

Report

Exploring Venues in Nelspruit, South Africa📍



Table of Content

- ◇ [Exploring Venues in Nelspruit, South Africa](#)
- [Table of Content](#)
- [1. Introduction](#)
 - [1.1 Background](#)
 - [1.2 Problem](#)
 - [1.3 Interest](#)
- [2. Data](#)
 - [2.1 Data source](#)
 - [2.2 Data collection](#)
 - [2.3 Data cleaning](#)
- [3. Methodology](#)
 - [3.1 Exploratory Data Analysis](#)
 - [3.1.1 Category](#)
 - [Social, Fun, and Adventure](#)
 - [Popular categories](#)
 - [Formatted view of Full list](#)

- [Categories with no venues listed](#)
- [3.1.2 Plot data](#)
 - [Just outside 20km](#)
 - [All venues per group](#)
- [3.1.3 Average distance from center](#)
 - [Per category](#)
 - [Per group](#)
- [3.2 Clustering Models](#)
 - [3.2.1 DBScan](#)
 - [Statistic](#)
 - [Venues per cluster](#)
 - [Average distance per cluster to center](#)
 - [Plotting clusters on the map](#)
- [4. Results](#)
- [5. Discussion](#)
- [6. Conclusion](#)
- [References](#)

1. Introduction

This project is intended towards exploring a specific area.

1.1 Background

Nelspruit is a small city in comparison to its neighboring cities, but well known for its wildlife reserve - Kruger National Park, home to elephants, zebras, rhinos, and other wildlife. The city's Lowveld National Botanical Garden features a man-made rainforest. Established in 1905 with ~110000 population.

I recently moved here and found a job working from home, I have visited some of the restaurants and malls, but then I wanted to know what are all the venues that I would be interested in.

And so I have thought about 3 groups of interest, Social, Adventure, and Fun venues. Also without travelling too far. I wanted to include a highly recommended venue at Kaapsehoop which then set the distance I'd travel the furthest for this reason - 20km.

1.2 Problem

What social, fun, and adventurous venues are in and around 20km radius from Nelspruit, South Africa?

So this project focuses on collecting, formatting, and presenting places that fall under the following categories: fun, adventurous, and social within a distance of 20km.

1.3 Interest

Target? This project is targeted to residents and visitors of these areas. But mostly myself.

Why? Residents can utilize the information to explore more entertainment options and as such can those who plan to visit too. Thus making it valuable to both categories of individuals. There are also value for entrepreneurs, governments, academics, etc. but that is beyond the scope of this assignment.

2. Data

The data I require are:

- Venue ID
- Venue Name
- Category
- Category ID
- Location Coordinates

2.1 Data source

The primary data source is Foursquare's Database, I also utilized GeoPy's Geolocator for Reverse Address lookup

Foursquare | <https://foursquare.com/>
GeoPy | <https://pypi.org/project/geopy/>

2.2 Data collection

Collecting the data was not a straight forward task a initially anticipated. I reviewed other notebooks from Google search in similar context and learned of KDTree and QuadTree algorithms to mine Foursquare, but I then suspected it might take to much time to learn and implement so I created a 2-dimensional looping algorithm with less elements or sets of elements.

My algorithm could be improved, but I was pressed with time and decided to stick to the working one. My algorithm extracts the required data from the JSON responses and stores them in lists until all categories are requested, then adds the lists to a dataframe.

2.3 Data cleaning

During the mining algorithm I encountered duplicates of venues in responses (with different catIDs) which I filtered out by keeping track of the Venue IDs recorded. Especially when requesting 50 more venues per venues' location to deepen the search, there found over 2000+ duplicates. The data was striped from the response as we got it and then added to a list in the desired format so the next phase we don't have to worry about cleaning that much.

After the mining algorithm the data cleaning was but a simple endeavor due to the control of the flow of data and the format from the API response to the dataframe.

3. Methodology

The methodology consists of the following:

- Establish data requirements
- Mine the data from the source and save to files
- Explore & analyze data
 - ◇ Categorize them
 - ◇ Plot them on a map
 - ◇ Calculate distances
- Prepare data for Machine Learning model
- Train Density-based model with geographical data
 - ◇ Review clusters
 - ◇ Calculate distances
 - ◇ Plot them on a map

3.1 Exploratory Data Analysis

At this part of the analysis I looked at the venues where they are assigned by category and/or groups.

