**Clustering Constituencies in Nairobi city, Kenya**

**Nyawate Hillary**

**1. Introduction**

**1.1 Background**

Nairobi city is a city that blends people from all cultures and walks of life. Adorned with modern skyscrapers, world class restaurants, fully equipped hospitals, modern shopping malls, schools, abundant private and public transportation, and universities and colleges that provide local and international curriculum. The city is also **home to numerous local and international businesses and organizations**. Nairobi bustles with activity. It's **a city that never sleeps**; the rhythm is fast, day and night. There's always something to do and see in Nairobi and its **people are friendly and hospitable**

**1.2 Problem**

Due to the strategic location of certain venues for commercial enterprises in Nairobi city, Kenya near the central business district, there is a rising number of restaurants and recreational facilities in these area and this has increased the population of Nairobi City County rapidly, which has led to traffic congestion and air/land pollution.

**1.3 Interest**

Therefore this project analyses the data about the geographical location of these venues in order to visualize the distribution of these facilities by clustering them into different categories that will inform investors on specific location in Nairobi to set up facilities that are less found in those areas. This analysis will inform the government on the strategy to improve city connectivity via better roads to and from these venues, whereas have a better understanding of the living standards of citizens from various locations in the city and reroute traffic towards other constituencies at rush hours.

**2. Data acquisition and cleaning**

**2.1 Data sources**

The data used in this project is a table of the current constituencies of Nairobi city county, Kenya with the corresponding wards in each constituency. The geographical location of these constituencies and the foursquare data on the venues of different facilities located in these constituencies. The data is collected from a web page through web scrapping using the python pandas library to read the table from the website. The data about the venues located in each of these constituencies is from foursquare which is requested using foursquare API calls in python libraries; requests and json.normlize. This is the [link](http://www.infohub.co.ke/2016/04/list-of-nairobi-constituencies-and-wards.html.) where I found the constituency data.

**2.2 Data cleaning**

Data scrapped from that website is structured into a data frame and I use *geopy* libraries to get the coordinates of each constituencies and used this new data frame with coordinates to explore the constituencies in Nairobi.

Figure 1.0 Sample cleaned data frame of the constituencies with their coordinates

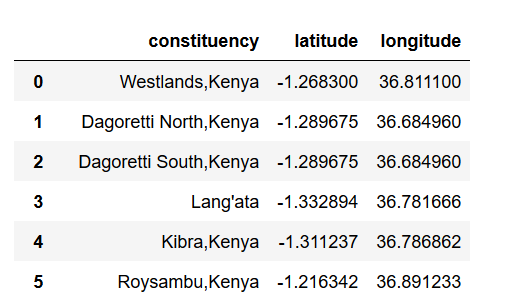


Figure 1.1 Sample data of the requested venues form one constituency

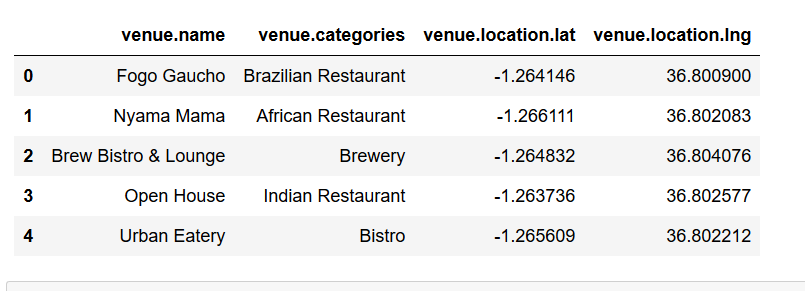
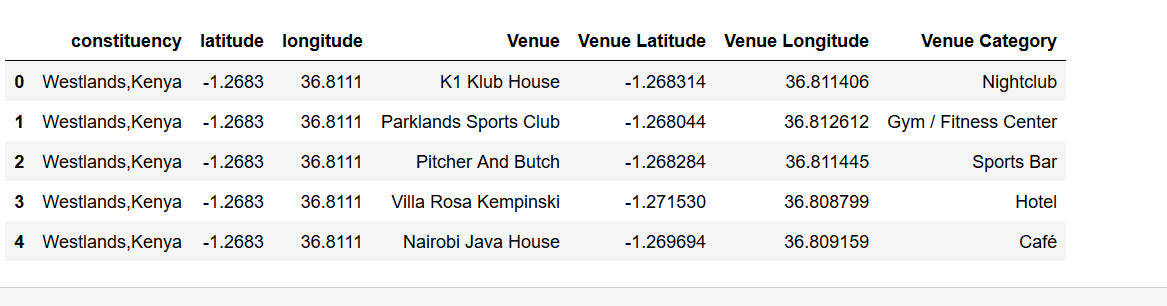


