Vacationing in Cape Town

An analysis of activities and accommodation rental prices

# Problem and Background discussion

- Cape town is one of South Africa's most visited tourist destinations, with a vast amount of options to choose from

- To maximise ones vacation utility by providing analysis into the following two questions

  - What are my options on things that I can do?

  - Where should I stay in relation to the activities that I want to perform and what is a reasonable price that I can spend on accommodation.

Data acquisition and cleaning

- Data was not readily available, thus quite bit of data cleansing was necessary

- Sources of data:

  - [Property website](#): to obtain the neighbourhood details

  - Air Bnb Website: to obtain the Accommodation details

  - [Mapit website](#): to obtain the GeoJson Files for each of the areas

  - Four Square API: obtain the activity data

- Tools utilised

  - Number of Python libraries: Pandas, Geopy, BeautifulSoup, Request

  - [GeoJson websites](#): to perform the consolidation of all the GeoJson files

Accommodation Analysis

- Average nightly rate calculated and imposed on to a map of the area

- This provided a view of how each of the accommodation areas compared to one another at high level

- Not a good representative of the underlying data further analysis was required
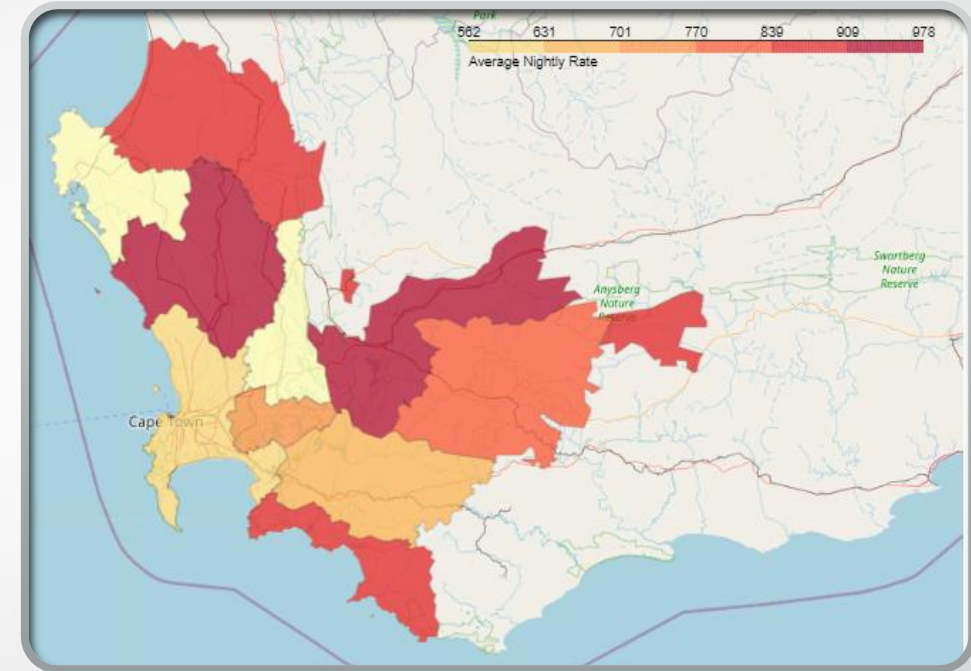
	Grouping	Price
0	Bergrivier	863.200000
1	Breedvalley	951.750000
2	Ceres	896.950000
3	City_of_Capetown	565.725000
4	Drakensfontein	721.625000
5	Langeberg	818.850000
6	Overstrand	855.566667
7	Saldana Bay	673.325000
8	Stellenbosch	757.225000
9	Strand	868.550000
10	Swartland	974.150000
11	Theewaterskloof	653.800000

More indepth Accommodation Analysis

- Swarm plot was constructed to provide further analysis into each of the areas

- Provides a view that the majority of the accommodation is largely grouped around the R500 mark

- However there is enough of a spread across the price range to offer adequate options for travelers

