IBM / COURSERA

CAPSTONE PROJECT
THE BATTLE OF NEIGHBORHOODS

Paulo Sousa

SUMMARY

- Introduction
- Data
- Methodology
- Results
- Observations
- Conclusion

INTRODUCTION

- **CONTEXTUALIZATION**
- This case is based on a **real case**: One year ago, we decided to move to Canada. That time we needed to decide *what* city should be the better choice between two options: **Toronto** and **Ottawa**. Of course, several factors can be considered, including objective or subjetive ones. Science cannot deal with the last kind, but if we consider numerical values under the first aspect, maybe it is possible to have a diagnosis about the better choice.
- In this work, I am using **DATA SCIENCE**, in order to colaborate with the decision that, in fact, we have made before. This tool can reinforce that when we consider personal aspects (most related with family and quality of life), our final decision was in fact correct.

- **BACKGROUND**
- This problem is a very common situation when we consider the doubt related with the best destiny for immigrant families that need to consider different aspects. The first problem is related with the *factors* that must be considered in order to get the result to the original problem.
- In this situation, getting *trusted* and *updated* information is crucial. The second problem is that you need to have similar data from both options: having a detailed data of only one city is not fair and sometimes it is a challenge.

INTRODUCTION

- **BUSINESS PROBLEM**
- This problem is closer to situation when you need to decide which city can be chosen for a particular purpose. In my specific (and realistic) scenarium, the purpose is for a *family immigration*. However, it can be widely used for different situations (maybe with minor personal impact), for instance, where is the best place to start a new small business company.

- **INTERESTS**
- Decision making is a very common situation that ML (Machine Learning) or simply data analysis can be used. In fact, the main purpose is having a better view in order to take a decison. It also important to say (and it is shown in the example) that is not only the *amount* of data, but also its **quality** (for instance, how updated it is) and the **relevance**, in order to avoid to make ponctuations without any particular criterion.

DATA

- *Source 1: FourSquare*
- As mentioned, similar data must be obtained. In fact, using the information from the laboratories in this specialization, I have got details about one city (Toronto). It should be relatively easy to have similar information about the other destiny (Ottawa).
- In fact, only changing the information about the city, the libraries and services from **Foursquare** (https://api.foursquare.com/) were enough to get the complementary information about the Capital City, considering the main Venues obtained from the free service. It was the *first* criterion used in this comparison exercise.

- *Source 2: Versus*
- The second criterion was chosen by the information recorded at **Versus** (https://versus.com/en/ottawa-vs-toronto), when we have several indicators in order to compare both cities. However, in order to have a fair situation, both aspects: advantages from both cities were considered. In this situation, it was necessary to combine two tables in a single one, with permutation of some elements, in order to get a whole picture.

DATA

- *Data detail 1: FourSquare*
- After request the service for Venues (https://api.foursquare.com/v2/venues/explore), it is peformed a filter in order to use only the following information:
  - Venue name;
  - Venue category;
  - Venue location (latitude / longitude).
- The category was specially important because it is used to mark some venues as important ("*top*") or not. It is also shown in the map of both cities.

- *Data detail 2: Versus*
- From this page, I used all the 16 initial indicators: / Population density;
- Monthly public transport ticket; / Unemployment rate;
- Cost of one-bedroom ap. in the center; / Younger population;
- Income inequalities; / Landmarks from UNESCO World Heritage list;
- Average minimum temperature; / Higher average maximum temperature;
- Higher average temperature; / Million inhabitants;
- Billion higher gross domestic product (GDP); / Higher average salary;
- 2 more big sport facilities; / Cost of one single transportation ticket;
- Billionaires from annual Forbes.

METHODOLOGY

- The methodology can be resumed in the following steps:
  - General preparation: When libraries, importation and functions are defined;
  - Data gathering and analysis: Using **Foursquare** and **Versus** information;
  - Conclusion: Combining the results in order to decide the best city.

- Some additional notes (detailed in the Solution description below):
- It is important to see that the information had to be prepared and combined from the Versus data;
- Method to extract numerical information from Versus was also implemented, in order to adjust the output of BeautifulSoup.

