



# Segmentation of Places

A cluster based machine learning approach to segment places and provide some valuable insights to the stakeholders

Saurabh Naik  
01-11-2022



# Problem Statement

- You will be given POI (Point of Interest) data from OSM (openstreetmaps).
- These POIs will include locations ranging from grocery stores, shopping malls to car dealerships within the city.
- You may also enrich your POI (point of interest, meaning, what type of place- showroom/ building/ outlet/ shop etc.) data using any location data available on the web (that you can extract/scrape).
- Your goal is to divide the city geographically into various blocks/localities. Next you are supposed to create clusters of similar localities and characterize each cluster so that these clusters are human interpretable.

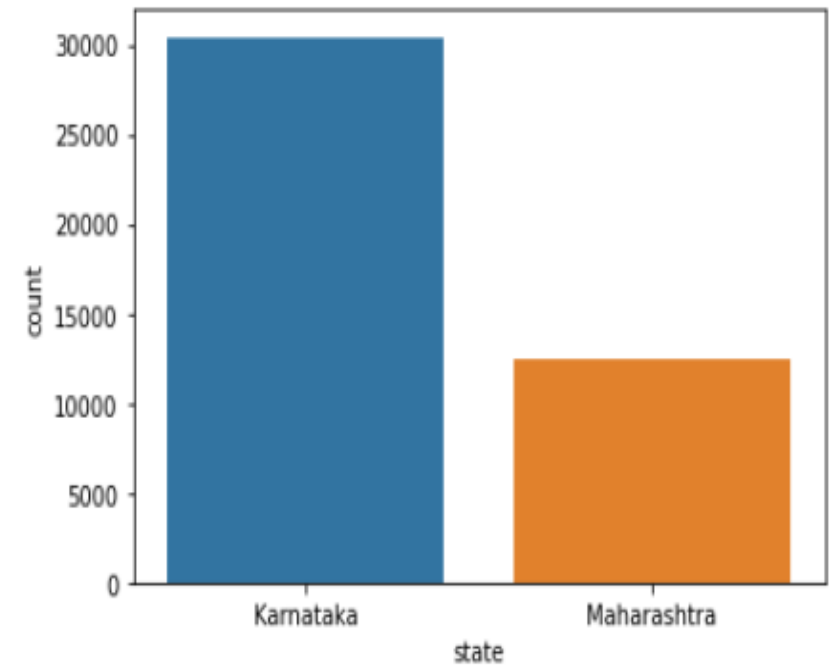
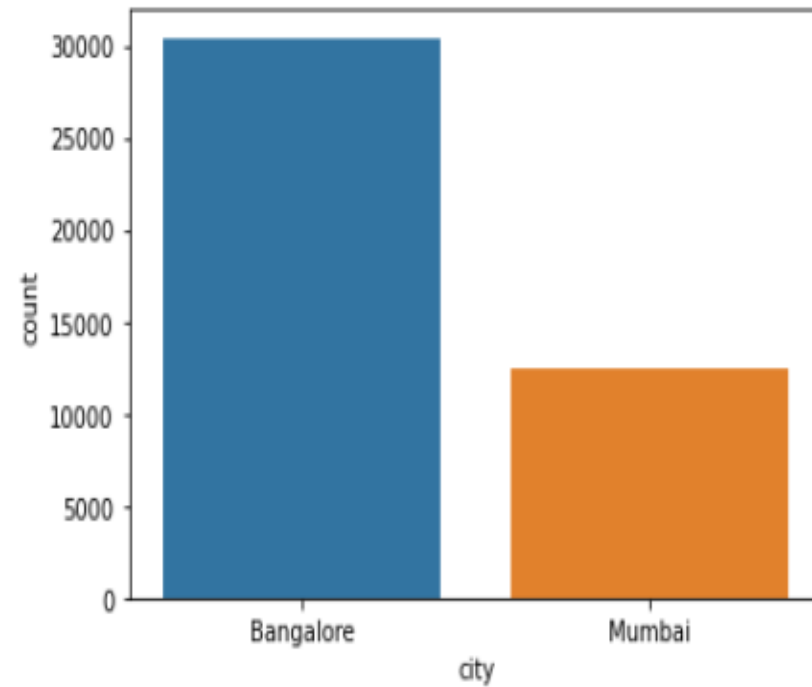


# Data Description

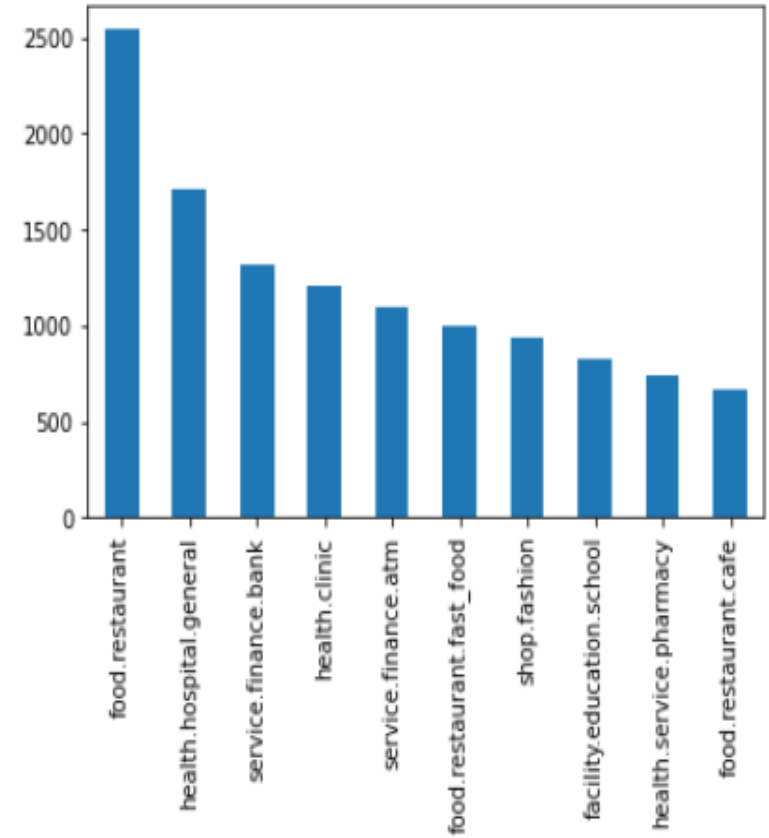
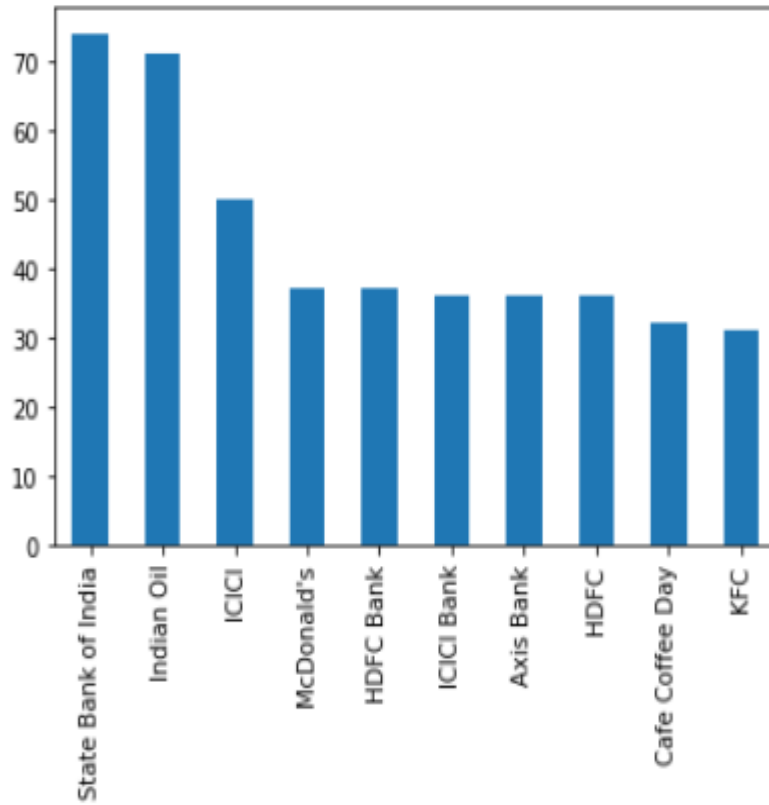
The schema of OSM data is given below:

