Launch delivery Startup Company

Capstone Project - The Battle of Neighborhoods

Author: Dmitry V.

Date: June, 2019

Background and description of the problem

- Nowadays we see developing and expansion of new technologies for delivery services. Online stores experiments with copters equipped with grippers to deliver items ordered by customers. It could potentially give big benefits for customers and companies.
- Following this trend, a new Startup Company has been established to piloting new delivery technologies for lunch delivery to city offices.
- The idea is to find best place to start business in Canada. For that it is obvious we should consider major cities: Toronto, Montreal and find best place there.
- **The assumption** is that we will make best profit in places where there are a lot of offices and their density is most high
- We have chosen major Canada cities: Toronto, Montreal. In order to get best start we would like to have an answer to questions below:
- **Problem/Question 1:**
- Which selected city is better in terms of potential revenue, meaning having greater number of customers?
- **Problem/Question 2:**
- After question 1 is answered, what would be the best area in chosen city to locate delivery parking facility?
- Data science project has launched to address both questions and get presentation and report to support business decisions.

Stakeholders

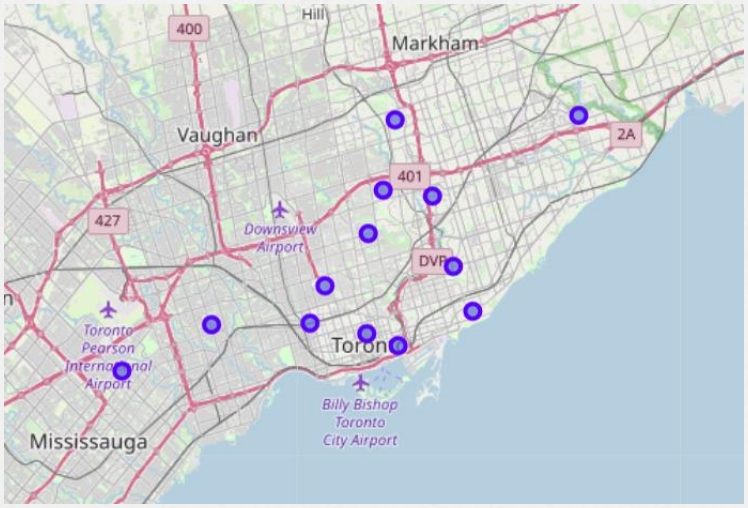
- Below are the stakeholders and their benefits:
- ***New Startup Company.***
  - To succeed pilot project, get maximum revenue to further extend new technologies of delivery to conquer market and increase revenues
- ***Restaurants.***
  - Fast delivery can attract more customers and help with competition
- ***Customers.***
  - Having launch delivery faster will make customers happier

Data Description

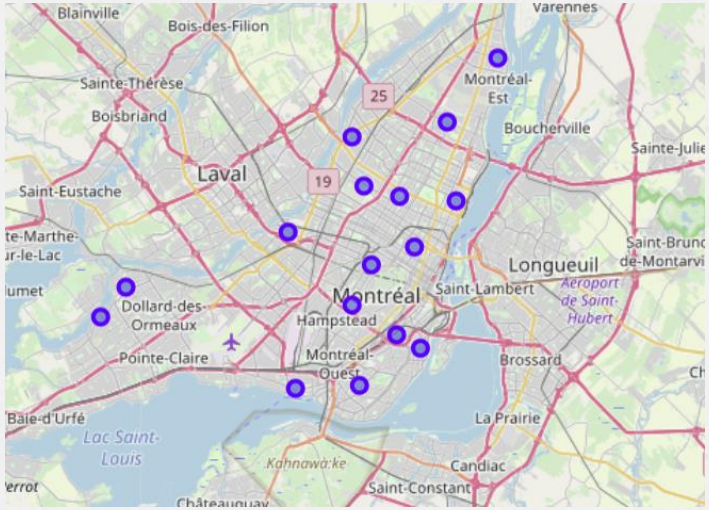
- **Problem 1:**
- ——————————
- Data collection utilized boroughs data from Wikipedia, geo location data from medium.com, Geopy/Nominatim. The Foursquare API was used to gather potential customer data in both cities.
- Foursquare data can be filtered to find potential customers using category ID's such as Offices, Business Centers etc. The extracted data can then be imported as .json files into data frames ready for preparation.
- Data extracted included features such as city/ borough location, business type, name along with their longitude and latitude co-ordinates. Borough postal codes were also scraped from Wikipedia. Collected data was then prepared and cleaned before analysis using data frames and mapping techniques.
- **Problem 2:**
- ——————————
- Potential customers within the selected city were segmented and clustered. The results overlaid on a map of the city. The data maps and clustering results established the ideal borough within the city for the clients business expansion

Analysis. Problem 1

- For each city selected we gathered from Wikipedia the boroughs of each city location, importing them as a .csv file before creating a data frame. Longitude and Latitude co-ordinates for each borough were also imported into a data frame. Both data frames were then merged together. Geopy library was utilized to get the latitude and longitude values of each city which was then used along with the merged Borough data frame to create maps with the boroughs superimposed on top.
- Toronto



Montreal



Analysis. Problem 1

- The Foursquare API was then used to create a .json file of potential customers within a radius of 750 meters of the center of each city. The resulting output was achieved by using a specific category ID within the foursquare URL, in this case ‘Offices’ the foursquare category included subcategories of ‘ Co-working Space, Tech Startups, Advertising Agency, Campaign Office, Corporate Amenity’. The results were then imported into its own data frame.
- Following data analysis of the two location choices of Montreal and Quebec, there was a clear winner. Montreal having a larger percentage of potential customer numbers available to the client

