

IBM Data Science Capstone Project

The Battle of Neighbourhoods: Analysing Business Opportunities in Sandakan

Presented by Dennis Lam (August 2019)

Agenda

- Business Problem
- Data
- Methodology
- Results
- Discussion
- Recommendations
- Conclusion

Business Problem

- Business owners want to find suitable location in neighbourhoods area
- Project objective is to analyze and recommend suitable areas

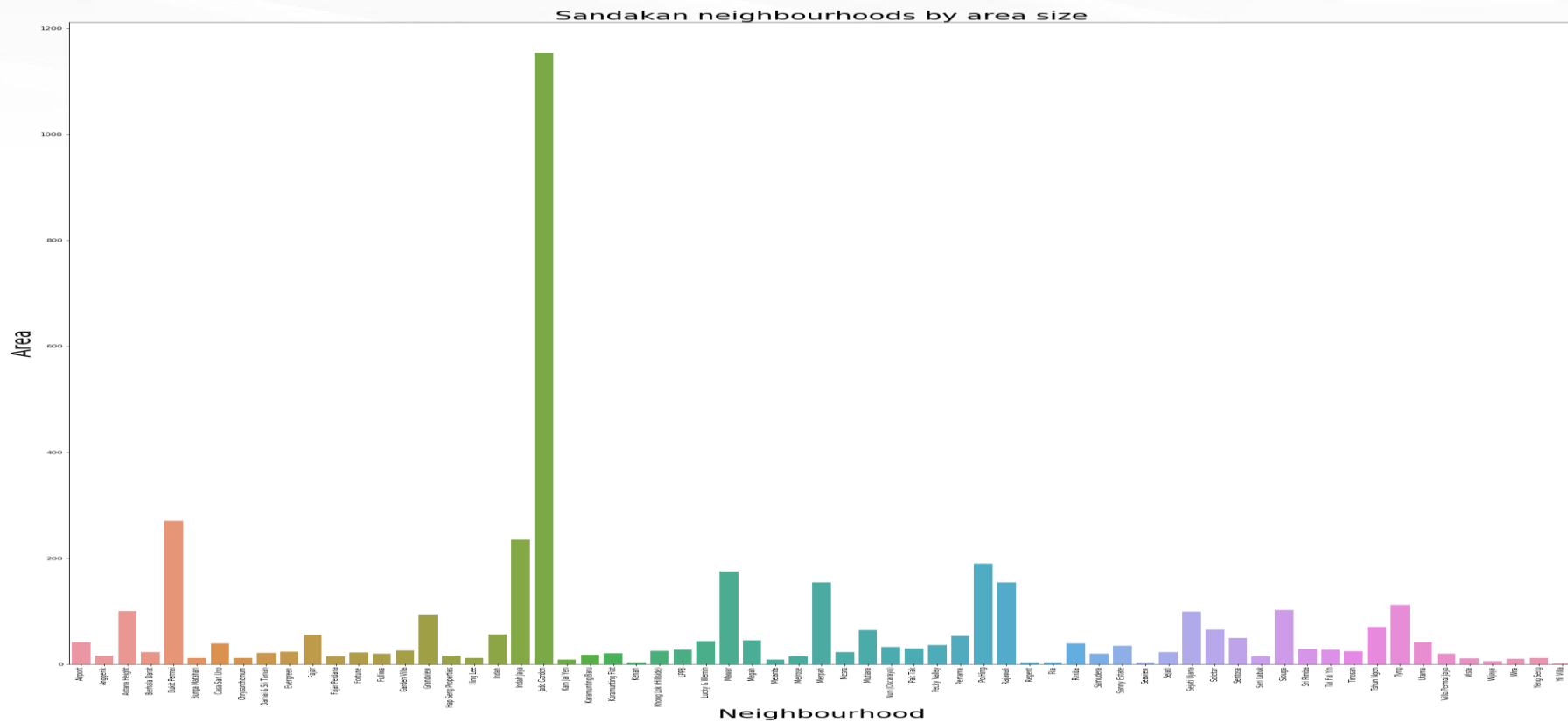
Data

- List of neighbourhoods in Sandakan
- Latitude and Longitude of each location
- Venues data
- Sources of data
 - Sandakan Municipal Council website
 - Geopy Python for location coordinates
 - Venues data using FourSquare API

