

2025

Where Should Air Taxis Land?

Finding Optimal Vertiport Locations in Seoul Metropolitan Area with K-Means



INTRODUCTION TO MACHINE LEARNING

2025

Chapter

1

Project Planning

2

Data Explanation

3

Preprocessing

4

Modeling

5

Evaluation

6

Conclusion

2025

Project Planning

R&R

PM

Planning and managing the overall project

Modeling Team

Designing machine learning models and evaluating their performance

Data Team

Preprocessing data to improve model performance



Goal: Implement the optimal machine learning model based on the proposed conditions.

2025

Project Planning

Member

name	Part	Major	Remarks
Jaeseong Choe	PM	Mass Communication	Modeling Team Lead
Subin Kim	Modeling	Management	
Jiyoung Lee	Modeling	Life Science	
Hakyeong OH	Data	Urban Engineering	Data Team Lead
Juen Kim	Data	Management	
Eunhye Kim	Data	Management	
Uiseong Ahn	Data	Urban Engineering	

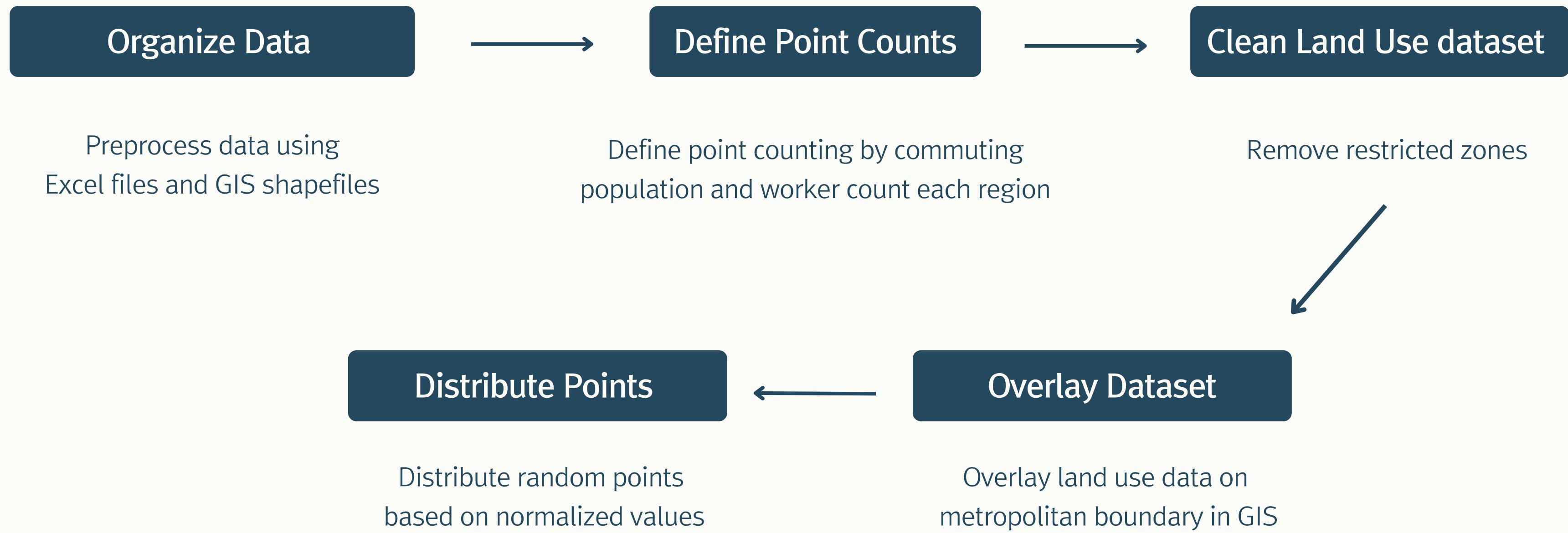
2025

Project Planning

Planning

Work	Todo	Due Date	R&R	Lead	Status	Etc
Kick-off	- Set the Project Scope - R&R - Part Lead Selection	~ 5/20	All	Jaeseong Choe	Done	
Data	- Data Searching	~ 5/28	Data	Hakyeong Oh	Done	
	- Data Preprocessing(lat/lon)	~5/31	Data	Hakyeong Oh	Done	
Modeling	- Develop the Skeleton Model	~5/30	Modeling	Jaeseong Choe	Done	
	- Generate the Pipeline - Tun the Parameter - Get the final result	~6/8	Modeling	Jaeseong Choe	Done	
Presentation	- Make the PPT of Data Team's Part	~6/6	Data	Hakyeong Oh	Done	
	- Make the PPT of Modeling Team's Part	~6/9	Modeling	Jaeseong Choe	Done	
	- Last Feedback at PPT	~6/10	All	Jaeseong Choe	Done	
	- Rehearsal	6/12	All	Jaeseong Choe	Done	
	- Final Presentation	6/13	All	Jaeseong Choe	In Progress	

Overall Preprocessing Flow



Data Organization

Data Name	Purpose	Source
Total commuting population	To determine the demand for commuting flows and normalize point distribution across districts	Statistics Korea