- 1. source: source from where the data was collected
- 2. poi\_code: unique identifier of the POI
- 3. name: name of the POI
- 4. poi\_type: type of POI (e.g car dealership, shopping mall, etc)
- 5. lat: latitude of the POI
- 6. long: longitude of the POI
- 7. address: address of POI
- 8. city: city of POI
- 9. state: state of POI
- 10. country: country of POI
- 11. pin\_code: pincode of POI
- 12. brand: brand information of POI

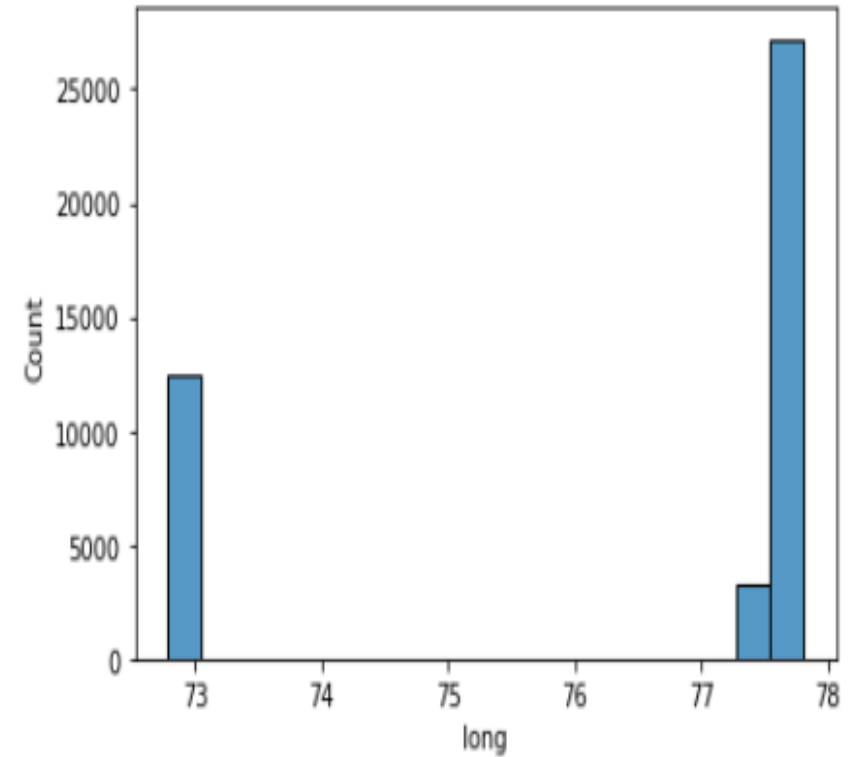
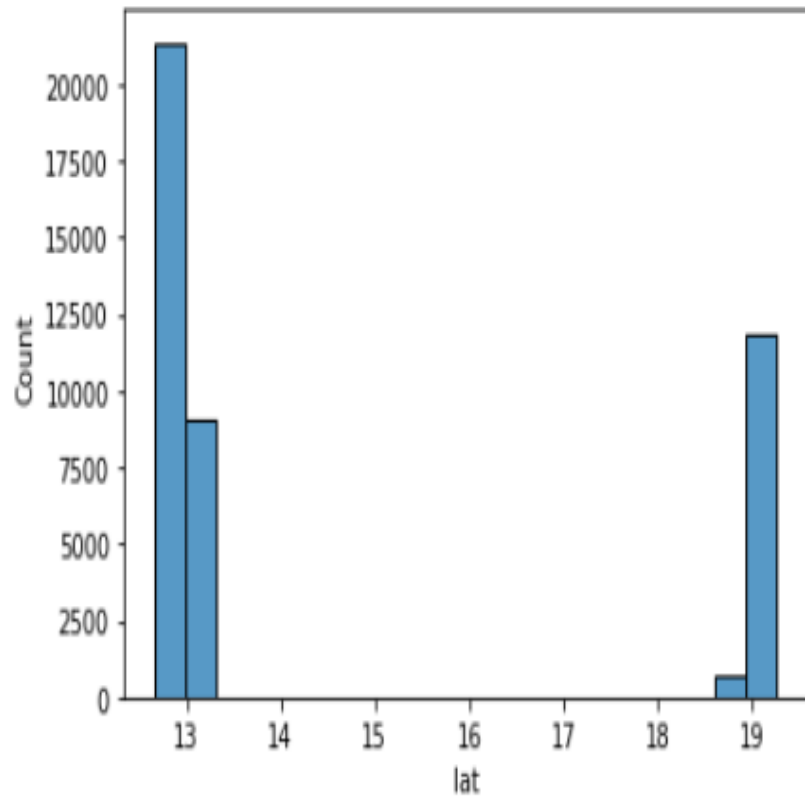
# Distribution of data into different cities and states



