Battle of the Neighbourhoods
New York vs Dublin

Coursera IBM Applied Data Science Capstone
Warren Shelley
31st of October 2020

Dublin Image (https://unsplash.com/photos/XI1yvtqKjl8)
New York Image (Photo by Quintin Gellar from Pexels)

Introduction

New York city is one of the most diverse and multicultural urban metropolises in USA and the world. It is the largest city in the USA with a population of approximately 8.3 million[1]. As a centre of multiculturalism, New York City has been a gateway for immigration to the USA throughout the years. Today it is the most linguistically diverse city in the world[2] that is home to 3.2 million immigrants[3].

Irish immigrants are historically one of the largest of these immigrant populations. This large pattern of immigration began during the Great Irish Famine of the 1840s and continued through much of the 20th century. As a result Irish culture has had a significant impact of the development of neighbourhoods within New York City.

Dublin is the largest city in Ireland and it is one of the hubs of Irish cultural. A comparison of the types of venues in these cities may reveal the similarity between neighbourhoods in the two cities.

Problem

A large company with its HQ in New York City has plans to open a European HQ in the centre of Dublin city. As one of several new employee of this company I have been assigned to a new role in the European HQ, but will have to move to New York for several months for job specific training. In order to prepare for the move I will investigate which neighbourhoods of New York are most similar to those in Dublin. Next I will investigate the average rent price for a one bedroom property in each of these neighbourhoods. This will be presented to the HR department of the company so that they can provide guidance to other employees making a transition from Dublin to New York. The main goal of this analysis is to find the most affordable New York neighbourhood which is similar to neighbourhoods in Dublin, so that those moving to New York for training purposes have a smooth as possible transition.

Data

To tackle this problem the following data sources are required:
- New York City data containing neighbourhood and borough information.
- New York venue information.
- Dublin data containing neighbourhood and borough information.
- Dublin venue data.
- New York average rent price per neighbourhood data.
- Dublin average rent price per neighbourhood data.

The data will be sourced from multiple open source sources. The New York neighbourhood information was extracted from the NYU dataset JSON file (https://geo.nyu.edu/catalog/nyu_2451_34572). The Dublin neighbourhood dataset was built by using the python BeautifulSoup library to web scrape the neighbourhood information from the Wikipedia page (https://en.wikipedia.org/wiki/List_of_Dublin_postal_districts) and reformatting into a

structured format. The geographical coordinates of neighbourhoods was found using the Python Geocoder package.

The venue information for each neighbourhood is retrieved using the Foursquare API. The Foursquare API has one of the largest databases of 105+ million places and is used by over 125,000 developers.

The rent data will be extracted from (https://streeteasy.com) which provides rental data for each neighbourhood in New York. The Dublin rental data is acquired from the Irish Central Statistics Office annual Rent Tenancy Board report.