**Introduction**

**Business Problem Description**

**Data used to solve the problem**

|  |  |
| --- | --- |
|  |  |
|  |  | The battle of the neighborhoods is a great opportunity to conduct exercises comparing neighborhoods anywhere in the world, |
|  |  | to know their multiculturalism, their length, size of habitants, distances, routes, most important avenues. |
|  |  | For that reason I am very interested in being able to make the comparison between Toronto and New York. |
|  |  | Both cities have millions of people living in them. |
|  |  | A capstone tool of geolocation analytics is very useful at the time of attracting investment from the government. |
|  |  | For the city of new york we already have the data set location available. In the city of Bogota a wikipedia page will be used |
|  |  | where it has important information of the main locations of the city, |
|  |  | additionally it would use the Foursquare API to be able to capture all the data and interpret them later. |

### **Data preparation and preprocessing**

**At this stage, we prepare our dataset for the modeling process, opting for the most suitable machine learning algorithm for our scope. Accordingly, we perform the following steps:**

**Rename the column names Format the date column Sort data by date of sale Select data only for the city of New York and Toronto . Make a list of street names in New York. Calculate the street-wise average price of the property Read the street-wise coordinates into a data frame, eliminating recurring word New York from individual names Join the data to find the coordinates of locations which fit into client's budget Plot recommended locations on London map along with current market prices**

**Methodology**

**The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises four stages:**

1. **Collect Inspection Data**
2. **Explore and Understand Data**
3. **Data preparation and preprocessing**
4. **Modeling**

**We can now proceed to the Modeling phase. We will analyze neighborhoods to recommend real estates where home buyers can make a real estate investment. We will then recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.**

### 

### **Modeling**

**After exploring the dataset and gaining insights into it, we are ready to use the clustering methodology to analyze real estates. We will use the k-means clustering technique as it is fast and efficient in terms of computational cost, is highly flexible to account for mutations in real estate market in London and is accurate.**

**After our inspection of venues/facilities/amenities nearby the most profitable real estate investments in London, we could begin by clustering properties by venues/facilities/amenities nearby.**

**Discussion**

* **To have a better results definitely more relevant information should be incorporated to this assessment. Some of the information that would be great to have is related to finance matter. Amount these additional data we have: predicted costs of the competition, estimated revenue from the sale of tickets and hospitality packages, estimated revenue from the sale of media and marketing rights, IT and location of the International Broadcasting Centre. Probably the results could change after incorporating these financial variables that are important on the decision making.**
* **The scope of this work was simple and since I was only looking to apply the methods learned during the course I did not incorporated the data mention before, but to have a thorough analysis with accurate results we should do it.**