**Measuring the similarity of New York City and Toronto**

Dennis Menze

May 06, 2019

# Introduction

## Background

New York City and Toronto are very diverse and are the financial capitals of their respective countries. But despite their diversity, how similar are they? How could this similarity be measured? Who could be interested in such a measurement?

## Problem

The goal of this project is to propose a solution for measuring the similarity of the neighborhoods of both cities.

## Interest

Different clients or groups of people could potentially be interested in the comparison of how similar or dissimilar both cities are in terms of their neighborhoods: citizens of both cities who consider to move to the other city; sociologists who study the relationships e.g. between the financial industry and city architecture; investors or traders who want information in order to decide in which city they will focus.

In the next sections, it will become more clear how the neighborhoods will be compared, and the results of the comparison will be presented.

# Data acquisition and cleaning

## Data sources

For the measurement of the similarity of the neighborhoods of New York City and Toronto, data is needed. Data to be used for this project will be venue data and geolocation data from Foursquare, data about the neighborhoods of New York City provided by the NYU Spatial Data Repository available on the Internet, and data about the neighborhoods of Toronto taken from Wikipedia.

This data will be joined and clustered based on the venue data. Then, the resulting clusters will be examined with respect to their contents in order to determine the similarity of the neighborhoods of New York City and Toronto. Specific details of the comparison follow in the next sections.