1. **Introduction**

When looking at different metropolitan cities, it can elude the casual observer as to why or how these places came to be. However, by looking at current venue data in a city’s neighborhoods, we can determine which areas may attract more tourists and which areas may be more suburban. When this information is contextualized in the city’s history, it is then possible to draw conclusions about how these cities developed from said history, and what trends they may follow in their future developments. In this project, I will be looking at the cities of Toronto, New York City, and Paris to determine how the current venue layouts in each of these cities relates to their respective histories.

1. **Data**

*Data Sources*

The dataset for this project were sourced from various places. For Toronto, postal code and neighborhood data were taken from a Wikipedia page with the boroughs and neighborhoods listed in a table. This table was scalped and formatted using Pandas. The respective latitude and longitude data for each of these neighborhoods was sourced from a previous project in this Capstone series. The New York City data was also sourced from a previous project in this Capstone series. As for Paris, the data was taken from a website curated by the city of Paris. In addition to the location data for each of these neighborhoods, venue location data was sourced from the Foursquare API.

*Data Cleaning*

Each of the previously mentioned datasets required a certain amount of cleaning and organization before any analysis could be done. For the Toronto dataset, it should be noted that if multiple neighborhoods shared the same postal code, then they were grouped together in the row labelled with the same borough and postal code. Longitudes and latitudes were then merged onto the data frame from a CSV based on the postal codes.

New York City neighborhood data was in a JSON file, and only the borough, neighborhood name, latitude, and longitude were taken from each entry in the file and converted to a Pandas data frame. Finally, the Paris data was also converted from a JSON file to a data frame with the arrondissement (borough), the neighborhood name, latitude, and longitude.

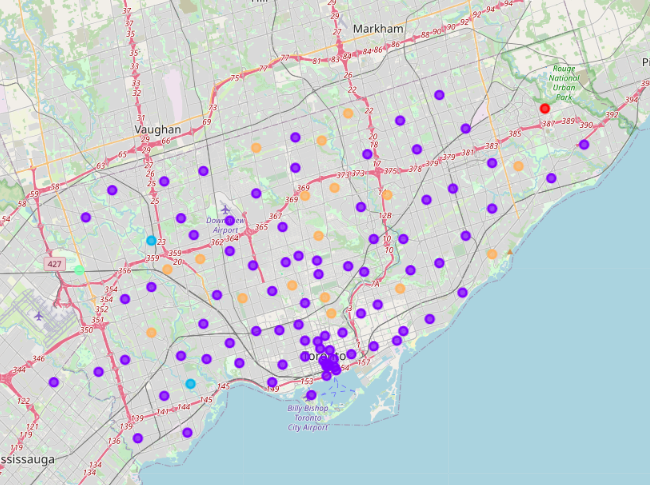
After the initial data cleaning was done, a method was written and used to collect data from the Foursquare API on what the top 100 venues are in an 800 meter radius of each neighborhood based on the latitude and longitude recorded for each area.

1. **Methodology**

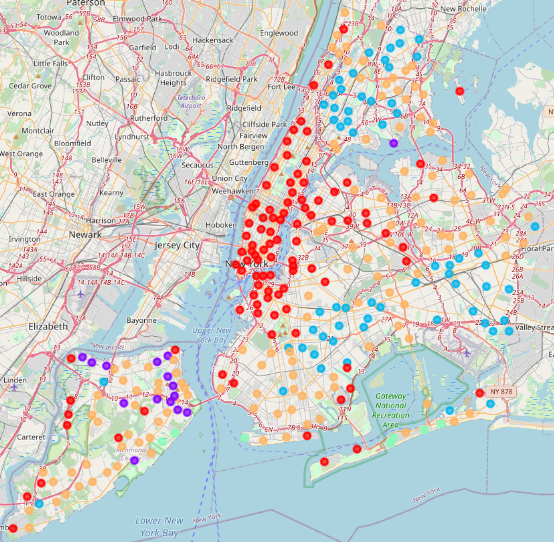
Analysis of the venue data from each city was done via k-means clustering. To do this, the venues were converted to dummies using the Pandas dummies function, then grouped together. The data was then sorted by the 10 most common venues in each neighborhood, then sorted into five different clusters based on the types of venues listed in the top 10. After clustering the data for each city was mapped onto their respective city map in the color of their respective cluster, and analysis was performed on the results.

1. **Results**

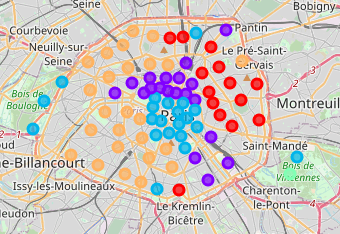
For Toronto, it seems that most neighborhoods fall into the second cluster (purple), with the most common venue being coffee shops or cafes. This cluster is spread out evenly across the area of the city, however, there is a concentration of points gathered around the downtown Toronto area. The first cluster (red) contains a single neighborhood where the most common venue is a fast food restaurant, and the second most common venue is a trail. This is likely because this neighborhood is located close to the Rouge National Urban Park. The second largest cluster is the fifth cluster (orange), which consists mainly of parks, grocery stores, and various types of restaurants, or in other words it contains many venues that may be used in people’s daily lives, suggesting that this area is more suburban.



