Battle of the Electronics Stores

Project submitted in partial fulfillment of the requirements for the certificate

of

Applied Data Science

by

Jose E. Jobel

Under the supervision of

Alex Aklson

FEB 2020

Chapter 1

Introduction

1.1 Background

The growth of cryptocurrency brought along new ways for individuals to make passive income. The idea of 'mining' cryptocurrency allows individuals to solve complex algorithms on their computers and are rewarded cryptocurrency for solving the algorithm, which can be exchanged to an equivalent amount of US dollars. The manner in which these algorithms are solved has changed drastically over the years where in the early stages of crypto mining anyone with a laptop was able to run these algorithms. Nowadays, with the advancement of computer graphic cards along with Application Specific Integrated Circuit (ASIC) equipment the use of laptops are not enough to successfully mine cryptocurrency. In 2017, we saw a rise in demand for graphics cards from two dominant groups, online gamers and crypto miners. This brought about a shortage in graphic cards supply which has extended into the present year, 2021. Many online retail stores have tried to implement strategies to try and tackle this issue by restricting one graphics card per household. In many instances this works, unfortunately an introduction of automated buying algorithms also known as 'bots' have made buying computer equipment much more difficult for consumers. It has become a race between consumers and bots to see who can buy an item the quickest. As with the one graphics card per household strategy, a few online retailers have put in place strict measures to prevent the use of bots on their web pages but this is not permanent or effective.

The demand for graphics cards along with many other hardware has made its voice heard through many frustrated online users many of whom express a frustration toward a lack of in-store supplies in their local electronics stores. What is interesting about this topic is that currently there are very few local computer electronics stores in the city of Los Angeles who offer products for online gamers and computer enthusiasts. Therefore, it is advantageous for stakeholders to identify what neighborhoods in the city of Los Angeles have electronics stores and which neighborhoods do not, such that opening an computer electronics store in an optimal location can meet the demands of many local consumers.

1.2 Problem

This project is geared toward stakeholders particularly interested in identifying an optimal location for a **computer electronics store** business in Los Angeles, California. The business is aimed at computer enthusiasts and professional online gamers, thus providing a large variety of computers, computer parts, electronics, software, and gaming supplies.

We will focus our attention on locations which **do not have an electronics store in the area and are not primarily residential neighborhoods**. We would also prefer to consider locations which **contain malls, shopping centers, and/or a large population of retail stores in the area**.

Five locations will then be recommended based on how close they satisfies the above criteria along with a description of the advantages of choosing a specific location versus other candidates.

Chapter 2

Data Acquisition and Cleaning

2.1 Data Sources

The names of neighborhoods and regions within the Los Angeles County were obtained through web scraping the Los Angeles Times' Mapping LA project at Mapping LA Project using the Beautiful Soup library. This data has been used previously by the City of Los Angeles Open Data portal to strictly map out neighborhoods exclusively in the city of Los Angeles, which we use as well. We obtain the longitude and latitude coordinates of each neighborhood using the geopy library.

Once we have the basic name and geographical coordinates of our neighborhoods, we leverage the Foursquare API to obtain venues that fall within their Shop & Service category. This means that common venues such as restaurants, neighborhoods, parks, schools, and churches will not be returned when we query each neighborhood. Please see LA Open Portal Data for more information about Foursquare categories.

2.2 Data Cleaning

The data scraped online includes areas that range from mountains to nearby cities. To remedy this issue I downloaded the dataset provided by the City of Los Angeles Open Data portal to remove 158 neighborhood or about 58% of the initial data. This left me with 114 relevant locations. The table (*df_la*) contains neighborhoods strictly in the city of Los Angeles and is organized such that neighborhoods with common regions are grouped together. This helps understand and verify what locations are being considered.

The longitude and latitude features were then added to the *df_la* dataframe using the *geopy* Python library. Once the geographical coordinates of each neighborhood were appended to *df_la*, I proceeded to use the Foursquare 'explore' API endpoint such that I get back a new dataframe called *la_venues*. Using the 'explore' API endpoint helped further remove residential neighborhoods along with locations without venues in the surrounding area. This process reduced the dataset by about 8%, leaving us with a total of 105 potential locations.

Generally, electronics stores do not offer in-store computer parts or hardware in their stores but do offer them on their websites for customers to pickup in-store or have the parts delivered to them. I will take a conservative approach toward the analysis and assume that all electronics stores pose a competition to our stakeholder's business. This in return allows us to remove neighborhoods that host electronics stores from our data and focus on recommending locations which do not have them. With this in mind, I went ahead and created a dataframe called *elec_neigh* which contains precisely 33 neighborhoods with electronic stores.

Finally, we derive the *la_clean* dataframe that provides information on the remaining 72 neighborhoods that do not have electronics stores. The information in this dataframe will be used to make our recommendation in our analysis.

Chapter 3

Methodology

The focus of this project is to recommend neighborhoods that host a high density of retail business that fall into the Foursquare Shop & Service category and do not have electronics stores in the surrounding area. We limit our search to a 500m radius around each neighborhood.

