NAIROBI'S NEIGHBOURHOODS: 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.

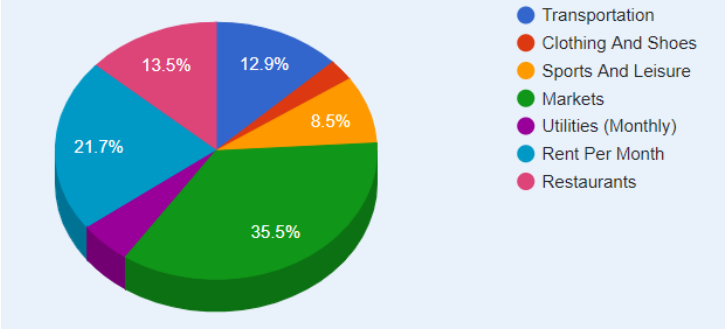
The city of Nairobi is growing consistently and currently stretches over an area of of 696 kilometres squared (269 square miles). This area size - in combination with the total number of residents - brings us the current population density which is now approximately 4,850 residents per square kilometre.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

POPULATION SIZE AND DENSITY

The city of Nairobi is growing consistently and currently stretches out over a surface area of of 696 kilometres 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 kilometre. (12,600 people living per square mile).

NAIROBI'S LIVING COST AND SPENDING HABITS. (A SUMMED-UP BRIEFING)

- 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 labour 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 ACQUISITION 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 neighbourhoods.

- 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/

METHODOLOGY

GEOCODER & FOLIUM

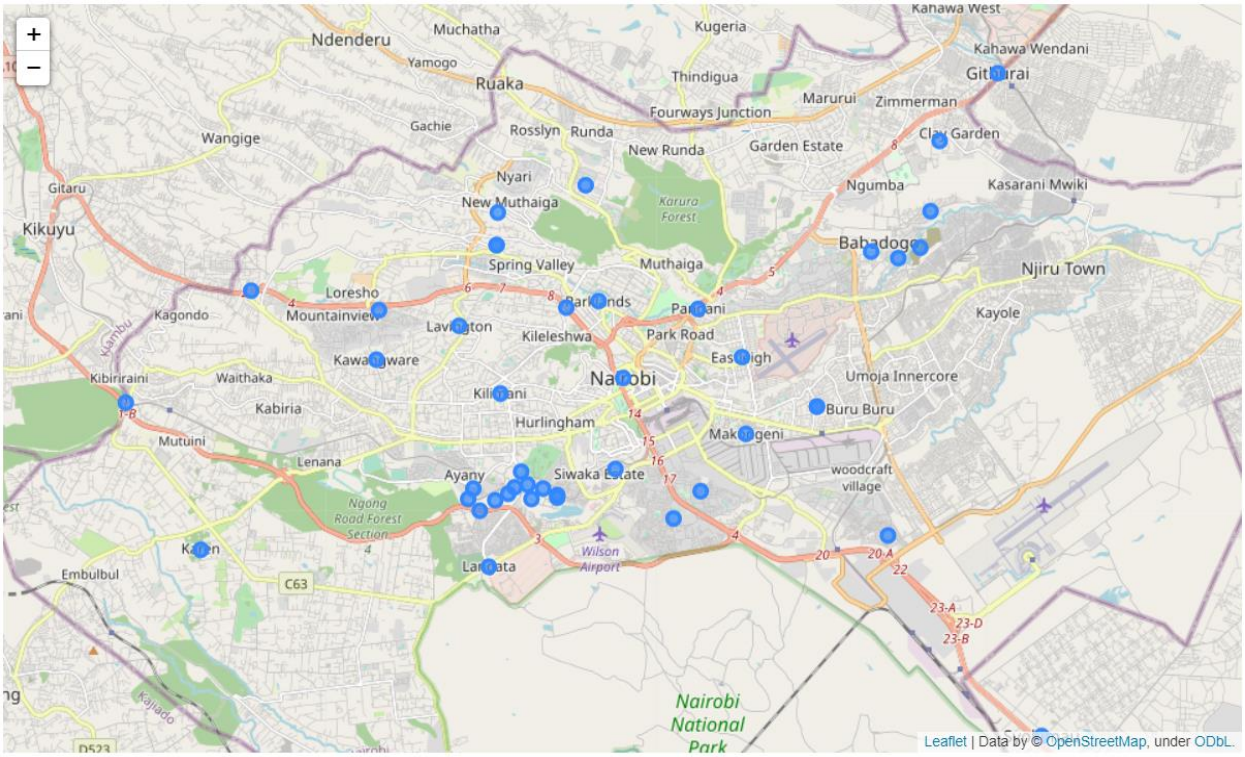
The coordinates of Nairobi were acquired through nominatim's geocoding service and were used to create a map of the city through folium's map. The same method to locate the coordinates of the neighbourhoods within

