1. Introduction

1.1 Background

The City of Cape Town is the largest metropolitan area in the Western Cape Province and falls apart of the Republic of South Africa. Cape Town is well known for many reasons internationally such as it's historical landmarks and museums, winelands, Fynbos ecosystem and the stunning Table Mountain.

Over the years, Cape Town and its city bowl in particular have become one of the main tourist attractions for international tourists to visit throughout the year.

On top of the large swarms of tourist, the city bowl is also the Central Business District (CBD) of the region which means large amounts of locals flocking to the CBD area for work. This allows for all sorts of small businesses such as telecommunications, restaurants and fast foods joints, informal markets and much more to take profit advantages of a high "people-traffic" area.

1.2 The problem:

After realizing the potential profit advantages of the Cape Town city bowl, a small business is looking to open a new coffee shop in the city bowl of Cape Town area.

They would like to determine the most ideal location to open their coffee shop and have three key requirements that need to be met when find the best location:

1. There must be **high foot traffic** in the vicinity of the shop to maximize number of potential buyers
2. There must be **few other coffee stores** to minimize competition
3. There must be **surrounding businesses** to further maximize foot traffic

These points are ordered according to the priority of the requirement being met.

This means that, for the purpose of this shop, high foot traffic is the most import factor to consider when determining where to locate the store, then the number of competitors, followed by the surrounding stores. What this means is that we are willing to endure a certain level of competitors for the chance of having higher foot traffic, and therefore, more customers.

One of the issues with minimizing other coffee stores in the area (mentioned in point 2 above) is that sometimes competitors actually bring people into the area and locating the store away from competitors might also move us away from potential high foot traffic areas. For this reason, we are willing to sacrifice having a low competition area for one that has high foot traffic, many surrounding businesses and does have a higher number of competitors.

This business also has no preference between targeting tourists and locals but has the intention of holding its market edge with being a low-cost coffee store. Generally, tourist travel with the expectation to spend money and do not necessarily have the same knowledge as locals in terms of what is cheap, averagely priced or expensive coffee. Also, they may gravitate to brands they know such as Settle Coffee, Starbucks or Mugg&Bean. For this reason, we will not be focusing on tourists but rather where the locals will be, for example, in the central business district who are more likely to return to a store that offers quality coffee at a fair price.

To sum it up, the ideal location for this store has high foot traffic, low competition and surrounding businesses that will attract more customers to our store, we will trade low competition area for one with higher foot traffic and surrounding businesses, and we are focusing on “local” areas instead of “tourist” areas.

1.3 Interest:

Obviously, the business mentioned above that is looking to open the store will be the most interested in this investigation. Others who may be interested include other individuals or companies that are also trying to open coffee stores or similar stores that have the same requirements location requirements as the business we are completing this analysis for.

2. Description of data and data sources

For this analysis we will be looking at several different datasets.

2.1 Transportation routes

We will use the Foursquare API to determine the transportation endpoints in the city. This will allow us to gage which areas are more likely to have more people moving about them. The assumption is that the more transportation routes feed into a specific area the more likely it is that there will be a high density of people moving from drop-off location to desired location.

We will use the following Foursquare categories for the API:

- **Parking** [4c38df4de52ce0d596b336e1]
- **Bus Station** [4bf58dd8d48988d1fe931735]
- **Bus Line** [4bf58dd8d48988d12b951735]
- **Bus Stop** [52f2ab2ebcbc57f1066b8b4f]
- **Light Rail Station** [4bf58dd8d48988d1fc931735]
- **Metro Station** [4bf58dd8d48988d1fd931735]
- **Train Station** [4bf58dd8d48988d129951735]
- **Transportation Service** [54541b70498ea6ccd0204bff]
- **Taxi Stand** [53fca564498e1a175f32528b]

2.2 Local coffee shops

After that, we will evaluate coffee shops in the city bowl to determine the high-density areas. Again, we will use the Foursquare API to pull in this information using the following categories:

- **Coffee Shop** [4bf58dd8d48988d1e0931735]
- **Café** [4bf58dd8d48988d16d941735]

This will allow us to assess locations with a high density of coffee stores and consider if it makes sense to be locate near the competition or far away. We will do this by evaluating what surrounds these stores and if area is too densely populate with coffee shops for our store to have a market edge.

For the purpose of this investigation, we will be focusing on stores that are considered exclusively coffee shops or cafés. While restaurants and fast food joints do generally provide coffee as one of the goods they sell, we will not be considering them for the investigation because their primary sale-point items are foods.

Nightlife locations such as bars will also not be taken into consideration as most of them are only open *after* our store's operational hours.

2.3 Surrounding businesses

Lastly, we will consider the surrounding businesses to get an idea of where the locals can be found. We will use the following categories to identify student and professional hotspots:

- **Professional & Other Places** [4d4b7105d754a06375d81259]
- **Office** [4bf58dd8d48988d124941735]
- **School** [4bf58dd8d48988d13b941735]
- **Shop & Service** [4d4b7105d754a06378d81259]

Using the datasets in points 1 - 3 we should be able to identity 1 or more location that closely resembles the ideal store location that has high foot traffic, low competition and surrounding businesses that will attract more customers to the store. As previously stated, we are willing to sacrifice a low competition area for one that has higher foot traffic and surrounding businesses.

Additional Note:

All data mentioned in points 1 - 3 will consist only of the name, category, location data and any other information from Foursquare that might aid our investigation.

