Finding the Most Suitable Neighborhood to Visit in Bali, Indonesia

Rafki Islamy

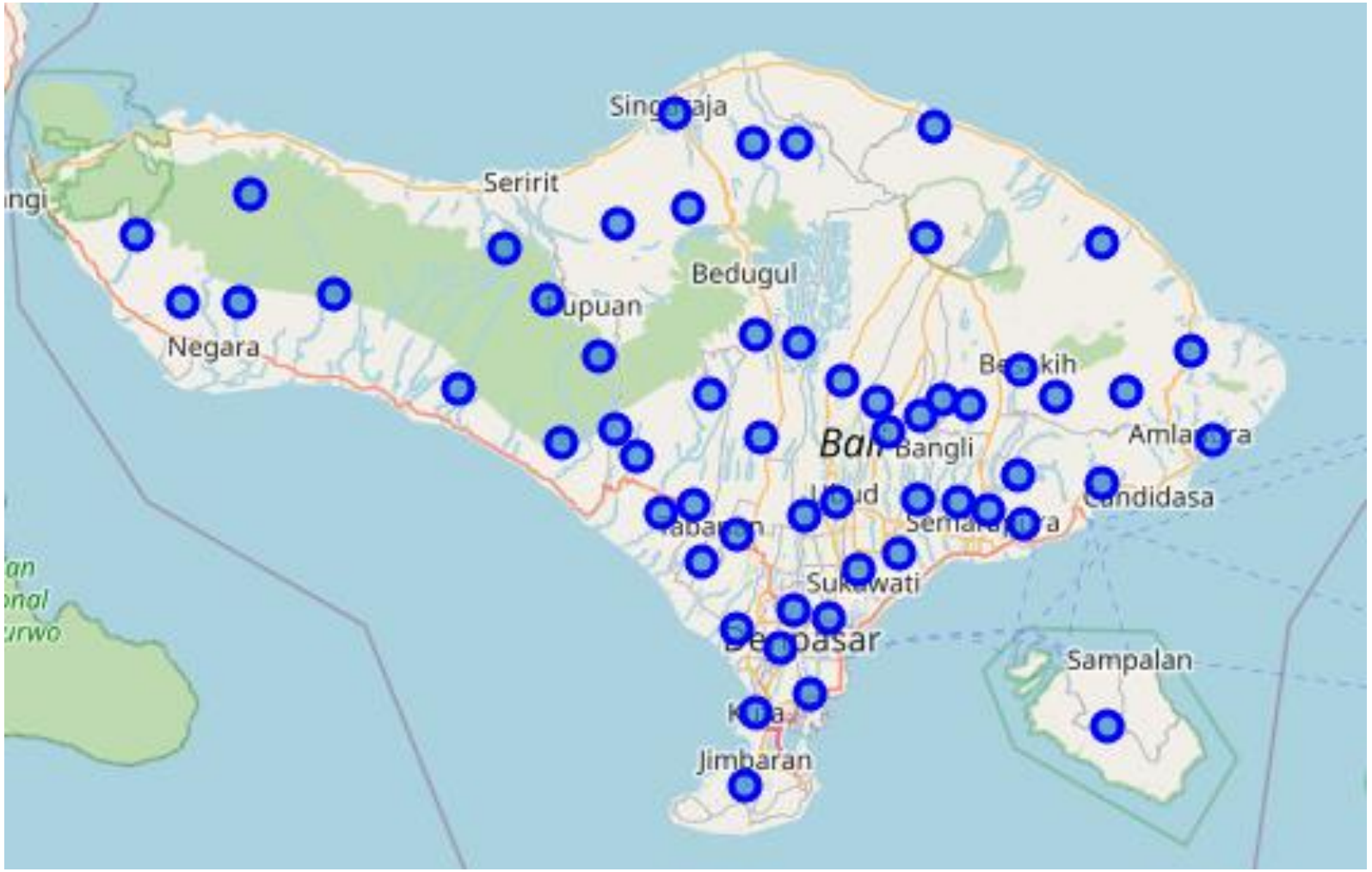
Introduction: Business Problem

- Each group of tourist has their on preferences on deciding where they want to go in Bali.
- It is because the differences between group of tourists 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, or the local tourist are most likely visit shopping areas.
- Finding the most suitable neighborhood to visit in Bali will help tourist in their holiday experience by recommending several neighborhoods that suits their preferences.
- The same things also be beneficial to the Investor who are looking for the best location to open their specific business.

Data Preparation and Cleaning

- Bali data containing the city, neighborhood and coordinates. Scrapped from: https://raw.githubusercontent.com/ArrayAccess/Indonesia-Postal-And-Area/master/data/csv/62/subDistricts.csv
- City size area in Bali. Scrapped from: https://en.wikipedia.org/wiki/List_of_districts_of_Bali
- All venues in Bali. Obtained from Foursquare API utilized via the request library in Python.
- In total, there are 1,489 row and 8 features in Datasets.

