INTRODUCTION

In recent decades, the world has experienced rates of urban growth unparalleled in any other period of history and this growth is shaping the community environment in which an increasing proportion of us live. Initially, this project explores how urban growth differs in the two cities namely, Toronto and New York. Specifically, two boroughs, Manhattan in New York and Downtown Toronto have been chosen for this study. Further, it aims to track and compare the urban activities in the two local areas, respectively. This provides an opportunity to study urban activity growth patterns at an intra-city level. Motivated by these observations, this study aims to identify that data from location-based services be utilized to identify areas in cities where there is a surge in urban development, and further exploit these areas for business investments in the city as well as in the traffic regulation between other areas.

METHODS AND DATASETS: To solve the problem at hand, the information was extracted from data available online. Specifically, this study takes the advantage of geo-localized activities exposed on Foursquare. An extensive exploration of the data collected leads me to choose venue as the elementary unit of information. Therefore, I represented them by the most common venues in the rest of this work. Besides this, this study also gathered neighborhoods latitudinal and longitudinal data.

To limit the scope of our study, this study focusses on cities located in the United States (New York) and Canada (Toronto). They were chosen for their high activity according to sqstat.com. The other parts of the world were deliberately excluded as I am not familiar enough with them to correctly evaluate our results. At the time of writing, this dataset is available online.

The New York neighborhood has a total of 5 boroughs and 306 neighborhoods. To segment the neighborhoods and explore them, a dataset containing the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. On the other hand, the Canada neighborhood has a total of 5 boroughs and 210 neighborhoods.

Data Exploration: The datasets were explored for micro-analysis, in order to discover features that could characterize places and help cluster them into similar groups. By examining these similar groups with respect to two specific boroughs/areas in the two cities, using k-means clustering analysis this project aims to reveal interesting trends in urban development where the role of geography and culture remains important, even in an era where centralized city planning is dominant.

Foursquare is the most popular LBSN. Its two main purposes are: (i) enable users discover new places, and (ii) let their friends or the world know where they are. The first goal is achieved through venues. Venues have their own web page that displays basic facts (like name, address and type), but also user contributed information (such as photos, reviews, likes and tips) which can be used for rating or recommendation.

K-Means clustering: Since urban growth was supposed to be defined by clustering with population density and local spatial entropy, a handy tool for local area classification would be preferred. Although there exist many statistical packages, the commonly used K-Means clustering was chosen according to Vicker’s comparison on cluster analysis methods. K-Means clustering is capable to partition units by spatial correlation, aiming to make differences among the units in a group, overall groups, is minimized. Furthermore, the K-means algorithm is suitable for the situation that the number of clusters has already been designated. Besides, the K-means algorithm is distance-based, taking distance as the evaluation index of similarity, that is, the closer the distance between two objects is, the larger the similarity is. The algorithm considers the cluster to be composed of objects that are close together, so the compact and independent cluster is the ultimate target.

RESULTS

## **The dataset**

The basis of the analyses are datasets from Foursquare describing movements between places in the two cities. For each Foursquare venue in a city, the dataset contains

* — unique ID,
* — latitude and longitude,
* — creation time,
* — general Foursquare category,
* — specific Foursquare category, and
* — total number of unique visitors.

One of the criteria for a venue to have been included in the dataset is that it must have at least 5 total top venues.

**Activity growth profiles**

The food category (restaurants, coffee shop, etc.) is shown to contribute a significant proportion of growth, ranging from 19% to 33% in Manhattan. This is to be expected due to the entrepreneurial nature of food establishments. Accordingly, the average lifespan of food venues is relatively low, and so the category undergoes significant churn. Although many features of the growth profiles are shared between cities, such as the prominence of food venues, we can also observe differences in the dynamics of urban activity profiles of the cities. Downtown Toronto’s travel and transport categories display relatively high growth, where the airport services are the most visited venues with 64% frequency, suggesting greater investment in transport infrastructure.