Questions that we analyzed the data with were:

- How many venues are per group?
- How many venues are per category?
- Formated view of all venues and their counts grouped by categories then grouped again by group.
- Which categories did Foursquare respond with no venues?
- Plotting of the Target City, the few venues just outside the 20km radius, and all the venues according to their groups.
- What are the average distance per group?

3.1.1 Category

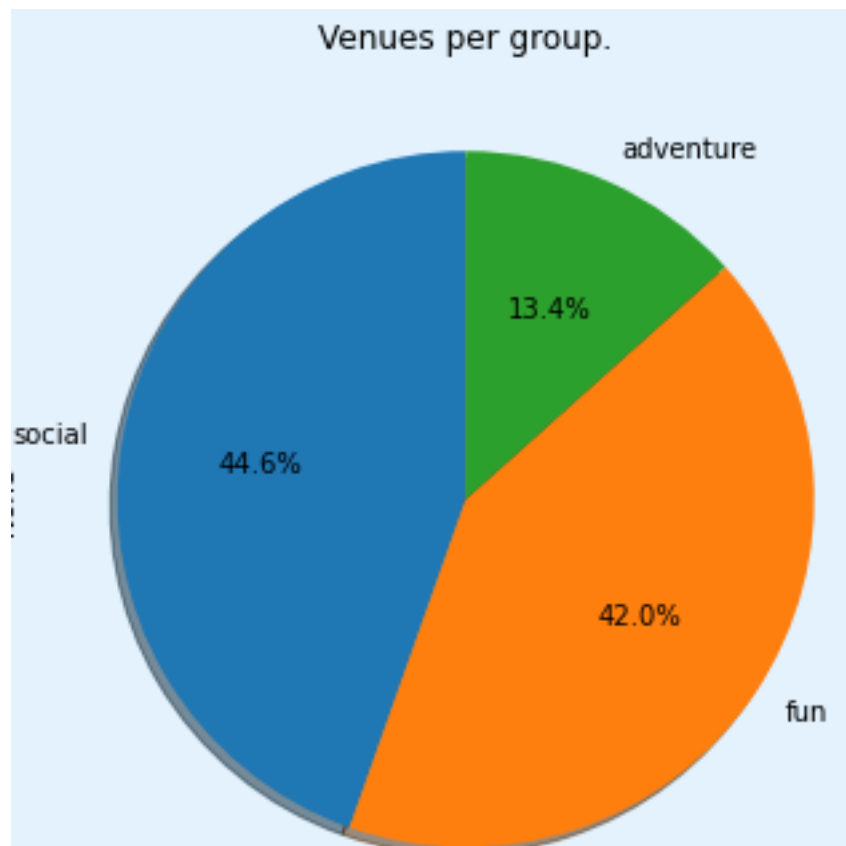
Here I explored the data inorder to answer the exploratory questions listed above.

Social, Fun, and Adventure

How many venues are per group?

Group	
social	70
fun	66
adventure	21

Proportionally graphed with Pie Chart



Popular categories

How many venues are per category?

