

Clustering Regions in Yogyakarta, Indonesia Based on It's Region Food Preferences

Rizky Oulia O. P. L.

July 27th, 2020

ouliaopl@gmail.com

Introduction

- Yogyakarta is a really diverse place in Indonesia. Food and beverage businesses is one of the most attractive and lucrative market.
- **Interests:** This project is meant to help entrepreneurs, local cooks and chefs to better understand the correlation between the location and the best type of restaurant to be opened based on local palate. By choosing the best location based on the type of restaurant they're going to open, hopefully they could minimize the risk of bankruptcy and maximize profit.
- **Problems:** To better understand the palate preferences in Yogyakarta, foursquare venue data will gives us relevant information about restaurant type and it's location. By using machine learning algorithm such as K-Modes clustering, we can group neighborhoods that have similar food palates based on type of restaurants thus giving us cluster of food preferences.

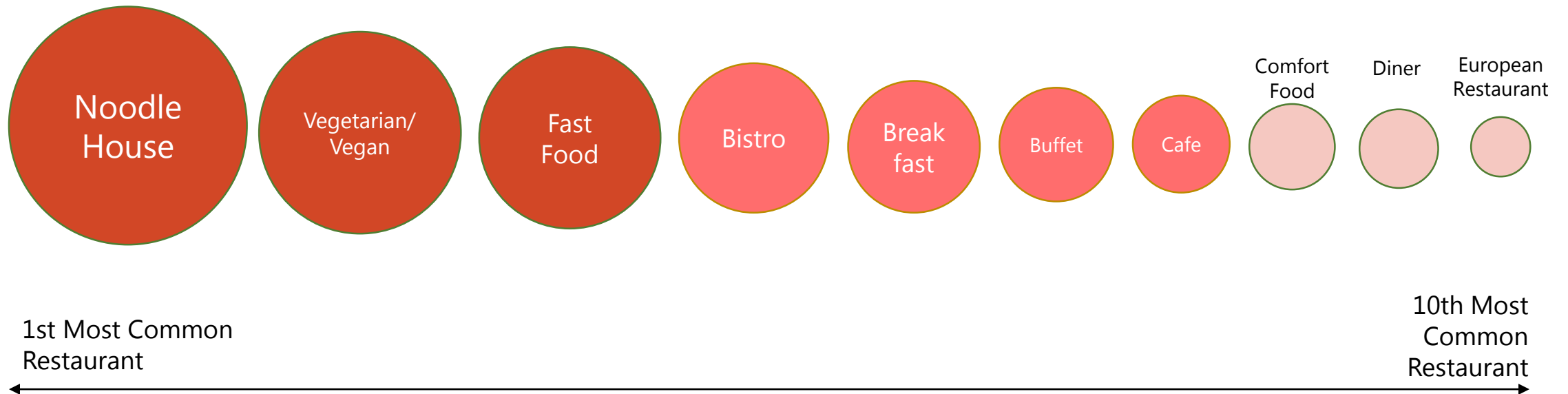
Data Acquisition and Cleaning

Data acquired are from:

- Data retrieved from [https://kodepos.nomor.net/kodepos.php? i=kota-kodepos&daerah=Provinsi&jobs=DI+Yogyakarta&perhal=400&urut=10&asc=00001111&sby=110000&no1=2](https://kodepos.nomor.net/kodepos.php?i=kota-kodepos&daerah=Provinsi&jobs=DI+Yogyakarta&perhal=400&urut=10&asc=00001111&sby=110000&no1=2) for **region names and postal code information.**
- Geopy library to find **geographical location information.**
- Using Foursquare API to find nearby **restaurants information such as their location and types of dishes they sell.**

