Predicting the Best Location to Build a Hotel in Bali

Lutfia Afifah

December, 2019

1. Introduction

1.1 Background

Bali is one of 34 provinces in Indonesia which is famous for its beautiful beaches. It is not surprising that Bali becomes the most visited destination in Indonesia by tourists, either domestic or foreign tourists. Its beautiful nature, especially the beaches, will attract anyone seeing it. Many tourists stay there for a long time to enjoy the nature. Even, there are some rich people intentionally purchasing a house in Bali to provide their accommodation when they and their families take a holiday in Bali. From the above description, it can be assumed that the most needed facilities in a tourist destination area is a place for staying. This project will explore some of the districts in Bali, specifically Southern Bali which have many tourist attraction areas, to predict the best location to build a new hotel to provide accommodation for tourists.

1.2 Business Problem

As a tourist destination, of course, there are already a lot of hotel in Bali. Therefore, this project will explore districts in south Bali and try to predict the optimal location to build a new hotel.

1.3 Target Audience

This project specifically will be targeted for stakeholders interested in opening a hotel in Southern Bali, Indonesia.

2. Data

2.1 Data Source and Description

Data which is be required for this exploration is a list of districts in Bali province, Indonesia. It can be found in Wikipedia from the following link: https://en.wikipedia.org/wiki/Kecamatans_of_Bali. Kecamatan is the Indonesian language means district. Wikipedia articles can be written by anyone, so it could be mistaken happened. Therefore, I use a PDF file from Indonesia's government website for data verification. Since I am an Indonesian native speaker, so it is easy for me to verify the data which comes from the local language. The file for verification can be found in the following link: https://www.kemendagri.go.id/files/2019-05/Kode&Data%20Wilayah/51.bali.fix.pdf.

3. Methodology

The stages of this project are started with web scraping, then preparing data, connecting to Foursquare, analyzing data, and clustering. For this project, I am going to explore the venues from the center of the districts as well as the beaches. So, it can be used by stakeholders who would like to build a hotel, either just in south Bali or specifically near a beach. The details of each stage will be discussed in the following.

3.1. Preparing Data

Data preparation is to set up the data will be used in exploration. The process is started with web scraping. The web page containing the name of district and regency in Bali to be scraped and moved into CSV file form. After the main data has created, the next step is to read the created CSV file into pandas dataframe and do some data transformations to beautify as well as make it easier to understand. In this stage, I also will generate the latitude and longitude coordinates using a Geocoder package in Python that will be needed to get the venues from Foursquare.

3.2. Connecting to Foursquare

This project needs to connect to Foursquare to acquire the venues will be used in exploring area in Bali. As I told above, I am going to make two exploration to predict the best location for a new hotel in any area in south Bali and in a specific area near a beach as Bali is popular with its beautiful beaches.

Since I have data of Bali districts, the first thing to do is to generate the venues within a particular radius from each center of the districts. These generated venues will be used to explore and predict the best district to build a new hotel. Then, for predicting the best area to build a hotel near a beach, the data from previous exploration will be used. The step to do is to filter the data which the category is Beach and then use them to generate new venues around the beaches.

3.3. Analyzing Data

Analyzing and exploring data in this project is also separated into two sections. First, to explore and analyze each district in south Bali and Nusa Penida island to see how many venues for each venue's categories there are and how many venues in which the category is hotel. So does for exploring and analyzing the venues around the beaches.

3.4. Clustering

The final stage is clustering the venues to get the view of how the districts or the beaches similar to one another based on the venue's categories they have. The clustering algorithm that I am going to use in this project is K-Means. K-Means is one of the simplest clustering algorithms which create k clusters. The first process of this algorithm is to randomly select k centroids to be used as the beginning points for each cluster. Next, it will perform iteratively calculation to optimize the position of the centroids.

For this project, I also will perform two clustering processes to cluster the districts as well as the beaches with different k because the number of data in each dataframe, which are district and beach dataframe, are also different.

4. Result

4.1. Preparing Data

First, I do web scraping to acquire a list of districts in Bali using BeautifulSoup package in Python and save it into a CSV file. Once CSV file have been created, the next step is to read that CSV file into pandas dataframe. Data displayed in the dataframe still does not in appropriate form. The name of districts indexed as column names, so I perform dataframe transposing to change the name of districts as rows and rename that single column as District.

The next step is to split the string of the columns in the dataframe. For example, the original form of string in the first index is *Abang, Karangasem*, which is *Abang* is a name of district and *Karangasem* is a name of regency. I split the district and regency name to make it clear which one is the districts and which one is the regencies. I put the regencies in a new column named Regency.

Next, data verification is performed to make sure that I use the proper data for exploring. I review the generated data from web scraping to be compared to the PDF file data from Indonesia's government website. As a result, I found some mismatches between the two data such as a mismatch between district and regency, a missed district, a should-not-be-there district, and some typo cases. Then, I clean and fix those problems and make sure that the dataset for exploring is already correct and resave it into a new CSV file contained 57 districts of Bali.

Since I am going to explore only the south part of Bali province, I shorten the dataframe into 9 rows containing the districts in the south Bali, included 8 districts in Bali island and 1 district in Nusa Penida island. For exploring the districts, I am going to use Foursquare location. Hence, the latitude and longitude coordinates are required for this exploration. To obtain the coordinates, I use Geocoder package in Python and put them into the two new columns named Latitude and Longitude. It causes some new problems. Not all of the generated-coordinates generated the proper values for each district. For example, for Kediri district in Tabanan regency, the coordinates refer to Kediri regency in East Java province, as you can see in the following figure.

```
[19]: geolocator = Nominatim(user_agent="my_explorer")

df['Latitude'] = df['District'].apply(geolocator.geocode, timeout=20).apply(lambda x: x.latitude)
df['Longitude'] = df['District'].apply(geolocator.geocode, timeout=20).apply(lambda x: x.longitude)
df
```

:[19]:

	District	Regency	Latitude	Longitude
0	Denpasar Barat	Denpasar	-8.662244	115.198067
1	Denpasar Selatan	Denpasar	-8.706134	115.225676
2	Denpasar Timur	Denpasar	-8.632914	115.246211
3	Denpasar Utara	Denpasar	-8.624513	115.209009
4	Kediri	Tabanan	-7.805898	112.168507
5	Kuta	Badung	-8.727835	115.173005
6	Kuta Selatan	Badung	-8.808849	115.169528
7	Kuta Utara	Badung	5.067268	97.041101
8	Nusa Penida	Klungkung	-8.745652	115.534745

Figure 1. Codes and outcome of generating Latitude and Longitude coordinates using Geolocator

Because of that, to fix this problem, in another cell in my notebook, I try to generate coordinates for Kediri with a clear address. So does for another district incorrectly produces coordinates, Kuta Utara. Then, replace the wrong values in the dataframe by the two new generated coordinates for each district.

```
[22]: df.loc[df['District'] == 'Kediri', ['Latitude','Longitude']] = [kediri_latitude, kediri_longitude]
      df.loc[df['District'] == 'Kuta Utara', ['Latitude','Longitude']] = [kuta_utara_latitude, kuta_utara_longitude]
      df
```

:[22]:

	District	Regency	Latitude	Longitude
0	Denpasar Barat	Denpasar	-8.662244	115.198067
1	Denpasar Selatan	Denpasar	-8.706134	115.225676
2	Denpasar Timur	Denpasar	-8.632914	115.246211
3	Denpasar Utara	Denpasar	-8.624513	115.209009
4	Kediri	Tabanan	-8.564455	115.132072
5	Kuta	Badung	-8.727835	115.173005
6	Kuta Selatan	Badung	-8.808849	115.169528
7	Kuta Utara	Badung	-8.678042	115.169861
8	Nusa Penida	Klungkung	-8.745652	115.534745

