Coursera

IBM Data Science Capstone Project

Bike Theft

Background

The Coursera IBM Data Science Professional Certification course consists of 9 online courses that covers topics including open-source tools and libraries, Python, databases, SQL, data visualization, data analysis, statistical analysis, predictive modeling, and machine learning algorithms. The course finishes with a Capstone project.

The purpose of this Capstone project is to demonstrate the use of the data science toolsets, methodologies, and skills that have been acquired during this course to help solve a business problem.

Problem

In this project I am a hypothetical bike GPS location device and services vendor looking to introduce the new GPS location device that is hidden within the seat tube and a subscription-based service that monitors the location of the bike in real time.

Although bike theft has always been a common issue, especially in urban areas, it has increased during the pandemic in a number of cities. The pandemic led to an unprecedented boom in bikes sales. The rising demand, increase in ridership, and shortage of bikes nationwide among other factors, has likely contributed to a rise in bike theft.

With the introduction of this new product, I will be using the theft data and known bike shop locations to target my approach to market with this new product.

Audience

This is hypothetical, but the target audience for the outcome of this project could be a real manufacturer of bike products that are promoted to reduce bike thefts.

This could be used to determine potential local businesses so that marketing campaigns can targeted to promote their products/services.

Data

For this project, the data that will be used:

- List of districts on Toronto
  - Data will be acquired through web scraping the Canadian Postal Code's Wikipedia page
    (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Geospatial coordinates of the neighborhoods and boroughs of Toronto
  - Data will be captured using the Geocoder package and stored into a csv file for easy consumption – the use of the Geocoder package is no longer free
- Bicycle theft data will be used to ascertain high theft areas using the coordinates of the neighborhoods and boroughs of Toronto
  - Data will be captured from the Toronto Police Services website and stored into a csv file for easy consumption
    (https://data.torontopolice.on.ca/datasets/TorontoPS::bicycle-thefts/about)
- Bike Shop venues within the neighborhoods and boroughs of Toronto
  - Data will be acquired through the use of the Foursquare API

Methodology

This project will compare suburbs and will determine similarities based on clustering techniques using location data services.

This project uses web scraping techniques to retrieve data from the Canadian Postal Code's Wikipedia page.

The data is then acquired and cleansed in preparation for clustering.

The geospatial locations data import will be merged with the post code data which will enable the data to be visualised over a map of the area.

The bicycle theft locations data import will also be merged with the post code data which will enable the data to be visualised over a map of the area. Due to the large volume of data, I have filtered for only the last 6 months of 2020 where the bike value was over $1000. The assumption is that that people with expensive bikes will be interested in this service.

The data will be clustered and plotted over the map.

The clustering is carried out by K Means and the clusters are plotted using the Folium Library.

The data will be mapped across Toronto and then focused/clustered in on boroughs.

Conclusion

The bike theft dataset:

