

Week 2 Assignment

For the second week, the final deliverables of the project will be:

1. A link to your Notebook on your Github repository, showing your code. **(15 marks)**
2. A full report consisting of all of the following components **(15 marks)**:
 - Introduction where you discuss the business problem and who would be interested in this project.
 - Data where you describe the data that will be used to solve the problem and the source of the data.
 - Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
 - Results section where you discuss the results.
 - Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
 - Conclusion section where you conclude the report.
3. Your choice of a presentation or blogpost. **(10 marks)**

PROJECT REPORT

1. Business Problem

Construction Company XYZ Investments is interested in building a massive shopping mall that will be housing mostly international and a number of local businesses, the mall will also have mostly high end restaurants. The mall should be located in one of the Gauteng cities but shouldn't be too close to an area where other malls are popular as this will increase competition especially on prices.

a. Stakeholders

- XYZ Investment Company (main investor)
- Government
- Investors

2. Data Description

The data that will be used is geographical codes and coordination of South African Provinces, this data will be vital in identifying the cities/neighborhoods across the provinces.

The second data will be coming from the Foursquare site; this data will be utilized to identify business types across the cities/neighborhoods. The data will also assist in clustering cities/neighborhoods according to the types of businesses.

a. Data Sources

- <https://dirkstrauss.com/south-african-cities/>
- <https://foursquare.com/>

3. Methodology/Approach

a. Machine Learnings Used

- K-means clustering was utilized to create 5 clusters utilizing the top businesses across cities/neighborhoods.

4. Results Discussion

- Cluster 1: The most common venues in these neighborhoods are mainly small businesses and a couples of entry level fast food restaurants.
- Cluster 2: The most common venues in these neighborhoods comprises of a good mix of small and big businesses such as shopping malls, food courts, super markets and gas stations.
- Cluster 3: The most common venues in these neighborhoods in this cluster are mainly convenient stores, fast food restaurants and flower shops which are fairly small businesses.
- Cluster 4: The most common venues in these neighborhoods are mainly basic necessities such as Pharmacies, Yoga Studios, Fast food restaurants etc.
- Cluster 5: The most common venues in this neighborhoods are small businesses and convenient stores, these include gas stations and Yoga Studios.

5. Observations/Recommendations

- Out of the five clusters neighborhoods we examined in the Gauteng Province, Cluster 3's neighborhoods comprise of a good mix of businesses. This is a sign of neighborhoods with a fair amount of economic activity that is worth exploring further.
- The mix of the types of restaurants in the neighborhoods in this cluster can be clustered in a shopping mall set up, by implication bring these types of venues together in a mall will be a viable business option as there is a need for them is already established.
- Whilst across the five clusters there is no explicit mention of clothing shops, this becomes a market that is worth tapping in a shopping mall set up. Cluster 3 therefore will make more business sense.

6. Conclusion

With the analysis performed on the all the clusters, it is worth noting that with the top 10 venues identified from the data set for all the clusters conclusion can be made that the results are a true reflection of the cities/neighborhoods overall data set.

XYZ Investment can therefore consider building a shopping mall in the neighborhoods within the Cluster 3 where there are diverse business types.