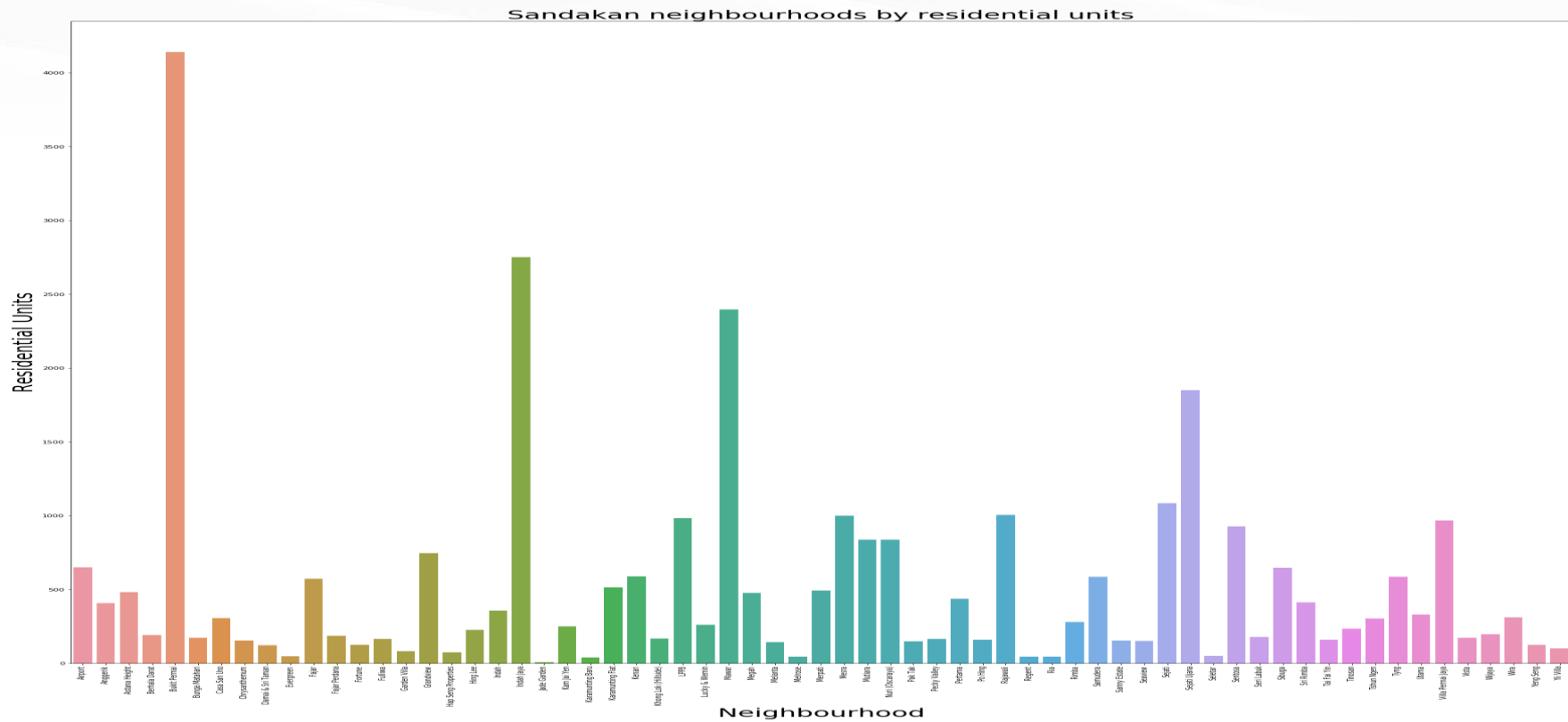
Methodology

- Gather all sources of data
- Create a csv file
- Create pandas dataframe and do data exploration
- Focus and map locations
- Use Foursquare API for venue data
- Perform clustering using K-means
- Visualise the clustering results