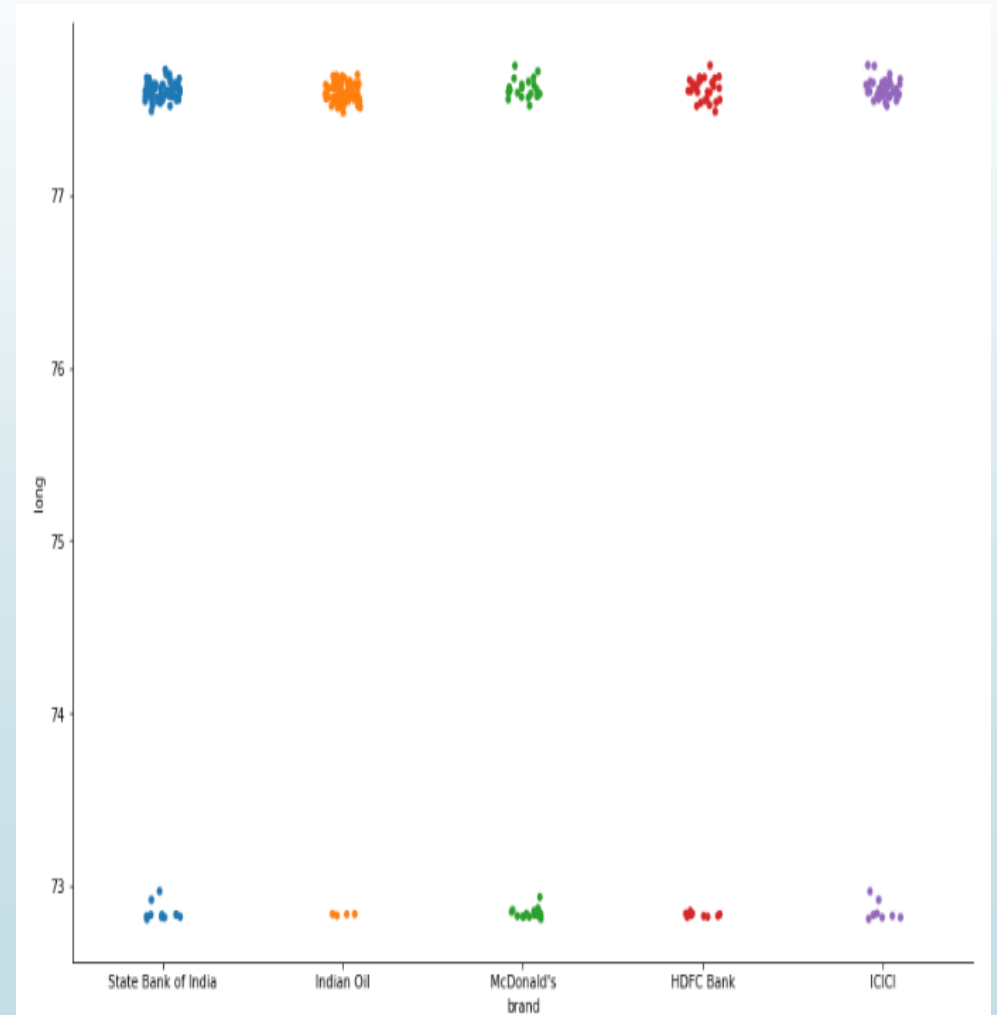
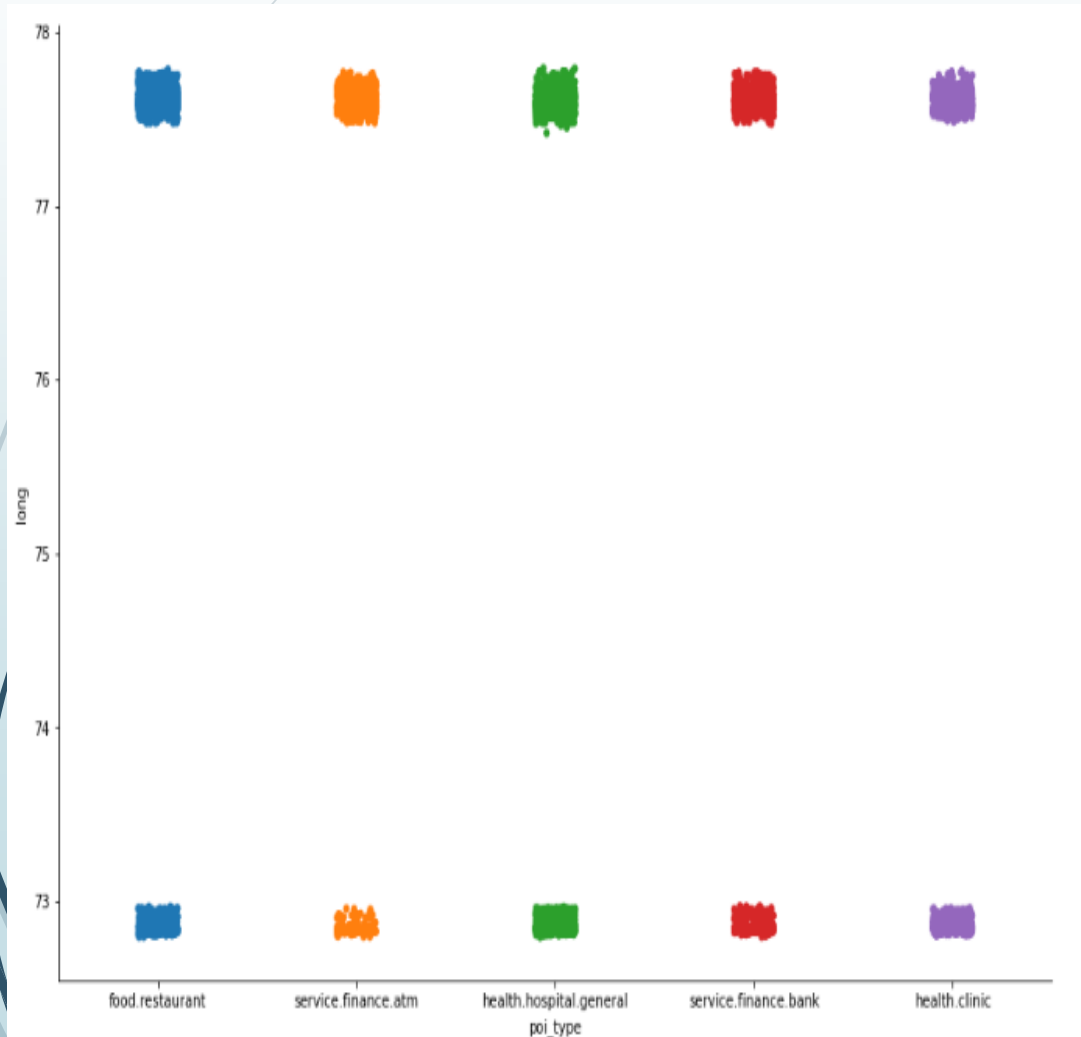
# Distribution of top 10 brands and poi\_type from dataset



