1.Introduction

1.1 Background

Tanzania is a country in East Africa within the African Great Lakes region. It borders Uganda to the north; Kenya to the northeast; Comoro Islands and the Indian Ocean to the east; Mozambique and Malawi to the south; Zambia to the southwest; and Rwanda, Burundi, and the Democratic Republic of the Congo to the west. Mount Kilimanjaro, Africa's highest mountain, is in northeastern Tanzania.

The population distribution in Tanzania is extremely uneven. Most people live on the northern border or the eastern coast, with much of the remainder of the country being sparsely populated. Density varies from 12 per square kilometer (31/sq mi) in the Katavi Region to 3,133 per square kilometer (8,110/sq mi) in Dar es Salaam. Approximately 70 percent of the population is rural, although this percentage has been declining since at least 1967

According to the 2012 census, the total population was 44,928,923 compared to 12,313,469 in 1967, resulting in an annual growth rate of 2.9 percent. The under 15 age group represented 44.1 percent of the population, with 35.5 percent being in the 15–35 age group, 52.2 percent being in the 15–64 age group, and 3.8 percent being older than 64. [1]

1.2 Problem

According to the Wikipedia [2] and [3], most of the regions in Tanzania, though not all, with high number of people (according to census done in 2012), contribute much to the national GDP. This project aims to find out the relationship between the population and national GDP and also explore some of the available markets in one of the high GDP contributing region so as to enable different stakeholders who are looking for where they can sell their products.

2. Data acquisition and cleaning

2.1 Data sources

Several sources were used for data collection. List of regions of Tanzania by GDP and population in each regions of Tanzania data was scraped from Wikipedia [2] [3], while the geographical

coordinates of each region were obtained from [4]. The coordinates were disorganized, so after obtaining the coordinated of each region, the CSV file was prepared which contained the list of all regions and thein coordinates. The geographical coordinates of the highest contributing region to the national GDP were utilized to collect venues data within 5000 m radius using Foursquare API.

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table. The regions such as Songwe, Geita, Simiyu, Katavi and others, which came into existence after the census done in 2012 were removed since there were no population data about them.

The remaining data were used to examine the relationship between the population trend in each region according to the census done in 2012 and the contribution of that region to the national GDP.

2.3 Feature selection

After data cleaning, there were 22 samples and 10 features in the data. Some features were kept for further usages but other features were discarded. The unrequired features were dropped because they don't provide any necessary information required for task completion. Table 1, shows the dropped features.

Table 1 Shows the Features dropped

S/No.	Feature dropped
1.	Rank
2.	GDP in mil USD (PPP)
3.	Equivalent country[3]

After dropping the unimportant features, the required features were combined so as to form a single table as shown in Table 2.

Table 2 Shows the Remained useful features for task completion

	Region	GDP in mil TZS	Population (2012)	Zone	Latitude	Longitude
0	Dar es Salaam	22577225	4364541	Coastal	-6.82349	39.26951
1	Mwanza	12731454	2772509	Lake	-2.51667	32.90000
2	Shinyanga	7540589	1534808	Lake	-3.66667	32.91667
3	Mbeya	7314302	2707410	Southern Highlands	-8.90000	33.45000
4	Morogoro	6191343	2218492	Coastal	-7.91667	37.25000

3. Methodology

In this project we will examine the relationship between number of people and the GDP. Also, we will use Foursquare to explore the most contributing region (Dar es Salaam as shown in Table 2) to the national GDP in Tanzania. The exploration will be limited to 5000 meters from Bibi Titi street in Dar es Salaam.

In the first step the data exploratory will be done so as to examine and visualize the relationship between the number of populations in each region in Tanzania according to the census done in 2012 and the contribution of each region to the national GDP.

The second step will involve the process of putting all the regions on the map and the number of people in each region.

The final step will involve the clustering of the regions by considering the number of people they contain and their contributions to the national GDP.

3.1 Exploratory Data Analysis

3.1.1 Relationship between population and GDP

The data were analyzed so as to examine the relationship between the number of people in a region and the region’s contribution to the national GDP. Figure 1 and Figure 2, shows the results.

