

# **Deciding to work with IT abroad**

Kelly Azevedo Borges Leal Neves

June 1<sup>st</sup>, 2020

## **1. Introduction**

### **i. Background**

The decision to immigrate to another country is a very tough one. However, it's common to see many people, especially from developing countries, that decide to do so. They usually follow this path in order to pursue a better quality of life and a better future for their family. People often go to countries that have good standards in terms of health, education and safety.

Most of the time, when somebody decides to immigrate, this means a big move which involves not only city or workplace changes but a whole new experience in terms of culture, food, language, friendship and much more.

Working as a Software Engineer or a Data Scientist may help when deciding to work abroad, because many countries lack professionals in these areas and therefore accept foreign people to fill these roles.

### **ii. Problem**

When people start studying the possibility to move abroad they want to gather information regarding job offers, city climate, city safety, school rating, average income, things to do there, rent values, and so on. In this study, I selected some of these variables to make an analysis comparing some cities from Canada and the USA (more specifically Vancouver, Toronto, New York, Seattle, Austin and San Francisco). The main goal is to help an IT worker - who is willing to move abroad - to select one of these cities.

### **iii. Interest**

People that work in the IT field and are considering to immigrate will be interested in reading this material as they might spend a lot of time planning ahead and doing research about all the topics involved.

## 2. Data acquisition and cleaning

### i. Data sources

Despite the fact that somebody planning to immigrate needs to gather a lot of different information as listed in (1.ii.), I decided to analyze the following topics: city temperature (highs and lows) throughout the year, number of job openings and their average annual income and the most common venues for the main neighborhoods for each city.

The cities that were chosen for analysis are: Vancouver, Toronto, New York, San Francisco, Austin and Seattle. All of them are cities famous for being home of many high tech companies.

The data sources used for this study are described as follow:

a) City temperature: web scraping on Wikipedia

1. <https://en.wikipedia.org/wiki/Toronto>
2. <https://en.wikipedia.org/wiki/Vancouver>
3. [https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)
4. [https://en.wikipedia.org/wiki/San\\_Francisco](https://en.wikipedia.org/wiki/San_Francisco)
5. <https://en.wikipedia.org/wiki/Seattle>
6. <https://en.wikipedia.org/wiki/Austin>

b) Number of job openings and their average annual income: web scraping on ziprecruiter

1. <https://www.ziprecruiter.com/Jobs/Data-Scientist/-in-Vancouver,BC>
2. <https://www.ziprecruiter.com/Jobs/Software-Engineer/-in-Vancouver,BC>
3. <https://www.ziprecruiter.com/Jobs/Data-Scientist/-in-Toronto,ON>
4. <https://www.ziprecruiter.com/Jobs/Software-Engineer/-in-Toronto,ON>
5. <https://www.ziprecruiter.com/Jobs/Data-Scientist/-in-San-Francisco,CA>
6. <https://www.ziprecruiter.com/Jobs/Software-Engineer/-in-San-Francisco,CA>
7. <https://www.ziprecruiter.com/Jobs/Data-Scientist/-in-New-York,NY>
8. <https://www.ziprecruiter.com/Jobs/Software-Engineer/-in-New-York,NY>
9. <https://www.ziprecruiter.com/Jobs/Data-Scientist/-in-Seattle,WA>
10. <https://www.ziprecruiter.com/Jobs/Software-Engineer/-in-Seattle,WA>
11. <https://www.ziprecruiter.com/Jobs/Data-Scientist/-in-Austin,TX>
12. <https://www.ziprecruiter.com/Jobs/Software-Engineer/-in-Austin,TX>

c) Some neighborhoods for each city: some neighborhoods of each city were chosen manually, as it was hard to get this information automatically for all cities. Hence there was not found a webpage where I could get this information from the same pattern, it was not a good choice to use web scraping for 6 totally different pages. However, for Toronto and New York data, I could use previous examples of the

Applied Data Science Capstone labs and the data were extracted automatically.

- d) The most common venues for each neighborhood: Foursquare.

## ii. Data cleaning

Since the data sources listed above are HTML pages, it was necessary to do a hard work in order to get the useful information. However, once I found out how to scrap the temperature data (from Wikipedia) and the job information (from ziprecruiter), it was easy to replicate for the cities chosen. After getting the data from the URLs above, I created some Dataframes to store them:

- a) A dataframe with the average temperature per month per city

	January	February	March	April	May	June	July	August	September	October	November	December
City												
Toronto	-4.0	-3.0	1.5	8.0	14.0	19.5	22.5	21.5	17.0	10.5	5.0	-0.5
Vancouver	4.0	5.0	6.5	9.5	13.0	16.0	18.0	18.0	15.0	10.5	6.5	3.5
San Francisco	11.0	12.0	13.0	13.0	14.5	15.5	15.5	16.5	17.0	16.5	13.5	11.0
New York	0.5	2.0	6.0	11.5	17.5	22.0	25.0	24.5	20.0	14.0	8.5	3.5
Seattle	4.5	6.0	7.5	10.0	13.0	16.0	17.0	18.0	15.0	11.5	8.0	5.5
Austin	10.5	12.5	16.5	21.0	24.5	27.5	29.0	30.0	27.0	22.0	16.5	11.5

Table 1 – AVERAGE TEMPERATURE PER MONTH PER CITY

- b) Two dataframes with the number of job openings and their annual income for each city, one for software engineer and other for data scientist roles.

	City	Income	Quantity
0	Toronto	85000.0	2078
1	Toronto	105000.0	1730
2	Toronto	120000.0	1337
3	Toronto	130000.0	963
4	Toronto	145000.0	414
5	Vancouver	90000.0	1013
6	Vancouver	110000.0	828
7	Vancouver	120000.0	684
8	Vancouver	135000.0	396
9	Vancouver	145000.0	223
10	San Francisco	110000.0	11232
11	San Francisco	130000.0	9467
12	San Francisco	145000.0	7351
13	San Francisco	160000.0	4786
14	San Francisco	175000.0	2458
15	New York	95000.0	12236
16	New York	115000.0	10219
17	New York	130000.0	7859
18	New York	145000.0	4926
19	New York	155000.0	2985
20	Austin	85000.0	3915
21	Austin	105000.0	3245
22	Austin	120000.0	2404
23	Austin	130000.0	1669
24	Austin	140000.0	975
25	Seattle	105000.0	7624
26	Seattle	125000.0	6329
27	Seattle	140000.0	4580
28	Seattle	145000.0	3723
29	Seattle	160000.0	1733

Table 2 - SOFTWARE ENGINEER JOB OPPENINGS

	City	Income	Quantity
0	Toronto	80000.0	1275
1	Toronto	105000.0	1012
2	Toronto	120000.0	775
3	Toronto	130000.0	570
4	Toronto	145000.0	280
5	Vancouver	90000.0	447
6	Vancouver	110000.0	369
7	Vancouver	120000.0	288
8	Vancouver	130000.0	219
9	Vancouver	145000.0	104
10	San Francisco	95000.0	9055
11	San Francisco	120000.0	7376
12	San Francisco	140000.0	5541
13	San Francisco	160000.0	3620
14	San Francisco	175000.0	2064
15	New York	80000.0	11831
16	New York	105000.0	9408
17	New York	120000.0	7668
18	New York	140000.0	4935
19	New York	155000.0	2846
20	Austin	75000.0	1876
21	Austin	100000.0	1534
22	Austin	115000.0	1174
23	Austin	130000.0	786
24	Austin	145000.0	372
25	Seattle	95000.0	4530
26	Seattle	120000.0	3692
27	Seattle	135000.0	2865
28	Seattle	150000.0	1829
29	Seattle	160000.0	1086

Table 3 – DATA SCIENTIST JOB OPPENINGS

- c) A dataframe with the city chosen neighborhoods and their location (latitude and longitude) in order to get their nearby venues (per neighborhood using Foursquare) and compare the data found in these six cities. The dataframe is shown below.

