

# Capstone Project

**Applied Data Science Capstone by IBM/Coursera**

Han, Donghee

# Table of contents

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

## Introduction : Business Problem

find an optimal location for delivery service company

This project is targeting Korean student in KAIST(Korea Advanced Institute of Science and Technology)

We are trying to make 'Drone Delivery Service'

looking for the optimal location for our 'Drone Base'

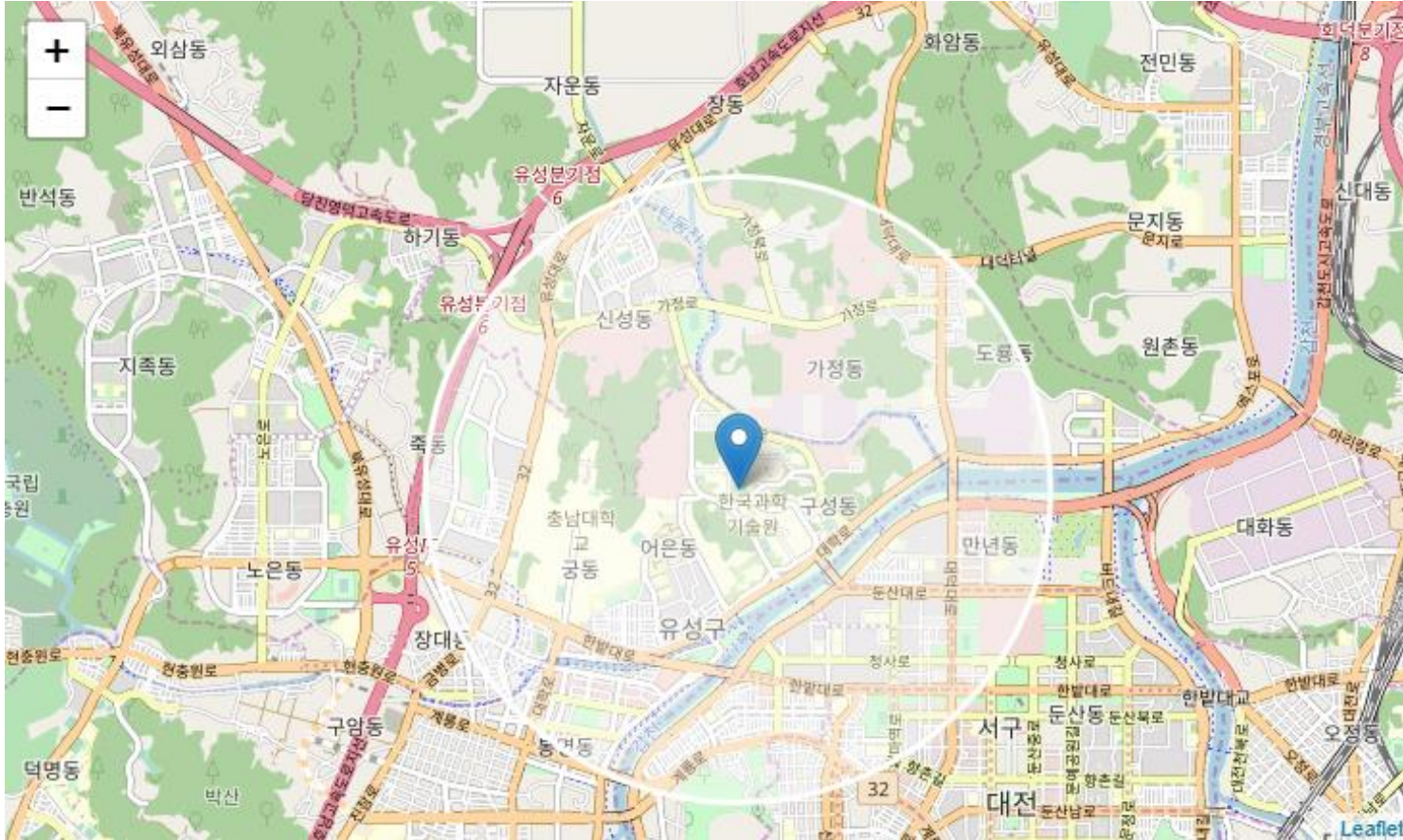
## Data

- number and location of restaurants clusters near KAIST
- distance of clusters from KAIST
- number of Chinese restaurants in the clusters (Chinese food is most popular delivery food)