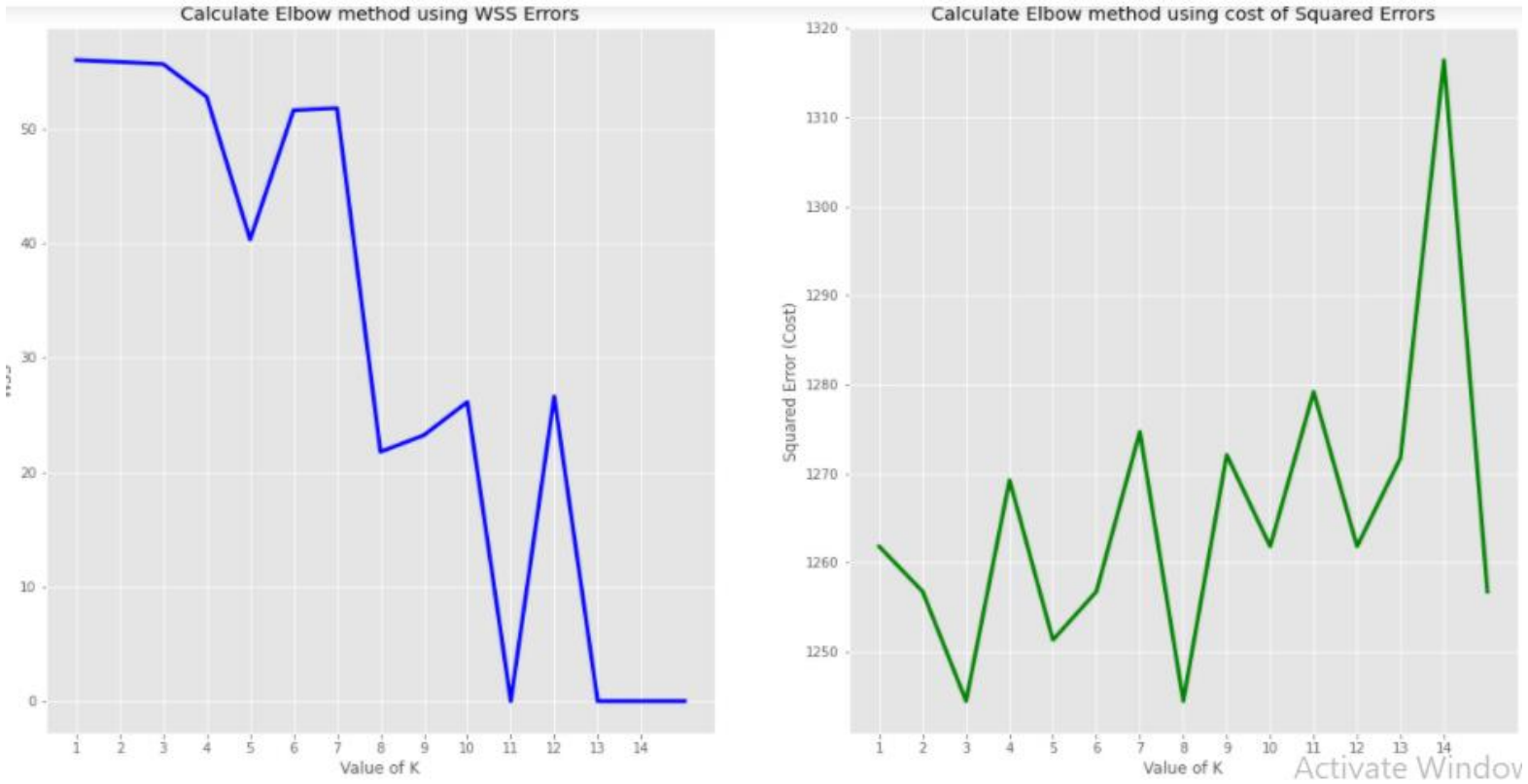
Bali Map Plotting Along With Neighborhood Labels