Group	Category	Name	Venue Count
adventure	Botanical Garden	Nelspruit Botanical Gardens	1
		Nelspruit Falls	1
	Campground	Nora Falls Lodge	1
	Castle	Beukes	1
	Cave	Sudwala Caves	1
	Lake	Wikidodam	1
	Mountain	ulitvk	1
	National Park	Chimp Eden - Jane Goodall Institute	1
	Other Great Outdoors	Isuzu KB Launch	1
		Klipspringer Lodge	1
		Kwanyoni Lodge	1
		Likweli Lodge & Sanctuary	1
		Lions Head Lodge	1
		adams calendar	1
		thaba chweu	1
	Park	Nelspruit Marathon Club	1
		Riverside Mall Park	1
	Trail	Timbavati	1
		Hilltop	1
		Kaapschehoop Hiking Trail	1
		Steillet - Nelspruit Over View	1
		Mr. Funtubbles	1
fun	Arcade	Orange	1
		Stables	1
		The Fun Companv	1
		The Magic Company	1
		Bundu Art	1
	Art Gallery	Mbombela Art Gallery	1
		Zebrina questhouse & gallery	1
	Athletics & Sports	Broham Padstall Driving Range	1
		Presidential Suite Mbombela Stadium	1
		Virgin active	1
	Casino	Whiteriver Stadium	1
		Emnotweni Casino	1
	Cocktail Bar	Riverside Casino	1
		News Cafe - Nelspruit	1
	General Entertainment	5fm Ultimix 3 Tour	1
		Edamini Carwash & Chillas	1
		Kruger Park SAFARI	1
		Riverside mall magic companv	1
		Tango Blue	1
	Golf Course	Umbali Events Venue	1
		Yoga 94 Serendib	1
		Leopard Creek	1
		Lowveld Country Club	1
	Gym	Matumi Golf Estate	1
		Nelspruit Golf Club	1
		Riverside Driving Range	1
		Training Edge Riverside	1
		Training Edge Westend	1
	Marital Arts School	Training Edge Whiteriver	1
		Virgin Active Nelspruit	1
	Movie Theater	k fit	1
		The gym	1
	Music Venue	4D xd theatre riverside mall	1
		Boulevard Cinema & Cafe	1
		Movies@Riverside	1
	Pool	Ster Kinekor	1
		Inniebos Music Fest	1
		Lowwagwala Fm	1
		Mthe's Music Production	1
		SABC MPUMALANGA	1
	Pool Hall	Sub Records	1
		Bull Pool Investments cc	1
		Bundu pool	1
		Matunyane	1
	Race-track	Summer Fun Water World	1
		Van Riebeck Swimming Pool	1
	Restaurant	Yoga next to the pool	1
		Jaika str 3	1
	Rock Club	Sasol Rally	1
		TAB	1
	Scenic Lookout	Bohemian Groove Cafe	1
		Bluemoon	1
	Skating Rink	Inniebos Rock Slane	1
		Hezekiah Lodge	1
	Ski Lodge	Kaapschehoop Escarpement Viewpoint	1
		Photos Of Africa	1
	Sports Club	Ice Rink	1
		Sunset Manor Lodge	1
	Tennis Court	Mbombela Stadium	1
		Nelsville	1
	Theme Park	LYCC	1
		Lowveld show Grounds	1
	Water Park	Indoors Action Cricket	1
		Matunyane	1
social	African Restaurant	Matunyane slide park riverside mall	1
		Donata African Food	1
	Bakery	Kuzuri	1
		Bakers bin	1
		30 something	1
		Paddy's pub	1
		Salvador	1
	Bar	Shooters	1
		ons hookantloor	1
	Beach Bar	Vibes - Pool, Bar & Grill	1
		Phat Boys Beer House	1
	Breakfast Spot	Arkansas Spur Steak Ranch	1
		Koek 'n Pan	1
	Brewery	Distell trade express	1
		Lions Pub & Grill	1
	Burger Joint	Rocco Mamas	1
		Wimpy	3
	Café	Capouccino's Cafe & Pizzeria	1
		Mugg & Bean	2
		The olive leaf	1
		Wild Bean Cafe	1
	Cocktail Bar	Capello	1
		Carisma Night Club	1
	Coffee Shop	Coffee, Cakes n Things	1
		Eat, Pray & Love	1
		Nadines Elmswood	1
		Nadines Montana	1
		Puzzle Coffee Shop	1
	Convenience Store	The Tea Garden Nelspruit Botanical Gardens	1
		Wiesenhof coffee shop	1
		SPAR Express at SHELL Chippos	1
		SPAR Mini	1
	Deli / Bodega	SUPERSPAR Riverside	1
		SPAR The Village	1
	Diner	Bon Appetito / The Grove Superspar	1
		Maxi	1
	English Restaurant	Fruit and Veg	1
		Gallito's Nelspruit Brown Street	1
	Fast Food Restaurant	Gallito's Nelspruit Halls Gateway	1
		McDonald's	1
		Steers	3
		Twinz Burger & Ribs - Pizza	1
		Mugg & Bean	1
	Food Truck	the chicken coop	1
		Wakaberry frozen voohurt	1
	Frozen Yogurt Shop	SPAR	1
		Nightclub	1
	Grocery Store	Capello	1
		Mbombela Chesanyama	1
	Other Nightlife	Debonairs Express	1
		Debonairs Pizza	1
	Pizza Place	Debonairs Westend	1
		Nando's	1
	Portuguese Restaurant	Nando's Ilango Mall	1
		Covote's Pub	1
	Pub	Jock & Java	1
		Martin Pub And Grill	1
	Restaurant	The Pub Nelspruit	1
		Cuban 8 Restaurant	1
		Feast	1
		Jos Macs	1
		Kuzuri Restaurant & Venue	1
	Seafood Restaurant	Max's White River Square	1
		Saffron Restaurant	1
	Sports Bar	3 Dara Seafood	1
		upper level	1
	Steakhouse	Turn n' Tender	1

Categories with no venues listed

Which categories did Foursquare respond with no venues?

