# ANALYSIS OF VENUES CATEGORIES IN THE NEIGHBORHOODS OF TORONTO

## DATA

| | | |
|---|---|---|
| Title | Analysis of venues categories in the neighborhoods of Toronto | |
| Reference | N/A | |
| Revision | 0 | |
| Location | Toronto (Canada) | |
| Date | April 4, 2020 | |
| Petitioner | N/A | |
| Author | Name | Louis Sang |
| | ID | N/A |
| | Qualification | MS Business Analytics |
| | Official Member's N. | N/A |

# 1. INTRODUCTION/BUSINESS PROBLEM

## 1.1. Business problem

How are venues distributed among the neighborhoods in Toronto? Where are the neighborhoods in which a certain venue category is specially usual? Which is the most common venue category in each neighborhood? Is there any "outlier" neighborhood in Totonto? Could this information be leveraged in any way? **All these questions define the business problem** around which this data science project revolves.

## 1.2. Stakeholders

Toronto is a city brimming with culture, leisure, catering and hostelry, and so it attracts many entrepreneurs and investors from these fields. **These are the stakeholders in this business problem**. To boost the chances of success of their enterprises and investments, many of them rely on market researches which try to answer questions such as:

- If a business of a specific type is to be put into operation, where are the neighborhoods in Toronto which should be targeted? How are they distributed in the city? In the suburbs, in the down town, in the old city?

- If a the stakeholder plans to put a business of any type into operation in a specific neighborhood of Toronto, what business types should they consider? Is this neighborhood saturated with a particular type of business? Does this neighborhood lack any type of business? What is the market niche?

- Is there any "special" neighborhood in Madrid regarding its venues? Does it offer any special market opportunity to the entrepreneurs and investors?

- Finally, considering that venues open, close or simply change overnight, would it be possible to have this information dynamically updated?

**This data science project aims to answer all of these questions for the stakeholders.**

## 1.3. Data

The following data is used:

- Toronto boroughs and neighborhoods data. They are freely available on Wikipedia and Foursquare API[1]. They come in the form of geospatial vector data files of Toronto boroughs and neighborhoods. Once downloaded, these files are loaded in QGIS, a free and open-source cross-platform desktop geographic information system, in order to be processed. The files are processed leveraging QGIS functionalities so that the corresponding attributes table contains the following information for each neighborhood of Madrid.
    - Borough name.
    - Neighborhood name.
    - Area.
    - Perimeter.
    - X coordinate (ETRS89/UTM zone 30) of its centroid.
    - Y coordinate (ETRS89/UTM zone 30) of its centroid.