Swarmplot depicting the per night pricing distribution
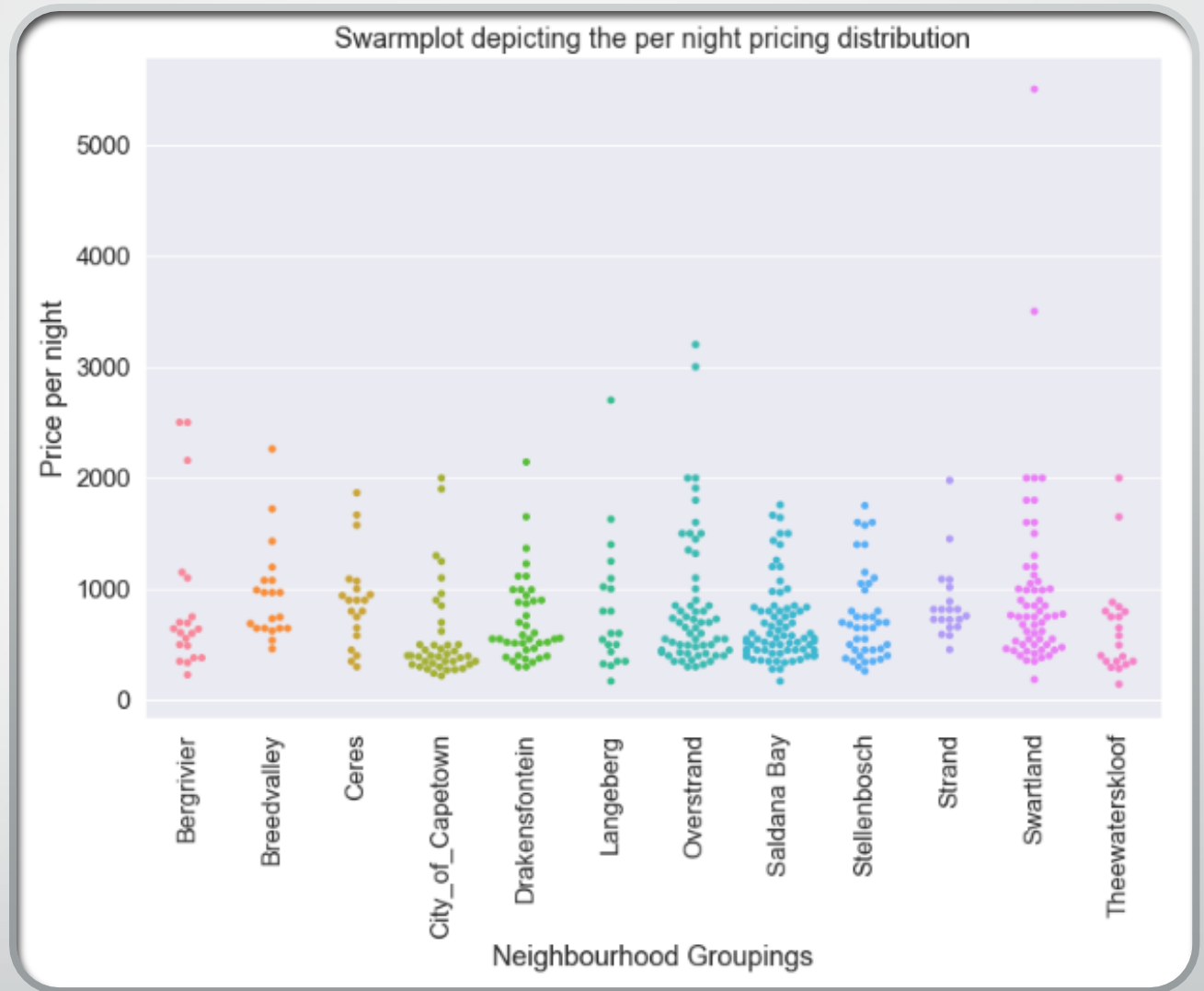
Activity Analysis