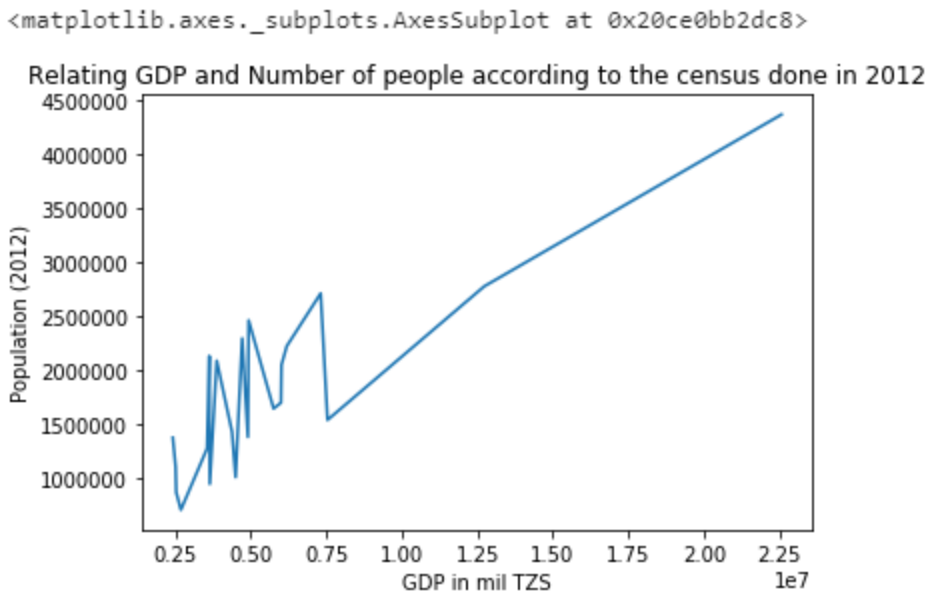

Figure 1 Shows the relationship between population and GDP in Tanzania

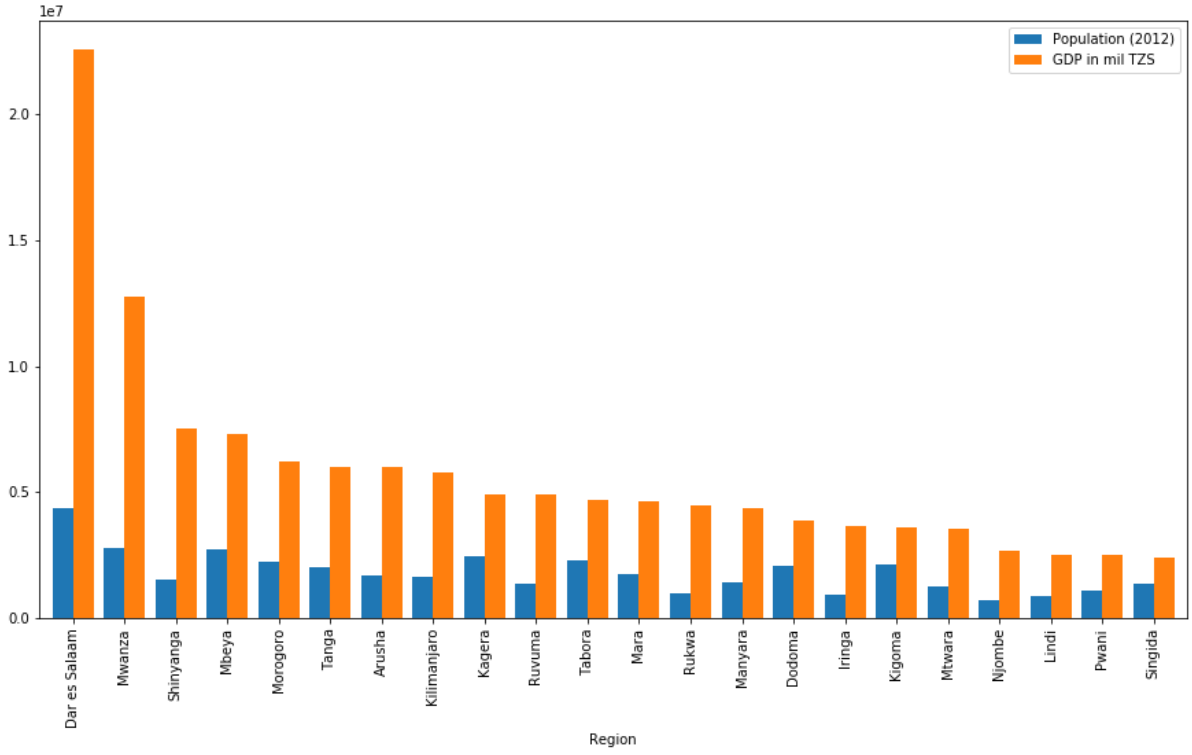

Figure 2 Shows the relationship between population and GDP in each region in Tanzania

3.2 Regions view

The map was used so as to enable the visualization of each region's location in Tanzania and their population (according to the census done in 2012). Figure 3 shows the results.

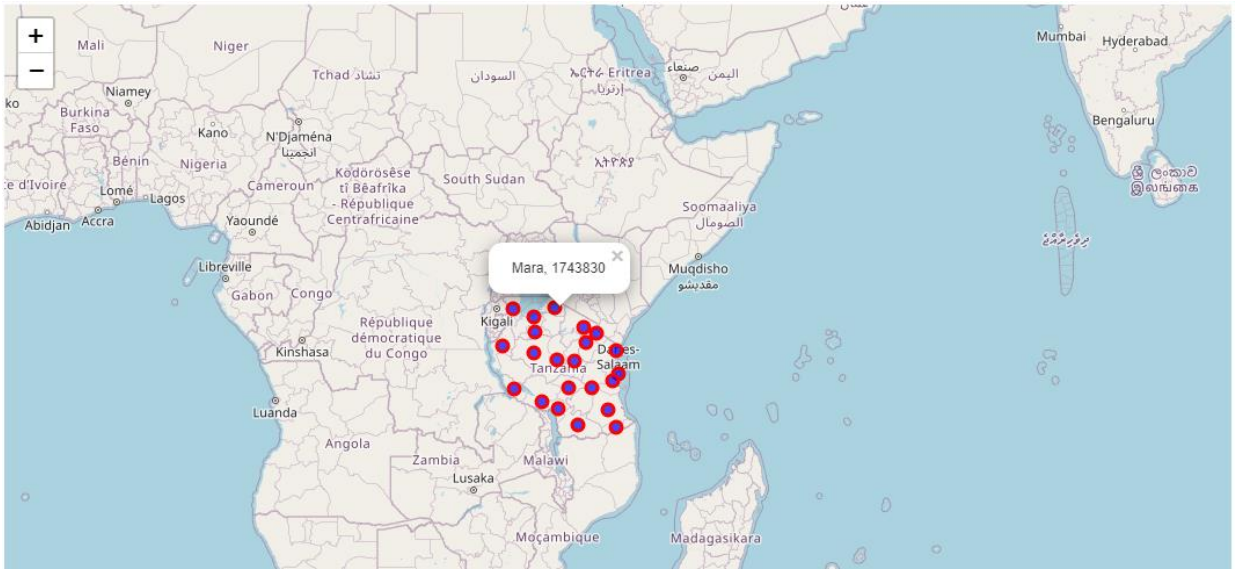

Figure 3 Shows the location of each region and population it contains

3.3 Clustering the Regions

This project uses clustering since it is the most common exploratory data analysis technique used to get an intuition about the structure of the data. Clustering helped to identify the subgroups available by considering the available data. In this project **KMeans**, which is the most used clustering algorithm was used to divide the data into non-overlapping subjects (clusters). The clusters were divided by considering the population size and the region's contribution to the national GDP. Table 3, shows the data which also contains the cluster labels, and, the Figure 4, shows the results on the map.

