CAPSTONE PROJECT- THE BATTLE OF NEIGHBORHOODS

HEMAJA PATOJU

1. INTRODUCTION

- Since the beginning of 2020, Jakarta and many other cities around the world have been under attack by an invisible army called ‘Novel Corona Virus’, also known as ‘Covid-19’.
- Every effort has been focusing on solving or minimizing problems, including Data Scientists.
- Data Scientists assessed the situations in places around the world, such as availability, amount, and geographical distribution (i.e. locations) of health infrastructures, such as virus testing centers and authorized hospitals to treat affected patients.
- I would like to present a simple analysis for determining strategic locations for the distribution of masks and medical devices for COVID-19 treatment, based on confirmed cases on May 28, 2020, and the red zone areas for “new normal” condition analysis.

2. DATA ACQUISITION AND CLEANING

- **A few Identified factors that influence our decision are:**
  - Covid-19 cases per district “Riwayat File Covid-19 DKI Jakarta”
  - Total population in DKI Jakarta 2020 statistik.jakarta.go.id
  - 10 most population in DKI Jakarta 2020 per district statistik.jakarta.go.id
  - Hospital for treatment covid-19 megapolitan.kompas.com

- ***The following data sources are needed to extract/generate the required information:***
  - Processed covid-19 positive case data collection on 28 May 2020 At 09.00.
  - The distribution of mask sales based on the population in the DKI Jakarta area.
  - The distribution of mask sales based on 5 districts with the most densely populated populations.
  - New datasets (to be created) from Hospital table that contains city, district, along with their latitudes and longitudes.

DATASETS

- Covid-19 cases per district data is collected from https://raw.githubusercontent.com/cahyati/Coursera_Capstone/master/Standar%20Keluraha n%20Data%20Corona%20(28%20MEI%202020%20Pukul%2009.00).csv")
- Total population data in DKI Jakarta 2020 data is collected from https://raw.githubusercontent.com/cahyati/Coursera_Capstone/master/population2020_DKI_ Jakarta.csv
- Data of covid hospitals in Jakarta is collected from: https://raw.githubusercontent.com/cahyati/Coursera_Capstone/master/Hospital%20for%20tre atment%20covid-19.csv

