HELPING ENTREPRENEURS CHOOSE BUSINESS STARTUPS BASED ON MOST POPULAR VENUES VIA FOURSQUARE

MUHAMMAD HAMZA ALI KHAN

JULY 4TH, 2021

1. Introduction

In these times, one cannot solely rely on working a white collar job from 9 to 5. These were the old times, where the pay scales were good and sufficient enough for a person to run his/her family. Now, to live a satisfied one must own a business, not only to make ends meet but to enjoy life with family, friends etc. However, starting a business is not a piece of cake. You need to do a lot of research like what type of business, what target market, who to target, thousands of survey and what not.

So in this project, we will try to find an optimal business idea. Specifically, this report will be targeted to stakeholders interested in opening business based on the **most visits** in New York City, USA. Since the area of NYC is really big, therefore, we would focus on Staten Island and its most visited venues in each neighborhood. We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible business category can be chosen by stakeholders.

2. Data

Based on definition of our problem, factors that will influence our decission are:
- number of existing businesses in the neighborhood (any type of business)

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- I will be using the New York City data provided in the previous weeks to generate a map as long and lat are already present there.

- number of businesses and their types and location in every neighborhood will be obtained using **Foursquare API**

- coordinates of NYC will be obtained using **Google Maps API geocoding**

Of course, the data that we have gathered will be cleaned with unwanted data thoroughly removed from the data set to ensure accuracy of the results and avoid any confusions. We have

defined a function to segregate our data into venues and categories along with the longitude and latitude which will be used for clustering and plotting of our map.

Below is a sample of what our data will look like after segregation:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	St. George	40.644982	-74.079353	A&S Pizzeria	40.643940	-74.077626	Pizza Place
1	St. George	40.644982	-74.079353	Beso	40.643306	-74.076508	Tapas Restaurant
2	St. George	40.644982	-74.079353	Shake Shack	40.643660	-74.075891	Burger Joint
3	St. George	40.644982	-74.079353	The Gavel Grill	40.642157	-74.076674	American Restaurant
4	St. George	40.644982	-74.079353	Staten Island September 11 Memorial	40.646767	-74.076510	Monument / Landmark

Please note that I have selected Staten Island as my preferred borough of NYC for setting up a business for illustration purposes only, and this project can select any borough or any other city in the USA.

3. Methodology

In this project we will direct our efforts on detecting what kind of businesses or attractions are most popular among visitors.

In first step we have collected the required **data: location and type (category) of every attraction/venue Staten Island**.

Second and final step in our analysis will be exploration and categorization of these venues **' across different neighborhoods of Staten Island - we will use** clustering** to categorize the venues to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4. Analysis

Lets look at our data after applying our function on all the data set. You can see in the table below how I have categorized the top 5 venues in every neighborhood in Staten Island.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Annadale	Pizza Place	Diner	Sushi Restaurant	Park	Pharmacy
Arden Heights	Deli / Bodega	Coffee Shop	Bus Stop	Home Service	Pharmacy
Arlington	Deli / Bodega	Boat or Ferry	Bus Stop	Grocery Store	Event Space
Arrochar	Pizza Place	Italian Restaurant	Bus Stop	Pharmacy	Supermarket
Bay Terrace	Italian Restaurant	Supermarket	Donut Shop	Insurance Office	Train Station

5. Result

From our analysis we have categorized the venues in 5 clusters which shows Top 10 venues in every venue.

6. Conclusion

Purpose of this project was for stakeholders/entrepreneurs to identify the top venues in every neighborhoods which can help them in choosing the best options for their business interests.

Final decision on optimal business idea will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.