FINDING THE MOST SUITABLE NEIGHBORHOOD TO VISIT IN BALI, INDONESIA

Rafki Islamy

TABLE OF CONTENTS

I. Introduction: Business Problem

According to Bali Local Tourism Officer, about 16.83 million of tourists come to Bali in 2019. Along with that number, 6.28 million of them are foreign tourists who come from various countries in the world. The number of visitor has grown significantly and has the potential to grow bigger in the future, with exception of pandemic period that would make a correction.

The number of tourist to Bali, which is expected to increase, will also result in an increase in the variation of tourist preferences for the place or area they want to visit. This may happen because each group of tourist has different demographics that affect the choice a particular place or area. For example, the tourist from non-tropical countries have a tendency to visit the uptown area with a view of forest or ricefield, the Asian tourist is more likely visit the area which have the Asian cuisine in surroundings, or the local tourist that tend to visit the shopping area. In this current condition, sometimes tourists don't have a clue on which place are the most suitable for them to visit in Bali. Therefore, in some case they don't get the best holiday experience in Bali.

This final project explores several areas in Bali to identify the characteristics of each area based on the most common venues. The output of the analysis will be the recommendation for tourists or travel agent to help them seeking out the most suitable areas for them to visit, given the list of area/district in Bali. This analysis is also beneficial for investor by providing recommendation of areas/districts that might be suitable to open a specific business unit. So, this project will attempt to answer the questions "Where is the most suitable areas/neighborhood to visit in Bali for each tourist?" and "Where should the investor open a specific business unit?

II. The Location

https://en.wikipedia.org/wiki/Bali

Bali is a province of Indonesia and the westernmost of the Lesser Sunda Islands. East of Java and west of Lombok, the province includes the island of Bali and a few smaller neighbouring islands, notably Nusa Penida, Nusa Lembongan, and Nusa Ceningan. The provincial capital, Denpasar, is the most populous city in the Lesser Sunda Islands and the second-largest, after Makassar, in Eastern Indonesia. Bali is Indonesia's main tourist destination, with a significant rise in tourism since the 1980s. Tourism-related business makes up 80% of its economy.

III. Data

In order to answer the above question, there are several data related on Bali that are required, as shown below.

1. Bali data containing the city and neighborhood, latitudes and longitudes. This data is used as the basis of analysis and to generate the Foursquare API requests. This data is obtained from the following data source: https://raw.githubusercontent.com/ArrayAccess/Indonesia-Postal-And-Area/master/data/csv/62/subDistricts.csv
2. City size area in Bali. This data is used to generate radius value that is required while running the Foursquare API requests. This data is obtained from the following data source: https://en.wikipedia.org/wiki/List_of_districts_of_Bali
3. All venues in Bali. This data is obtained from Foursquare API utilized via the request library in Python.