The neighborhoods were clustered into 3 clusters each for Manhattan and Downtown Toronto Boroughs by k-means clustering method. Each cluster was examined and the discriminating venue categories that distinguish each cluster were determined. Based on the defining categories, a name was assigned to each cluster. In accordance with the growth profiles, the clusters confirmed that the food category contributed the most to urban growth in Manhattan, New York, while the airport services or the travel and transport category contributed to the surge in Downtown Toronto.

DISCUSSION

The present project demonstrated that data crowdsourced from location-based services can be used to identify cities and regions where particular urban activities are currently experiencing strong growth.

It has to be noted that the term growth somewhat loosely in this context to mean the number of venues in a particular neighborhood. A more satisfactory way to measure growth would be to look at the change in the number of venues (venues opening minus venues closing). However, the limitations of the dataset mean that there is no reliable way of detecting opening and closing venues.

To begin with, the study investigated the clustering of the cities’ neighborhoods according to the number of the most common venues, within the time period covered by the dataset. The main questions addressed were: Can we use crowdsourced data about places to profile cities in terms of urban activity growth? And moreover, in what way do these profiles of urban growth compare and what is the role of geography in this relationship?

While the number of different place types in a city’s neighborhood can shed light on historic and cultural aspects of urban growth, it can also be viewed as a snapshot of current investment in that neighborhood of the city. This can further highlight the priorities of local government and public spending (*colleges and universities*, *schools*, *government buildings,* etc.) as well as where growth in the private sector is currently focused (*restaurants*, *food*, *offices,* etc.).

Before the analyses, the possibility of heterogeneous Foursquare representation across different categories was noted. For example, it may be that the average Foursquare user is more interested in coffee shops than bowling alleys and, therefore, the former might be overrepresented in the data when compared with the latter. The extent of this potential bias could also be affected by such factors as geography and culture. Given the relatively large size of the Foursquare user base and the length of time the dataset spans, I hope that the effect of this bias is relatively small. Nonetheless, I cannot discount it entirely, and the reader should keep it in mind as a possibility.

The food category is shown to contribute a significant proportion of all new place growth, ranging from 19% to 33% in Manhattan. This is to be expected due to the entrepreneurial nature of food establishments. Accordingly, the average lifespan of food venues is relatively low, and so the category undergoes significant churn. Although many features of the growth profiles are shared between cities, such as the prominence of food venues, we can also observe differences in the dynamics of urban activity profiles of the cities. Downtown Toronto’s travel and transport categories display relatively high growth, where the airport services are the most visited venues with 64% frequency, suggesting greater investment in transport infrastructure.

The results of this study provided a cross city overview of urban activity growth patterns based on the most frequently visited places (food outlets and airport services). The study shows that local areas or neighborhoods which belong to the same borough are more likely to feature similar urban activity profiles. However, the data available through location-based services feature very high spatial granularity and today the position of real-world places is known with accuracy down to 5 m or less. This provides an opportunity to study urban activity growth patterns at an intra-city level. Motivated by these observations, this study identified that data from Foursquare location-based services could be exploited to identify areas in cities (for e.g. the travel and transport category, airport services in DowntownToronto), where there is a surge in urban development.

CONCLUSION

The present study identified areas in cities where we can detect a surge in urban growth. Therefore, in future, we can turn our investigation to how a new venue can influence traffic between other venues in these two local areas. Answering this question is of interest for the sake of understanding the impact of investment on certain urban activities on a large scale. It has also implications for important problems in retail geography, such as deciding where to open a new business in a city. This study further demonstrates how the digital datasets emerging from location-based technologies can become an important tool for monitoring urban development on large scale and with unprecedented geographical accuracy.

Limitations of the study: These two datasets suffer from general problem related to online sources. First, as its the case in most of psychological studies, the user distribution of online services is biased toward young people, living in cities and with a high level of life. Finally, as this study is problem and not data-driven, I conclude that this project needs to be further exploited using adequate user-survey based data towards achieving this goal of solving for important problems in retail geography, such as business investments in a city.