Economic Census worker count	To evaluate the density of business activity and allocate points accordingly with appropriate weighting	Statistics Korea
Urban Area Land Use Zone	To filter based on the UQA land use code table, and to exclude restricted areas	V-world (Digital Twin Korea)
Capital region boundary	To define the spatial boundary and limit the analysis and point generation to the capital region	Statistics Korea and MOIS

Defining Point Count Standard

 Reference Variables:

- ① Total commuting flow
- ② Economic Census worker count

 Method:

Use Ongjin-gun as the minimum reference value

 Example:

Gapyeong-gun's commuting volume = 58,095
 Ongjin-gun's commuting volume = 22,062
 $\triangleright 58,095 \div 22,062 = 2.63 \rightarrow 3$ points assigned

SIG_CD	SD_SGG_Eng	CommutingPopulation	WorkerCount	CommutingPopulation PointCount	WorkerCount PointCount
28720	Ongjin-gun	22062	1457148	1	1
41800	Yeoncheon-gun	44191	4557720	2	3
28710	Ganghwa-gun	58247	5131140	3	4
41820	Gapyeong-gun	58095	5382624	3	4
41250	Dongducheon-si	88557	6307848	4	4
...
11680	Gangnam-gu	1188287	190680504	54	131

Land Use Code Filtering

UQA Code	Land_Use_Tyep
UQA123	Type 3 General Residential
UQA121	Type 1 General Residential
UQA122	Type 2 General Residential
...	...
UQA111	Type 1 Exclusive Residential
UQA310	Exclusive Industrial
UQA120	General Residential Area
UQA300	Urban Planning Facility
UQA200	Mixed-use District
UQA110	Exclusive Residential Area

✓ Data Source:

v-world (Digital Twin Korea) → Land Use Zoning Information
› Urban Area Shapefile

✓ Classification Basis:

- Based on UQA land use code table
- Exclude restricted areas (e.g., green zones, unspecified)

✓ Excluded Areas

- Conservation green zone, Production green zone, Natural green zone
- Unassigned, Planning areas, Agricultural/Forestry zones, Environmental Protection Areas

