Scenic Lookout	47.604595	-122.330816
9	Smith Tower Observation Deck	Scenic Lookout	47.601877	-122.331866
10	Monorail Espresso	Coffee Shop	47.604651	-122.331094
11	Elm Coffee Roasters	Coffee Shop	47.600152	-122.330944
12	Flatstick Pub	Beer Bar	47.600220	-122.331310
13	Delicatus	Deli / Bodega	47.601579	-122.334253
14	Red Bowls	Noodle House	47.604654	-122.333211
15	Cherry Street Public House	Coffee Shop	47.600450	-122.332908
16	Bill Speidel's Underground Tour	Tour Provider	47.602124	-122.334000
17	Damn the Weather	Cocktail Bar	47.601004	-122.334119
18	Cherry Street Coffee House	Coffee Shop	47.602767	-122.334151

With the machine learning algorithm, the most common venues are gathered and the 7 first most common venues are shown for every community in the district.

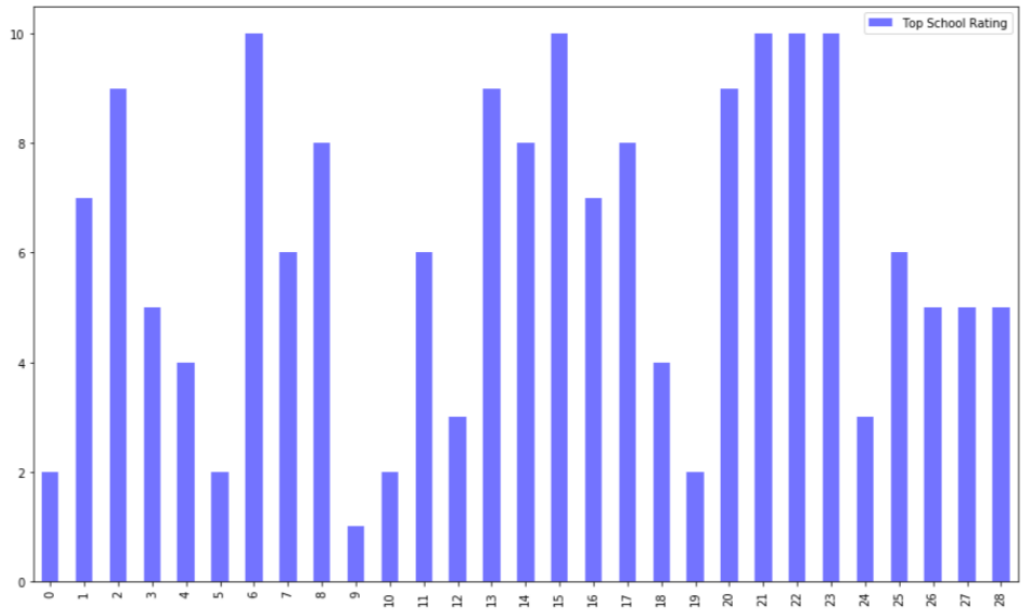
```
----Laurelhurst----
                venue  freq
0         Coffee Shop  0.09
1          Playground  0.09
2         Pizza Place  0.06
3            Pharmacy  0.06
4                Park  0.06
5                 ATM  0.03
6  Mexican Restaurant  0.03

----Lynnwood----
                venue  freq
0         Coffee Shop  0.09
1          Playground  0.09
2         Pizza Place  0.06
3            Pharmacy  0.06
4                Park  0.06
5                 ATM  0.03
6  Mexican Restaurant  0.03

----Madrona----
                venue  freq
0         Coffee Shop  0.09
1          Playground  0.09
2         Pizza Place  0.06
3            Pharmacy  0.06
4                Park  0.06
5                 ATM  0.03
6  Mexican Restaurant  0.03
```

```
----Belltown----
                venue  freq
0         Coffee Shop  0.09
1          Playground  0.09
2         Pizza Place  0.06
3            Pharmacy  0.06
4                Park  0.06
5                 ATM  0.03
6  Mexican Restaurant  0.03

----Capital Hill----
                venue  freq
0         Coffee Shop  0.09
1          Playground  0.09
2         Pizza Place  0.06
3            Pharmacy  0.06
4                Park  0.06
5                 ATM  0.03
6  Mexican Restaurant  0.03

----Columbia City----
                venue  freq
0         Coffee Shop  0.09
1          Playground  0.09
2         Pizza Place  0.06
3            Pharmacy  0.06
4                Park  0.06
5                 ATM  0.03
6  Mexican Restaurant  0.03
```