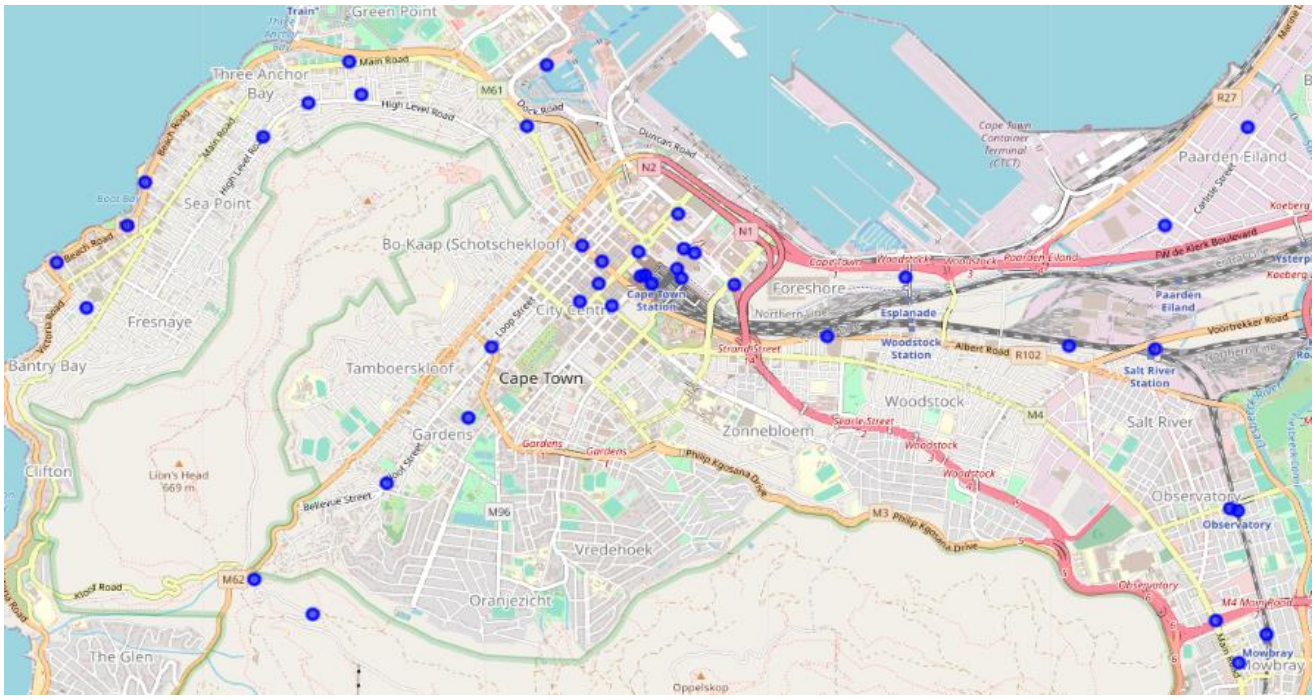
3. Data Analysis (Methodology)

3.1 Folium Mapping Analysis

3.1.1 Transport Routes

Using the Foursquare API to bring in the transport key-locations (such as stations and parking garages) and the Folium library again the below map was generated depicting the key-transport location in and around the City Bowl.

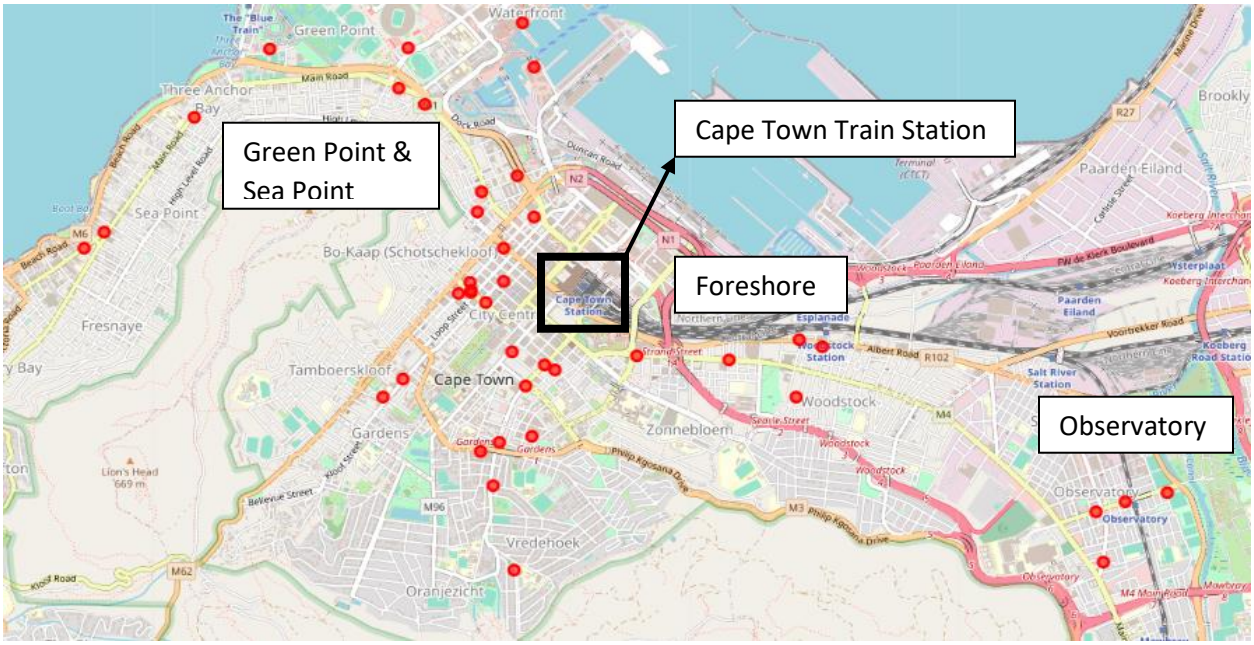
Looking at these data points we can see there are several potential clusters of data. For example, on the West side of the map we can see a fair chunk of transport locations in the Green Point-Sea Point area which is mostly considered locally a residential area for younger people and tourists. The is a large cluster in the centre of the CBD where many of the train lines, bus stations and taxi stations converge. Lastly, there is a fair cluster spread over the East side of the map – which is a mix of an industrial and local business area.