Figure 2. Codes and outcome of generating after fixing the Latitude and Longitude coordinates

Data preparation have been done. As you can see above, the coordinates of Kediri and Kuta Utara have changed. To make it easy for viewing the location of the districts, the map of Bali created using Folium package as you can see in the below figures.
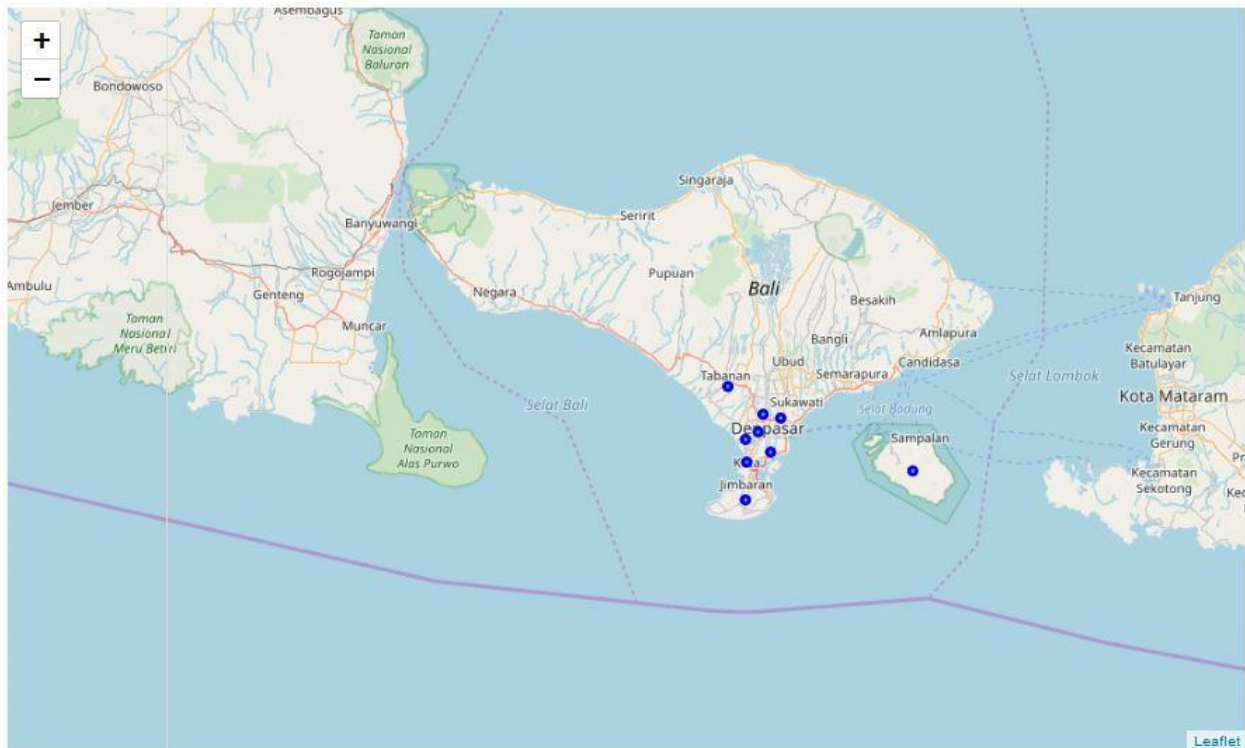
Figure 3. Generated map of Bali with mark of each district created using Folium

4.2. Connecting to Foursquare

To determine the best location to build a hotel, I am going to connect to Foursquare to obtain the venues as well as the detail information of them, such as the coordinates and categories, within a specific radius.

4.2.1. Generating Venues Around the Center of Districts

There are 9 districts in south Bali will be used for the exploration, 8 districts in Bali island and 1 district in Nusa Penida island. Nusa Penida district is larger than other districts because it covers all of the Nusa Penida island. Because of that, the radius for exploring Nusa Penida to be set 13 km from its center. For other districts in Bali island, the radius to be set 5 km each from its center.

The generated-venues from Foursquare for districts in Bali island have total number of venues 739 venues with 140 unique categories, while for district in Nusa Penida island has 73 venues with 25 unique categories. The below figures show a part of the results of generated-venues.

```
print(south_bali_venues.shape)
south_bali_venues
```
(739, 7)

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Denpasar Barat	-8.662244	115.198067	Bebek Goreng H.Slamet	-8.670825	115.191436	Indonesian Restaurant
1	Denpasar Barat	-8.662244	115.198067	Quest San Hotel Denpasar	-8.677209	115.190134	Hotel
2	Denpasar Barat	-8.662244	115.198067	Bakso Soto IMBO	-8.659310	115.209071	Breakfast Spot
3	Denpasar Barat	-8.662244	115.198067	Kopi Bali Cap Kupu-Kupu Bola Dunia	-8.655491	115.213394	Coffee Shop
4	Denpasar Barat	-8.662244	115.198067	Indoraya Grosir	-8.671957	115.208982	Mobile Phone Shop
5	Denpasar Barat	-8.662244	115.198067	Depot Umum "Batan Sabo"	-8.658079	115.214675	Chinese Restaurant
6	Denpasar Barat	-8.662244	115.198067	Balinese Therapy Massage	-8.666056	115.215917	Spa
7	Denpasar Barat	-8.662244	115.198067	Marugame Udon	-8.676021	115.207144	Noodle House
8	Denpasar Barat	-8.662244	115.198067	9/11 Cafe & Concept Store	-8.681070	115.188435	Bistro

Figure 4. Generated venues for each district in Bali island

```
print(nusa_penida_venues.shape)
nusa_penida_venues
```
(73, 4)

	Venue	Venue Category	Venue Latitude	Venue Longitude
0	Peguyangan Waterfalls	Waterfall	-8.781282	115.517186
1	Kelingking Beach	Beach	-8.750886	115.475774
2	Penida Colada	Beach Bar	-8.676435	115.514370
3	Crystal Bay	Beach	-8.715771	115.458927
4	Angel's Billabong	Scenic Lookout	-8.733337	115.448826
5	Broken Beach (Pasih Uug)	Scenic Lookout	-8.732653	115.450927
6	Secret Penida Cafe	Café	-8.677506	115.526102
7	Le Pirate Beach Club	Resort	-8.700540	115.443122
8	Warung Putu	Indonesian Restaurant	-8.693702	115.447575
9	Mangrove Beach	Beach	-8.666345	115.464161

Figure 5. Generated venues in Nusa Penida island

After the two dataframes have been generated, the next step is to concatenate those two dataframes into a single dataframe to make it easier for the next process. The result of concatenating dataframe generate 812 venues with 145 unique categories.

