Bali Indonesia Regency Recommendation for Food Lover Tourist

By: Bima Putra Pratama

1. INTRODUCTION

1.1 Background

Bali is an Indonesian island known for its forested volcanic mountains, iconic rice paddies, beaches and coral reefs. The island is home to religious sites such as cliffside Uluwatu Temple. To the south, the beachside city of Kuta has lively bars, while Seminyak, Sanur and Nusa Dua are popular resort towns. The island is also known for its yoga and meditation retreats.

Other than that, Bali also have many choice of food venue that worth to try. In order to experienced it better, living near the location that surrounded by food can be a good when visiting Bali.

1.2 Purpose

The intention of this analysis is to give suggestion for tourist where is the best location to stay in Bali based on the surrounded food venue.

1.3 Interest

For this project, we will clustering administrative division of Bali based on the food venue around. So this analysis result can be used by tourist to pick their place to stay before visiting Bali. More over the result of this analysis also can be used by Ministry of Tourism as a guide to develop the specific administrative location for better food tourism experience.

2. DATASET

The data for this project is the most popular administrative regency in Bali. This data were picked through Bali Wikipedia Page that can be found through this link https://en.wikipedia.org/wiki/Bali. The popular regencies in Bali are Denpasar, Tabanan, Gianyar, Klungkung, and Bangli.

In addition, I got the latitude and longitude location for each popular regencies through OpenStreetMap API by using *Nominatim* function. Then I used foursquare API to get the top venue surrounded in each regency.

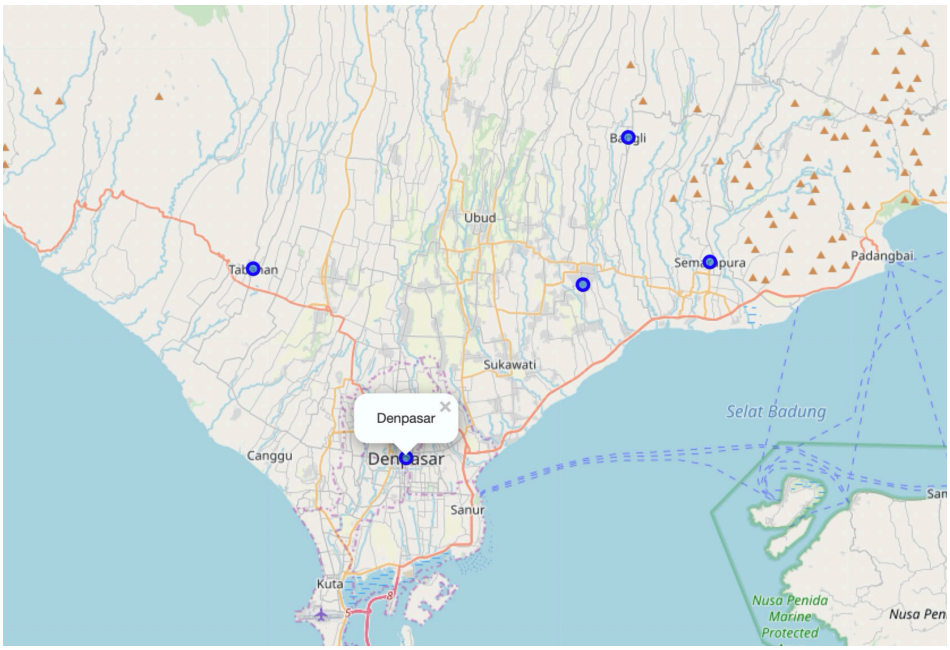
3. METHODOLOGY

3.1 Data Preparation

There are several steps involved to do this analysis. First I prepared the dataset by creating the data frame that have popular regencies inside that. Then I did geocoding with *Nominatim* function to get the coordinate for each regencies.

	Regency	Lat	Long
0	Denpasar	-8.652497	115.219117
1	Tabanan	-8.539231	115.126568
2	Gianyar	-8.548236	115.326054
3	Klungkung	-8.535017	115.403276
4	Bangli	-8.460312	115.353520

I also plot it on map using leaflet based on latitude and longitude value each regencies.



Second, I used foursquare API to get the venue around this regency. The detail that I captured for each venue includes Venue Name, Venue Latitude, Venue Longitude and Venue Category. Then I join the data with the current dataset.

	Regency	Regency Latitude	Regency Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Denpasar	-8.652497	115.219117	Dapoer Pemuda	-8.653717	115.217061	Indonesian Restaurant
1	Denpasar	-8.652497	115.219117	Kober Mie Setan Kaliasem	-8.655080	115.218596	Noodle House
2	Denpasar	-8.652497	115.219117	Voltvet Eatery & Coffee	-8.653594	115.217020	Café
3	Denpasar	-8.652497	115.219117	Warung Wardani	-8.651159	115.215957	Indonesian Restaurant
4	Denpasar	-8.652497	115.219117	Patung Catur Muka	-8.656036	115.216993	Monument / Landmark

In total there are 34 unique categories for all regencies and I did one hot encoding followed by aggregating by Regency to get mean values in each regency. This data then will be used

for modelling in order to get where is the recommended place for tourist to stay.