Nairobi was also implemented. Unfortunately, not all neighbourhoods could be located properly because the process of converting their addresses to coordinates was somewhat limited by the geocoder, and as a result, not all listed neighbourhoods were mapped.

```
adress = 'Nairobi, Kenya'
location = geolocator.geocode(adress)
latitude = location.latitude
longitude = location.longitude

print(f'Nairobi coordinates are {latitude},{longitude}')
```

```
Nairobi coordinates are -1.2832533,36.8172449
```

```
nairobi_map = folium.Map(location=[latitude,longitude], zoom_start=12.)

for lat,long,neigh in zip(data_filt['latitude'], data_filt['longitude'], data_filt['Neighborhood']):
    label = f'{neigh}'
    label = folium.Popup(label,parse_html=True)
    folium.CircleMarker([lat,long], radius=5,popup=label,
                        colour = 'cyan', fill=True,fill_colour ='#3186cc',
                        fill_opacity=0.7,).add_to(nairobi_map)
```

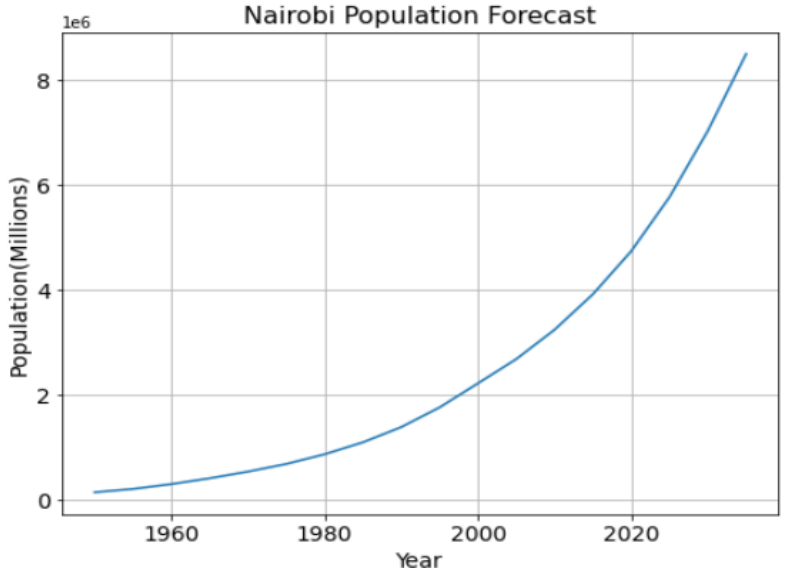


NAIROBI'S POPULATION

A visual on how Nairobi's population is constantly increasing is needed to emphasise the importance of how this will impact the unemployment rate, especially on the youth, considering the fact that a large percentage of Nairobi's population is young.