Get Data from Google Maps API & Foursquare API

## Data

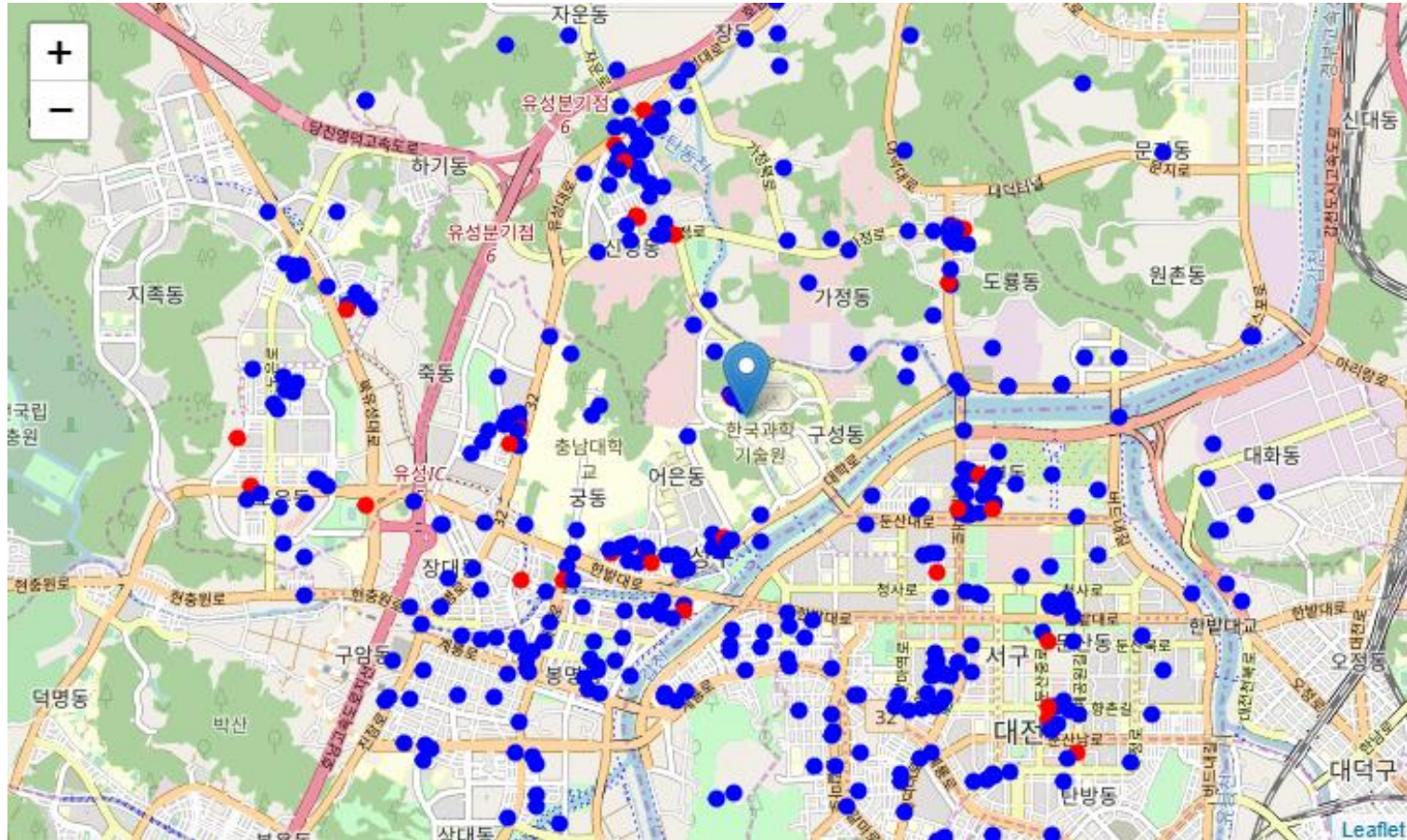
Get KAIST Location Data from Google Maps API





