The Battle of Neighborhoods- Bali Indonesia

1. **Introduction where you discuss the business problem and who would be interested in this project:**

I am going to be exploring the island of Bali in Indonesia as a travel destination first and then looking for areas where hotel investments would be sound. I will be looking at the different cities and utilizing foursquare to get more insights on services available: hotels, cafes, restaurants.. Based on the result of the analysis, I will look at which area would be better for a new hotel based on the need, potentially the population and number of tourists visiting over the last years. The target for this analysis would be a potential investor in the hospitality industry looking to tap into Bali as a travel destination. So the basic question is: should the investor invest in Bali? Why? And If yes, potentially where?

2. **Data where you describe the data that will be used to solve the problem and the source of the data:**

To solve this problem, I will be using the following data:

a. List of Bali cities and provinces with their latitudes sand longitudes (a combination of data scrapped from the web and a data file used as asset)
b. Foursquare data to retrieve venues around Bali: hotels, restaurants, cafes..
c. Bali population and visitor data from the web and potentially also from a csv file
d. Comparison data from another location

3. **Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.**

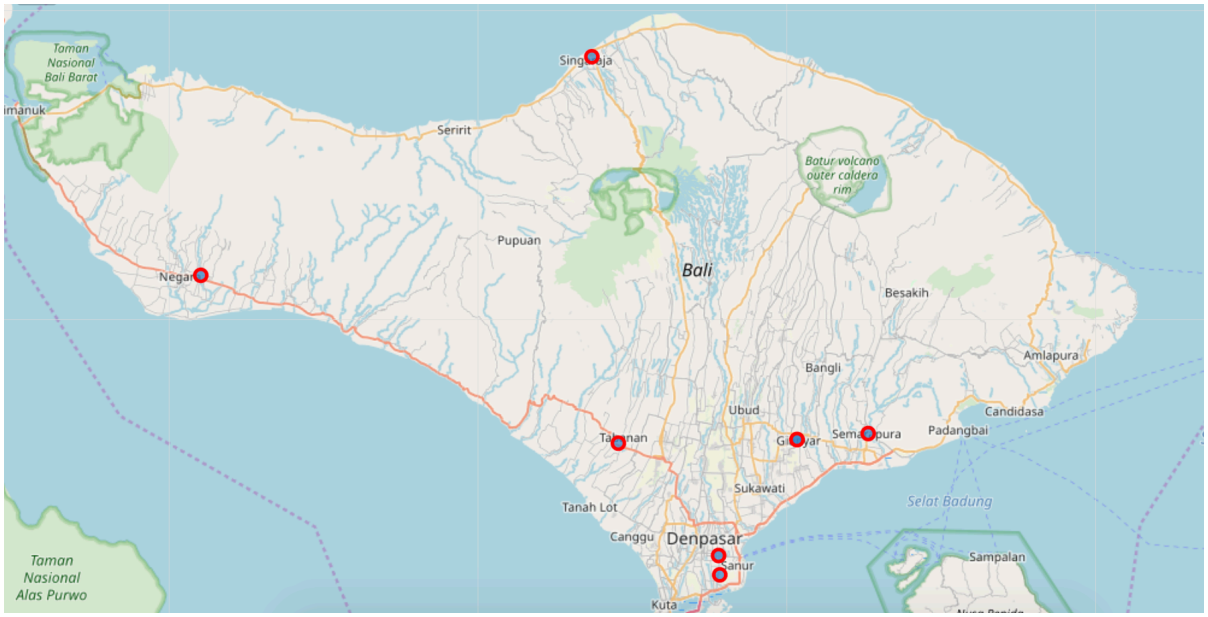
The data preparation step consisted of two main actions:

a. Scrap the list of Bali cities from a webpage and saving that into a dataframe
b. Extract the list of cities and their coordinates from a data asset (csv) file

The two data sets were then merged to obtain one dataframe with the list of Bali regions, cities and their coordinates.