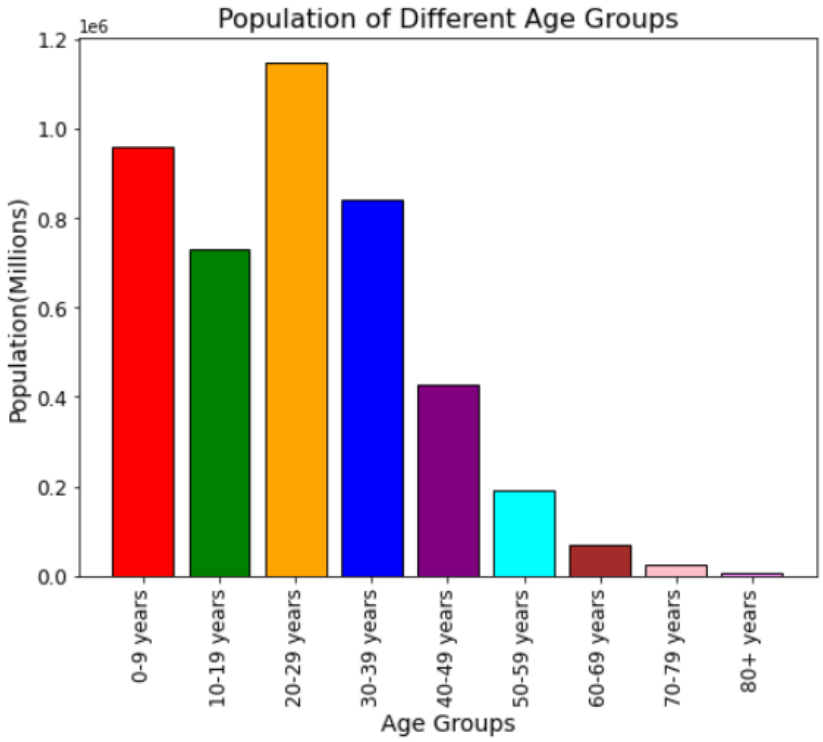
	Year	Population	GrowthRate	Growth
0	2035	8499403	0.0387	1468512
1	2030	7030891	0.0404	1263902
2	2025	5766989	0.0402	1032108
3	2020	4734881	0.0388	821369
4	2015	3913512	0.0387	676923

```
plt.figure(figsize=[8,6])
plt.plot('Year','Population', data=population_forecast)
plt.grid()
plt.title('Nairobi Population Forecast', fontsize=16)
plt.xlabel('Year', fontsize = 14)
plt.ylabel('Population(Millions)', fontsize=14)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.show()
```



	Age Distribution (2019)	Population
0	0-9 years	957699
1	10-19 years	730403
2	20-29 years	1146567
3	30-39 years	841266
4	40-49 years	428094

```
plt.figure(figsize=[8,6])
colours = ['red','green','orange','blue','purple','cyan','brown','pink','violet']
plt.bar(age_dist['Age Distribution (2019)'],age_dist['Population'], color= colours, edgecolor='Black')
plt.xticks(fontsize = 12, rotation = 90)
plt.yticks(fontsize = 12)
plt.title('Population of Different Age Groups', fontsize=16)
plt.xlabel('Age Groups', fontsize=14)
plt.ylabel('Population(Millions)', fontsize=14)
plt.show()
```



FOURSQUARE'S API

Using foursquare personal client credentials, nearby venues within a 500m radius of each neighbourhood were established. However, this calling function was, once again, limited in some neighbourhoods, reason being that some of them did not have any nearby venues within their 500m radii. I attempted using different circumferential values to try get more data on common venues, but a 500m proved to be the most ideal option.

```
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: REMOVED FOR PRIVACY
CLIENT_SECRET:REMOVED FOR PRIVACY
```

```
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={}'.format(
            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)
```

ONE HOT ENCODING

This refers to splitting the column which contains numerical categorical data to many columns depending on the number of categories present in that column. Each column contains **0** or **1** corresponding to which column it has been placed. In the case of the established nearby venues, 88 of these were formatted into one hot encoding. The rows were then grouped by neighbourhood and by taking the mean frequency of occurrence of each category.

```
nairobi_onehot = pd.get_dummies(nairobi_venues[['Venue Category']], prefix="", prefix_sep="")
nairobi_onehot['Neighborhood'] = nairobi_venues['Neighborhood']
fixed_columns = [nairobi_onehot.columns[-1]] + list(nairobi_onehot.columns[:-1])
nairobi_onehot = nairobi_onehot[fixed_columns]
nairobi_onehot.head(3)
```