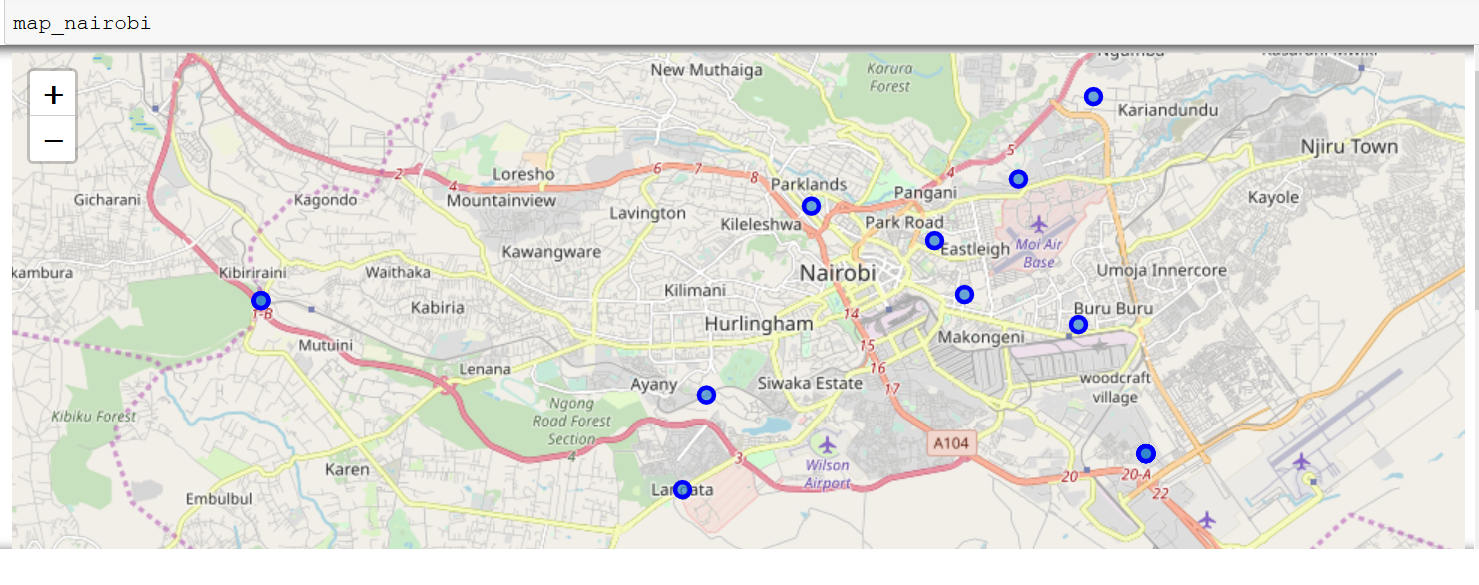
Figure 1.2 Sample of merged data frame of constituencies with their venues



**3. Geo-location analysis**

In order to visualise the location and distribution of these constituencies in Nairobi city I used python libraries to get the geographical location of Nairobi city using the ***Nomatim*** function in the ***geopy*** library and applied the ***folium*** library to plot the geographical map of Nairobi city and superimpose the coordinates of the constituencies on it as pop ups.