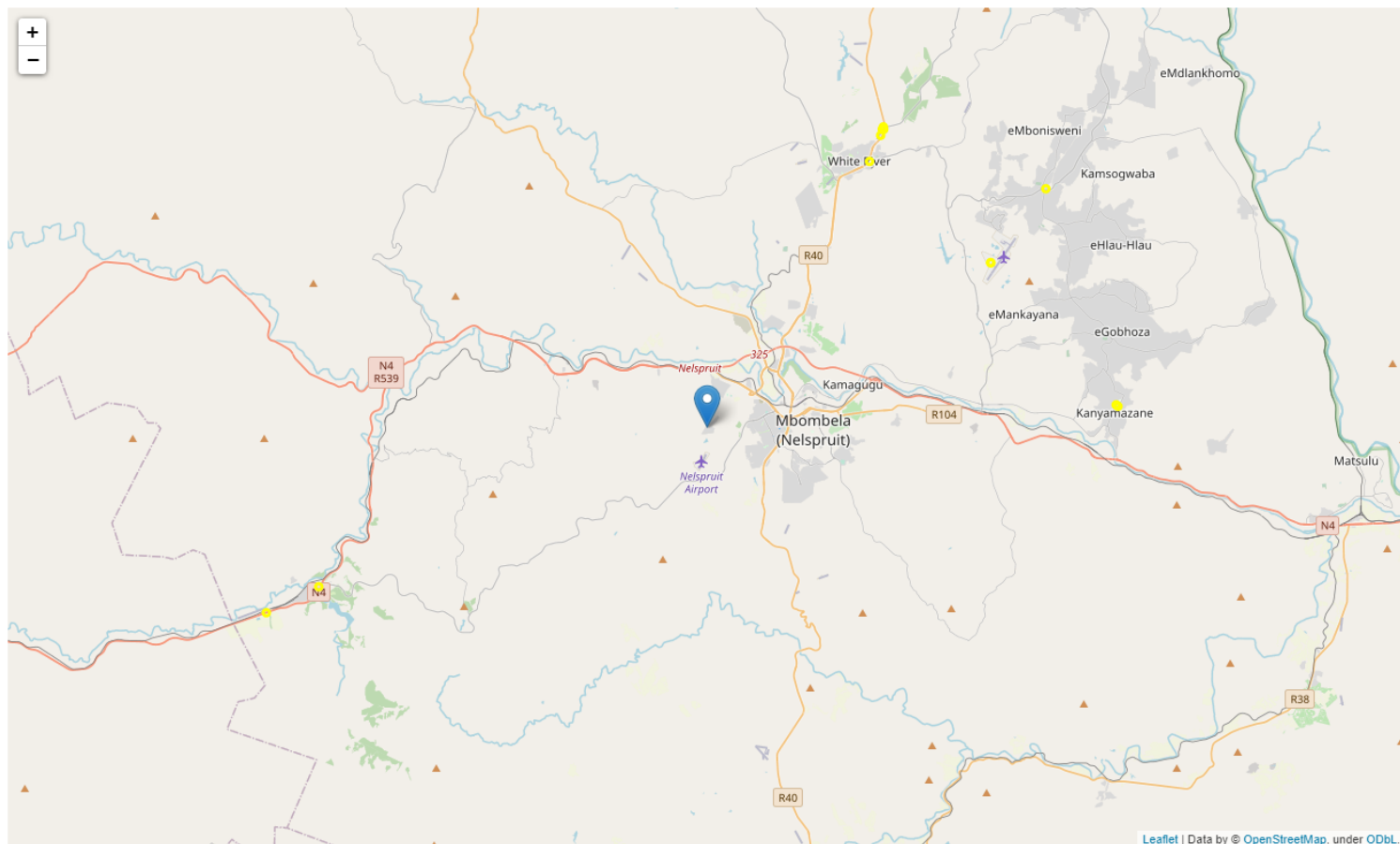
Category Name	Response Count
Events	0
Recreational Center	0
Roof Deck	0
Sculpture Garden	0
Vineyard	0
Gaming Cafe	0
Bike Trail	0
Dive Spot	0
Fishing Spot	0
Forest	0
Gun Range	0
Nature Preserve	0
Palace	0
Rafting	0
Rock Climbing Spot	0
Skydiving Drop Zone	0
Summer Camp	0
Volcano	0
Waterfall	0

