**A Glance of Major Cities in Canada**

* A Tourist’s point of View

Wenjing Huang

Contents

[Abstract 2](#_Toc18276942)

[1. Introduction 2](#_Toc18276943)

[1.1. The inspiration of the business problem 2](#_Toc18276944)

[1.2. The generalization and application of the analysis 2](#_Toc18276945)

[2. Data 2](#_Toc18276946)

[2.1. Zip Code and Neighborhood 3](#_Toc18276947)

[2.2. Coordinates 4](#_Toc18276948)

[2.3. Venue Information 4](#_Toc18276949)

[3. Methodology 5](#_Toc18276950)

[3.1. Overview of the cities 5](#_Toc18276951)

[3.2. Neighborhood Clusters 6](#_Toc18276952)

[3.3. Deep-dive into food selection 6](#_Toc18276953)

[3.4. Deep-dive into activities 6](#_Toc18276954)

[4. Results 6](#_Toc18276955)

[4.1. Overview of the cities 6](#_Toc18276956)

[4.2. Neighborhood Clusters 7](#_Toc18276957)

[4.2.1 Toronto 8](#_Toc18276958)

[4.2.2 Vancouver 9](#_Toc18276959)

[4.2.3. Ottawa 10](#_Toc18276960)

[4.3. Deep-dive into food selection 11](#_Toc18276961)

[4.4. Deep-dive into Activities 11](#_Toc18276962)

[5. Discussion 11](#_Toc18276963)

[6. Conclusion 12](#_Toc18276964)

# Abstract

In this report, we will compare and contrast the 3 major cities in Canada – Toronto, Vancouver, and Ottawa. We will use Foursquare’s venue data as the primary dataset, discuss the similarity, the dissimilarity of the 3 cities, and the unique characteristic of each city.

# Introduction

## The inspiration of the business problem

As tourist visiting Canada for the first time, the major cities Toronto, Vancouver, and Ottawa are on my list. While planning my trip, I want to learn about the cities on the high level, so that I can prioritize the places I want to visit, and plan the time accordingly. In particular, I would like to look at the 3 cities in the following areas:

* + 1. **Downtowns**: How similar/dissimilar are the downtown areas?
    2. **Neighborhoods**: What are the different neighborhoods in each city?
    3. **Food Selection**: What are the most popular cuisines in each city?
    4. **To-do**: What are the most popular recreational activities in each city?
    5. **Other**: What are some of the unique characteristics of each city?

## The generalization and application of the analysis

This analysis can be generalized and packaged as a product, which can be a customizable trip planning service that’s embedded in many of the travel sites. The distinguishable features of this service include:

* Customers can select any combination of cities they want to visit, rather than most of the current travel sites that provide plans for only fixed combination of cities.

Rather than just giving out plans, compare and contrast the cities from various perspective to give travelers more background information. Travelers can then refine their own plans to include/exclude, or increase/reduce time in certain cities based on what they see and what interest them.

* In each category (such as food, culture, activities), customers can have an overview of what’s most popular and what’s most unique to each city, so that they can prioritize based on their own preferences.

# Data

To Achieve this, the following data will be used

* Data that contains the zip code and neighborhood information of Toronto, Vancouver, and Ottawa.
* Latitude and Longitude information by zip code
* Foursquare data that contains venue information in the 3 cities

I will discuss each of these in the upcoming section.

## Zip Code and Neighborhood

To start out, we need the basic neighborhood location and name for each city. This information can be found online. I will be obtaining this data by scraping a few Wikipedia pages.

* + 1. **Toronto – well-formatted data**

For Toronto, there’s a Wikipedia page that contains well-formatted data source that contains zip code – neighborhood information: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

  
Fig 2.1.1

Although some processing and cleansing is needed for missing data, overall the processing is easy and straightforward.