	Regency	Asian Restaurant	BBQ Joint	Bakery	Bike Shop	Breakfast Spot	Café	Chinese Restaurant	Coffee Shop	Convenience Store	...	Monument / Landmark	Night Market	Noodle House	Park	Pet Store	Pizza Place
0	Bangli	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Denpasar	0.035714	0.035714	0.035714	0.035714	0.00	0.035714	0.00	0.107143	0.035714	...	0.035714	0.000000	0.107143	0.000000	0.035714	0.035714
2	Gianyar	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.666667	...	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000
3	Klungkung	0.000000	0.000000	0.000000	0.000000	0.00	0.090909	0.00	0.000000	0.000000	...	0.000000	0.090909	0.000000	0.000000	0.000000	0.000000
4	Tabanan	0.000000	0.000000	0.250000	0.000000	0.25	0.000000	0.25	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

5 rows × 35 columns

3.2 Modelling

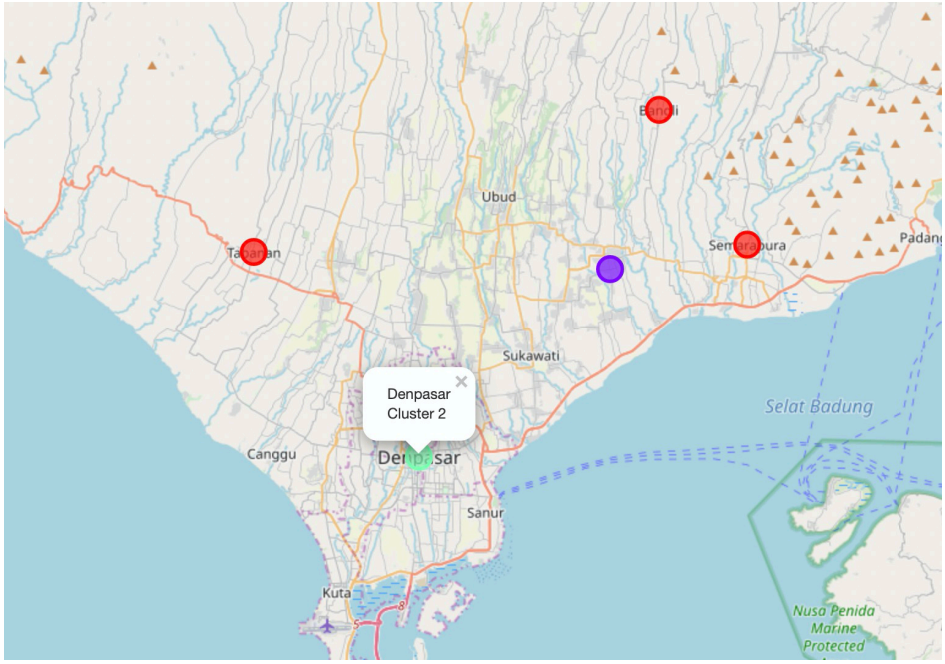
K-Means clustering was used for modelling using the data that already prepared. Because I want to find a natural group of each area in Bali. So we can suggest which area is better for food tourist. Then due to I only check 5 regency then I used k = 3 for the number of cluster that will be generated by the K-Means algorithm.

As a result we there are 3 regency that become cluster 0, then 1 regency for each cluster 1 and 2 respectively. Here's I also append the Most Common Venue for each Regencies. So we can describe the cluster more easily.

	Regency	Lat	Long	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Denpasar	-8.652497	115.219117	2	Indonesian Restaurant	Coffee Shop	Noodle House
1	Tabanan	-8.539231	115.126568	0	Bakery	Breakfast Spot	Chinese Restaurant
2	Gianyar	-8.548236	115.326054	1	Convenience Store	Park	Video Game Store
3	Klungkung	-8.535017	115.403276	0	History Museum	Historic Site	Night Market
4	Bangli	-8.460312	115.353520	0	Gym	Flea Market	Snack Place

4. ANALYSIS

Based on the clustering results, we can say that cluster 2 is the best place for tourist to stay in order to gain best experience on Indonesian food. In this case they should find an accommodation in Denpasar regency.



Deep down In Denpasar, tourist can find Indonesian Restaurant nearby like can be seen in table below.

	Regency	Regency Latitude	Regency Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Denpasar	-8.652497	115.219117	Dapoer Pemuda	-8.653717	115.217061	Indonesian Restaurant
3	Denpasar	-8.652497	115.219117	Warung Wardani	-8.651159	115.215957	Indonesian Restaurant
8	Denpasar	-8.652497	115.219117	Depot Karna Sari	-8.654098	115.216454	Indonesian Restaurant
25	Denpasar	-8.652497	115.219117	Warung Adnyana	-8.649157	115.219420	Indonesian Restaurant

5. CONCLUSION

Finally, based on the K-Means clustering method we can give recommendation to tourist that love food to stay in Denpasar (Cluster 2). This is due to Denpasar have more food venue compared to other Regency.

	Regency	Lat	Long	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Denpasar	-8.652497	115.219117	2	Indonesian Restaurant	Coffee Shop	Noodle House
1	Tabanan	-8.539231	115.126568	0	Bakery	Breakfast Spot	Chinese Restaurant
2	Gianyar	-8.548236	115.326054	1	Convenience Store	Park	Video Game Store
3	Klungkung	-8.535017	115.403276	0	History Museum	Historic Site	Night Market
4	Bangli	-8.460312	115.353520	0	Gym	Flea Market	Snack Place

6. FUTURE DIRECTION

Furthermore, this analysis can be expanded by finding the best accommodation that the tourist can stay in Denpasar. So they can find a place near a top restaurant arounds. Moreover if the tourist also want to visit other venue than food we also can give recommendation to them.