4.2.2. Generating Venues Near Beaches

From the generated-venues which define their names, coordinates, and categories, I filter the venues in which the category is Beach. Then, I will use it as the center of another exploration. There are 24 beaches generated from 9 districts. Here is the picture of the first 10 beaches in the new dataframe.

```
bali_beach = bali_venues[bali_venues['Venue Category'] == 'Beach']
print(bali_beach.shape)
bali_beach.head(10)
```
(24, 5)

	District	Venue	Venue Category	Venue Latitude	Venue Longitude
127	Denpasar Selatan	Mercure beach	Beach	-8.711357	115.254103
135	Denpasar Selatan	Pantai Mertasari	Beach	-8.712437	115.251110
459	Kuta	Discovery Beach	Beach	-8.727725	115.167094
470	Kuta	Boardwalk Kuta	Beach	-8.727078	115.167793
509	Kuta	Pantai Kuta	Beach	-8.718751	115.169037
579	Kuta Selatan	Pantai Muaya	Beach	-8.781143	115.163808
611	Kuta Selatan	Four Seasons Beach	Beach	-8.782794	115.161406
612	Kuta Selatan	Pantai Melasti	Beach	-8.848288	115.160215
616	Kuta Selatan	Tegal Wangi Beach (Pantai Tegal Wangi)	Beach	-8.782205	115.140860
618	Kuta Selatan	Pantai Pandawa	Beach	-8.845491	115.185822

Figure 6. Generated venues with Beach category

From 24 beaches generated from Foursquare, it turns out that not all of the beaches appear automatically in the dataframe. The most popular and visited beaches, such as Seminyak beach, Legian beach, Jimbaran beach, and other beautiful and popular beaches, did not generated automatically by Foursquare. This is a problem as I will optimize the exploration, especially around the popular beaches. So, I manually search the list of popular beaches, then insert the data and use the Geocoder package in Python to generate the coordinates. The number of added beaches is 11 beaches, so the total beaches in the dataframe now is 35. These 35 beaches will be used as the center in exploring venues around.

The next step is to generate the venues within a radius of 500 meters from each beach. There are 656 venues with 96 unique categories from all of 35 beaches. Here is the picture of the first 10 venues in the dataframe.

```
print(beach_venues.shape)
beach_venues.head(10)

(656, 7)
```

	Beach	Beach Latitude	Beach Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mercure beach	-8.711357	115.254103	Power Of Now Oasis	-8.710909	115.254747	Yoga Studio
1	Mercure beach	-8.711357	115.254103	Genius Cafe - Home to the Entrepreneur Beachclub	-8.711490	115.253829	Restaurant
2	Mercure beach	-8.711357	115.254103	Mercure beach	-8.711357	115.254103	Beach
3	Mercure beach	-8.711357	115.254103	Mercure Resort Sanur Bali	-8.709372	115.253974	Hotel
4	Mercure beach	-8.711357	115.254103	Prama Sanur Beach Hotel	-8.709834	115.256470	Hotel
5	Mercure beach	-8.711357	115.254103	Bamboo Bar	-8.710628	115.256778	Beach Bar
6	Mercure beach	-8.711357	115.254103	Pantai Mertasari	-8.712437	115.251110	Beach
7	Mercure beach	-8.711357	115.254103	Sudamala Suites & Villas	-8.707675	115.255988	Resort
8	Mercure beach	-8.711357	115.254103	RipCurl School Of Surf	-8.710887	115.256043	Surf Spot
9	Mercure beach	-8.711357	115.254103	Baby Monkey	-8.707431	115.256223	Balinese Restaurant

Figure 7. Generated venues around each beach

4.3. Analyzing Data

Analyzing and exploring data are performed to know how many venues in each category which are located there. This exploration also is separated into two sections, which are for around the center of districts and near beaches.

4.3.1. Exploring and Analyzing Venues Around the Center of Districts

The first exploration is to know what category of venues is the most common in south Bali. As we know, there are 812 venues all around the districts with 145 unique categories. The first step to do is to count the number of venues grouped by each category. Then, sorting in ascending is performed based on the number of venues. Not all of the categories to be displayed, only the top 15 categories displayed. The part of the codes can be seen in the following figure.

```
bali_venues_count = bali_venues.groupby(['Venue Category'])['Venue'].count().reset_index()
bali_venues_count = bali_venues_count.sort_values(by='Venue', ascending=True)
bali_venues_count.set_index('Venue Category', inplace=True)
bali_top15 = bali_venues_count['Venue'].tail(15)
bali_top15

Venue Category
Seafood Restaurant      13
Pizza Place             13
Italian Restaurant      13
Spa                     15
Chinese Restaurant      17
BBQ Joint               20
Restaurant              21
Bakery                  23
Beach                   24
Asian Restaurant        25
Café                    30
Coffee Shop             45
Resort                  47
Hotel                   55
Indonesian Restaurant   79
Name: Venue, dtype: int64
```

Figure 8. A part of codes to generate the top 15 venue's categories

As you can see above, the number of hotels generated from Foursquare in south Bali is 55. Then, from those 55 hotels in south Bali, let's specifically explore the number of hotels in each district. The first thing to do is filtering the dataframe which the category is hotel. The result of filtering, there are not any single hotel in two districts, Denpasar Utara and Kediri. So, I manually insert the 0 value for those districts in order that we can see all of the results. The next step is to count the number of the hotels grouped by the districts and perform ascending sort order based on it. The part of the codes can be seen in the following figure.

```
hotel_bali_count = hotel_bali.groupby('District')['Venue'].count().reset_index()
hotel_bali_count = hotel_bali_count.append([{'District': 'Denpasar Utara', 'Venue': 0},
                                            {'District': 'Kediri', 'Venue': 0}])

hotel_bali_count = hotel_bali_count.sort_values(by='Venue', ascending=True)
hotel_bali_count.set_index('District', inplace=True)
hotel_bali_count = hotel_bali_count['Venue'].tail(10)
hotel_bali_count
```

```
District
Denpasar Utara      0
Kediri              0
Denpasar Timur      1
Denpasar Barat      5
Kuta Selatan        5
Nusa Penida         5
Denpasar Selatan    9
Kuta Utara         12
Kuta               18
Name: Venue, dtype: int64
```

Figure 9. A part of codes to generate the number of hotels in each district

4.3.2. Exploring and Analyzing Venues Near Beaches

Same as the exploration data for each district, for exploration data near beaches, is also performed analyzing to know what the categories of venues are most near beaches in south Bali. The first step is also the same, which is to count the number of venues grouped by each category. Then, sorting in ascending is performed based on the number of venues. Not all of the categories to be displayed, only the top 15 categories displayed. The part of the codes can be seen in the following figure.

```
beach_venues_count = beach_venues.groupby(['Venue Category'])['Venue'].count().reset_index()
beach_venues_count = beach_venues_count.sort_values(by='Venue', ascending=True)
beach_venues_count.set_index('Venue Category', inplace=True)
beach_top15 = beach_venues_count['Venue'].tail(15)
beach_top15
```

```
Venue Category
Lounge                 10
Italian Restaurant     10
Balinese Restaurant    10
Ice Cream Shop         11
Spa                    12
Asian Restaurant       15
Coffee Shop            20
Café                   20
Beach Bar              20
Indonesian Restaurant  22
Restaurant             26
Seafood Restaurant     60
Beach                  60
Resort                 65
Hotel                  69
Name: Venue, dtype: int64
```

Figure 10. A part of codes to generate the top 15 venue's categories near beaches

