IBM Applied Data Science – Final Project Submission

Use of Economic Indicators and Population Density to optimize franchise location

Prepared By: Stephan G. Maher

_____

Part 1 – Problem definition and Data Sources

1. Question to be answered

Can we use economic indicators, population density, and information on competitors for each zip-code, to find an optimal location for our franchise?

1. Problem Background

My children work for a local delicatessen. Besides providing locally sourced cold- cuts, the deli also makes sandwiches and hoagies and either a soup or hot sandwich selection ranging daily from hot roast beef, pork, or meatballs. If I were to secure franchise rights, one of the first problems would be where would I locate the franchise? Before creating a financial profile, I would like to get the economic outlook and demographic data to see if I can optimize a location. The original store is in the Roxborough section of Philadelphia, specifically the "19128" zip code. I am also located in the same zip code.

2. Uses for this data

The problem to find a new location affects everyone. The real estate industry compiles data from the Census, IRS, and other sources to match home buyers with the best options. Recently Amazon accepted bids from locations all over the USA to open a second major hub. The small business association has statistics about the overall business outlook. Many factors can go into the decision for the optimal location, depending on the industry or personal preferences. While searching for data, I noted that small businesses make up over 95% of all U.S. businesses and 47% of U.S. annual GDP. Only about 40% of these businesses are profitable. By applying data science, I hope to find the optimal location for a franchise. This would be a good starting point for the business plan.

3. Description of the Data

For this evaluation, I would like to keep the commute within 20 miles, but far enough away from the current location so as not to be in competition. Some neighborhoods while close, are separate communities that would not directly compete with the target store. There are 3 suburbs of Philadelphia which fall within my initial radius. These are Montgomery, Chester, and Delaware counties. I will probably discount many Philadelphia zip codes due to the number of restaurants per zip code, but I will need to review population data before doing so. I'm also excluded New Jersey, which is within the 20-mile radius, due to the fact that some data is State specific. I would like to get information about the economic stability for each zip code, as well as rental cost information. Also, I would like information about competing stores, and population density.

I was able to find the following information by zip codes

- Zip Code with Spatial Latitude and Longitude per zip
- Population and population density by zip code
- Personal financial information by zip code.
- Residential housing and rent pricing by zip code – From this historic Data I computed a one and three year growth by zip code.
- As part of the IRS Data, I was able to find information on small business income. This is included with the Personal Financial information

4. Data and Sources

Location Data https://simplemaps.com/static/data/us-zips/1.73/basic/simplemaps_uszips_basicv1.73.zip This site contains a free table of US Zip codes in excel. The free version comes with Geo Spatial data by zip code, population and population density information. I found 180 zip codes with their spatial reference in the four-county area, and 138 of those exist within a "20 mile" radius of the initial store zip code of 19128.

The next data set comes from the IRS. This data is made up of financial demographics by zip code for 2018. The IRS data was found at the following site SOI Tax Stats - Individual Income Tax Statistics - 2018 ZIP Code Data (SOI) | Internal Revenue Service (irs.gov) The most current year's data is 2018.

Census and Small Business Association data relies on NAICS codes. The specific code for a delicatessen comes from the following site, NAICS Code: 722513 Limited-Service Restaurants | NAICS Association. The basic 2-digit code for food services is 72. I was only able to find limited information by zip code and industry. There are two tables, one that matches % of the number employees per job per zip code and another that provides business statistical profiles by NAICS codes. I created a list of the counts of all food services then 2 more lists food services with less than 20 employees and specifically the delicatessen category for deli's with less than 20 people.

The last dataset that I imported was from Zillow. I have historical data on the median house price from Zillow from 1996 to the present. I computed the upward and downward trends for 1 year and 3 years prior to 2018, to matvh the census data. Zillow data can be found here. Housing Data - Zillow Research.

I searched all day today to find a geo-json file for the counties in PA that has the county and zip code layers. I was able to find state and county with FIPS code. I reviewed the original zip code file that I had downloaded. According to week 5 forum, this course is being revised and we do not need to

use Foursquare data. Had I not already invested so much time, I would have revised my project. In any event, I will attempt to find the best place to relocate using the data that I currently have. It just may not be possible to visually demonstrate the results on a map.