	Name	Latitude	Longitude
0	Denpasar City	-8.672180	115.233550
1	Badung Regency	-8.694210	115.235809
2	Bangli Regency	-7.589500	112.790138
3	Buleleng Regency	-8.110320	115.089394
4	Gianyar Regency	-8.541600	115.323051
5	Jembrana Regency	-8.357130	114.645592
6	Karangasem Regency	-6.702870	108.482811
7	Klungkung Regency	-8.534860	115.404213
8	Tabanan Regency	-8.545040	115.119957
9	Ubud	-8.506854	115.262482

The next step was to visualize the Bali cities on a Folium map after using the geolocator to get the coordinates of Bali.



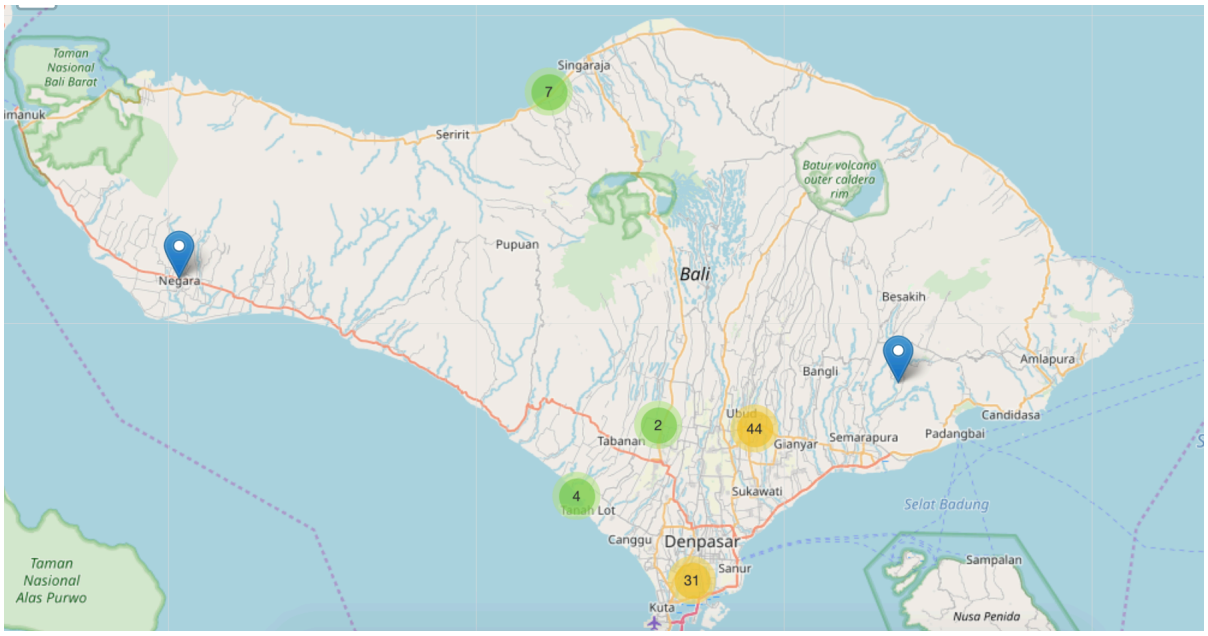
Foursquare was then used to extract the list of venues for every city in Bali as listed in the dataframe prepared. The radius was set to 10000 to get the maximum number of venues due to the size of the whole island. A total of 717 venues was obtained mapped to 137 categories.

Capital	Capital Latitude	Capital Longitude	Venue	Venue id	Venue Latitude	Venue Longitude	Venue Category
Amlapura	100	100	100	100	100	100	100
Bangli	19	19	19	19	19	19	19
Denpasar	100	100	100	100	100	100	100
Gianyar	100	100	100	100	100	100	100
Mangupura	100	100	100	100	100	100	100
Negara	12	12	12	12	12	12	12
Semarapura	68	68	68	68	68	68	68
Singaraja	48	48	48	48	48	48	48
Tabanan	70	70	70	70	70	70	70
Ubud	100	100	100	100	100	100	100