	Neighborhood	African Restaurant	Arcade	Athletics & Sports	BBQ Joint	Bakery	Bar	Beer Garden	Bistro	Bookstore	...	Szechuan Restaurant	Tapas Restaurant	Tea Room	Tourist Information Center	Trail	T 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

```
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	S
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	

VENUE FREQUENCY

Creating a Data Frame of the top 10 most common venues in each neighbourhood.

```
def most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

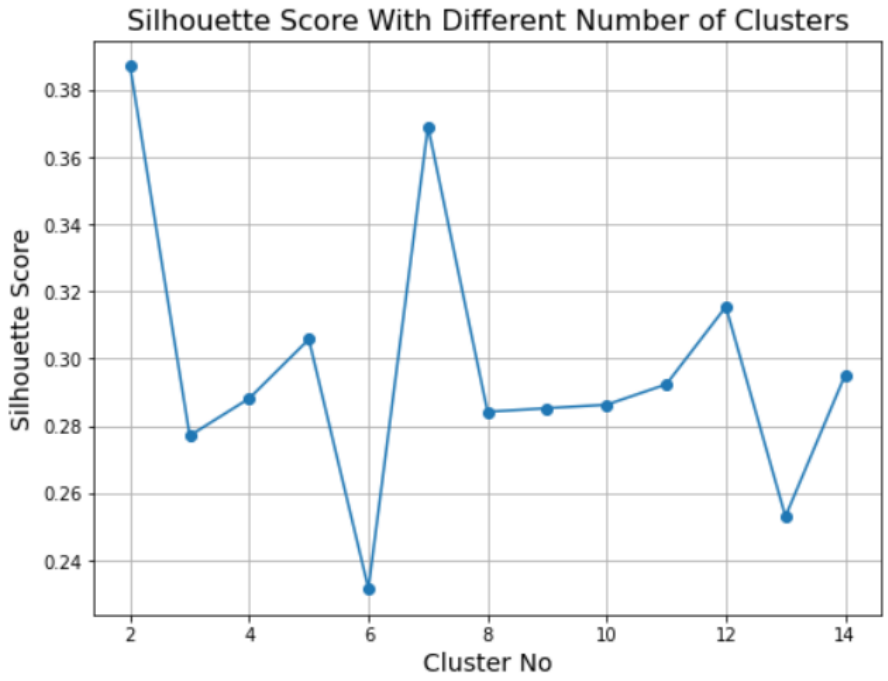
```
num_top_venues = 10
indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append(f'{ind+1}{indicators[ind]} Most Common Venue')
    except:
        columns.append(f'{ind+1}th Most Common Venue')

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = nairobi_grp['Neighborhood']

for ind in np.arange(nairobi_grp.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = most_common_venues(nairobi_grp.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

	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

SILHOUETTE SCORE TO FIND THE OPTIMAL NUMBER OF CLUSTERS

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). The optimal number of clusters to be used in KMeans in this case is 2.

