SANDTON VENUES

AT MTHIMUNYE

13 JUNE 2020

Table of content:

0. Introduction

0.1. Business Problem

Johannesburg is the largest city in South Africa, with a population of 957 441. Since Johannesburg is one of the wealthiest cities in Africa, most entrepreneurs and entertainers come to Johannesburg to seek opportunities, and because of its diversity, a lot of tourists like to explore the city.

Sandton is a suburb located in Johannesburg and has a population of 222 415. Since Sandton is the most elite suburb in Johannesburg, most luxury hotels and restaurants are located there. Usually high profiled people seek accommodation in Sandton when they come to do business in Johannesburg. Sandton is a big suburb filled with different venues and business operations taking place.

In this project we will explore 50 different venues in Sandton within a radius of 500 meters, so that entrepreneurs, entertainers and tourists who come to Sandton can get an idea of the shopping options they have and available venues they can explore.

0.2. Data Description

1.1. **Johannesburg Data**
- Since Sandton is located in Johannesburg we are going to start by scraping The City of Johannesburg Wikipedia page to get the information about Sandton.
- Wikipedia page: https://en.wikipedia.org/wiki/City_of_Johannesburg_Metropolitan_Municipality

1.2. **Johannesburg Coordinates**
- Since the Johannesburg table on Wikipedia didn't have coordinates for the different suburbs, I created a csv file containing the coordinates of the different suburbs in Johannesburg using google maps.

1.3 **Venues Data**
- I found the 50 different venues in Sandton by using the Foursquare API.

1. Methodology

After scraping the City of Johannesburg table from Wikipedia, I created a dataframe containing all the columns I will need to implement the project. The columns were Suburbs, Code, Population and Area.

	Suburbs	Code	Population	Area
1	Alexandra	798014	179,624	6.91
2	Chartwell	798011	1,728	9.07
3	City of Johannesburg (non-urban)	798002	9,933	289.84
4	Dainfern	798012	6,601	4.08
5	Diepsloot	798003	138,329	12.00

Since the Johannesburg table didn't contain coordinates of each suburb. I created a csv file containing the coordinates using google maps, and merged the dataframe with the Johannesburg dataframe.

|	Suburbs	Code	Population	Area	Latitude	Longitude
0	Alexandra	798014	179624	6.91	-26.1033	28.0976
1	Chartwell	798011	1728	9.07	-25.9870	27.9715
2	City of Johannesburg (non-urban)	798002	9933	289.84	-26.1704	27.9718
3	Dainfern	798012	6601	4.08	-25.9833	27.9981
4	Diepsloot	798003	138329	12.00	-25.9396	28.0128

