Capstone Project: Battle of Neighborhoods

F&B Business Prospect and Venues Data Analysis in Bali, Indonesia

by Nur Cahyo Nugroho

www.nurcahyonugroho.com

Problem Statement

Business owner will need to perform research and exploration of the neighborhoods before open any business in specific location to understand many factors, like nearby residential area, tourist place & office, another similar competitor, public facilities, and most importantly average rental fare.

In this project I have chosen Bali as place to explore if business owner/investor want to open F&B (food and beverage) business in this famous tourism place.

This project will also help to compare different sub-districts in Bali to choose best suited location based on top 5 venues surround it.

Objective

This project will give insight to restaurant business owner/investor to comparing neighborhoods based on sub-districts in Bali to choose best suited location based on top 5 common venues surrounding it.

Methods:

- Web scraping the list of district and population in Bali from Wikipedia.
- Extracting top trending venues using Foursquare API.
- Forming sub-district clusters based on venue categories using unsupervised *k-means* clustering algorithm.
- Understanding the similarities and differences between district to retrieve more insights and to conclude which sub-district is best suited for business prospect.

Data Reference & Library

Data Reference for Dataset:

- Wikipedia: https://id.wikipedia.org/wiki/Daftar_kabupaten_dan_kota_di_Bali.
- Sub-district geolocation: https://raw.githubusercontent.com/bluearticuno/Coursera_Capstone/master/Bali_Geospatial.csv

Libraries:

- Panda libraries for dataframe and other dataset manipulation.
- Numpy for any scientific computation.
- Requests to call Foursquare API.
- KMeans cluster from sklearn for clustering.
- Matplotlib for plotting modules.
- Seaborn for bar graph plotting.
- Folium for map plotting.

Workflow

- Web scrapping, data extraction, and data wrangling.
- Extracting nearby venues and cluster using K-means algorithm.
- Data visualization based on clustered result.

Data Preparation

1 Web Scrapping for district list and population from Wikipedia

	District	District Population
0	Kabupaten Badung	468.346
1	Kabupaten Bangli	264.945
2	Kabupaten Buleleng	814.356
3	Kabupaten Gianyar	492.757
4	Kabupaten Jembrana	323.211
5	Kabupaten Karangasem	545.389
6	Kabupaten Klungkung	215.852
7	Kabupaten Tabanan	466.647
8	Kota Denpasar	638.548

2 Read csv file for geolocation for sub-districts from GitHub

	Kabupaten/Kota	Kecamatan	Latitude	Longitude
0	Kabupaten Badung	Abiansemal	-8.5584	115.2245
1	Kabupaten Badung	Kuta	-8.7238	115.1752
2	Kabupaten Badung	Kuta Selatan	-8.8063	115.1533
3	Kabupaten Badung	Kuta Utara	-8.6489	115.1593
4	Kabupaten Badung	Mengwi	-8.5617	115.1771
5	Kabupaten Badung	Petang	-8.3366	115.2245
6	Kabupaten Bangli	Bangli	-8.4330	115.3608

