Demographic Projection & Venues Data Analysis of Dakar – SENEGAL

Amadou SOULEY

April 23, 2019

A. Introduction

Before describing our business problem and its interest, we first go to discover a historical city in Africa, the city of Dakar, which will be the subject of study for this project. The city of Dakar (*in* ***Wolof: Ndakaaru***) is the capital of the Republic of Senegal and the Dakar region. It has ***1,363,444*** *inhabitants* out of a total *population of* ***3,732,284*** (2019 estimate) in the Dakar region as a whole. It is one of Senegal's four historic municipalities and the former capital of *French West Africa (AOF)*. Under the dual influence of migration flows from the countryside and natural increase, the Dakar region has developed very rapidly. It has thus increased from ***400,000*** *inhabitants in* ***1970 to 3.6 million****inhabitants in 2018, an increase of nearly* ***5% per year***. Occupying only *0.28% of the national territory*, the Dakar region covers ***550 km2, 25%*** *of the population* and concentrates ***80%*** *of the country's economic activities*.

As part of its activities, the ***National Agency for Statistics and Demography*** (ANSD – Dakar) has published its estimates of Senegal's population by ***2025***. When a municipality sees its population increase, should it be considered, a priori, as an advantage or a disadvantage? Whatever the cause, the decrease in the population of a municipality very generally has financial consequences in terms of budget. On the one hand, the financial allocations paid by the State, which are based in particular on the number of inhabitants, are expected to decrease.

A.1 Problem

Indeed, when you think about it, demographic has an impact on pension and health insurance. In addition for the bourgeois class you expect them to prefer the chic neighborhoods but also the less

populated districts. At the same time, they may want to choose neighborhoods based on the density of social places. This can also be the choice of large companies that want to set up in the city.

A.2 Interest

The target audience and who would care about these problems are specially ***the authorities of each neighborhood, the United Nations*** but also ***the Insurers, investors*** as well as ***the principales players*** *in the economy*. However, it is difficult today to obtain information that can guide them in this direction.

To address these business issues we will present the current demographics as well as its projection for 2025 by gender for each neighborhood of this city. In order to know the popularity of each borough, we will analyze the venues of Dakar city. We will create a map and information chart where the real percentage by sex is placed on Dakar city and each neighborhood is clustered according to the venue density.

B. Data Description

To examine the problem, we can list the data as follows:

- To have the Dakar Neighborhood Data the following features was extracted: ***Borough, Neighborhood, Men_2019, Women_2019, Men_2025, Women_2025****(Name_year contain the number of population by sex and year and these 2025 population projection) and the geospatial coordinates L****atitude****and* ***Longitude***.
- For the features Borough, Neighborhood, I required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format.On the website of ***the National Agency for Statistics and Demography ANSD***, for each neighborhood of the city,I had to retrieve the demographic projections by gender and year and its ***2025 demographic projection***.
- I used ***Google Map, 'Search Nearby'*** option to get the geospatial coordinates latitude and Longitude of each Neighborhood. Coordinate of Dakar Center will be obtained by using ***Goog Maps API geocoding*** of well known location.

- I used ***Forsquare API*** to get the most common venues of given Neighborhood of Dakar city.
- Finally I will explore and cluster the neighborhoods of Dakar.