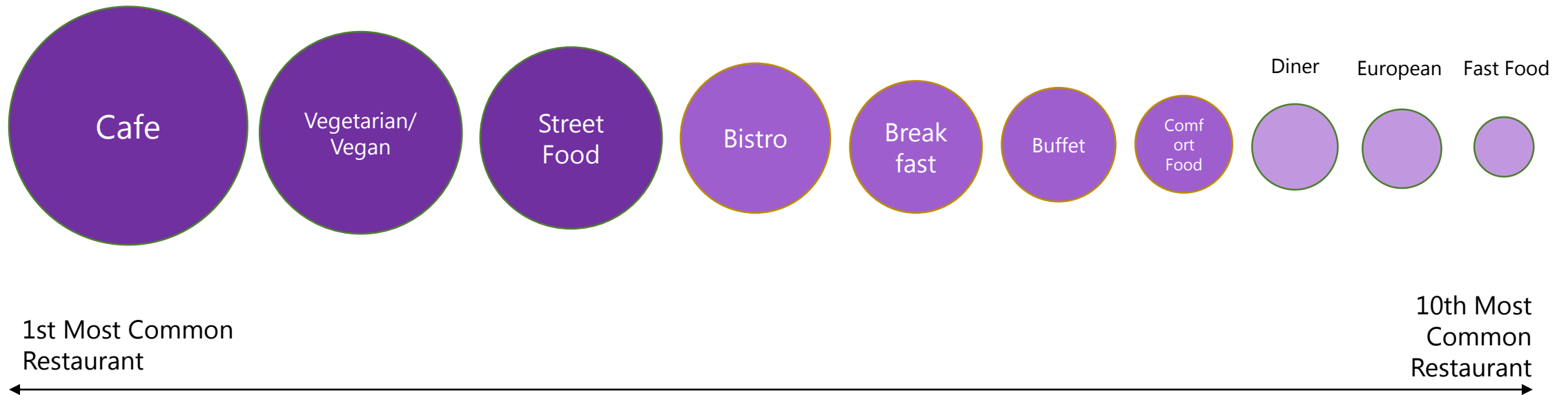
Clusters That Were Obtained Using K-Mode Method

Cluster 1



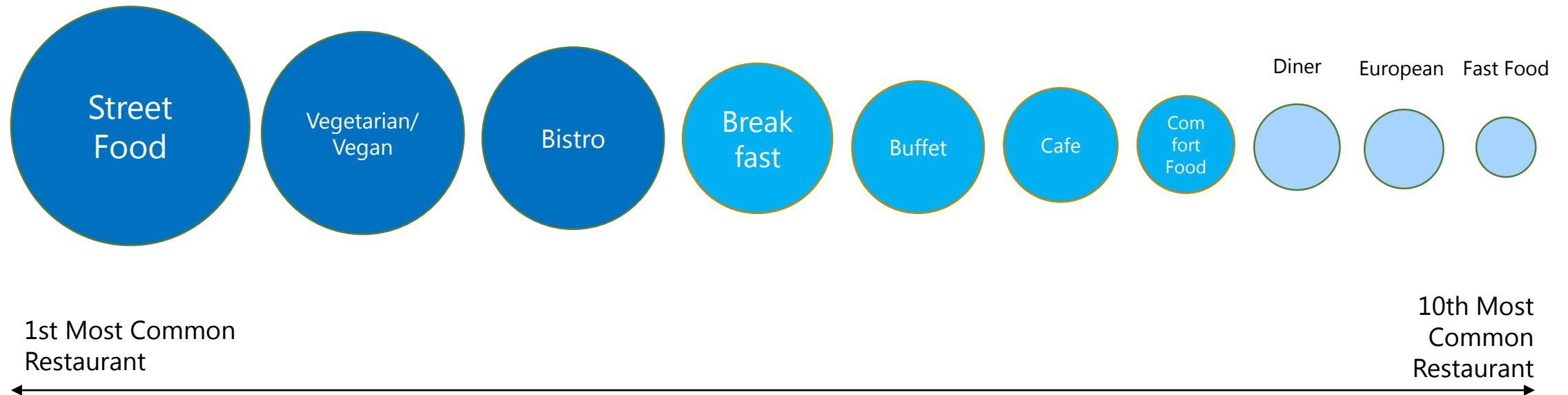
Clusters That Were Obtained Using K-Mode Method