## Data

Get Restaurant Location Data near KAIST from Foursquare API

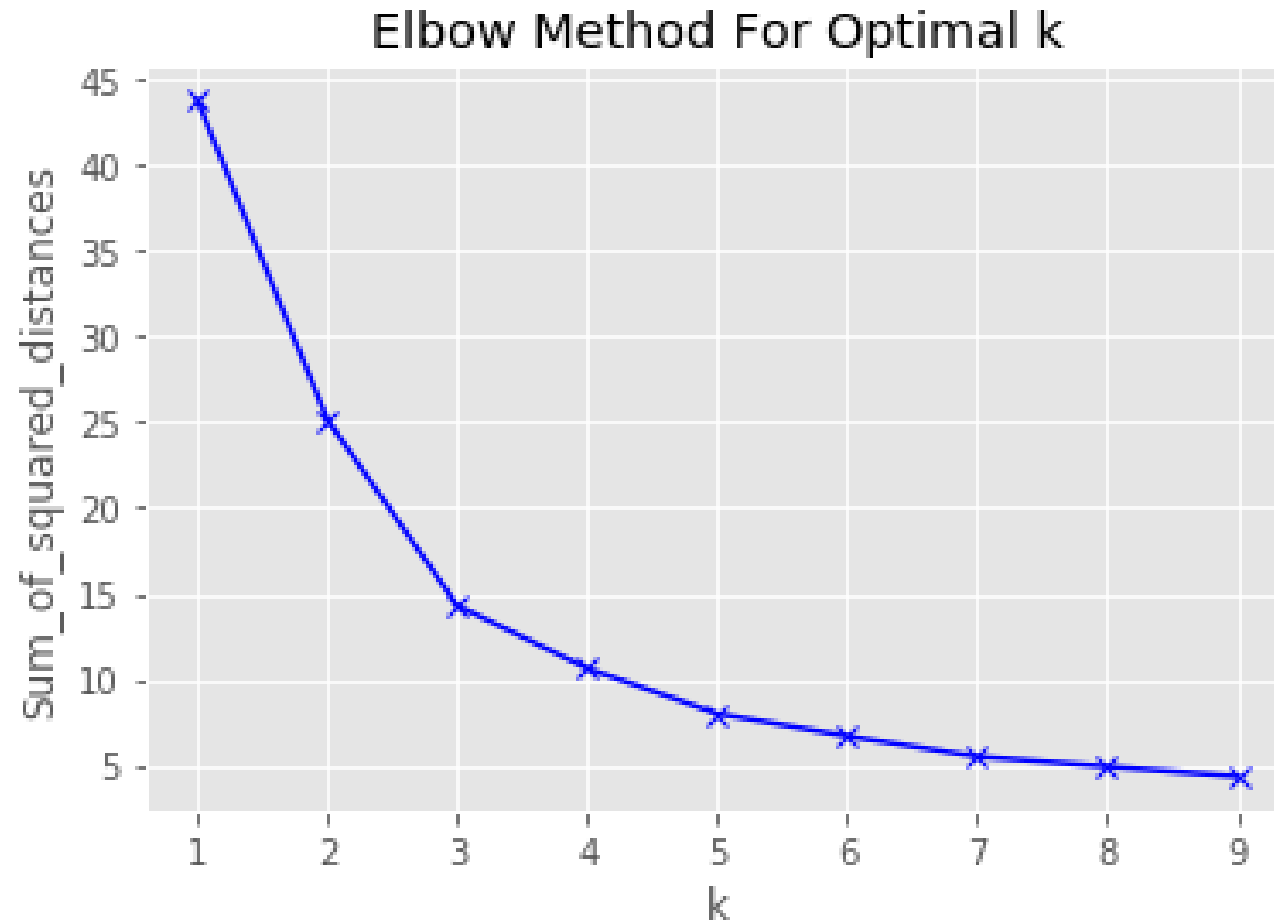


## Methodology

1. We will use k-means & DBSCAN to find clusters of restaurants
2. Using visualization optimize clustering.
3. find center location of the clusters and number of Chinese restaurants in the clusters
4. find center location of the clusters and number of Chinese restaurants in the clusters
5. select important clusters
6. find best location for 'Drone Base' based on the distances from the clusters.