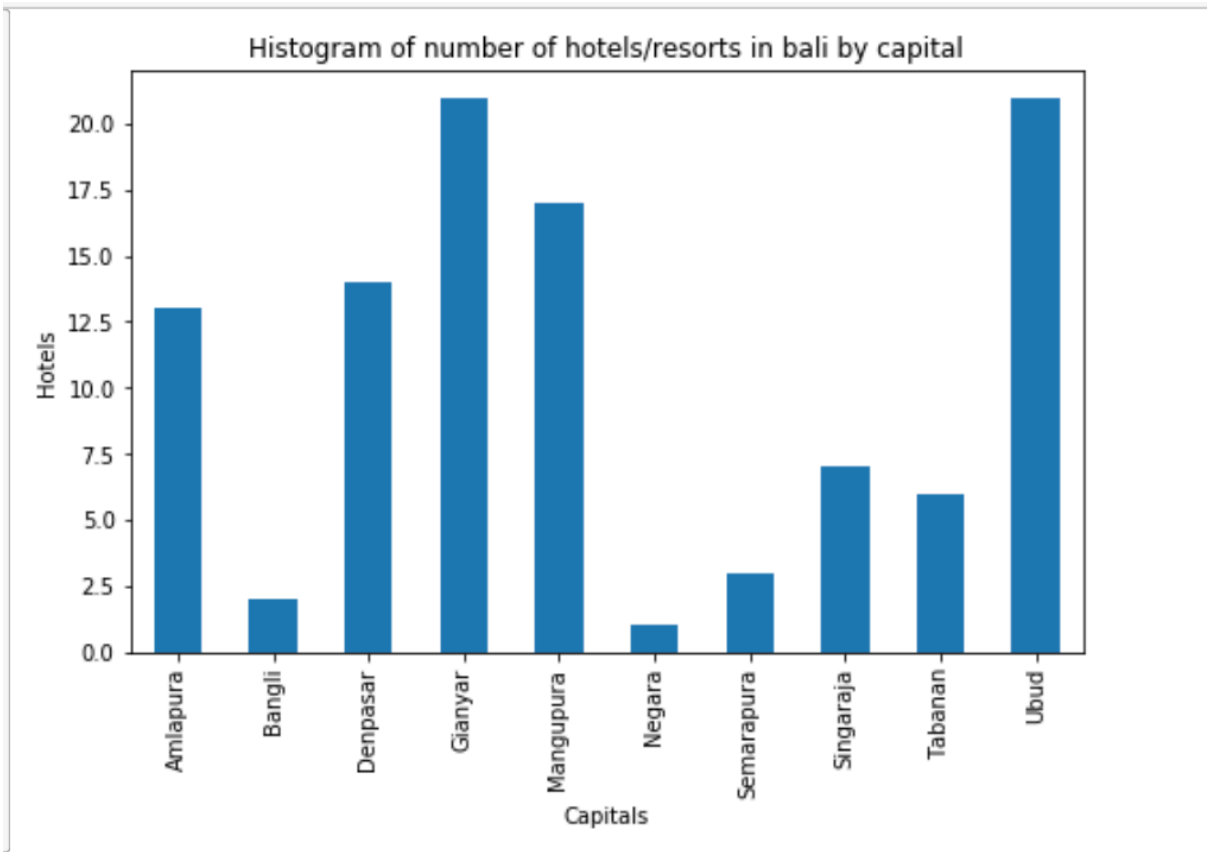
The next steps consisted of:

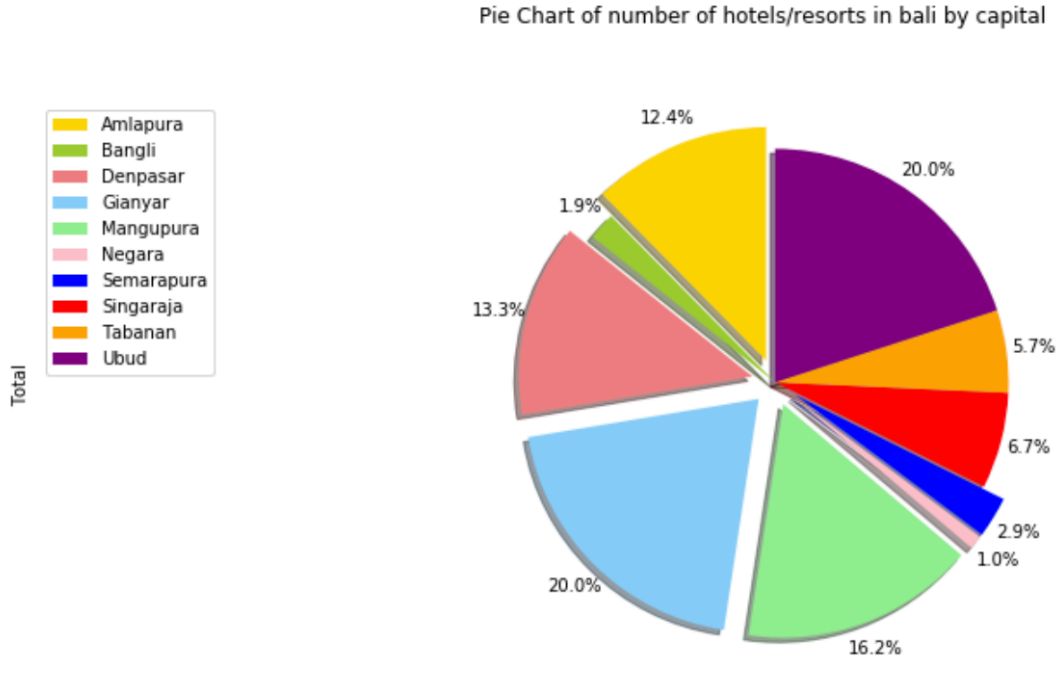
a. Exploring hotel and resorts venue categories to infer the cities that should be considered for a new hotels based on the existing competition:

The list of venues returned from Foursquare was narrowed down by categories: hotel and resorts and yielded 105 venues for all cities. These venues were represented on a map that was then clustered for visual convenience.



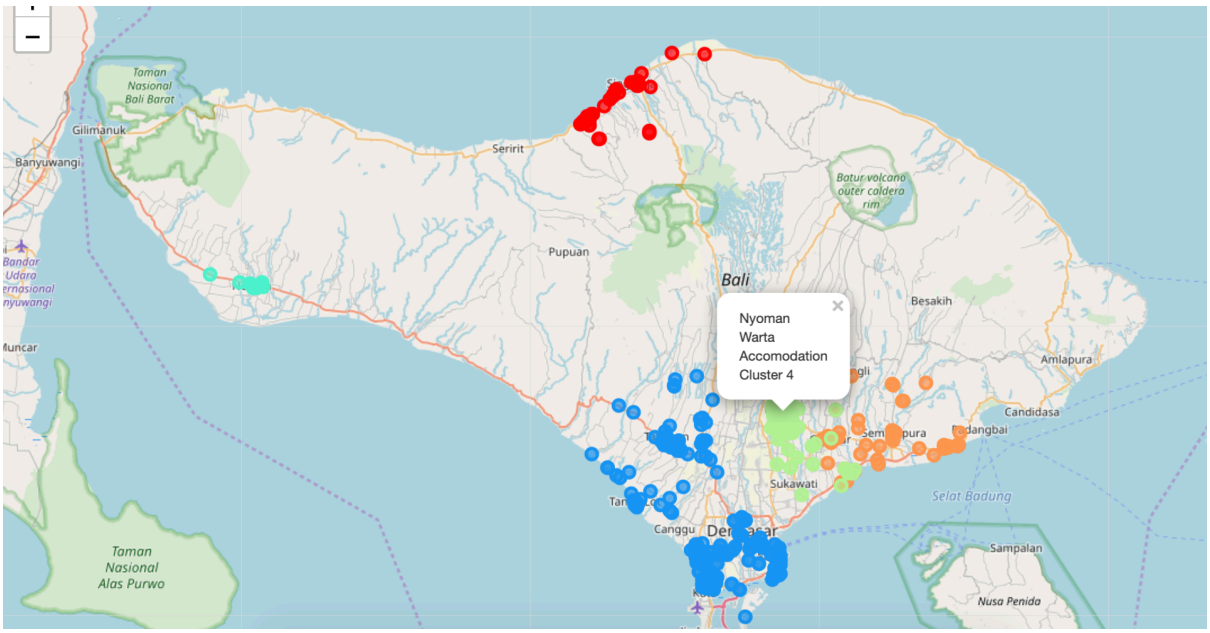
By counting the number of hotels/resorts per city we can infer that Gianyar and Ubud are the two cities with the most venues in this category. These counts were represented as histograms and Pie charts.