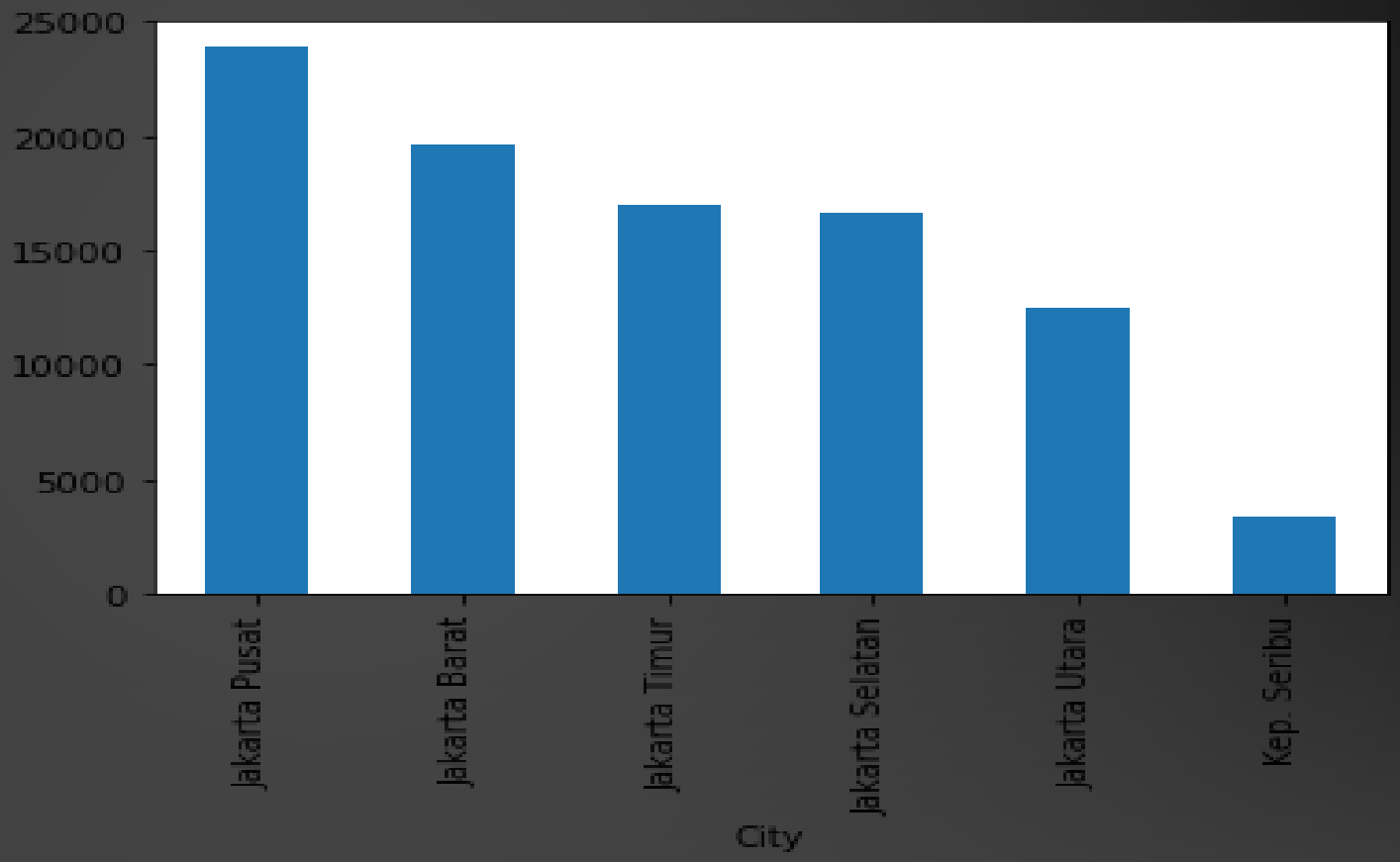
3. METHODOLOGY

- Using OpenCageGeocode, retrieved an API key to get the latitude and longitude positions from the address and vice versa.
- Then, we retrieved the latitude and longitude of all Covid-19 testing centers in Jakarta that we have checked from the source.
- Foursquare credentials to get the nearby areas of a given latitude and longitude value with the query.