As you can see above, the category of hotel is in the first place with 69 venues from all of the venues near beaches. Now, let's specify the exploration to know the number of hotels in each beach. The first step to do is the same as we did in the previous exploration, which is to filter the dataframe which the category is hotel. The result of filtering, there are not any single hotel in nine beaches, so I manually insert the 0 value for those beaches in order that we can see all of the results. The next step is to count the number of the hotels grouped by the beaches and perform ascending sort order based on it. The part of the codes can be seen in the following figure.

```python
hotel_beach_count = hotel_beach.groupby('Beach')['Venue'].count().reset_index()
hotel_beach_count = hotel_beach_count.append([{'Beach': 'Tegal Wangi Beach', 'Venue': 0},
                                              {'Beach': 'Pantai Pandawa', 'Venue': 0},
                                              {'Beach': 'Green Bowl Beach', 'Venue': 0},
                                              {'Beach': 'Kubu Beach', 'Venue': 0},
                                              {'Beach': 'Kelingking Beach', 'Venue': 0},
                                              {'Beach': 'Crystal Bay', 'Venue': 0},
                                              {'Beach': 'Mangrove Beach', 'Venue': 0},
                                              {'Beach': 'Atuh Beach', 'Venue': 0},
                                              {'Beach': 'Diamond Beach', 'Venue': 0}])

hotel_beach_count = hotel_beach_count.sort_values(by='Venue', ascending=True)
hotel_beach_count.set_index('Beach', inplace=True)
hotel_beach_count = hotel_beach_count['Venue'].tail(50)
hotel_beach_count
```

```
Beach
Diamond Beach          0
Mangrove Beach         0
Crystal Bay            0
Kelingking Beach       0
Kubu Beach             0
Green Bowl Beach       0
Pantai Pandawa         0
Tegal Wangi Beach      0
Atuh Beach             0
Pantai Melasti         1
Balangan Beach         1
Pantai Mertasari       1
Four Seasons Beach     1
Pantai Jungutbatu      2
Bingin Beach           2
Dream Beach            2
Jimbaran Beach         2
Pantai Muaya           2
Pemedal Beach          2
Sanur Beach            2
Secret Point Beach     2
Segara Beach           2
Dreamland Beach        2
Tanjung Benoa Beach    2
Nusa Dua Beach         3
Pantai Kuta            3
Mercure beach          3
Canggu Beach           4
Discovery Beach        6
Boardwalk Kuta         6
Legian Beach           9
Seminyak Beach         9
Name: Venue, dtype: int64
```

Figure 11. A part of codes to generate the number of hotels in each beach

4.4. Clustering

Clustering data is to group similar objects into a set known as cluster. This process is used to get some cluster contain similarity of venues they have. For this project, the K-Means algorithm is used as a method to get some clusters.

4.4.1. Clustering Districts in South Bali

As we would like to know the similarity of districts, we have to know what venues the most are in each district. The first step to do is to encode the venue categories using one-hot encoding method. It turns out value 0 or 1 in venue categories for each venue. For example, for a hotel, it turns out value 1 in column Hotel, while it generates value 0 in other columns.

Once we get the encoding dataframe, the next step is to calculate the average or mean for each venue grouped by the district. Then, we generate the top 5 venues having the highest average. The higher average score, the most the venues are in a district. To be able to see clearly about what the common venues in each district, we create a new dataframe of the 1st to 10th most common venues for each district.

The most common venues for each district are created. The next step is to cluster the district into 3 clusters using the K-Means algorithm. In python, clustering data using the K-Means algorithm can be automatically generated by using a machine learning library, Scikit-learn. The result of each cluster can be seen in the following figures.

	District	Regency	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Denpasar Selatan	Denpasar	0	Hotel	Indonesian Restaurant	Café	Resort	Coffee Shop	Restaurant	Bakery	Asian Restaurant	Dive Shop	Korean Restaurant
5	Kuta	Badung	0	Hotel	Coffee Shop	Clothing Store	Resort	Restaurant	Spa	American Restaurant	Multiplex	Indonesian Restaurant	BBQ Joint
7	Kuta Utara	Badung	0	Hotel	Resort	Café	Asian Restaurant	Restaurant	Italian Restaurant	Coffee Shop	Indonesian Restaurant	French Restaurant	Ice Cream Shop

Figure 12. The result of the first district cluster, cluster label 0

As we can see in Figure 13, there are three districts in cluster label 0, Kuta and Kuta Utara in Badung regency as well as Denpasar Selatan in Denpasar. The 1st most common venue in cluster label 0 is hotel.

	District	Regency	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Kuta Selatan	Badung	1	Resort	Beach	Café	Indonesian Restaurant	Hotel	Italian Restaurant	Coffee Shop	Asian Restaurant	Seafood Restaurant	Beach Bar
8	Nusa Penida	Klungkung	1	Beach	Resort	Indonesian Restaurant	Hotel	Scenic Lookout	Dive Shop	Café	Beach Bar	Coffee Shop	Bed & Breakfast

Figure 13. The result of the second district cluster, cluster label 1

There are two districts in cluster label 1, Kuta Selatan in Badung and Nusa Penida in Klungkung. The 1st and 2nd most common venues are resort, and beach.

	District	Regency	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Denpasar Barat	Denpasar	2	Indonesian Restaurant	Coffee Shop	Hotel	Asian Restaurant	Chinese Restaurant	Bakery	Pizza Place	Javanese Restaurant	Restaurant	Ice Cream Shop
2	Denpasar Timur	Denpasar	2	Indonesian Restaurant	Coffee Shop	Bakery	BBQ Joint	Asian Restaurant	Art Gallery	Chinese Restaurant	Pizza Place	Snack Place	Music Venue
3	Denpasar Utara	Denpasar	2	Indonesian Restaurant	Pizza Place	Coffee Shop	BBQ Joint	Bakery	Chinese Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Asian Restaurant	Kids Store
4	Kediri	Tabanan	2	Food Court	Indonesian Restaurant	Café	BBQ Joint	Chinese Restaurant	Convenience Store	Boutique	Shopping Mall	Museum	Department Store

Figure 14. The result of the third district cluster, cluster label 2

In Figure 15, there are 3 districts in cluster label 2, which are Denpasar Barat, Denpasar Timur, and Denpasar Utara in Denpasar, as well as Kediri in Tabanan. The 1st and 2nd most common venues are Indonesian restaurant, coffee shop, pizza place, and food court.

4.4.2. Clustering Beaches in Southern Bali

Same as what we have done above, the first step to for clustering beaches in southern Bali is to encode the venue categories using one-hot encoding method. Once the one-hot encoding dataframe is ready, the next step is to calculate the average or mean for each venue grouped by the beach. Then, we generate the top 5 venues having the highest average scores. Also, we save it into a new dataframe containing the 1st common venues to 10th common venues for each beach.

The most common venues for each beach have created. As we did before, the next step is to cluster the beaches. For the beach clustering, the number of clusters is 5 as we have more than 30 data in the dataframe. We still use the same process as we did above, this clustering also use K-Means clustering algorithm which can automatically generate the result using Scikit-learn library in Python. The result of each cluster can be seen in the following figure.