3.1.2 Plot data

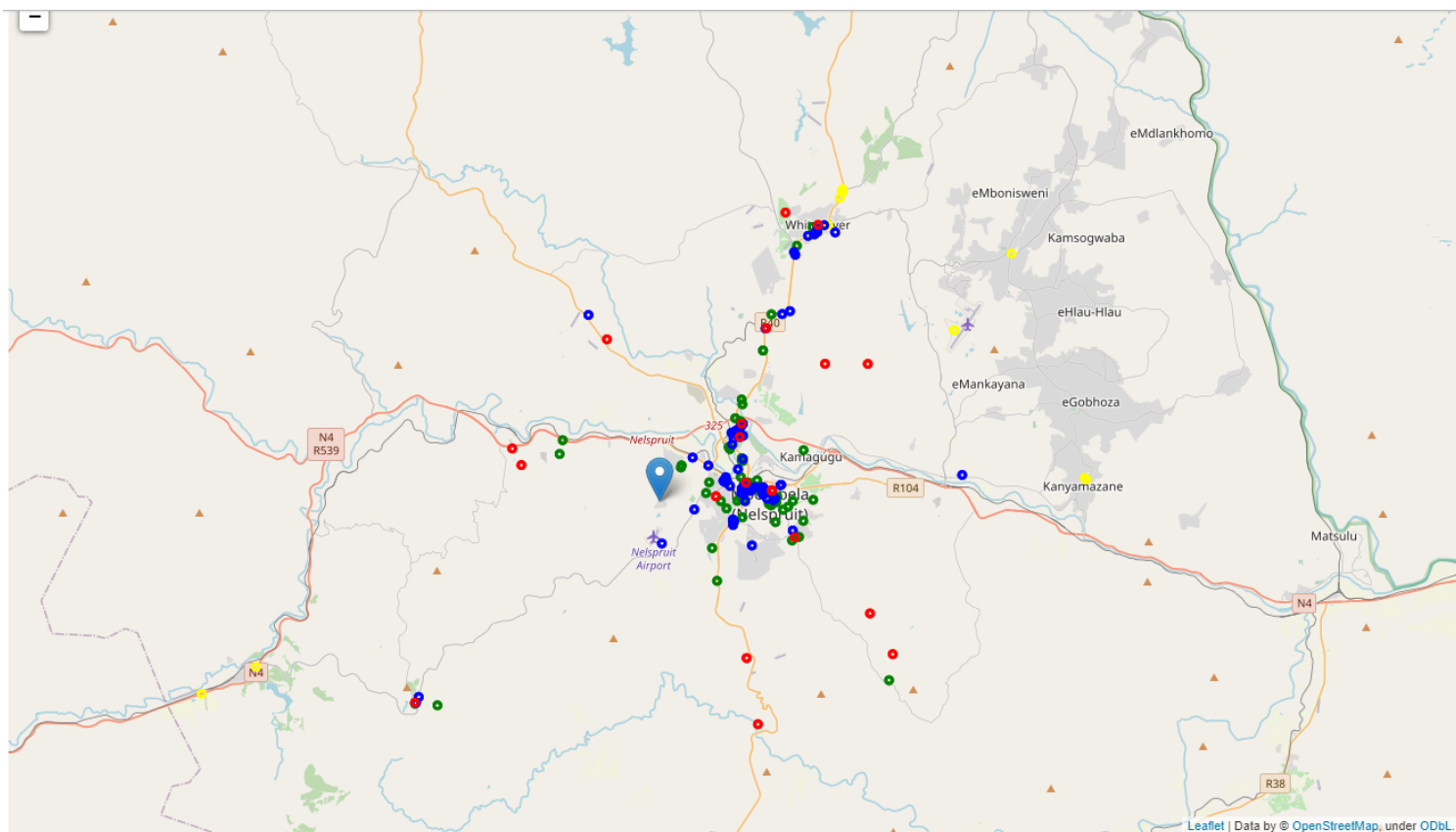
Here we explored the data by plotting them on a map

For 'Plotting of the Target City, the few venues just outside the 20km radius, and all the venues according to their groups.'