## Analysis : k-Means

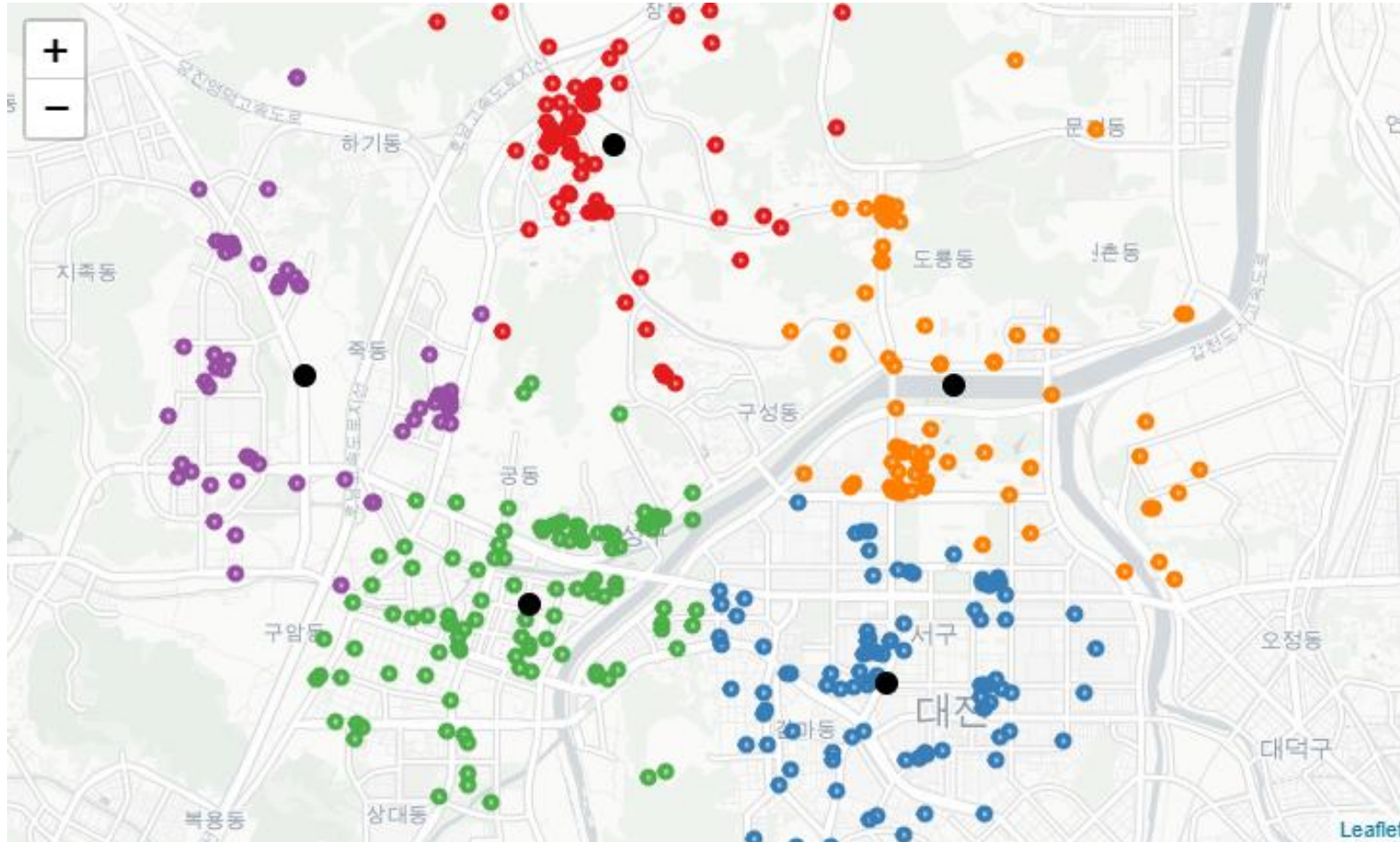
Use Elbow Method for Optimal k. From this we can find the best k is 5.





