New Café Location Analysis

Investment Regions Analysis - Capstone Project Final

Isaac Injeti – January 14th, 2019

Abstract:
Investors have been investing in various cafes in the area based on projected growth in the city. Unfortunately, a number of those ventures have failed because of their locations. As we know, the key to success is typically location, location, and location. In order to reduce the chance of failure again, investors are wanting to understand what areas of the city would have the most potential for the success of a new café venture based on the characteristics of the area in which the top cafe currently operates in the city.

Audience: Cafe Investors seeking to capitalize on the recent growth of the Vizag (Visakhapatnam) city region within Andhra Pradesh, India.

Introduction and Assumptions:
A major factor for the success of a venture such as a café is the volume of people moving through the area in which the café is located and the demographic of that area. Often there is a direct correlation between the venues or businesses that are in an area and the demographic of the area. An example might be industrial parks may have more business venues near by than recreational, etc. This correlation is outside of this analysis but is taken as an assumption to better understand various areas in the city and find recommendations for investors based on the existing café ventures.

The Data:

- Extract of Areas (neighborhoods) of Vizag from Wikipedia: http://bit.ly/vskp_wikipedia
- Extract of zip codes from: http://bit.ly/vskp_pincodes
- geocode extract of coordinates
- Foursquare exploration data

The multiple datasets will be combined into a consumable framework from which areas will be clustered using the k-means algorithm based on types of venues such as businesses, restaurants, etc within a 10 km radius of the city hub. This will then be visualized using a folium map to show the areas that are most similar to the area in which the top cafe is located.

Methodology:
The analysis began with some online searching to find a website that would contain the names for the various areas of our target region, Visakhapatnam. After trying a few different sites, the Wikipedia page for Neighborhoods of Visakhapatnam had an alphabetical list of area names. The area names were then input into the geocode function to query latitude and longitude coordinates for each area using OpenStreetMap's Nominatim.

Once the coordinates were found, any areas that did not return coordinates were dropped and the remaining data was plotted to a Folium map for visualization. When reviewing the map and areas returned, it was found that there were many inaccuracies in the area names and the points where they were plotted. In light of these issues, after further troubleshooting, an alternative approach using Zip Codes (pins) of the target region was used instead. All work is shown in the included Jupyter Notebook.

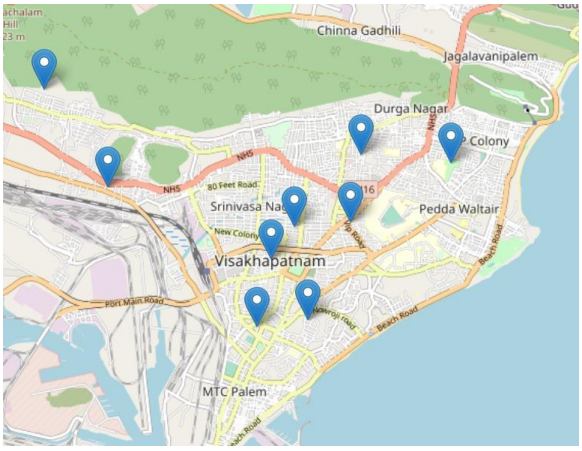

Figure 1 - Folium map of Vizag Areas

After consolidating all the data, an area radius of 10 km from the primary city postal code was determined and used to query venue locations from Foursquare. Having all venues and categories

captured from Foursquare for each area, a OneHot Encoded numeric dataset was created and then prepared for clustering based on the mean of the grouped venues per area.

Algorithm:
K-means clustering was chosen to cluster the different areas together based on the venue category data. The k-means algorithm was set to the following parameters which produced the best results: init='k-means++', n_clusters=10, and n_init=15. The final results of the seven clustered labels were then mapped by color onto the Folium map and also merged into the original dataset.

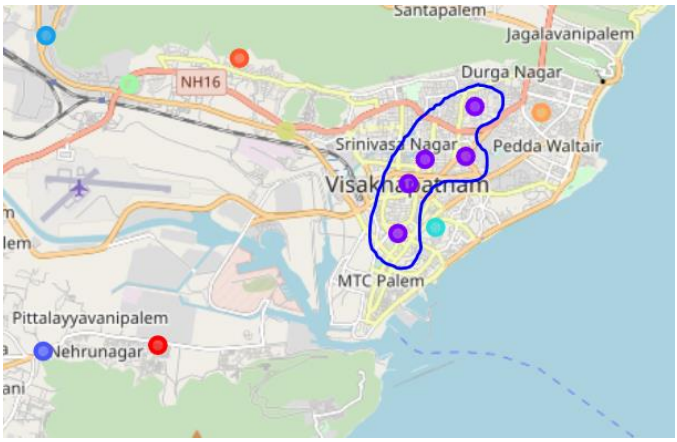

Figure 2 - Clustered Areas by Venues

Discussing the Results:
The outcome of this analysis proved to be very insightful in understanding not only what areas of Visakhapatnam were in scope for consideration but also understanding how the areas were alike regarding venue categories. Though the study focused specifically on new café ventures locations, the methodology and framework of this study could serve multiple other objectives such as understanding demographics of the areas, identifying contrasting characteristics of areas, or even creating a model for identifying characteristics that contribute to successful businesses operating in the region.

Regarding the objectives of this project, the following observations stood out as most meaningful:

1) As assumed, the already existing cafes were in similar areas and were within the same cluster as the Top Cafe
2) 'Breakfast Spot', 'Convenience Store', 'Fabric Shop', 'Snack Place' were the categories of venues most commonly associated to the Top Café cluster (see Figure 3)
3) Akkayyapalem, P&T Colony, and Visakhapatnam (central) are the areas in which investors may have greater success based on similarity to the area in which the Top Café was located
4) The Top Café is *Café Coffee Day- The Lounge* and is located in the A.U. Eng College area

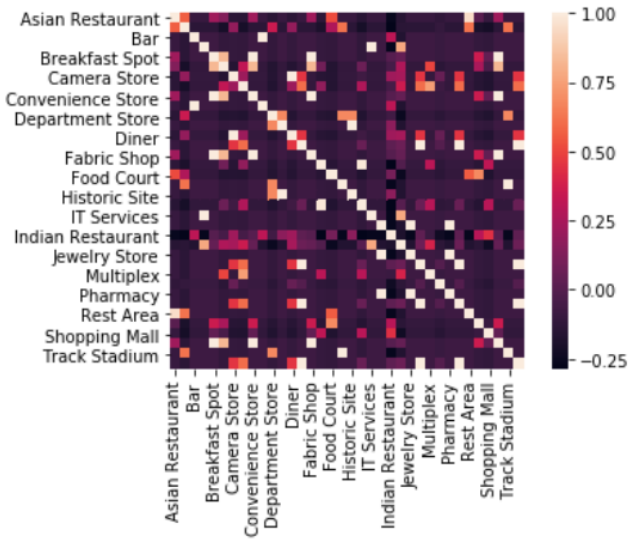

Figure 3 - Venue Correlation Map

Conclusion:
In conclusion, based on the analysis and results described above, given the success of *Café Coffee Day- The Lounge* in the A.U. Eng College area, investors should consider investing in either Akkayyapalem, P&T Colony, or Visakhapatnam (central). An additional insight to consider in finding the optimal region would be to consider if any of the following categories of businesses are found in the target area: 'Breakfast Spot', 'Convenience Store', 'Fabric Shop', or 'Snack Place'.

Link to Notebook:

https://github.com/iamzic/DS-Capstone/blob/master/Investment%20Regions%20Analysis%20-%20Capstone%20Project%20Final%20-%20Injeti.ipynb