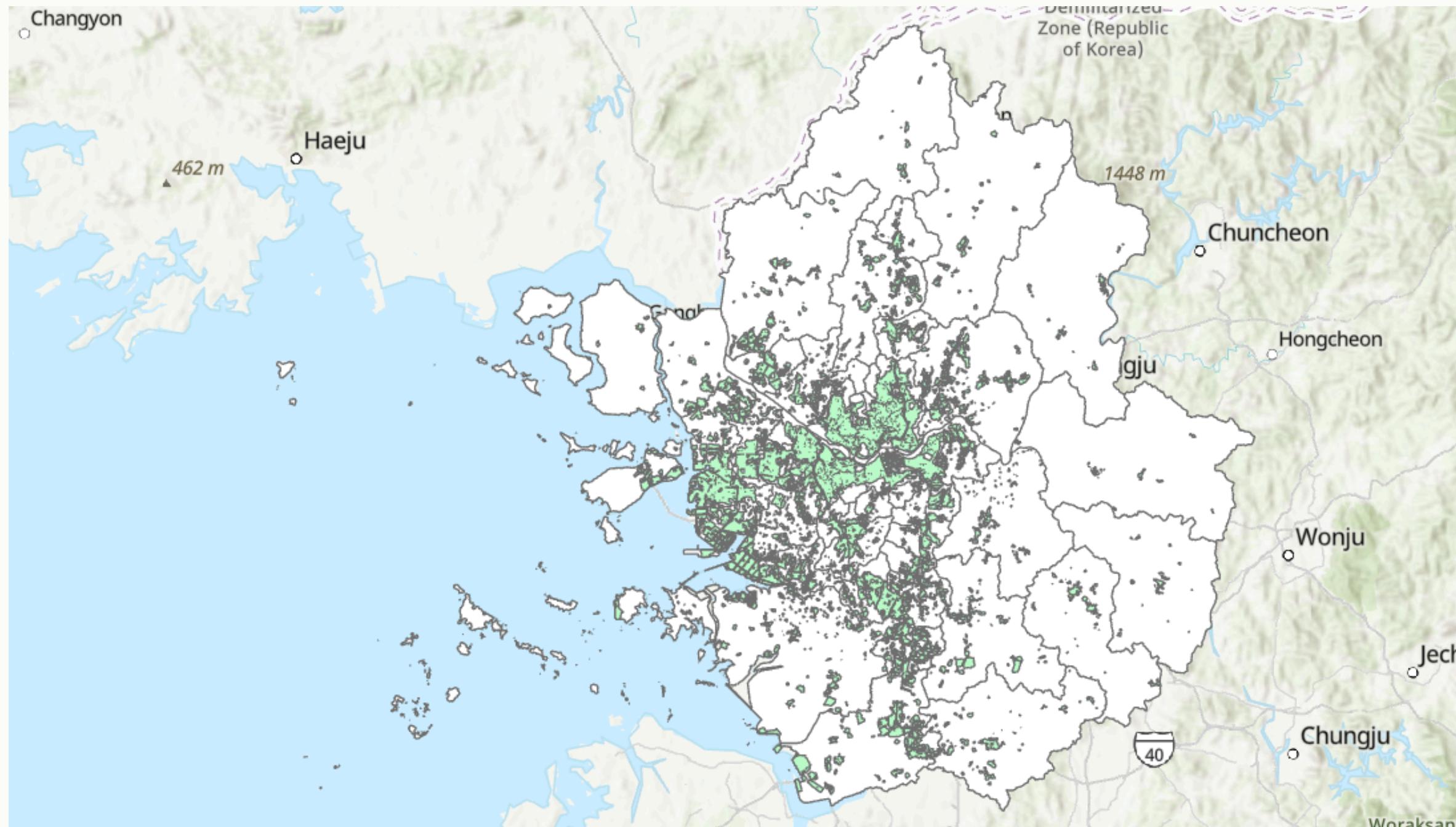
✓ Included Urban Land Use Zones

- Type 1, 2, 3 General Residential
- Semi-residential, Exclusive Residential
- General Commercial, Central Commercial, Neighborhood Commercial
- Distribution Commercial, Mixed-use District
- General Industrial, Semi-industrial, Exclusive Industrial
- Urban Planning Facility