3.1.2 Other local coffee shops

Using the Foursquare API and Folium library in Python again we see that there is a large concentration of coffee shops dispersed throughout the city centre and with another smaller cluster in the Observatory (East) area and in the Green Point-Sea Point area (North to North-West).

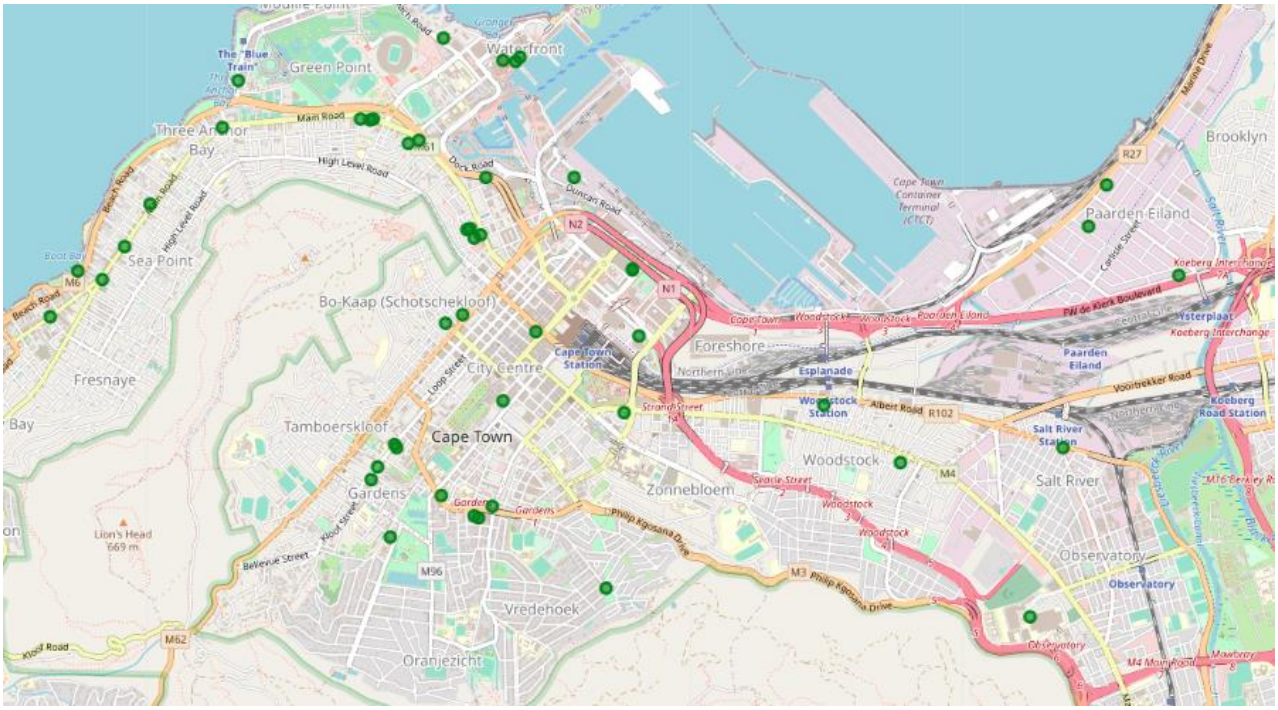
What is particularly interesting to note is that there are no coffee shops located at or near to the Cape Town Train Station (indicated with black box) and in the Foreshore region.



3.1.3 Surrounding businesses

Using the Foursquare API again to gather the data of the businesses in the area in Python. Again, we visual this data using the Folium library where the transport endpoints are depicted

It is interesting to take not of the scattered-ness of the businesses in this map and that majority of them are found along the main roads of the city.