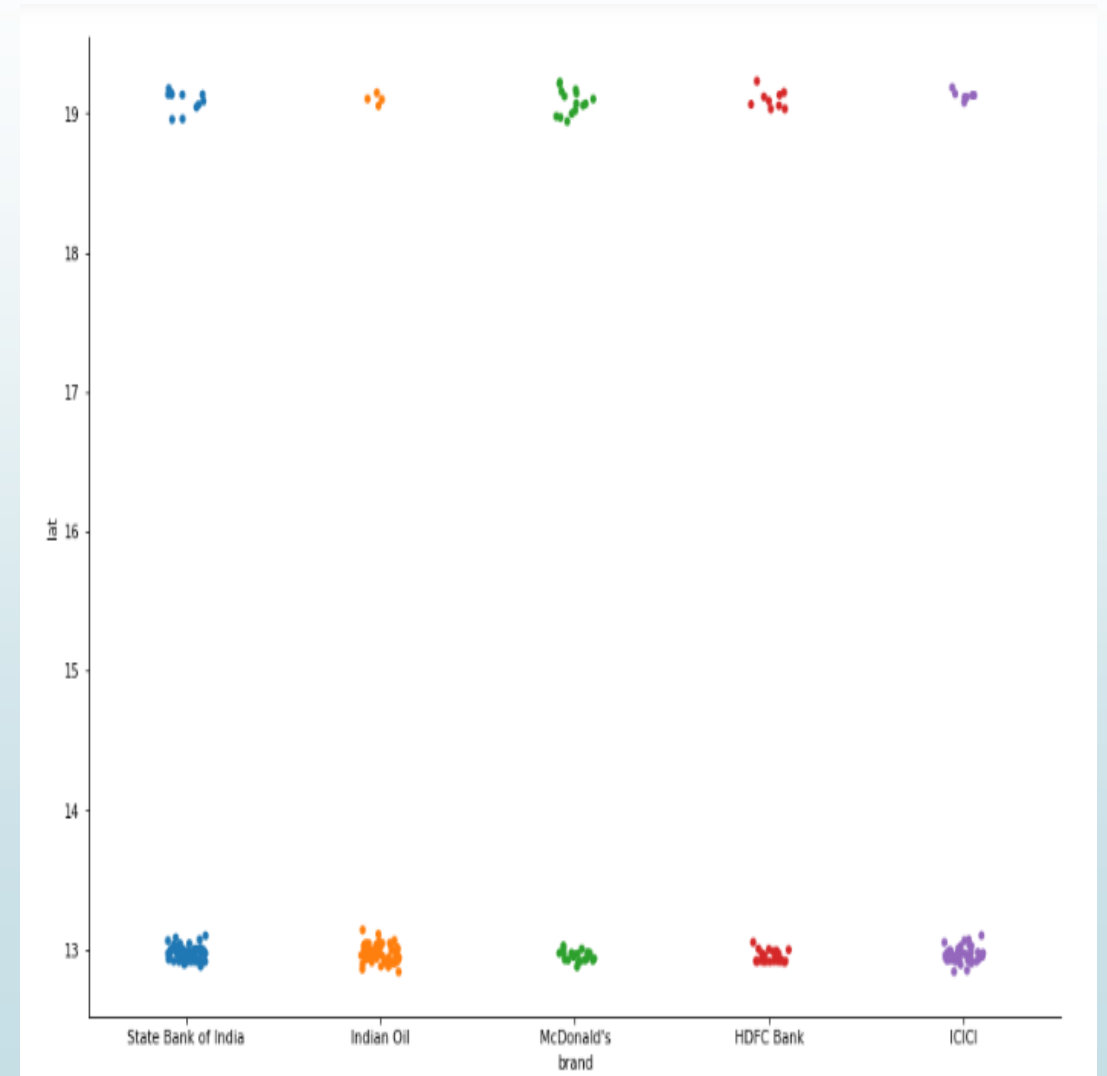
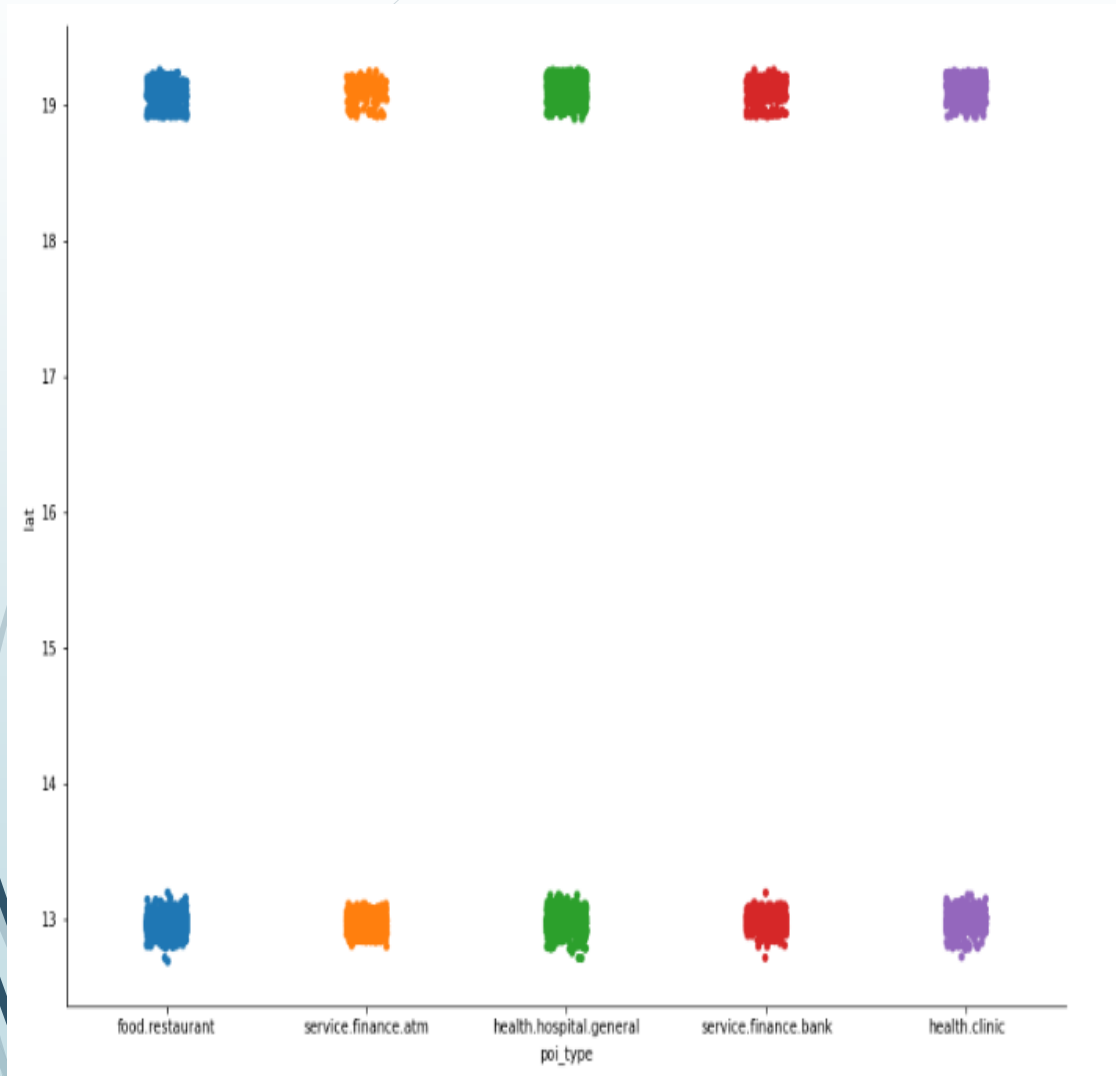
# Distribution of latitude and longitude from the given dataset