	City	Neighborhood	Latitude	Longitude
0	Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	Toronto	Garden District, Ryerson	43.657162	-79.378937
3	Toronto	St. James Town	43.651494	-79.375418
4	Toronto	The Beaches	43.676357	-79.293031
...	...	...	...	...
123	Austin	Downtown	30.268054	-97.744764
124	Austin	Hancock	30.295896	-97.724768
125	Austin	Heritage	30.345796	-97.690991
126	Austin	West Downtown	30.268054	-97.744764
127	Austin	Rosedale	30.313390	-97.744898

Table 4 – CITIES CHOSEN NEIGHBORHOODS WITH THEIR LOCATION

- d) After creating this last dataframe (Table 4), I used Foursquare to get nearby venues for each neighborhood in order to compare the chosen neighborhoods for each city and see which town offers the best experience for those who are considering to immigrate and establish a life there.

### 3. Methodology

The methodology chosen for this study is to make a comparative analysis between the cities of Austin, New York, Seattle, Vancouver, Toronto and San Francisco in terms of annual average temperature, the number of job openings and their average annual income for Software Engineer and Data Scientist roles.

I also selected some neighborhoods from each city and searched for their most common nearby venues in order to compare them, using the k-means clustering algorithm.

### 4. Results

The annual temperature is the first type of comparison between the six cities. From Wikipedia, using web scraping, it was possible to create the charts below, which represent the low and high averages temperatures per month for each city.