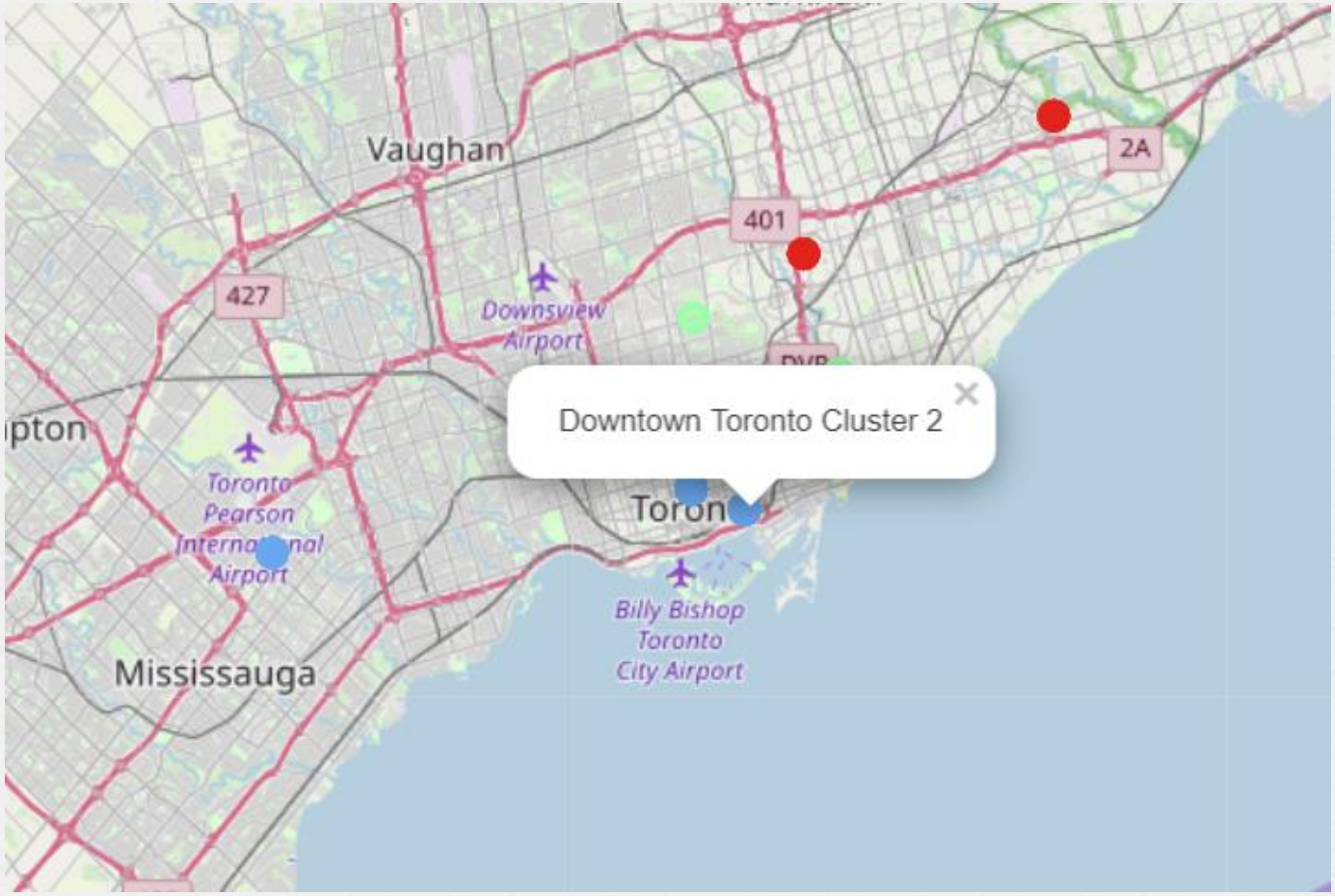
```
print('{} Nearby offices in Montreal.'.format(nearby_offices_montreal.shape[0]))
print('{} Nearby offices in Toronto.'.format(nearby_offices_toronto.shape[0]))
```

```
13 Nearby offices in Montreal.
97 Nearby offices in Toronto.
```

Analysis. Problem 2

- Further analysis of the boroughs of Toronto revealed a cluster containing more potential offices/customers than the others. **Cluster 3** contained the greatest concentration of offices. Customers in this cluster are in Offices/ Tech Startup Offices.
- **Cluster 3** consists of three boroughs which are Downtown Toronto, Queen's Park and Mississauga. We are looking at the one borough location, so we showed each cluster onto a map of Toronto. (**Cluster 3** boroughs are in blue on the map) From this data, the city borough which would offer the greatest place for the business are Downtown Toronto and Queen's Park. he third option while having a large concentration of potential customers is isolated and would reduce opportunities for future business growth. Between first two options Downtown Toronto would be preferable having more potential offices/customers.
- Our winner is Downtown Toronto.

Analysis. Problem 2



Results

- After data collection, cleaning and analysis we can now answer the initial data science problem posed in the introduction of this project.

- **Problem/Question 1:** Which out of two possible locations (Toronto and Montreal) has the greater number of potential offices/customers?
  - **Answer**: Toronto as it offers the greater number of potential offices/customers for starting the business.

- **Problem/Question 2:** In the city chosen in answer to Problem/Question 1, in this case Toronto, which city borough would offer the best place for the business?
  - **Answer**: The city borough of Montreal which would offer the best place for starting the business is Downtown Toronto.

Conclusion

- Data science project implemented providing all necessary data, analysis and reasoning to make decision on where to better start business and locate launch delivery facilities.
- It should be noted that there are assumptions and limitations (described in Report) which should be taken into account for further investigation and data analysis refinement