IBM Data Science Capstone Project

Neighbourhoods of Nairobi

William Odhiambo

Introduction

- Nairobi is the capital of Kenya, which is located in Africa. In addition to being the capital, Nairobi is also the country's largest city by population. The last official population was taken in 2009 and at that time was 3,138,369 in the city proper. That number has since grown to approximately 3.5 million. The metro area has over 6.5 million residents. This "Green City In The Sun" has a history dating back to 1899 and continues to grow as rural residents make their way to this big city for employment opportunities.

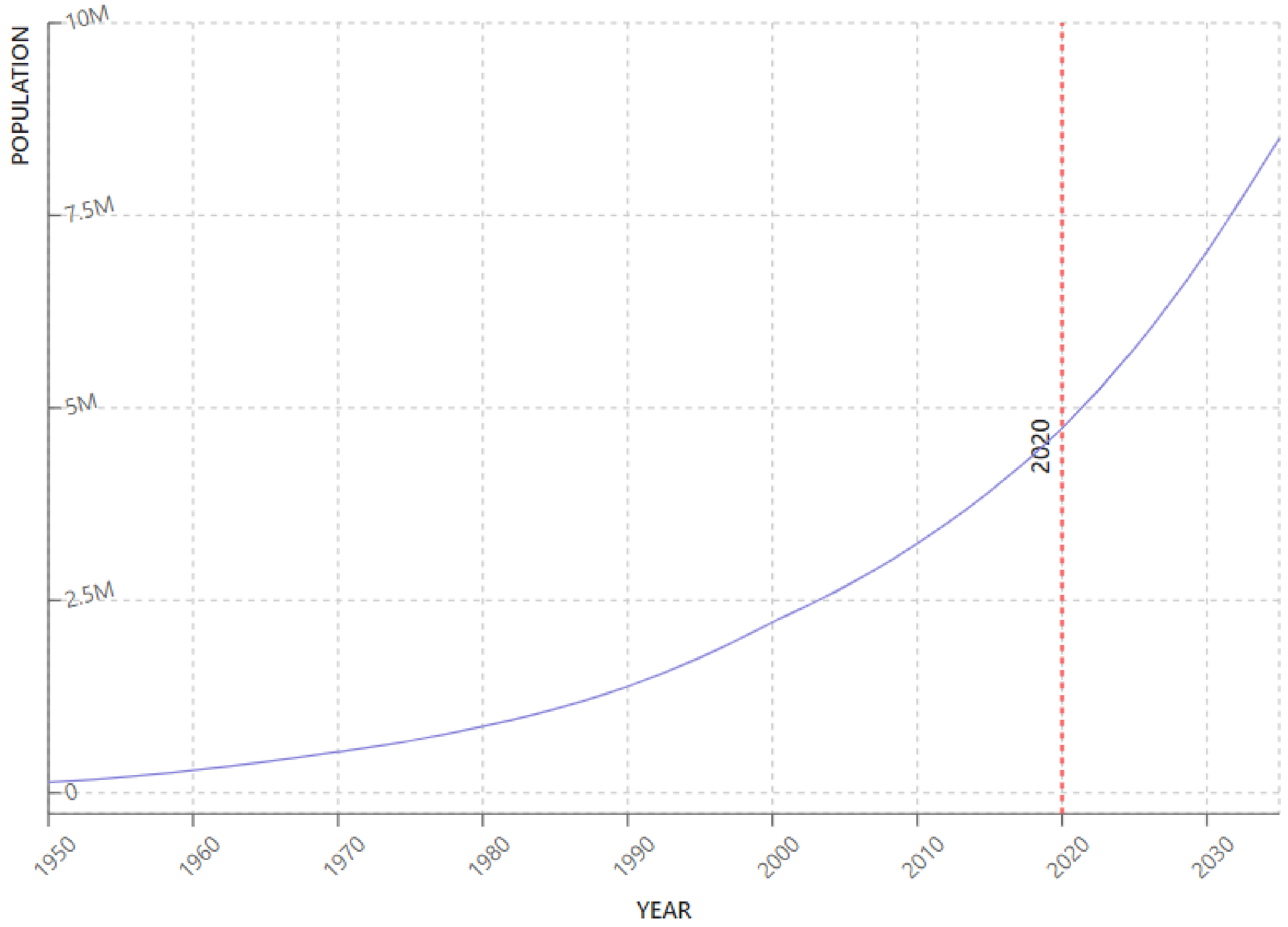
Population

- Nairobi's 2020 population is now estimated at **4,734,881**. In 1950, the population of Nairobi was **137,456**. Nairobi has grown by 821,369 since 2015, which represents a 3.88% annual change. These population estimates and projections come from the latest revision of the UN World Urbanization Prospects. These estimates represent the Urban agglomeration of Nairobi, which typically includes Nairobi's population in addition to adjacent suburban areas.
- The city of Nairobi is growing consistently and currently stretches over an area of of 696 kilometers squared (269 square miles). This area size - in combination with the total number of residents - brings us the current population density which is now approximately 4,850 residents per square kilometer. (12,600 people living per square mile).

