

Market Sizing – Identifying potential customers in Gwinnet County

Jude Omozegie

December 26, 2019

1. Introduction

1.1. Background

Zambezi Consult is a start-up consulting company with experience in technology, brand management, retail and channel marketing and go-to-market strategy and planning. Zambezi Consults helps small and medium (SMB) sized companies reduce their time to market and shorten their learning curve by providing expert advise on how to grow their business with ROI, strategic marketing services that deliver returns for their brands, business and data analytics that wold provide insights. Zambezi Consult also helps provide access to emerging markets in Africa. We do this by providing built-for-purpose services to existing companies, based on their industry sector, product category and business objectives.

Zambezi Consult plans to expand its services into the US, with a pilot in Gwinnet County, Georgia. Each year, new companies are established in the county. To ensure they grow their businesses and deliver on returns, they need expert advice and a shorter learning curve. Zambezi Consults helps by allowing these new companies leverage their expertise and domain knowledge. We also offer our service to existing businesses, who look for opportunities to increase consumer traction, increased sales and market share. A number of these business will not survive in the long term without adequate guidance and assistance.

1.2. Problem

As a pilot, it is important to understand the type and locations of businesses in Gwinnet, where and how they are clustered and what unique feature influences their co-location. This project aims to predict which locations are suitable for new businesses to establish in, cross sales opportunities to leverage and market data to assist new and existing companies in marketing and growing their brands.

Interest

Obviously, new businesses would be very interested in accurate prediction of suitable locations to establish their businesses, to leverage unique opportunities existing in those locations. Existing companies would be interested in clusters they can leverage to cross sell their products.

2. Data acquisition and cleaning

2.1. Data sources

Database of small businesses in Gwinnet can be found in multiple sources. However, we are using data from the U.S. Small Business Administration (SBA) database. The data available on this site includes the Name/Trade name of the business entity, Contact Information, address and a synopsis on the business describing the nature of the business. However, not all the business has their description provided. Some are blank (not provided) while some do not provide definite detail on their nature of business. GPS Coordinates obtained using Geocode By Awesome Addon on Google

2.2. Data cleaning

Data downloaded or scraped from SBA was downloaded into a table. A total of 273 SMB contact information was downloaded. The table had the following rows: Name of Firm; Trade Name; Contact; Address; City, Zip; Business Type (Service, Construction, etc); Capabilities Narrative. There were lots of missing values and poor formats that needed to be corrected. The Trade Name column had a lot of missing data. As the Name of Firm column was available, there was no need for this column in the analysis, so it was removed. The Capabilities Narrative column was wordy. It would be difficult to manually read, edit and categorize the over 273 rows of data. So, for the purpose of this analysis, I decided to drop this column and use only the Business Type (Service, Construction, etc) column for analysis. The Contact column was also removed, as it was unnecessary for the purpose of this analysis.

3. Methodology

We downloaded the list of small businesses in Gwinnett, as our dataset for this analysis. The data source is the US Small Business Administration www.sba.gov. As part of initial analysis, to obtain the latitude and longitude of the locations, we had used the Geocode By Awesome Addon on Google sheets, to resolve the coordinates of the street addresses for each small business location. This turned out easier and less cumbersome than using Geopy or Gocoder. Several challenges were faced while trying to obtain the coordinates using Geopy or Geocode on Python. I then decided to use the easier option available.

The needed file were uploaded and loaded into pandas dataframe using ExcelWriter. We verified the dataframe for missing or NaN values, and made necessary clean-up of the dataset where needed.

The dataset has columns with data that are unnecessary for our analysis, such as the 'Capabilities Narrative' column. It also has multiple index entries. These will be removed. The 'Business Type (Service, Construction, etc)' column name is too long for analysis. This will be reduced to 'Business Type'. The dataset also includes columns with missing or NaN values. Rows with missing or NaN values will be dropped.

We finally produced a dataset that had relevant data necessary for our analysis.

4. Result

The purpose of this analysis is to show that there are sufficient small businesses situated in Gwinnett and that they are well dispersed within the county. Being able to categorize, identify their locations/coordinates and visualize these on the map, is a good start towards kicking off the business. Doing further analysis on this dataset, to identify their suitability and relevance to new prospects we consult for, was not considered as part of this project. This map make is easier for the team to easily identify and locate businesses to engage with.

5. Conclusion

My analysis shows that there are significant number of new businesses situated in Gwinnett county to start a pilot program. After identifying and classifying each business location, we see from the map that they are there are significant opportunities available to commence the pilot.

However, further analysis need to be done to ensure the businesses are in need of our services and currently do not use competing solutions.