event_unique_id	Report_Year	Report_Month	City	Hood_ID	NeighbourhoodName	Postal Code	Location_Type	Premises_Type	Cost_of_Bike	Status
GO-20201211930	2020	July	Toronto	76	Bay Street Corridor (76)	M1K	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1245	STOLEN
GO-20209016767	2020	July	Toronto	75	Church-Yonge Corridor (75)	M4L	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1808	STOLEN
GO-20201230882	2020	July	Toronto	76	Bay Street Corridor (76)	M1K	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1500	STOLEN
GO-20209017022	2020	July	Toronto	77	Waterfront Communities-The Island (77)	M5V	Bar / Restaurant	Commercial	1250	STOLEN
GO-20209017022	2020	July	Toronto	77	Waterfront Communities-The Island (77)	M5V	Bar / Restaurant	Commercial	1250	STOLEN
GO-20209016980	2020	July	Toronto	77	Waterfront Communities-The Island (77)	M5V	Other Commercial / Corporate Places (For Profit, Warehouse, Corp. Bldg	Commercial	1200	STOLEN
GO-20201279923	2020	July	Toronto	95	Annex (95)	M4G	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	2000	STOLEN
GO-20201272925	2020	July	Toronto	95	Annex (95)	M4G	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1400	STOLEN
GO-20201296593	2020	July	Toronto	51	Willowdale East (51)	M5X	Parking Lots (Apt., Commercial Or Non-Commercial)	Outside	1100	STOLEN
GO-20201256255	2020	July	Toronto	75	Church-Yonge Corridor (75)	M4L	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1200	STOLEN
GO-20209017155	2020	July	Toronto	95	Annex (95)	M4G	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1800	STOLEN
GO-20209017195	2020	July	Toronto	76	Bay Street Corridor (76)	M1K	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1400	STOLEN
GO-20209017195	2020	July	Toronto	76	Bay Street Corridor (76)	M1K	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1400	STOLEN
GO-20209017206	2020	July	Toronto	95	Annex (95)	M4G	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1200	STOLEN
GO-20201693165	2020	September	Toronto	136	West Hill (136)	M9R	Gas Station (Self, Full, Attached Convenience)	Commercial	2000	STOLEN
GO-20209017395	2020	July	Toronto	76	Bay Street Corridor (76)	M1K	Bar / Restaurant	Commercial	1500	STOLEN
GO-20209017593	2020	July	Toronto	85	South Parkdale (85)	M9R	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	2000	STOLEN
GO-20201310355	2020	July	Toronto	66	Danforth (66)	M1C	Other Commercial / Corporate Places (For Profit, Warehouse, Corp. Bldg	Commercial	3000	STOLEN
GO-20201322559	2020	July	Toronto	75	Church-Yonge Corridor (75)	M4L	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1650	STOLEN
GO-20209017611	2020	July	Toronto	81	Trinity-Bellwoods (81)	M1K	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	2500	STOLEN
GO-20209017856	2020	July	Toronto	82	Niagara (82)	M5T	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1100	STOLEN
GO-20201323363	2020	July	Toronto	53	Henry Farm (53)	M5P	Parking Lots (Apt., Commercial Or Non-Commercial)	Outside	3500	STOLEN
GO-20201361719	2020	July	Toronto	100	Yonge-Eglinton (100)	M9C	Parking Lots (Apt., Commercial Or Non-Commercial)	Outside	1100	STOLEN
GO-20209017953	2020	July	Toronto	100	Yonge-Eglinton (100)	M9C	Other Commercial / Corporate Places (For Profit, Warehouse, Corp. Bldg	Commercial	1300	STOLEN
GO-20201336667	2020	July	Toronto	31	Yorkdale-Glen Park (31)	M5V	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1842	STOLEN
GO-20209019259	2020	July	Toronto	76	Bay Street Corridor (76)	M1K	Streets, Roads, Highways (Bicycle Path, Private Road)	Outside	1600	STOLEN

This data was then processed and the following shows the clustering of these bike thefts across Toronto.
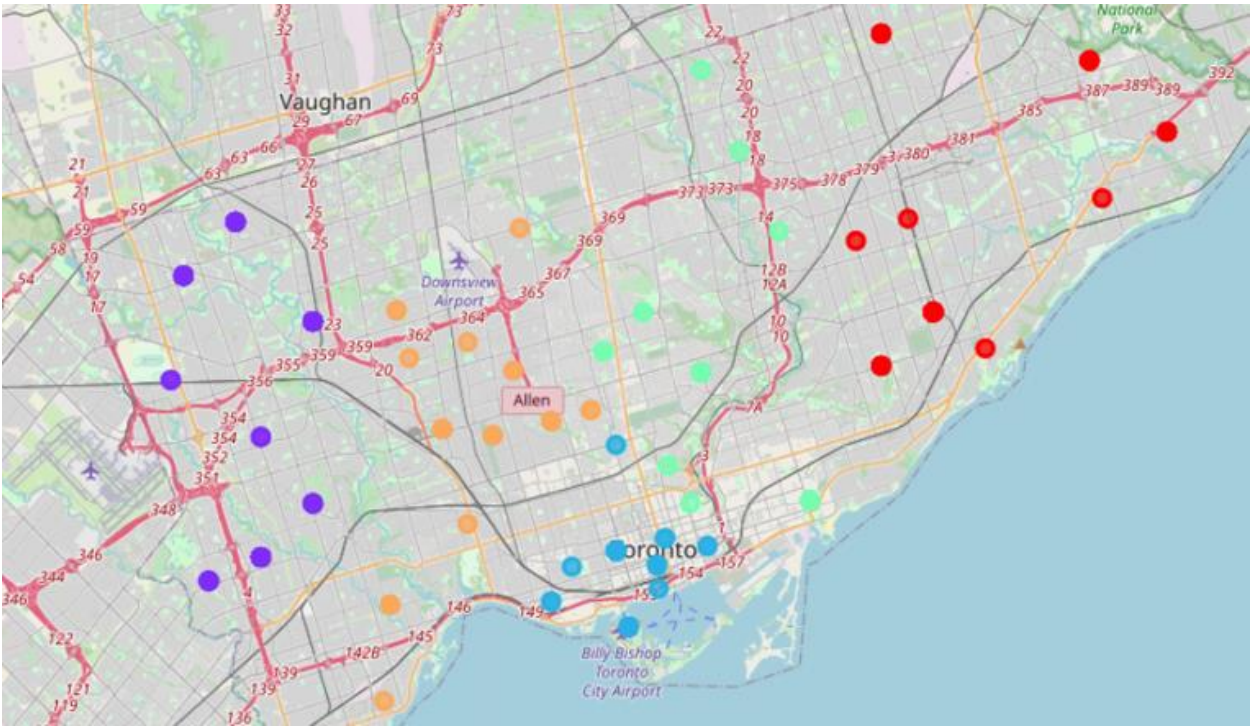
Using the Foursquare API, I wanted to be able to find the bike shops within the Toronto area, so that I could determine where to target my product/service. I was able to deduce that there only two bike shops within the Toronto area:

	PostalCode	Borough	Neighborhood	BoroughLatitude	BoroughLongitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
230	M3C	North York	Don Mills South	43.72590	-79.340923	Skiis and Biikes	43.726351	-79.342977	Bike Shop
436	M4G	East York	Leaside	43.70906	-79.363452	Enduro Sport	43.706059	-79.361835	Bike Shop
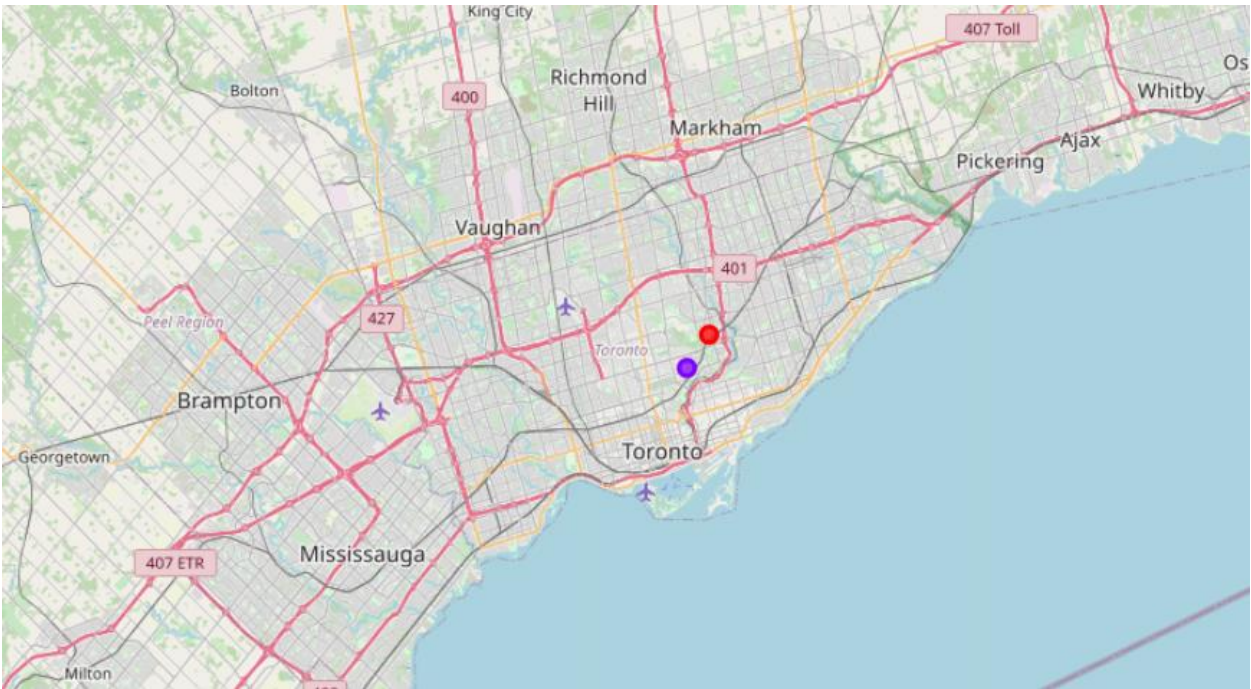
We can see that these shops sit within the Cluster 2 and Cluster 3 of the bike theft data, therefore both of these shops would be approached to start a conversation about stocking my product.