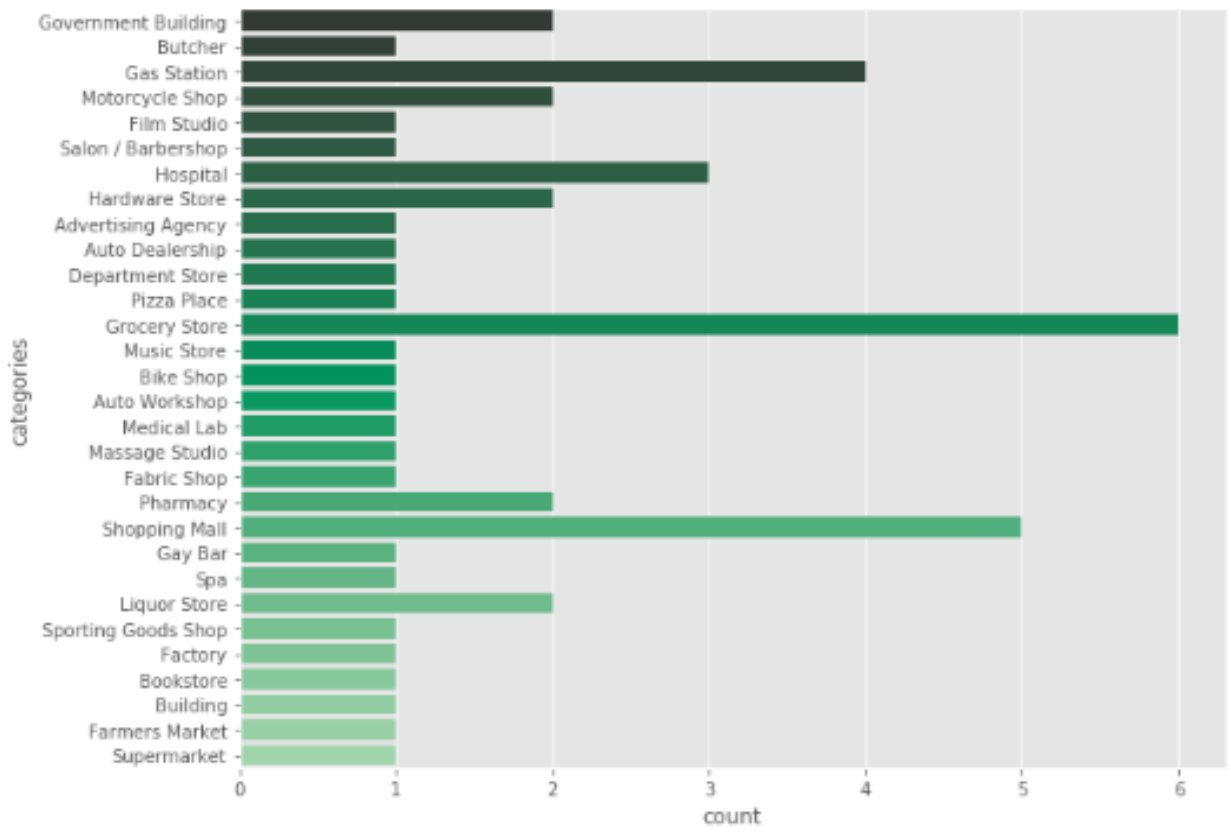
Additionally, in the below graph we can see the different categories of businesses classified in the above map.

The top 3 most commonly occurring categories include:

1. Grocery Stores
2. Shopping Malls
3. Gas Stations

These categories make sense when considering that, as previously mentioned, the North to North-West region of the map is considered predominately residential and the South is also considered a very residential area.

While residential areas can't typically be considered an ideal location for a coffee shop that requires high traffic of people, it may be beneficial to locate the store in or near a shopping mall or grocery store that will already attract people to the area.



3.2 Cluster analysis

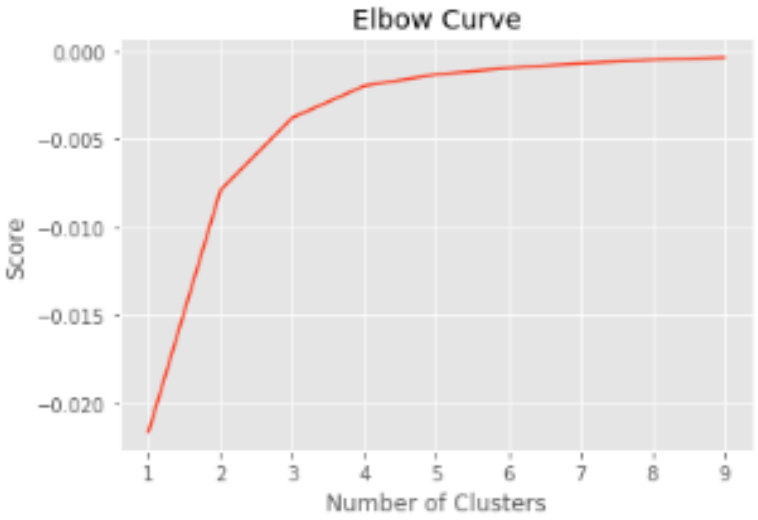
As the final part of our analysis we completed a Kmeans cluster analysis to try and determine which location in the City will be best suited for opening our store as per the requirements mentioned in section 1.2 of this report.

3.2.1 Determining optimal number of clusters

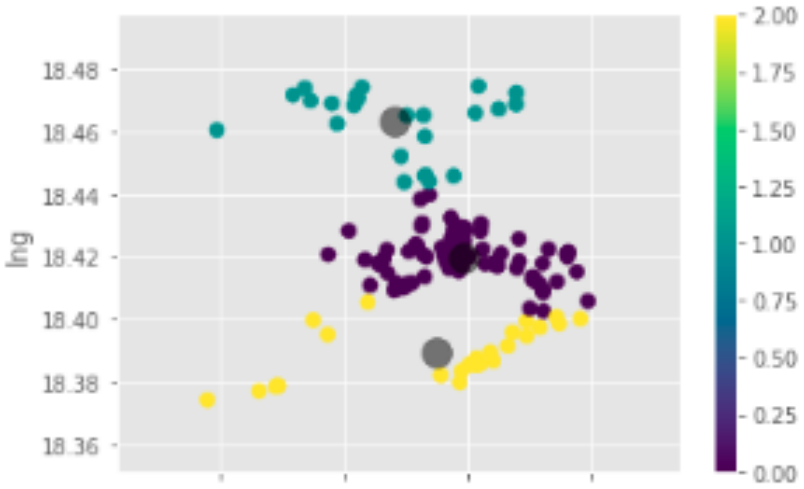
Here we determined that the optimal number of clusters for our Kmeans analysis was three clusters. The Elbow Method was used to determine this number, and this can be seen in the below graph.