Just outside 20km



All venues per group



Social venues are colored **blue**
 Adventure venues are colored **red**
 Fun venues are colored **green**

3.1.3 Average distance from center

Per category

```
avg_dst_grp.set_index('Category').head(15)
```

	Group	Distances
Category		
Botanical Garden	adventure	6.769587
Campground	adventure	12.656354
Castle	adventure	3.582697
Cave	adventure	7.093683
Lake	adventure	19.550869
Mountain	adventure	14.843822
National Park	adventure	7.089274
Other Great Outdoors	adventure	13.827276
Park	adventure	8.681033
Trail	adventure	13.137054
Arcade	fun	6.819588
Art Gallery	fun	7.377321
Athletics & Sports	fun	8.616603
Casino	fun	7.073137
Cocktail Bar	fun	5.365071

Per group

Distances	
Group	
adventure	11.457194
fun	7.474814
social	8.488651

3.2 Clustering Models

Here I prepared the data for the clustering model - Density-based model.

Then I created the model applicable to geographical data, then fitted it.

Questions to answer:

- What are the statistics of the clusters?
- How many venues are per cluster?
- What is the average distance between cluster and city center?
- Plot the clusters on a map.

3.2.1 DBScan

Known as DBSCAN for Density Based Spatial Clustering of Application with Noise

Finds core samples of high density and expands clusters from them. Good for data which contains clusters of similar density.

There were 4 parameters important to setting the **model, epsilon, min_samples, algorithm, and metric**.

Epsilon - The maximum distance between two samples for one to be considered as in the neighborhood of the other. This is not a maximum bound on the distances of points within a cluster.

Min Samples - The number of samples (or total weight) in a neighborhood for a point to be considered as a core point. This includes the point itself.

Algorithm - The algorithm to be used by the NearestNeighbors module to compute pointwise distances and find nearest neighbors.

Metric - The metric to use when calculating distance between instances in a feature array.

The parameters were set to use the haversine formula to calculate distance of instances after being converted to radians.

Statistic

What are the statistics of the clusters?

The model Clustered 157 points down to 4 clusters

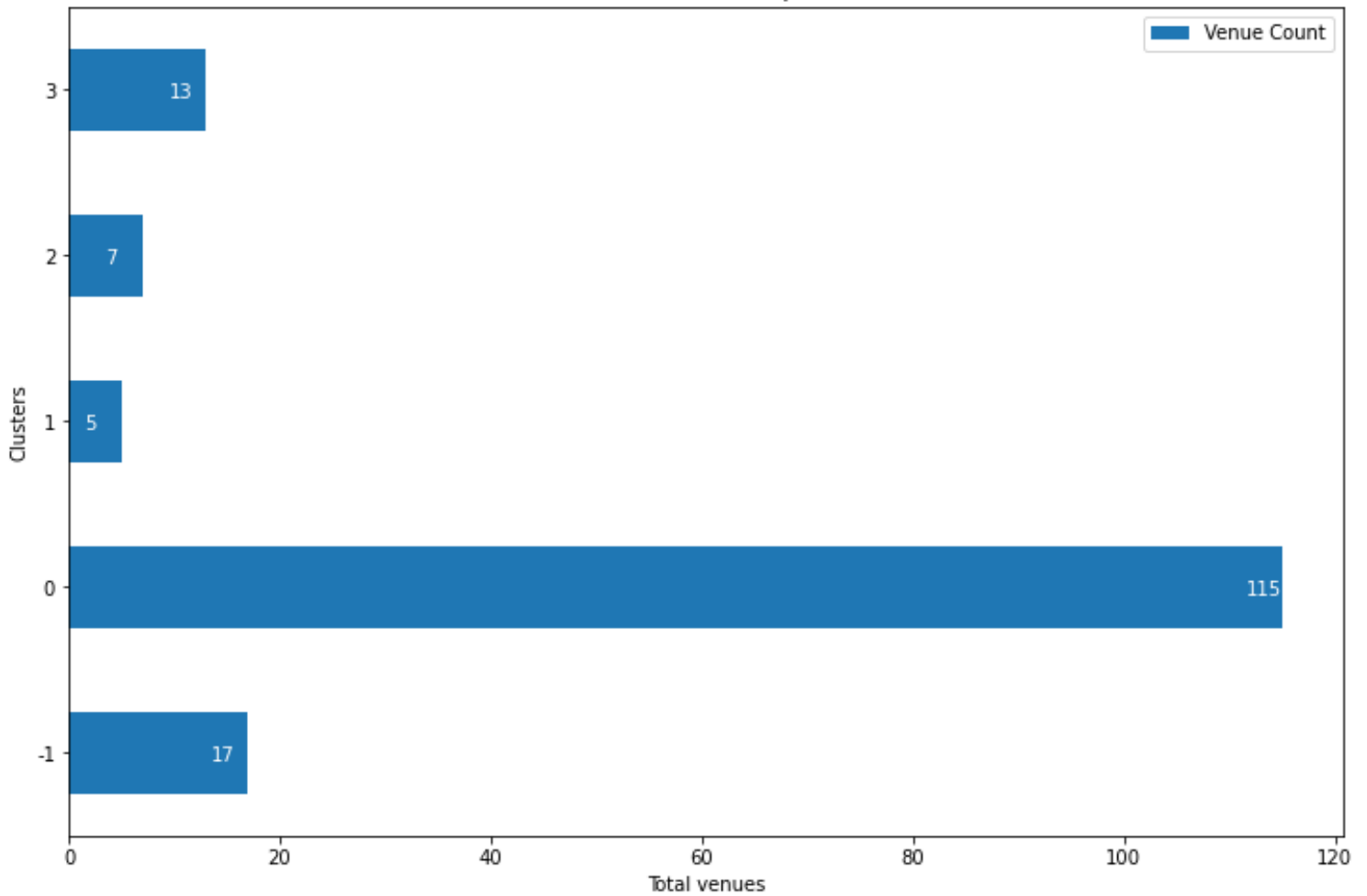
and received a Silhouette coefficient score of 0.559

