IBM Data Science

# CAPSTONE PROJECT

Opening a Pizza Place in Oakland, CA

Charlotte Barelle
8-18-2020

# 1. 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. 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?".

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.

# 2. 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.

# 3. Methodology

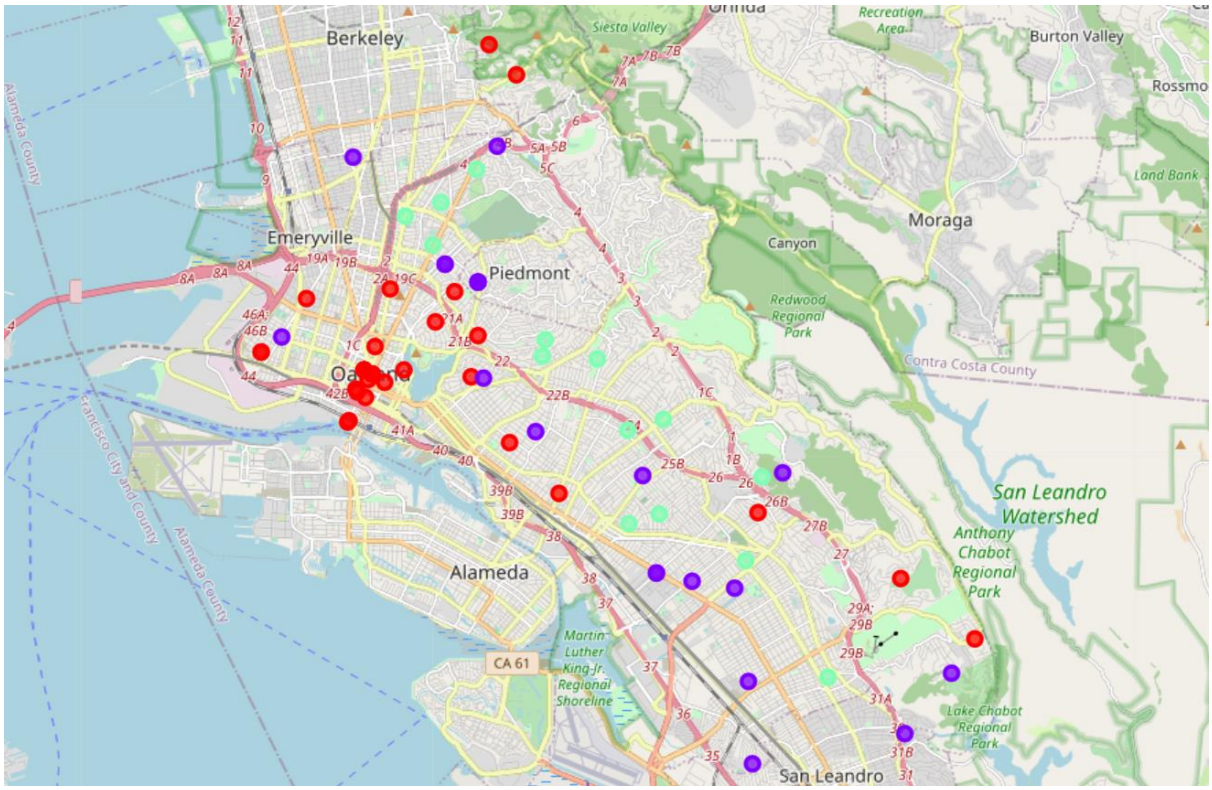
Initially, the list of neighborhoods in Oakland, CA was obtained from the Wikipedia page https://en.wikipedia,org/wiki/Category:Neighborhoods_in_Oakland,_California. The data was extracted through web scraping using Python requests and BeautifulSoup packages and converted into a Pandas DataFrame. Next, the geographical coordinates of these neighborhoods were obtained using the Geocoder package. The respective latitude and longitude values of each neighborhood were added to the DataFrame. Following this, a map of Oakland with the relevant neighborhoods superimposed on top was created using the Folium package.

Next, Foursquare API was used to make calls passing in the neighborhood geographical coordinates. Foursquare returned the relevant venue data in JSON format. The returned venues were converted into a Pandas DataFrame. The rows of this DataFrame were grouped by neighborhood and by taking the mean of the frequency of occurrence of each venue category. The DataFrame was then filtered for "Pizza Place" data entries only.

Next, a k-means clustering algorithm was employed in order to cluster the neighborhoods into k=3 clusters based upon their frequency of occurrence of Pizza Places. The resulting clusters were visualised using a Folium map and the results analysed.

## 4. Results

The Oakland neighbourhoods were successfully categorized into 3 clusters based upon the frequency of occurrence of Pizza Places in the respective neighborhoods using k-means clustering. The results are depicted in the map below where Cluster 1 is shown in red, Cluster 2 in purple, and Cluster 3 in green. Cluster 1 is situated in the city centre of Oakland and has a low concentration of Pizza Places with some neighborhoods having zero Pizza Places. Clusters 2 and 3 are situated around the outskirts of the city centre with Cluster 2 having an intermediate concentration of Pizza Places and Cluster 3 having a high concentration of Pizza Places.




## 5. Discussion

The relevant data was successfully acquired and analysed. Clustering by k-means enabled the thorough analysis of the neighborhood and respective venue data in order to obtain key observations. The most saturated neighborhoods fall into Cluster 3 where there exists a high abundance of Pizza Places. Therefore, due to high levels of competition with other Pizza Places, this project would advise respective businesspeople to avoid these areas when considering opening a new Pizza Place. Neighborhoods in Cluster 1 represent potentially good locations for a new Pizza Place due to the lack of competition from other Pizza Places due to the very low concentration of Pizza Places in this cluster. However, being in the dense city centre, the business may see threat from other food places

which may exist in a higher concentration in these neighbourhoods. Further analysis would be required to determine the saturation of other food places in this area in order to determine if these neighbourhoods are suitable for a new Pizza Place's success. Neighbourhoods in Cluster 2 may be the optimum locations for a new Pizza place as they are located away from the busy city centre typically saturated with other food places, and also only have moderate concentrations of Pizza Places. Thus, providing the opportunity for a new Pizza Place to be successful. Therefore, this project would recommend neighbourhoods in Cluster 2 as the optimum areas to open a new Pizza Place in Oakland.

As mentioned above, further analysis may be recommended in order to gain inside into other aspects of the problem. This may include competition from other types of venues such as other restaurants and cafes. This is out with the scope of this project.

# 6. Conclusion

This project successfully demonstrated the clustering of Oakland neighborhoods into 3 distinct clusters based upon the frequency of occurrence of Pizza Places in the respective neighborhoods. This allowed insights to be drawn from the observations gathered in order to propose an answer the 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?". The proposed solution is to recommend the opening of a new Pizza Place in Oakland to be in a neighbourhood which falls into Cluster 1. The observations and recommendations of this project can help future businesspeople make an informed and insightful decision to avoid competition when opening a new Pizza Place in Oakland.