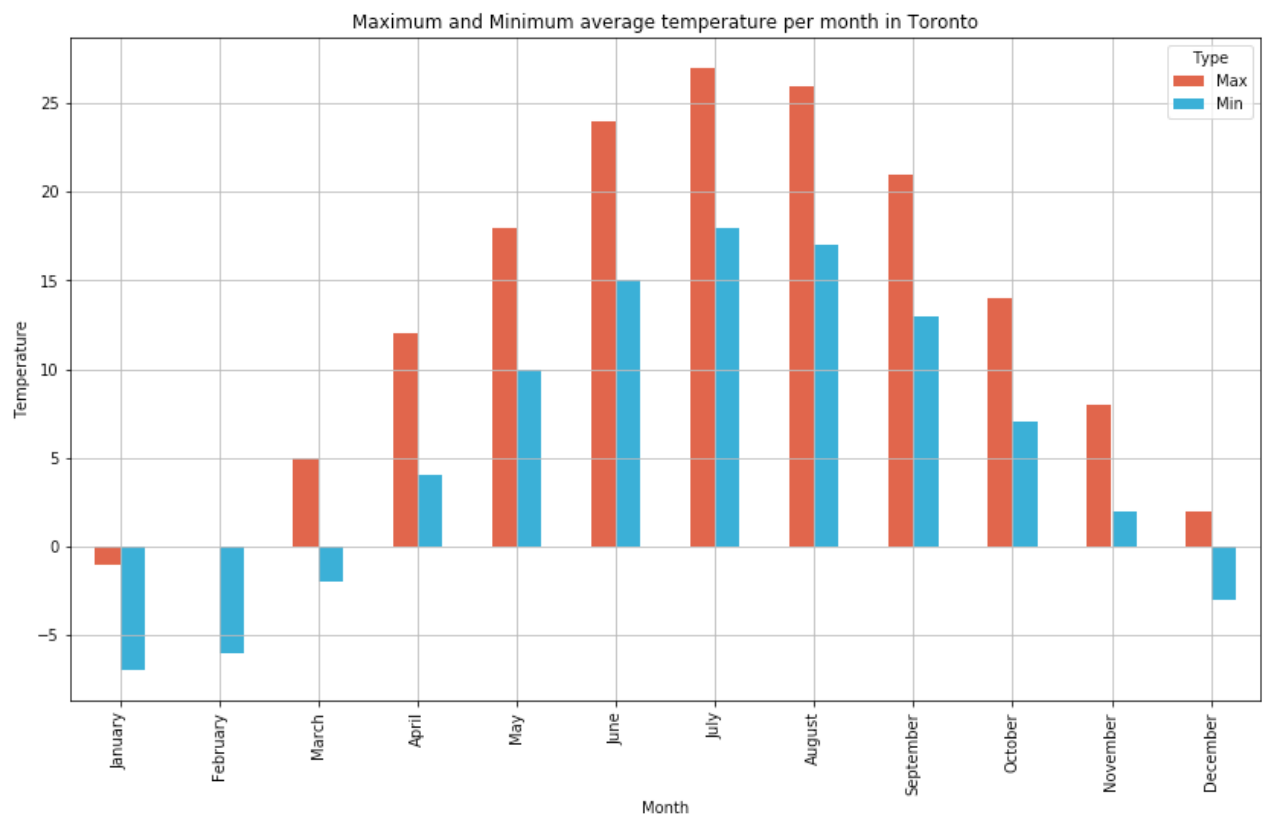


Figure 1 – MAXIMUM AND MINIMUM AVERAGE TEMPERATURE PER MONTH IN TORONTO

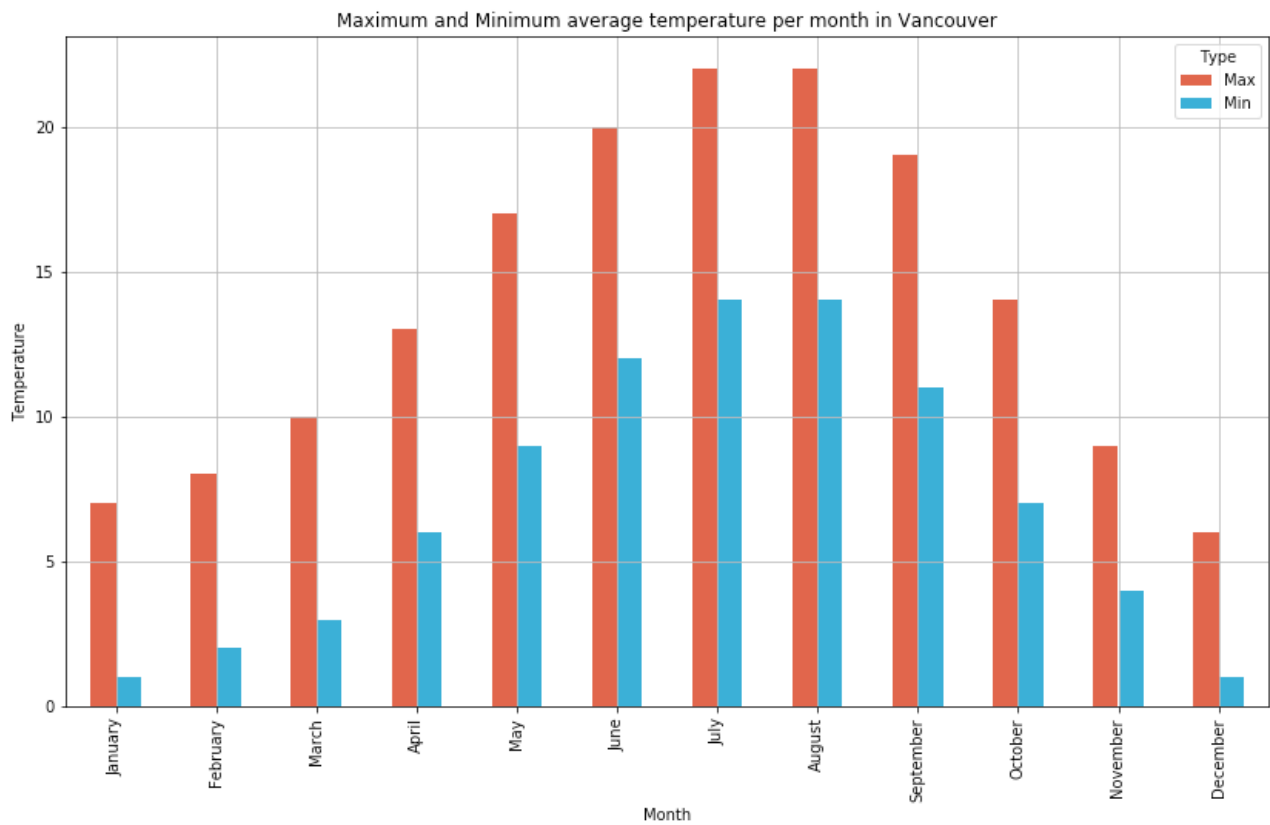


Figure 2 – MAXIMUM AND MINIMUM AVERAGE TEMPERATURE PER MONTH IN VANCOUVER

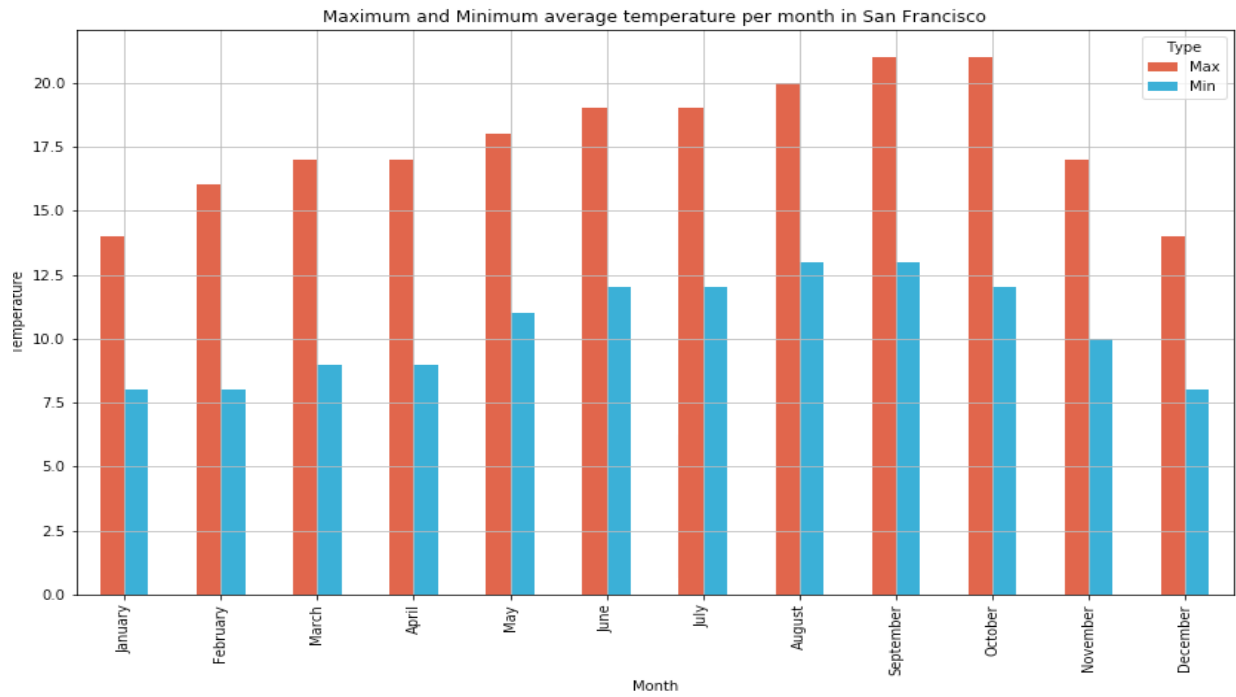


Figure 3 – MAXIMUM AND MINIMUM AVERAGE TEMPERATURE PER MONTH IN SAN FRANCISCO

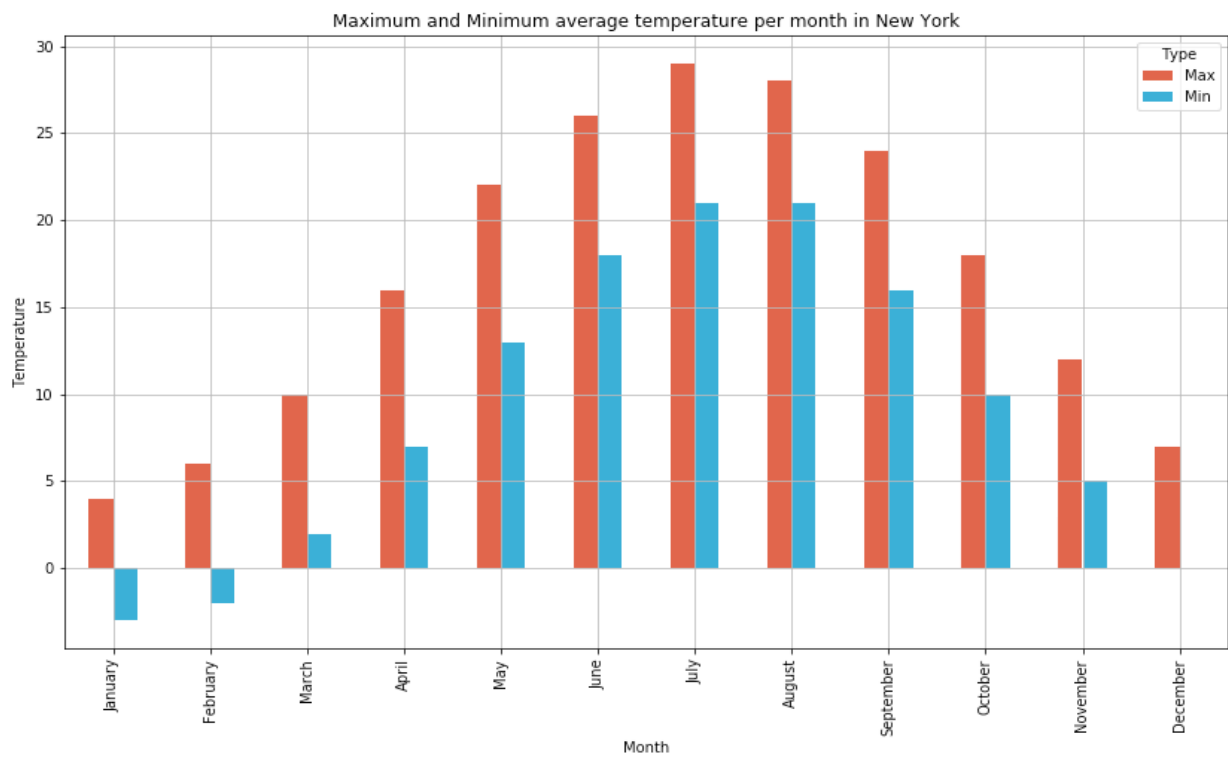


Figure 4 – MAXIMUM AND MINIMUM AVERAGE TEMPERATURE PER MONTH IN NEW YORK



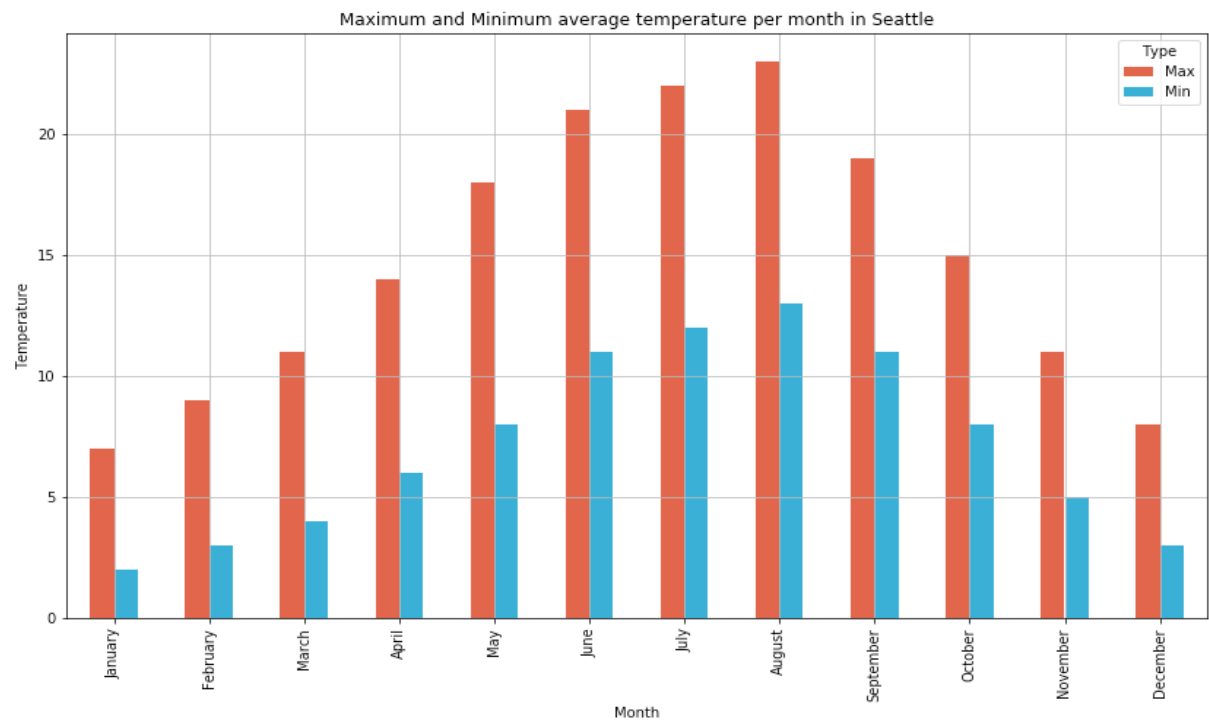


Figure 5 – MAXIMUM AND MINIMUM AVERAGE TEMPERATURE PER MONTH IN SEATTLE