- The Four-square data set yielded 231 different venues that consists of 81 categories

- The data (as seen in the 1st table) provided a mixed bag of options with some areas having a number of options and others not having much

- The data was also grouped into the most common categories in that area (2nd table)

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bettys Bay	8	8	8	8	8	8
Cape Town	5	5	5	5	5	5
Ceres	5	5	5	5	5	5
Franschhoek	36	36	36	36	36	36
Gansbaai	8	8	8	8	8	8
Grabouw	7	7	7	7	7	7
Kleinmond	4	4	4	4	4	4
Langebaan	11	11	11	11	11	11
Malmesbury	6	6	6	6	6	6
Paarl	3	3	3	3	3	3
Paternoster	14	14	14	14	14	14
Riebeek Valley	4	4	4	4	4	4
Robertson	11	11	11	11	11	11
Saldanha	6	6	6	6	6	6
Somerset West	25	25	25	25	25	25
Stellenbosch	50	50	50	50	50	50
Velddrif	2	2	2	2	2	2

	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	Bettys Bay	Coffee Shop	Inn	Campground	Restaurant	River	Scenic Lookout	Park	Hostel	Football Stadium	French Restaurant
1	Cape Town	Scenic Lookout	Gift Shop	Café	Cable Car	Wine Shop	Flea Market	Department Store	Diner	Discount Store	Fast Food Restaurant
2	Ceres	Grocery Store	Breakfast Spot	Fast Food Restaurant	Burger Joint	Convenience Store	Department Store	Diner	Discount Store	Fish & Chips Shop	Flea Market
3	Franschhoek	Hotel	French Restaurant	Coffee Shop	Restaurant	Brewery	African Restaurant	Burger Joint	Italian Restaurant	Mediterranean Restaurant	Chocolate Shop
4	Gansbaai	Seafood Restaurant	Burger Joint	Harbor / Marina	Restaurant	Bed & Breakfast	Convenience Store	Shopping Mall	Wine Shop	Department Store	Diner
5	Grabouw	Grocery Store	Convenience Store	Department Store	Fried Chicken Joint	Discount Store	Shopping Mall	Flea Market	Deli / Bodega	Diner	Fast Food Restaurant

Activity Analysis

- With the data set provided by four square the neighborhoods where clustered using the unsupervised K-means algorithm, 5 clusters was deemed to be appropriate

- The 5 classifications where named as follows and can be seen over laid on a map of the area

  - Seafood restaurant (red)

  - Restaurant (purple)

  - General (blue)

  - General Retail (light green)