## Analysis : k-Means

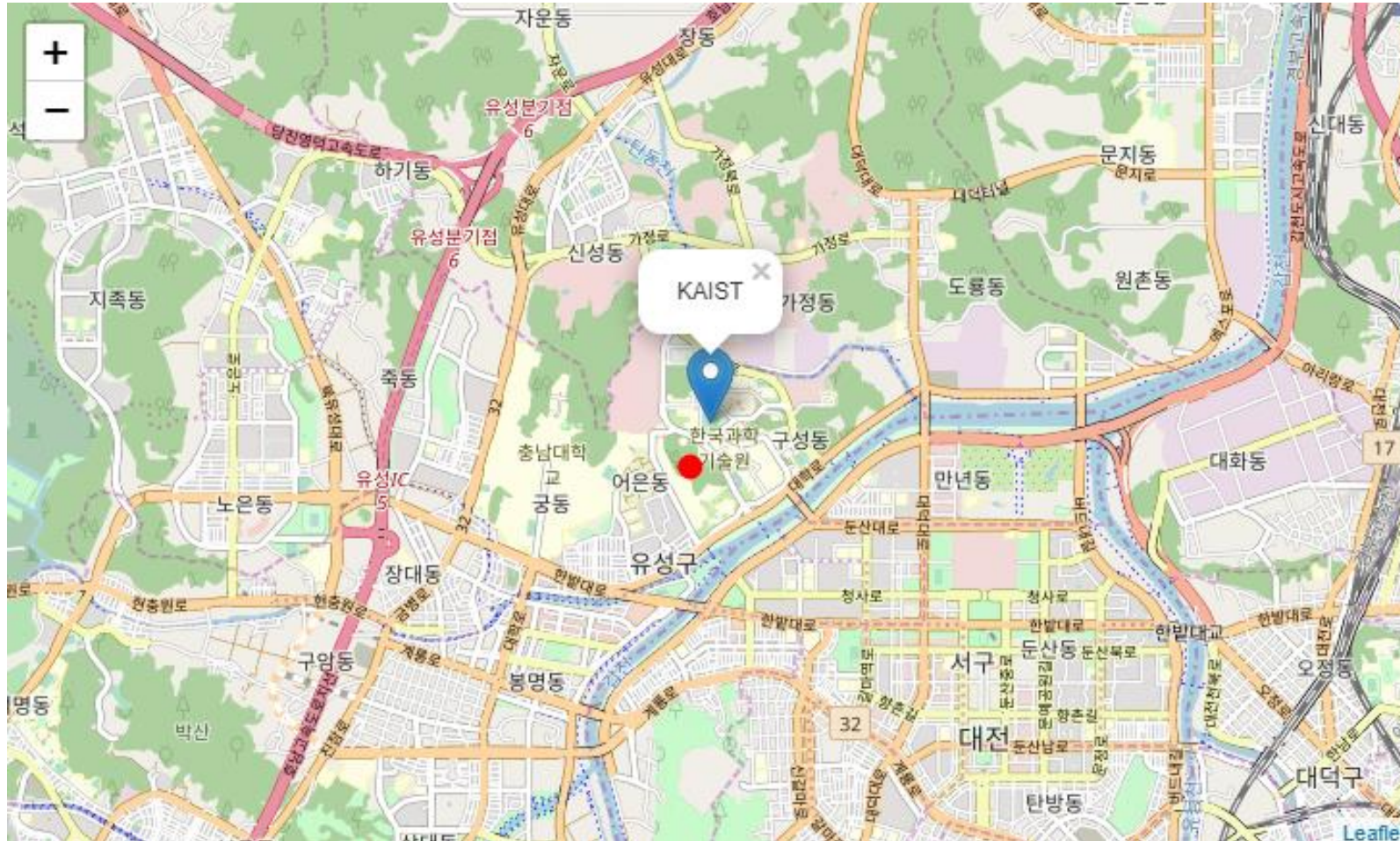
Draw clusters on the map and check the number of Chinese restaurants in each clusters.



```
num of Chinese restaurants in 0 : 9  
num of Chinese restaurants in 1 : 11  
num of Chinese restaurants in 2 : 11  
num of Chinese restaurants in 3 : 11  
num of Chinese restaurants in 4 : 11
```

## Analysis : k-Means

From the location data of each clusters, we can get center of the centers of the clusters.