IV. Methodology

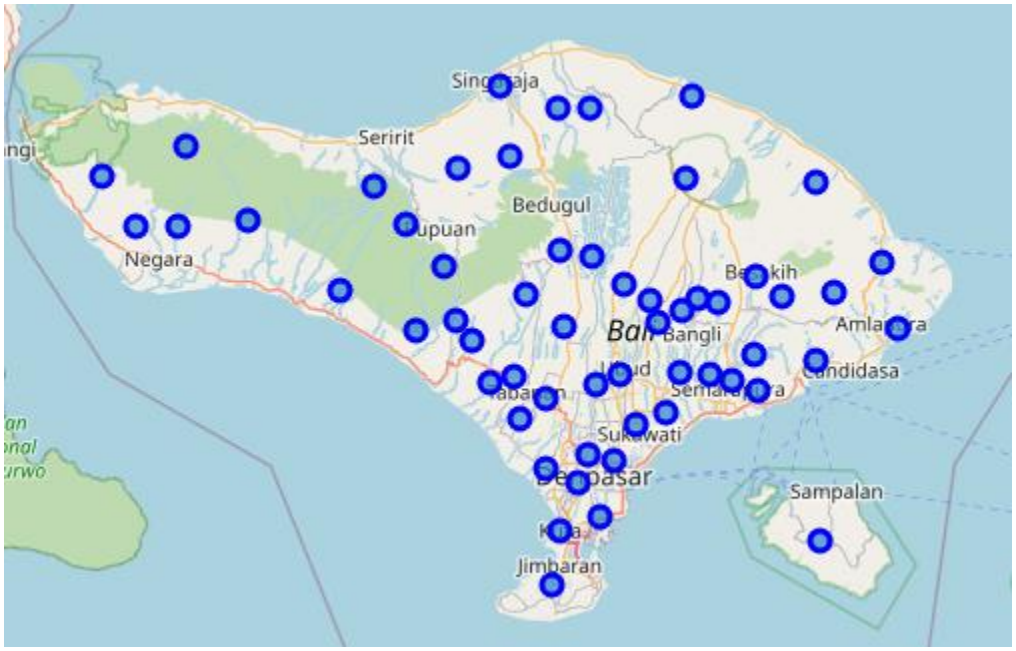
4.1 Data Preparation and Cleaning

The Bali data containing neighborhood and coordinates that has been described in previous chapter are being scraped and transformed into data frame using Pandas library. These data are still raw because there are several irrelevant data that are still in the data frame. Therefore, we drop the irrelevant data. Then, we make adjustment to the dataframe by renaming columns name and the city column values. The part result of data cleaning is shown below.

	Code	City	Neighborhood	Latitude	Longitude	Postal
0	5101010	JEMBRANA	MELAYA	-8.247269	114.542118	82252
1	5101020	JEMBRANA	NEGARA	-8.314523	114.589049	82251,82212,82218
2	5101021	JEMBRANA	JEMBRANA	-8.315036	114.646487	82211,82216,82217,82218,82219
3	5101030	JEMBRANA	MENDOYO	-8.307245	114.743959	82261
4	5101040	JEMBRANA	PEKUTATAN	-8.401172	114.868995	82262

4.2 Exploratory Data Analysis

Firstly, lets map the neighborhood in Bali using Geopy and Folium library.



Fetch all the venues for each neighborhood from Foursquare API.

```
bali_venues = getNearbyVenues(names=bali_neigh['Neighborhood'], cities=bali_neigh['City'],
                              latitudes=bali_neigh['Latitude'],
                              longitudes=bali_neigh['Longitude']
                             )
```

```
print(bali_venues.shape)
bali_venues.head()
```

```
(1459, 8)
```

	Neighborhood	City	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	MELAYA	JEMBRANA	-8.247269	114.542118	Bendungan Palasari	-8.254218	114.542237	Lake
1	MELAYA	JEMBRANA	-8.247269	114.542118	Taman Wana Villas & Spa	-8.257664	114.551894	Hotel
2	MELAYA	JEMBRANA	-8.247269	114.542118	GKPB Jemaat PNIEL Blimbingsari	-8.237667	114.518530	Church
3	MELAYA	JEMBRANA	-8.247269	114.542118	Bendungan Palasari	-8.294341	114.511383	River
4	NEGARA	JEMBRANA	-8.314523	114.589049	Pasar Senggol Terminal Negara	-8.359315	114.625844	Diner

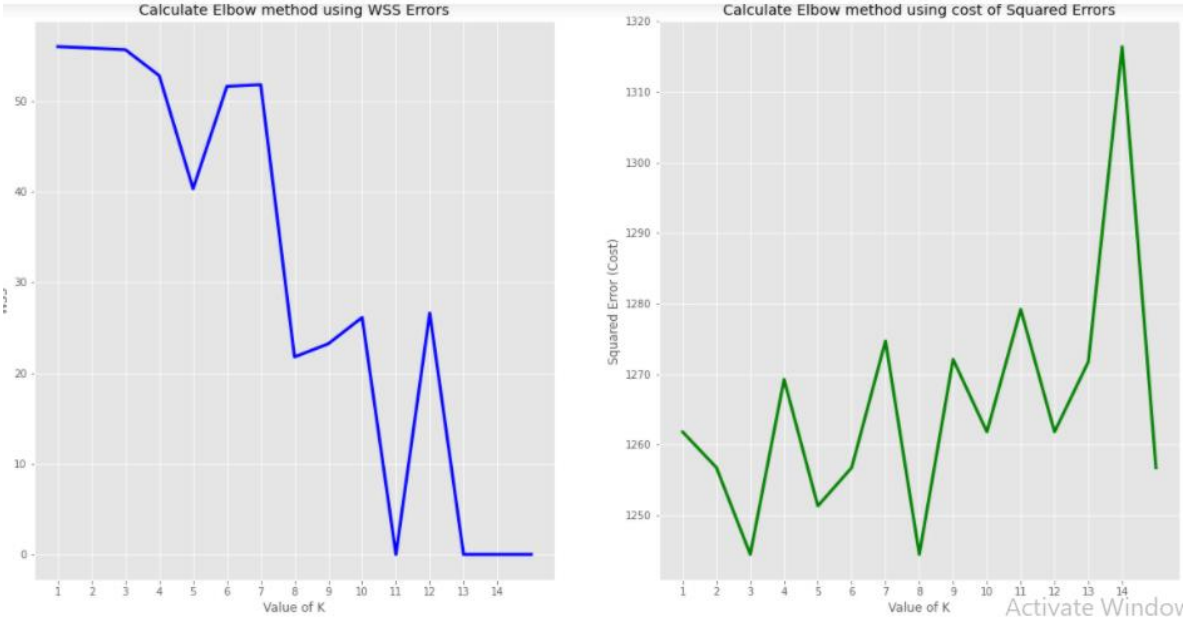
From above table, we got 1459 venues. After seeking the data more further, there are some of the neighborhood that have less number of venue. To prevent the biased analysis which is caused by the inadequacy of the data, we need to eliminate those neigborhood. In this case, we drop the neighborhood who have less than 18 venues.