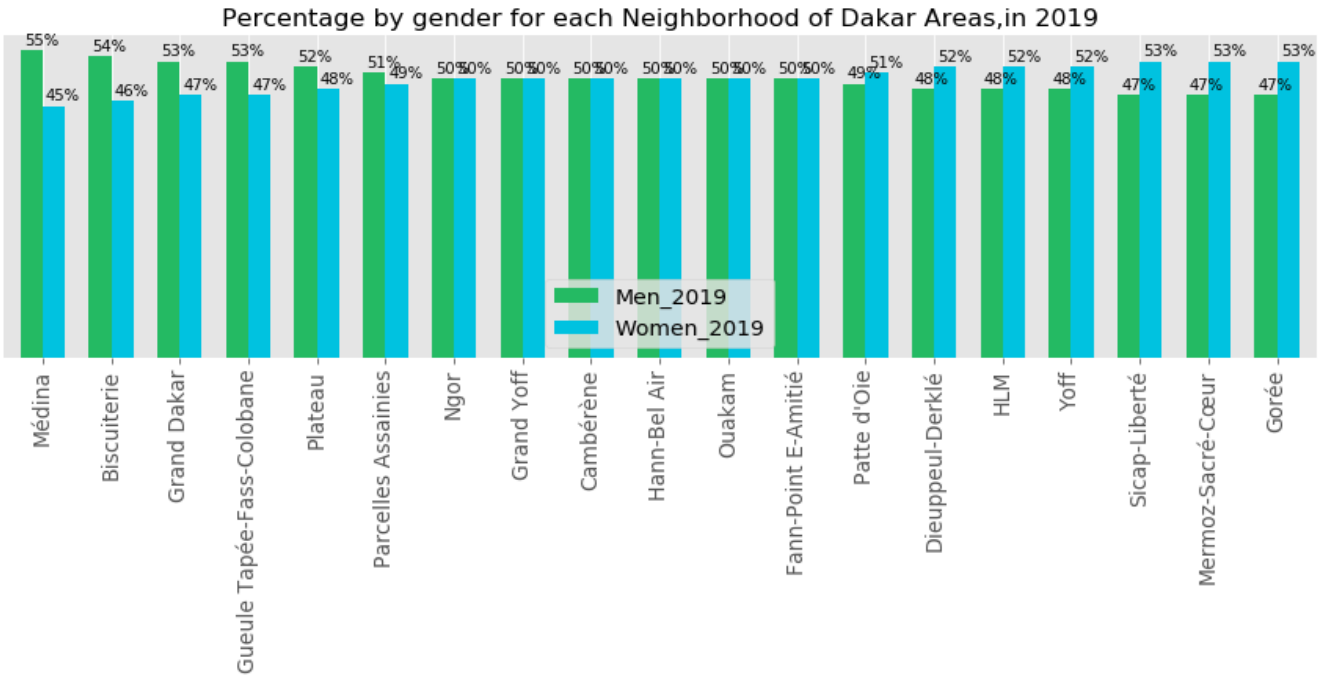
C. **Methodology**

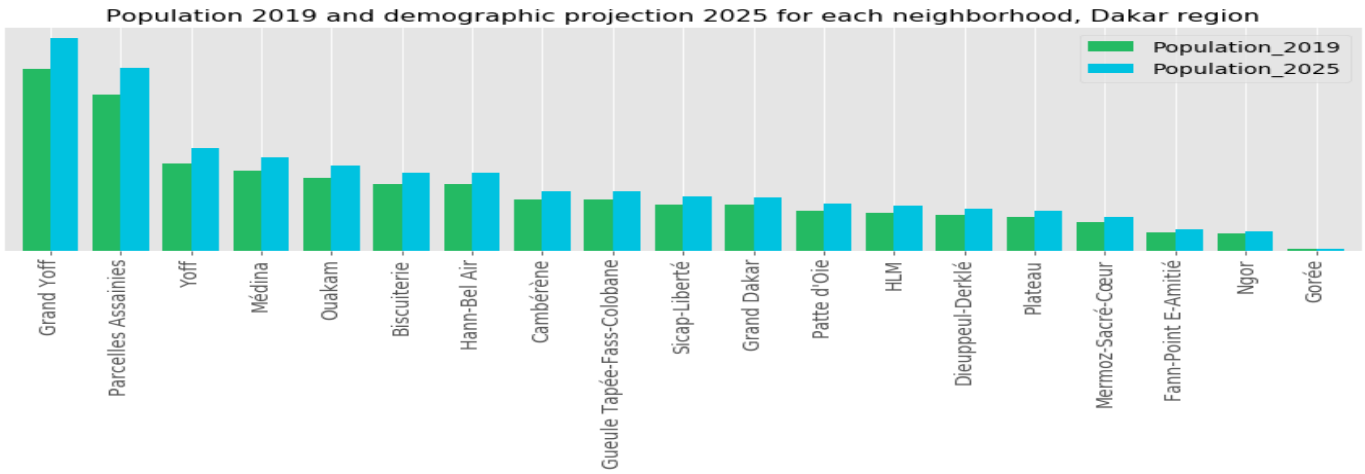
C.1 Population Projection 2025

After wrangle the data, clean it, and read it into a *pandas dataframe*. I obtained the following database:

	Neighborhood	Borough	Men_2019	Women_2019	Men_2025	Women_2025	Latitude	Longitude
0	Mermoz-Sacré-Cœur	Almadies	16640	18810	19424	21957	14.712995	-17.471159
1	Ngor	Almadies	10391	10290	12129	12011	14.748792	-17.514961
2	Ouakam	Almadies	44281	44579	51689	52037	14.724737	-17.485066

I used ***Matplotlib*** to visualize the percentage of each neighbourhood in Dakar by gender and year as can be seen here:



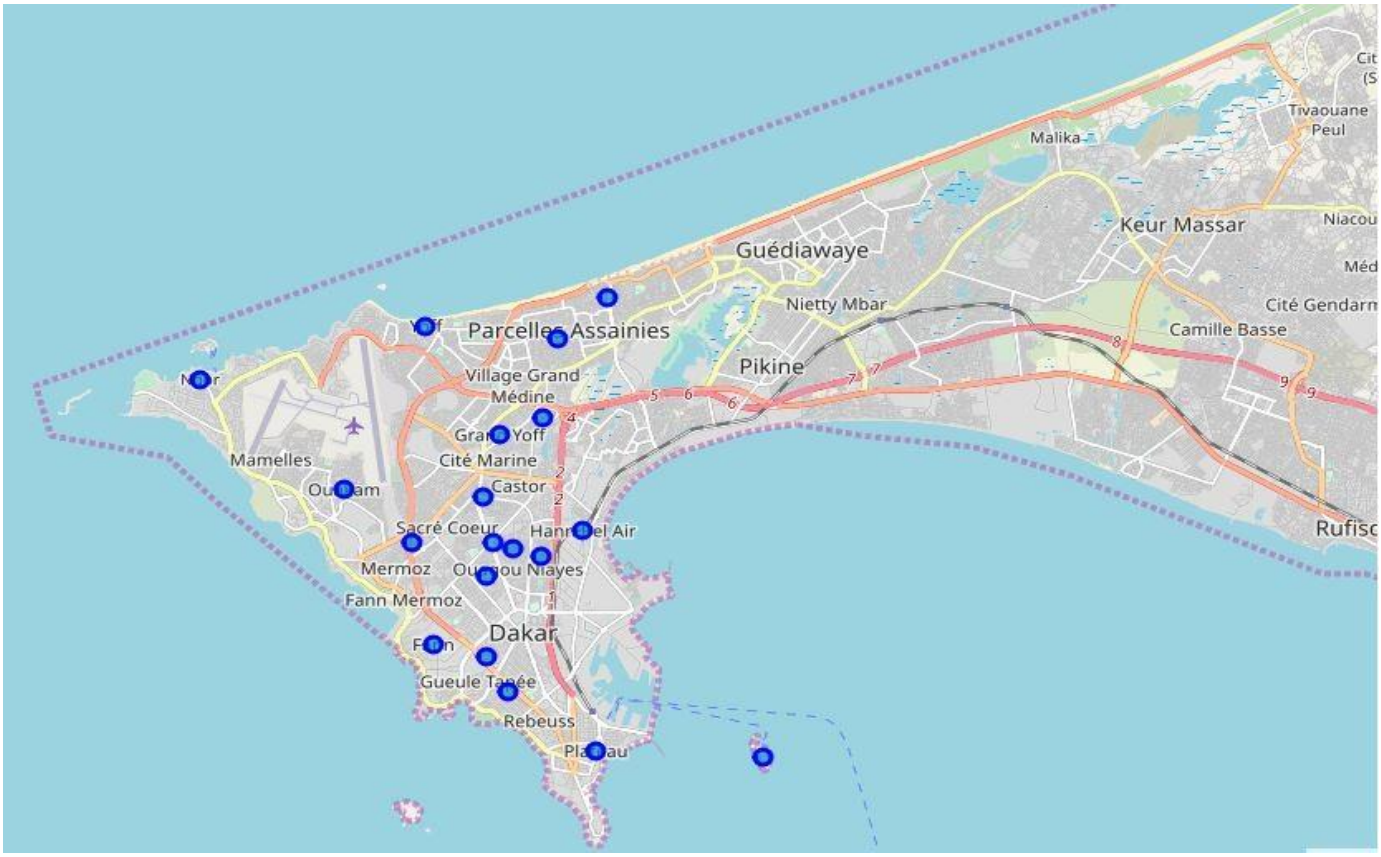


According to the graphs above, there is a clear increase in the population and women will outnumber men in each Dakar's neighborhood in 2025. Except Cambérene and Grand Yoff where we saw an equal number of women and men. Below we can see the demographic rate in each Dakar's borough:

	Borough	Neighborhood	Population_2019	Population_2025	demographic_rate
0	Almadies	Ngor,Mermoz-Sacré-Cœur,Yoff,Ouakam	251399	293456	0.167292
1	Dakar Plateau	Gorée,Plateau,Médina,Fann-Point E-Amitié,Gueul...	225428	263142	0.167300
2	Grand Dakar	Sicap-Liberté,Biscuiterie,Grand Dakar,Dieuppeu...	364907	425956	0.167300
3	Parcelles Assainies	Cambérène,Patte d'Oie,Parcelles Assainies,Gran...	521711	608992	0.167298

