

Food and Nightlife Venue Data Analysis of Cape Town

Applied Data Science Capstone Project

1. Introduction

1.1 Background

Cape Town is a coastal city on the west coast of South Africa. The city possesses a number of tourist attractions and in 2018 was voted the World's Best city by readers of Travel and Leisure. The city's most iconic landmark is undoubtedly the UNESCO World Heritage site, Table Mountain [1]. Cape Town's urban geography is influenced by the contours of Table Mountain, its surrounding peaks, the Durbanville Hills, and the expansive lowland region known as the Cape Flats. These geographic features in part divide the city into several commonly known groupings of suburbs (equivalent to districts/boroughs outside South Africa). These districts/boroughs are: Atlantic Seaboard, Cape Flats, City Bowl, Helderberg, Northern Suburbs, South Peninsula, Southern Suburbs, West Coast [1].

1.2 Problem

Aside from international and local tourists, the city also sees a relatively high amount of domestic migration from South African citizens moving in from other parts of the country. This is because the city is perceived as having a higher standard of living than other cities and better employment opportunities. The beautiful natural surroundings also serve as a significant draw card for domestic migrants. The majority of these migrants are recently graduated individuals and young professionals who are able to move to a new city because they have greater freedom to do so. There are a number of factors that a typical young person may consider when moving into a new neighbourhood. For the purpose of this analysis we will assume that the two most important factors these individuals consider are the following:

- The type of eateries that exist around the neighbourhood.
- The number of nightlife spots that are available within a short distance

We will therefore create a map that young people can use to decide which neighbourhood they want to move into based on the above mentioned factors. For the purpose of this analysis we will also only consider neighborhoods within the districts/boroughs of: Atlantic Seaboard, City Bowl, and Southern Suburbs because these form the economic hub of the city where all our main venues of interest are located.

2. Data Sources

To create the map we will need the following data sources:

- Food and Nightlife spot venues for each neighbourhood will be obtained from the Foursquare API [2].

- The list of suburbs in each district/borough (In South Africa, the term "suburb" does not necessarily mean "residential area on the edge of a city"; rather, it is used synonymously with neighbourhood to refer to the smallest geographical subdivision of the city.) [3]
- Google Maps to obtain the gps coordinates of the each neighbourhood. [4]
- We want to create a choropleth map to display the number of nightlife spots per neighbourhood. The highest level boundary file I could find was a .json file for South Africa was a Fourth-level Administrative Divisions from NYU spatial data repository [5]. The .json goes down to ward level. In South Africa, wards are geopolitical subdivisions of municipalities used for electoral purposes.
- Lastly we need to determine which ward each neighbourhood belongs to. For this we can use the map viewer provided by the city of Cape Town [6].