# Top 5 poi\_type and brands with respect to longitude

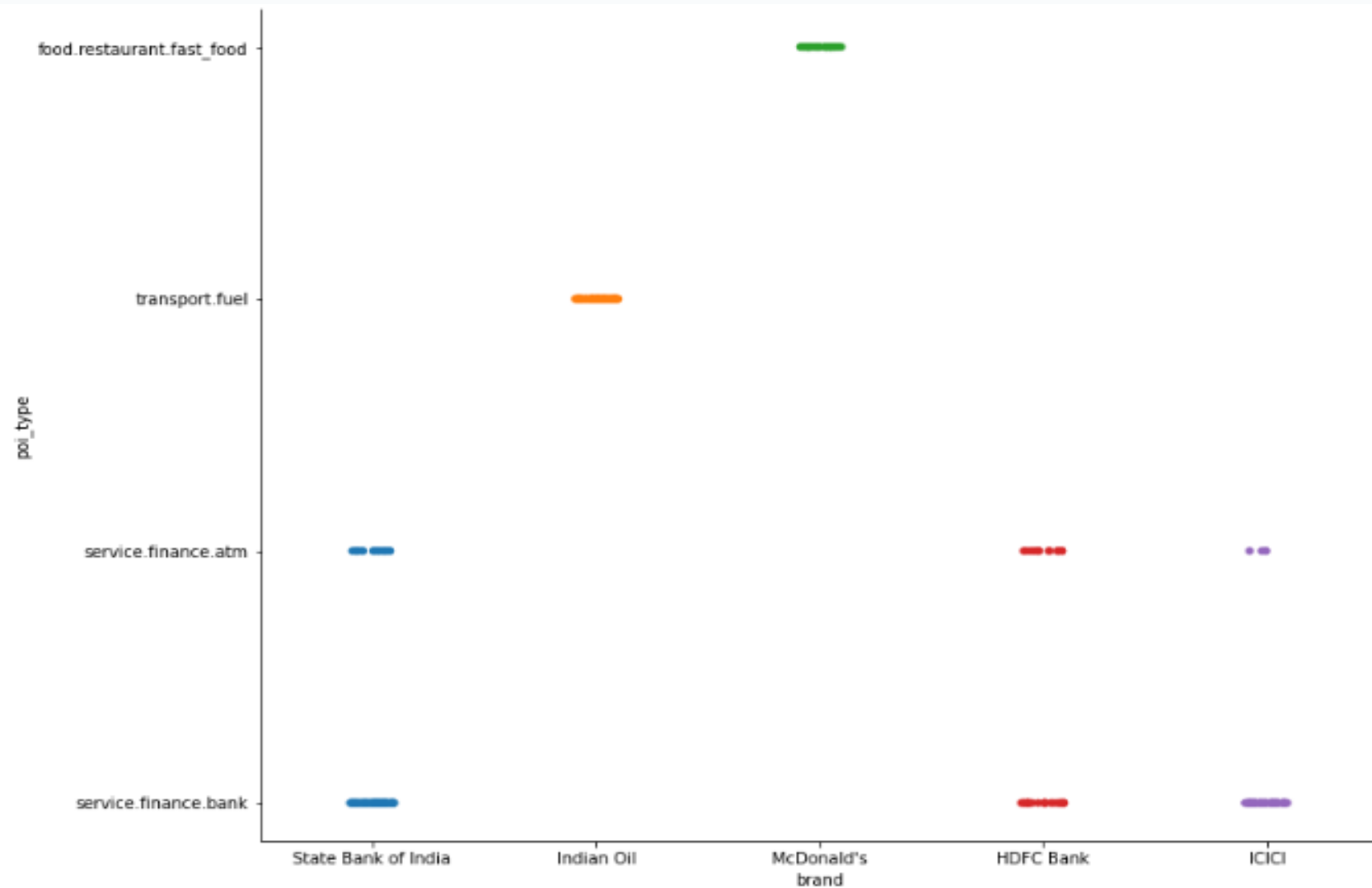


# Top 5 poi\_type and brands with respect to latitude

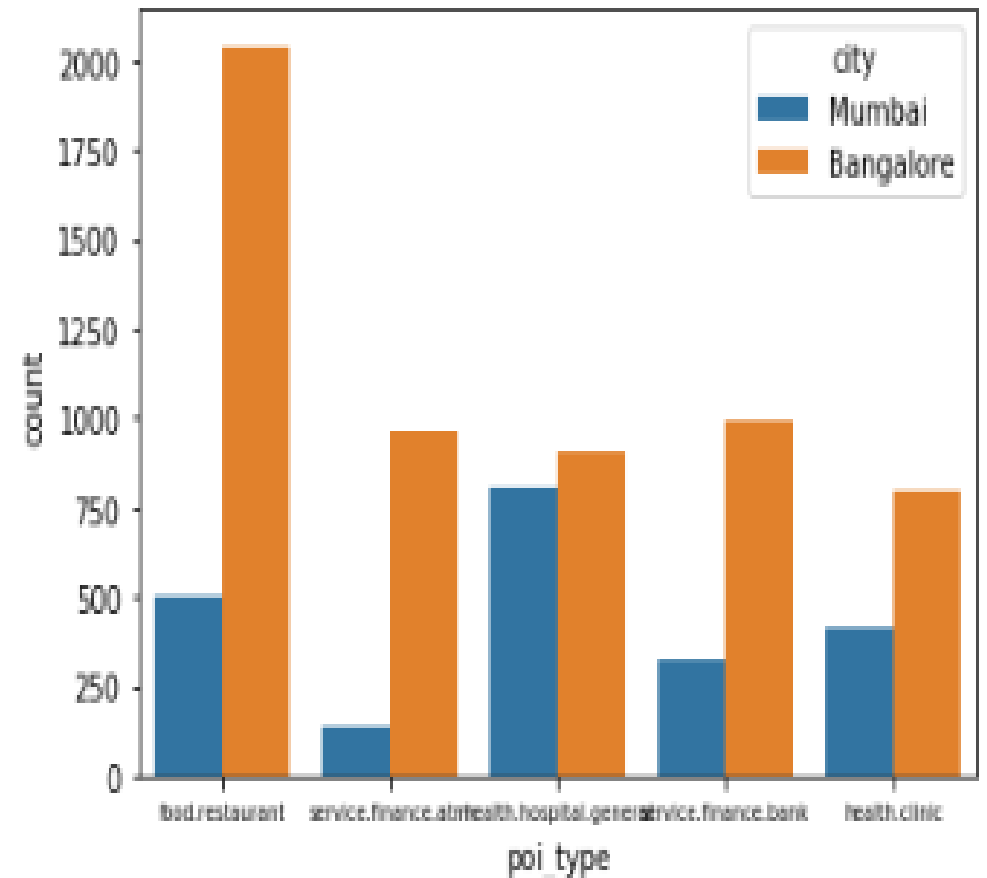
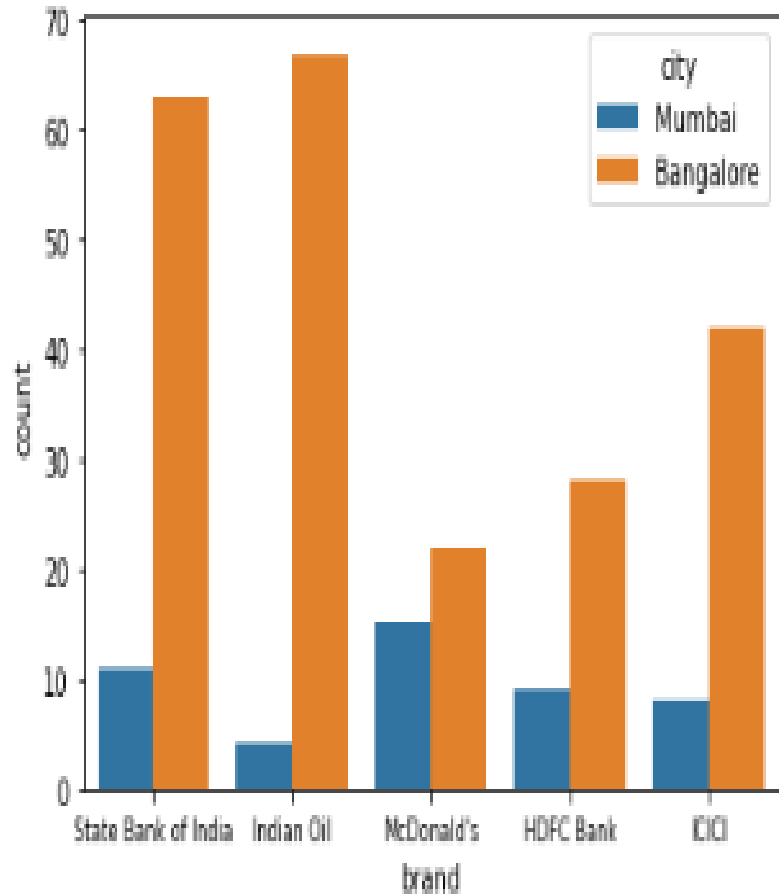




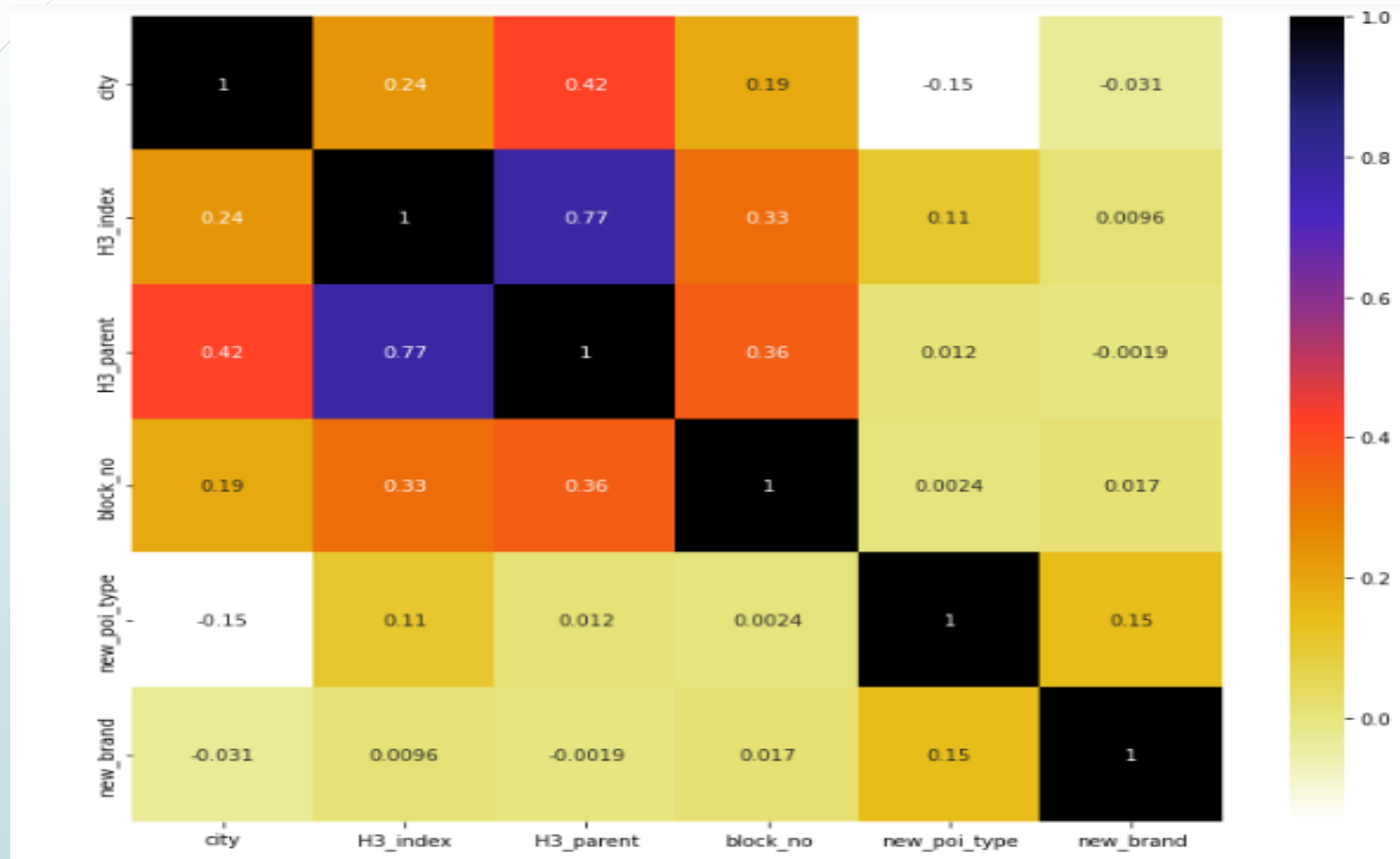
# Which top 5 brands represents which poi\_types?