C.2. Venues Data Analysis

To analyze venues, I used ***python folium library*** to visualize geographic details of Dakar and its neighborhoods and I created a map of Dakar with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:

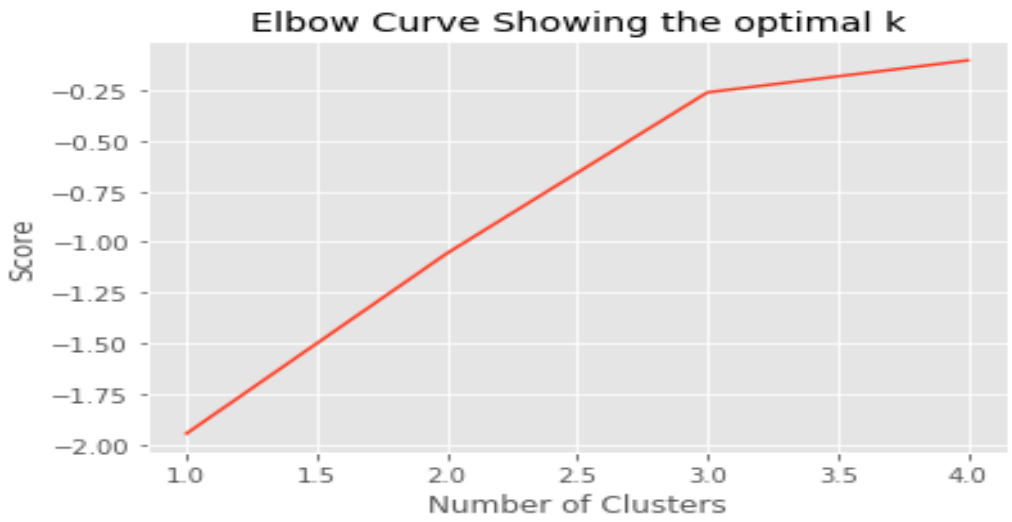


I used the ***Foursquare API*** to explore **Dakar Plateau neighborhoods** and segment them. I designed the limit as **100 venues and the radius 600 meter** for each neighborhood from their given latitude and longitude informations. I obtained 42 sites including 33 unique categories distributed as follows: 31 in Plateau, 5 in Gorée, 4 in Guele Tapée-Fass-Colobane, 1 in Fann-Point E-Amitié and 1 in Medina.. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Fann-Point E-Amitié	14.690791	-17.466559	Tennis & Squash(UCAD)	14.687146	-17.468209	Tennis Court
1	Gorée	14.666410	-17.398279	Mémorial de Gorée du Castel	14.666089	-17.398053	History Museum

The result doesn't mean that inquiry run all the possible results in Dakar Plateau Neighborhoods. We can increase the possibilities with Neighborhood informations and more latitude and longitude informations.