	Beach	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Mercure beach	0	Restaurant	Hotel	Surf Spot	Indonesian Restaurant	Italian Restaurant	Coffee Shop	Resort	Café	Yoga Studio	Bar
1	Pantai Mertasari	0	Yoga Studio	Restaurant	Hotel	Resort	Coffee Shop	General Entertainment	Dive Spot	Hawaiian Restaurant	Gym	Convenience Store
2	Discovery Beach	0	Hotel	Ice Cream Shop	Seafood Restaurant	Resort	Mediterranean Restaurant	Café	Shopping Mall	Japanese Restaurant	Mexican Restaurant	Lounge
3	Boardwalk Kuta	0	Hotel	Restaurant	Japanese Restaurant	Ice Cream Shop	Seafood Restaurant	Cocktail Bar	Mexican Restaurant	Mediterranean Restaurant	Massage Studio	Coffee Shop
4	Pantai Kuta	0	Coffee Shop	Hotel	Clothing Store	Restaurant	Café	Resort	Balinese Restaurant	Italian Restaurant	Surf Spot	Convenience Store
7	Pantai Melasti	0	Resort	Asian Restaurant	Hotel	Diner	New American Restaurant	Café	Yoga Studio	Fast Food Restaurant	Dessert Shop	Dive Shop
9	Pantai Pandawa	0	Café	Resort	Vacation Rental	Beach Bar	Yoga Studio	Fish & Chips Shop	Cosmetics Shop	Dessert Shop	Diner	Dive Shop
10	Greenbowl Beach	0	Yoga Studio	Flea Market	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Dive Spot	Donut Shop	Eastern European Restaurant	Fast Food Restaurant
13	Kelingking Beach	0	Balinese Restaurant	Scenic Lookout	Comfort Food Restaurant	Yoga Studio	Fish & Chips Shop	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Dive Spot
16	Atuh Beach	0	Yoga Studio	Flea Market	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Dive Spot	Donut Shop	Eastern European Restaurant	Fast Food Restaurant
17	Dream Beach	0	Hotel	Residential Building (Apartment / Condo)	Restaurant	Indonesian Restaurant	Bed & Breakfast	Fish & Chips Shop	Cosmetics Shop	Dessert Shop	Diner	Dive Shop
18	Pantai Jungutbatu	0	Resort	Dive Shop	Hotel	Yoga Studio	Asian Restaurant	Boarding House	Café	Coffee Shop	Eastern European Restaurant	Indonesian Restaurant
19	Diamond Beach	0	Scenic Lookout	Yoga Studio	Comfort Food Restaurant	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Dive Spot	Donut Shop	Eastern European Restaurant
20	Secret Point Beach	0	Scenic Lookout	Hotel	Hotel Pool	Resort	Indonesian Restaurant	Fast Food Restaurant	Cosmetics Shop	Dessert Shop	Diner	Dive Shop
21	Pemedal Beach	0	Resort	Hotel	Indonesian Restaurant	Spa	Lounge	Dive Spot	Juice Bar	Restaurant	Dive Shop	American Restaurant
23	Segara Beach	0	Resort	Dive Shop	Boat or Ferry	Hotel	Yoga Studio	Boarding House	Café	Coffee Shop	Eastern European Restaurant	Indonesian Restaurant
24	Sanur Beach	0	Convenience Store	Asian Restaurant	Indonesian Restaurant	Café	Seafood Restaurant	Hotel	Coffee Shop	Chinese Restaurant	Restaurant	Pub
25	Seminyak Beach	0	Hotel	Resort	Restaurant	Indonesian Restaurant	Spa	Cosmetics Shop	Asian Restaurant	Ice Cream Shop	Italian Restaurant	Bar
26	Legian Beach	0	Hotel	Resort	Café	Surf Spot	Restaurant	Hotel Pool	Spa	Food Truck	Indonesian Meatball Place	Indonesian Restaurant
30	Nusa Dua Beach	0	Hotel	Beach Bar	Resort	Cocktail Bar	Hotel Pool	Balinese Restaurant	Asian Restaurant	Indonesian Restaurant	Steakhouse	Seafood Restaurant
31	Tanjung Benoa Beach	0	Resort	Indonesian Restaurant	Water Park	Hotel	Bar	Coffee Shop	Breakfast Spot	Snack Place	Spa	Restaurant
32	Canggu Beach	0	Hotel	Yoga Studio	Restaurant	Balinese Restaurant	Cocktail Bar	Surf Spot	Bed & Breakfast	Beer Garden	Breakfast Spot	Café
33	Dreamland Beach	0	Hotel	Spanish Restaurant	Resort	Indonesian Restaurant	Café	Mediterranean Restaurant	Lounge	Fast Food Restaurant	Dessert Shop	Diner
34	Bingin Beach	0	Hotel	Resort	Juice Bar	Rest Area	Restaurant	Building	Fast Food Restaurant	Cosmetics Shop	Dessert Shop	Diner

Figure 15. The result of the first beach cluster, cluster label 0

There are 24 beaches in cluster label 0, Mercure Beach, Pantai Mertasari, Discovery Beach, Broadwalk Kuta, Pantai Kuta, Pantai Melasti, Pantai Pandawa, Greenbowl Beach, Kelingking Beach, Atuh

Beach, Dream Beach, Pantai Jungutbatu, Diamond Beach, Secret Point Beach, Pemedal Beach, Segara Beach, Sanur Beach, Seminyak Beach, Legian Beach, Nusa Dua Beach, Tanjung Benoa Beach, Canggu Beach, Dreamland Beach, and Bingin Beach. Hotel is the most common beach in this cluster. Most of the beaches have hotel in their 1st most common venues.

	Beach	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Pantai Muaya	1	Seafood Restaurant	Resort	Pool	Lounge	French Restaurant	Hotel	Spa	Beach Bar	Coffee Shop	Pizza Place
6	Four Seasons Beach	1	Seafood Restaurant	Resort	Asian Restaurant	French Restaurant	Spa	Ice Cream Shop	Beach Bar	Lounge	Pizza Place	Pedestrian Plaza
27	Jimbaran Beach	1	Seafood Restaurant	Resort	Pool	Beach Bar	Coffee Shop	Lounge	Spa	Hotel	French Restaurant	Food Truck

Figure 16. The result of the second beach cluster, cluster label 1

In figure 16, there are 3 beaches which come in to cluster label 1, which are Pantai Muaya, Four Seasons Beach, and Jimbaran Beach. The 1st most common venues is seafood restaurant, while the 2nd is resort.

	Beach	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Kubu Beach	2	Resort	Beach Bar	Breakfast Spot	Yoga Studio	Flea Market	Dessert Shop	Diner	Dive Shop	Dive Spot	Donut Shop
22	Broken Beach	2	Beach Bar	Yoga Studio	Flea Market	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Dive Spot	Donut Shop	Eastern European Restaurant
28	Blue Point Beach	2	Resort	Beach Bar	Restaurant	General Entertainment	Indonesian Restaurant	Convenience Store	Breakfast Spot	Yoga Studio	Fast Food Restaurant	Dessert Shop

Figure 17. The result of the third beach cluster, cluster label 2

In the figure above, we can see that there are three beaches in cluster label 2, which are Kubu Beach, Broken Beach, and Blue Point Beach. The most common venues in that beach are beach bar and resort.

	Beach	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	Crystal Bay	3	Dive Spot	Yoga Studio	Flea Market	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Donut Shop	Eastern European Restaurant	Fast Food Restaurant
15	Mangrove Beach	3	Balinese Restaurant	Dive Spot	Forest	Yoga Studio	Flea Market	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Donut Shop

Figure 18. The result of the fourth beach cluster, cluster label 3

For cluster label 3, there are two beaches in this cluster, which are Crystal Bay and Mangrove Beach. The most common venue in those two beaches is dive spot, in which it is the 1st most common venues for Crystal Bay and the 2nd most common venues for Mangrove Beach.

	Beach	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Tegal Wangi Beach (Pantai Tegal Wangi)	4	Resort	Pool	Cocktail Bar	Breakfast Spot	Dessert Shop	Tea Room	Italian Restaurant	Convenience Store	Cosmetics Shop	Diner
12	Karma Private Beach	4	Resort	Indonesian Restaurant	Beach Bar	Spa	Mediterranean Restaurant	Yoga Studio	Fast Food Restaurant	Cosmetics Shop	Dessert Shop	Diner
29	Balangan Beach	4	Resort	Hotel	Seafood Restaurant	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Dessert Shop	Diner	Dive Shop	Dive Spot

Figure 19. The result of the fifth beach cluster, cluster label 4

Lastly, for the cluster label 4, there are three beaches in this cluster, which are Tegal Wangi Beach, Karma Private Beach, and Balangan Beach. The 1st most common venue for these three beaches is resort.