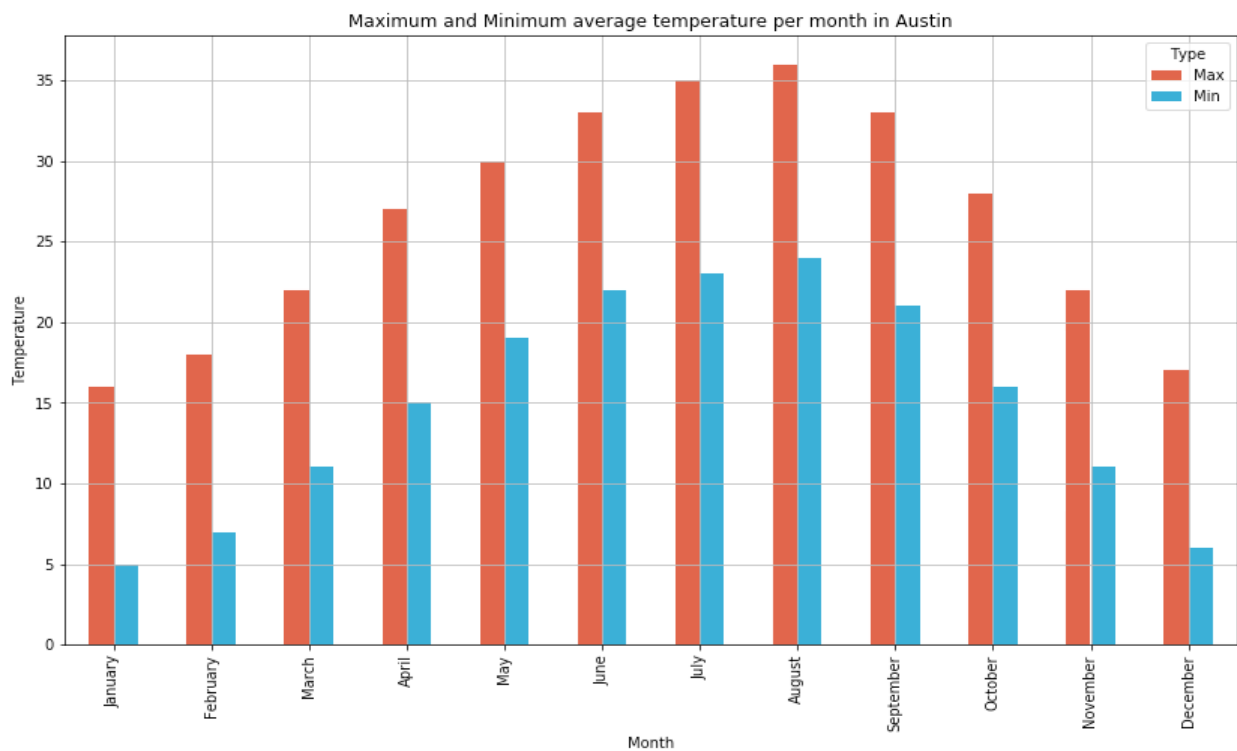


Figure 6 – MAXIMUM AND MINIMUM AVERAGE TEMPERATURE PER MONTH IN AUSTIN

With the data from these six charts, it's possible to get good information about the averages temperatures in each city throughout the year. It helps when choosing a city to live, as a person who doesn't like cold weather, for example, would not be very willing to move to Toronto. On the other hand, a worker who doesn't feel comfortable in places that have warm weather would think twice about living in Austin.