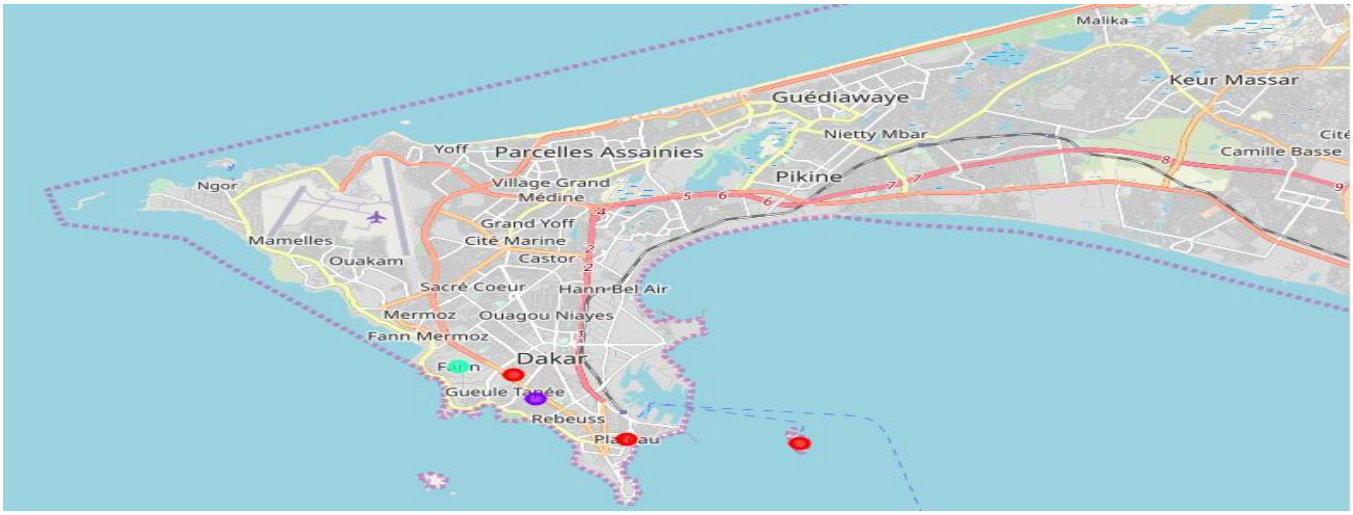
I have created a new dataframe and I displayed the 10 top venues for each district of the Dakar Plateau. According to the result, in the neighbourhoods, there are a few categories of common venues . For that , as I want to cluster the neighborhoods, I used one of the most common cluster method of unsupervised learning: K-means algorithm . First I check the optimal k with the elbow method as follows:



The optimal k is therefore 3. It is important to note that the elbow method is only one approach among others that can allow us to decide on the optimal k. Here is my table merged with the cluster labels and the 10 top venues for each neighborhood:

	Neighborhood	Borough	Latitude	Longitude	Population_2019	Population_2025	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Con Ven
4	Fann-Point E-Amitié	Dakar Plateau	14.690791	-17.466559	22415	26164	2	Tennis Court	Tea Room	Ame Res
5	Gorée	Dakar Plateau	14.666410	-17.398279	1998	2333	0	Boat or Ferry	History Museum	Hote
6	Gueule Tapée-Fass-Colobane	Dakar Plateau	14.688142	-17.455588	62184	72588	0	American Restaurant	Cosmetics Shop	Sho Mall
7	Médina	Dakar Plateau	14.680487	-17.450928	97533	113850	1	African Restaurant	Tea Room	Ame Res
8	Plateau	Dakar Plateau	14.667566	-17.432894	41298	48207	0	American Restaurant	Café	Hote