  - Coffee shops (orange)
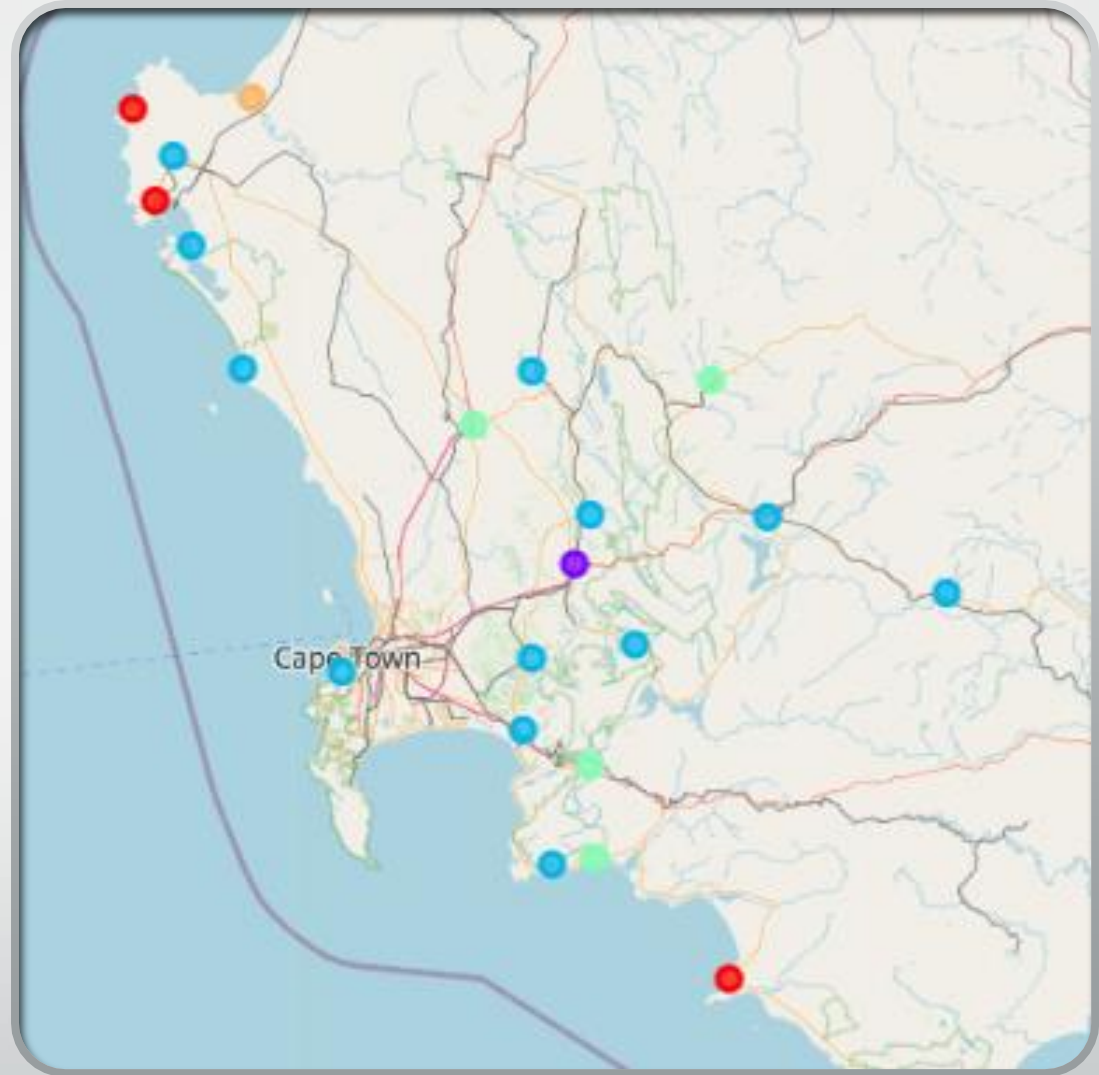
Summary of Results

Accommodation analysis

- As can be seen in the Choropleth map certain areas have a higher average pricing than others, however after a more in-depth analysis; prices are evenly distributed and provide a wide range of options depending on customers budget requirements.