Nairobi Population 2020 | Source: World Urbanization Prospects

Nairobi Population 2020

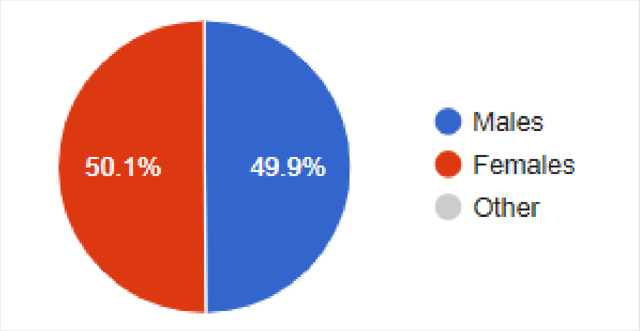
4,734,881



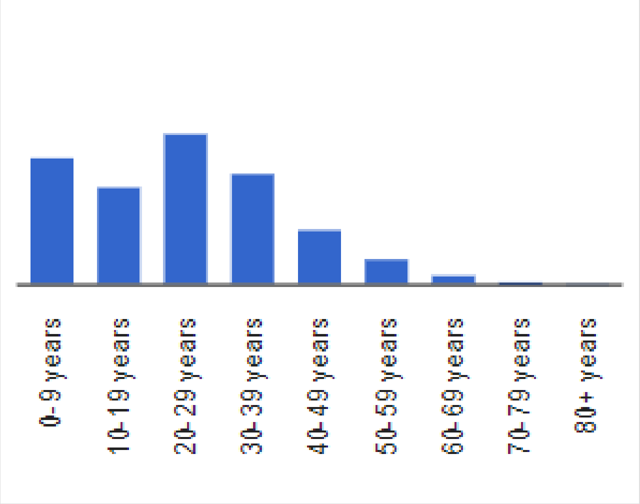
Country	Kenya
Density (km^2)	6803
Growth Rate	3.88%
Area	696 km^2

2019 Population Based on Sex, Age, and Urbanization

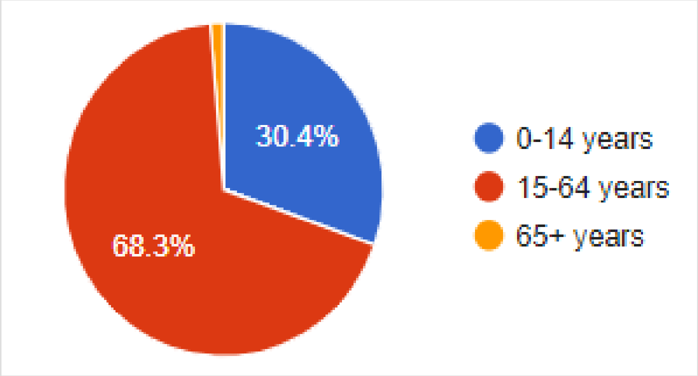
Further information about the population structure:



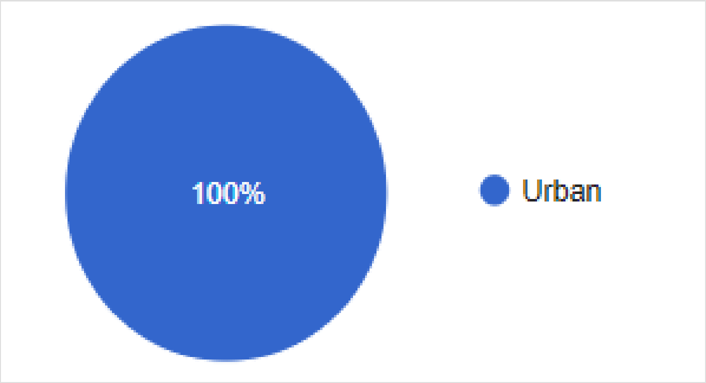
Gender (C 2019)	
Males	2,192,452
Females	2,204,376
Intersex	245



Age Distribution (C 2019)	
0-9 years	957,699
10-19 years	730,403
20-29 years	1,146,567
30-39 years	841,266
40-49 years	428,094
50-59 years	190,550
60-69 years	70,367
70-79 years	23,791
80+ years	7,948