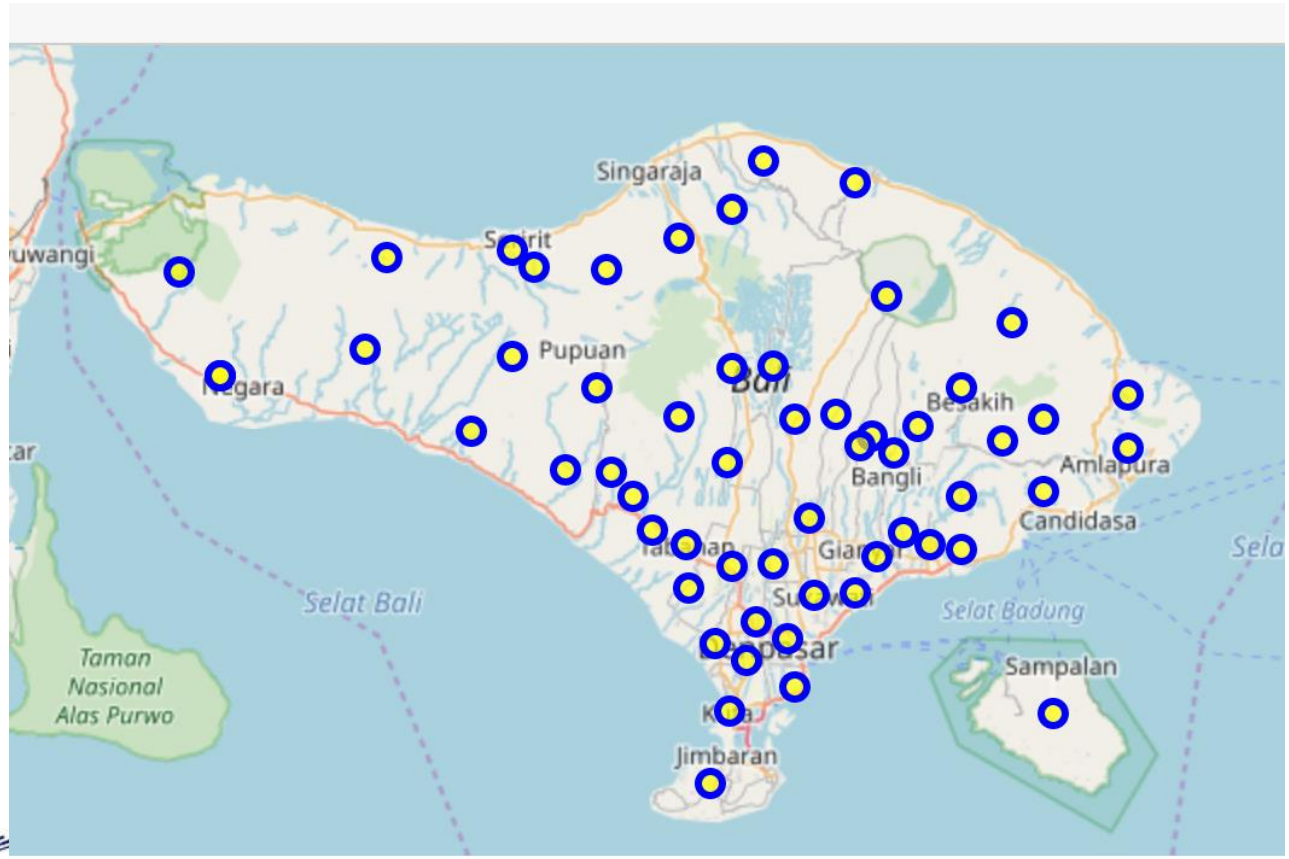
merged dataset

3

	District	District Population	Sub-District	Latitude	Longitude
0	Kabupaten Badung	468346	Abiansemal	-8.5584	115.2245
1	Kabupaten Badung	468346	Kuta	-8.7238	115.1752
2	Kabupaten Badung	468346	Kuta Selatan	-8.8063	115.1533
3	Kabupaten Badung	468346	Kuta Utara	-8.6489	115.1593
4	Kabupaten Badung	468346	Mengwi	-8.5617	115.1771
5	Kabupaten Badung	468346	Petang	-8.3366	115.2245
6	Kabupaten Bangli	264945	Bangli	-8.4330	115.3608