2025

Preprocessing

Land Use Code Filtering



- ✓ Green : Land Use Zone
- ✓ White : Seoul Metropolitan Area boundary

Point Placement Logic

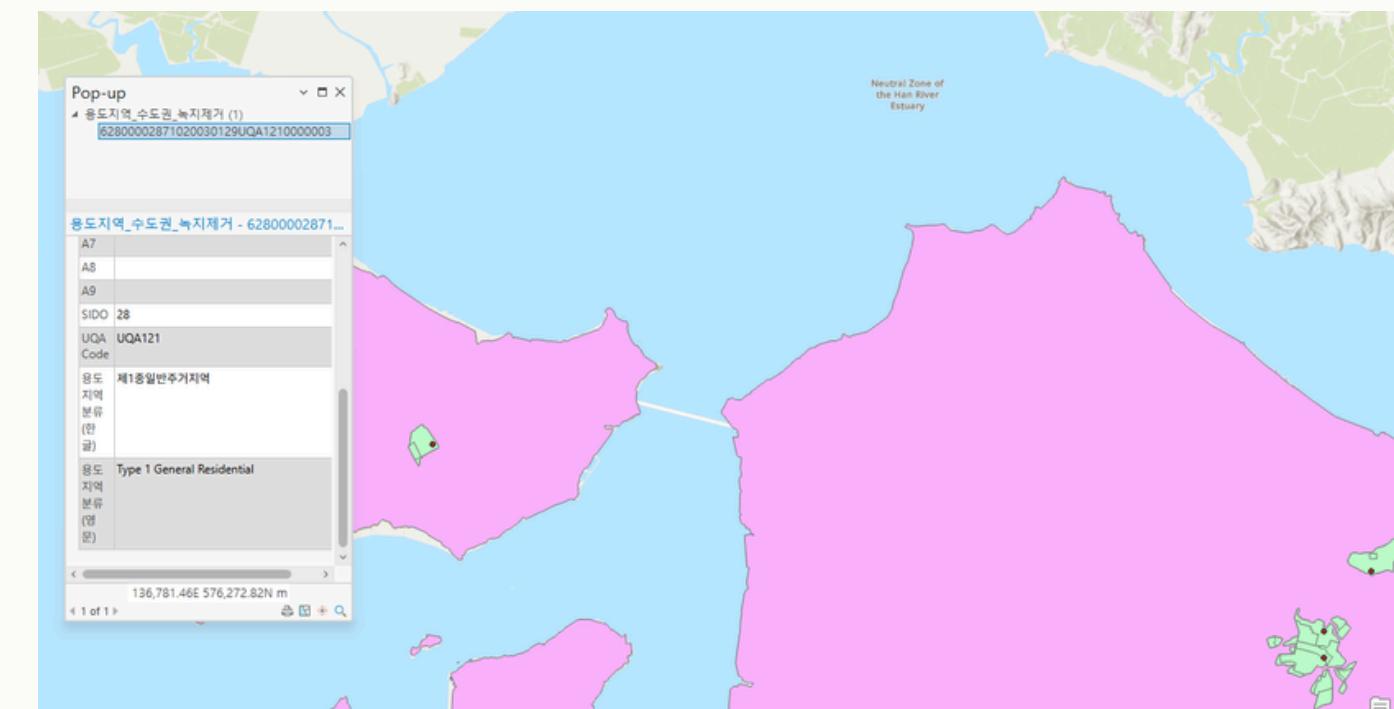
- ✓ Refer to the “00_PointCount” column per district to decide **how many points to assign**

Example: If Gapyeong-gun has a count of 3, then 3 points are distributed within its boundary