Age Groups (C 2019)	
0-14 years	1,336,249
15-64 years	3,002,314
65+ years	58,122



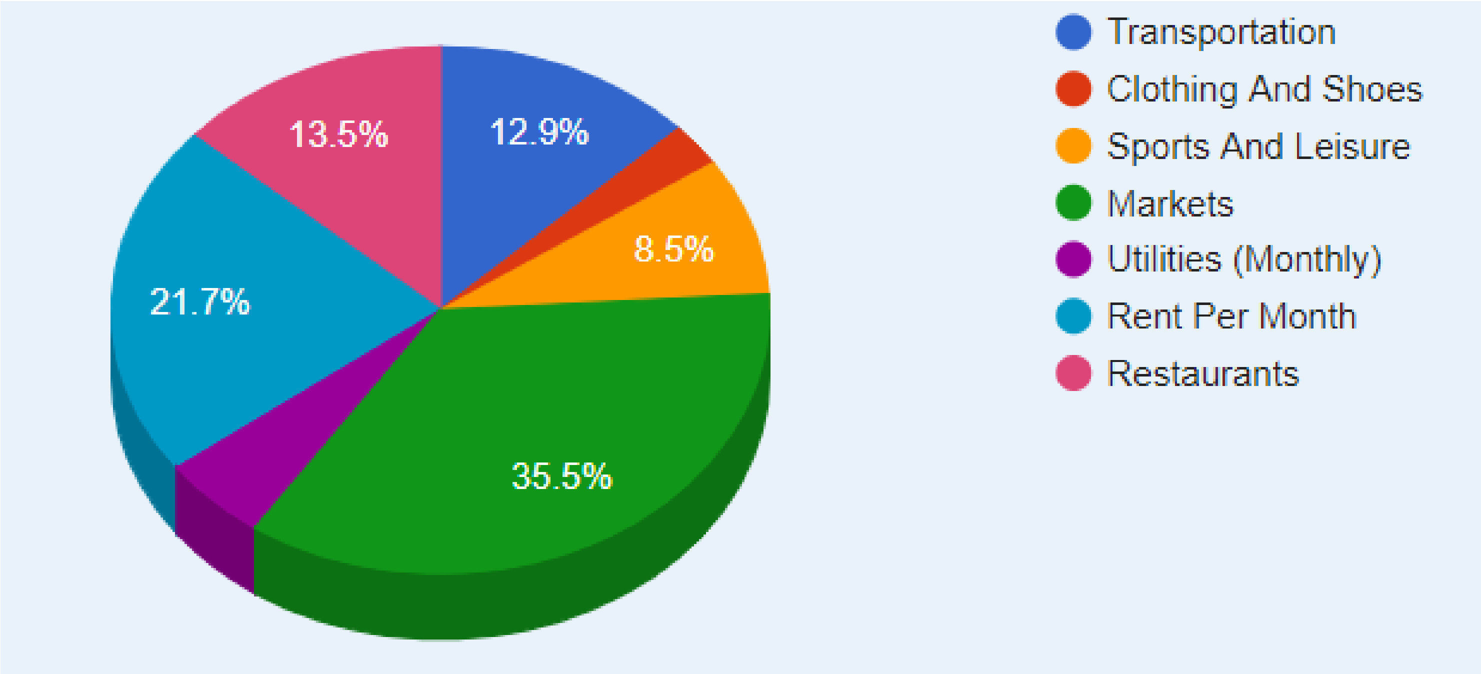
Urbanization (C 2019)	
Urban	4,397,073

City Size and Population Density

- The city of Nairobi is growing consistently and currently stretches out over a surface area of of 696 kilometers squared (269 square miles). This area size - in combination with the total number of residents - brings us the current population density which is now approximately 4,850 residents per square kilometer. (12,600 people living per square mile).

Brief Summary Of Nairobi's Living Costs and Spending Habits

- Four-person family monthly costs: **(198,633.63KSh)** without rent.
- A single person monthly costs:**(55,581.81KSh)** without rent.
- Cost of living rank **316th out of 488** cities in the world.
- Nairobi has a cost of living index of **39.40**.



Business Problem

- Nine million individuals are expected to enter the labor force in a decade between 2015 and 2025, further pushing up the country's unemployment rate which stood at 9.3% in 2017

- According to the Kenya Economic Survey 2019, 840,600 new jobs were created in 2018 compared to 909,800 reported in 2017.

- Kenya has to create at least 900,000 jobs annually between 2019 and 2025 to absorb the high number of youths joining the job market, according to the latest World Bank report

- The ten year World Bank survey projects unemployment rate in Kenya was to rise to 10.5 per cent in 2019 before slowing to 10 per cent in 2020

- *The core objective is to establish businesses that will create Jobs - especially for the youth specifically in Nairobi, seeing that a large percentage of Nairobi's population seem to be youthful. These businesses, also, need to be the type that appeal and build spending habits to customers of a young age, particularly from the teenage years to the late youth.*

Data Collection and Pre-processing

- Data of Nairobi's Neighbourhoods was obtained/scraped from https://en.wikipedia.org/wiki/Category:Suburbs_of_Nairobi and formatted into a pandas Data Frame.
- Population Information, Visualizations and CSV file was obtained from Nairobi Population 2020 (Demographics, Maps, Graphs), and Nairobi (County, Kenya) - Population Statistics, Charts, Map and Location.
- Geocoding: Nominatim Geocoding service, which is built on top of OpenStreetMap data, was used to find the latitudinal and longitudinal values of Nairobi's neighborhoods. Unfortunately, not all neighborhoods could be mapped because the process of converting their addresses to coordinates was somewhat limited by the Geocoder.
- Folium: This builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the leaflet.js library. The library has a number of built-in tile sets from OpenStreetMap
- Foursquare Developers Access to venue data: https://foursquare.com/

