IBM Data Science

CAPSTONE PROJECT

Opening a Pizza Place in Oakland, CA

Charlotte Barelle
8-18-2020

Introduction

Pizza places have become a standard type of restaurant in many countries, including the USA. These restaurants are extremely popular for tourists and locals, wanting a nice meal out or wanting a quick bite after work. Due to the abundance and popularity of pizza places, including large chains and local restaurants, competition is booming in many cities as new pizza restaurants are opened. Opening any kind of business, especially one so popular, requires many considerations. A key decision to be decided is the location of such a restaurant, as this can directly determine its success.

This project aims to determine the most suitable areas in Oakland, CA, to open a new pizza place. The target audience of this project are businesspeople interested in opening a new Pizza Place in the area. This is relevant right now as pizza places are extremely popular and so opening a new one in the optimum area would be of interest to certain businesspeople. To this aim, data science methologies and machine learning techniques will be employed to analyse and cluster Oakland neighbourhoods in order to solve the business problem: "If a business person is interested in opening a new pizza place in Oakland, CA, in which area(s) would they be recommended to do so?".

Data

The following data will be required for this project:

1. A list of neighborhoods in Oakland, CA, USA. This defines the strict region this project is focusing on. This data can be obtained from the Wikipedia page https://en.wikipedia,org/wiki/Category:Neighborhoods_in_Oakland,_California which contains a list of these neighborhoods. This data will be scraped from the Wiki page using Python requests and BeautifulSoup packages.
2. Latitude and longitude coordinates of the respective neighborhoods in Oakland. This is required in order to create a map and obtain venue data for the region. This data will be obtained using the Python Geocoder package.
3. Venue data for these neighborhoods, particularly focusing on Pizza Place venue data. This is required in order to cluster the neighborhoods. This venue data will be obtained using Foursquare API, providing many venue data categories for the associated neighborhoods.

Methodology

Results

Discussion

Conclusion