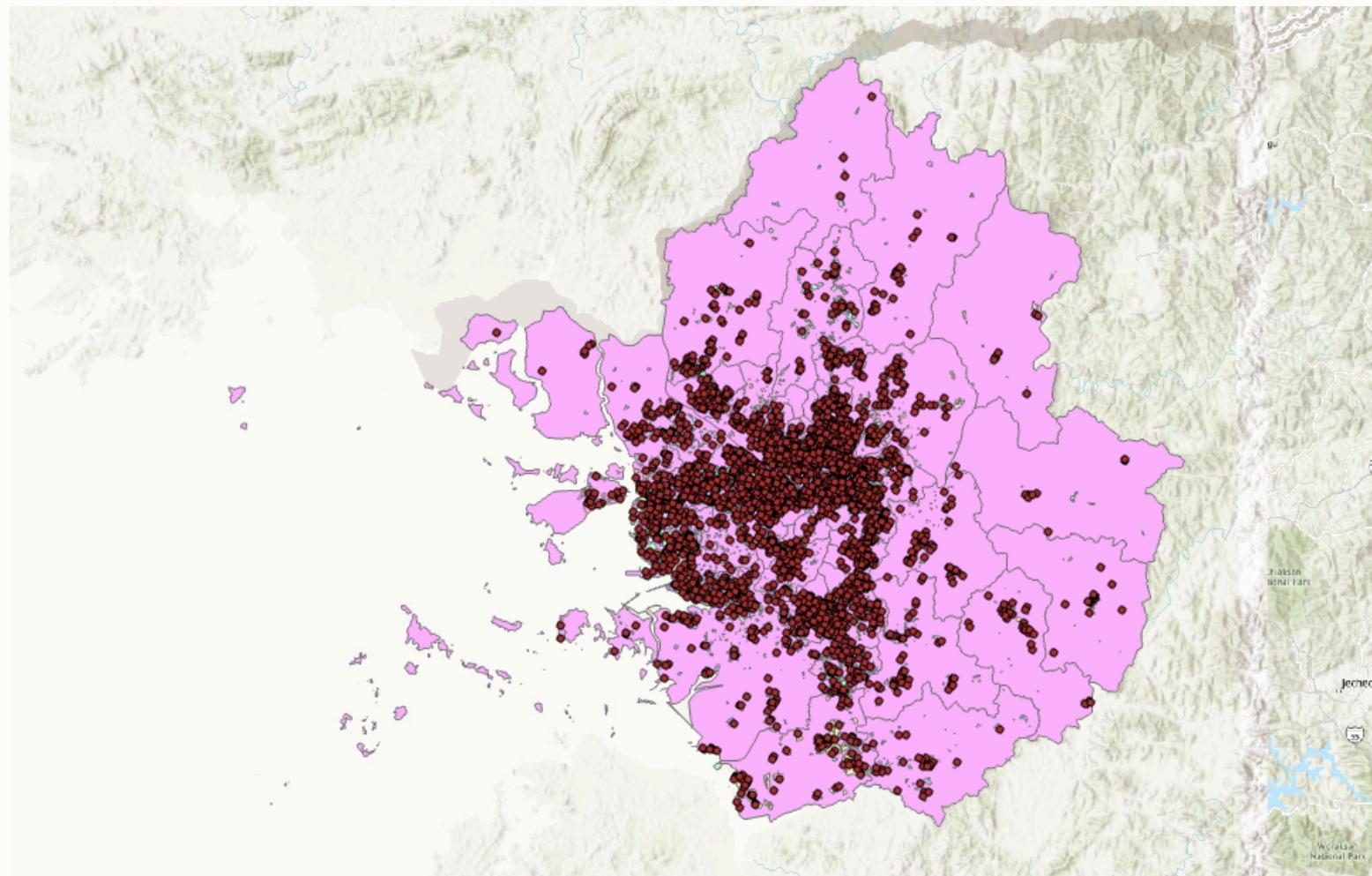
- ✓ Points are generated randomly but prioritized toward **residential or business-dense zones**

Use ArcGIS Pro's “Create Random Points” tool within each district's shapefile

SD_SGG_Eng	CommutingPopulation PointCount	WorkerCount PointCount
Ongjin-gun	1	1
Yeoncheon-gun	2	3
Ganghwa-gun	3	4
Gapyeong-gun	3	4
Dongducheon-si	4	4
...
Gangnam-gu	54	131