Cleaning the Neighbourhood CSV File and locating the Coordinates of Each Neighbourhood

```
print(data['Neighborhood'].unique())
print(data.shape)
```

```
['B' 'Bahati, Nairobi' 'Buruburu' 'D' 'Dagoretti' 'Dandora' 'E'
 'Eastleigh, Nairobi' 'Embakasi' 'G' 'Gatwekera' 'Gigiri' 'Githurai' 'H'
 'Highridge' 'Huruma' 'J' 'Jericho, Nairobi' 'K' 'Kambi Muru' 'Kamulu'
 'Kangemi' 'Karen, Kenya' 'Kariobangi' 'Kasarani' 'Kawangware' 'Kiambiu'
 'Kibera' 'Kichinjio' 'Kilimani' 'Kisumu Ndogo' 'Kitisuru' 'Korogocho' 'L'
 'Laini Saba' "Lang'ata" 'Lavington, Nairobi' 'Lindi, Nairobi'
 'Lucky Summer Estate' 'M' 'Madaraka Estate' 'Majengo, Nairobi'
 'Makongeni' 'Mashimoni' 'Mathare' 'Mathare Valley' 'Matopeni' 'Mugumoini'
 'Muirigo' 'Mukuru kwa Njenga' 'O' 'Ofafa' 'P' 'Pangani, Nairobi'
 'Parklands, Nairobi' 'Pumwani' 'R' 'Raila' 'Runda' 'S' "Sarang'ombe"
 'Shilanga' 'Siranga' 'South B' 'South C' 'Soweto East' 'Soweto West'
 'Syokimau' 'U' 'Upper Hill, Nairobi' 'Uthiru' 'W' 'Westlands, Nairobi']
(73, 1)
```

Cleaning '

```
data[data['Neighborhood'].isin(['B','D','E','G','H','J','K','L','M','O','P','R','S','U','W'])]
```

4]:

	Neighborhood
0	B
3	D
6	E
9	G
13	H

```
geolocator = Nominatim(user_agent='my-application')
```

```
# 1 - conveneint function to delay between geocoding calls
geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1)
# 2- - create location column
data['location'] = data['Neighborhood'].apply(geocode)
# 3 - create longitude, laatitude and altitude from location column (returns tuple)
data['point'] = data['location'].apply(lambda loc: tuple(loc.point) if loc else None)
# 4 - split point column into latitude, longitude and altitude columns
data[['latitude', 'longitude','altitude']] = pd.DataFrame(data['point'].tolist(), index=data.index)
```

```
data.head()
```

	Neighborhood	location	point	latitude	longitude	altitude
1	Bahati, Nairobi	(Bahati, Chahafi, Kisoro, Western Region, Ugan...	(-1.3, 29.766667, 0.0)	-1.300000	29.766667	0.0
2	Buruburu	(Buruburu, Amolatar, Northern Region, Uganda, ...	(1.766667, 32.8, 0.0)	1.766667	32.800000	0.0
4	Dagoretti	(Dagoretti, Dagoretti Road, Kabiria, Nairobi, ...	(-1.2896931, 36.6849829, 0.0)	-1.289693	36.684983	0.0
5	Dandora	(Dandora, Salamat سلامات, Tchad / 11.08) ,تشاد...	(11.0857, 20.3518667, 0.0)	11.085700	20.351867	0.0
7	Eastleigh, Nairobi	(Eastleigh, Nairobi, 00611, Kenya, (-1.2778285...	(-1.2778285, 36.8486835, 0.0)	-1.277829	36.848683	0.0

```
data.drop(['location','point','altitude'], axis=1, inplace = True)
print(data.shape)
```

```
(58, 3)
```

- Further Cleaning of Null Values as well as latitudinal and longitudinal values that seemed to be quite off.
- Cleaning of these values was done twice, where the second time was after the Neighbourhoods were renamed in order to get more accurate coordinates. This attempt was partially successful; some coordinates still had odd, and null values.

```
odd_values = data.loc[[5,14,15,26,27,29,38,48,49,57,61,62,64,69],] #Include Null Coordinates
odd_values.reset_index(drop=True, inplace=True)
odd_values.head()
```

	Neighborhood	latitude	longitude
0	Dandora	11.085700	20.351867
1	Highridge	54.033333	-114.133333
2	Huruma	51.362249	9.469233
3	Kiambiu	NaN	NaN
4	Kibera	57.747092	26.988900

```
odd_values = odd_values.set_index('Neighborhood')
```

```
odd_values.rename(index={'Dandora':'Dandora,Kenya', 'Highridge':'Highridge,Kenya','Huruma':'Huruma,Kenya',
                        'Kiambiu':'Kiambiu,Kenya','Kibera':'Kibera,Kenya','Kilimani':'Kilimani,Kenya',
                         'Lucky Summer Estate':'Lucky Summer,Kenya','Muirigo':'Muirigo,Kenya','Mukuru kwa Njenga':'Mukur
                        'Raila':'Raila,Kenya','Shilanga':'Shilanga,Kenya','Siranga':'Siranga,Kenya','South C':'South C,K
                        'Upper Hill, Nairobi':'Upper Hill, Kenya'}, inplace=True)
```

```
odd_values['location'] = odd_values['Neighborhood'].apply(geocode)
odd_values['point'] = odd_values['location'].apply(lambda loc: tuple(loc.point) if loc else None)
odd_values[['latitude', 'longitude','altitude']] = pd.DataFrame(odd_values['point'].tolist(), index=odd_values.index)
```

```
odd_values.drop(['point','location','altitude'],axis=1,inplace=True)
odd_values.shape
```

(14, 3)

```
odd_values.head()
```

	Neighborhood	latitude	longitude
0	Dandora,Kenya	-1.249073	36.896289
1	Highridge,Kenya	54.033333	-114.133333
2	Huruma,Kenya	51.362249	9.469233
3	Kiambiu,Kenya	NaN	NaN
4	Kibera,Kenya	-1.314787	36.799647

- The coordinates that were successfully revalued were concatenated into the data frame that had the rest of the coordinates which had no issues.

```
cleaned_data = odd_values.drop(index=[1,2,3,7,8,10,13])
cleaned_data.shape
```

```
(7, 3)
```

```
cleaned_data.head()
```