Problem Approach Using K-Means Clustering

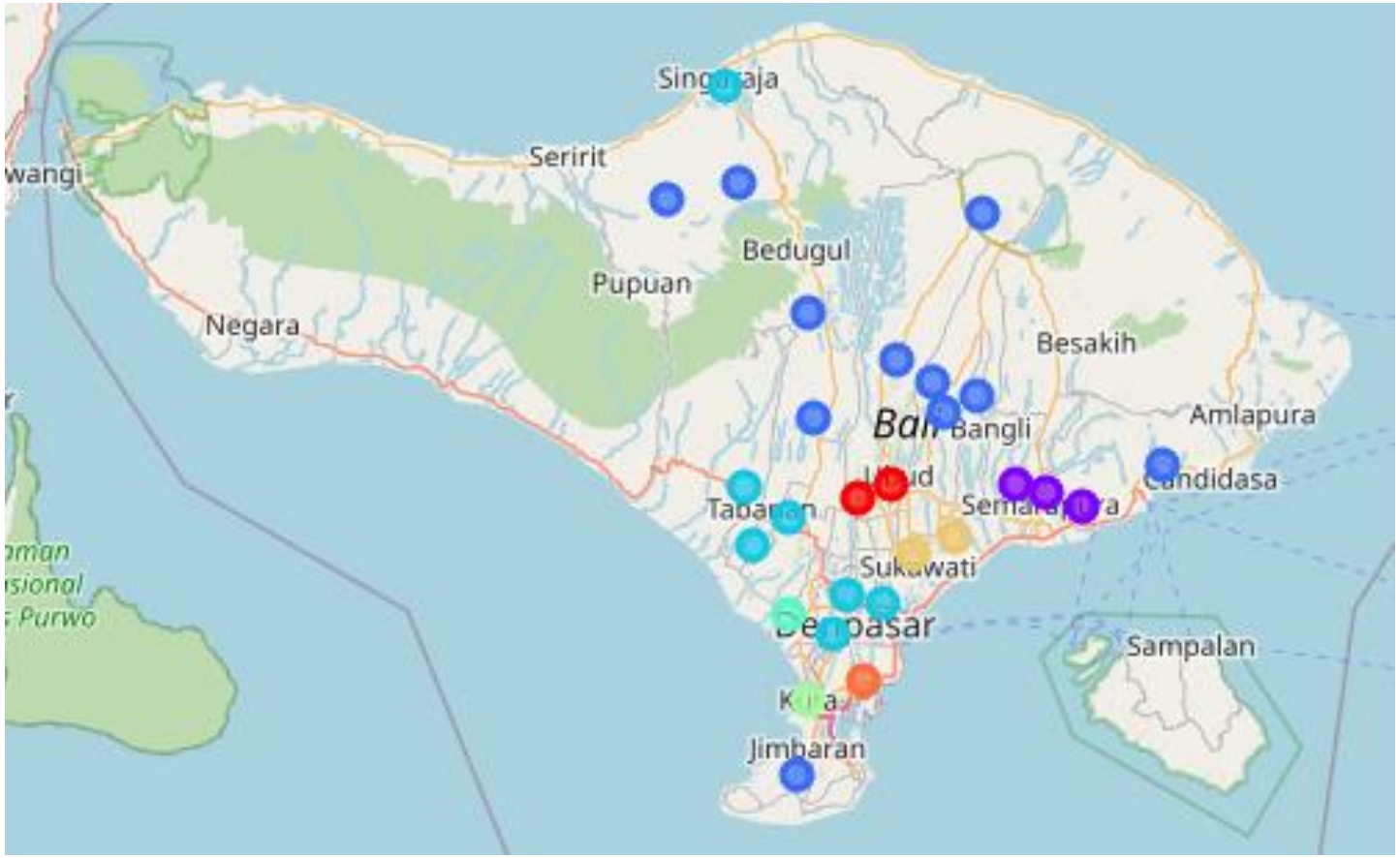
- This problem can be solved using machine learning algorithm that is clustering analysis.
- This kind of analysis is used because we need to clearly group the neighborhood based on the most common venues for each neighborhood.
- K-Means is used because of it is widely used in clustering analysis.

Finding the Best K-Value



Using elbow method, we get k-value = 8

Clustering Results



Discussion

- Main Characteristics of Each Cluster:
- Cluster 1 (Red) : Resort, Vegan Restaurant
- Cluster 2 (Purple) : Historical Tourism (Site, Museum)
- Cluster 3 (Dark Blue) : Natural Tourism (Farm, Mountain, Beach, Waterfall)
- Cluster 4 (Light Blue) : Asian Cuisine (Indonesian, Asian, Chinese Restaurant)
- Cluster 5 (Tosca) : Cafe, Hotel
- Cluster 6 (Green) : Coffee Shop, Clothing Store
- Cluster 7 (Yellow) : Art (Gallery, Theather, Craft store)
- Cluster 8 (Light Red) : Food Destination
- Based on those main characteristics, the tourist can be more easily make a choice about which neighborhood to visit based on their own preferences.

Conclusion and Future Direction

- This analysis can helped the stakeholder in giving the recommendation about what location that suit their expectation.
- This analysis is performed on limited data. This may affect in the result accuracy that might be low.
- This analysis could be better by adding other variables such as venue ratings and tourism demography, as the other potential variable for analysis.