```
plt.figure(figsize=[8,6])
plt.plot(indices, scores, 'o-')
plt.title('Silhouette Score With Different Number of Clusters', fontsize = 16)
plt.xlabel('Cluster No', fontsize=14)
plt.ylabel('Silhouette Score', fontsize=14)
plt.grid()
plt.show()
```



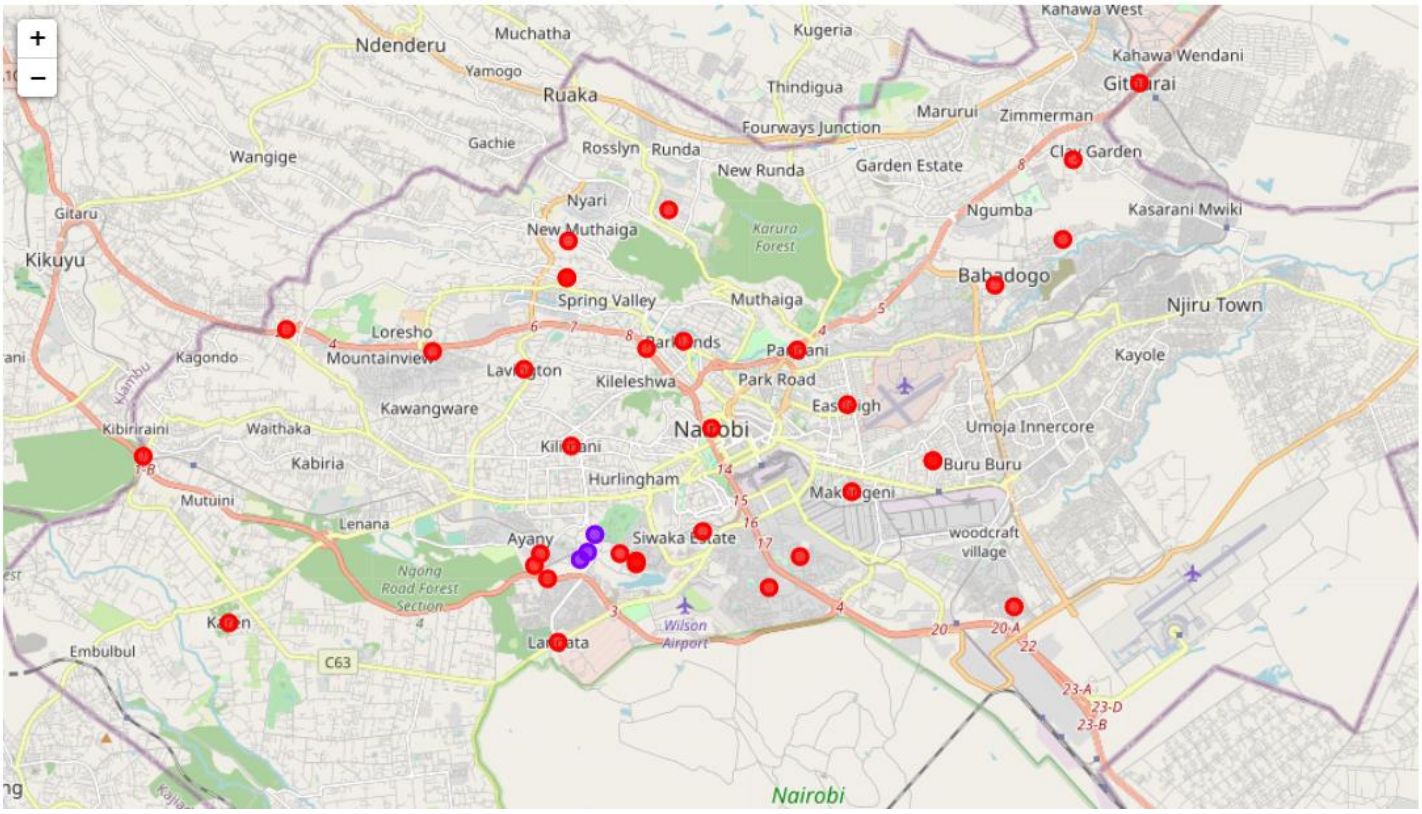

```
kmeans = KMeans(n_clusters=2, init='k-means++', random_state=0).fit(clustered_nairo_grp)
kmeans.labels_
```

```
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, 0])
```

The venue data was trained using K-means clustering algorithm as the venue category features are quite a lot, and in such situations K- means will be computationally faster than other algorithms.

VISUALISING THE NIGHBOURHOODS IN ARCCODANCE TO THEIR LABELLED CLUSTERS