	Neighborhood	latitude	longitude
0	Dandora,Kenya	-1.249073	36.896289
4	Kibera,Kenya	-1.314787	36.799647
5	Kilimani,Kenya	-1.287442	36.784523
6	Lucky Summer,Kenya	-1.239308	36.898786
9	Raila,Kenya	-1.318348	36.779293
11	Siranga,Kenya	0.230146	34.231529
12	South C,Kenya	-1.320401	36.830525

```
data2 = data.drop([5,14,15,26,27,29,38,48,49,57,61,62,64,69])
data2.shape
```

```
(44, 3)
```

```
data_filt = data2.append(cleaned_data)
```

```
data_filt = data2.append(cleaned_data)
```

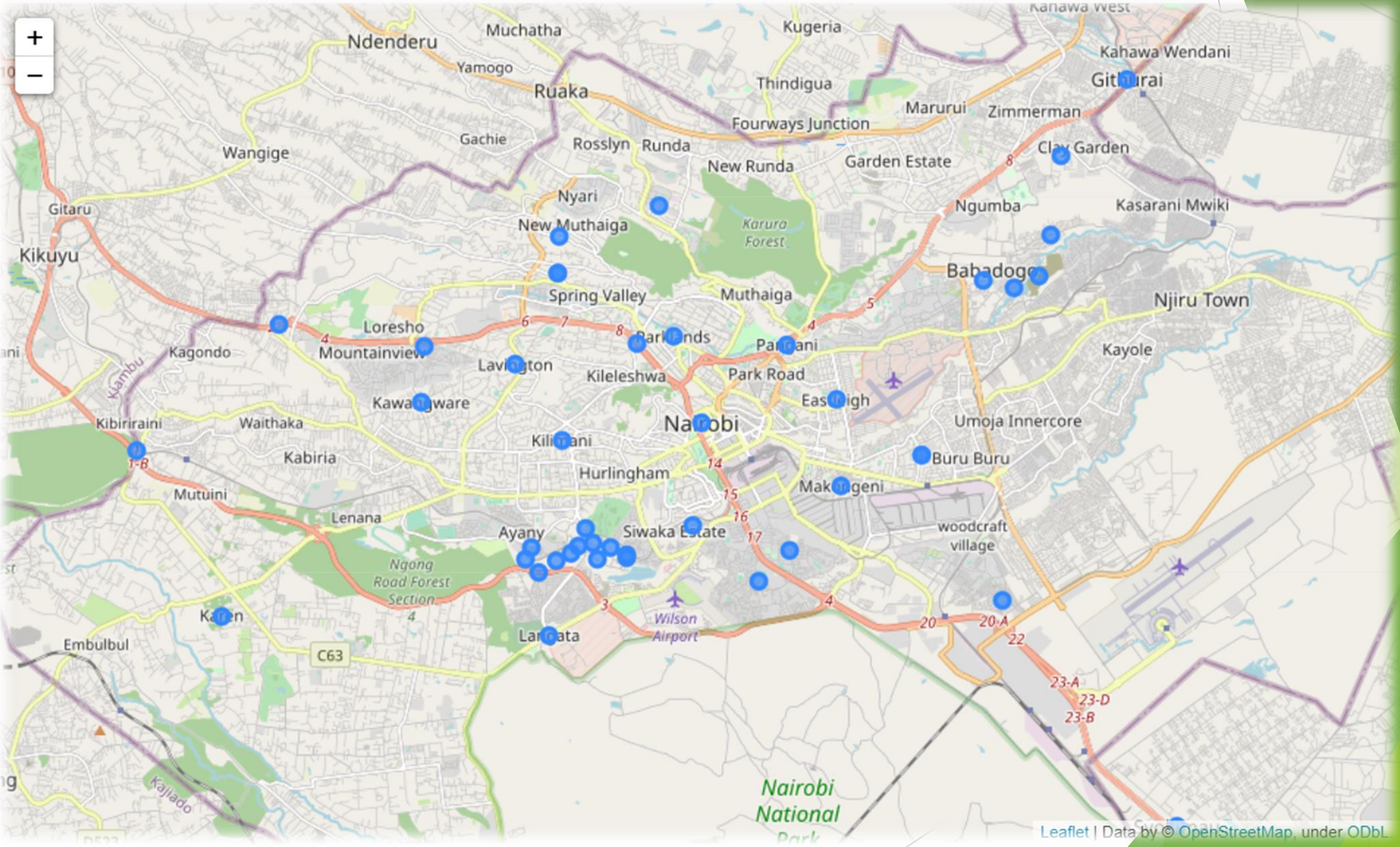
```
print(data.info())
print(data_filt.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 58 entries, 1 to 72
Data columns (total 3 columns):
Neighborhood    58 non-null object
latitude        54 non-null float64
longitude       54 non-null float64
dtypes: float64(2), object(1)
memory usage: 4.3+ KB
None
<class 'pandas.core.frame.DataFrame'>
Int64Index: 51 entries, 1 to 12
Data columns (total 3 columns):
Neighborhood    51 non-null object
latitude        51 non-null float64
longitude       51 non-null float64
dtypes: float64(2), object(1)
memory usage: 1.6+ KB
None
```

```
print(data.shape)
print(data2.shape)
print(data_filt.shape)
```

```
(58, 3)
(44, 3)
(51, 3)
```