GIS Processing View



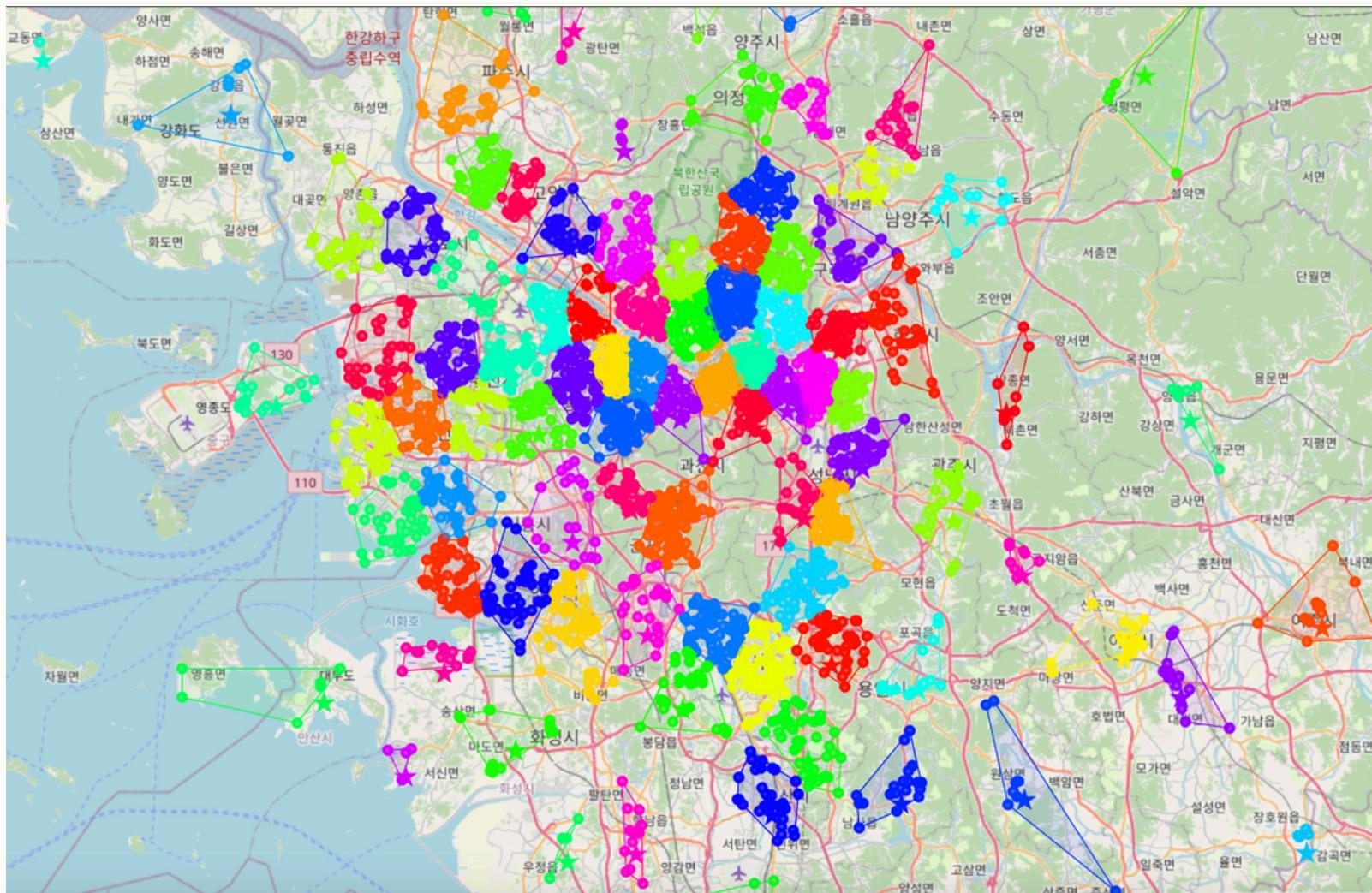
- ✓ In ArcGIS Pro, merged capital region shapefile with filtered land use zones
- ✓ Points generated only on valid parcels
- ✓ Final points verified via attribute table in export layer

K-Means Clustering

$$\begin{aligned}\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_K^{(0)} &\in \mathbb{R}^d \\ C_i^{(t)} &= \left\{ x_j \in \mathbb{R}^d \mid \|x_j - \mu_i^{(t)}\|^2 \leq \|x_j - \mu_k^{(t)}\|^2, \forall k \in \{1, \dots, K\} \right\}, \quad \forall i \in \{1, \dots, K\} \\ \mu_i^{(t+1)} &= \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j, \quad \forall i \in \{1, \dots, K\}\end{aligned}$$

