

|  |
| --- |
| Coursera Capstone Report  IBM Data Scientist Professional |
|  |
| April 27, 2021  Authored by: Reza Sadeghi Jafari |



# Coursera Capstone Report

## IBM Data Scientist professional

This Report developed based on Coursera capstone project which represent the business problem, data gathering, data analysis, ML solutions, the result, discussion and the conclusion.

Criteria

This capstone project will be graded by your peers. This capstone project is worth 70% of your total grade. The project will be completed over the course of 2 weeks. Week 1 submissions will be worth 30% whereas week 2 submissions will be worth 40% of your total grade.

For this week, you will required to submit the following:

A description of the problem and a discussion of the background. (15 marks)

A description of the data and how it will be used to solve the problem. (15 marks)

For the second week, the final deliverables of the project will be:

A link to your Notebook on your Github repository, showing your code. (15 marks)

2. A full report consisting of all of the following components (15 marks):

Introduction where you discuss the business problem and who would be interested in this project.

Data where you describe the data that will be used to solve the problem and the source of the data.

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

Results section where you discuss the results.

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

Conclusion section where you conclude the report.

3. Your choice of a presentation or blogpost. (10 marks)

Business Problem:

Every city needs to be controlled by a team, to be ensure all the essential needs of the city's residents are well available in all parts of the city. Concentrating facilities in a part of the city will cause complex problems for urban management such as unwanted traffic on morning and evening, caricatural growing up of city and their facilities, there is some unhappy citizens and so much. Therefore, in municipalities, a sector is generally considered as urban development, and this team needs to control the distribution of facilities in different urban sectors.

In this project, the team play the role, as a customer, has requested that the necessary facilities be explored as follows throughout the city of **Toronto**, and that different parts of the city be segmented accordingly.

1- Medical centers

2- Police stations

3- Fire stations

So, Sectors with fewer facilities, that will need more development, will be a priority in next year's development plans.

**Data gathering:**

To achieve this goal, we need data from different urban areas as follows

1- Urban areas

2- Facilities available in each area

3- the population on each area

To provide this information, we fetch urban areas from the information provided on the web - Wikipedia - as well as urban venues information from the foursquare.com based on the geographical location of the areas. At last, we need the population of Toronto on each area, after some goggle search I found it at https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&SR=1&S=22&O=A&RPP=9999&PR=0

**Data Analysis:**

In data analysis, dirty information needs to be cleaned first. Then, the geographical information of urban areas will be added to it, next step is, we will add the information of the venues which specified in the program for each area. Finally we need to add

Now we can examine the distribution of facilities in each part of the city and prioritize the facilities that need further development based on each of the needs.