```
markers_colors = []
for lat, lon, poi, cluster in zip(nrb_merged_clean['latitude'], nrb_merged_clean['longitude'], nrb_merged_clean['Neighborhood'], nrb
    label = folium.Popup(str(poi) + ' (Cluster ' + str(cluster+1) +')', parse_html=True)
    map_clusters.add_child(folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7))

map_clusters
```

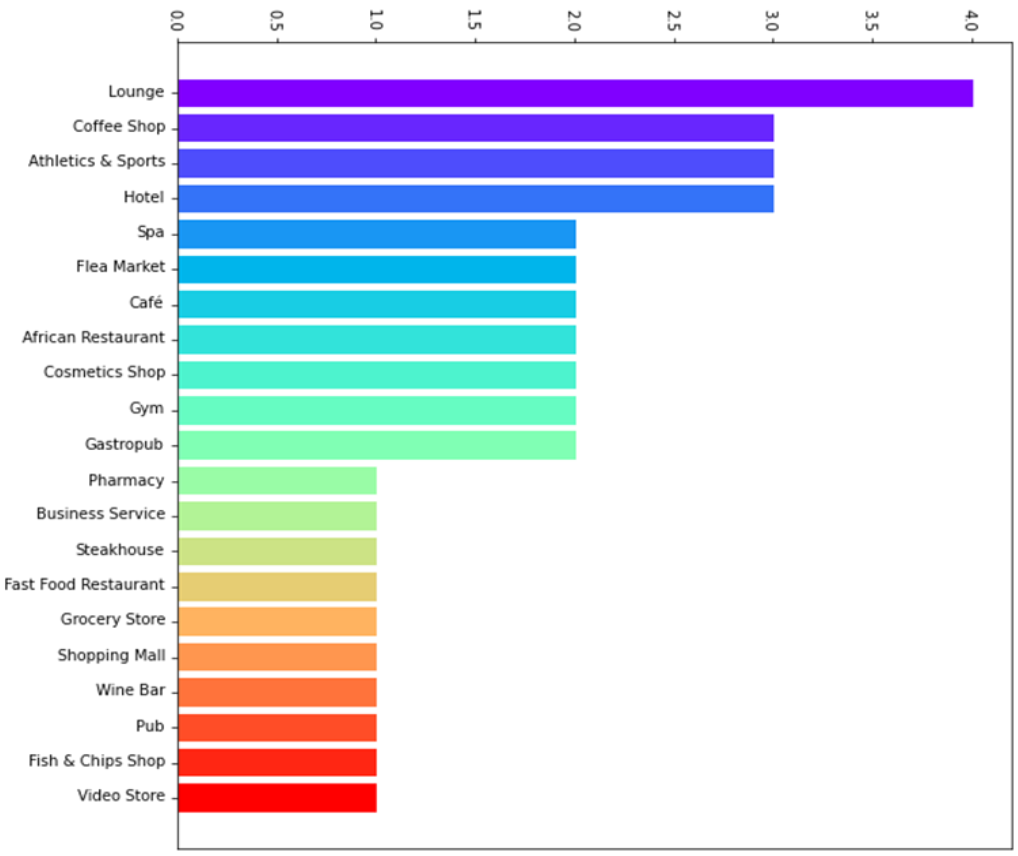


RESULTS

- 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.

- Two specific Venues, - 'Athletics & Sports' and 'Gym' – have a summed appearance of 7 times as common venues in the 1st most common venue. These two venues will be important in trying to find a solution to the problem regarding youth unemployment and spending habits.



DISCUSSION

- 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://KENYANWALLSTREET.COM/CENSUS-2019-DATASHOWS-KENYA-HAS-A-YOUTHFUL-RURAL-POPULATION/

- 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.

SOURCE: HTTPS://WWW.OHCHR.ORG/DOCUMENTS/ISSUES/YOUTH/D_ODONDI_KENYA.PDF

- The two *'Athletics & Sports' and 'Gym'* venues have a summed appearance of 7 times as common venues in the 1st most common

venue (from the bar chart in the results section). 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% (from the introductory section).

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.