5. Discussion

In the exploring and analyzing stage of the districts in south Bali, Indonesian restaurant stands in the first place of the most common venues with having total number of 79 venues for all around the 9 districts. The second place is the hotel with 55 venues followed by resort, coffee shop, and café in the third, fourth, and fifth in order. The chart of the top 15 venues in all of 9 districts can be seen in the following figure.
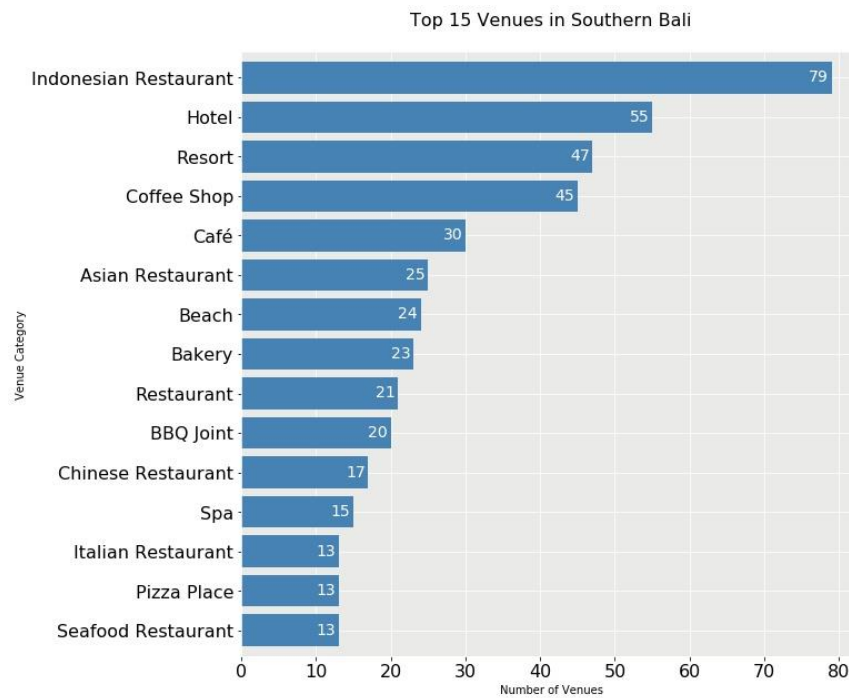

Figure 20. Top 15 venues in southern Bali

In general, there are quite a number of hotels in south Bali. Now, let's take a look at the chart of the number of hotels for each district.
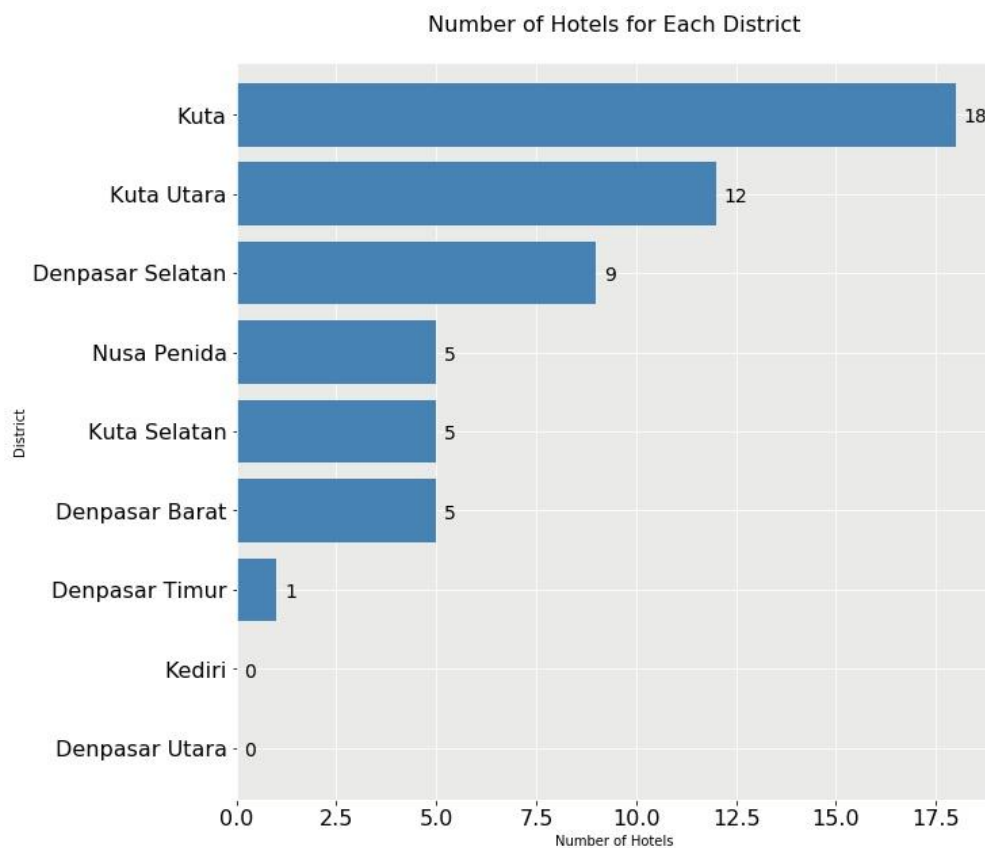
Figure 21. Number of hotels for each district

Kuta has the most number of hotels with 18 hotels, while Kuta Utara follows in the second most number of hotels with 12 hotels around the district. As we see above, there are two districts which do not have any hotel, which are Kediri and Denpasar Utara. Now, let's see the map of the district clustering.
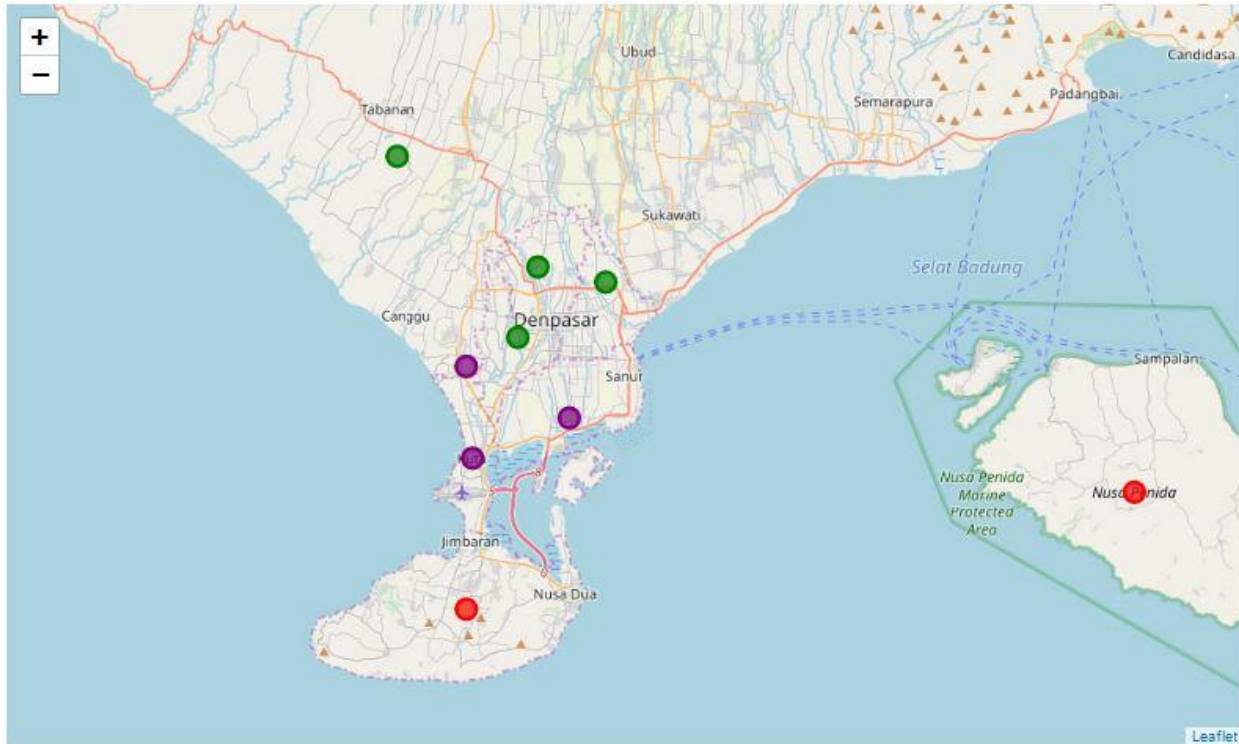
Figure 22. The map of district clustering in southern Bali