And below is the map showing the three clusters:



Finally, I examined each cluster to determine the discriminating venue categories that distinguish each cluster. As can be seen below:

Cluster 1:

	Neighborhood	Population_2025	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 Most Common Venue
5	Gorée	2333	0	Boat or Ferry	History Museum	Hotel	Restaurant	Italian Restaurant	Bookstore	Café	Burger Joint	Breakfast Spot	Tennis Court
6	Gueule Tapée-Fass-Colobane	72588	0	American Restaurant	Cosmetics Shop	Shopping Mall	Hotel	Tennis Court	Bookstore	Café	Burger Joint	Breakfast Spot	Bar
8	Plateau	48207	0	American Restaurant	Café	Hotel	Fast Food Restaurant	Tea Room	Argentinian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar

Cluster 2:

	Neighborhood	Population_2025	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 Most Common Venue
7	Médina	113850	1	African Restaurant	Tea Room	American Restaurant	Argentinian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Boat or Ferry	Bookstore

Cluster 3:

	Neighborhood	Population_2025	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 Most Common Venue
4	Fann-Point E-Amitié	26164	2	Tennis Court	Tea Room	American Restaurant	Argentinian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Boat or Ferry	Bookstore

D. Results and Discussion

It is important to note that the stagnation or decline in the number of inhabitants in a municipality must generally be interpreted as negative factors related to poorer living conditions, a particularly weak fertility context, a situation considered repulsive or a poor anticipation of housing needs, to the point of explaining migration losses. As I have already said, under the dual influence of rural migratory flows and natural increase, the Dakar region has developed very rapidly. So will we expect a population boom in this region by *2025?* Of course, because the results of our analysis reveal the following: In *2019*, the number of women is greater than or equal to that of men in *12 of Dakar's 19 neighborhoods*. Unlike the neighborhood of *Medina, Biscuitérie, Grand Dakar, Guele Tapée-Fass-Colobane, plateau and Parcelle Assainies*, where men outnumber women. In addition, the three most populated neighborhoods are respectively *Grand Yoff, Parcelle Assainies and Yoff* and the least populated are *Fann-Point-E-Amitié, Ngor and Gorée* . And in *2025*, the same observation is made according to the projections except for Gorée where the population will remain static. However, the population will increase in all boroughs by ***16.73%* in *2025***. It is also worth noting that the most populated boroughs are successively Parcelles Assainies, Grand Dakar, Almadies and Dakar Plateau. Thus, Dakar Plateau is the least populated borough. In addition, the

analysis of this borough revealed that it contains ***42 venues, 33****of which are unique*, but also its neighborhoods can be grouped into three cluster of three, one and one. These results seem quite correct because in the least populated neighborhoods there are more intellectuals and bourgeois people than in the most populated ones where the working class resides. Which also explains more venues in the least populated.

E. Conclusion

The objective of this project is to identify relevant information that can facilitate decision-making by our target audience as well as stakeholders. We can say that we have achieved our objectives well and well because we were able to obtain the following key information: By ***2025***, the population of the Dakar region will increase by ***16.73%*** (and this is the same population rate in all boroughs). In addition, the number of women will remain approximately higher than that of men almost everywhere in the different neighborhoods of Dakar. Dakar Plateau is the least populated borough and then Almadies. This can be explained by the larger number of bourgeois residents in these boroughs. Therefore, given the relatively high population rate, it can be expected that there will be insecurity in the most populated district municipalities. But also an impact on pension and health insurance systems. In any case, demography is an economic asset if it is controlled.We hope that these results will allow decision-makers to better orient their strategic decisions.

F. References:

. [1] https://en.wikipedia.org/wiki/Dakar_Region

. [2] http://www.ansd.sn/

. [3] Forsquare API

. [4] Google Map