Silhouette Coefficient determines how accurate the model has assigned points to a cluster based on the coordinates. With -1 being the worst, 0 meaning overlapping clusters, and 1 the best score. Our model seems to have done fine. I have not tested with varying epsilon and min_samples due this assignment already being submitted late.

Venues per cluster

How many venues are per cluster?

Total amount of venues per cluster



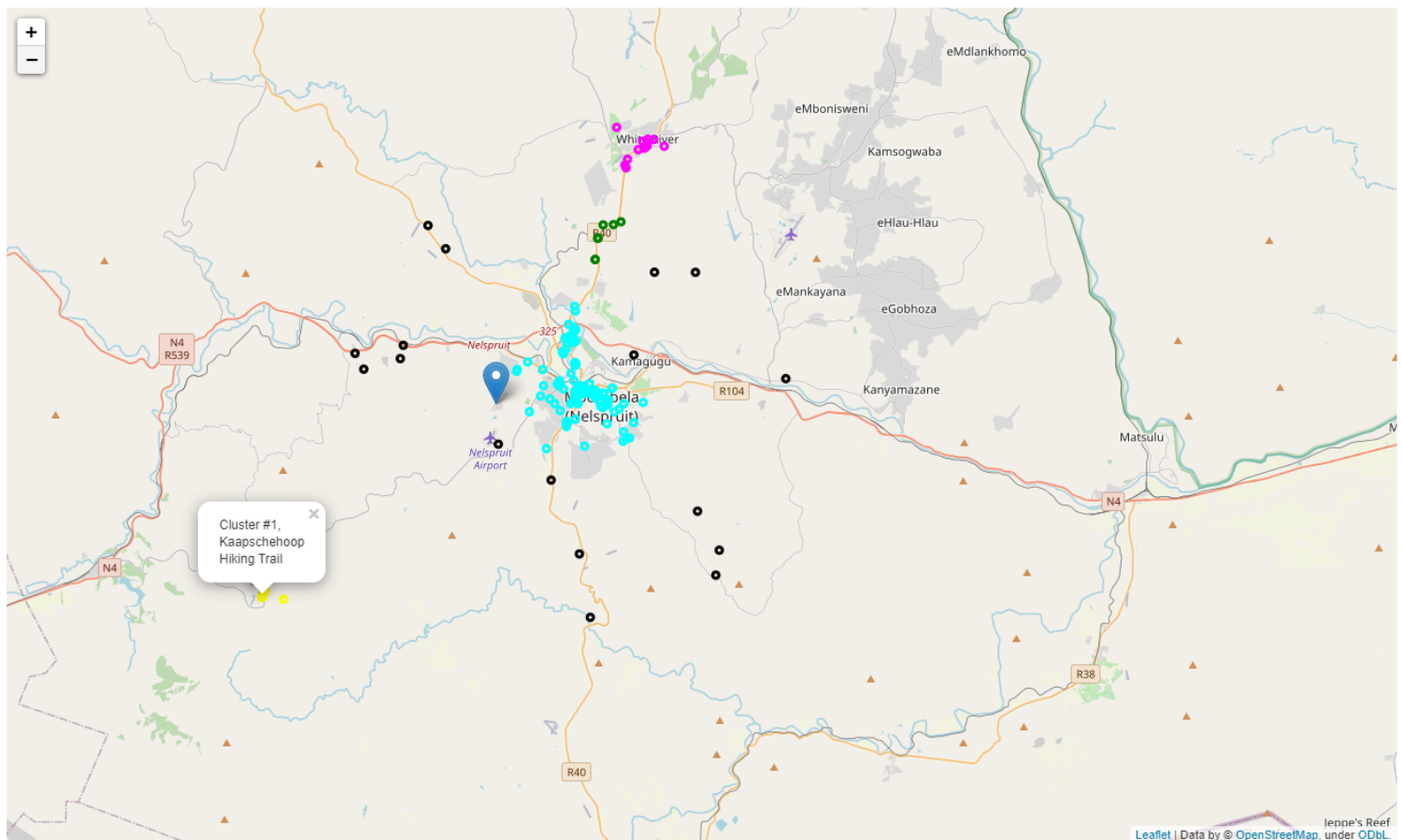
Average distance per cluster to center

What is the average distance between cluster and city center?

Distances	
Clusters	
-1	11.639542
0	6.049674
1	19.299549
2	12.838859
3	19.091523

Plotting clusters on the map

Plot the clusters on a map.



4. Results

We have solved the problem, we now have a dataset containing venues of the 3 main groups of entertainment, in my opinion. It seems we have less than 180 venues within 20km radius from Nelspruit. Analysis shows that venues in Witrivier, Kaapsechoop, and Kanyamezane also.

There are 70 venues in the Social group, 66 venues in the Fun group, 21 venues in the Adventure group.

These are more than enough venues to provide a selection for quality entertainment and socializing. For further interest we used a density-based clustering model to cluster the points on the map and calculate their average distance from the city center: Cluster #0 is 6km, #1 19.2km, #2 12.8km, and #3 19km Cluster 0 being the closest to Nelspruit would be less costly to travel to.

5. Discussion

According to locals, Kaapsechoop's Trails and Nelspruit's social places come recommended.

Looking at the definitions of the groups