Cluster 1



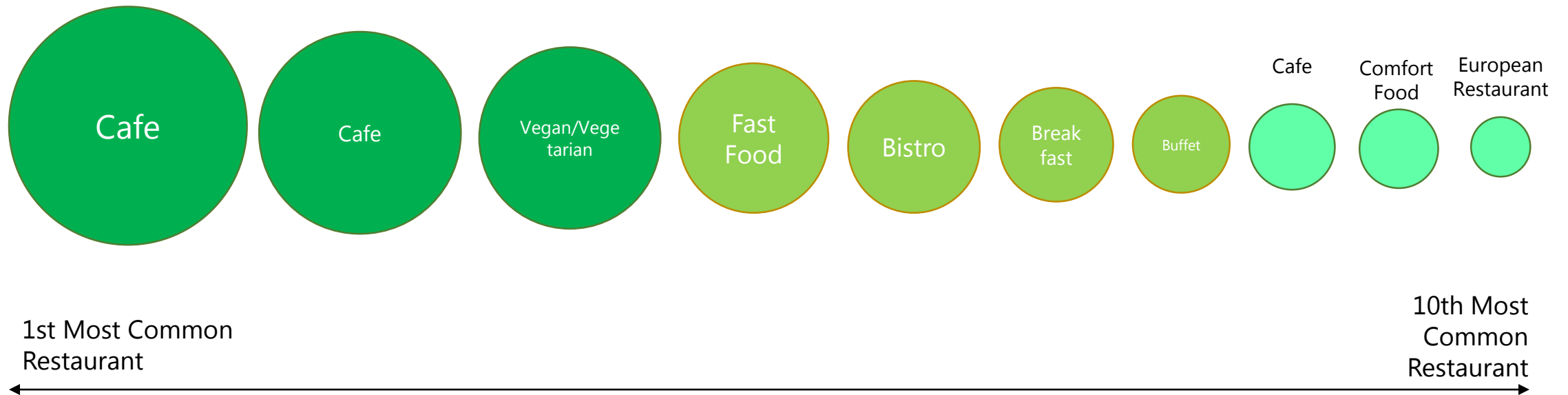
Clusters That Were Obtained Using K-Mode Method

