**Clustering of Restaurants in Downtown Atlanta**

Arjun Thiruppathi

November 3, 2020

1. **Business Problem**

Downtown Atlanta is home to a wide range of restaurants and bars, from Southern Soul to Latin American. Geographically, the restaurants share similar business problems, and serve the same customers. By understanding the commonalities between the restaurants in the Downtown Atlanta area, through data provided by Foursquare, restaurants can better understand their customers and improve business metrics. A deeper understanding of the key metrics of success, described in this paper, enables restaurants to target weak areas in their business and adjust accordingly.

1. **Data**

The data used in this report was provided by the Places API provided by Foursquare. This API is well regarded as accurate and up to date, containing crowd-sourced information of many locations across the world. Quality data is vital to meaningful business insights, so I subscribed to Foursquares personal API plan, which allows for 950 Regular API calls per day. For the purpose of the project, I only used data points of rating number and number of photos uploaded by the target restaurants. I chose to explore the data point of number of photos uploaded by the restaurant because it is a factor that the restaurant can directly control, whereas a data point like number of likes is completely user controlled.

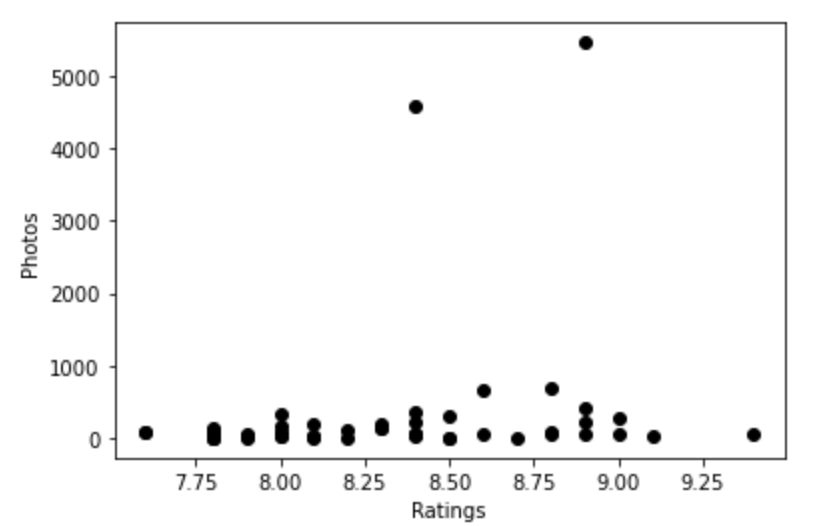
One limitation to the scope of data in this project is the limitation of the type of data a personal, Foursquare developer account can access. A premium developer subscription provides access to endpoints that would allow for deeper insights, such as venue statistics and venue links.

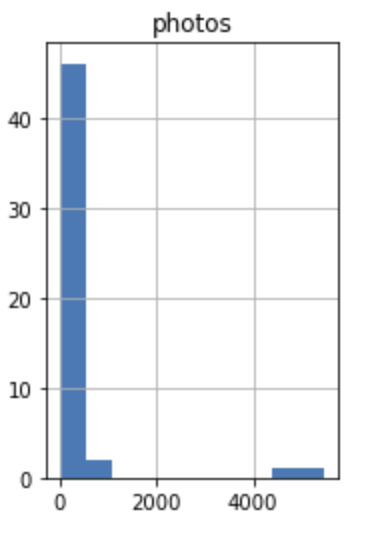
1. **Data Methodology**

I used the geopy package to get the latitude and longitude of Atlanta, GA, then used these coordinates to create a data frame of the 100 closest venues from the center of Downtown Atlanta. I filtered this initial data frame to exclude venues that are not restaurants and bars. The resulting data frame consisted of 53 datapoints, with a corresponding venue name, id, category, latitude, and longitude.

For the data pre-processing stage, I extracted the rating number (on a scale of 1 to 10) and the corresponding number of photos uploaded per restaurant. Then, I removed unnecessary columns, so the resulting data frame only consists of a rating number and number of photos; the name of the restaurant and id are unnecessary for the purpose of clustering data.

After normalizing the data, I plotted the relationship between rating and number of photos on a scatterplot, shown below.





There seems to be no clear relationship between the number of photos uploaded by the venue and the venue’s rating, on the Foursquare app. The number of photos uploaded is skewed by the two outliers, each with over 4000 photos uploaded.

For clustering, I used scikit-learn library to perform k-means clustering. I chose a k of 2, meaning I chose to group the data into 2 clusters. The normalized data frame produces an array with two feature sets, rating, and photos, that make it possible for the k-means clustering algorithm to run.

1. **Results**

The clustering algorithm produced two clusters of data, one consisting of the points with a low number of photos. The other cluster consists of the two outlier points, with a very high number of photos uploaded. The business applications of these results mean that restaurants do not need to spend the time and resources to upload hundreds of photos to their venue page. Uploading a high number of photos has no relationship to a high rating on the Foursquare app; there are other related factors that may or may not be tracked by Foursquare.

1. **Discussion**

The results show a lack of relationship between photos uploaded and rating. More data from premium endpoints would be useful in exploring more relationships, but that is beyond the scope of this project. Another algorithm, like a simple linear regression algorithm, would be useful in accounting for the outliers in data.

1. **Conclusion**

The results from this project show that restaurants in Downtown Atlanta do not need to upload many pictures to have a high rating.