Clustering Analysis of Potential Locations for Virtual Hotel Business Expansion in South Jakarta, Indonesia

Capstone Project for IBM Data Science Professional Certificate

By Imam Hadrazi Baihaki

Introduction

Preface

- Low budget traveling and staycation has become a common needs for young people in Indonesia, especially in its capital city, Jakarta
- To accommodate this needs, several virtual hotel operators (e.g., OYO, RedDoorz, and Airy) have rapidly developed their business in this city.
- This type of accommodation has become a favorite choice amongst travelers because its simplicity, easiness to book, and lower price compared to other hotel operators.

Business Problem

- South Jakarta is a very promising area to develop this kind of business.
- There are plenty of popular restaurants, cafes, shopping malls, nightlife entertainments, and other interesting places in this area that would attract guests to stay in this area.
- Opening a new business unit in this area does not guarantee that it will be a pleasant place to stay for the guests.
- It must be opened in the right, strategic location so guests can have great experience while staying.

Guests Satisfactory Measures

- Google Map, a popular navigation app can accommodate its users to give to rate any places, including hotels.
- The rating value ranges from 0 to 5, where 0 means very unpleasant and 5 means excellent satisfactory.
- These ratings is considerable to be used as satisfactory measurement by hotel guests.

Project Goals

Develop a classification machine learning model to find promising location to open a new VHO franchise.

Create clusters of potential locations within South Jakarta area

Purpose

The result of this project is expected to bring insights for virtual hotel operators that interested in opening a new franchise in South Jakarta Area.

Several locations recommended in the result section can be used as starting points for further analyses prior to make a business decision.

Project Workflow

- Data acquisition
- Developing classification machine learning
- Generating new potential locations
- Create clusters of potential area

Data Acquisition

Required Data

Existing VHO franchises located within the area of observation.

Average Google rating for all existing VHO franchises.

Points of Interest located within the area of observation.

Data Sources

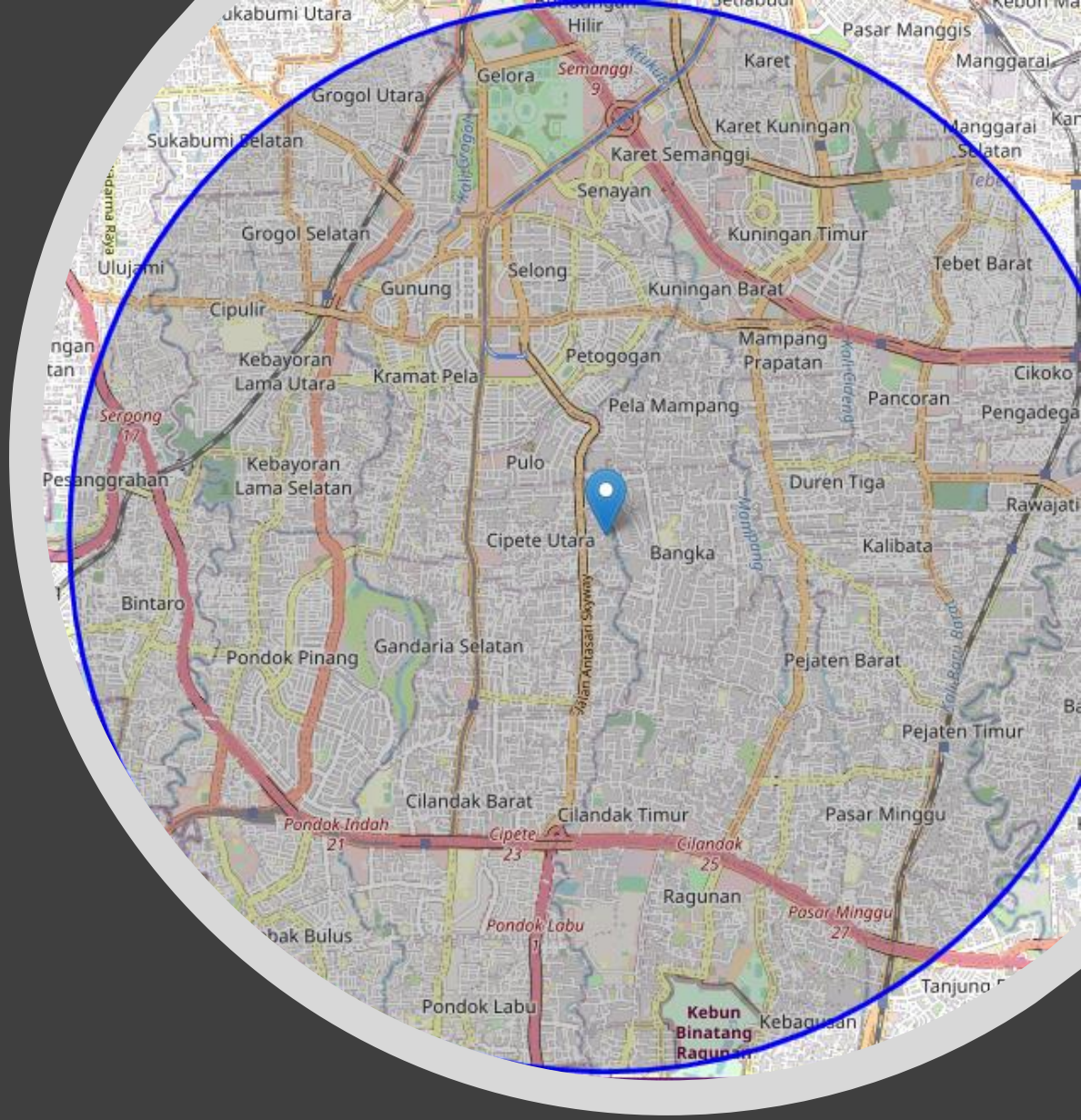
- Google Geocoding API (https://developers.google.com/maps/documentation/geocoding/overview)
- Google Places API (https://developers.google.com/places/web-service/overview)
- Here Map API (https://developer.here.com/)