The next chart compares the annual average temperature between all the cities and gives more clear information about the differences between them. It's possible to see that Austin is the warmest city throughout the year and Toronto is the coldest one. It's also possible to infer that in San Francisco the temperature doesn't vary a lot in a year.

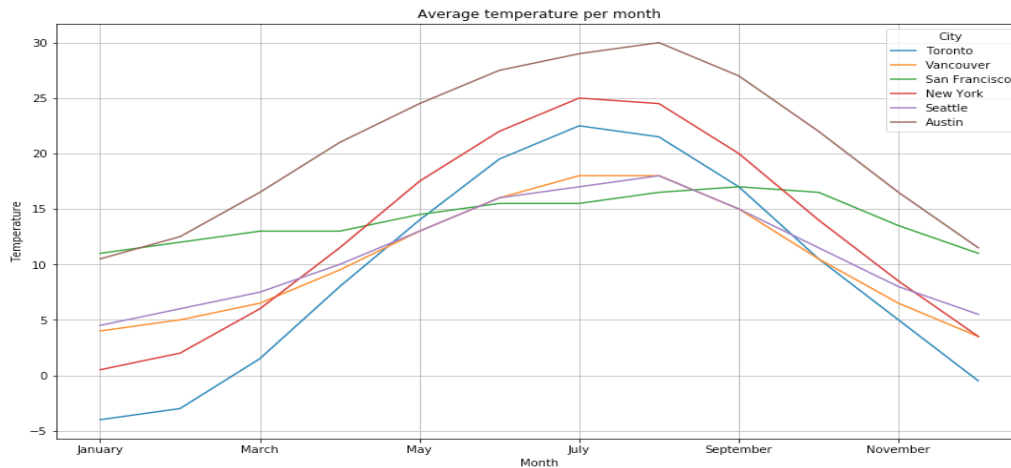


Figure 9 – COMPARISON BETWEEN THE AVERAGE TEMPERATURE PER MONTH IN EACH CITY

The next source of comparison between the cities is the job offers shown in ZipRecruiter. With the data extracted from this website, it was possible to create the two charts below.

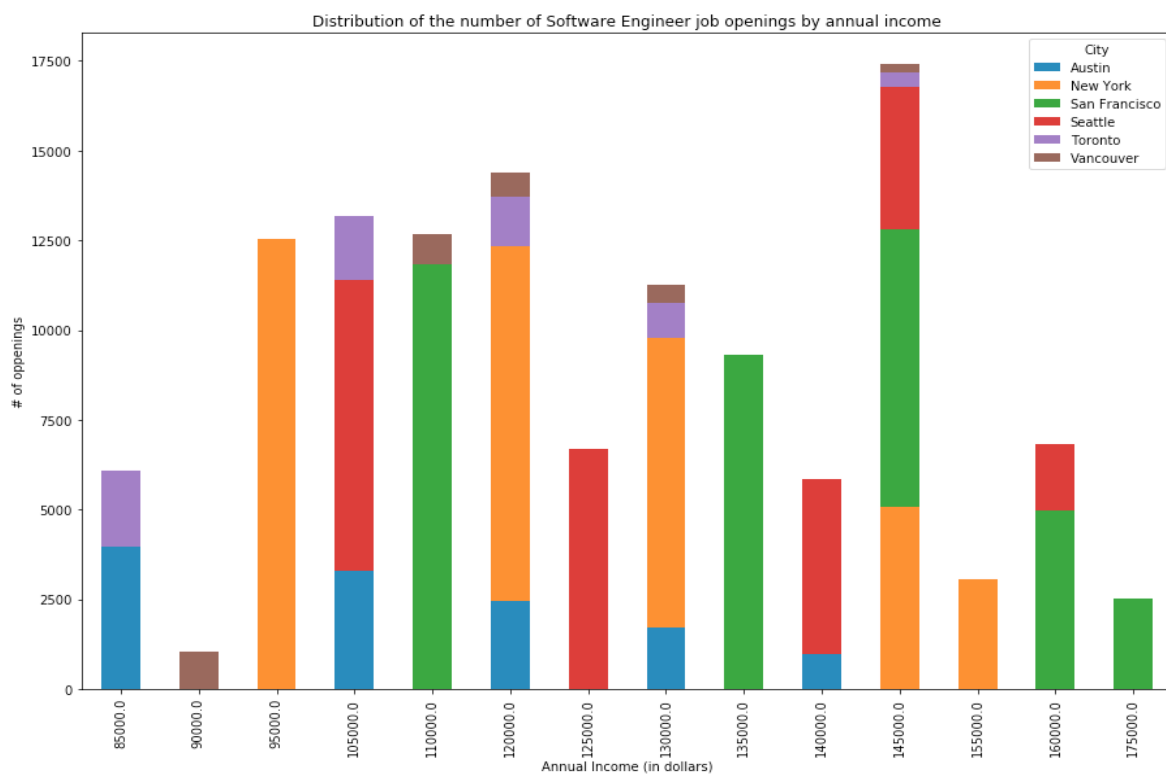


Figure 10 – SOFTWARE ENIGEER JOB OPENINGS

It is really clear that New York and San Francisco concentrate the greatest part of the job openings for Software Engineer roles. Seattle comes just behind them. Austin, Toronto and Vancouver are cities that also have many available positions, but their quantity doesn't compare to NYC and San Francisco numbers.

Another information that can be concluded analyzing the graphic is that San Francisco, New York and Seattle usually pay higher incomes.

The number of job openings and the value of average annual income in a city for a determinate position are good resources for helping to decide to move in. Another information is also very important, for example, the average apartment rental and other monthly expenses, as it is closely connected to the job payings. But in this study, this type of information was not analyzed.