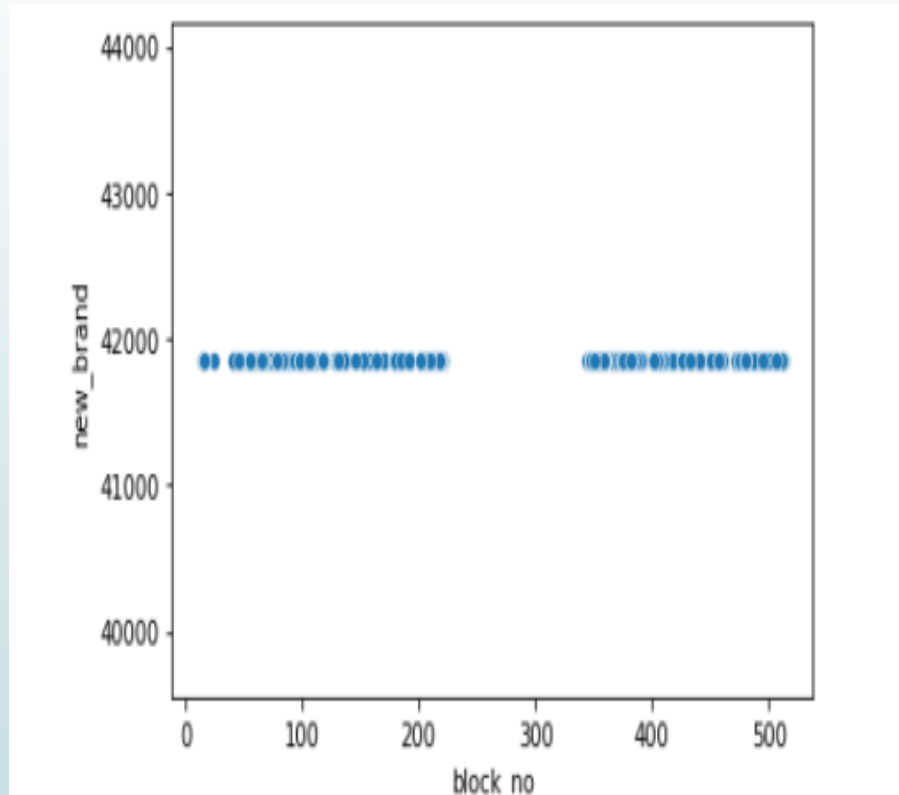
# Distribution of top 5 brands and poi\_type with respect to the cities



# Correlation Matrix for all features used for clustering

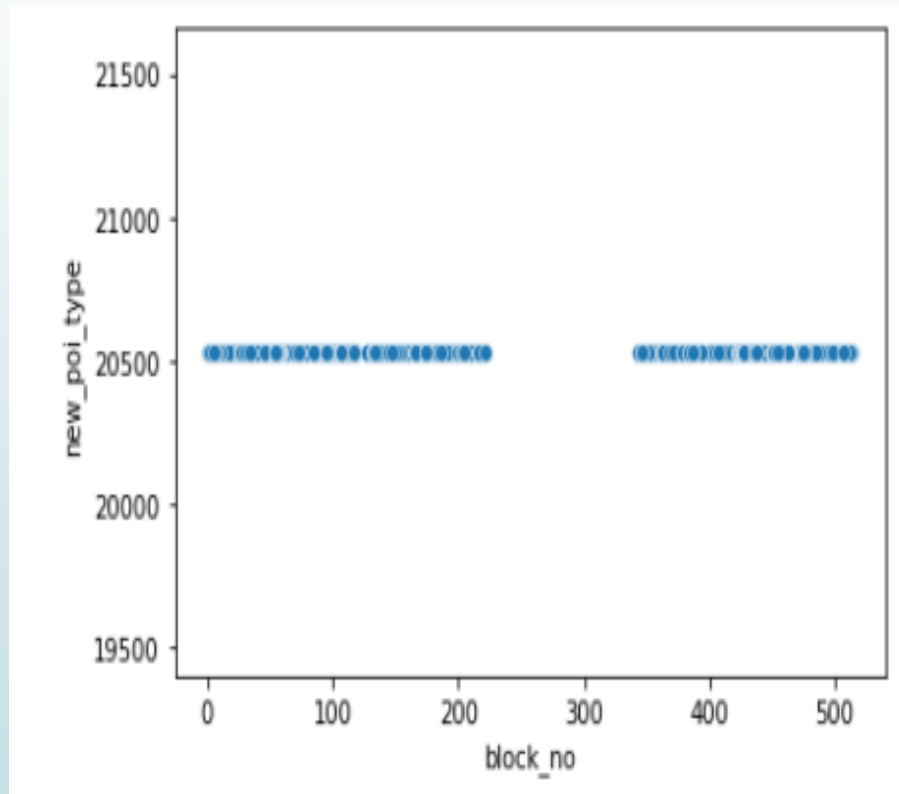


# Cluster 0



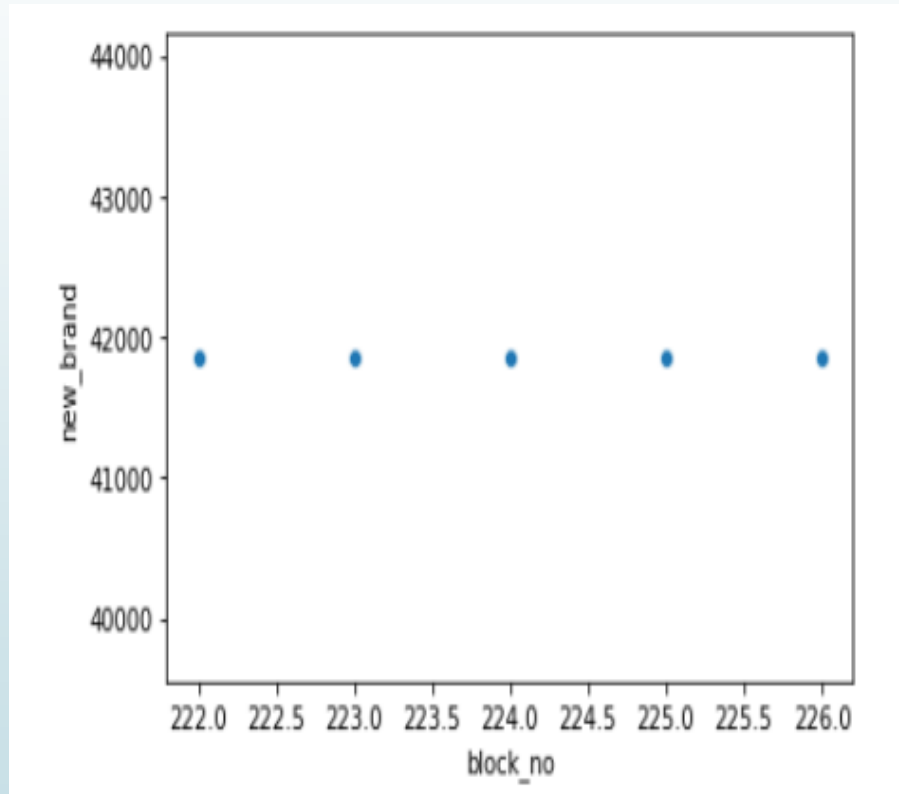
- From this plot we can conclude that block/locality of this cluster0 only contains one type of brand which is represented by 42000 in numerical format

# Cluster 1



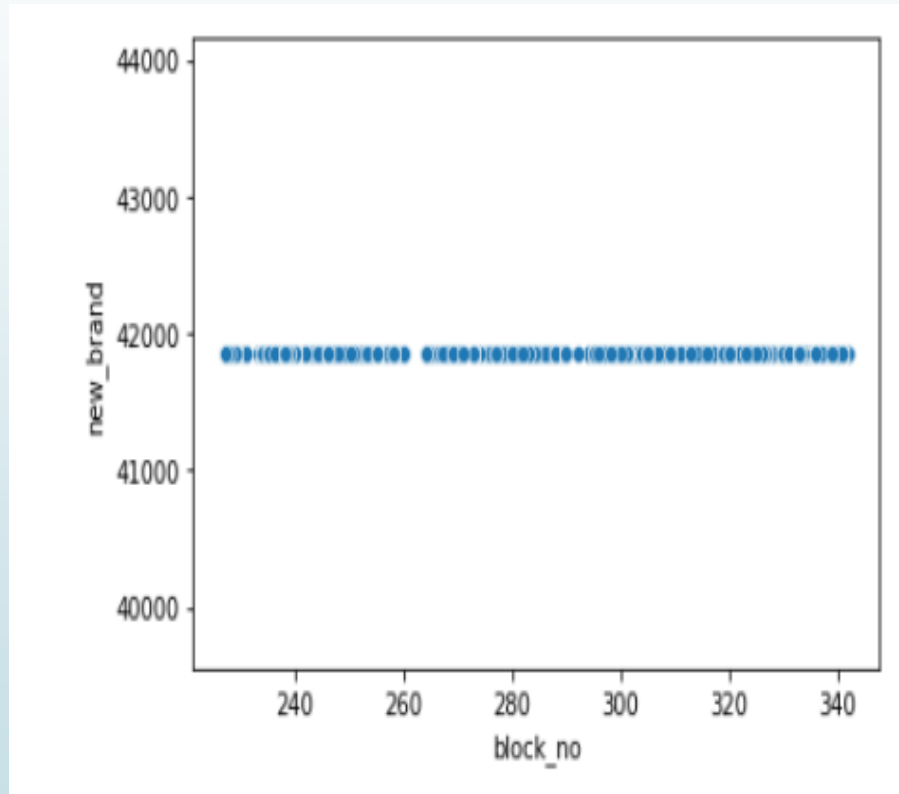
- From this plot we can conclude that block/locality of this cluster only contains one type of poi\_type which is represented by 20500 in numerical format