For interest sake we also clustered the data using four cluster and unsurprisingly this did not output results that we significantly improved to the point where using four cluster could have been considered better than three.

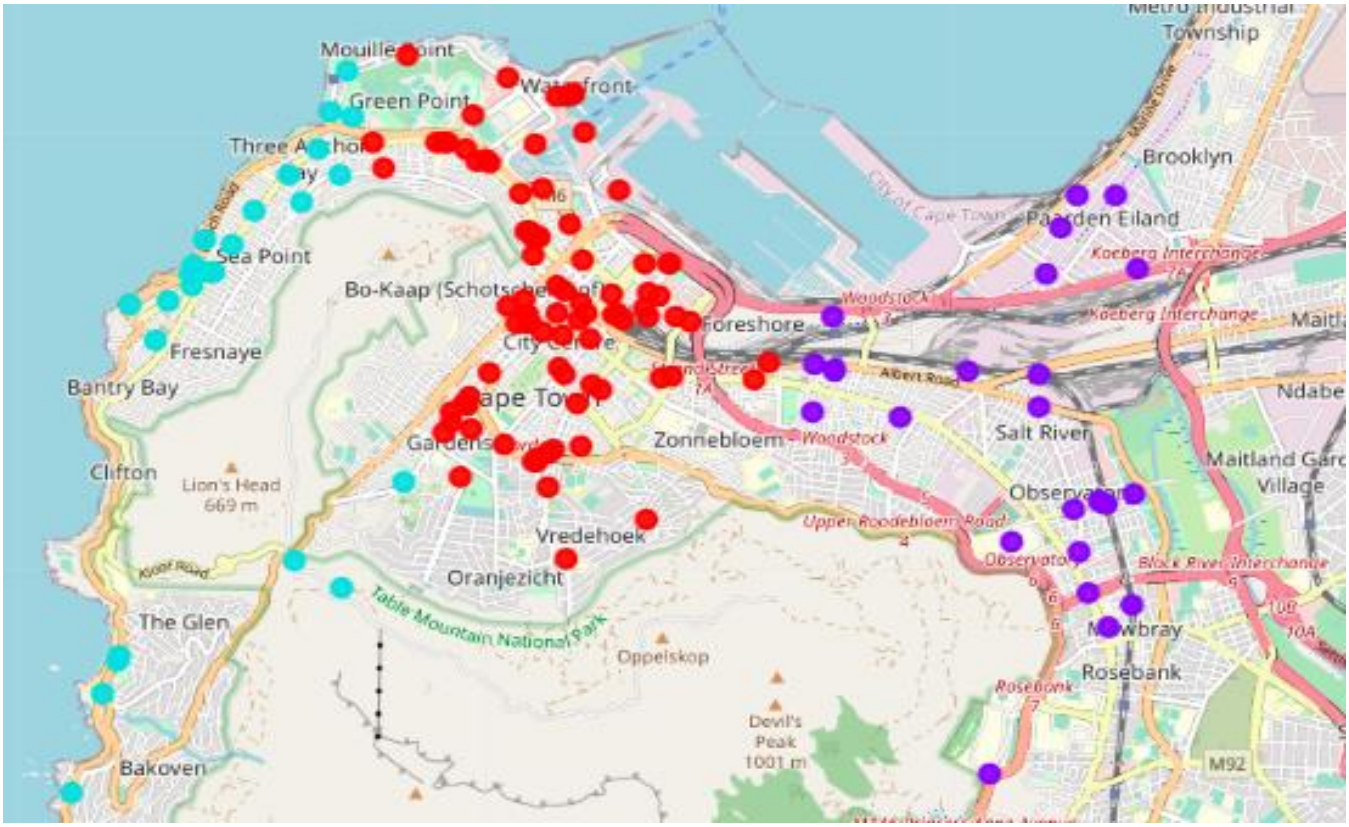
Because suburb information was not available, we used the coordinates of each location to cluster them. Later in this report we will evaluate frequency of occurrence of each of the categories.



The following scatter plot was generated to visualize the clusters. While we can definitely see the three distinct cluster, this map does not provide proper context in terms the city.



Here we see the same clusters now represented on a map using the Folium package again.