Nairobi's Neighbourhoods (Visualized through Folium)



Using Foursquare's API to determine all the neighbourhood's nearby venues, and their coordinates, that are within a 500m range

```python
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        results = requests.get(url).json()["response"]['groups'][0]['items']

        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

	Neighborhood	African Restaurant	Arcade	Athletics & Sports	BBQ Joint	Bakery	Bar	Beer Garden	Bistro	Bookstore	...	Szechuan Restaurant	Tapas Restaurant	Tea Room	Tourist Information Center	Trail	Sta
0	Dagoretti	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
1	Eastleigh, Nairobi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Eastleigh, Nairobi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

3 rows × 89 columns

```python
nairobi_grp = nairobi_onehot.groupby('Neighborhood').mean().reset_index()
print(nairobi_grp.shape)
nairobi_grp.head(4)
```

```
(37, 89)
```

	Neighborhood	African Restaurant	Arcade	Athletics & Sports	BBQ Joint	Bakery	Bar	Beer Garden	Bistro	Bookstore	...	Szechuan Restaurant	Tapas Restaurant	Tea Room	Tourist Information Center	Trail
0	Dagoretti	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0
1	Eastleigh, Nairobi	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0
2	Embakasi	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0
3	Gigiri	0.0	0.0	0.0	0.0	0.052632	0.0	0.0	0.0	0.0	...	0.0	0.052632	0.0	0.0	0.0

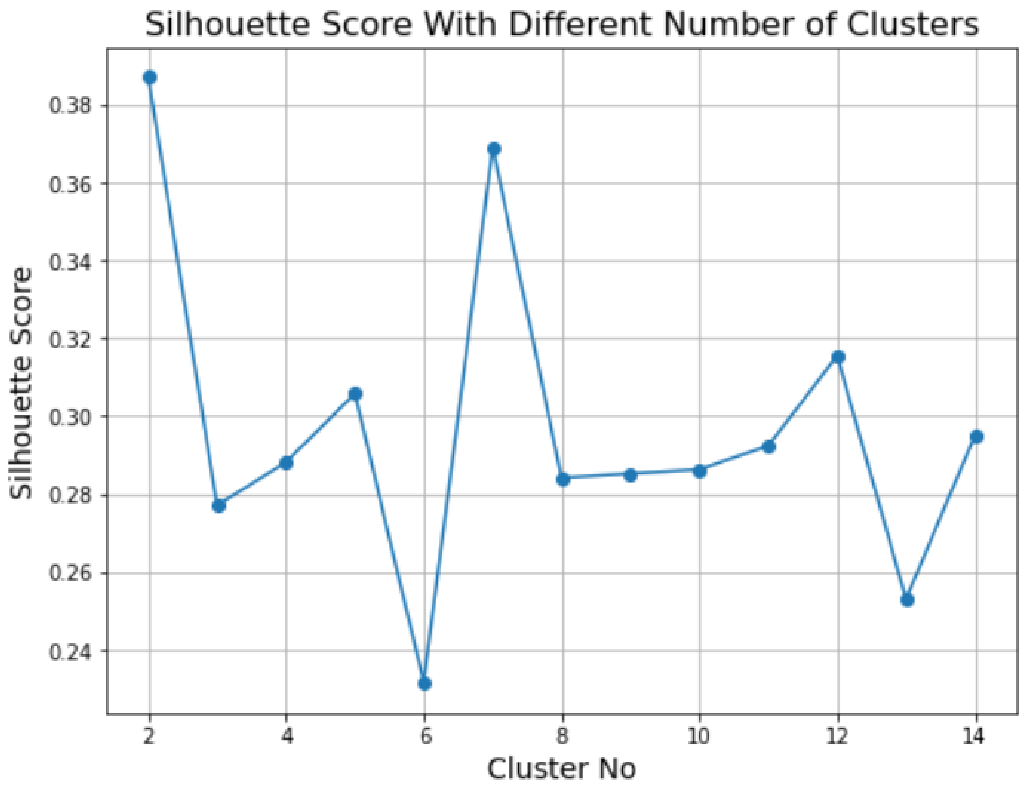
4 rows × 89 columns

Data frame of the top Ten Most common Venues in each neighbourhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Dagoretti	Flea Market	Zoo Exhibit	Fast Food Restaurant	Cricket Ground	Deli / Bodega	Department Store	Dessert Shop	Dry Cleaner	Eastern European Restaurant	Electronics Store