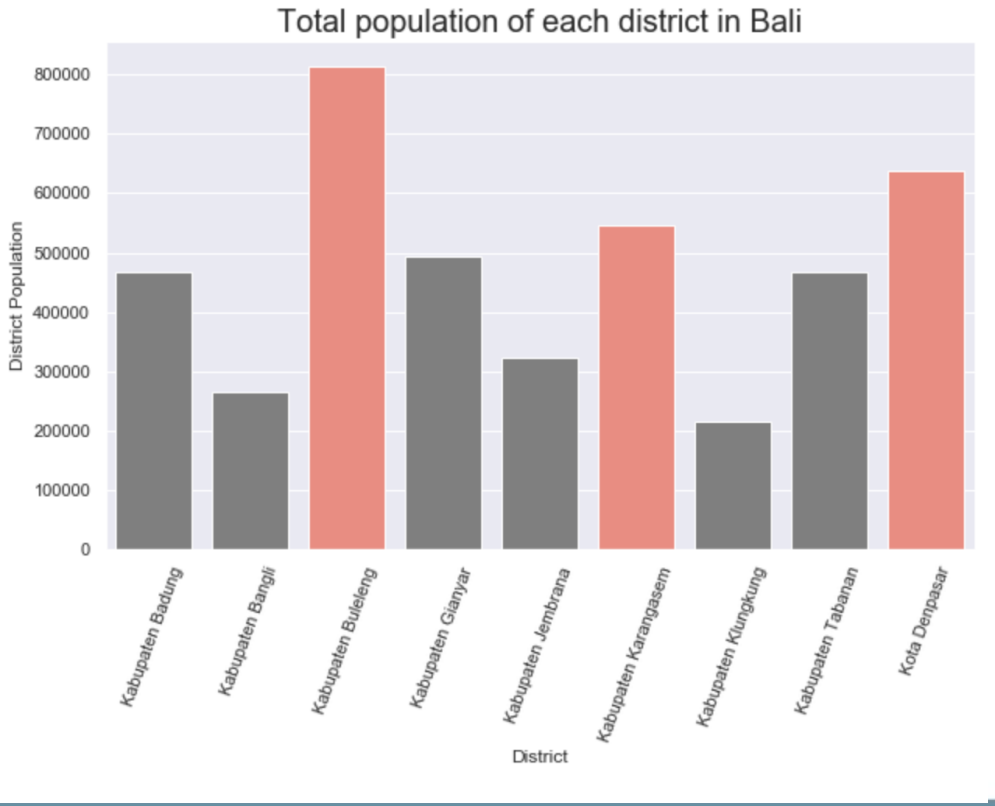
Data Visualization

- Plot using folium for all districts in Bali map.



Data Visualization

- Plot using seaborn bar graph to show population on each district, and highlighted district that has population more than 500,000 people.



Extract Top 5 Venues

1 Call Foursquare API to get nearby venues

	name	categories	lat	lng
0	Munduk Moding Plantation Resort Bali	Resort	-8.240019	115.070711
1	Munduk	Mountain	-8.263634	115.059512
2	Damai Lovina Villas Bali	Hotel Bar	-8.181425	115.043972
3	Puri Lumbung Cottages	Hotel	-8.264693	115.059353
4	Sanak Retreat Restaurant	Balinese Restaurant	-8.253628	115.029540

2 Merge with sub-district data

	Sub-District	Sub-District Latitude	Sub-District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Sawan	-8.1587	115.1771	Cafe cocacola beten bingin	-8.156245	115.178318	Bar
1	Seririt	-8.2053	114.9280	Santal	-8.201068	114.927315	Lake
2	Seririt	-8.2053	114.9280	Pantai Seririt	-8.200984	114.927610	Beach
3	Karangasem	-8.4297	115.6273	Jasri beach	-8.428145	115.630224	Beach
4	Denpasar Barat	-8.6666	115.1948	Starbucks Coffee	-8.664219	115.198003	Coffee Shop
5	Denpasar Barat	-8.6666	115.1948	Terang bulan dan martabak 'Sedap Mantap'	-8.665786	115.198667	Food Truck
6	Denpasar Barat	-8.6666	115.1948	Coco Express	-8.664558	115.195004	Convenience Store
7	Denpasar Barat	-8.6666	115.1948	Indomaret	-8.664604	115.197508	Convenience Store
8	Denpasar Barat	-8.6666	115.1948	EKA Print Rinjani	-8.665185	115.198821	Print Shop
9	Denpasar Selatan	-8.6983	115.2482	Cupa-Cupa Corner (CCC)	-8.697910	115.251220	Fast Food Restaurant

3 Top 5 venues

	Sub-District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Denpasar Barat	Convenience Store	Print Shop	Food Truck	Coffee Shop	Skate Park
1	Denpasar Selatan	Skate Park	Noodle House	Fast Food Restaurant	Asian Restaurant	Print Shop
2	Denpasar Timur	Indonesian Restaurant	Bookstore	Balinese Restaurant	Bakery	Skate Park
3	Denpasar Utara	Noodle House	Convenience Store	American Restaurant	Skate Park	Print Shop
4	Karangasem	Beach	Skate Park	Print Shop	Noodle House	Lake