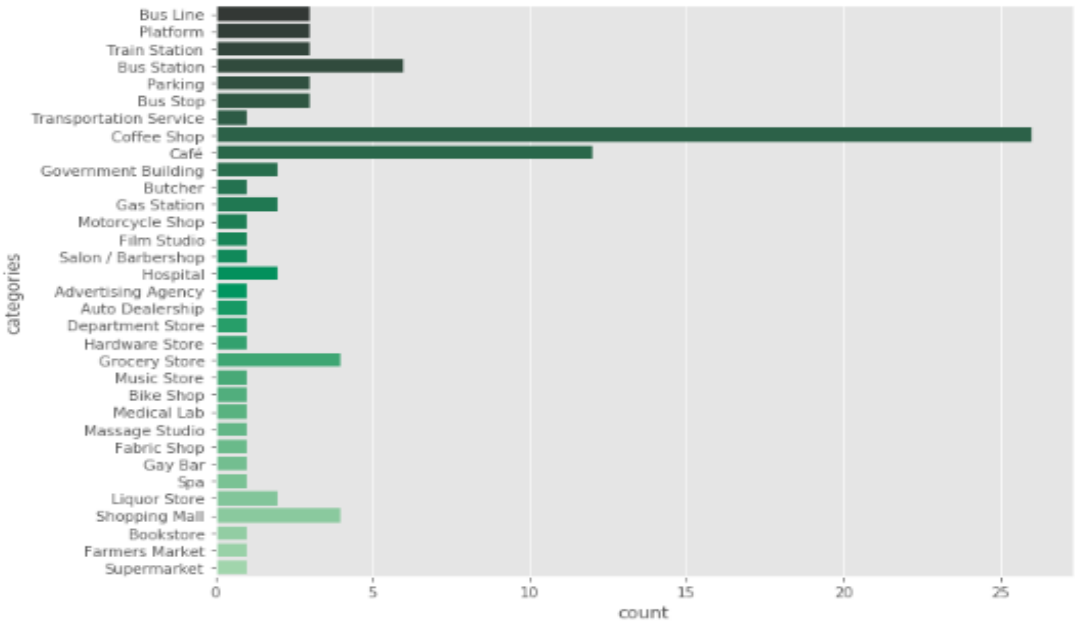
4.Results

4.1 Analysis of individual clusters

In the next section we will analysis each cluster individually. This will allow us to determine which Cluster is most suitable for opening the new coffee shop.

4.1.1 Cluster 1

The first cluster we are evaluating is the most densely populated in terms of number of categories and number of endpoints in each of these categories



The top 3 categories in terms of frequency of occurrence are:

1. Coffee Shops
2. Cafés
3. Bus Stations

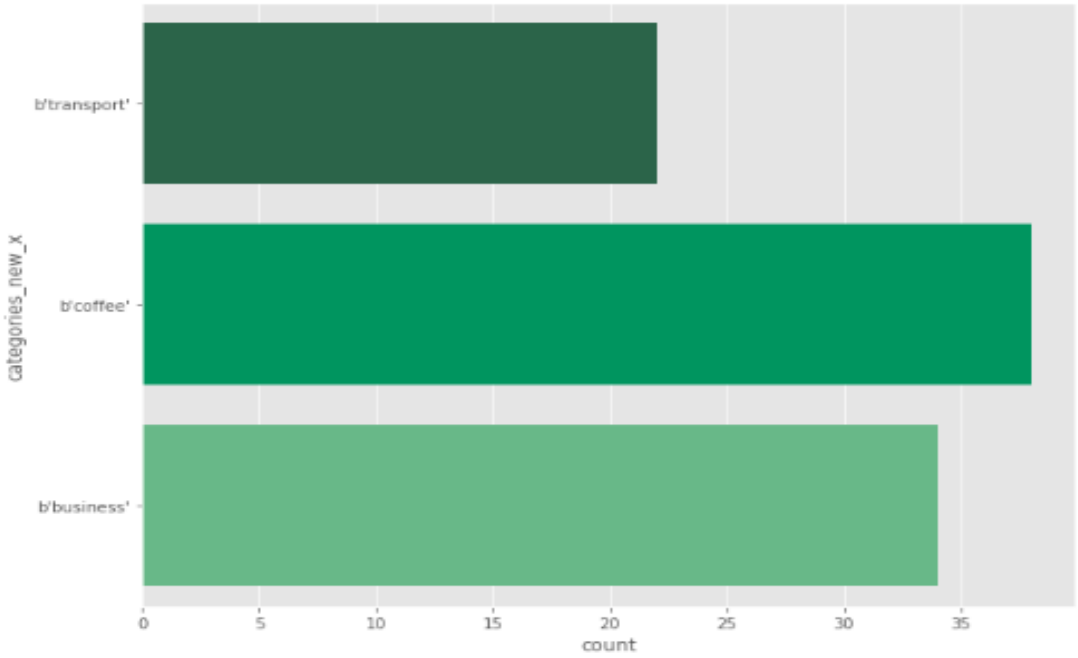
The first thing to take note of here is the high amount of competition with both “coffee” categories (coffee shop and café) being first on our top 3 list. For this cluster along there are 38 coffee stores and cafés and while this cluster is certainly the largest of the three, this is a lot of potential competition.

These broad categories are not particularly helpful when trying to understand if this area does meet the three criteria mentioned in our problem statement (section 1.2). In the below graph we have divided the locations in our clusters into three high-level categories:

1. Transport
2. Coffee (meaning coffee store)
3. Business

Here we see that the Coffee category is first in terms of frequency of occurrence, followed by surrounding businesses and finally transport.

It is important to note here that the number of coffee stores found this area might be a good reason to look elsewhere for this store but let’s review all the clusters first.