1	Eastleigh, Nairobi	Hotel	Men's Store	Food Court	Shopping Mall	Electronics Store	Zoo Exhibit	Cricket Ground	Deli / Bodega	Department Store	Dessert Shop
2	Embakasi	Lounge	Convenience Store	Fast Food Restaurant	Cricket Ground	Deli / Bodega	Department Store	Dessert Shop	Dry Cleaner	Eastern European Restaurant	Electronics Store
3	Gigiri	Café	Frozen Yogurt Shop	Spa	Deli / Bodega	Lounge	Pool	Restaurant	Sandwich Place	Burger Joint	Ethiopian Restaurant
4	Githurai	Flea Market	Train Station	Moving Target	Bus Station	Zoo Exhibit	Ethiopian Restaurant	Deli / Bodega	Department Store	Dessert Shop	Dry Cleaner

Finding the optimal number of clusters with Silhouette analysis on KMeans clustering:

- The silhouette plot displays a measure of how close each point in one cluster is to points in the neighbouring clusters and thus provides a way to asses parameters like number of clusters visually. This measure has a range of (-1,1)

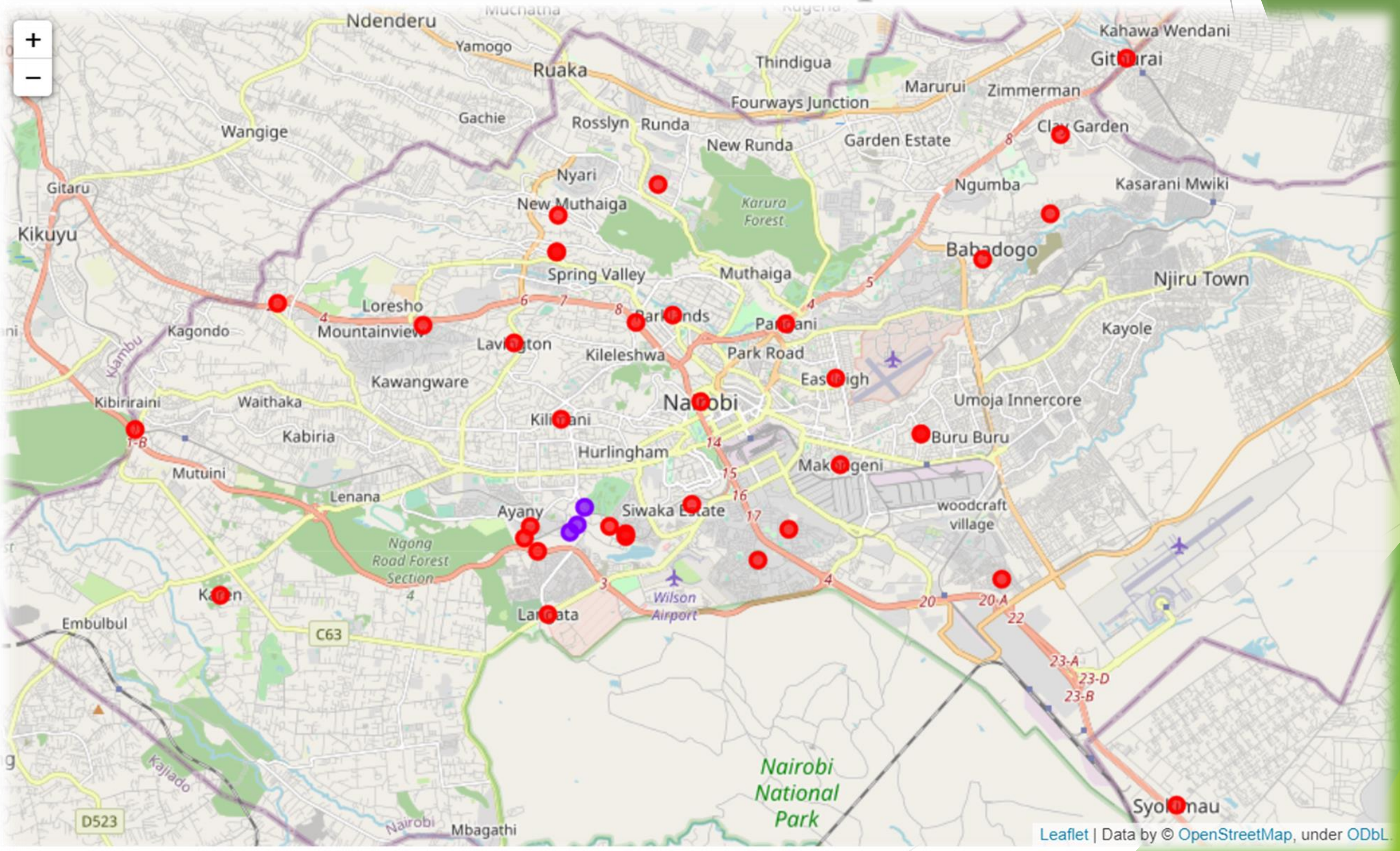


Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mc Comm Ven
Dagoretti	-1.289693	36.684983	0	Flea Market	Zoo Exhibit	Fast Food Restaurant	Cricket Ground	Deli / Bodega	Department Store	Dessert Shop	Dry Cleaner	Eastern European Restaurant	Electron Stc
Eastleigh, Nairobi	-1.277829	36.848683	0	Hotel	Men's Store	Food Court	Shopping Mall	Electronics Store	Zoo Exhibit	Cricket Ground	Deli / Bodega	Department Store	Dess Sh
Embakasi	-1.324728	36.887724	0	Lounge	Convenience Store	Fast Food Restaurant	Cricket Ground	Deli / Bodega	Department Store	Dessert Shop	Dry Cleaner	Eastern European Restaurant	Electron Stc