4. RESULTS

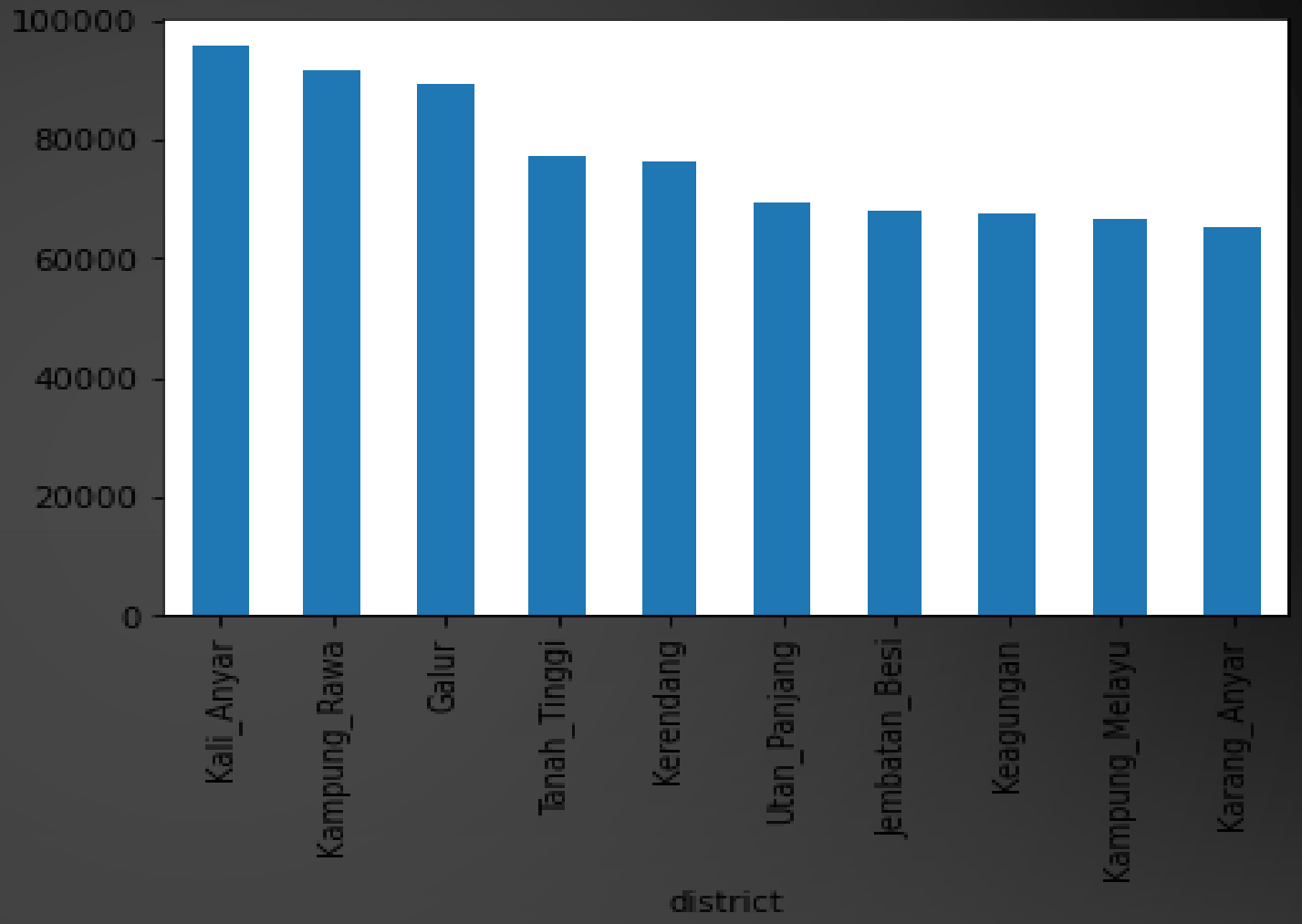
Population density in Jakarta :

Based on the graph results shown that areas need the distribution of masks the most is Central Jakarta (Jakarta Pusat) with the most populated areas