RESULTS

- Toronto Venues

- Ottawa Venues

RESULTS

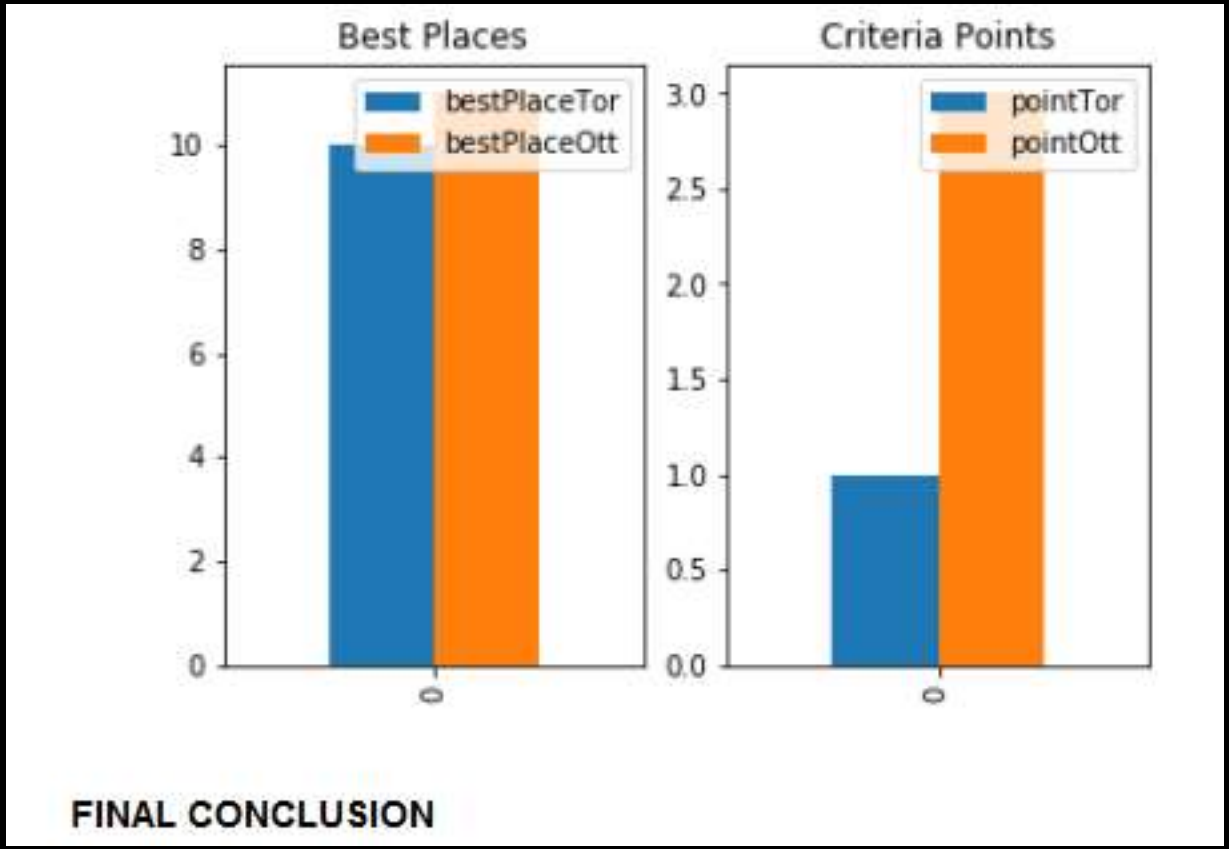
- Versus

	City	PopDensity	MonthlyPubTrans	UnemploymentRate	CostOneBedroom	YoungPop	IncomeIneq	Heritage	AvgMinTemp
0	Ottawa	1680 people/km²	81.26$	6.3%	902.6$	36.7 years	0.440	1	1.5°C
1	Toronto	4150 people/km²	105.01$	9.9%	1199.1$	36.9 years	0.441	0	6.2°C

AvgMaxTemp	HighAvgTemp	MilInhabitants	BilDGP	HighAvgSal	SportsFacility	CostSingTrans	ForbesBil
10.9°C	6.2°C	0.88 million	40$ billion	2441.68$	1	2.62$	0
11.9°C	9.1°C	2.62 million	270$ billion	2574.92$	3	2.41$	5

RESULTS

- CONCLUSION



OBSERVATIONS

- **BASIC OBSERVATIONS**
- The first aspect is that data need to be understood and appropriately used. For instance, we need to mark whici indicators were more appropriated to the current proof of concepts.

- **RECOMENDATIONS**
- Evaluation under different criterium, for instance, changing the categories of venues considered priority, are also relevant and it depends from each case.

CONCLUSION

- **2.2.6. Conclusion section where you conclude the report.**
- This study was based in a real fact: the decision where to go, considering particular aspects.
- Maybe in the future, newcomers can use this example, with their own data, in order to help to take a decision like this.

Thank you

- Paulo Sousa