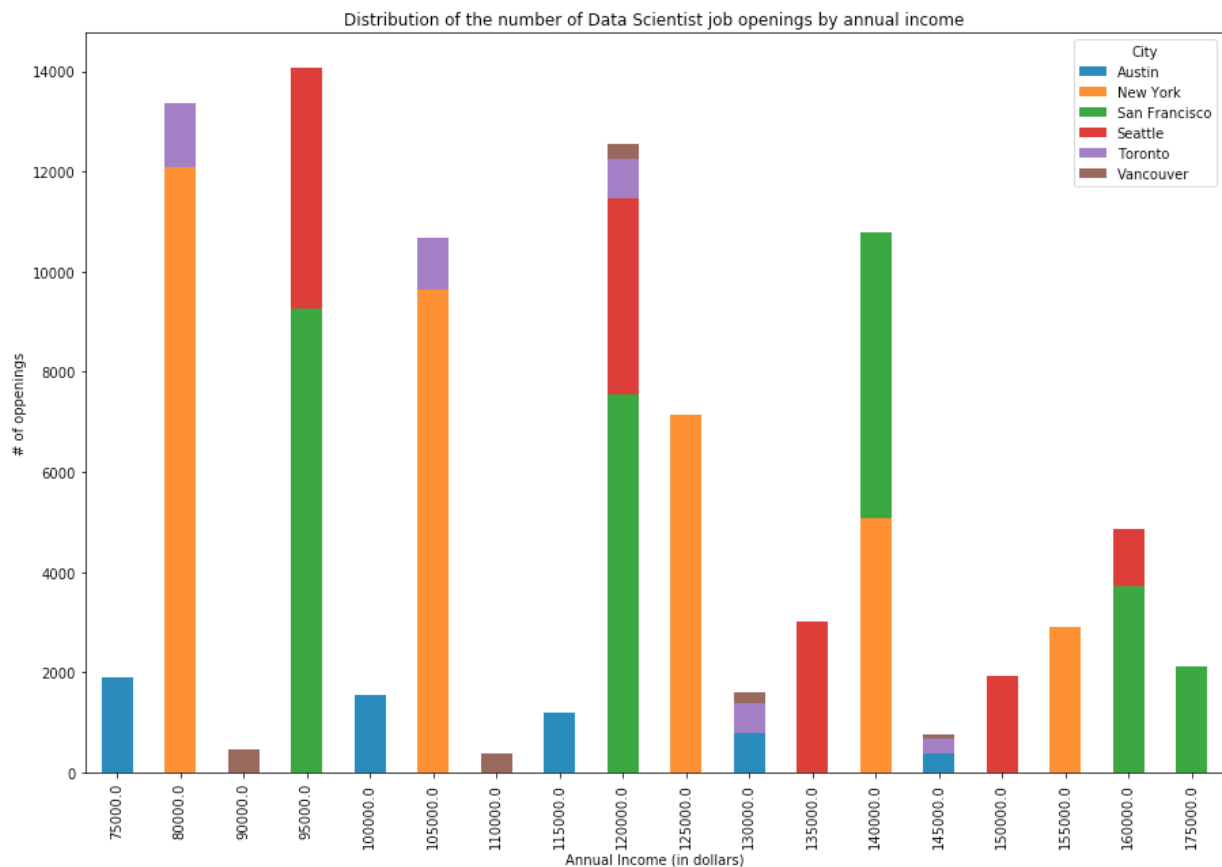


Figure 11 – DATA SCIENTIST JOB OPENINGS

In this chart, again, it's clearly expressed that New York, San Francisco and Seattle have the greatest number of job openings for Data Scientists. They also pay the largest annual incomes. It was already expected that the two charts would present similar data in terms of the number of job openings and annual income since Software Engineer and Data Scientist are both roles that are quite similar and have people moving from one to another.

The next few maps show the neighborhoods classified by the k-means algorithm and the comparison method was their most common nearby venues. The algorithm used 5 classes, which are represented by the colors red, orange, green, purple and blue. As we can see in the maps, the neighborhoods are very similar regarding their venues, as all of the maps have many purple circles. That means that it doesn't matter the city somebody chooses to live, it will have very similar venues. Naturally, what can change is the number of venues, as in bigger cities people have many more options of entertainment.