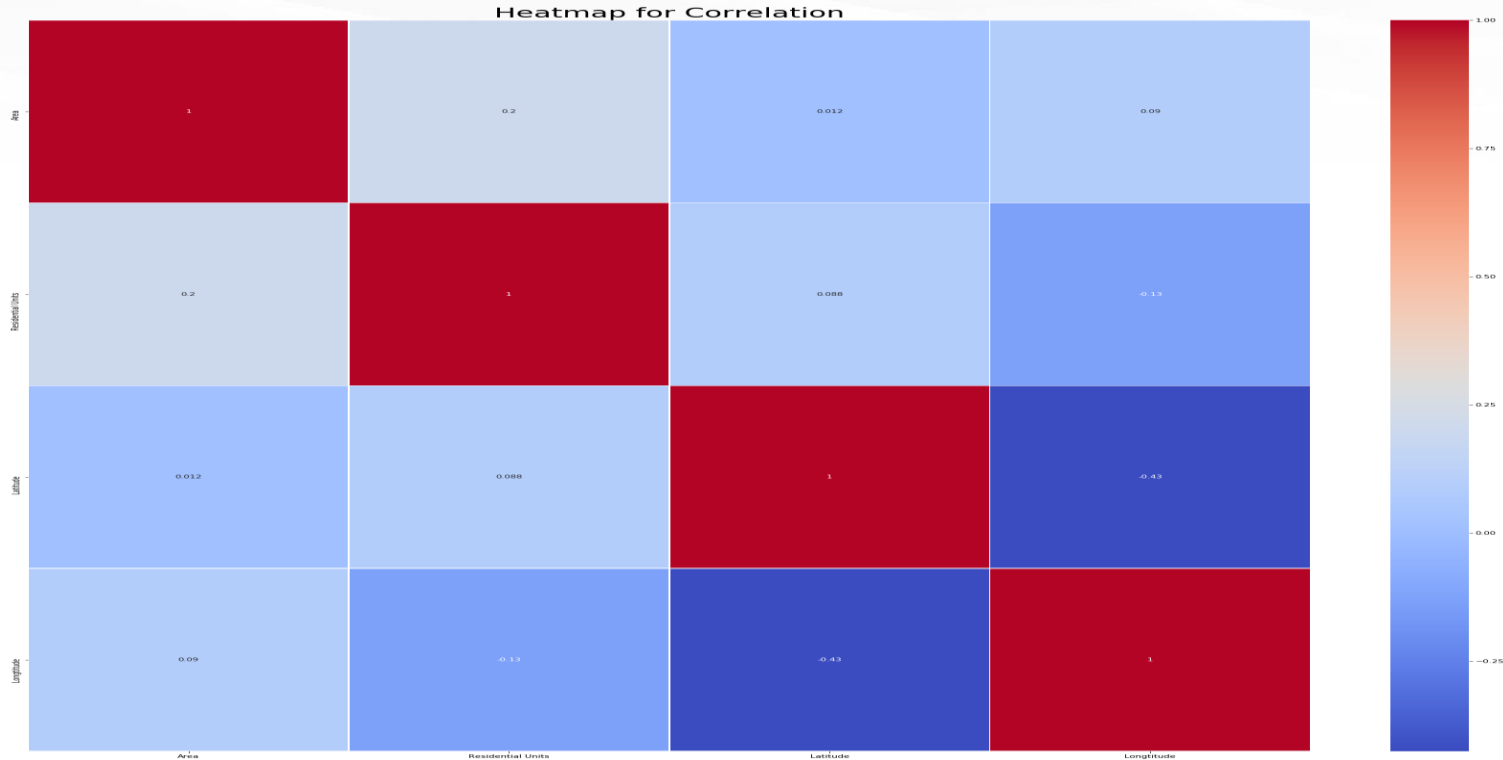
Sandakan neighbourhoods by area size



Sandakan neighbourhoods by residential units

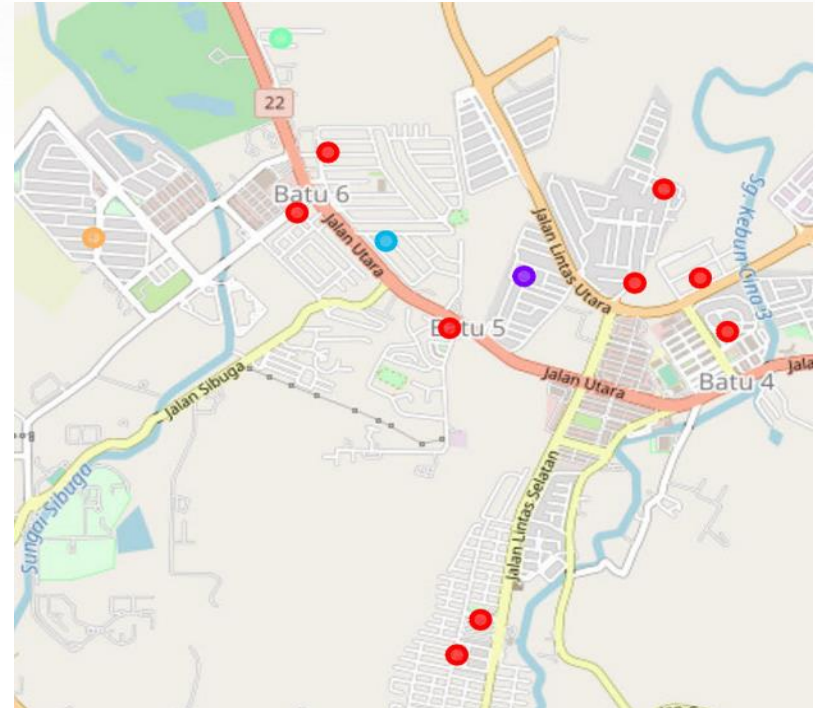


Heatmap to illustrate Correlation



Results

- Cluster 0: These neighbourhoods are concentrated
- Cluster 4: Good potential in near future
- Cluster 1,2 & 3: Stand-alone neighbourhoods



Discussion

- Cluster 0 has highest number of neighbourhoods
- Cluster 1, 2, 3 and 4 has one neighbourhood each

Recommendations

- Cluster 0 is most desirable place to set up businesses.
- Cluster 4 has good business potential in future as new buildings are being constructed there
- The remaining clusters are not recommended due to stand-alone and lower population

Conclusion

- The project has achieved its objective of classifying neighbourhoods with common venues each
- Cluster 0 is identified as best location to set up new business
- The project findings can assist business people to make decisions on business setup

Thank you