Figure 1.3 Map of Nairobi city with location of the constituencies

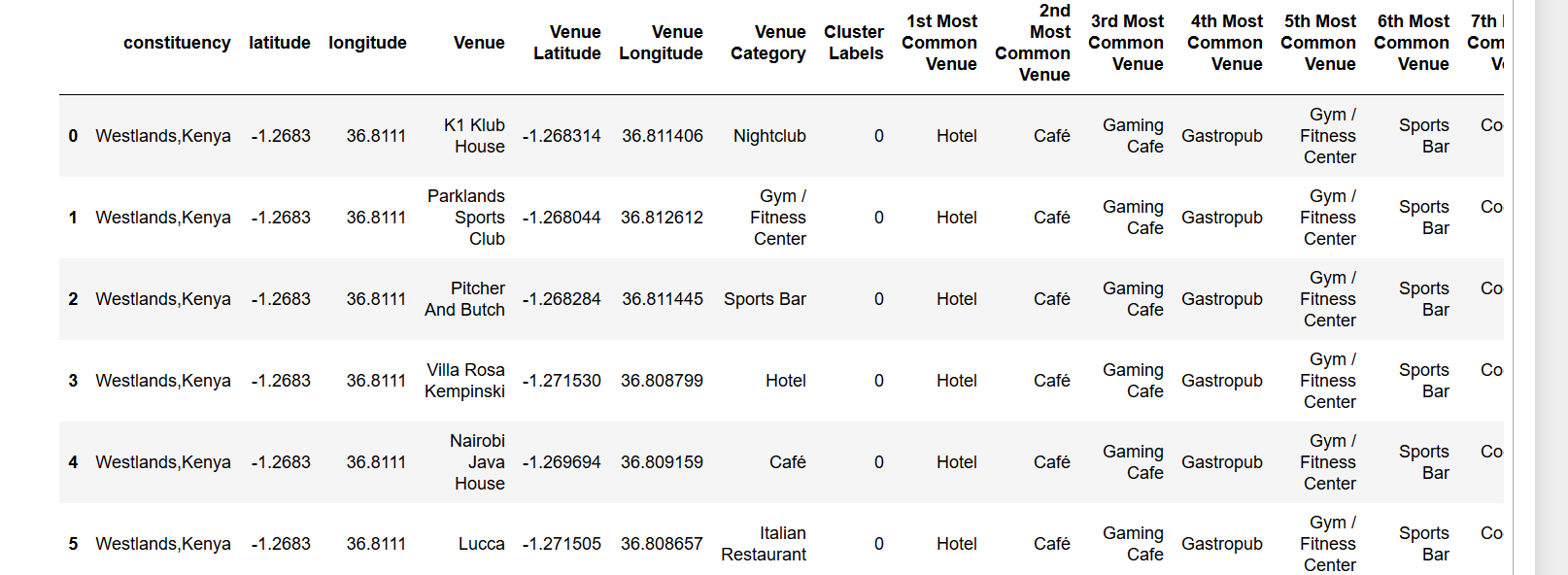


Next, I utilized the Foursquare API call from python to explore the constituencies through requesting for data on the venues located in these constituencies based on their coordinates. Then I got the top 100 venues that are in Nairobi within a radius of 500 meters in each constituency. Then from here the file that is returned is a json file hence has to be cleaned and converted to a data frame and all categories of the venues are extracted and the data is structured a=into a new data frame