From the map above describing each cluster, the purple circles define the cluster label 0, the red circles define the cluster label 1, and the green circles define the cluster label 2. As we saw in the result section, the cluster label 0, which is the district with the purple circles, have much more hotel than two other cluster labels. The districts are in it are Denpasar Selatan, Kuta, and Kuta Utara. Based on the chart shown in Figure 21, those three districts are in the top 3 based on the number of hotels.

The two other clusters, the red circles and the green circles, do not have many hotels. Both Denpasar Utara and Kediri, which do not have any hotel at all, become parts of cluster label 2, the green circles, as those two districts have many restaurants as the 1st most common venues like the other member in that cluster, which are Denpasar Barat and Denpasar Timur. For the cluster label 1, the red circles, the 1st and 2nd most common venue in those two districts is resort and beach. As we know, the resort is somewhat similar to the hotel but it has a little bit of difference in facilities and services.

From those results, I can assume that the best location for a new hotel in south Bali if no more criteria have to be required is in districts which are parts of cluster label 2, especially Kediri district in Tabanan regency or Denpasar Utara in Denpasar which has no any single hotel. Denpasar Timur also can be one of the candidates as it has only one hotel.

Now, let's assume that the new hotel must near a beach as Bali is popular with its beautiful beach so that it likely will get more visitors than other places. Regarding the exploration we had before, let's take a look at the chart we have made.
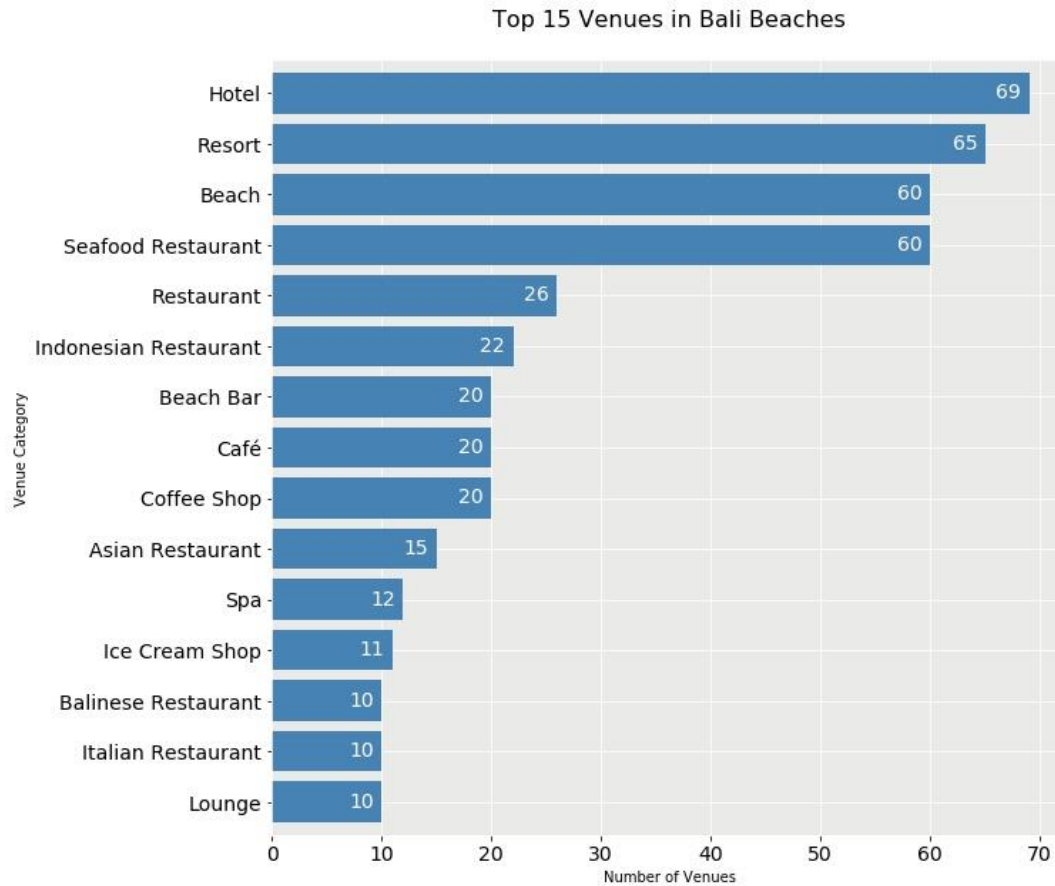
Figure 23. Top 15 venues in Bali beaches

As we can see above, hotel is in the first place of the most common venues with 69 venues. It can be said that the hotel is the most important venue near a beach as the tourists who come to Bali definitely visit the beaches and they prefer to stay near a beach. Now, let's see which beach having the most hotels is.
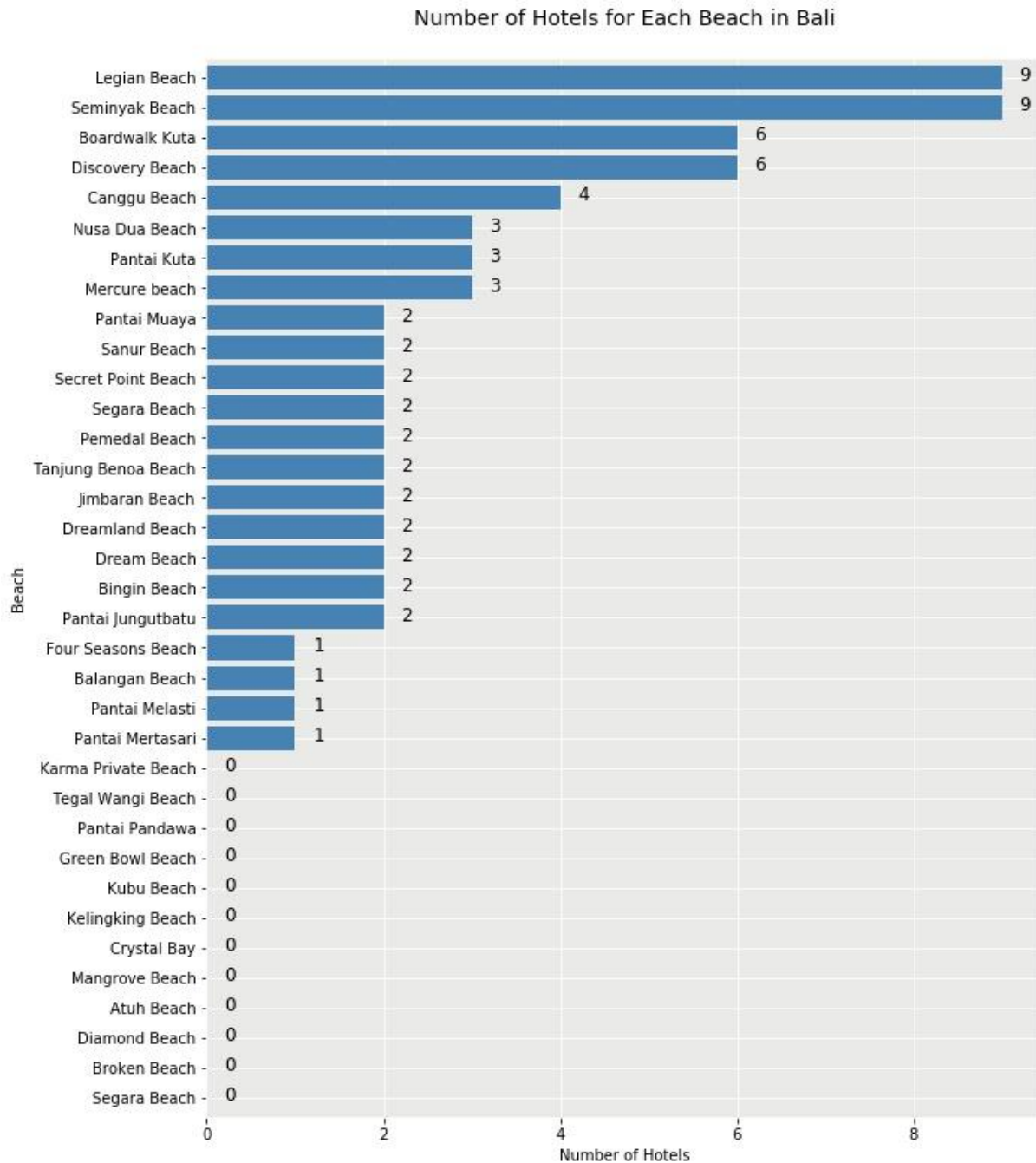
Figure 24. Number of hotels for each beach in Bali