I used python's folium library to visualize geographic details of each suburb in Johannesburg. I created a map of Johannesburg with suburbs superimposed on top. I used latitude and longitude values to get the visual below:
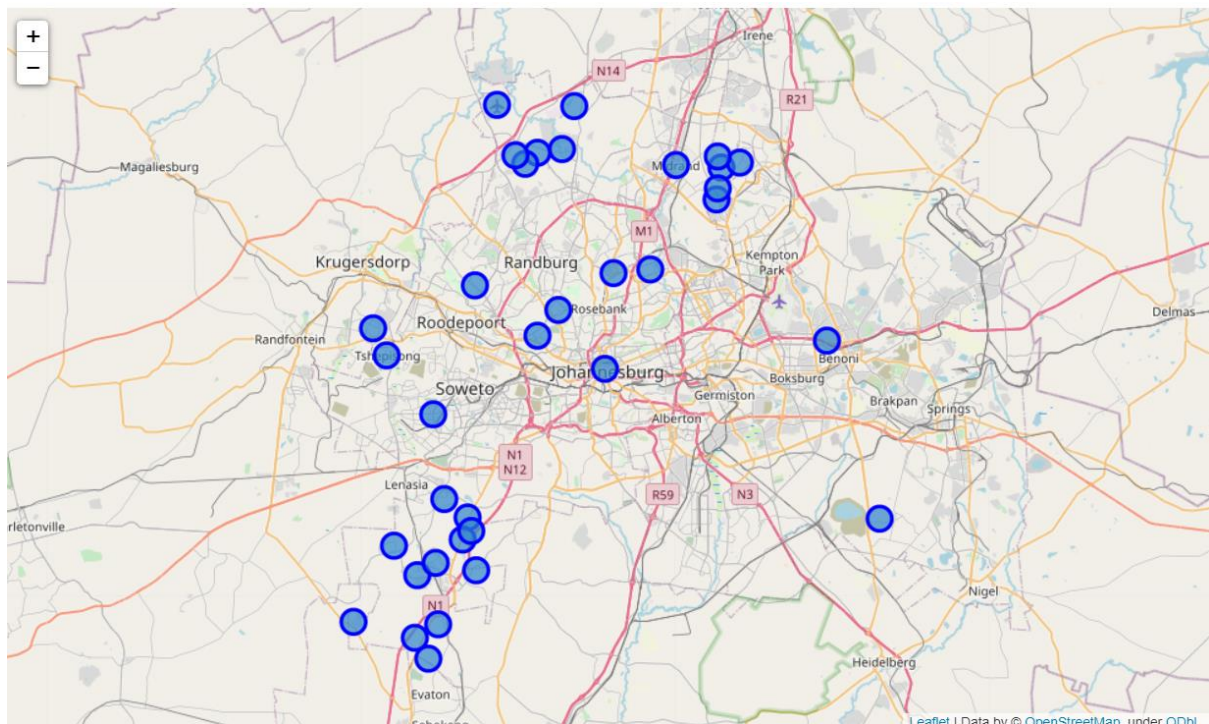
Since this project is mainly focused on obtaining the venues in Sandton. I created a dataframe that only has the details for Sandton.

	Suburbs	Code	Population	Area	Latitude	Longitude
31	Sandton	798013	222415	143.54	-26.1076	28.0567

Sandton was plotted on the Johannesburg map.
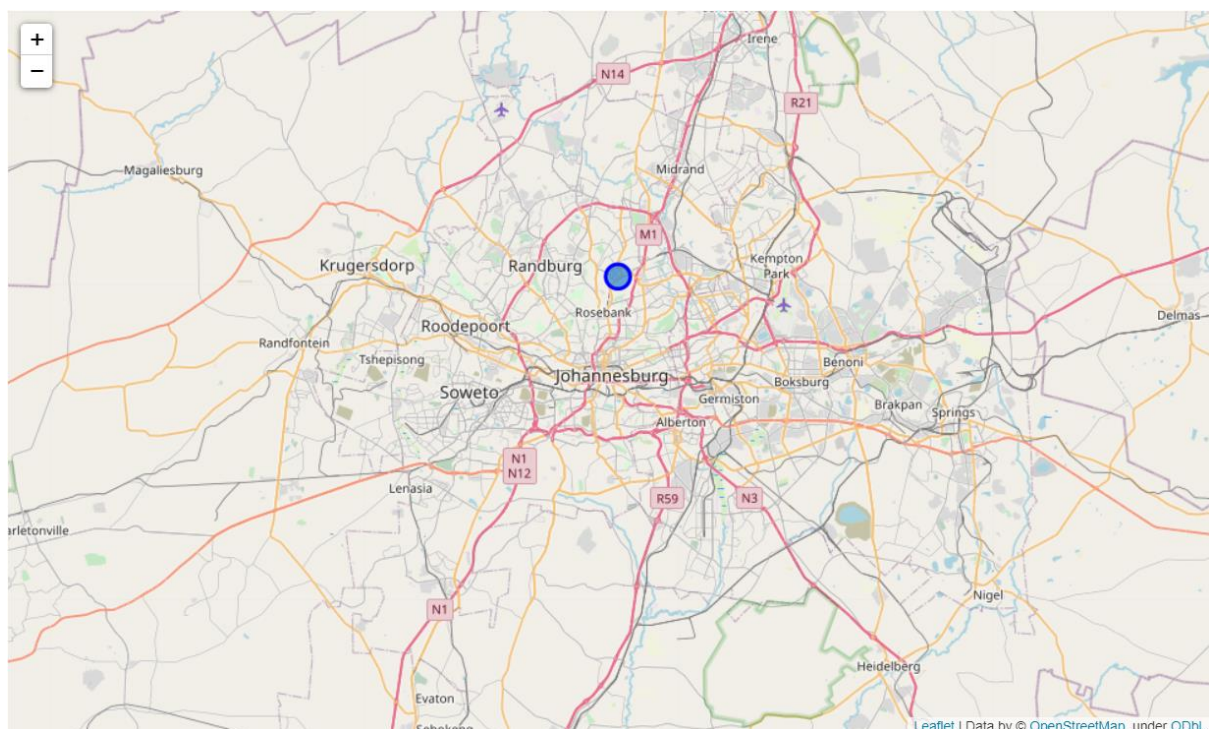
I used the Foursquare API to get the 50 venues in Sandton that are within a radius of 500 meters. Here is a head of the list of Venue Names, Latitude, Longitude and Categories that are located in Sandton.