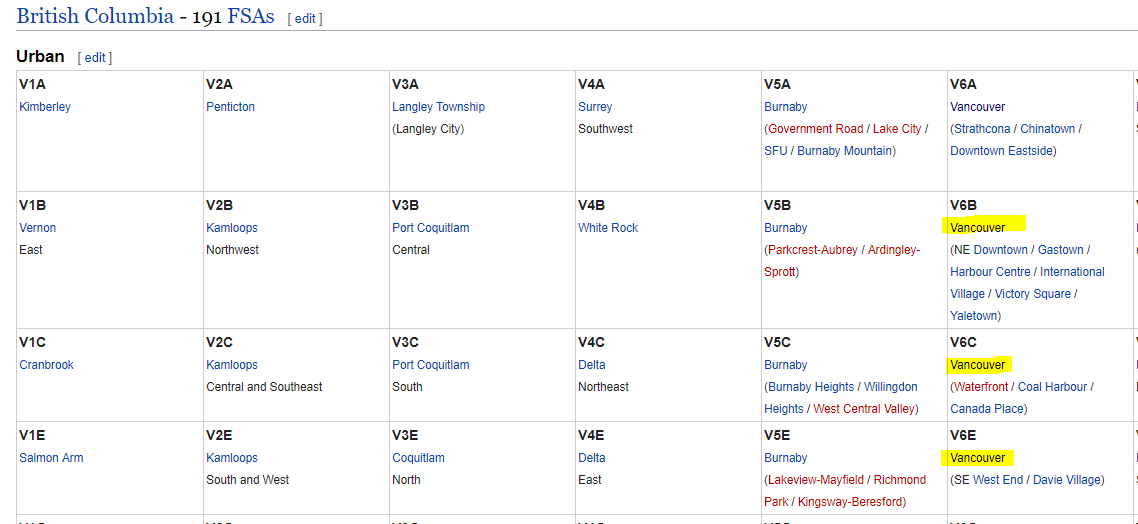
* + 1. **Vancouver and Ottawa – more processing needed**

Similar zip code – neighborhood information can be found for Vancouver and Ottawa:

Vancouver: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V>

Ottawa: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_K>

But the data is not organized in a format that’s easy for us to use.

  
Fig 2.1.2

As shown in the screenshot, we need to parse the data into a tabular format for the analysis in the next steps; and it needs to be limited to the areas that’s related to the analysis.

## Coordinates

Part 2.1 will give us basic neighborhood information. In order to initiate API searches for venues in each neighborhood, we also need their location information in latitudes and longitudes.

This information can also be found online. For this analysis, the data from this Google fusion table can be used:  
<https://fusiontables.google.com/DataSource?docid=1H_cl-oyeG4FDwqJUTeI_aGKmmkJdPDzRNccp96M&hl=en_US&pli=1>

We can filter the data by the first letter in the zip code (FSA-1) to download only data that’s relevant to this analysis.

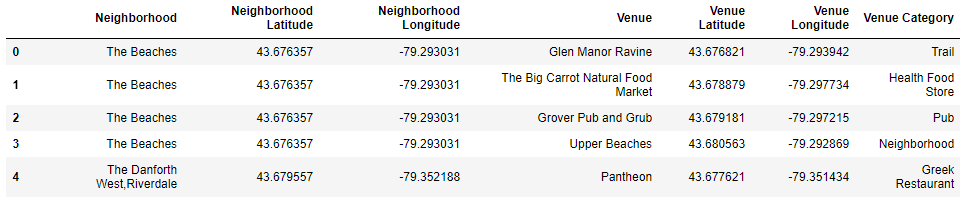
The information will then be imported and combined with the neighborhood dataset in preparation for the next steps.

## Venue Information

To compare and contrast the cities, we will look at what kind of venues there are in each of their downtown areas. We will use Foursquare API data to obtain this information.

The format of the API is:  
[https://api.foursquare.com/v2/venues/explore?client\_id={}&client\_secret={}&ll={},{}&v={}&radius={}&limit={}](https://api.foursquare.com/v2/venues/explore?client_id=%7b%7d&client_secret=%7b%7d&ll=%7b%7d,%7b%7d&v=%7b%7d&radius=%7b%7d&limit=%7b%7d)

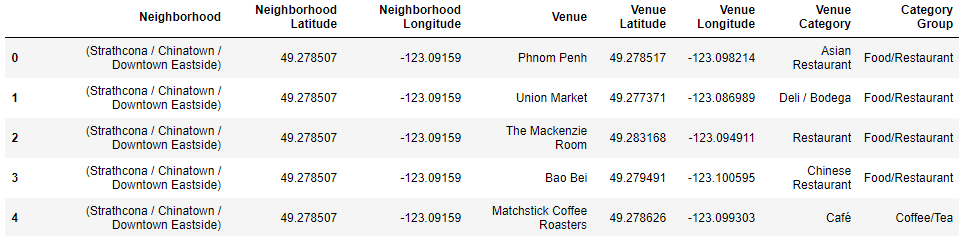
Fig 2.3.1 shows a sample of venue data obtained for the neighborhoods and locations extracted in step 2.1 and 2.2:

  
Fig 2.3.1

Upon scanning some more lines of the data, it’s noticeable that the “Venue Category” data is very granular. For example, we saw food and restaurants are categorized into very detailed cuisine specific groups. The analysis may need some more general grouping, and we will do some manual work to put the “Venue Category” items in higher level groups.

For this particular analysis, there are only a finite amount of venue categories, so I will do the work manually in excel. If the analysis is to be generalized, we may need to study the data more and come up with a more intelligent and repeatable way to do this.

The higher-level group will be called “Category Group”, and will be imported back and merged with the venue dataset. Here’s some examples of the data records

  
Fig 2.3.2

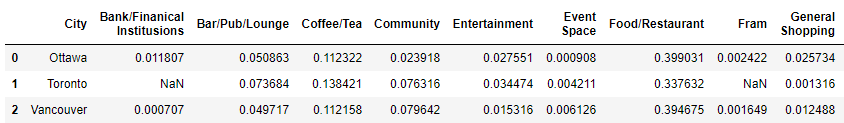
The data is now ready to be used. We will now perform some analyses to answer the business questions raised at the beginning of this report.

# Methodology

## Overview of the cities

In this section, we are going to look at the top venue types (Category Groups) in each of the 3 cities. By comparing and contrasting the venue type ranks and densities, we can gain an overview of each cities and their characteristics.

To achieve this, the frequency of each category group is calculated. Fig 3.1.1 shows a sample of the data:

  
Fig 3.1.1

Then we sort the category groups by their frequency for each city. Fig 3.1.2 shows a sample of the data:

  
Fig 3.1.2

## Neighborhood Clusters

In this section we are going to look at what different neighborhoods there are in each city. Through this we will know what city functionality is in what relative location.

We will be using k-nearest neighbors (KNN) algorithm to cluster the neighborhoods based on the venue category groups that present in each of them.

For each city, we will group the neighborhoods into 4 clusters. For each cluster, we will compare and contrast the venue category group frequencies, top groups, and observe their location on a map.

## Deep-dive into food selection

In this section, venue data will be used and limited to “Food/Restaurant” and “Specialty Food” Category Groups. We will compare and contrast each city’s food selection based on venue categories’ frequency ranks.

## Deep-dive into activities

In this section, venue data will be used and limited to "Outdoor", "Tourism", and "Historic/Museum" Category Groups. We will compare and contrast each city’s activity options based on venue categories’ frequency ranks.

# Results

## Overview of the cities

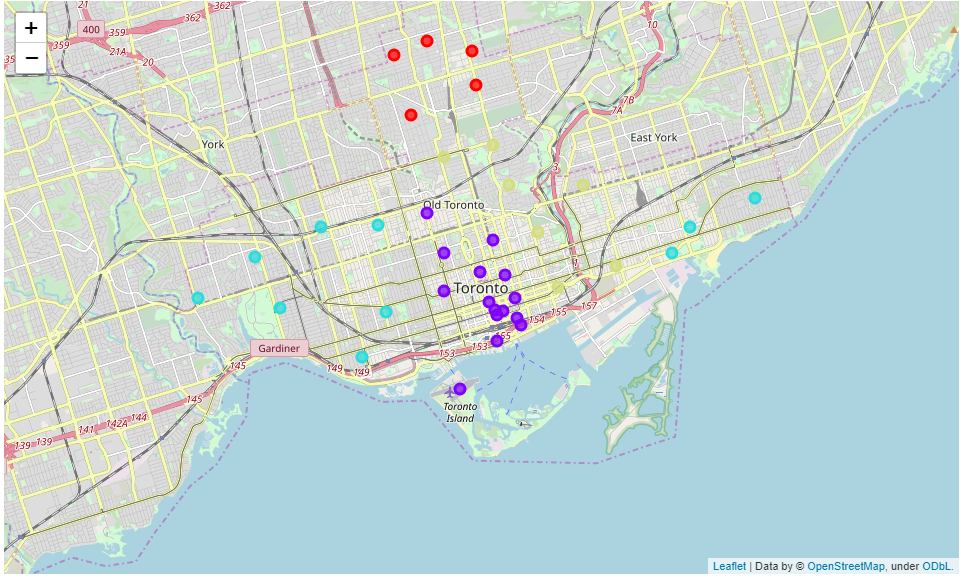
By looking at the individual frequencies of each category group and their ranks, we observed the following:

* + Top 3 Venue Types (Category Groups)
    - No.1 Venue Type: For all 3 cities, Food/Restaurant is the most common venue type in the downtown area.
    - Other members in Top 3 Places: Coffee/Tea places are also among the top 3 venue types for all 3 cities. The remaining member in the top 3 places in both Ottawa and Vancouver is "Specialty Store", whereas in Vancouver "Specialty Food" made the 3rd place. This makes the density of food venues in downtown Vancouver the highest among the 3 cities.
  + Shopping
    - With "Specialty Store" being the no.2 venue type, and General Shopping (which primarily comprises of department stores, shopping malls & etc.), Ottawa appears to be the city that will be most attractive to shoppers.
  + Car Dependency
    - With a significantly higher density of public transportation facility (Category Group = Transportation), Ottawa is the least car dependent city, which usually means more flexibility and convenience to travelers.
  + Culture
    - Ottawa and Toronto have more Cultural venues with a combined Entertainment and Historic/Museum density of 5% (compared to 1.8% in Vancouver)
  + Outdoor
    - Vancouver has much more Outdoor activity options (5% venues being Outdoor compared to 2% in Toronto and only 1.2% in Ottawa).
  + Business
    - Vancouver has much higher density of Office spaces.
    - Vancouver also has the highest Event Spaces and Hotel/Accommodation venues, which makes it a place very suitable for conventions, exhibitions, and other events.
  + Other
    - Toronto may be a more comfortable and convenient city to live in with higher density of Service venues (salon, spa, auto shop, & etc.), and a good balance of other venues including public facilities (Category Group = Community), grocery store/supermarket, gym/fitness, and good amount of shopping and food selection in general. It also seems to be an interesting place to live in with its high density of bars, pubs, and coffee/tea places, and the good amount of cultural venues including museums and entertainment venues.

## Neighborhood Clusters

We will now take a look at the clustering results for each city.

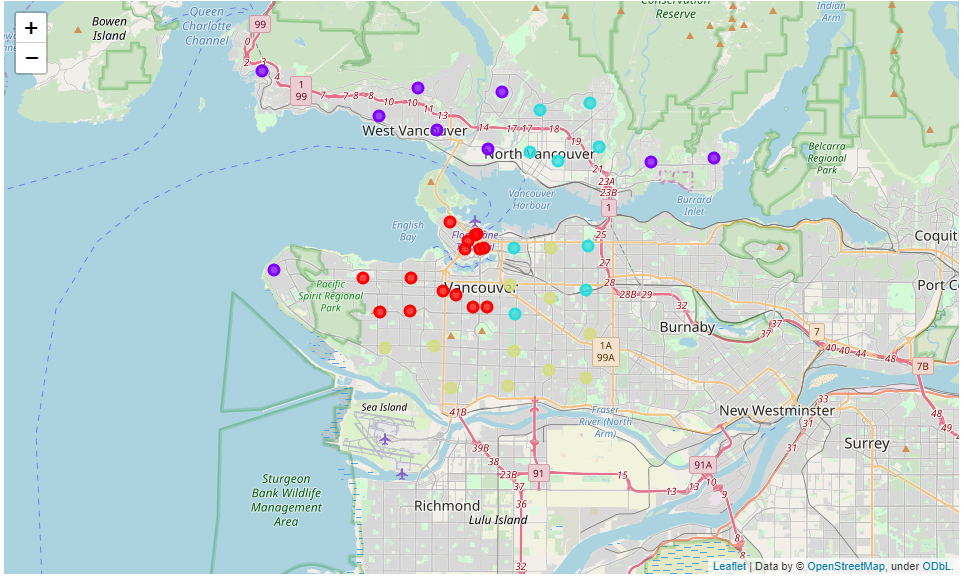
### 4.2.1 Toronto

  
Fig 3.2.1

From the map visualization and some analysis on the venue category group frequencies and ranks, we have the following observation for the 4 clusters in Toronto:

* + **Cluster 0: Residential (red)**
    - Venues include Shopping, Pharmacy, Service, and a decent amount of Grocery/Supermarket, Outdoor, and Sports/Fitness (including gyms and studios) facilities
  + **Cluster 1: Event/Tourism (purple)**
    - High concentration of Event Space, Hotel/Accommodation, Transportation, Tourism, Entertainment, Sports/Fitness (including stadiums)
  + **Cluster 2: Business/City Center (teal/blue)**
    - High Concentration of Bar/Pub/Lounge, Food/Restaurant & Specialty Food. Office buildings are also found in this area, and it also has a decent amount of public facilities
  + **Cluster 3: Culture/Education (lime/green)**
    - Center of culture and education. Where you can find historic sites and museums, training and education institutions. There's also a decent amount of public facilities, entertainment, and hotel/accommodations.

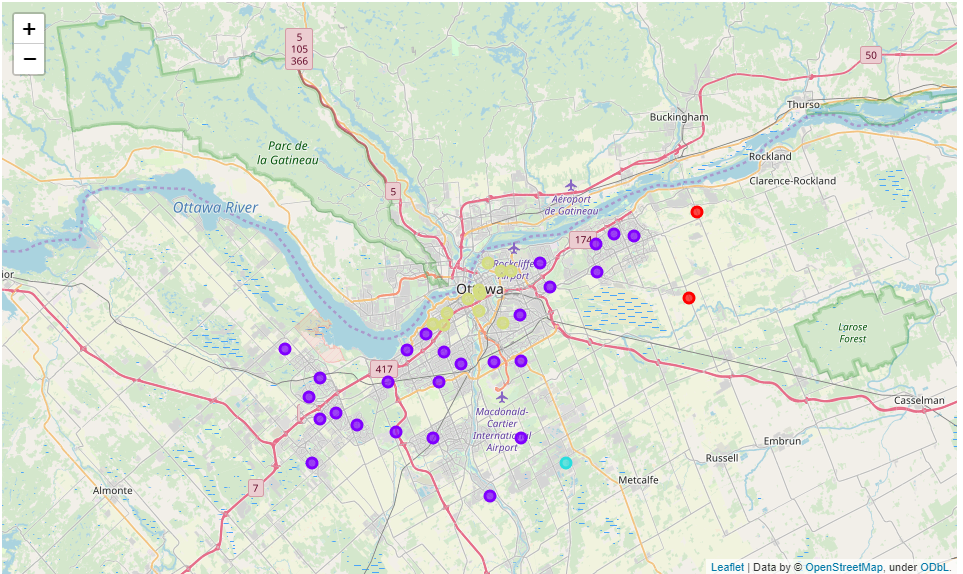
### 4.2.2 Vancouver

  
Fig 3.2.2

From the map visualization and some analysis on the venue category group frequencies and ranks, we have the following observation for the 4 clusters in Vancouver:

* + **Cluster 0: City Center (red)**
    - Center of entertainment and hotel/accommodation for the city, also has good amount of food selection including specialty food and general restaurant. Some training and education institution and outdoor facilities can also be found.
  + **Cluster 1: Residential, Harbor, Tourism & Outdoor (purple)**
    - Higher centration of banks, public facilities, general shopping, grocery store/supermarket and pharmacies make it convenient for residents. It has a good amount of office buildings, making commute easier for office workers.
    - Unlike residential areas in Toronto, this function area in Vancouver also has a good amount of outdoor facilities, tourism sites, historical sites and museums.
    - The harbor and public transportation facilities make it convenient for both residents and visitors.
  + **Cluster 2: Culture & Entertainment (teal/blue)**
    - This area is a good location for local cultural events and entertainment, with high concentration of bars, pubs, and lounges, decent entertainment, food venues, and historic sites/museums.
  + **Cluster 3: Event & Entertainment (lime/green)**
    - This is the event center for the city. It has the highest density of event spaces, great food selection including both general restaurants and specialty food, and decent entertainment options.