From 69 hotels, the most are in Legian Beach and Seminyak Beach. It is not surprising as those two beaches are the most popular in Bali. It turns out that there are still some beaches which do not have any hotel at all, such as Atuh Beach, Tegal Wangi Beach, Pantai Pandawa, Greenbowl Beach, Kubu Beach, Kelingking Beach, Crystal Bay, Mangrove Beach, Diamond Beach, Karma Private Beach, Broken Beach, and Segara Beach. Now, let's see the map of the beach clustering.
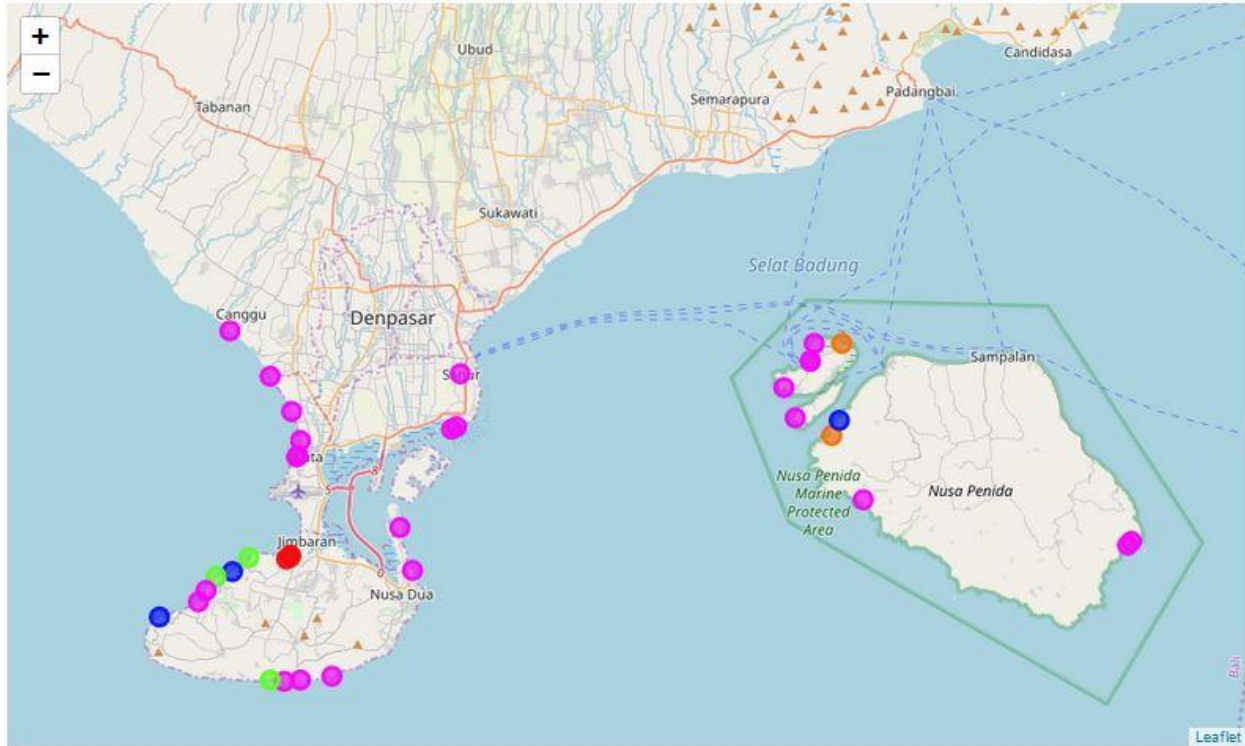
Figure 25. The map of beach clustering in southern Bali

From the map above describing each beach, the pink circles define the cluster label 0, the red circles define the cluster label 1, the blue circles define the cluster label 2, the orange circles define the cluster label 3, and the green circles define the cluster label 4. As we saw in the result section, the cluster label 0, have more hotels than four other cluster labels. Legian Beach, Seminyak Beach, Broadwalk Kuta, Discovery Beach, Canggu Beach, Tanjung Benoa Beach, Pantai Kuta, and Nusa Dua Beach, which are in the top 8 beach in number of hotels, are members of that cluster.

Cluster labels 1, 2, and 4, the red, blue, and green circles in order, have fewer hotels than cluster label 0, but they have many resort which is similar to hotel with somewhat different facilities and services. Meanwhile, as we saw in Figure 18 in the Result section, the beaches in cluster label 3 do not have any hotel or resort in their 1st to 10th most common venues.

From these results, I assume that the best location for a new hotel near a beach in south Bali is in the beaches of the cluster label 3 members, which are Crystal Bay and Mangrove Beach. The other choice to build a new hotel near beaches are on the beaches which do not have any hotel yet or the beaches which have only one hotel. You can see the list of those beaches in Figure 24 above.

6. Conclusion

Purpose of this project is to identify the areas in Bali with low number of hotels in order to help stakeholders in searching for optimal location to build a new hotel in Bali. After exploring the venues in south Bali using Foursquare data, we can conclude some of the points in the following.

- The best location for a new hotel in south Bali if no more criteria have to be required is in districts which are parts of cluster label 2, especially Kediri district in Tabanan regency or Denpasar Utara in Denpasar which has no any single hotel.
- The best location for a new hotel near a beach in south Bali is in the beaches of the cluster label 3 members, which are Crystal Bay and Mangrove Beach.
- The other choices are on the beaches which do not have any hotel yet or the beaches which have only one hotel.

The final decision will be made by stakeholders by considering all aspects required. This result depends on the time you access. This project used Foursquare, where the data will be probably changed because of the additional venues or other possibilities.