

Deciding to work with IT abroad

Kelly Azevedo Borges Leal Neves

June 1st, 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 times, 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, schools 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 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 of the fact that somebody planning to immigrate needs to gather a lot of different information as listed in (1.ii.), I decided to analyse 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
4. 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) The main neighborhoods for each city: the main neighborhoods for 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 oppenings and their anual 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