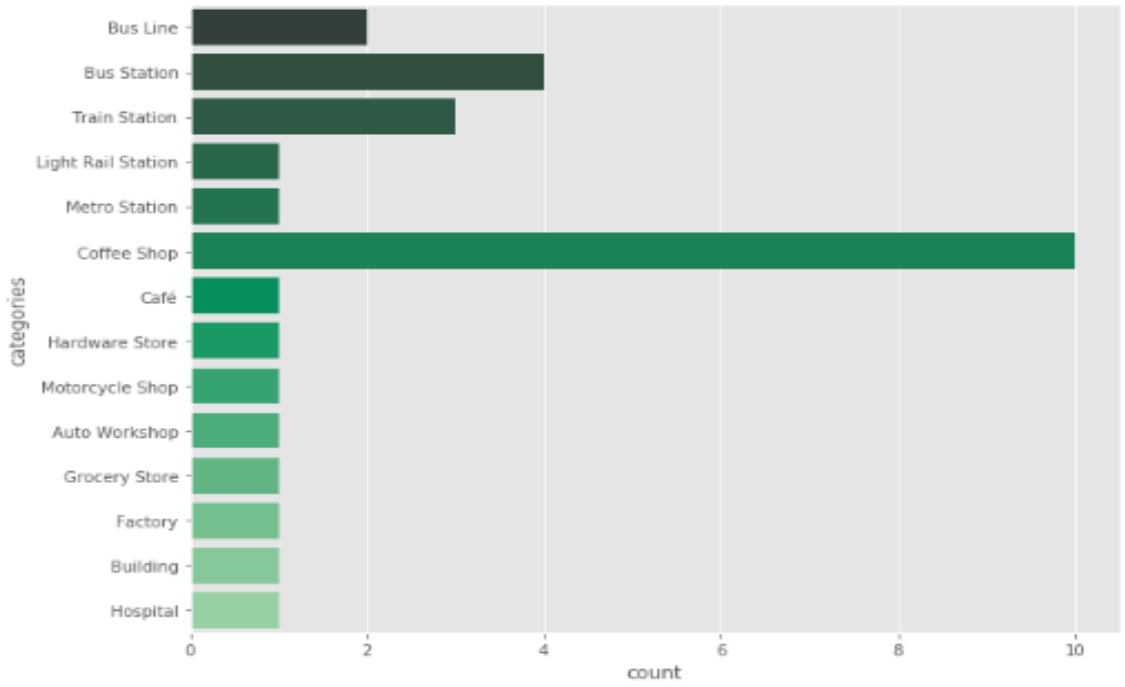
4.1.2 Cluster 2

Next, we evaluate Cluster 2.

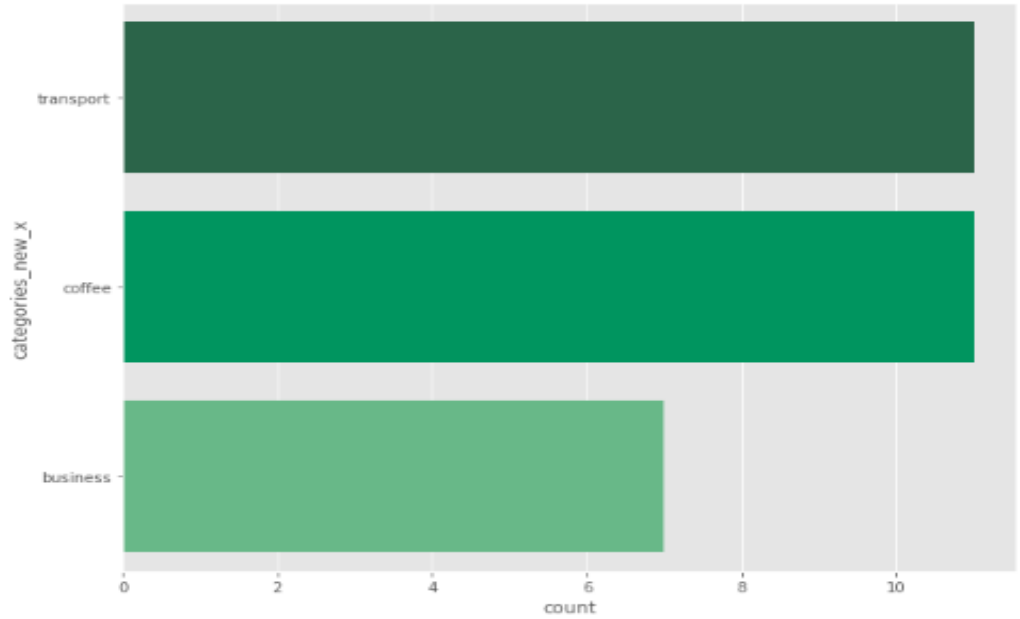
Our Top 3 categories in terms of frequency of occurrence for this cluster is:

1. Coffee Shop
2. Bus station
3. Train Station

Again, we see many coffee stores found in this cluster, totaling at 11. While this still seems to be a high number of stores for this cluster (which is significantly smaller than Cluster 1), there seems to be more transportation endpoints in this cluster.



Again, we added high-level categories to this cluster's dataset of transport, coffee and businesses, and came up with the following results.