Cluster 1

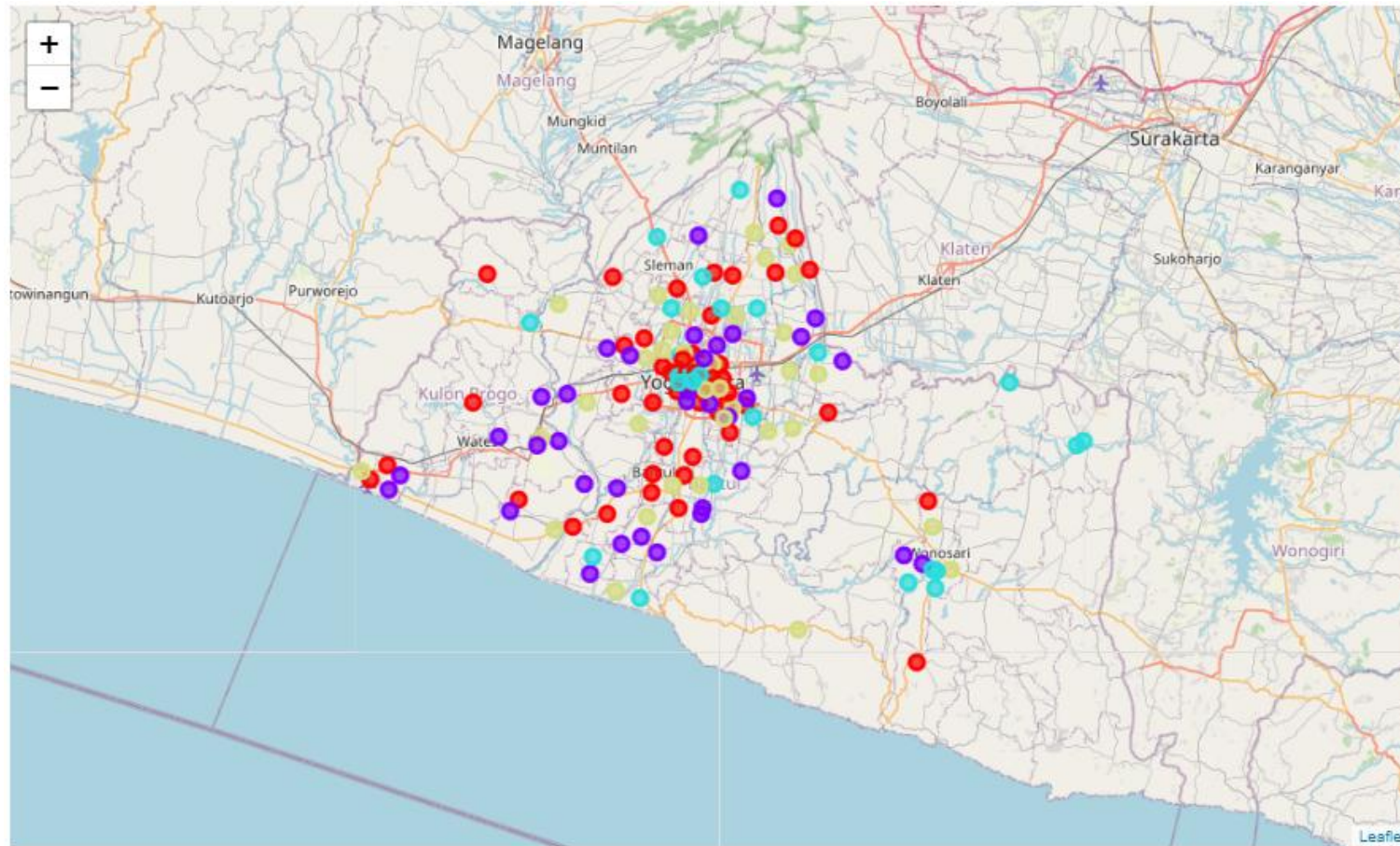


Clusters That Were Obtained Using K-Mode Method

Cluster 1



Distribution of Clusters on the Map



MAP LEGEND

- Cluster 1 - Red dots
- Cluster 2 - Purple dots
- Cluster 3 - Blue dots
- Cluster 4 - Green dots

Conclusion and Future Directions

Restaurant Recommendations:

Cluster 1:

Types: Bistro, Breakfast Spot, Buffet;

Group of People: All;

Price: Low-Medium-High

Cluster 2:

Types: Bistro, Breakfast Spot, Buffet;

Group of People: Tourist;

Price: Medium-High

Cluster 3:

Types: Breakfast Spot, Buffet, Cafe;

Group of People: Tourists;

Price: Medium-High

Cluster 4:

Types: Fast Food, Bistro and Breakfast Spot;

Group of People: Students, Residents;

Price: Low-Medium

Future Directions:

- This model only gives **62% of accuracy**
- I think **there may be better approaches other than the K-Modes** that was used in this study. Furthermore, **more data might be the answer** since I observed that small and rural regions does not have many data or even worse, no data at all from the Foursquare API.
- **This study also excluded the frequency/number** of restaurants on each region since I only cluster the region by ranking. **This may cause overvalue-ing regions with lack of data and undervalue-ing regions with lots of data.** And **more study about other attributes such as prices and group of people is needed** to better understand the clustering. Since the foursquare API have limited access to such data, I did not include it in the study.