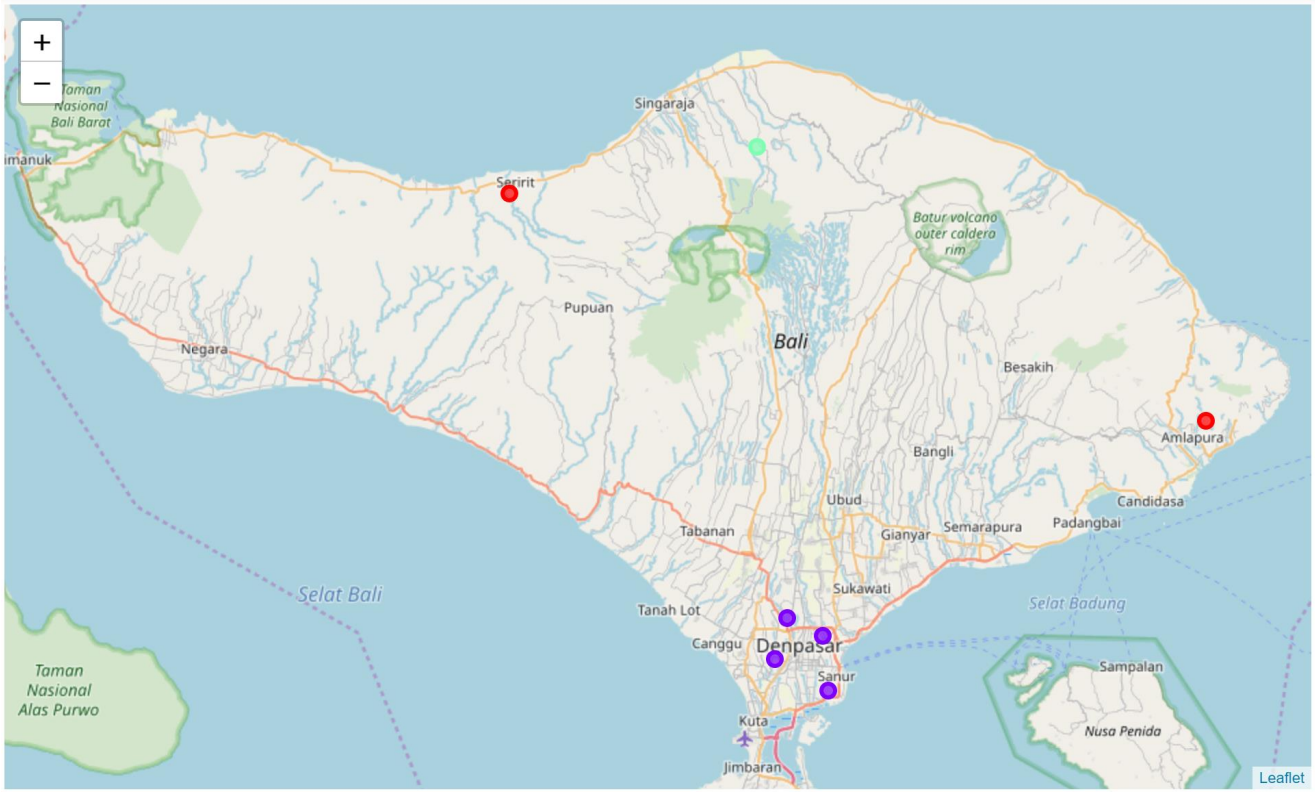
Cluster with K-means

- Use Kmeans cluster from sklearn library, with input cluster = 3. Merged data of sub-district with cluster label.

Sub-District	District	District Population	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Sawan	Kabupaten Buleleng	814356	-8.1587	115.1771	2.0	Bar	Skate Park	Print Shop	Noodle House	Lake
Seririt	Kabupaten Buleleng	814356	-8.2053	114.9280	0.0	Lake	Beach	Skate Park	Print Shop	Noodle House
Karangasem	Kabupaten Karangasem	545389	-8.4297	115.6273	0.0	Beach	Skate Park	Print Shop	Noodle House	Lake
Denpasar Barat	Kota Denpasar	638548	-8.6666	115.1948	1.0	Convenience Store	Print Shop	Food Truck	Coffee Shop	Skate Park
Denpasar Selatan	Kota Denpasar	638548	-8.6983	115.2482	1.0	Skate Park	Noodle House	Fast Food Restaurant	Asian Restaurant	Print Shop