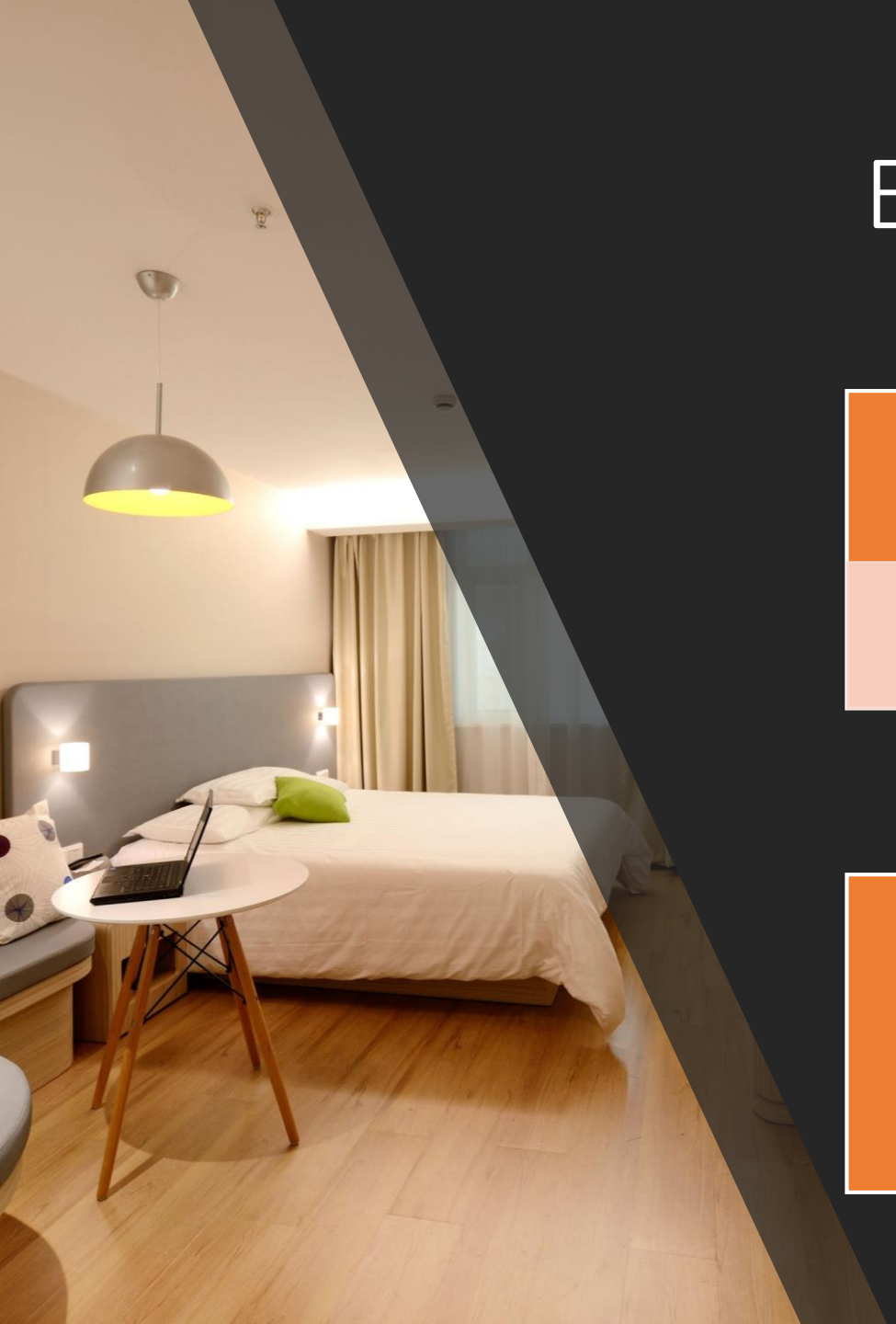
Area of Observation

Centre’s Latitude: -6.26149

Centre’s Longitude: 106.81059

Radius: 6000 meters

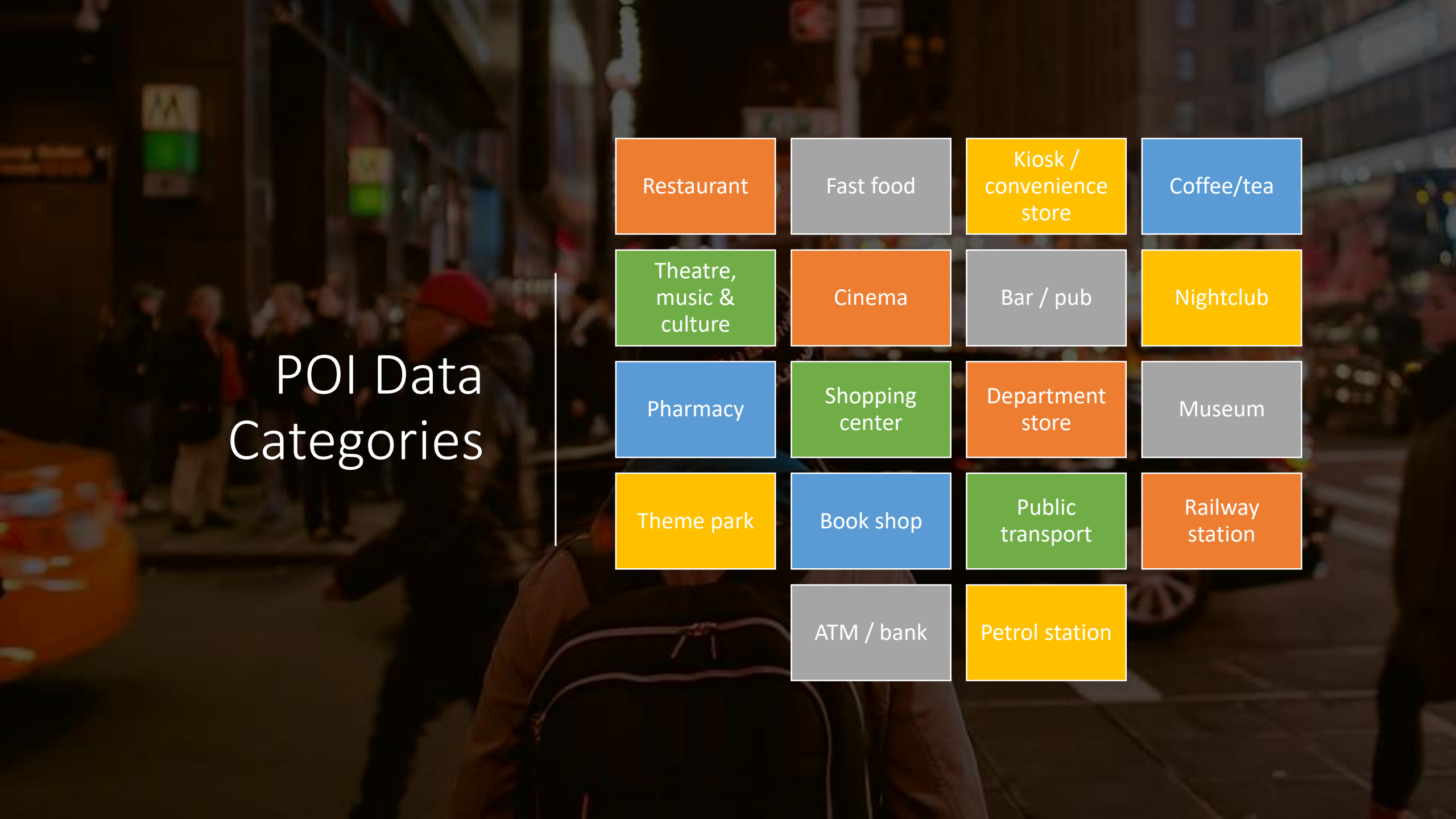




Extracted Hotels Data

Providers:		
Airy Rooms	OYO	RedDoorz

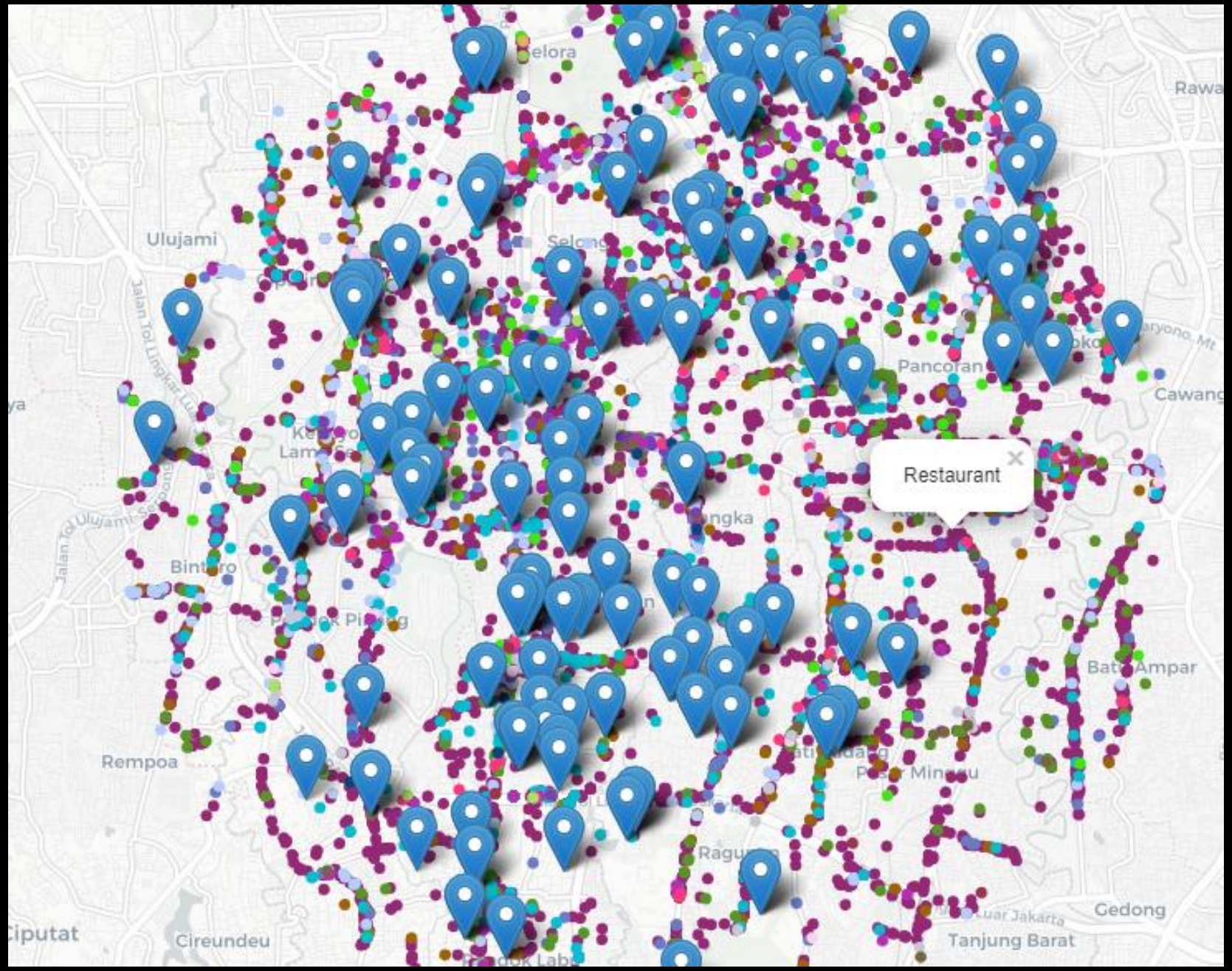
Number of hotels information: 275



POI Data Categories

Restaurant	Fast food	Kiosk / convenience store	Coffee/tea
Theatre, music & culture	Cinema	Bar / pub	Nightclub
Pharmacy	Shopping center	Department store	Museum
Theme park	Book shop	Public transport	Railway station
	ATM / bank	Petrol station	

Hotels and POIs Data Combined



Nearby POI Counts

- Each hotels will be given a 500 meters explore radius