3. Methodology

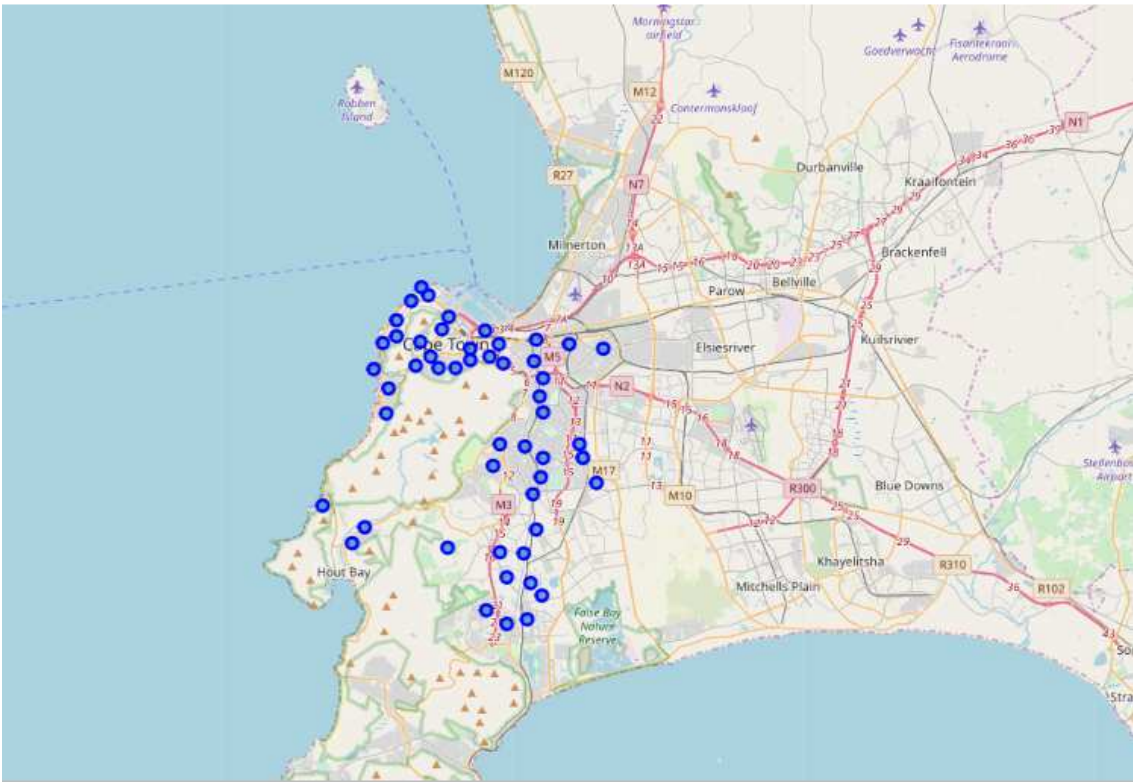
As mentioned above our analysis will focus on the districts/boroughs of: Atlantic Seaboard, City Bowl, and Southern Suburbs. We firstly need to scrape the “List of Cape Town suburbs” Wikipedia page in order to get obtain all the neighbourhoods around Cape Town. We also rename some neighbourhoods, drop some columns, and drop some neighborhoods for the reasons listed in the table below.

	Dropped Column	Dropped Neighbourhoods	Renamed Neighbourhoods
Column/Neighbourhood	Postal Code, Street Column	Lower Vrede (District Six), Schotse Kloof (Malay Quarter), Kreupelbosch, SouthField	Bo-Kaap(Malay Quarter): Bo-Kaap Devil's Peak Estate: Devils Peak Estate Walmer Estate (District Six): Walmer Estate Woodstock (including Upper Woodstock): Woodstock Zonnebloem (District Six): Zonnebloem
Reason	No need to keep these columns as we won't be using them in our analysis	All of these neighbourhoods are part of neighbourhoods which are already in our dataframe.	To make it easier to merge our list of suburbs df with the GPS coordinates df later

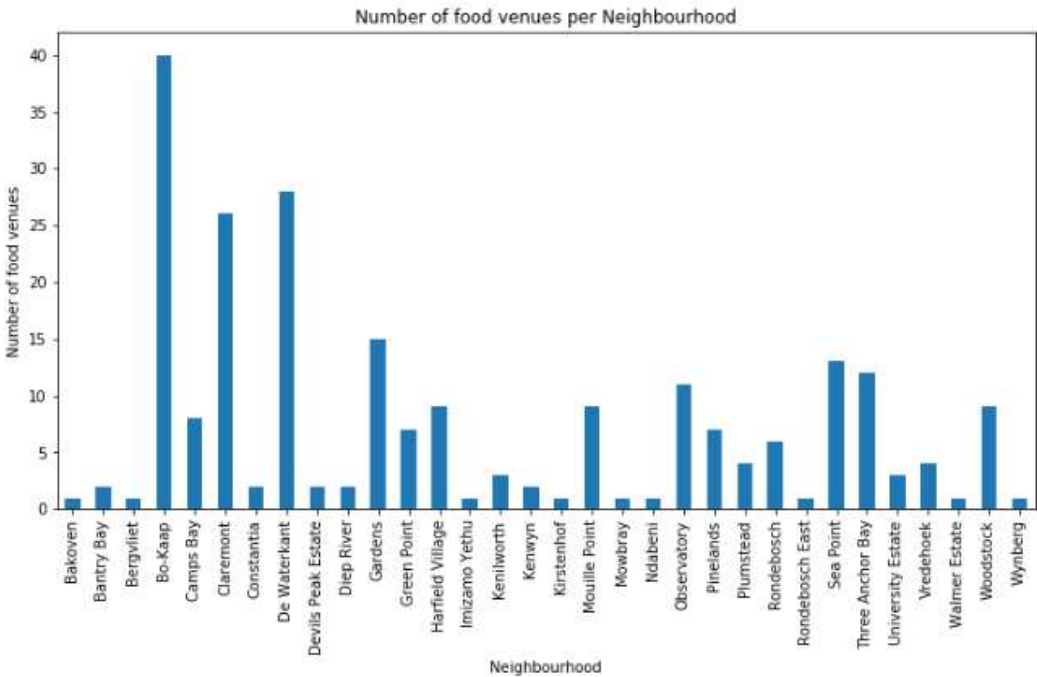
Next we need to append the gps coordinates to each neighbourhood. The CSV files containing the gps coordinates of each neighbor are stored in my GitHub repository. The resultant df looks as follows:

	Neighbourhood	Area	latitude	longitude
0	Bo-Kaap	City Bowl	-33.9216	18.4150
1	Devils Peak Estate	City Bowl	-33.9370	18.4321
2	De Waterkant	City Bowl	-33.9154	18.4194
3	Foreshore	City Bowl	-33.9222	18.4410
4	Gardens	City Bowl	-33.9353	18.4083

The neighbourhoods were then visualized geospatially using the folium library.



Next we use the Foursquare API to extract the number of food venues within a 500m radius of each neighbourhood's location. To do this we first request all venues and then filter out only those venue categories related to food venues. Foursquare provides a list of available food related venue categories [7]. The bar chart below shows the number of food venues in each neighbourhood:



In summary there are 42 unique food venue categories spread across 32 of the 51 neighbourhoods in our study area. There are 233 food venues in total. We can see that Bo-Kaap, Claremont and De Waterkant are host to most of our food venues.

The data frame below displays the 5 most common venue types in each neighbourhood.

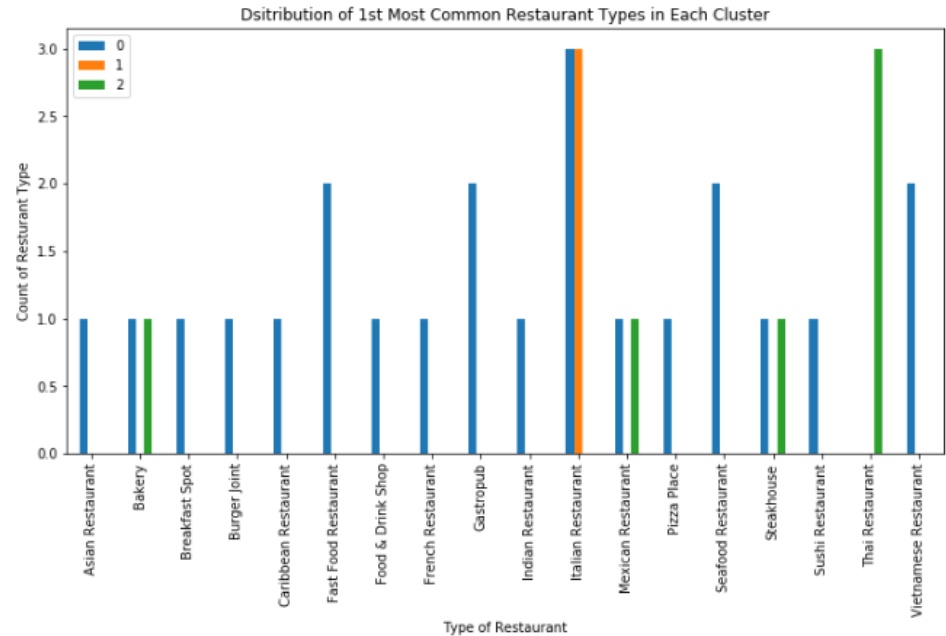
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bakoven	Steakhouse	Vietnamese Restaurant	Cuban Restaurant	Food Truck	Food & Drink Shop
1	Bantry Bay	Seafood Restaurant	Deli / Bodega	Vietnamese Restaurant	Cuban Restaurant	Food Truck
2	Bergvliet	Breakfast Spot	Vietnamese Restaurant	Deli / Bodega	Food Truck	Food & Drink Shop
3	Bo-Kaap	Burger Joint	Steakhouse	Indian Restaurant	Ethiopian Restaurant	Tapas Restaurant
4	Camps Bay	Pizza Place	Gastropub	Deli / Bodega	Breakfast Spot	Seafood Restaurant