```
kmeans = KMeans(n_clusters=2, init='k-means++', random_state=0).fit(clustered_nairo_grp)
kmeans.labels_
```

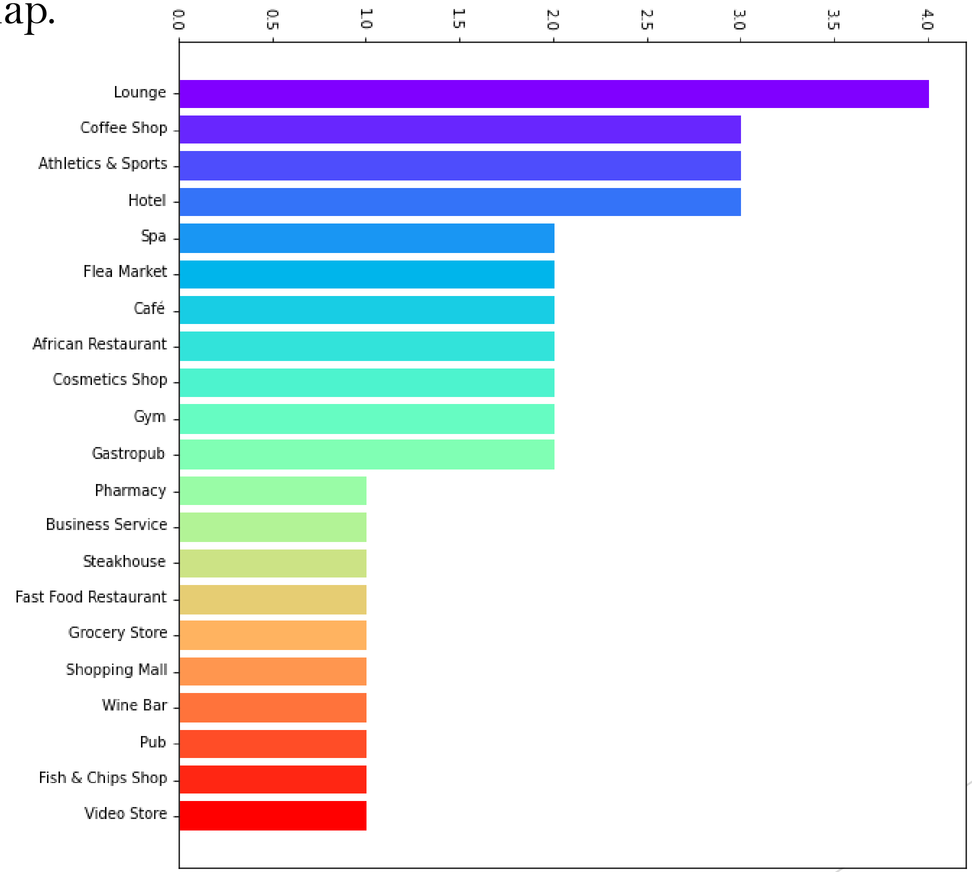
```
array([0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

Nairobi's (clustered) neighbourhoods visualised with Folium's Map



Frequency of the **first** most common venues in Nairobi's neighbourhoods

- Clustering Nairobi's neighbourhoods into only 2 divisions may have been based on an optimal score, however these clusters do not seem to provide sufficient information regarding common venue data that will be needed for further analysis. An extra step had to be taken by finding venues that appear often in the 1st most common venue
- The 'Athletics & Sports' venue is a unique case, in the sense that it all came from one labelled cluster which consisted of 3 Neighbourhoods, as shown on the previous page containing the clustered map.



Discussion

- Source: https://kenyanwallstreet.com/census-2019-datashows-kenya-has-a-youthful-rural-population/
- The conventional population of youth in Kenya aged 18 to 34 in 2019 was 13.7 million, out of which 61% were working while 1.6 million were seeking work or indicated that there was no work available. This implies youth unemployment stands at 39%

- Source: https://www.ohchr.org/Documents/Issues/Youth/D_Odondi_Kenya.pdf
- Leisure, recreation and community service are important for the psychological and physical development of the youth. It contributes to their personal development by promoting good health, personal discipline, leadership and team building skills. It also provides opportunity for appreciation, participation and creative experience in leisure, music, art, dance, drama crafts, novelty events service and cultural activities. This helps engaging the youth to make good use of their leisure time, express their beliefs and values as well as promote and preserve local art and culture for the benefit of the future youth. However, current investment in leisure and recreation has not reflected its importance. The sector suffers from inadequate funds and facilities while the talented youth lack motivation and are often exploited by organizations. Due to these constraints, it has not been possible to tap fully the talents of many youth.

- Two specific Venues, - 'Athletics & Sports' and 'Gym' – have a summed appearance of 7 times as common venues in the 1st most common venue (from the bar chart in the previous page). However, this cannot be considered to frequent enough. There also needs to be in increment of spending habits of 'Sports and Leisure' which is currently standing at 8.5%

Conclusion

With our core objective being able to create more jobs for the youth, as well as trying to divert spending habits to the younger generation (both dependants and independent individuals), an effective way of achieving this desired goal is by creating retail business shops which sell sportswear to appeal both male and female individuals who are in their teenage years, up until those who are in their late youth.

More Gym and Fitness centres can also be developed as this will not only aid in creating jobs, but it is also a benefit to one's health both psychologically and physically.

More Gym centres, which would result in more clients wanting to join a certain club, will therefore result in an increase in demand of consumables that aid to an individual's fitness, e.g. Protein shakes. More retail shops can also be established in order to accommodate this demand, and thus creates more jobs.