- Within the radius, number of POIs will be counted, grouped by its category

Sample Data

Hostel ID	Hostel Latitude	Hostel Longitude	Hostel Rating	Hostel Total Rating	Restaurant	Snacks/Fast food	Convenience Store	ATM/Bank/ Exchange	Coffee/Tea	Theatre, Music & Culture
1	-6.2380528	106.8169531	3.6	218	106	1	3	5	10	0
2	-6.2341022	106.8171694	5	2	80	1	0	3	5	0
3	-6.2432143	106.7815005	4.5	20	88	3	7	13	9	2
4	-6.244997	106.78035	2.7	3	99	4	6	18	9	4
5	-6.2433616	106.7806216	3.2	45	92	3	7	13	8	2
6	-6.2886455	106.8003567	4.2	9	79	4	1	3	17	0
7	-6.2871132	106.8027785	4.4	72	14	0	1	0	3	0
8	-6.286095	106.799798	4.5	48	23	1	2	7	4	0
9	-6.2755727	106.7973817	3.8	40	92	17	2	21	11	4
10	-6.2754433	106.7982942	4.4	38	99	17	2	22	16	4

Developing Classification Machine Learning

Variables

EXPLANATORY VARIABLE:

NUMBERS OF POIS LOCATED NEARBY THE HOTEL

RESPONSE VARIABLE:

CUSTOMER'S SATISFACTION LEVEL IN A FORM OF GOOGLE RATING

Feature Engineering

- Convert rating value type from continuous to discrete, example:
  - 2.3 → 2.0 to 2.5
  - 3.1 → 3.0 to 3.5
  - 4.6 → 4.5 to 5.0
- Applying standard scaler to feature's values

Train and Test Dataset Split

TOTAL RECORDS:

201 DATA

TRAIN SET:

140 RECORDS

TEST SET:

61 RECORDS

Classification Algorithms and Accuracy Measurement

- Algorithms:
  - K-Means Clustering
  - Decision Tree
  - Support Vector Machine