Adventurous - "willing to take risks or to try out new methods, ideas, or experiences." - Google. Simpler, "outdoor or risking", the categorization is relative to the individual, thus only the outdoor venues that relate to be active by nature have I considered as adventurous, but artificial passive participation like browsing a plant nursery or garden does not qualify as adventurous, but botanical gardens qualifies due to the size and "wildness".

Fun - "enjoyment, amusement, or light-hearted pleasure." - Google.

Social - "an informal social gathering." - Google.

A recommendation from my side will be to update Foursquare's database in order to get more venues that might not be listed in this part of the world.

6. Conclusion

In this study, I explored venues in a City in South Africa which are of specific categories within groups. I've extracted all the venues pertaining to Social, Fun, and Adventurous types. Then further analyzed their position and distance from the center of target city. How much venues there are per group and category. I then used a Density-based Machine Learning model to cluster the venues close to each other. I calculated their distances from city center. Plotted them on a map to see where the clusters are situated.

In conclusion, Nelspruit has the most venues and to explore further the clusters in Whiteriver, Kaapsehoop, and then Rocky's Drift. With Venues just further outside of the city.

Now I am able to view the venues on a map or search from a list whenever I am looking for Social, Fun, or Adventurous time.

References

Using geographical data in DBSCAN model - [\[link\]](#)[\[link\]](#) - notebook]

Grouping categories - [\[link\]](#)

Pandas - Merging dataframes - [\[link\]](#)

Silhouette Coefficient - [\[link\]](#)