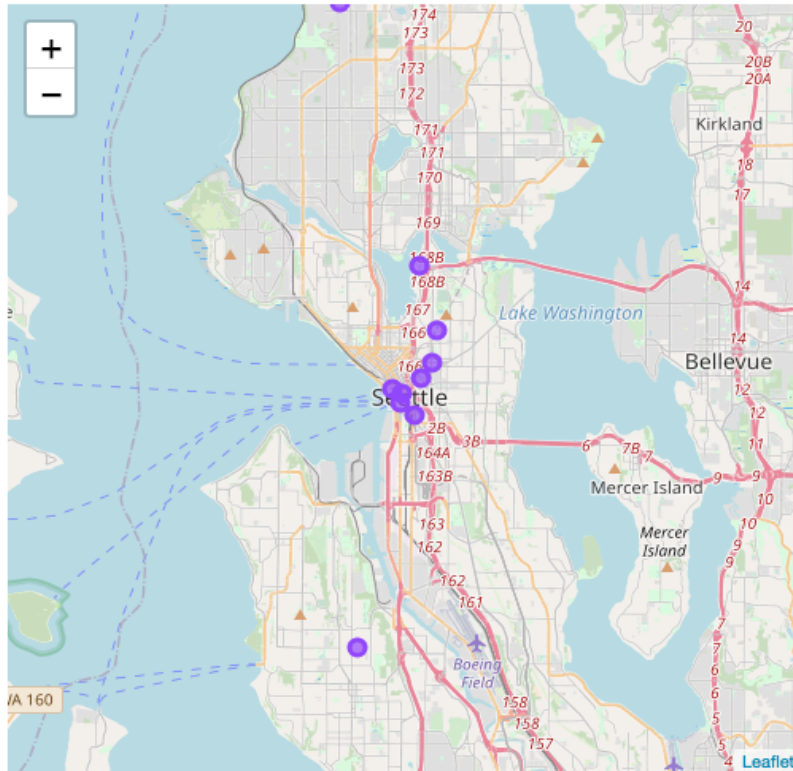


Figure 12 – MAP OF SEATTLE NEIGHBORHOODS CLASSIFIED USING K-MEANS CLUSTERING

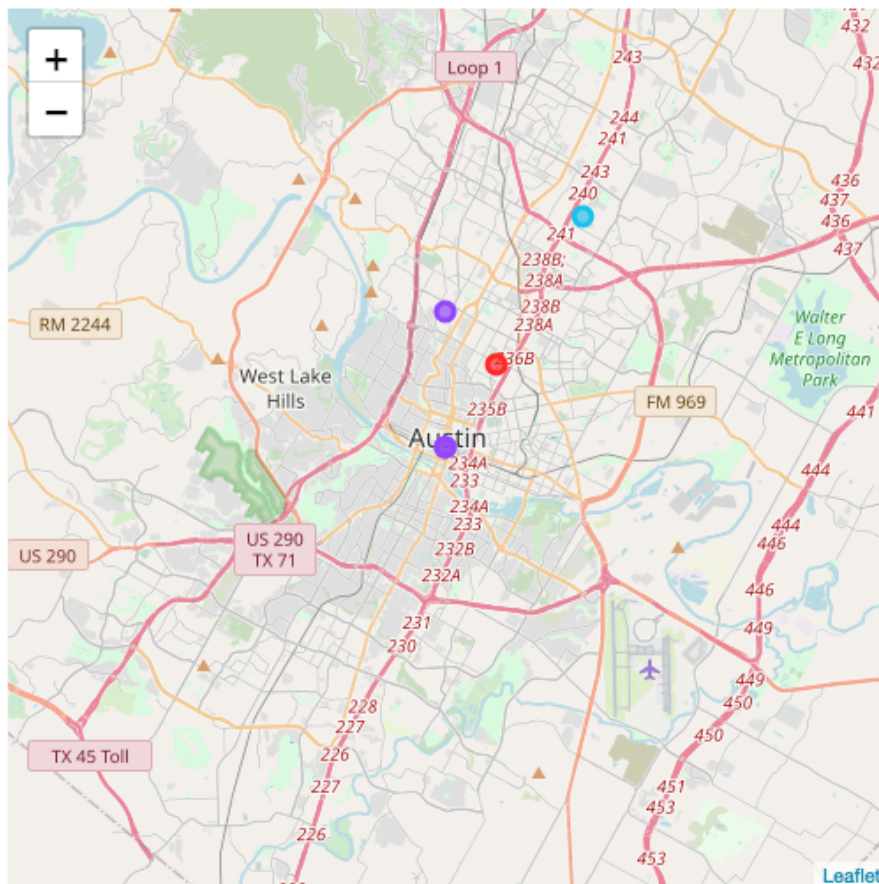


Figure 13 – MAP OF AUSTIN NEIGHBORHOODS CLASSIFIED USING K-MEANS CLUSTERING

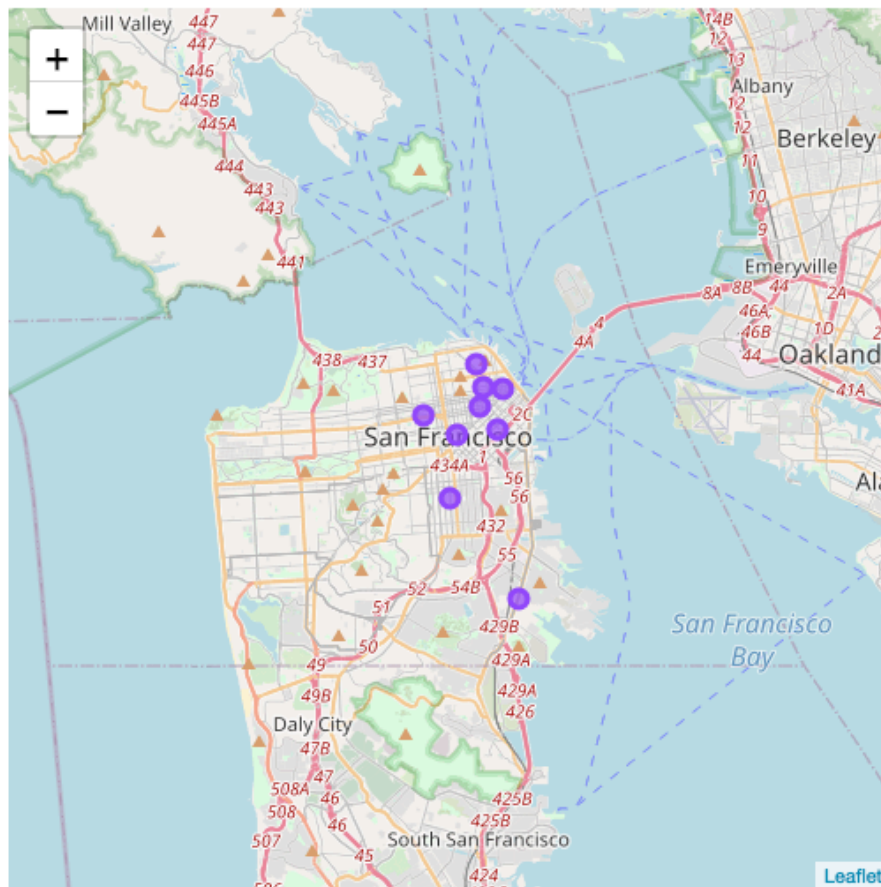


Figure 14 – MAP OF SAN FRANCISCO NEIGHBORHOODS CLASSIFIED USING K-MEANS CLUSTERING

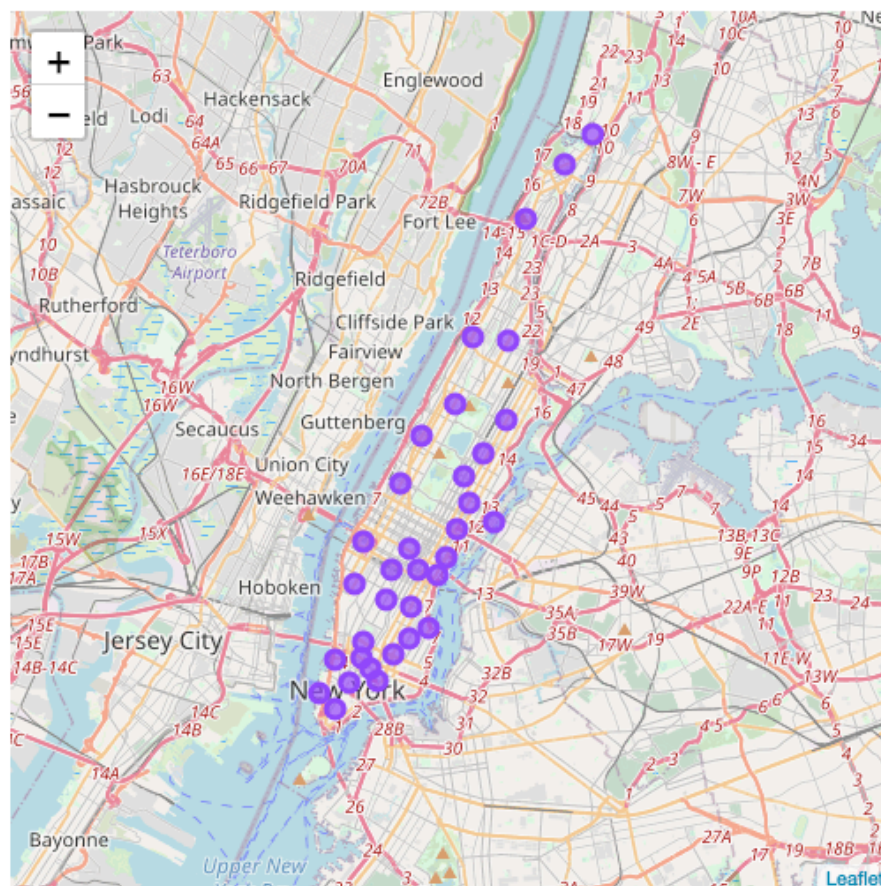


Figure 15 – MAP OF NEW YORK NEIGHBORHOODS CLASSIFIED USING K-MEANS CLUSTERING



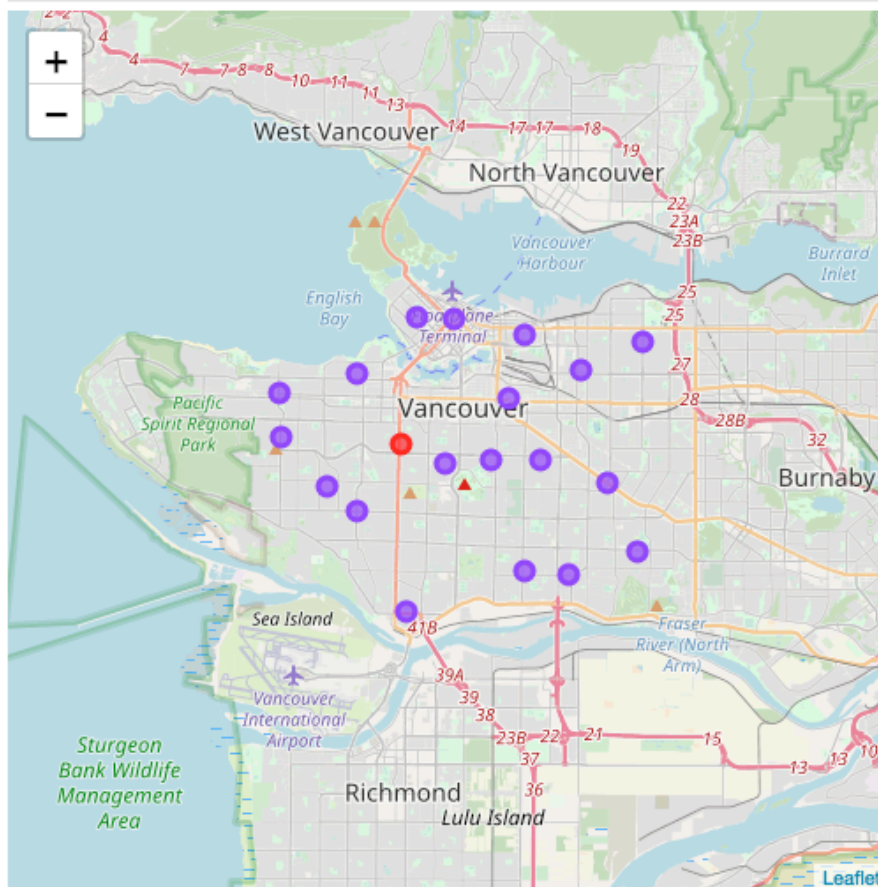


Figure 16 – MAP OF VANCOUVER NEIGHBORHOODS CLASSIFIED USING K-MEANS CLUSTERING

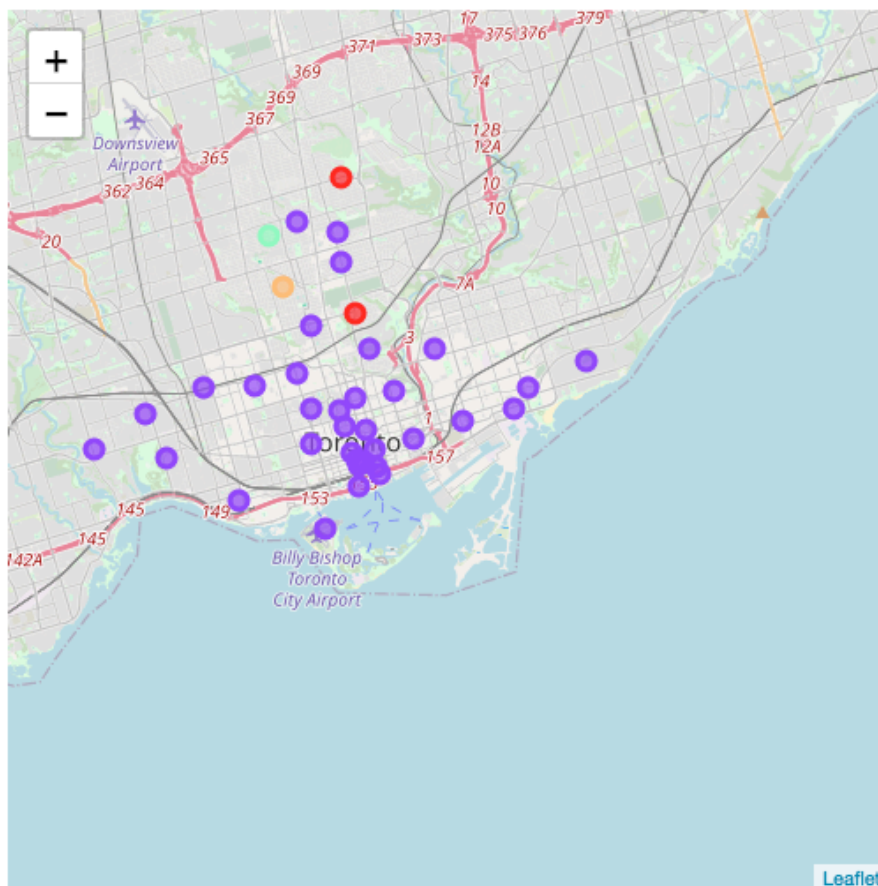


Figure 17 – MAP OF TORONTO NEIGHBORHOODS CLASSIFIED USING K-MEANS CLUSTERING

## **5. Discussion**

As I decided to compare cities in two very big countries, which are far away from each other, it was possible to see a big difference in terms of annual temperature between them. For sure this is a piece of very important information for someone who is studying the possibility to live abroad. A person who likes warm wheater would be inclined to choose Austin between these six options. In the other hand, somebody who would like to live in a very cold place, where it gets to snow very often, would prefer to choose Seattle, Toronto, Vancouver or New York.

Moreover, another source of information would be to look at the number of job openings and their average annual income. If somebody only cares about getting the highest payings, he/she would definitely choose New York, San Francisco or Seattle. However, in these cities, the cost of living may also be bigger and maybe it's not worth it to get paid more, as the monthly expenses can also be higher. So, in this case, it could be helpful to look at other variables, as rental costs, for example. But the chart built regarding this information of number of job openings and annual income can be very supportive when negotiating a salary with an employer.

Additionally, in terms of venues available in each city, the maps above demonstrate that the six cities are very similar. That was expected as they are in the same continent and from only two different countries, which are very linked in terms of culture and people's behavior.

## **6. Conclusion and future work**

In this study, the goal of helping an IT person who is willing to work abroad (in the six cities included) as a Software Engineer or a Data Scientist was achieved, as one can get a lot of relevant information about each one of these cities. Needless to say that when thinking of moving abroad, as it is a very big change, many others variables should be analyzed.

Therefore, to enhance this source of information, it would be good to make more research about the topic, for example: include other cities and more neighborhoods for each city, get rental information, get school data for the worker's children, and city rank about quality of life. So, giving weights for each type of information, related to what somebody thinks its more relevant, it would be even possible to create a machine-learning algorithm to help the decision making.