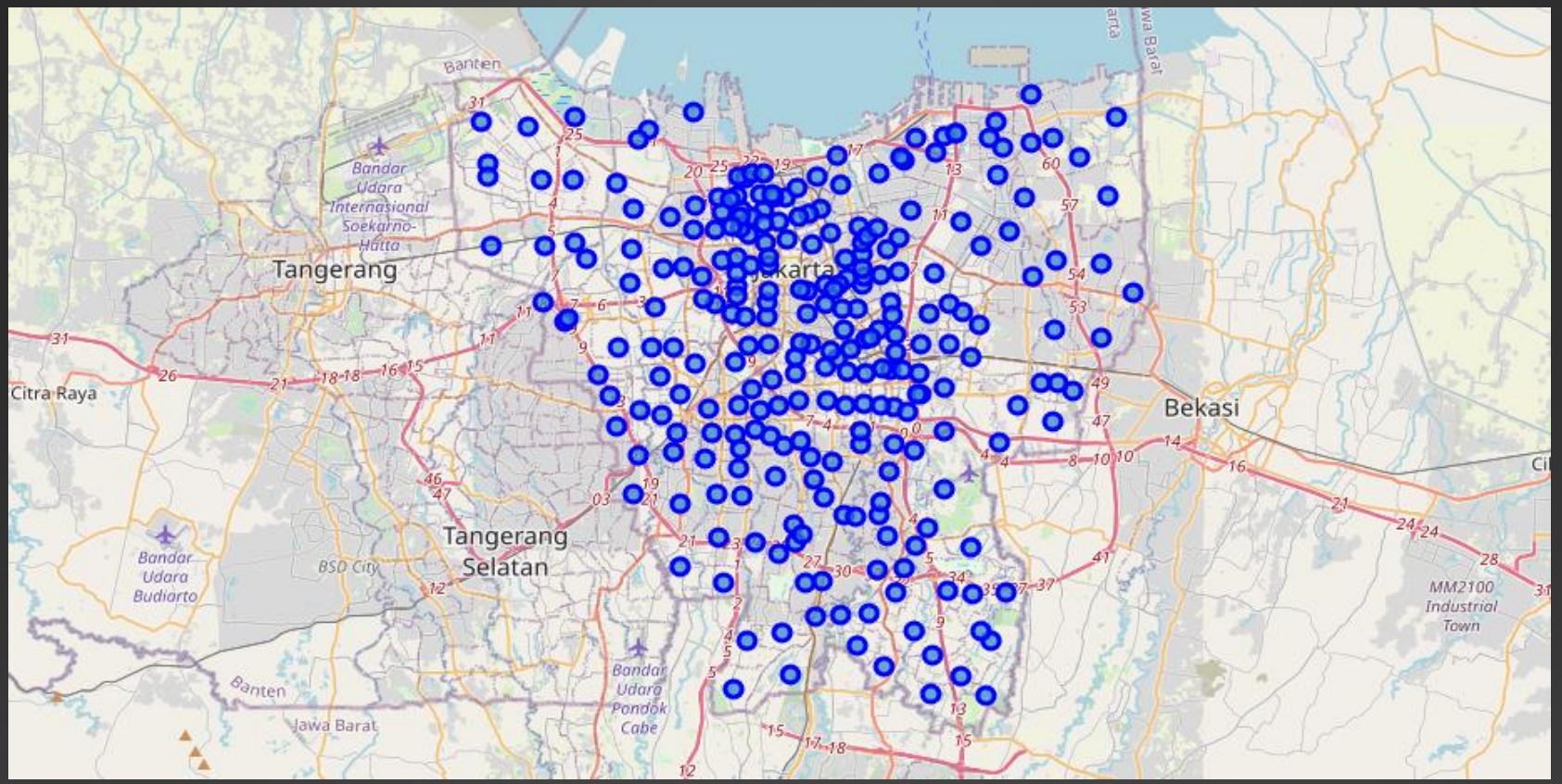


The population density in Jakarta, per district:

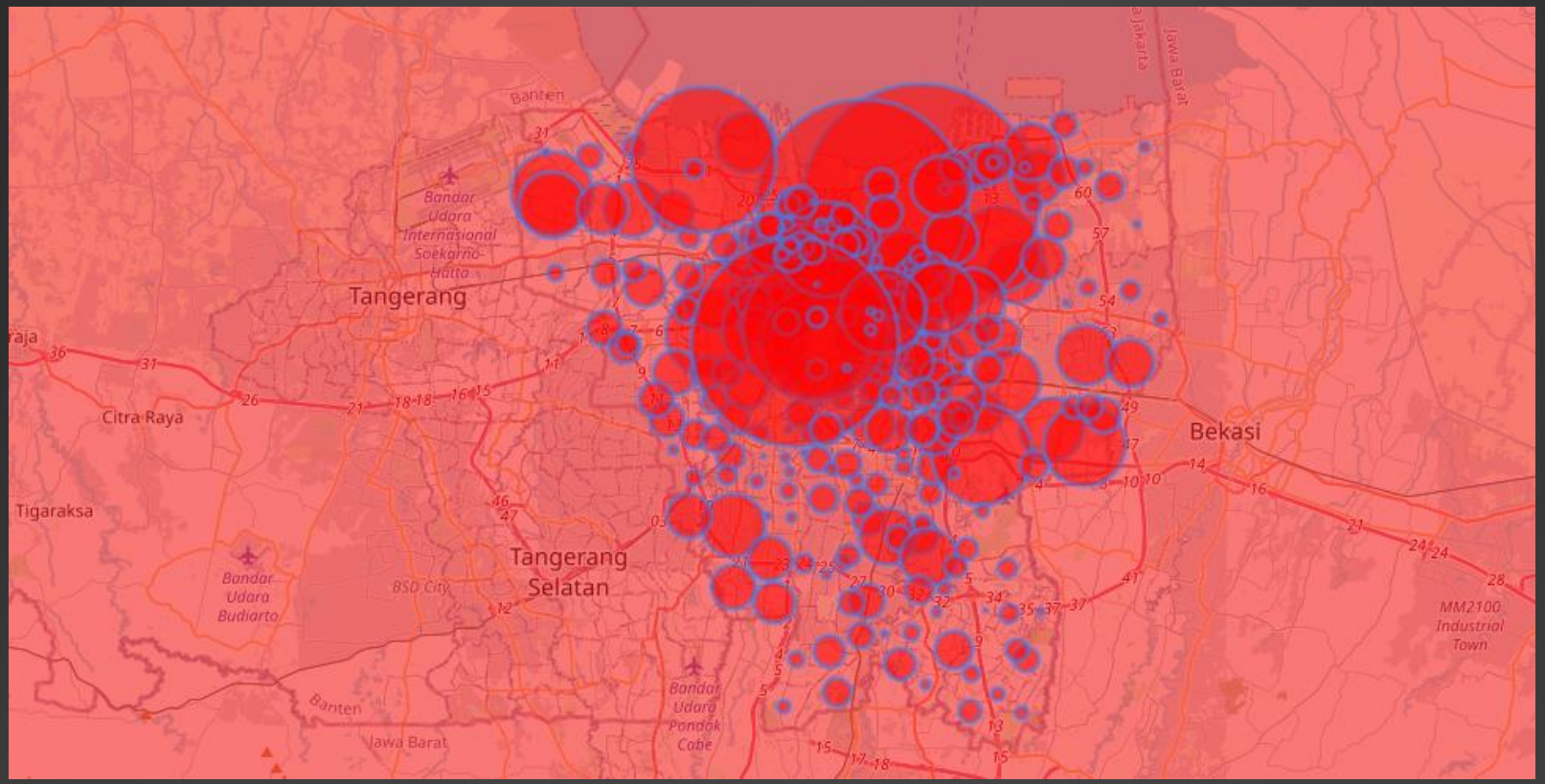
5 districts that mostly need for a distribution of masks are Kali Anyar, Kampung Rawa, Galur, Tanah Tinggi, and Kerendang.