**4. Clustering**

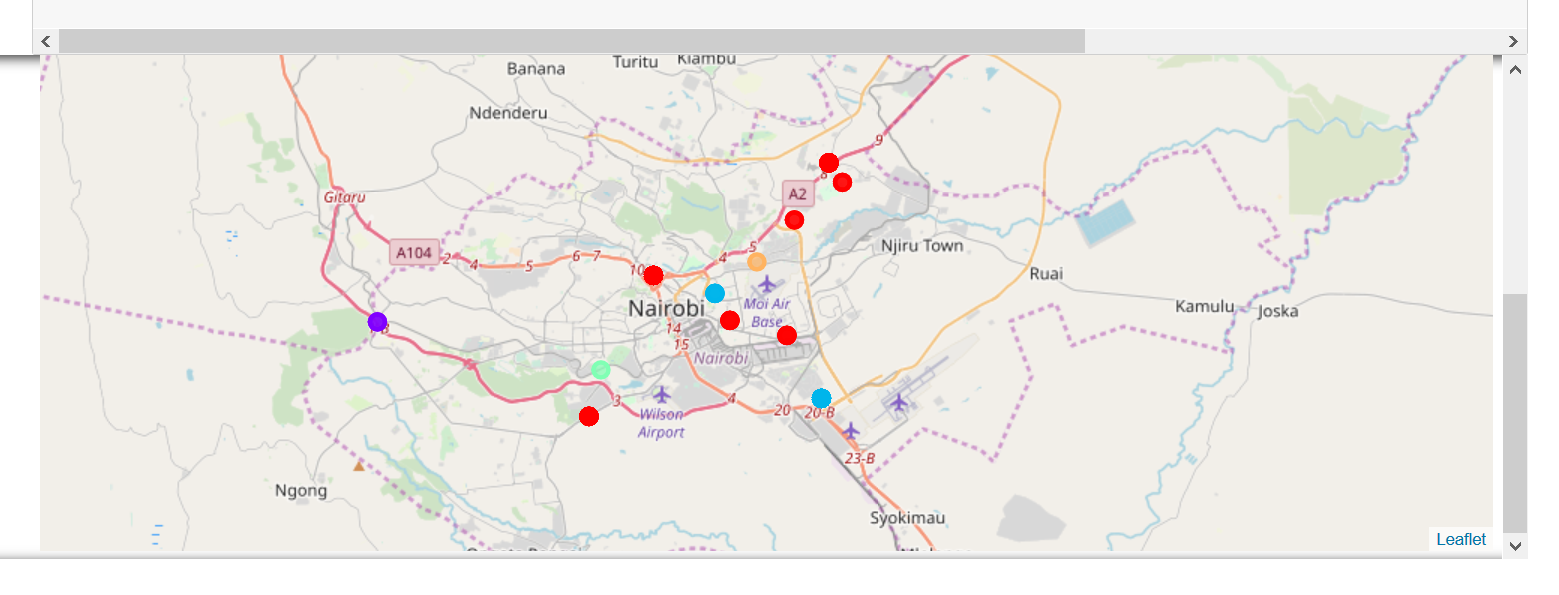
Before grouping constituencies, I performed a cross calculation of each constituency based on every attribute of the venues in order to get the number of each attribute in each constituency. Then perform the one hot encoding to get the dummies of the categories of the venues and cross tabulate with constituencies to create a final data frame that is used to fit the K-means clustering model to group venues into clusters based on similarities of these venues. The following Figure displays the final data frame that sorts all venues based on their trend and common venue in each constituency.

Figure 1.4 Final data frame of the clustered and sorted constituencies based on the venues



The folium library once again is applied to display the clustered constituencies based on the colour of the pop up in the map below.

Figure Final map of the clustered constituencies in different colors



**Key for the map**

**Red: (cluster 0) - Food outlet venues with food related services**

**Purple :( cluster 1) – Market place**

**Blue: (cluster 2) - Resting places**

**Green: (cluster 3) – Sporting venues**

**5. Conclusion**

In this project I analysed the venues of various facilities in Nairobi city per constituency and clustered the constituency based on the categories and trend and frequency of these venues so that I could compare these constituencies with each other and make decisions on which to find any type of facility in Nairobi city. Also I found out that most constituencies have restaurants and eating venues as most common, whereas very few had venues had sporting facilities as most common.

**6. Recommendations**

To begin with the geo locator did not perform very accurately hence some constituencies that are too near to each other were sharing coordinates, this made the pop ups on the map to be fewer thus not very representative of all the data. If the geo locator could be modified in order to get unique coordinates for all the constituencies then we would view all constituencies and cluster them accordingly.