- Accuracy Measurement:
  - Jaccard Similarity Index
  - F1 Score

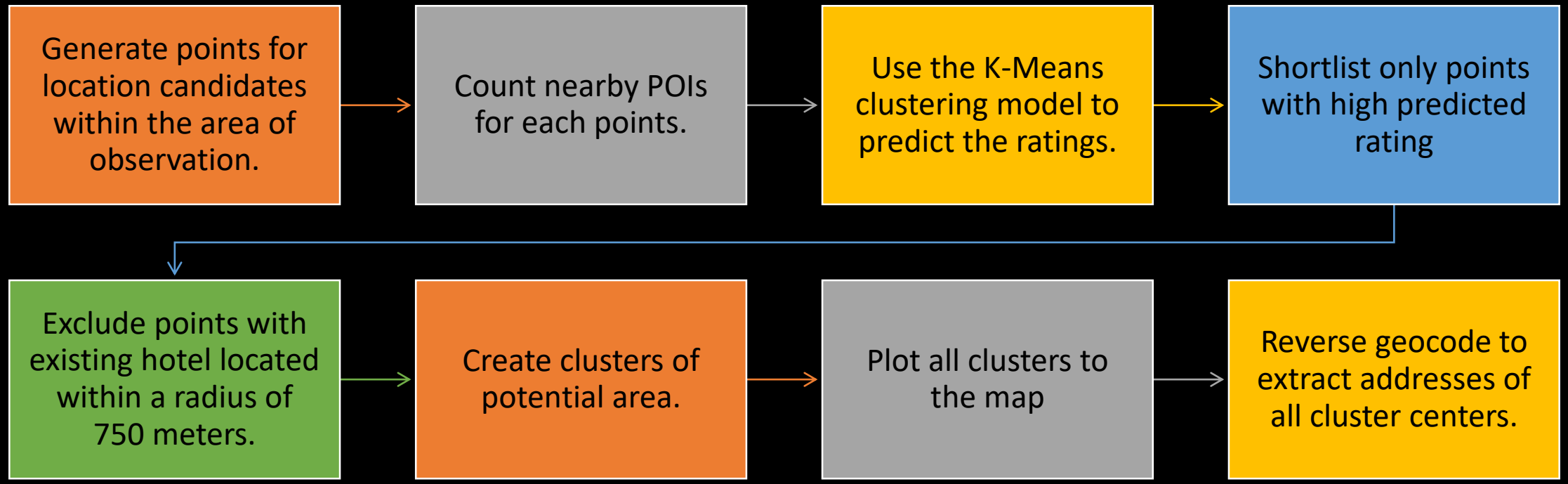
Model's Performances

- Table below shows the accuracy score for each algorithm
- Based on these result, K-Means clustering model will be used to find new potential locations

Algorithm	Jaccard Similarity Index	F1 Score
K-Means Clustering	0.806	0.892
Decision Tree	0.540	0.691
Support Vector Machine	0.539	0.697