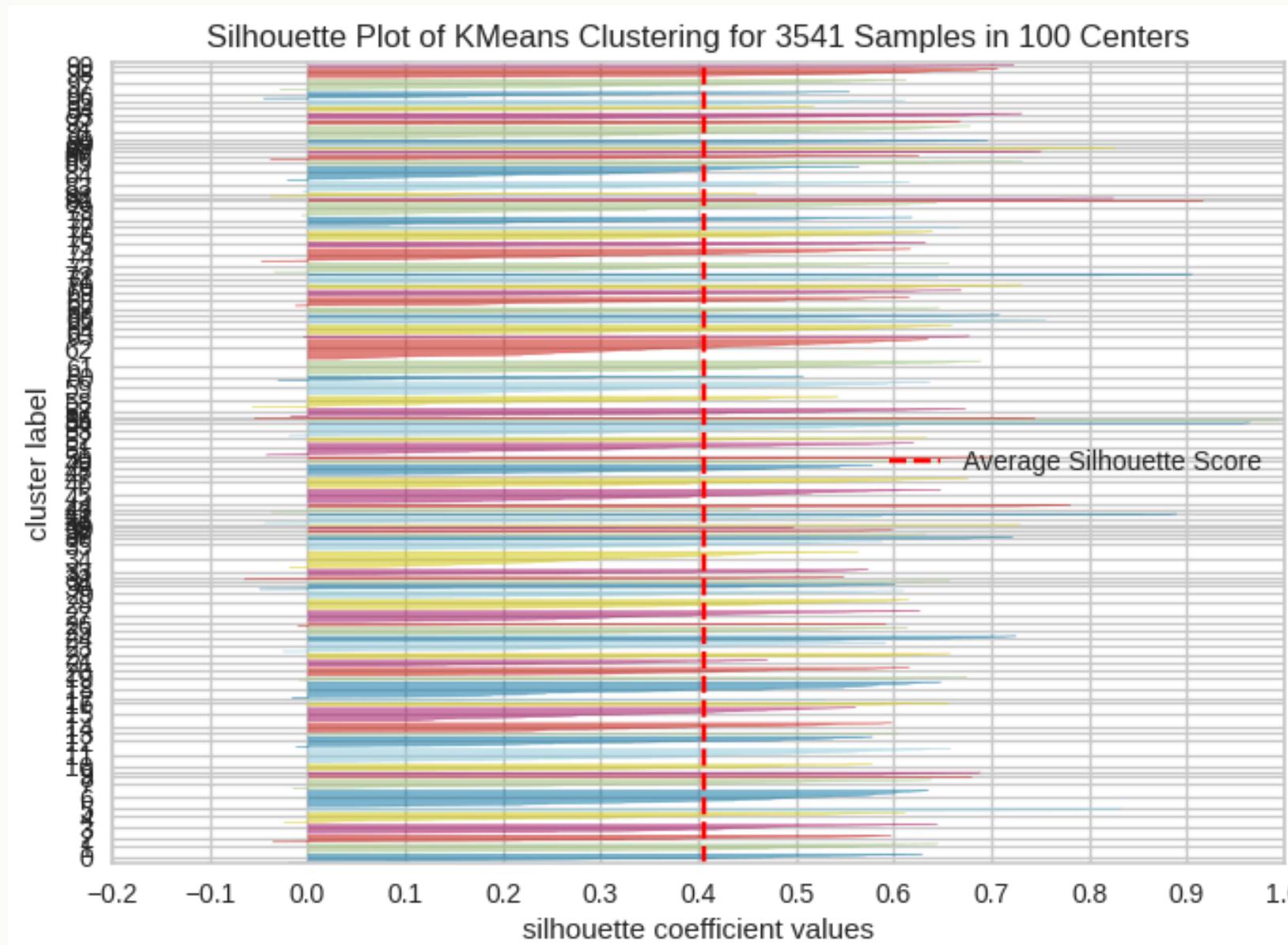
- Clusters based on Euclidean distance
- K-Means Clustering (scikit-learn)
- Clusters: 100
- Iterations: 1000
- Random seed: 123
- Manual adjustment of centroids in non-installable areas

Map Visualization



- Assigned cluster labels to the original data
- Visualized results using Folium
- Displayed point locations, cluster boundaries, and centroids

Silhouette Coefficient



- Evaluated using silhouette coefficient
- Average silhouette score 0.4
- Possible improvements: fewer clusters or K-Means++ initialization

Project Result

Need to Determine the Optimal Number of Clusters

it is necessary to identify the optimal number of clusters using the Elbow Method.

Consider Alternative Models

K-Means++, Weighted K-Means, K-Medoids, DBSCAN, Mean-Shift, etc.

Improve Distance Calculation Methods

In spatial clustering, it is important to consider the Earth's curvature when calculating distances. Therefore, models that use Haversine distance or geodesic distance instead of Euclidean distance should be considered.

Consider Project Feasibility

South Korea is technically at war (armistice state), so the use of aerial vehicles within Seoul is not permitted.

2025

Thank you