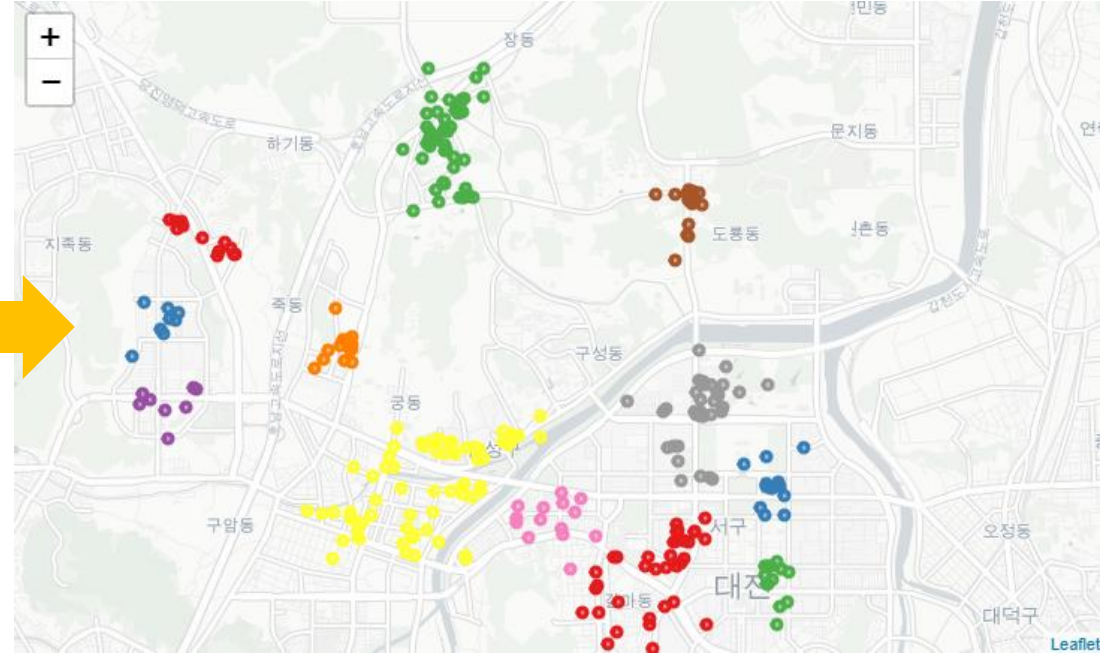
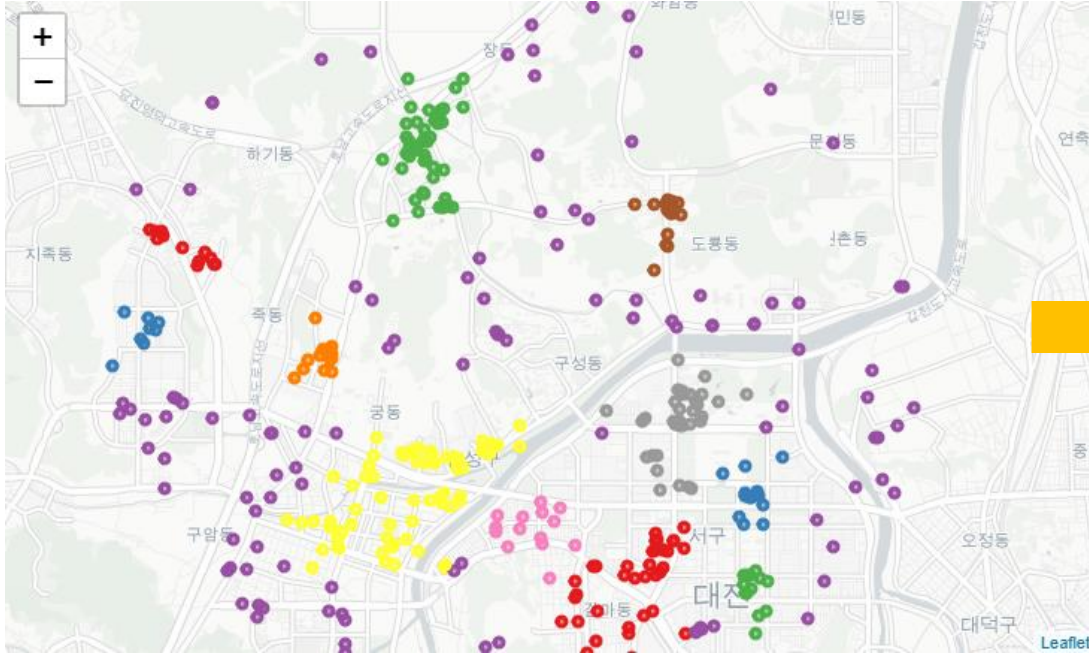