Exploring New Potential Locations

Workflows

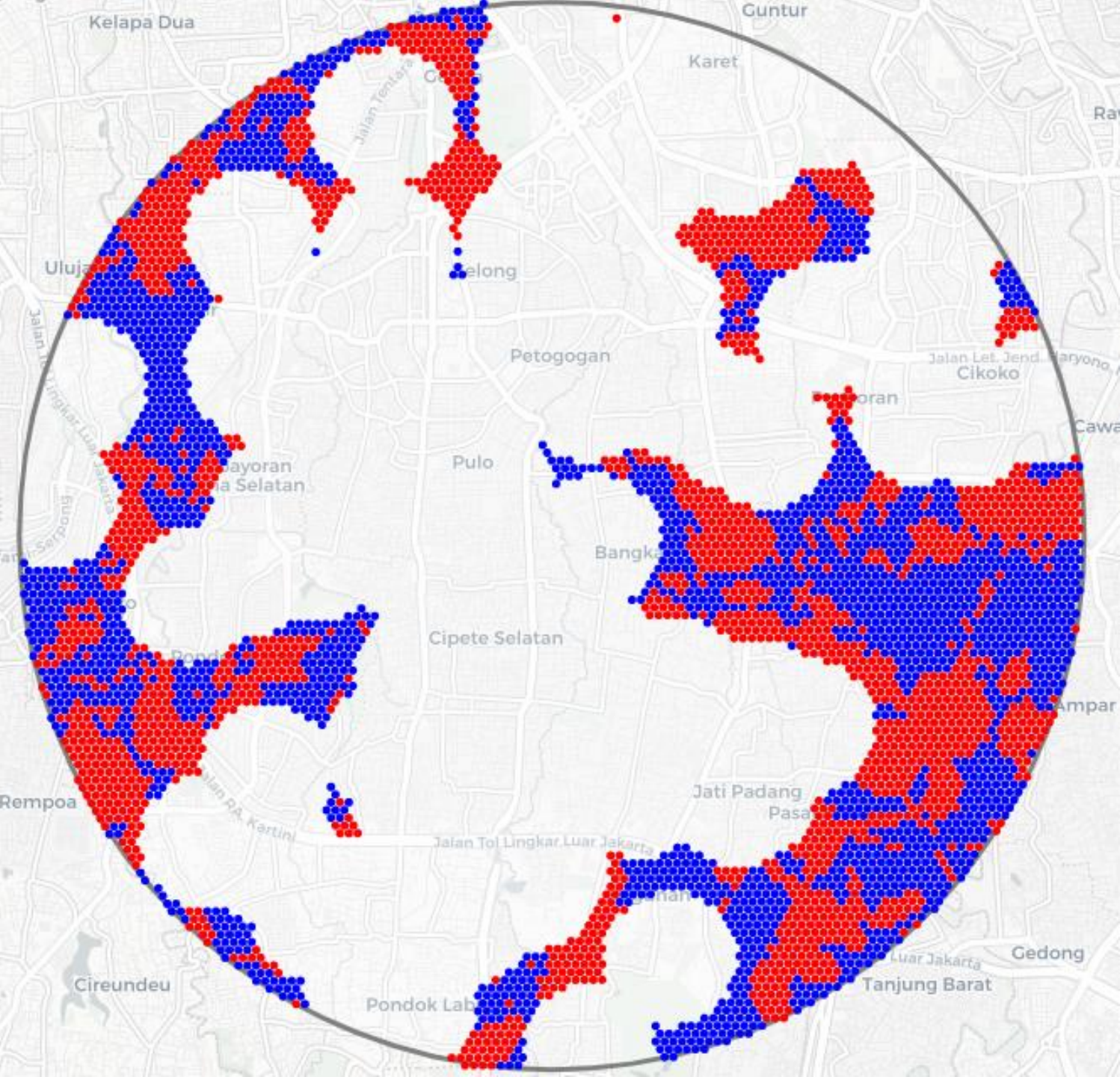

Generate points for location candidates within the area of observation.
Count nearby POIs for each points.
Use the K-Means clustering model to predict the ratings.
Shortlist only points with high predicted rating
Exclude points with existing hotel located within a radius of 750 meters.
Create clusters of potential area.
Plot all clusters to the map
Reverse geocode to extract addresses of all cluster centers.

Example for Locations with High Predicted Rating

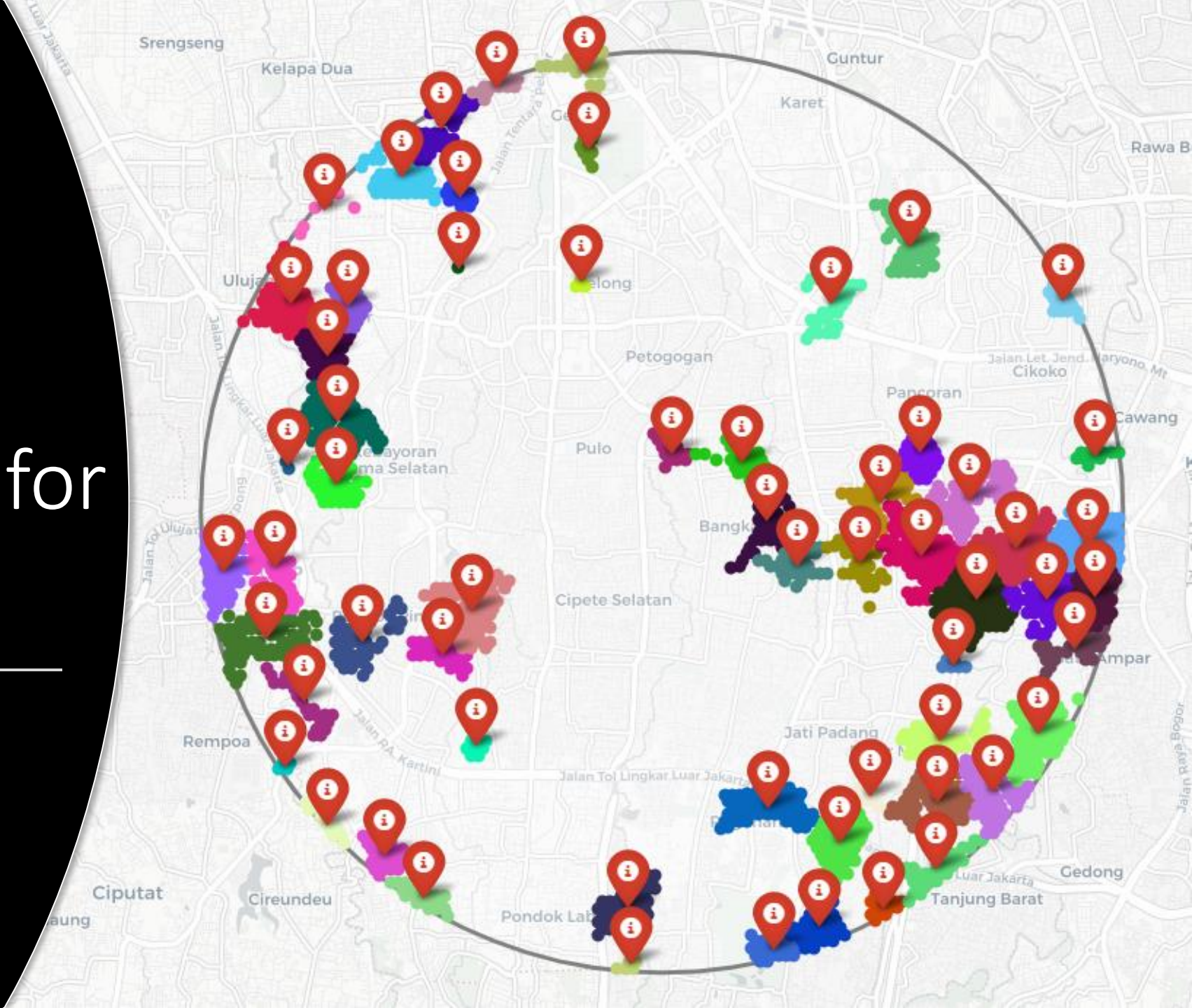
POINT ID	POINT LATITUDE	POINT LONGITUDE	POINT X	POINT Y	PREDICTED RATING
Point 1	-6.267378	106.756855	694363.269	-693086.6074	4.0 to 4.5
Point 2	-6.265812	106.75685	694363.269	-692913.4023	4.0 to 4.5
Point 3	-6.264245	106.756845	694363.269	-692740.1972	4.0 to 4.5
Point 10	-6.273639	106.75778	694463.269	-693779.4277	4.0 to 4.5
Point 11	-6.272073	106.757775	694463.269	-693606.2226	4.0 to 4.5