We then used unsupervised machine learning K-means algorithm to cluster the neighbourhoods. K-Means algorithm is able to cluster unlabeled data into user specified number of cluster. The cluster labels obtained are then are merged with the 2 table above to obtain the following table:

	Neighbourhood	Area	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Top Three
0	Bo-Kaap	City Bowl	-33.9216	18.4150	0	Burger Joint	Steakhouse	Indian Restaurant	Ethiopian Restaurant	Tapas Restaurant	Burger Joint, Steakhouse, Indian Restaurant
1	Devils Peak Estate	City Bowl	-33.9370	18.4321	1	Italian Restaurant	Vietnamese Restaurant	Deli / Bodega	Food Truck	Food & Drink Shop	Italian Restaurant, Vietnamese Restaurant, Deli...
2	De Waterkant	City Bowl	-33.9154	18.4194	0	Sushi Restaurant	Italian Restaurant	African Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Sushi Restaurant, Italian Restaurant, African ...
3	Gardens	City Bowl	-33.9353	18.4083	0	Italian Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	French Restaurant	Bakery	Italian Restaurant, Mediterranean Restaurant, ...
4	University Estate	City Bowl	-33.9386	18.4521	0	Caribbean Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Deli / Bodega	Food Truck	Caribbean Restaurant, Mediterranean Restaurant...

We’ve created another column to display the top 3 food venues in each neighbourhood. This column will be used as a pop up label in our map.

Before creating the map let’s analyze the distribution of restaurant types in each cluster. The bar chart below show a distribution of restaurant types for each of the 3 clusters. We are just considering restaurant types under the 1st Most Common Venue column here:

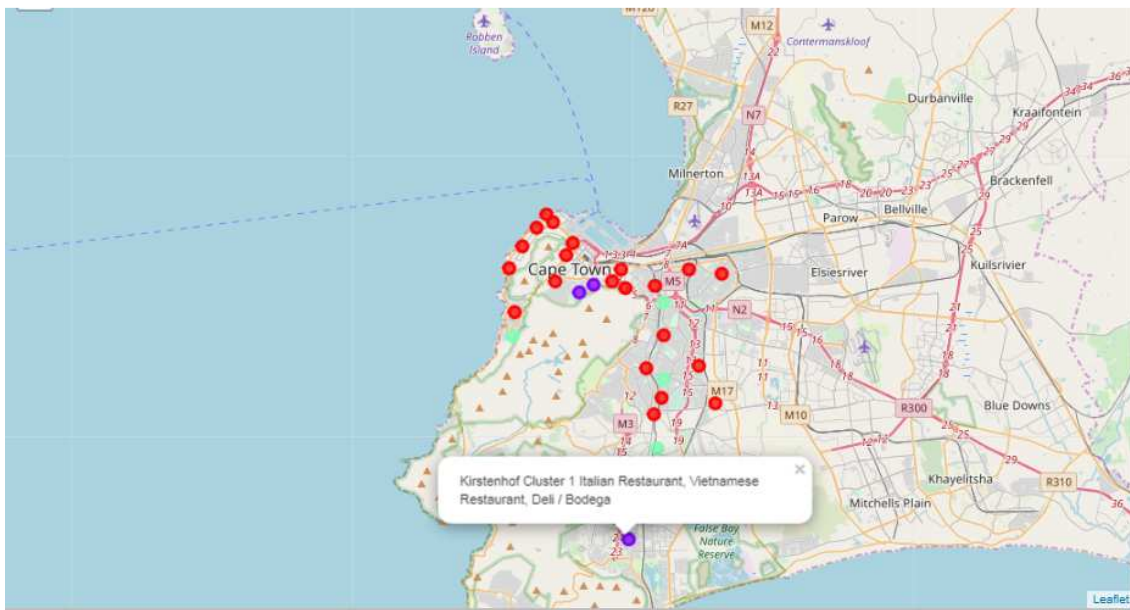


As can be seen above, most neighbourhoods fall into cluster 0. Cluster 0 has a high variety of different restaurant types. Increasing the number of clusters does not change this. Increasing clusters only creates cluster with only one or two neighbourhoods.

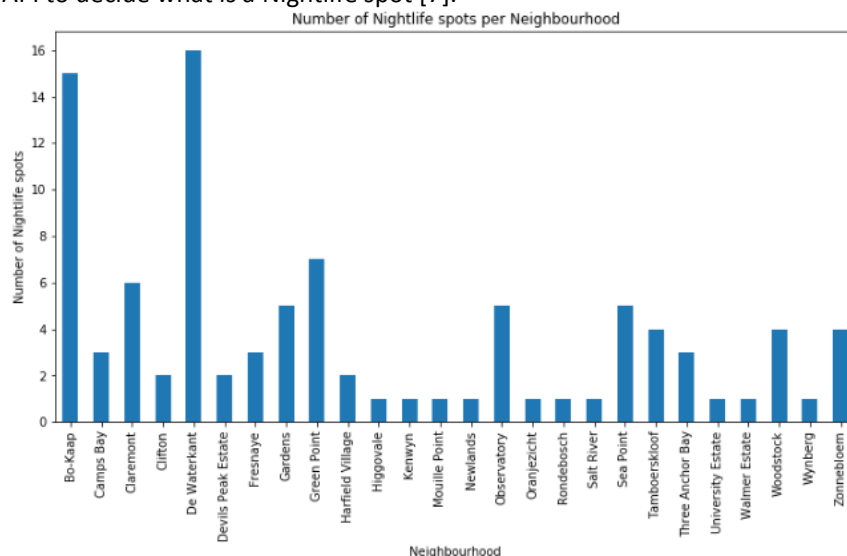
Cluster 1 is neighbourhoods with Italian restaurants as the most common. There are 3 neighbourhoods that fall into this cluster namely: Vredehoek, Devil's Peak and Kirstenhof. Interestingly, Vietnamese restaurants also fall into the top 3 most common restaurants in these neighbourhoods. So we can say this cluster represents Italian and Vietnamese restaurants.

Cluster 2 is most Thai restaurants. Steakhouses also feature prominently in the top three most common restaurants in the six neighbourhoods that are part of this cluster.

Now let view a map of the clusters:



Now we obtain the number of nightlife spots and again we use the categories provided by the Foursquare API to decide what is a Nightlife spot [7].



We now need to create the choropleth by determining the number of nightlife spots in each ward.