## Cluster 2



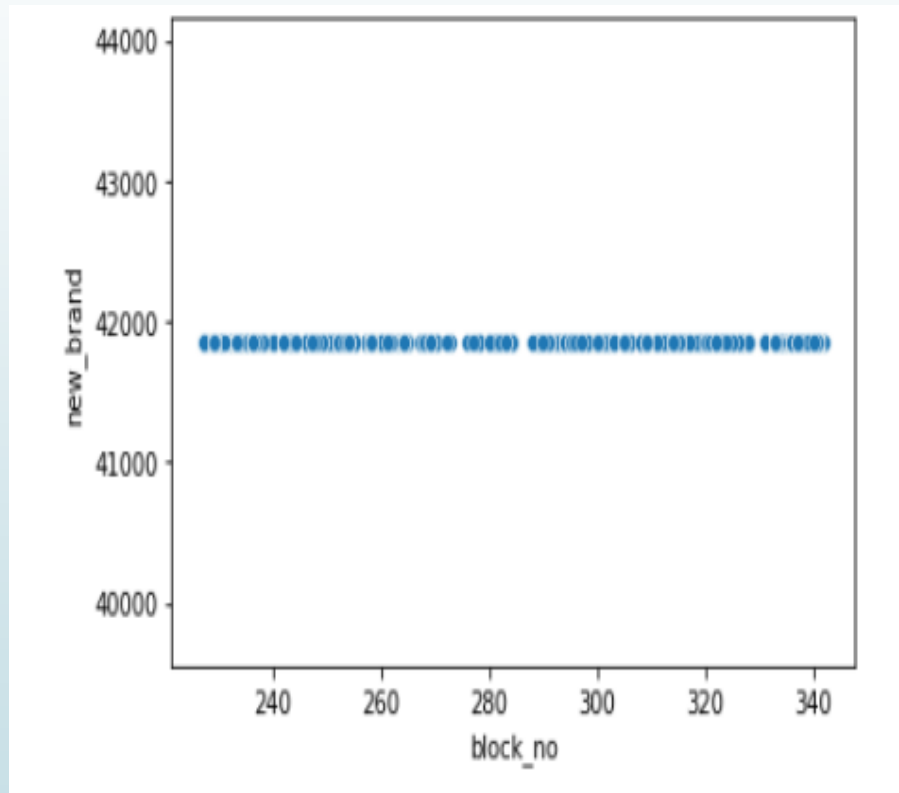
- From this plot we can conclude that block/locality of this cluster are ranging from 222 to 226 so it is a small region and it only contains one type of brand which is represented by 42000 in numerical format

## Cluster 3



- From this plot we can conclude that block/locality of this cluster only contains one type of brand which is represented by 42000 in numerical format

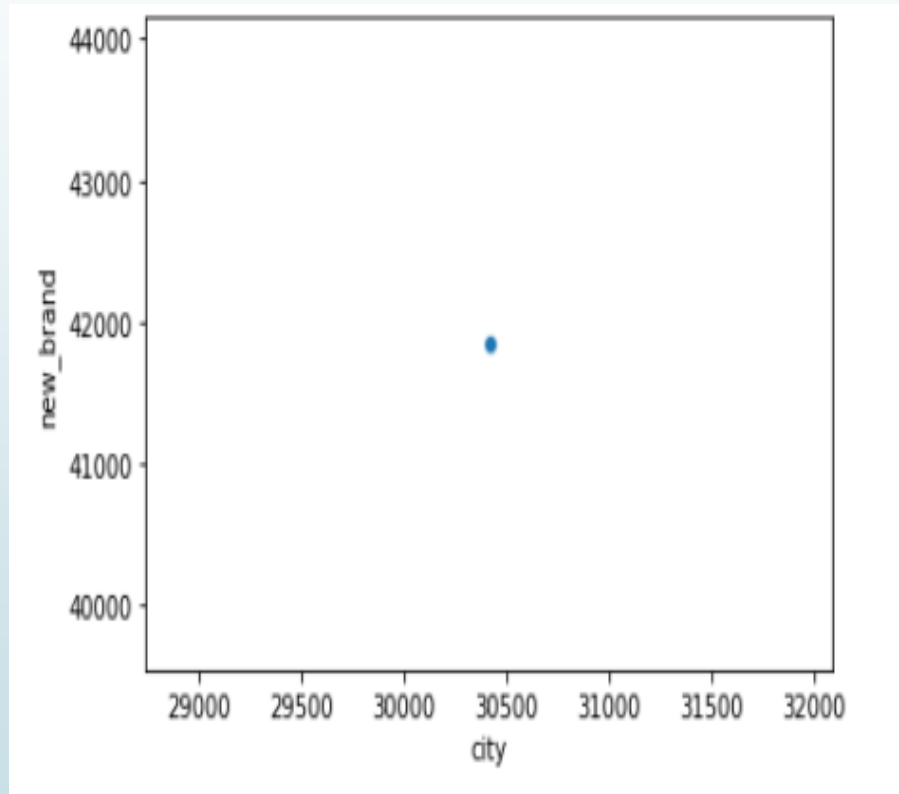
## Cluster 4



- From this plot we can conclude that block/locality of this cluster only contains one type of brand which is represented by 42000 in numerical format