In New York City, we can see that the majority of the first cluster (red) is in the borough of Manhattan. This is the second largest cluster after the fifth cluster (orange) which covers most of the other four boroughs (Queens, The Bronx, Brooklyn, Staten Island). Within the first cluster, the most common venues are Italian restaurants and coffee shops, however, there also seem to be a large variety of other ethnic restaurants and tourist attractions/venues such as parks, theaters, and hotels. On the other hand, the fifth cluster contains mostly pizza places and Italian restaurants, as well as some other ethnic restaurants and bodegas. The third cluster (blue) also contains many restaurants and bodegas, although there also happen to be many Caribbean restaurants and pizza places. The second cluster (purple) has mostly bus stops as the most common venue, and the fourth cluster (green) contains beaches.



For Paris, the western half of the city is dominated by the fifth cluster (orange), where the first and second most common venues in nearly every neighborhood are French restaurants and hotels respectively. In the eastern half, where the entirety of the first cluster (red) resides, there are mainly French restaurants and bars. The second (purple) also contains many French restaurants and hotels, but also has a variety of other ethnic cuisines, as well as shopping and entertainment venues. The third cluster (blue) has a similar make up of venues to the second cluster but contains more museums and galleries than the second and is also more densely clustered in the center of the city. The third cluster also include more neighborhoods that are located on the outskirts of the city.



1. **Discussion**

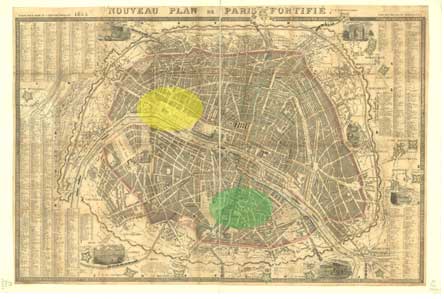
To understand the locations of the clusters in each of the cities, it is important to learn more about the historical context of each of the locations. In New York, the clusters have been divided such that Manhattan is dominated by the first cluster, and the other boroughs are mostly claimed by the third and fifth. As mentioned previously, the first cluster contains a variety of ethnic restaurants and coffee shops, as well as tourist attractions and entertainment venues. On the other hand, the third and fifth clusters consist of Italian restaurants, pizza places, other eateries, and bodegas.

From the types of venues in these clusters, we can infer that Manhattan, where the first cluster resides, has been built mainly as a place for visitors, whether that be tourists or people working, while the other boroughs have been constructed to be more for living in. This makes sense when we consider the history of NYC and how the city was created and built in the first place.

Manhattan is the tourist destination of NYC, and it is quintessentially what people think of when they hear New York. It is the home of Wall Street and the Empire State building, and a plethora of other tourist attractions including Broadway and Central Park. This is not to say that this borough does not also have places of residence. In fact, because of historical economic dichotomy, both luxurious townhouses and several ethnic enclaves can be found in the borough. However, it is clear from the analysis that the city has grown in a way to account for the mass amount of tourists that flood Manhattan each year, such that a majority of neighborhoods in the borough have developed to accommodate visitors.

Based off the data, there seems to be a clear divide between eastern and western Paris, with the fifth cluster and its French restaurants and hotels on the east side, and the first cluster and its restaurants and bars on the west. This divide can be attributed to the historical factions within Parisienne politics and how those factions were divided within the city. Historically speaking, eastern Paris has always been more liberal leaning, while western Paris is more conservative leaning. This divide is reflected in the monuments that are found on either side of the city, with the Bastille and the Pantheon, both monuments of Revolutionary pride, on the east, while military monuments like the Arc de Triumph and the Hotel des Invalides in the west.

The more touristy destinations, like shopping centers and museums are located in areas dominated by the purple and blue clusters (2 and 3) which are primary located in and around the center of the city. This is also in line with how the city was built as the areas to the northwest and southeast of the city center have always been considered the social centers of Paris. In other words, these are the chic and trendy places within the city.



Finally, when we look at Toronto we can see that the majority of neighborhoods fall within the second cluster, and that within this cluster the most common venues are coffee shops and cafes, with a concentrated grouping of this cluster in the area around the university and near the waterfront. This makes sense when we remember that Toronto was initially a trading city for the fur trade. It is likely that the settlement started on the waterfront with a harbor and gradually expanded more inland as time passed. In addition to this, Toronto’s population is much less dense than that of New York City or Paris, which is likely another factor as to why there is just one major cluster rather than five distinct clusters.

1. **Conclusion**

While the analysis of city venues in neighborhoods can be a bit abstract, putting it in the context of each cities respective history can give us context and explain how and why metropolitan areas grow and develop the way they do. This work could be refined further by sorting and counting the venue types in each city, and in the future it could be used to discover new trends for tourism and business ventures.