Then, we need to display five most common venue for each neighborhood. The result is shown below.

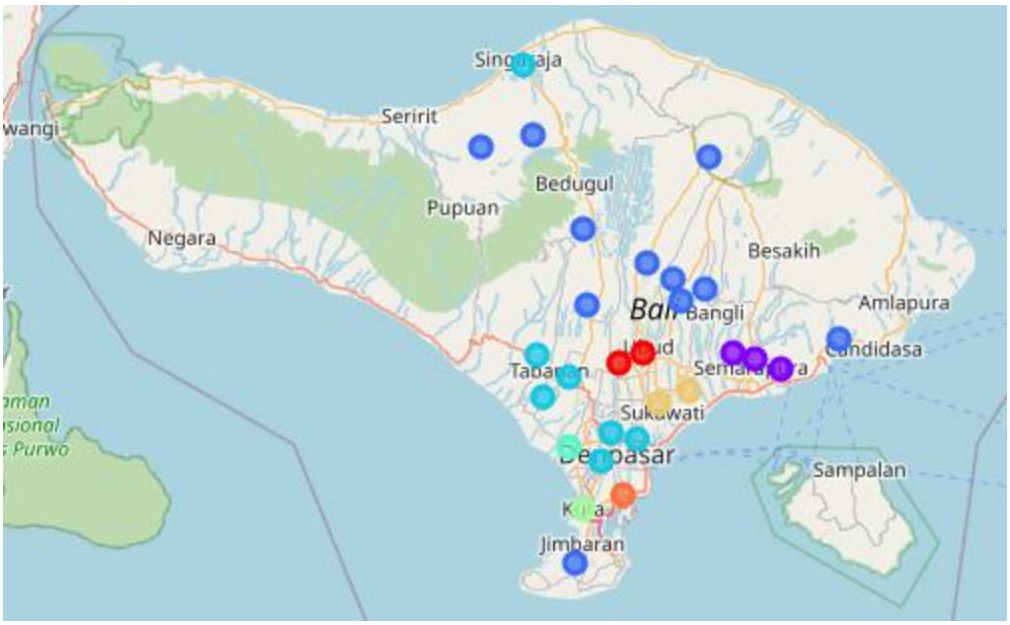
	City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	BADUNG	ABIANSEMAL	Resort	Vegetarian / Vegan Restaurant	Indonesian Restaurant	Asian Restaurant	Bistro
1	BADUNG	KUTA	Coffee Shop	Clothing Store	Hotel	Spa	Multiplex
2	BADUNG	KUTA SELATAN	Resort	Café	Indonesian Restaurant	Seafood Restaurant	Beach
3	BADUNG	KUTA UTARA	Café	Hotel	Asian Restaurant	Yoga Studio	Vegetarian / Vegan Restaurant
4	BADUNG	MENGWI	Indonesian Restaurant	Convenience Store	Food Court	Café	BBQ Joint

4.3 Problem Approach Using K-Means Clustering Analysis

The problem of finding the most suitable neighborhood in Bali can be solved using machine learning algorithm that is clustering analysis. This kind of analysis is used because we need to clearly group the neighborhood based on the most common venues for each neighborhood. Then the result will help the stakeholder in their decision making. K-Means is one of the most used and popular clustering algorithm. This algorithm can be used in various case, including this problem. This algorithm is required k-value that can be find through evaluation using elbow method for several possible k-value.



From the above evaluation, we can take the value of k=8. Then after running the k-means clustering algorithm, we can label neighborhood in Bali map based on their respective cluster, as shown below.



V. Results and Discussion

Based on the result of neighborhood clustering, we can analyze the main characteristics of each cluster by extracting what is the most common venue that are in each cluster compare with other clusters. Hence, below are the main characteristics of each cluster.