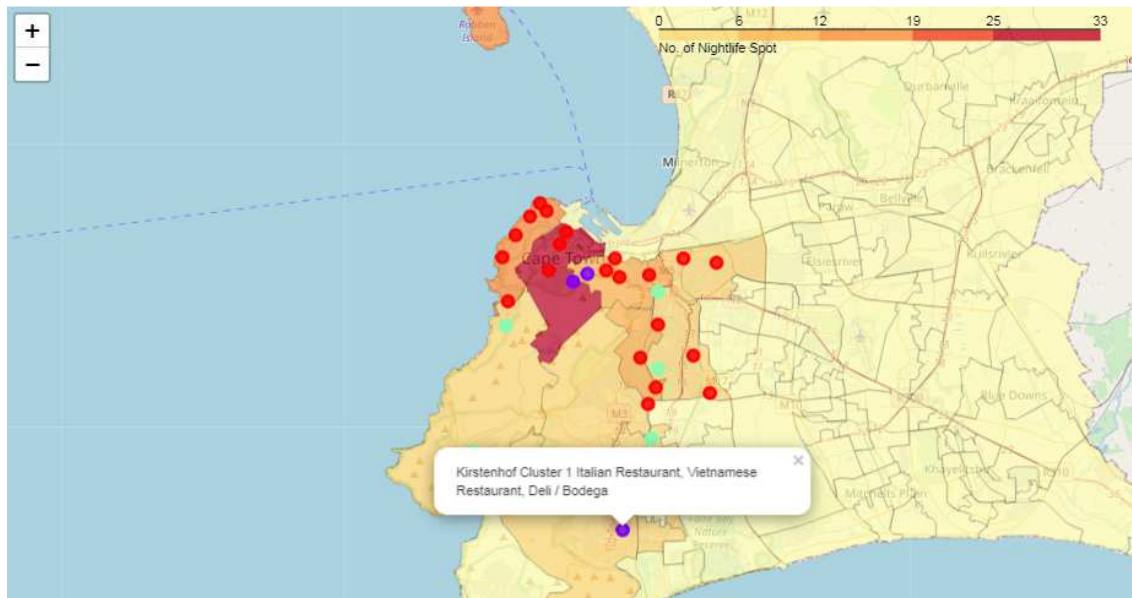
1. First we count the number of nightlife spots in each neighbourhood
2. Add a column which specifies which ward a neighbourhood belongs to. Ward are made up of multiple columns
3. Group by ward to see the total number of nightlife spots in each ward

The three tables below show the 3 stages explained above:

	Venue Category	Neighbourhood	Ward	Venue Category	Ward	No. of Nightlife Spots		
Neighbourhood		0	Bakoven	54	NaN	0	53	0.0
Bo-Kaap	15	1	Bantry Bay	54	NaN	1	54	13.0
Camps Bay	3	2	Bergvliet	73	NaN	2	57	7.0
Claremont	6	3	Bishopscourt	62	NaN	3	58	2.0
Clifton	2	4	Bo-Kaap	77	15.0	4	59	8.0
De Waterkant	16							

4. Results

We are now ready to create the choropleth style map showing the clustered map based on food venue and a choropleth layer which tell us the number of nightlife spots in each ward. The pop up also shows us the three most common food venue category in each neighbourhood.



The number of nightlife spots per ward ranges from 0 to 32. The city center area has the most restaurants and nightlife spots as expected. The suburban areas in the southern suburbs have very few restaurants and nightlife spots. The city bowl area seems most ideal for someone looking for a variety of restaurants and nightlife spots.

5. Discussion

Ideally, if we wanted to accurately compare neighbourhoods against each other it would have been better to find a .json which shows us the actual neighbourhood boundaries. The ward .json we used bundles multiple neighbourhoods and some detail is probably lost. The wards vary greatly in size and sometimes cuts a neighbourhood in two so we had to decide which ward to assign a neighbourhood to in such cases.

The clustering method used ended up cluster most neighbourhoods into one cluster no matter how many clusters we used. Perhaps further mapping the food categories into higher level categories could solve the problem and create better distributed clusters. For example we could say Thai, Japanese, Chinese restaurants are mapped to Asian Cuisine. And create European Cuisine, African etc.

However, a user of the map can still get a sense of what type of neighbourhood's restaurant profile because we have a pop up which displays the top 3 restaurants available in a neighbourhood.

6. Conclusion

The choropleth map created successfully achieves the goal of using data analytics to assist a young person moving into Cape Town decide which neighbourhood to move into based on:

- The type of eateries that exist around the neighbourhood.
- The number of nightlife spots that are available within a short distance

References

- [1]. [Cape Town - Wikipedia](#)
- [2]. [Foursquare API](#)
- [3]. [List of Cape Town suburbs - Wikipedia](#)
- [4]. [Google Maps](#)
- [5]. [Fourth-level Administrative Divisions, South Africa, 2015](#)
- [6]. [City of Cape Town Map Viewer](#)
- [7]. [Foursquare - Venue Categories](#)