b. Exploring all venue categories to infer which city should be considered for new hotels based on existing hospitality services in the city such as restaurants, malls…

After encoding and looking at the top 5 most frequent venue categories per city, we K-Means to cluster the boroughs into clusters and then visualize the same on a map which led us to infer that Gianyar and Ubud cities are dense with diverse venue categories.



The last step of the exploration was to use Foursquare to get the number of likes for each venue of type hotel/resort to be able to identify the most liked ones based on user reviews which would support a deeper competitive analysis for new hotels.

	Capital	Capital Latitude	Capital Longitude	Venue	Venue id	Venue Latitude	Venue Longitude	Venue Category	Venue Likes
52	Mangupura	-8.69421	115.235809	Sheraton Bali Kuta Resort	4dff735e628421556008ee8c	-8.717966	115.169126	Hotel	230
112	Amlapura	-6.70287	108.482811	Aston Cirebon Hotel & Convention Center	50873793e4b095f041ac562d	-6.722276	108.539745	Hotel	212
53	Mangupura	-8.69421	115.235809	The Stones Hotel	50481fb4e4b06b582141a247	-8.711445	115.169402	Hotel	183
76	Gianyar	-8.54160	115.323051	Komune Resort and Beach Club	509f153ee4b0df8f7f4d0409	-8.598222	115.338460	Resort	165
77	Gianyar	-8.54160	115.323051	Komune Resort and Beach Club	509f153ee4b0df8f7f4d0409	-8.598222	115.338460	Resort	165
144	Ubud	-8.54160	115.323051	Komune Resort and Beach Club	509f153ee4b0df8f7f4d0409	-8.598222	115.338460	Resort	165
145	Ubud	-8.54160	115.323051	Komune Resort and Beach Club	509f153ee4b0df8f7f4d0409	-8.598222	115.338460	Resort	165
16	Denpasar	-8.67218	115.233550	Ramada Encore Hotel & Resort, Seminyak-Bali	54704d91498ed714fb972282	-8.685195	115.176169	Hotel	120
17	Denpasar	-8.67218	115.233550	Ramada Encore Hotel & Resort, Seminyak-Bali	54704d91498ed714fb972282	-8.685195	115.176169	Hotel	120
46	Mangupura	-8.69421	115.235809	Ramada Encore Hotel & Resort, Seminyak-Bali	54704d91498ed714fb972282	-8.685195	115.176169	Hotel	120

4. Results section where you discuss the results.

Looking at the frequency and distribution of venues in the Bali cities we can make the initial inferences:

a. Gianyar and Ubud being the two cities with the most hotels are attractive to visitors and should be considered for potential new hotels. Although the instinct would be to go to a less populous city, but a different strategy is to introduce a new player in the same ones
b. The same two cities seem to have a high number of diverse venues which makes them attractive to visitors and complements the hotel business
c. Having data about the most like venues can be used to dig deeper into those specific venues to understand them more from a competitive advantage perspective

5. **Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.**

My recommendation would be to get visitor data mapped to the cities and analyze which cities attract more visitors during the different seasons. Such information would support (or not) the choice of cities made. Another suggestion to take this further is to build a model to predict potential revenues of a new hotel based on market data and expected visitor figures.

6. **Conclusion section where you conclude the report.**

The data is clear on which cities have the highest number of venues by type and based on the strategy for a new hotel this can be used to either launch into a new market/city or introduce competition.