## Analysis : DBSCAN

*DBSCAN* from *sklearn.cluster*

**eps = 0.22, min\_samples = 8, algorithm='kd\_tree'**

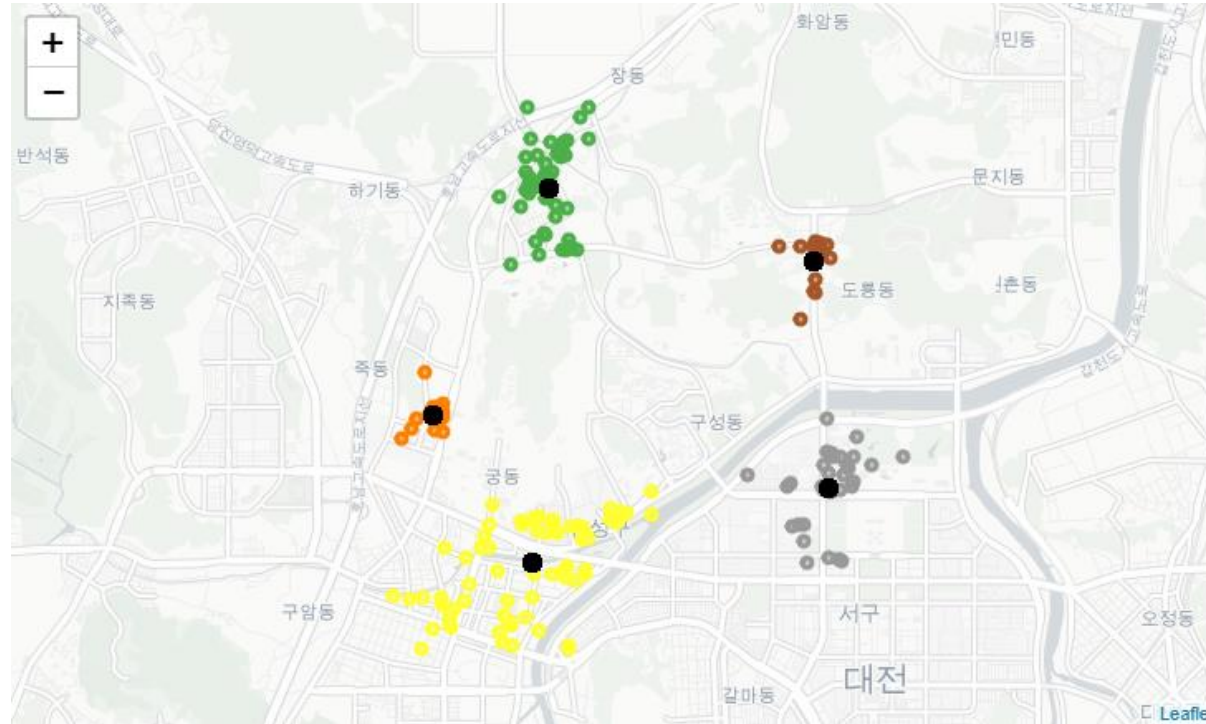


Remove outliers

## Analysis : DBSCAN

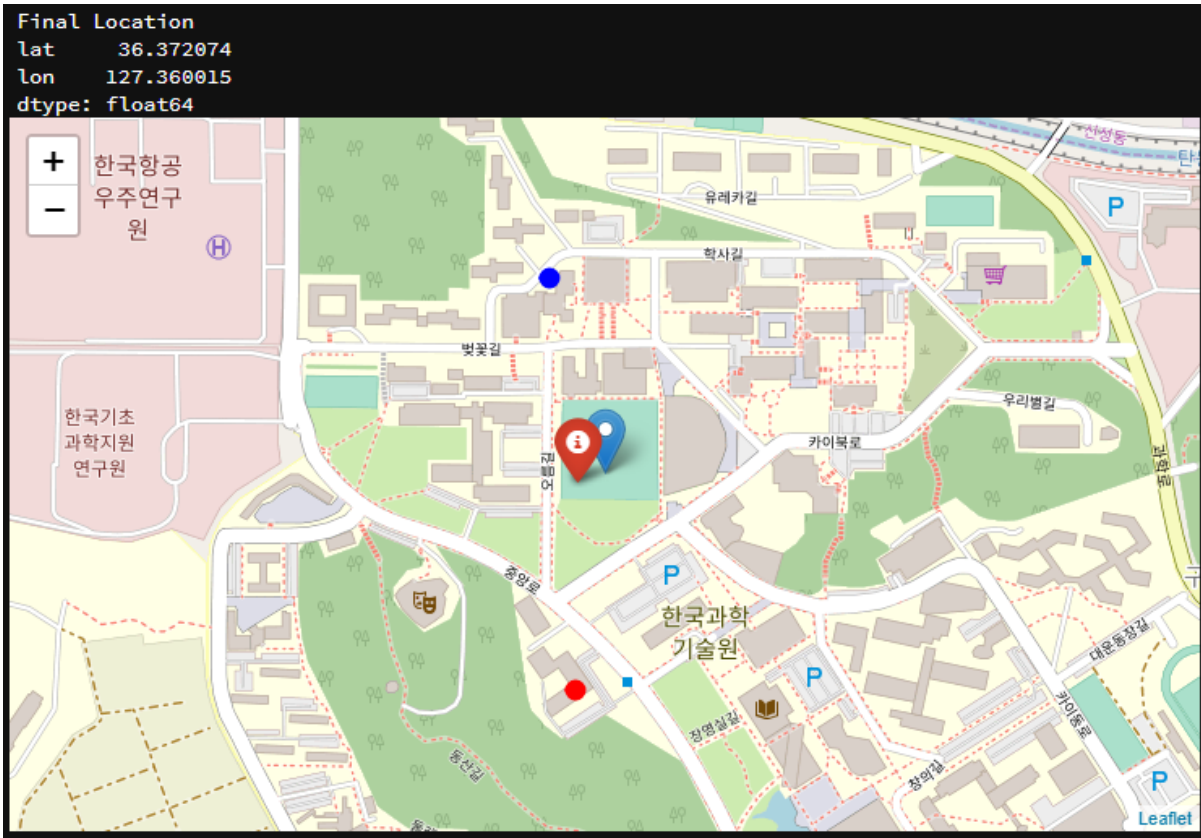
Check the number of Chinese restaurant and total number of restaurant.

0:	2	/	12
1:	1	/	8
2:	7	/	47
3:	1	/	8
4:	2	/	15
5:	8	/	73
6:	3	/	16
7:	0	/	14
8:	4	/	34
9:	0	/	37
10:	1	/	14
11:	3	/	13



Remove group size under 15 and group without Chinese restaurant. → Selected group : 2,4,5,6,8.

## Results and Discussion



1. from k-means : [ 36.369084 , 127.358620 ]

2. from DBSCAN : [ 36.374233 , 127.359663 ]

Final Location : [ 36.372074 , 127.360015 ]

## Conclusion

- Figure out suitable place for the Food Delivery Service Drone Base.
- Get data from Google API and Foursquare
- Using k-means & DBSCAN, we find out main restaurants clusters around KAIST.
- Count the number of Chinese restaurants in the clusters.
- The result of k-means & DBSCAN was different but, the center of the clusters center was almost same.
- From this project we get meaningful data for the business.
- If we try to develop this project, we can consider about other data in real world.