Table 3 Shows the dataset and the cluster labels

	Cluster Labels	Region	GDP in mil TZS	Population (2012)	Zone	Latitude	Longitude
0	1	Dar es Salaam	22577225	4364541	Coastal	-6.82349	39.26951
1	2	Mwanza	12731454	2772509	Lake	-2.51667	32.90000
2	4	Shinyanga	7540589	1534808	Lake	-3.66667	32.91667
3	4	Mbeya	7314302	2707410	Southern Highlands	-8.90000	33.45000
4	4	Morogoro	6191343	2218492	Coastal	-7.91667	37.25000

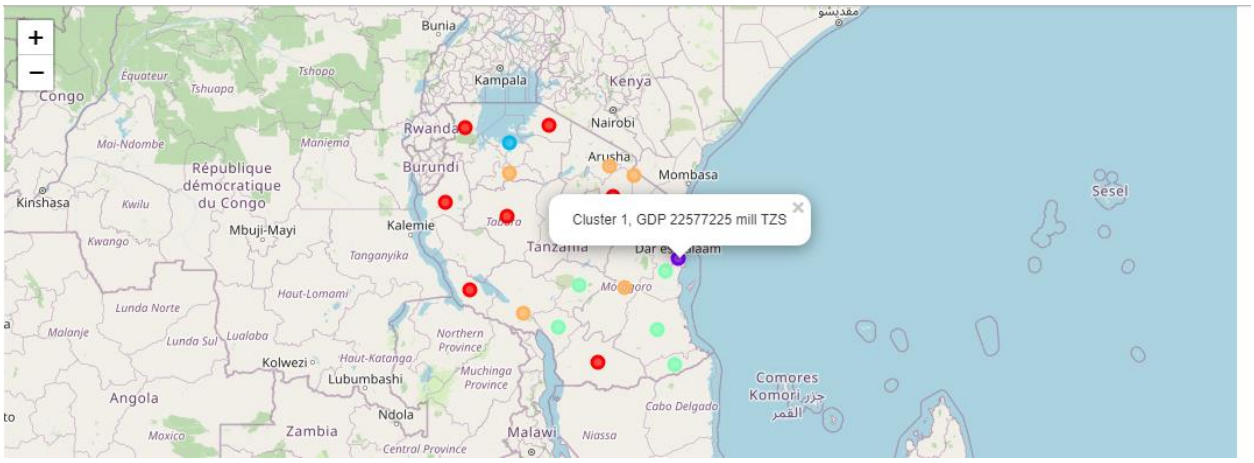

Figure 4 Shows the clustering results

The city of Dar es Salaam seems city to be the most populated region in Tanzania (4364541 people in 2012) and the highest region that contribute to the national GDP, is the most attractive city which many stakeholders target to sell their products and run their business activities and create more profits.

This project uses **Foursquare** to explore the available markets within a range of 5000 meters from city center. The request was sent using the Foursquare API and the results were analyzed and the relevant information were extracted and transformed into pandas Data Frame as shown in Table 4.

Table 4 Shows the Foursquare search results

	name	categories	id	location
0	Kisutu Market	Market	5198eb12498edba16e1d4dec	{'lat': -6.817484192449897, 'lng': 39.28042191...
1	Kisutu Market	Miscellaneous Shop	4e43cdd71f6e0a1ba5d503a5	{'lat': -6.817233932077075, 'lng': 39.28171989...
2	Kivukoni Fish Market	Fish Market	50445f37e4b068ce7e21da99	{'lat': -6.8189232426173625, 'lng': 39.2997570...
3	Kariakoo Market	Market	4e44f6aa7d8bee2823a3651c	{'lat': -6.8189300322379625, 'lng': 39.2752145...
4	Dallas Mini Market	Grocery Store	501857a0e4b0072c28d91b1c	{'address': 'Dar Es Salaam', 'crossStreet': 'L...

The location column was separated from the data frame and analyzed further so as to extract the coordinates of each market identified. After getting all of the required coordinates, the result was then merged to the data frame as shown in the Table 5.