	Suburb	Suburb Latitude	Suburb Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Sandton	-26.1076	28.0567	Nelson Mandela Square	-26.107411	28.054767	Plaza
1	Sandton	-26.1076	28.0567	The Michelangelo Hotel	-26.107369	28.054795	Hotel
2	Sandton	-26.1076	28.0567	Krispy Kreme Doughnuts	-26.107423	28.057398	Donut Shop
3	Sandton	-26.1076	28.0567	Trump's Grill	-26.107484	28.054812	BBQ Joint
4	Sandton	-26.1076	28.0567	cilantros	-26.107331	28.054742	Tapas Restaurant

According to the 50 venues returned by the Foursquare API, there were 31 unique venue categories. Which are:

Category Venue	Category Venue	Category Venue	Category Venue	Category Venue
African Restaurant	Cocktail Bar	Indian Restaurant	Pharmacy	Tapas Restaurant
American Restaurant	Coffee Shop	Juice Bar	Plaza	Thai Restaurant
BBQ Joint	Donut Shop	Monument/Landmark	Restaurant	Train Station
Bakery	French Restaurant	Multiplex	Salad Place	Null
Bookstore	Greek Restaurant	Music Store	Seafood Restaurant	Null
Cafe	Hotel	Nightclub	Shopping Mall	Null
Clothing Store	Ice Cream Shop	North Indian Restauran	Steakhouse	Null

As we are trying to determine the type of shops and venues that are found in Sandton. We will analyse the venue categories in Sandton. I used the unsupervised learning K-Means algorithm to cluster the venue categories. This resulted in similar venue categories grouped together. I ran K-Means to cluster the venue categories into 5 clusters. The following results were obtained:

Cluster 5

The dataframe for cluster 5 after running K-Means was the following:

	Code	Population	Area	Latitude	Longitude	Cluster Labels	African Restaurant	American Restaurant	BBQ Joint	Bakery	...	Pharmacy	Plaza	Restaurant	Salad Place	Seafood Restaurant	Sho
31	798013	222415	143.54	-26.1076	28.0567	4	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	4	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	4	0	0	0	0	...	0	0	0	0	0	

3 rows × 37 columns

To get a better view of the dataframe, I removed the Code, Population, Area, Latitude, Longitude and Cluster Labels columns. I took the transposed view of the dataframe and set head view to 7. The new dataframe was:

	31	31	31
African Restaurant	0	0	0
American Restaurant	0	0	0
BBQ Joint	0	0	0
Bakery	0	0	0
Bookstore	0	0	0
Café	0	0	0
Clothing Store	1	1	1

Based on the new dataframe, cluster 5 contains 3 clothing store venue categories. A bargraph representation of the results is:

Cluster 4

The dataframe for cluster 4 after running K-Means was the following:
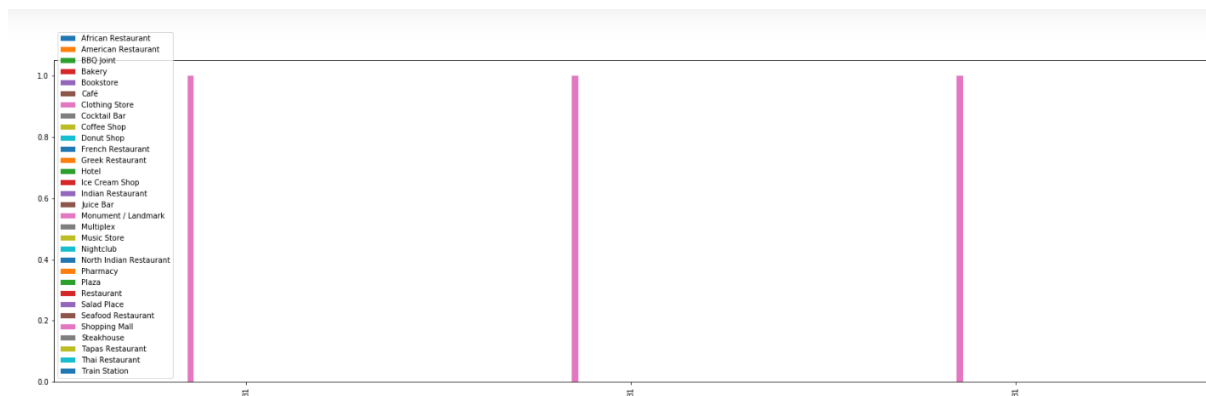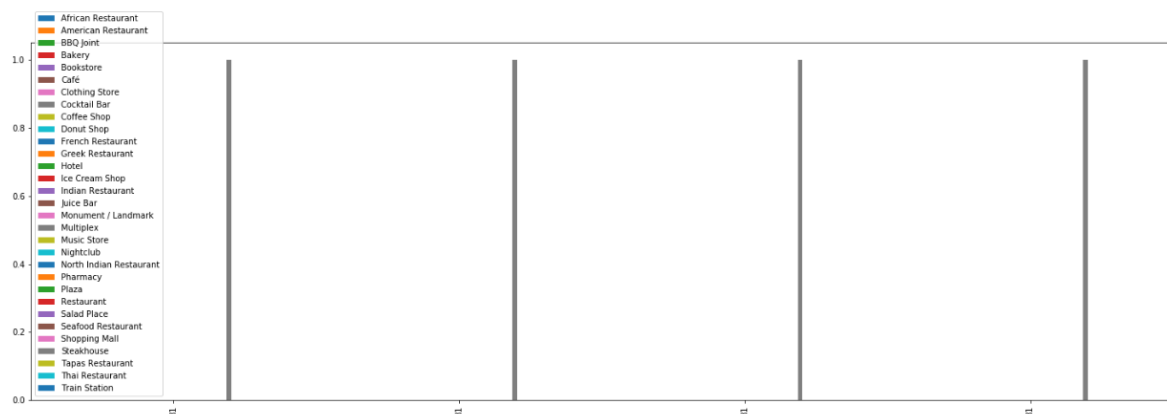
	Code	Population	Area	Latitude	Longitude	Cluster Labels	African Restaurant	American Restaurant	BBQ Joint	Bakery	...	Pharmacy	Plaza	Restaurant	Salad Place	Seafood Restaurant	Sho
31	798013	222415	143.54	-26.1076	28.0567	3	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	3	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	3	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	3	0	0	0	0	...	0	0	0	0	0	

4 rows × 37 columns

To get a better view of the dataframe, I removed the Code, Population, Area, Latitude, Longitude and Cluster Labels columns. I took the transposed view of the dataframe. The new dataframe was:

	31	31	31	31
Shopping Mall	0	0	0	0
Steakhouse	1	1	1	1
Tapas Restaurant	0	0	0	0
Thai Restaurant	0	0	0	0
Train Station	0	0	0	0

Based on the new dataframe, cluster 4 contains 4 Steakhouse venue categories. A bargraph representation of the results is:

Cluster 3

The dataframe for cluster 3 after running K-Means was the following:

	Code	Population	Area	Latitude	Longitude	Cluster Labels	African Restaurant	American Restaurant	BBQ Joint	Bakery	...	Pharmacy	Plaza	Restaurant	Salad Place	Seafood Restaurant	Sho
31	798013	222415	143.54	-26.1076	28.0567	2	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	2	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	2	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	2	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	2	0	0	0	0	...	0	0	0	0	0	

5 rows × 37 columns

To get a better view of the dataframe, I removed the Code, Population, Area, Latitude, Longitude and Cluster Labels columns. I took the transposed view of the dataframe and set head view to 15. The new dataframe was:

	31	31	31	31	31	31	31
African Restaurant	0	0	0	0	0	0	0
American Restaurant	0	0	0	0	0	0	0
BBQ Joint	0	0	0	0	0	0	0
Bakery	0	0	0	0	0	0	0
Bookstore	0	0	0	0	0	0	0
Café	0	0	0	0	0	0	0
Clothing Store	0	0	0	0	0	0	0
Cocktail Bar	0	0	0	0	0	0	0
Coffee Shop	0	0	0	0	0	0	0
Donut Shop	0	0	0	0	0	0	0
French Restaurant	0	0	0	0	0	0	0
Greek Restaurant	0	0	0	0	0	0	0
Hotel	1	1	1	1	1	1	1
Ice Cream Shop	0	0	0	0	0	0	0
Indian Restaurant	0	0	0	0	0	0	0

Based on the new dataframe, cluster 3 contains 7 Hotel venue categories. A bargraph representation of the results is:

Cluster 2

The dataframe for cluster 2 after running K-Means was the following:
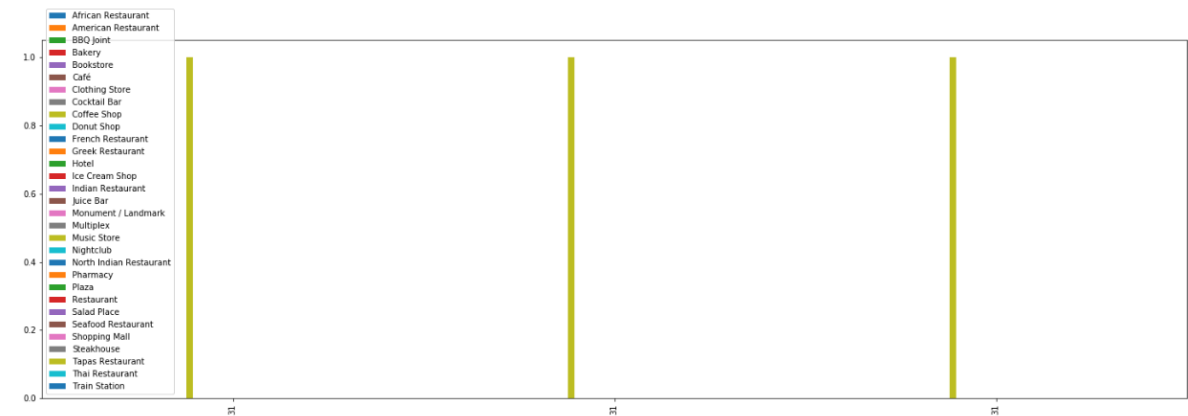
	Code	Population	Area	Latitude	Longitude	Cluster Labels	African Restaurant	American Restaurant	BBQ Joint	Bakery	...	Pharmacy	Plaza	Restaurant	Salad Place	Seafood Restaurant	Sho
31	798013	222415	143.54	-26.1076	28.0567	1	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	1	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	1	0	0	0	0	...	0	0	0	0	0	