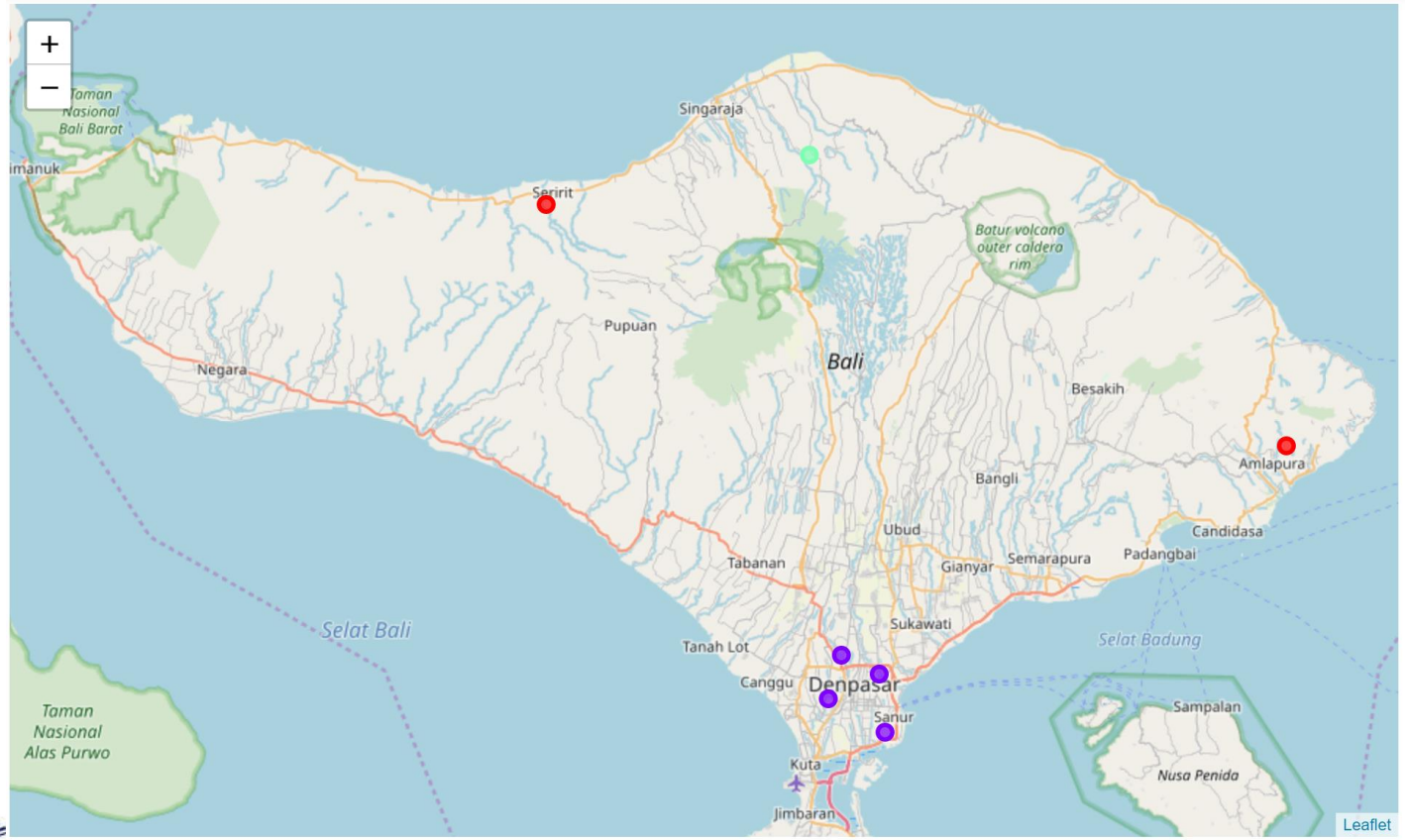
Result

- Plot cluster in Bali map:



Result

- Plot cluster in Bali map:



Result

- Comparison for sub-districts that has nearby F&B venues:

Sub-District	Denpasar Timur	Denpasar Barat	Sawan
District	Kota Denpasar	Kota Denpasar	Kabupaten Buleleng
District Population	638548	638548	814356
Latitude	-8.6431	-8.6666	-8.1587
Longitude	115.242	115.195	115.177
Cluster Labels	1	1	2
1st Most Common Venue	Indonesian Restaurant	Convenience Store	Bar
2nd Most Common Venue	Bookstore	Print Shop	Skate Park
3rd Most Common Venue	Balinese Restaurant	Food Truck	Print Shop
4th Most Common Venue	Bakery	Coffee Shop	Noodle House
5th Most Common Venue	Skate Park	Skate Park	Lake

Conclusion

As a result of clustering and comparison, people will have more insight if they want to open business in specific location. With this information, business owner can achieve better decision through their access availability to the reports/platforms where such information is provided.

Not only for business owner/investor, it will be usefull for local government as well to understand about the business diversity and prospect in their local area.

Reference

Visit my GitHub for detail of the python notebook:
https://github.com/bluearticuno/Coursera_Capstone/

or alternatively can visit:
https://nbviewer.jupyter.org/github/bluearticuno/Coursera_Capstone/blob/master/Capstone-business_prospect_in_Bali_v2.ipynb

Thank You