

# GREEK RESTAURANT IN TORONTO, CA

---

IBM DATA SCIENCE  
CAPSTONE PROJECT

JAVIER ARNAU

# INTRODUCTION & BUSINESS PROBLEM DEFINITION (BPD)

---

- The City of Toronto, is one of the most populous city in Canada. It is multicultural and provides a lot of business opportunities and a business friendly environment. It has attracted many different players into the market. This capstone project is focused on providing such an analysis for a Greek restaurant business in Toronto.
- Toronto's food culture includes an array of international cuisines influenced by the city's immigrant history. Greek restaurants have become popular in Canada. New Greek restaurants should be open in an area that inadequate neighborhood in this way the bar can attract more customers. This project aims to answer the business question: "In Toronto, if an entrepreneur wants to open a Greek restaurant, where should they consider opening it?"

# DATA COLLECTED FOR PROJECT

---

- List of neighborhoods in Toronto:
  - [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada)
- Latitude and longitude the neighborhoods:
  - [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data) (CSV extension file)
- Venue data Greek Restaurants:
  - Using Foursquare API

# METHODOLOGY - DATA, ANALYSIS AND FEATURES

---

- Get the list of neighborhoods in Toronto (Wikipedia)
- Apply exploratory analysis and feature engineering to the original DataFrame.
- Retrieve coordinates ([http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)) to then utilize Foursquare to pull the list of venues near these neighborhoods.
- Visualize the map of Toronto using Folium package
- Use Foursquare API with personal Client ID/Secret dev credentials
- From Foursquare, I pulled the names, categories, LAT& LONG of the venues and checked how many unique categories exist.
- Then, analyze each neighborhood by grouping the rows by neighborhood.



# METHODOLOGY - RUN K-MEANS

---

- Perform the clustering method by using **k-means clustering**, for  $k=3$ . I clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for "Greek restaurant".
- The results from k-means clustering show 3 clusters based on how many Greek restaurants:
  - **Cluster 0:** Neighborhoods with no Greek restaurants
  - **Cluster 1:** Neighborhoods high number of Greek restaurants
  - **Cluster 2:** Neighborhoods with little or no Greek restaurants

# RECOMMENDATIONS AND LIMITATIONS

---

- Recommendations: Most of the Greek restaurants are concentrated in cluster 1. Also, the recommendation is to avoid neighborhoods in cluster 1 which already have high concentration of reek restaurants.
- Limitations: In this project, there are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new restaurant.