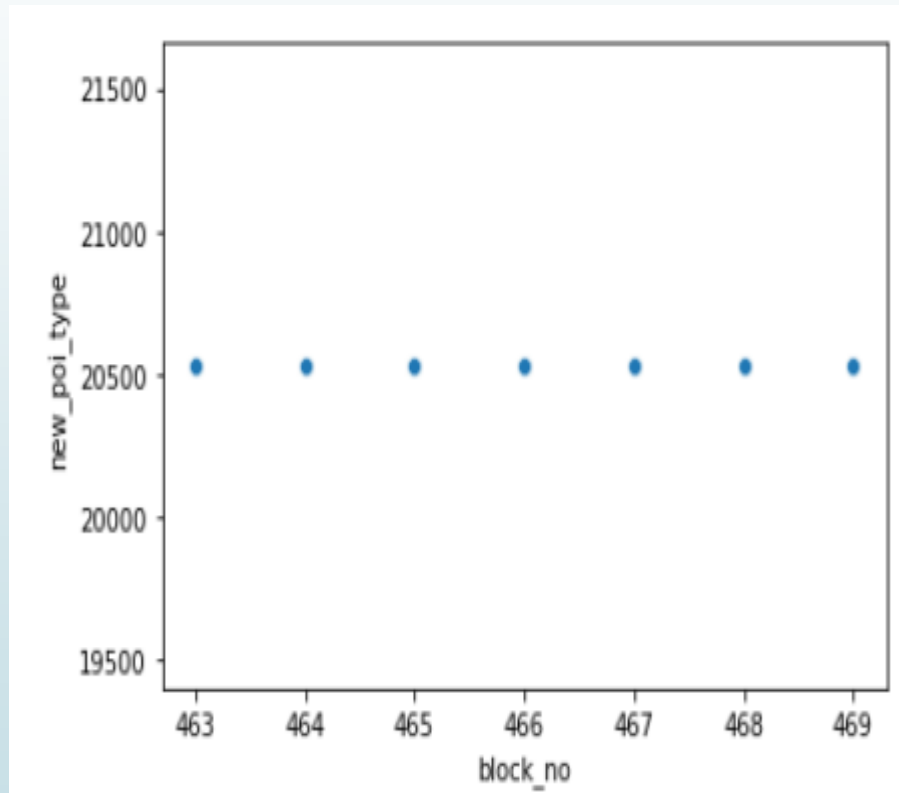


## Cluster 5



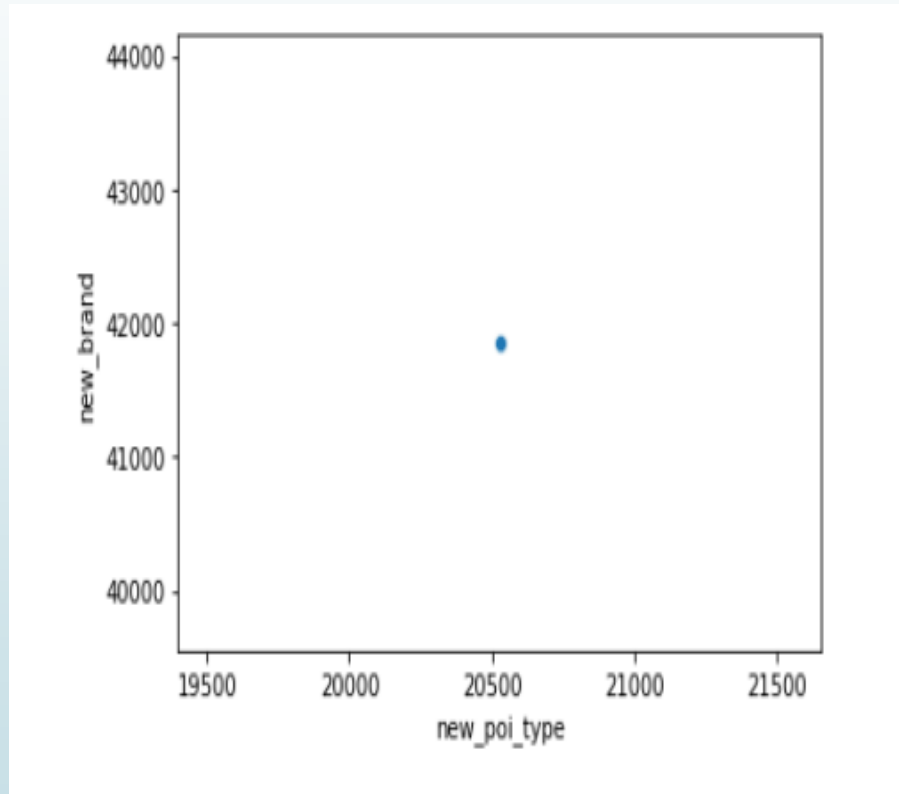
- From this plot we can conclude that block/locality of this cluster only contains one type of brand which is represented by 42000 in numerical format and this cluster belong to only 1 city represented by 30500

## Cluster 6



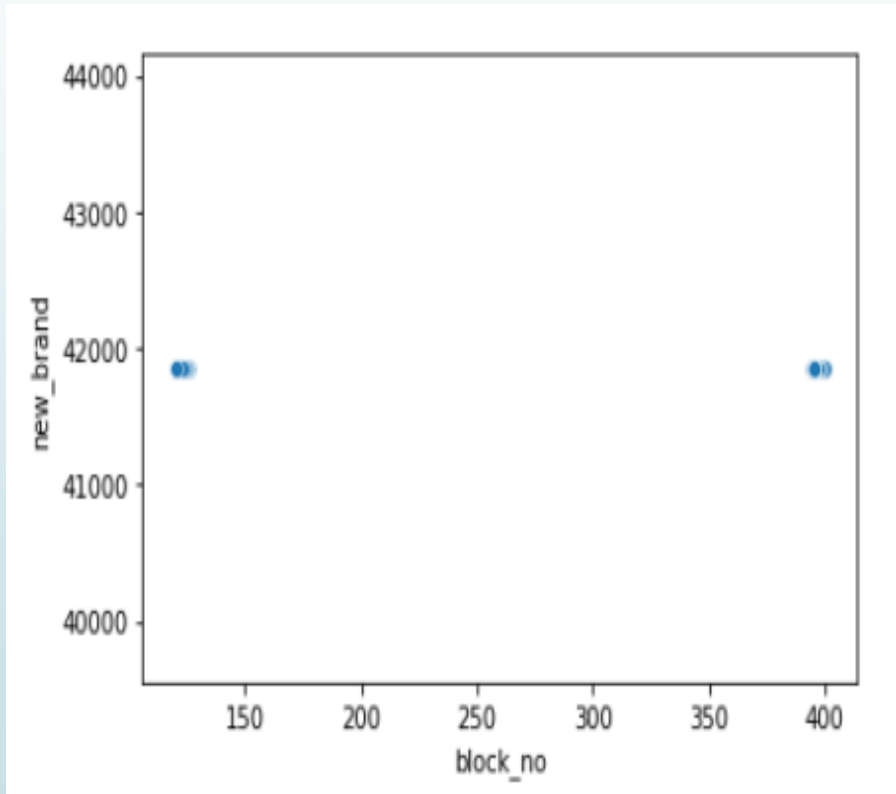
- From this plot we can conclude that block/locality of this cluster which is of small size ranging from 463 to 469 only contains one type of poi\_type which is represented by 20500 in numerical format

## Cluster 7



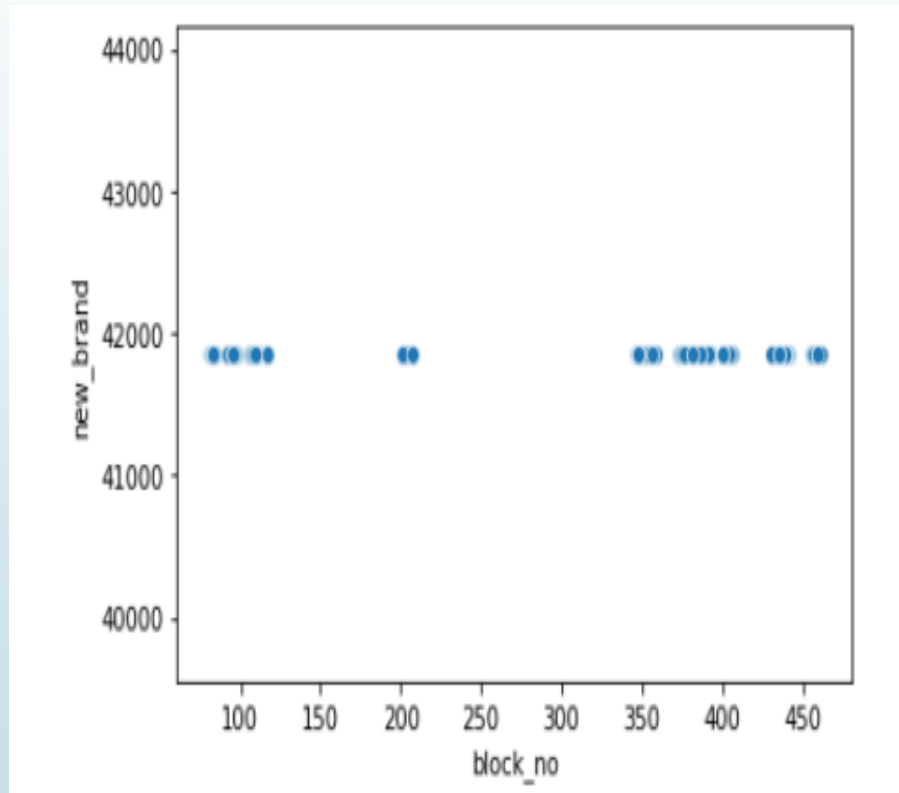
- From this plot we can conclude that this cluster has only 1 poi\_type which is represented by 20500 and 1 type of brand represented by 42000

## Cluster 8



- From this plot we can conclude that block/locality of this cluster does not contains any brand from range 150 to almost 400.

## Cluster 9



- From this plot we can conclude that block/locality of this cluster does not contains any brand from range 150 to around 200 and 210 to 350. And locality that do contain brand has only 1 type of brand represented by 42000



Thank You!