3 rows × 37 columns

To get a better view of the dataframe, I removed the Code, Population, Area, Latitude, Longitude and Cluster Labels columns. I took the transposed view of the dataframe and set head view to 10. The new dataframe was:

	31	31	31
African Restaurant	0	0	0
American Restaurant	0	0	0
BBQ Joint	0	0	0
Bakery	0	0	0
Bookstore	0	0	0
Café	0	0	0
Clothing Store	0	0	0
Cocktail Bar	0	0	0
Coffee Shop	1	1	1
Donut Shop	0	0	0

Based on the new dataframe, cluster 2 contains 3 Coffee Shop venue categories. A bargraph representation of the results is:

Cluster 1

The dataframe for cluster 1 after running K-Means was the following:

	Code	Population	Area	Latitude	Longitude	Cluster Labels	African Restaurant	American Restaurant	BBQ Joint	Bakery	...	Pharmacy	Plaza	Restaurant	Salad Place	Seafood Restaurant	Sho
31	798013	222415	143.54	-26.1076	28.0567	0	0	0	0	0	...	0	1	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	0	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	0	0	0	1	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	0	0	0	0	0	...	0	0	0	0	0	
31	798013	222415	143.54	-26.1076	28.0567	0	0	0	0	0	...	0	0	0	0	0	

5 rows × 37 columns

Cluster 1 contains all the other venue categories in Sandton that are less than 3.

2. Results

The Foursquare API returned 50 venues in Sandton, within a radius of 500 meters. The 50 venues had 31 unique categories, which are:

Category Venue	Category Venue	Category Venue	Category Venue	Category Venue
African Restaurant	Cocktail Bar	Indian Restaurant	Pharmacy	Tapas Restaurant
American Restaurant	Coffee Shop	Juice Bar	Plaza	Thai Restaurant
BBQ Joint	Donut Shop	Monument/Landmark	Restaurant	Train Station
Bakery	French Restaurant	Multiplex	Salad Place	Null
Bookstore	Greek Restaurant	Music Store	Seafood Restaurant	Null
Cafe	Hotel	Nightclub	Shopping Mall	Null
Clothing Store	Ice Cream Shop	North Indian Restauran	Steakhouse	Null

I used K-Means to cluster the venue categories. The venue categories were placed in 5 clusters. the 5th cluster contained 3 Clothing stores. The 4th cluster contained 4 Steak houses. The 3rd cluster contained 7 hotels. The 2nd cluster contained 3 Coffee Shops. And the 1st cluster contained the other venue categories that are less than 3.

The top 5 most common venues were found. The first most common venue category was hotels, the second most common venue category was Steakhouse, the third most common venue was Coffee Shop, the fourth most common venue category was Clothing Store, The fifth most common venue was Bookstore.

	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Sandton	Hotel	Steakhouse	Coffee Shop	Clothing Store	Bookstore

3. Discussion

As a resident of Johannesburg that frequently visits Sandton, I can say that Sandton has more diversed venues. For this project I only focused on 50 venues and found the most common. Sandton is mostly known for its vibrant clubs and stylish boutiques.

More venues could be explore for detailed and accurate results. The results obtain are sufficient and outline the important venues in Sandtan that visitors are interest in.

The results could be presented as a web application that users can navigate through to find the closest venue to them.

4. Conclusion

In conclusion, entrepreneurs, entertainers and tourists that come to Sandton can expect to find a large number of hotels to choose accommodation from. There are also a variety of steak houses and Coffee Shops. Clothing stores and Bookstores are also available to explore.

5. References

- The City of Johannesburg Municipality wikipedia page: https://en.wikipedia.org/wiki/City_of_Johannesburg_Metropolitan_Municipality
- Johannesburg wikipedia page: https://en.wikipedia.org/wiki/Johannesburg
- Sandton wikipedia page: https://en.wikipedia.org/wiki/Sandton
- Google Maps
- Foursquare API