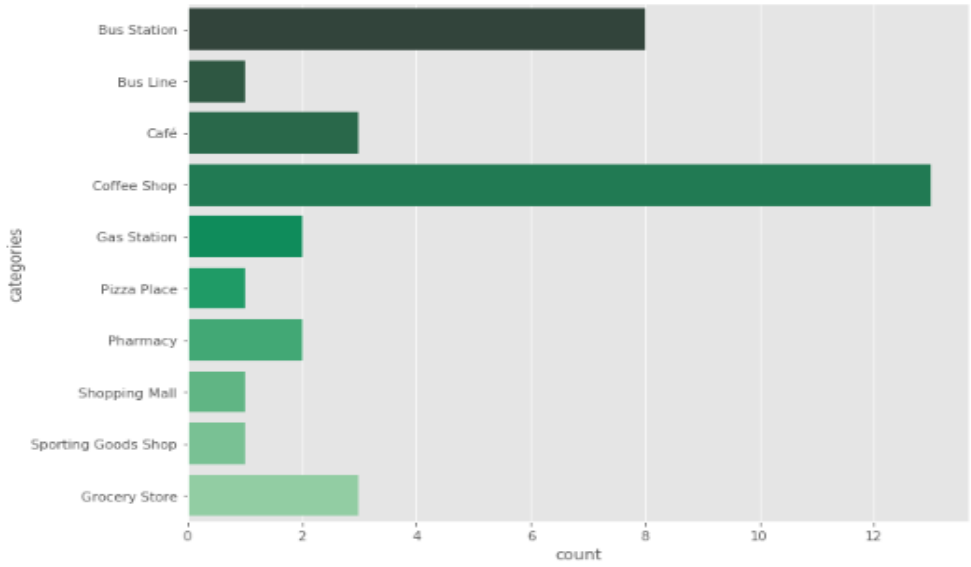
This cluster appears to have an equal amount of Transportation routes as coffee shops. It also has few surrounding businesses in terms of the ratio we saw in Cluster 1. This cluster is also located on the East side of the city, which we know to be more of an industrial and small business area.

4.1.3 Cluster 3

Finally, we evaluate the last cluster, Cluster 3. This cluster is about the same size as cluster 2.

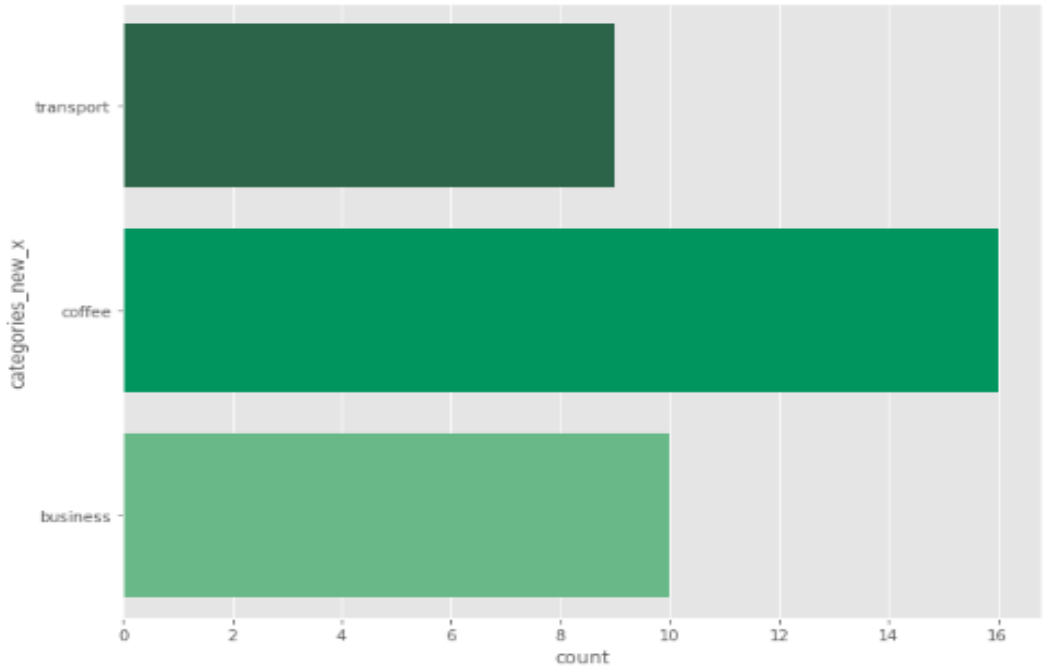
Our top 3 categories based on frequency of occurrence are:

1. Coffee Shop
2. Bus Station
3. Café and Grocery store (joint 3rd)



The total number of coffee stores come to 16 which is more than we saw in Cluster 2.

Let's evaluate the high-level categories for this cluster from the below graph. Here we see that the largest category is coffee, then business and finally transport. In terms of ratio of coffee stores to business and transport this cluster has far more stores than even Cluster 1.



It may be important to note here that this cluster is located on the Western side of the map city which is also the beachside of the city.

4. Conclusion and recommendations

In summary, Cluster 1 has a very large number of competitive coffee stores, many surrounding businesses and few transportation points in relation to the other two clusters.

Cluster 2 had a fair amount of coffee stores and just as many transportation routes with fewer surrounding businesses in relation to the other clusters.

Cluster 3 had the largest number of coffee stores in relation to the other cluster's ratio between coffee shops and surrounding businesses and transportation points.

We suggest that the owner looks to open his coffee store in the vicinity of Cluster 2. While this cluster does not have as many surrounding business in relation with the other two cluster we believe the high number of transportation points and the lower competitive environment will make this the ideal area to open the coffee store.

Having said that, as previously mentioned, Cluster 2 is both an industrial area and a business district. For the purpose of flow of people, it may make sense to open the shop in the more "business" part of the district and additionally to locate on a main street to further exposure.

5. Limitations of analysis

We would like to state here that we were severely limited by the allowance of the Foursquare API to bring in data of the various location categories which only allows 50 endpoints to be brought back per API call.

As seen in the Section 3: Methodology of this report there were only 3 API calls completed which means that there was only a maximum of 150 data points capable of being brought back. Additionally, even if we had called each Foursquare category used individually there still would not have been enough data points to fully encompass all the Transportation points, Coffee Shops and Surrounding Businesses of the Cape Town Metropolitan area.

Regardless of this, it was an extremely interesting analysis to complete and there is definitely merit in completing this analysis again but with datasets that truly encompass all the relevant location data required.