# The best area for a pub in Toronto

*Applied Data Science Capstone*

**Vitor Machado**

## INTRODUCTION

Following the suggestions and teachings from the courses, this project uses Foursquare data to discover the best location for pubs in Toronto. This knowledge could be useful to someone who wants to open a new pub or related businesses such as hotels and nightclubs.

## DATA

The data used in this project is derived from Wikipedia and Foursquare only. To get the data from Wikipedia, which is relative to the postal codes of Toronto Neighbourhoods, I used web scraping (BeautifulSoup) as learned in the previous courses. The data from Foursquare is relative to the location of pubs in each neighbourhood. Both datasets were combined into a single dataset to determine the best location for pubs using sklearn.cluster library.

## METHODOLOGY

For this project, the following python libraries were used: *pandas, numpy, matplotlib, seaborn, folium, requests, json, bs4 and sklearn.cluster and yellowbrick.cluster*. Using the Toronto neighborhoods dataset from Wikipedia and the location data from Foursquare, one dataset was generated and used for clustering. *Yellowbrick.cluster* was used to determine the ideal value of K. New dataframes were created for each cluster and then compared using *matplotlib*.

## RESULTS

After analyzing each cluster dataset and comparing the values, the results showed that the best location is East Toronto since it has the highest average number of pubs.

## DISCUSSION

The project can be helpful to pub owners in Toronto but it is important to add that the analysis included only venues provided by Foursquare. It is also relevant to add that other businesses that operate in the same category but are not labeled as pubs were not included.

## CONCLUSION

The results suggest that there is a considerable demand in the area, considering the proportion of pubs. This information could be useful for new business owners interested in starting pubs or related enterprises.