### 4.2.3. Ottawa

  
Fig 3.2.3

From the map visualization and some analysis on the venue category group frequencies and ranks, we have the following observation for the 4 clusters in Ottawa:

* + **Cluster 0: Outer City (red)**
    - Only two neighborhoods are categorized in this cluster. This cluster is primarily comprised of farms, historical sites and museums, transportation facilities, and sports facilities.
  + **Cluster 1: Residential (purple)**
    - This is the residential area with easy access to banks, entertainment, food and restaurant, shopping (general and specialty goods), grocery store and supermarket, pharmacies, service facilities, and gyms and fitness facilities.
    - There are also some office buildings which makes it convenient for commuters.
  + **Cluster 2: Suburb (teal/blue)**
    - Only one neighborhood (Greely) is in this cluster. Only four venues are found in this neighborhood, which is a zoo, a post office, a trailer park, and a bazaar. This is a suburban neighborhood.
  + **Cluster 3: City Center (lime/green)**
    - This is the city center with high concentration of bars, pubs, and lounges, entertainment, event spaces, great food selections including general restaurant and specialty stores, and hotels and accommodations.
    - This is also the center of public cultural functions including public service facilities, historical site and museums, and libraries.

## Deep-dive into food selection

From the venue category frequency analysis, we have the following observation:

* + Relative to the other two cities, Ottawa has less international food selections and more casual eating options (pizza, fast food, and sandwiches)
  + Toronto has more international food by having Japanese, Italian, Mexican, French, Mediterranean, and Spanish (Tapas) restaurants as top venues.
  + Vancouver has more Asian food by having Japanese, Vietnamese, and Indian restaurants in the top venue list.

## Deep-dive into Activities

From the venue category frequency analysis, we have the following observation:

* + Ottawa has relatively limited outdoor activity options. There is ample amount of cultural activities in Ottawa such as museums, art, historical sites & etc.
  + There are plenty to look at in the city of Toronto, in addition to museums and historical sites like Ottawa, Toronto also has more varieties like Castles and Aquariums.
  + In Vancouver, visitors have many outdoor activity options. There are trails, beaches, scenic lookouts, national parks & etc. in Vancouver area.

# Discussion

While working on this project, it came to my attention that the following areas will affect the accuracy and quality of the analyses:

**Venue data categorization**

the current venue category values seem to be quite liberal, sometimes inaccurate categorization may happen. This is common in user-submitted data.

When the data volume is big (in big cities like Vancouver), usually it won’t affect the overall quality of the analyses too much. But if the data volume is limited (like the few suburban neighborhoods in Ottawa), a few mis-classification may have bigger impact on overall accuracy of analyses. In this case, more human intervention/review of the data will be needed.

**Granularity of venue category data**

As discussed in the Data section, the current “Venue Category” data from Foursquare API is too detailed and granular. This will post challenges in clustering analysis. In this project, a manual grouping of the Venue Category data is performed, and it greatly helped the effectiveness of the clustering algorithm (the clusters based on Category Groups are much cleaner and more explainable than those based on the raw Venue Categories).

If this analysis were to be generalized to be used on more cities, a better method to create the Category Group will be needed. This can be done through a few different ways:

* Improvement of Foursquare data submission (make a mandatory Category Group field with pre-determined value selections upon venue creation)
* A more intelligent algorithm to create the Category Group upon data processing. Regular Expression and/or other string analysis methods will need to be used.

# Conclusion

Through the analyses of venues in the 3 major cities of Canada – Toronto, Vancouver, and Ottawa, we have gained knowledge of the similarities and dissimilarities of the cities.

* Toronto is a very diversified city, it has many facilities that make it convenient and comfortable for residents to live in, and it also has many cultural and entertainment activities which makes it a very interesting place for tourists to visit.
* Vancouver is a city with significant immigrant presence and it has great food selections. It has many outdoor attractions which makes it a wonderful destination for nature lovers. It is also a great city for conventions and events.
* Ottawa as the capital of the country, it has many cultural and historical sites. The public transportation system is advanced here which makes it easy for visitors to get around. It is also a good place for shopping with a high concentration of department stores, shopping malls and specialty stores.

As a tourist, this project provided me with clear pictures of the unique characteristics of the 3 cities, and I now have a clear vision and expectation when I visit the cities.