Plot for Potential and Non-Potential Points

Note:
Blue dots → High predicted rating
Red dots → Low predicted rating



Clusters of Potential Area for Development



Central Addresses for Potential Area (1 – 10 out of 60)

- H.Engkik, RT.1/RW.12, Bintaro, Kec. Pesanggrahan
- Gg. H. Nawi No.31, RT.12/RW.8, Pd. Pinang, Kec. Kby. Lama
- Jl. Kalibata Timur I No.76, RT.2/RW.8, Kalibata, Kec. Pancoran
- Jl. Perdagangan Blok Talib Maazis No.21D, RT.4/RW.7, Bintaro, Kec. Pesanggrahan
- Jl. Budaya Blok Haji Naiman No.51, RT.10/RW.6, Batu Ampar, Kec. Kramat jati
- Jl. Bukit Golf I Blok PD No.8, RT.2/RW.16, Pd. Pinang, Kec. Kby. Lama
- Jl. Gedung Hijau III No.2, RT.2/RW.13, Pd. Pinang, Kec. Kby. Lama
- Jl. TB Simatupang No.41, RT.2/RW.1, Cilandak, Kec. Jagakarsa
- Jl. Batu Kinyang 4 No.25, RT.12/RW.4, Batu Ampar, Kec. Kramat jati
- Jl. Tj. Barat Raya No.3, RW.7, Ps. Minggu, Kec. Ps. Minggu

Central Addresses for Potential Area (11 – 20 out of 60)

- Jl. Gunuk Raya No.35, RT.11/RW.11, Pejaten Tim., Kec. Ps. Minggu
- Jl. H. Minun II Jalan Haji Minun No.2, RT.6/RW.6, Bintaro
- Jl. Kemang Utara VI No.37, RT.13/RW.4, Bangka, Kec. Mampang Prpt.
- Jl. Komp. Bangdes No.7, RT.7/RW.1, Pejaten Bar., Kec. Ps. Minggu
- Jl. Kebagusan III No.2, RT.7/RW.5, Kebagusan, Kec. Ps. Minggu
- Jl. Bunga Mayang II No.21, RT.2/RW.1, Bintaro, Kec. Pesanggrahan
- Gg. Melati No.11, RT.7/RW.3, Bangka, Kec. Mampang Prpt.
- JL Dermaga No. 1, Muara Angke, North Jakarta, RT.10/RW.3, Bale Kambang, Kramat Jati
- Jl. Salak Blok M No.1, RT.9/RW.6, Rawajati, Kec. Pancoran, Kota Jakarta Selatan
- Gg. Komp. Koperasi II No.72, RW.9, Kalibata, Kec. Pancoran, Kota Jakarta Selatan

Central Addresses for Potential Area (21 – 30 out of 60)

- Jl. Delman Utama No.26-29, RT.2/RW.10, Kby. Lama Sel., Kebayoran Lama
- Jl. Bangka XII No.17A, RT.3/RW.7, Pela Mampang, Kec. Mampang Prpt.
- Jl. Bumi Asih I No.a2/16, RT.13/RW.3, Lb. Bulus, Kec. Cilandak
- Jl. Kalibata Utara II No.116, RT.11/RW.2, Kalibata, Kec. Pancoran
- Jl. Flamboyan Gg. I No.27, RT.2/RW.10, Menteng Dalam, Kec. Tebet
- Jl. Kb. Mangga I No.8, RT.4/RW.2, Grogol Utara, Kec. Kby. Lama
- Jl. Gatot Subroto No.Kav 51, RT.5/RW.4, Kuningan Tim., Kecamatan Setiabudi
- Jl. H. Marzuki No.11, RT.2/RW.1, Kebagusan, Kec. Ps. Minggu
- Lemigas, Cipulir, Kec. Kby. Lama, Kota Jakarta Selatan
- Jl. Tj. Sanyang No.11, RT.8/RW.3, Cawang, Kec. Kramat jati

