Moving to Raleigh from Indore

24.06.2020

—

Anmol Lunavat

IBM Applied Data Science Capstone

1. Introduction

Each year there are a number of people who move from one place to another for variety of reasons such as a new job, education, or retirement. Moving to a new city though exciting comes with many challenges. Since we are accustomed to the place and neighborhood we have been staying in, it is easier to make this shift if the new city has a familiar environment. There are multiple factors which constitute in deciding whether two cities are similar such as food, transportation, climate, culture etc. Though, factors such as climate or culture aren't under our control, we can analyze our new city based on food preferences and venues. Therefore, it will be beneficial to predict similar neighborhoods between two cities based on their venues.

Venues surrounding a neighborhood can be analyzed on location and category. Additionally, venue categories can be used to classify similar neighborhoods together. The aim of this project is to predict neighborhoods in Raleigh which are similar to neighborhoods Vijay Nagar and Old Palasia in Indore. The project implements a clustering method which utilizes different categories of venues to group neighborhoods in Raleigh.

The findings of this project will interest people planning to move from Indore to the city of Raleigh.

2. Data Collection and Strategy

To analyze the similarity between neighborhoods of Raleigh with Vijay Nagar and Old Palasia in Indore we will gather a list of neighborhoods and venues in Raleigh.

2.1 Neighbourhoods

A list of neighborhoods in Raleigh could be found on the Wikipedia page Raleigh, North Carolina Neighborhoods. Python web scraping techniques with URL handling and beautifulsoup packages will be utilized to extract the list and convert it into a pandas dataframe. Following this, we will utilize Python geocoder package to collect latitude and longitude values for all the above collected neighborhoods.

2.2 Venues

In the second part of data collection we will be utilizing Foursquare API to gather a list of venues in the neighborhoods of Raleigh. Particularly we will be using Places by Foursquare, a database of more than 105 million places worldwide and API services that enable retrieval of location data. Format of the data received is not appropriate for the purpose of analytics. Thus, we will be using data wrangling techniques to transform the data into a more appropriate and valuable format.