

IBM Applied Data Science Specialization

Capstone Project

By Arshiya Tripathy

1. Introduction

A lot of people who have to travel to different cities for work or vacation for a longer duration would want to understand how similar or different is that place in comparison to their current residential city or a city they have lived in past or are very familiar with. Someone would like to visit/stay in a city very similar to their hometown and someone would want the complete opposite. Adare, County Limerick in Ireland is a very quiet town full of heritage whereas Dublin is modern metropolitan city filled with top end restaurants, theatres, shopping centres etc. A city can be defined by the venues and places it has to offer, some can be rich in culture with lots of museums , chic cafes and other can be fast and modern with high-end bars, fast-food chains etc. Thus, two cities can be compared based on similarities or differences between what venues each has to offer to its residents. This profiling can be used by travellers to choose the city based on their preferences, making their stay more comfortable/enjoyable.

2. Problem Statement

Cities and towns in the island of Ireland (Norther Ireland and Republic of Ireland) are selected for this analysis. These cities are then clustered to determine which set of cities are similar or different than each other based on the venues present in them. This clustering will help decide the traveller, based on his/her preference which city should he/she visit next. The traveller can compare their current city or any other city they are familiar with to decide the next city they want to visit by looking at the cities present in the same cluster of their current or other know city.

3. Data for analysis

A list of 60 largest town and cities on Island of Ireland is collected by web-scraping the Wikipedia page https://en.wikipedia.org/wiki/List_of_settlements_on_the_island_of_Ireland_by_population. This gives the name of the city, province, county and country's name along with a small description and population. The city name is used to query foursquare database to get a list of popular venues around the latitude and longitude of the city's centre. All this information will be used to create clusters of cities based on the venues they have to offer.