Table 5 Shows the final merged result

	name	categories	id	lat	lng
0	Kisutu Market	Market	5198eb12498edba16e1d4dec	-6.817484	39.280422
1	Kisutu Market	Miscellaneous Shop	4e43cdd71f6e0a1ba5d503a5	-6.817234	39.281720
2	Kivukoni Fish Market	Fish Market	50445f37e4b068ce7e21da99	-6.818923	39.299757
3	Kariakoo Market	Market	4e44f6aa7d8bee2823a3651c	-6.818930	39.275215
4	Dallas Mini Market	Grocery Store	501857a0e4b0072c28d91b1c	-6.815720	39.283058

The map was then used to visualize the location results of the discovered markets which are located within a range of 5000 meters from the city center. Figure 5, shows the map consisting of all the discovered markets.

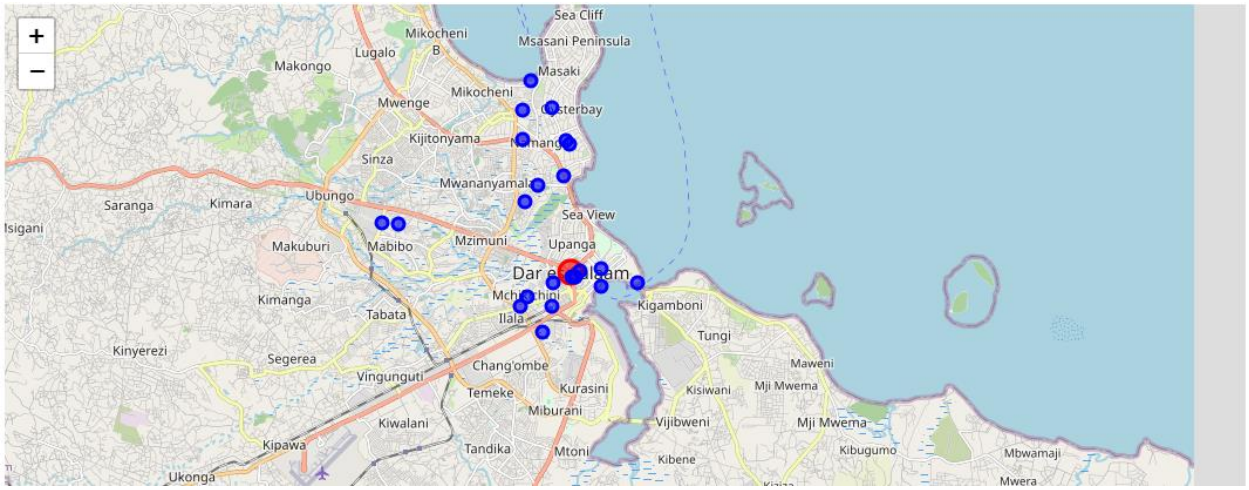

Figure 5 shows the map consisting of all the discovered markets

4. Results and Discussion

4.1 Results

The result obtained reveals that, the national GDP is mostly depends on the number of people available in each region of Tanzania, though not all. This means there are other factors that might affect the national GDP when combined with the population factor.

4.2 Discussion

There could be more venues as the foursquare API only gives venues based on the geographical coordinates, set limit, and the radius. The number of venues could be increased by increasing the radius, limit, and changing the latitude and longitude values.

Since the population data used was that of 2012, the current population now could be higher than the expected. This might bring some changes on the results. High availability of large market places in a region, reveals that there is a lot of business processes conducted around those areas. I would like to recommend that, the government of Tanzania should consider to build and create enough market places so as to help the growth of business processes in each region and finally the growth of the national GDP.

5. Conclusion

In this study, I analyzed the relationship between population in each region in Tanzania and the contribution of the country to the national GDP. The analysis went further and picked Dar es Salaam as it is the most populated region in Tanzania and it contribute much to the national GDP than other regions. One of the factors which lead to high GDP is the availability of the places where the stakeholders can run their businesses, and one of them is the market places.

The availability of large markets such as Kariakoo market, Kisutu market and others in Dar es Salaam, have made the region to highly contribute to the national GDP due to the businesses conducted.

References

[1] "Demographics of Tanzania".

[2] "List of regions of Tanzania by GDP".

[3] "Regions of Tanzania".

[4] Latitude.to, "regions," 2020. [Online]. Available: https://latitude.to/map/tz/tanzania/regions.