To start we collect our data through two methods. First we scrap neighborhood names online using Beautiful Soup and read in the parsed data into a Pandas dataframe. Secondly, we download a publicly available CVS file containing neighborhood names belonging to the city of Los Angeles. Geographical coordinates are then adding using the geopy library. With the coordinates of each neighborhood readily available in our dataframe we use the Foursquare API to continue building on top of our data by returning and appending all venues found in a 500m radius around a neighborhood. All returned venues fall within Foursquare's Shop & Service category.

Once the data is collected we move on to exploring the data by separating our initial dataframe into three separate dataframes namely, *la_venues*, *elec_neigh*, and *la_clean*. Each dataframe will be used to visually represent their corresponding neighborhoods on a map. Using the *la_clean* dataframe we will determine which neighborhoods have the largest density of venues by taking into consideration the category of businesses in those neighborhoods.

Ultimately we will create clusters of locations using k-means clustering to identify neighborhoods with common venue types and optimal location recommendations for the stakeholder's business.

3.1 Data Analysis

We start the analysis by mapping out the locations of all the neighborhoods along with a separate map displaying the locations of all 3531 venues in *la_venues*. Next we determine how many venues are in each neighborhood and the average number of venues. We then use this information to create a bar chart to display neighborhoods with number of venues greater than the average. By doing this, we refine our search to neighborhoods with large numbers of retail stores in their vicinity. We repeat the above process for *elec_neigh* with the exception of understanding on average how many electronics stores are in a given neighborhood.

To get a clear idea of potential locations to recommend, we first display the venues in *la_clean* on a map and provide basic descriptive statistics for our results. As a final step, we group neighborhood into 5 separate clusters based on similar venues. From those clusters, we will recommend locations with their neighborhood name, region, and which cluster they belong to as results of our analysis.

This conclude our analysis. We created 5 clusters representing neighborhoods with similar venues and containing no electronics stores. In our analysis, we searched for venues using a 500m radius around each neighborhood but it is important to keep in mind venues outside of this range may exist including electronics stores not recognized by Foursquare. We note that about 79% of the neighborhoods in *la_clean* belong to *cluster_2* for the reason that they share similar venues such as clothing stores, furniture/home stores, department stores, and convenience stores just to name a few.

Chapter 4

Results and Discussion

Our analysis shows that roughly 79% of neighborhoods fall under the *cluster_2* group making it the only cluster with the largest number of neighborhood membership. Locations in this cluster share similar venues such as 'Business Services', 'Photography Studios', 'Home Services', and 'Furniture/Home Store' are examples of such venues. We'd also like to focus our attention to locations strictly in *cluster_2* since many venues compliment a computer electronics store compared to venues in all other clusters. From these, we will present five optimal locations for a computer electronics store business and at the same time provide reasons why we chose each of them.

The first location is Venice. Venice is one of the most popular beaches in Southern California and hosts roughly 30,000 people daily according to laparks.org. This neighborhood is a top choice on our lists not only for the 10 million visitors it receives per year but rather for the 97 Shop & Service venues that compliment our stakeholder's business. Venues in several categories such as home services, men's/women's stores, and bookstores all offer services that many computer enthusiasts, online gamers , and the general public benefit from.

The second location we recommend is Chinatown. Well known as a commercial center for its Asian businesses, Chinatown sits roughly 1.6 miles North of Downtown Los Angeles and is home to 94 Shop & Service venues. One of the main benefits of this neighborhood is the Dynasty Shopping Mall and the Chinatown Saigon Shopping Mall both of which are situated around a large population of retail stores. This location offers a great opportunity to attract additional customers outside of our stakeholder's targeted audience.

The third and final location we recommend is Panorama City. This locations hosts 75 Shop & Service venues including a shopping mall and a shopping center. Panorama City having been one of the largest major retail store outlets in the San Fernando Valley region, it also offers one of the youngest age range populations. The median age in Panorama City is approximately 34 years of age censusreporter.org. According to a statista.com 2020 survey, roughly 38% of online gamers fall into the 18-34 age range. Thus Panorama City is an optimal location that offers a young audience to cater a computer electronics store along with its reputation as a major retail store outlet.

Each location was chosen carefully by taking into account ability to attract an audience outside of our targeted customers through the attention it would provide our stakeholder's business by means of its popularity and the number of retail stores in the area. As another requirement we decided to take into account the number of venues in the area and whether a computer electronics store would thrive in such an area. Ultimately, we decided Venice was our top choice followed by Chinatown and Panorama City. While the locations we recommended are our top choices there are of course many other locations in which a computer electronics store may thrive in by taking into account the number of electronics stores previously established in each neighborhood, what specific addresses within a neighborhood are optimal locations, or what category of venues most compliments a computer electronics store. These are only a few directions we could consider during future follow ups.

Chapter 5

Conclusion

In this project we identified three optimal locations for a computer electronics store in the city of Los Angeles, CA. We identified a total of 3535 venues within 105 neighborhoods using the Foursquare API. During this processes we remove 33 neighborhoods that contained electronics stores and used the remaining 72 neighborhoods for our recommendation. We used k-means clustering to group similar locations based on common venues. The decision process made use of the candidate location's population of venues, location popularity, and the category of venues offered. From this we recommended to our stakeholders three optimal neighborhoods: Venice, Chinatown, and Panorama City.