	Neighborhood	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
0	Alki Beach	Playground	Coffee Shop	Pizza Place	Pharmacy	Park	Video Store	Italian Restaurant
1	Bainbridge Island	Playground	Coffee Shop	Pizza Place	Pharmacy	Park	Video Store	Italian Restaurant
2	Ballard	Playground	Coffee Shop	Pizza Place	Pharmacy	Park	Video Store	Italian Restaurant
3	Bellevue	Playground	Coffee Shop	Pizza Place	Pharmacy	Park	Video Store	Italian Restaurant
4	Belltown	Playground	Coffee Shop	Pizza Place	Pharmacy	Park	Video Store	Italian Restaurant

For comparation of communities the school ratings were used. The communities were organized in the descending order of school ratings to give the client the opportunity to find the community with the best school.

	Neighborhood	Top School Rating
23	Alki Beach	10
15	University District	10
21	Madrona	10
6	Renton	10
22	West Seattle	10
20	Bainbridge Island	9
2	Kirkland	9
13	International District	9
14	Pioneer Square	8
8	Downtown	8
17	Ballard	8
1	Bellevue	7
16	Laurelhurst	7
7	Seattle	6
11	Freemont	6
25	Belltown	6
26	Northgate	5
27	Mount Baker	5



The second used parameter to sort the communities was the price of the property. The communities were organized in the descending order of the house pricing.

	Neighborhood_Name	Price_of_housing
8	Downtown	2356900.0
17	Ballard	453000.0
2	Kirkland	256000.0
4	Mercer Island	253695.0
5	Redmond	253640.0
28	Magnolia	240000.0
3	Lynnwood	235940.0
1	Bellevue	235478.0
6	Renton	200000.0
0	Federal Way	152956.0
25	Belltown	135000.0
27	Mount Baker	132000.0
24	Columbia City	125000.0
16	Laurelhurst	125000.0
15	University District	120653.0
11	Freemont	120358.0
9	Capital Hill	120354.0
21	Madrona	120333.0
18	South	120000.0

Then client can choose the five preferred communities and to compare them more detailed as shown.

Neighborhood_Name	Lynnwood	West Seattle	Pioneer Square	Madrona	Redmond
Post_Code	98037	98116	98104	98110	98052
Latitude	47.6422	47.6422	47.6422	47.6422	47.6422
Longitude	-122.4	-122.4	-122.4	-122.4	-122.4
Cluster Labels	0	0	0	0	0
1st Most Common Venue	Playground	Playground	Playground	Playground	Playground
2nd Most Common Venue	Coffee Shop	Coffee Shop	Coffee Shop	Coffee Shop	Coffee Shop
3rd Most Common Venue	Pizza Place	Pizza Place	Pizza Place	Pizza Place	Pizza Place
4th Most Common Venue	Pharmacy	Pharmacy	Pharmacy	Pharmacy	Pharmacy
5th Most Common Venue	Park	Park	Park	Park	Park
6th Most Common Venue	Video Store	Video Store	Video Store	Video Store	Video Store
7th Most Common Venue	Italian Restaurant	Italian Restaurant	Italian Restaurant	Italian Restaurant	Italian Restaurant

Conclusion

The decision about the buying of the property is regularly made with response of the different factors such as the property price, the amount of public places nearby and so on.
The machine learning algorithms can help the real estate agents to search for the property and communities, that will be the best choice for the particular client, because different parameters can be used to analyze the needs and requirements. Moreover the information obtained from the machine learning analyze can be used by clients to decide, what of the parameters have the most impact on the price.