Coursera Capstone

IBM Applied Data Science

Optimal Locations to Open a Bakery Food Chain Business

Akshit Wankhede

August 2021

Introduction

In this modern world, bakery products can be considered as essential food items. Most importantly, bakery products are a symbol of happiness, cheerful events, and they fill your important days. Not just birthdays, but we celebrate almost every special day with a cake, be it anniversary, or Christmas Day.

As an excellent baker, a person wants to start her own bakery. But her goal is to get her baking skills noticed as quickly as possible. So she wants to take advantage of data science to help her achieve that in an efficient way. That is, to start her first shop in an optimal location.

In this project, we will use various data science tools and a dataset, and provide a solution to the bakers requirements.

The target audience could be anyone who wants to start a food business, as this analysis does not confine to a particular food product. The analysis results can be also used in other food categories such as cafes, restaurants, coffee shops as well as grocery marts.

Problem Statement:

A professional baker wants to open a bakery chain. But she doesn't know where to start. She wants to start from her hometown *(Nagpur, IN)*. But her hometown is a big city. She requires knowledge as to where her bakery would get recognition quickly, and she could make best profits. Where should she start from?

Data

All the neighbourhoods, and coordinates of *Nagpur*. *Nagpur* is the second capital of the state of *Maharashtra* in *India*. *Mumbai* is the first capital in this state. This data can be scraped through the internet. *Wikipedia* has a page with this data. Note that instead of "neighborhood/borough", we use the term "locality" in India.

https://en.wikipedia.org/wiki/List_of_localities_in_Nagpur

We can get the location coordinates using a geocoder service. After gathering the coordinates we will clean the data and then use foursquare to gather more information about the cleaned data.

Nearby venues, their ratings and reviews, as well as pictures can be helpful in predicting the best possible locations to set up our business. We will use *foursquare* Apis to fetch this data. *Foursquare* will provide categorical data which we can use to our advantage.

For example, we are interested in locations where there are:

1. Colleges and institution
2. Offices and organisations
3. Tourist spots

More often, the age group would be young adults and working professionals. So using the foursquare places data, we find neighborhoods with colleges and offices to target more customers to our shop. We can find this information by analyzing the neighborhoods and the places that are in it.

Once we gather this, we will use some machine learning techniques such as *k-means clustering*. Using this algorithm we can find the coordinates of optimal locations.

One we find the coordinates we will visualise it on the map using a data visualisation library such as *folium*.