Central Addresses for Potential Area (31 – 40 out of 60)

- Pulo Melati 6 No.9, RT.2/RW.16, Grogol Utara, Kec. Kby. Lama
- Jl. Poncol Lestari No.59, RT.8/RW.2, Lb. Bulus, Kec. Cilandak
- Jl. Cipulir Permai No.3, RT.6/RW.12, Grogol Sel., Kec. Kby. Lama
- RT.1/RW.3, Gelora, Kecamatan Tanah Abang
- Jl. Bango II No.78, RT.6/RW.3, Pd. Labu, Kec. Cilandak
- Jl. TB Simatupang, RT.10/RW.4, Tj. Bar., Kec. Jagakarsa
- Jl. Lkr. Gelora No.1, RT.1/RW.3, Gelora, Kecamatan Tanah Abang
- Jl. H. Batong 1 No.36, RT.2/RW.6, Cilandak Bar., Kec. Cilandak
- Gg. H.Riin No.12A, RT.1/RW.3, Duren Tiga, Kec. Pancoran
- Jl. Ir H. Juanda No.56, Rempoa, Kec. Ciputat Tim.

Central Addresses for Potential Area (41 – 50 out of 60)

- Jl. Paninggaran Tim. I No.65, RT.11/RW.6, Cipulir, Kec. Kby. Lama
- Jl. Poltangan Raya Blok Delima No.41, RT.1/RW.4, Tj. Bar., Kec. Jagakarsa
- Jl. Ruby No.34, RT.9/RW.13, Grogol Utara, Kec. Kby. Lama
- Jl. Pattimura No.29, RT.4/RW.1, Selong, Kec. Kby. Baru
- Jl. Kong ali rt. 1 rw. 3 no 48, Kebagusan, Pasarminggu, RT.2/RW.3, Kebagusan, Kec. Ps. Minggu
- Jl. Simprug Garden II h No.12, RT.2/RW.2, Grogol Sel., Kec. Kby. Lama
- Jl. Panjang No.32, RT.10/RW.9, Grogol Utara, Kec. Kby. Lama
- Jl. Tebet Timur Dalam VII No.1, RT.5/RW.6, Tebet Tim., Kec. Tebet
- Jl. Raya Condet No.34, RT.9/RW.6, Cililitan, Kec. Kramat jati
- Jl. Panjang Kp. Baru No.2, RT.2/RW.3, Grogol Sel., Kec. Kby. Lama

Central Addresses for Potential Area (51 – 60 out of 60)

- Jl. Panjang Kp. Baru No.2, RT.2/RW.3, Grogol Sel., Kec. Kby. Lama
- Jl. Cilandak 6 No.35, RT.4/RW.4, Lb. Bulus, Kec. Cilandak
- Jalan Damai Musawaroh No.39, RT.2/RW.3, Pd. Labu, Kec. Cilandak
- Jl. Kebagusan Raya No.48, RT.10/RW.7, Ragunan, Kec. Ps. Minggu
- Kencana Permai III No.15, RT.2/RW.15, Pd. Pinang, Kec. Kby. Lama
- Jl. Kemuning Dalam I No.89, RT.5/RW.6, Pejaten Tim., Kec. Ps. Minggu
- Jl. Batu Ampar II No.97, RT.12/RW.3, Batu Ampar, Kec. Kramat jati
- Jl. Cipulir II No.28, RT.4/RW.4, Cipulir, Kec. Kby. Lama
- Jl. Pejaten Mas I Blok I No.H-06, RT.4/RW.2, Ps. Minggu, Kec. Ps. Minggu
- Jl. H. Samali No.39A, RT.4/RW.4, Kalibata, Kec. Pancoran
- Jl. Aup Bar. I No.30, RW.6, Jati Padang, Kec. Ps. Minggu

Results and Conclusion

Along the project, we have done these things:

EXTRACTING LOCATION INFORMATION OF EXISTING VIRTUAL HOTEL OPERATOR FRANCHISE WITHIN THE AREA OF OBSERVATION

EXTRACTING VARIOUS CATEGORIES OF POINTS OF INTEREST THAT LOCATED WITHIN THE AREA OF OBSERVATION.

CREATING A MACHINE LEARNING MODEL TO PREDICT HOTEL'S RATING BASED ON THE NUMBERS OF VARIOUS CATEGORY OF POIS NEARBY.

USING THE MACHINE LEARNING MODEL TO ANALYZE PROMISING LOCATION FOR OPENING A NEW VIRTUAL HOTEL FRANCHISE WHEREAS NO EXISTING FRANCHISE IS LOCATED NEARBY.

CREATE CLUSTERS OF POTENTIAL DEVELOPMENT AREA AND EXTRACTING ADDRESS FOR EACH CLUSTER'S CENTER