Cluster 1: Resort, Vegan Restaurant

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
BADUNG	ABIANSEMAL	Resort	Vegetarian / Vegan Restaurant	Indonesian Restaurant	Asian Restaurant	Bistro
GIANYAR	UBUD	Resort	Indonesian Restaurant	Vegetarian / Vegan Restaurant	Café	Asian Restaurant

Cluster 2: Historical Tourism (Site, Museum)

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
KLUNGKUNG	BANJARANGKAN	Historic Site	Indonesian Restaurant	History Museum	Miscellaneous Shop	Cultural Center
KLUNGKUNG	KLUNGKUNG	BBQ Joint	History Museum	Miscellaneous Shop	Historic Site	Outdoors & Recreation
KLUNGKUNG	DAWAN	Beach	Seafood Restaurant	Indonesian Restaurant	Harbor / Marina	History Museum

Cluster 3: Natural Tourism (Farm, Mountain, Beach, Waterfall)

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
TABANAN	MARGA	Indonesian Restaurant	Park	BBQ Joint	Satay Restaurant	Restaurant
TABANAN	BATURITI	Indonesian Restaurant	Restaurant	Asian Restaurant	Garden	Snack Place
BADUNG	KUTA SELATAN	Resort	Café	Indonesian Restaurant	Seafood Restaurant	Beach
GIANYAR	TAMPAKSIRING	Café	Indonesian Restaurant	Balinese Restaurant	Resort	Pedestrian Plaza
GIANYAR	TEGALLALANG	Café	Resort	Indonesian Restaurant	Balinese Restaurant	Coffee Shop
GIANYAR	PAYANGAN	Resort	Indonesian Restaurant	Coffee Shop	Paintball Field	Balinese Restaurant
BANGLI	SUSUT	Farm	Indonesian Restaurant	Garden	Coffee Shop	Chinese Restaurant
BANGLI	KINTAMANI	Indonesian Restaurant	Scenic Lookout	Mountain	Bed & Breakfast	Restaurant
KARANGASEM	MANGGIS	Resort	Hotel	Beach	Restaurant	Indonesian Restaurant
BULELENG	BANJAR	Hotel	Resort	Indonesian Restaurant	Balinese Restaurant	Coffee Shop
BULELENG	SUKASADA	Hotel	Scenic Lookout	Waterfall	Spa	Cajun / Creole Restaurant

Cluster 4: Asian Cuisine (Indonesian, Asian, Chinese Restaurant)

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
TABANAN	TABANAN	Indonesian Restaurant	Chinese Restaurant	Café	Farmers Market	Food Truck
TABANAN	KEDIRI	BBQ Joint	Café	Indonesian Restaurant	Convenience Store	Coffee Shop
BADUNG	MENGWI	Indonesian Restaurant	Convenience Store	Food Court	Café	BBQ Joint
BULELENG	BULELENG	Resort	Hotel	Beach	Indonesian Restaurant	Fast Food Restaurant
DENPASAR	DENPASAR TIMUR	Indonesian Restaurant	BBQ Joint	Asian Restaurant	Bakery	Chinese Restaurant
DENPASAR	DENPASAR BARAT	Indonesian Restaurant	Asian Restaurant	Coffee Shop	Food Truck	Hotel
DENPASAR	DENPASAR UTARA	Indonesian Restaurant	BBQ Joint	Seafood Restaurant	Asian Restaurant	Donut Shop

Cluster 5: Cafe, Hotel

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
BADUNG	KUTA UTARA	Café	Hotel	Asian Restaurant	Yoga Studio	Vegetarian / Vegan Restaurant

Cluster 6: Coffee Shop, Clothing Store

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
BADUNG	KUTA	Coffee Shop	Clothing Store	Hotel	Spa	Multiplex

Cluster 7: Art (Gallery, Theater, Craft store)

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
GIANYAR	SUKAWATI	Art Gallery	Arts & Crafts Store	Indonesian Restaurant	Restaurant	Coffee Shop
GIANYAR	BLAHBATUH	Theater	BBQ Joint	Bakery	Balinese Restaurant	Park

Cluster 8: Food Destination

City	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
DENPASAR	DENPASAR SELATAN	Seafood Restaurant	Indonesian Restaurant	BBQ Joint	Korean Restaurant	Noodle House

Based on those main characteristics, the tourist can be more easily make a choice about which neighborhood to visit based on their own preferences. For example, if the tourist is more likely to visit the historical areas, we recommend them to visit neighborhood that are in cluster 2. This recommendation process could also be used for investor in making decision about where they want to open their business based on its business unit.

VI. Conclusion

This analysis is performed on limited data. This may affect in the result accuracy that might be low. If good amount of data is available there is scope to come up with better results. But in overall, this analysis can helped the stakeholder in giving the recommendation about what location that suit their expectation. This analysis could be better by adding other variables such as venue ratings and tourism demography, as the other potential variable for analysis.