Activity Analysis

- There are certain areas that lend them self to certain activities, however due to the lack of depth in the data I don't believe this provides a true reflection of activities that reside with in each of these areas. Further data is required.
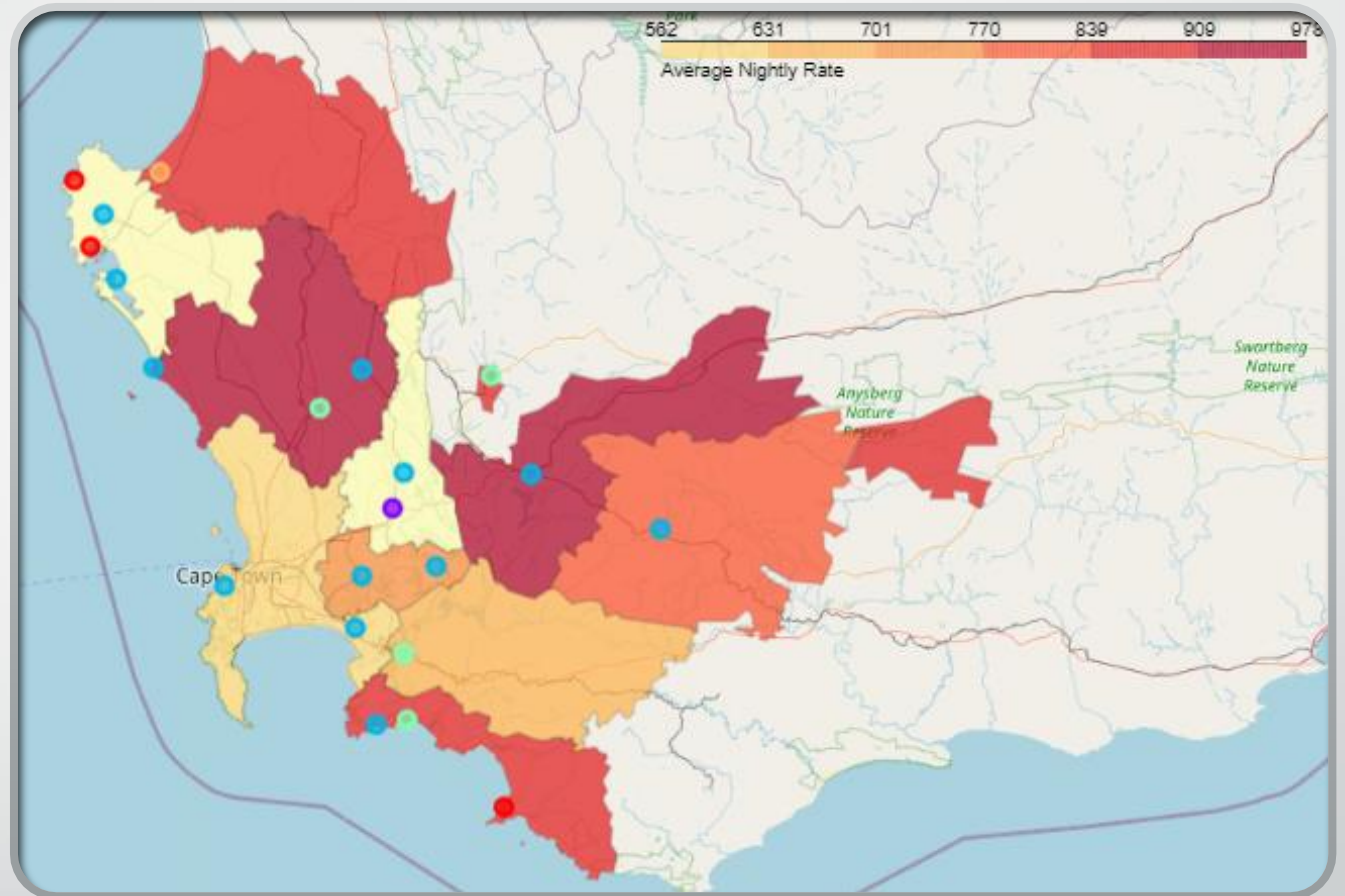
Conclusion and further research

What are my options on things that I can do?

- Data here was inconclusive, however from the data that was available 5 classifications where obtained, which were Seafood restaurants, Restaurants, General, General retail and Coffee shops. More data is required to provide more value add options in this space.

Where should I stay in relation to the activities that I want to perform and what is a reasonable price that I can spend on accommodation?

- Quite insightful, and found that there are areas that differ from an average pricing perspective

- Most of the options with in each of the areas reside on the R500 mark.

- In analysing each of the areas it was found that there was a price premium on accommodation for the more popular areas to the lesser known ones, even though they were not far (km) from one another

Further Study

I believe that if further study is performed on this piece of work it will yield some valuable results. Areas where I feel addition data can be added:

- Western Cape is a wine region, more wine venues would need to be added

- A quick view at Trip Advisor highlighted a multitude of items available, the activity data could be enriched with this information.

- Additional accommodation Sources could be added, like hotel rates, and addition self catering sources to provide broader options and possible pricing differentials