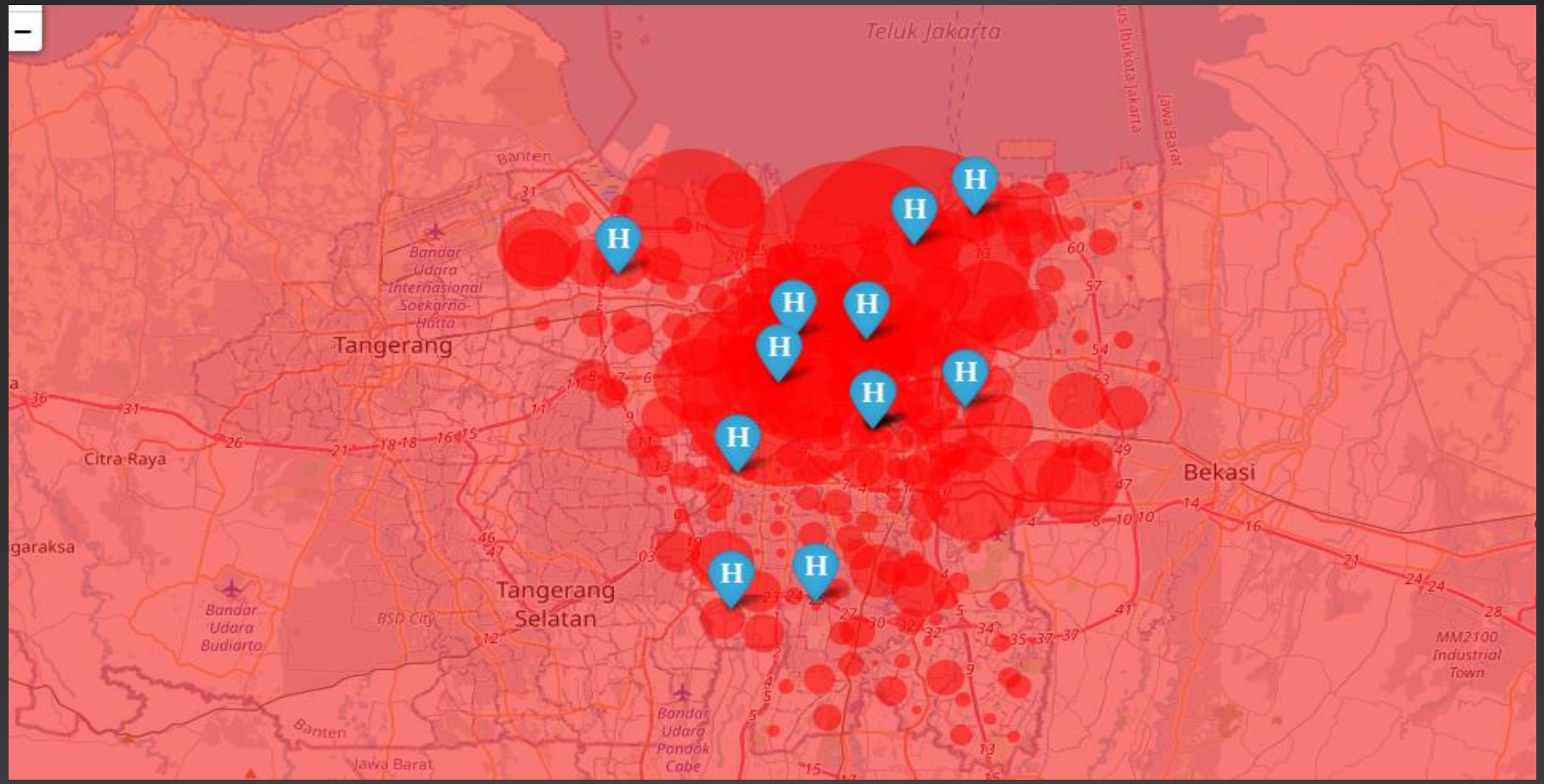
- We can see from the above map, most of the districts are within the main outer ring roads surrounding the city, and others are situated outside the main ring roads.



- We can see that most of the regions in Jakarta are now in the ‘RED’ zone, with the radius of the circle represent the relative extent of Covid-19 distribution in the City of Jakarta



- Finally, We can see from the results of the distribution of COVID-19 cases and the location of hospitals, almost all hospitals require a lot of medical equipment for COVID-19 treatment.



5. DISCUSSION

- Based on the results generated by the FOURSQUARE API, we can locate the business site around Tarakan hospital and identify affected business locations in the red zone.
- To simplify our analysis, we will just use the Euclidian (distance-based) clustering technique which is part of the unsupervised machine learning technique.
- In particular, we will use K-means clustering. And with the available data by drawing the curve we find out the best k.
- The result of analysis is the location of the business which is in the Tarakan hospital neighborhood and is within a radius of 500 meters.

- we have divided the data into 3 clusters randomly and assigned the cluster numbers using K means Clustering and different colors accordingly.
- This map shows us the clustered areas of the TARAKA hospital neighborhood

- The project aims to provide information to local people who must be alerted to go out of the house from the distribution of the COVID-19 case in Jakarta.
- It also aims to provide information on areas that are most needed for a lot of mask distribution, according to population density in the area.
- Further, it provides information on which hospitals that need the most medical equipment's for COVID-19 treatment, possibly even additional medical personnel's (doctors and nurses).
- It also provides information on the business neighborhood which shall implement Covid-19 health protocol with a high discipline when “new normal” comes

6.CONCLUSION

- This project helps mask sellers to understand potential distribution areas according to population density in Jakarta.
- It also helps the distribution of medical devices for corona care to hospitals that are estimated to have a large number of patients or even helps analyzing which hospitals need additional medical personnel (doctors and nurses).
- It will also provide awareness to help business owners who run businesses surrounding the adjacent clusters to be better informed, with the density of people within the business neighborhood.

THANKYOU

HEMAJA PATOJU