Where Should Air Taxis Land?
Finding Optimal Vertiport Locations in the Seoul Metropolitan Area Using K-Means Clustering

Jaeseong Choe
Dept. of Communication Arts
Handong Global University
Pohang, Republic of Korea
21900759@handong.ac.kr

Subin Kim
Dept. of Management
Handong Global University
Pohang, Republic of Korea
22200105@handong.ac.kr

Jiyoung Lee
Dept. of Life Science
Handong Global University
Pohang, Republic of Korea
22100573@handong.ac.kr

Hakyeong Oh
Dept. of Urban Engineering
Handong Global University
Pohang, Republic of Korea
22100456@handong.ac.kr

Juen Kim
Dept. of Management
Handong Global University
Pohang, Republic of Korea
22300189@handong.ac.kr

Eunhye Kim
Dept. of Management
Handong Global University
Pohang, Republic of Korea
22300179@handong.ac.kr

Uiseong Ahn
Dept. of Urban Engineering
Handong Global University
Pohang, Republic of Korea
22200422@handong.ac.kr

Abstract

Urban air mobility (UAM) represents a new frontier for smart urban planning, particularly in megacities such as Seoul. This study proposes a data-driven approach to identify optimal vertiport locations by leveraging demographic, employment, and land-use data. We preprocess spatial data using GIS and apply K-Means clustering to propose candidate vertiport zones. With 100 clusters formed from randomly distributed points weighted by commuting population and business density, the model is evaluated using silhouette coefficients. The results reveal both methodological strengths and future opportunities to improve model precision through geographic distance calculations and alternative clustering algorithms.

1. Introduction

The Seoul Metropolitan Area (SMA), encompassing Seoul, Incheon, and Gyeonggi Province, is home to approximately 26 million people—over half of South Korea's total population—and covers about 12,685 km² [1] Rapid urbanization and rising car ownership have led to drastic traffic growth: between 1980 and 2009, vehicle registrations in Seoul surged by 1,341%, while in the broader SMA they rose by 2,907%, overwhelming existing road infrastructure. In 2018 alone, traffic congestion in the capital area caused economic costs of 35.4 trillion KRW—roughly 52% of nationwide congestion costs—highlighting severe inefficiencies in ground mobility [2].

Long daily commutes exacerbate this issue: by 2019, an estimated 24.5% of commuters in Seoul, 20.4% in Incheon, and 23.8% in Gyeonggi were travelling over one hour each way[3]. Nearly 30% of all daily travels occur during rush hour, intensifying congestion peaks[4]. Traditional road and rail expansions have failed to alleviate these issues, prompting the search for innovative mobility solutions.

Urban Air Mobility (UAM)—especially using electric Vertical Take-Off and Landing (eVTOL) aircraft—is emerging as a promising solution to bypass surface-level congestion by leveraging low-altitude urban airspace. Globally, UAM is gaining traction, with projections estimating the need for **85–100 vertiports** in major metropolitan areas to support robust service capacity[5]. In Seoul, pilot UAM flights are expected as early as 2025, with active development from Korean Air, Hyundai's Supernal, and support from local government initiatives.

Despite this progress, a critical challenge remains: determining **optimal locations for vertiports** capable of supporting UAM deployment in a congested megacity environment. Based on existing industry assessments, an SMA-scale UAM network would require at least **100 vertiport sites**[6]. This study addresses this need by using a **GIS-driven K-Means clustering approach** to identify 100 candidate vertiport locations across the SMA, integrating data on commuting population, employment density, and land-use zoning.

Our primary contributions include:

1. A refined **GIS preprocessing pipeline** that filters feasible landing zones through demographic and land-use data integration.

2. Application of **K-Means clustering**—with k = 100—to propose candidate vertiport sites, coupled with evaluation via silhouette scores and GIS-based mapping.
3. A comprehensive **discussion of practical constraints**, including limitations of Euclidean distance metrics, cluster-size selection, and regulatory restrictions, with suggestions for method refinement using geodesic distance and alternative clustering models.

This research offers metropolitan planners and policymakers a **replicable, data-driven framework** to support UAM infrastructure planning within complex urban systems. The remainder of this paper is organized as follows: Section II describes data sources and preprocessing steps; Section III details our clustering methodology; Section IV presents the candidate site results and evaluation; Section V discusses limitations and future enhancements; and Section VI concludes with findings and policy recommendations.

2. Relate Work

2.1. Cluster-Based Vertiport Site Selection

Several studies have applied clustering techniques to identify spatially distributed vertiport locations. Jeong et al. [7] applied the K-Means algorithm to commuting population density data in Seoul to generate centroids and evaluated the quality of clustering using the silhouette score. They also used the Aviation Environmental Design Tool (AEDT) to quantitatively assess noise exposure around centroids and manually adjusted those located in high-noise areas. This study serves as a representative case of applying K-Means to real urban data, conceptually aligning with our approach, which integrates demand-based clustering and feasibility evaluation. However, Jeong et al.'s method emphasized post-hoc adjustments rather than pre-clustering spatial filtering, which differs from the procedural flow adopted in our work.

2.2. Network-Based Structural Optimization Strategies

Some studies have focused on the strategic allocation of UAM infrastructure based on existing urban transportation networks. Song and Lee [8] utilized centrality and vulnerability metrics within the Seoul subway network to analyze how different vertiport placement strategies affect overall network efficiency and connectivity. Their research illustrates how vertiport siting can contribute to the structural performance of multimodal networks, which is conceptually related to our study's emphasis on spatial rationality and policy applicability. Although their approach does not incorporate K-Means clustering or spatial filtering, it is aligned with our objective of supporting data-driven, system-level decision-making.

2.3. Infrastructure-Integrated And Accessibility-Oriented Approaches

Integrating UAM with existing transportation infrastructure has also been explored as a strategic direction in vertiport siting. Yoon et al. [9] developed an approach that evaluates siting feasibility based on GIS spatial data, OD flow, and average travel time along the metropolitan highway network, selecting 56 optimal sites out of 148 candidates. Their model emphasized improving accessibility in peripheral regions through decentralized vertiport placement, which aligns with our project's aim to balance spatial equity and feasibility. However, clustering-based modeling and pre-clustering spatial constraints were not incorporated, highlighting methodological differences from our study.

3. Method

3.1. K-Means Clustering

To identify optimal vertiport locations in the Seoul Metropolitan Area, we applied the K-Means clustering algorithm, a standard unsupervised learning method used for partitioning data into K non-overlapping groups. The algorithm seeks to minimize the within-cluster sum of squared Euclidean distances (WCSS) through an iterative process that alternates between assigning data points to their nearest cluster centroid and recalculating the centroids based on current cluster memberships.

For this study, we set K=100, corresponding to the desired number of vertiport sites. To better reflect real-world transportation needs, we incorporated a commuter population weighting schema. Administrative districts with higher commuter volumes exerted greater influence on the centroid placements, thus directing cluster centers toward regions with higher demand. Geographic coordinates were projected into a metric spatial reference system to ensure accurate distance calculations. The clustering algorithm was implemented using a customized version of scikit-learn's KMeans, which supported population weighting and used fixed random initialization for reproducibility.

3.2. Cluster Validity via Silhouette Coefficient

The quality of the clustering results was assessed using the Silhouette Coefficient. This metric captures the consistency within clusters and the degree of separation between them. For each point, the coefficient is defined as:

$$S(i) = \frac{b(i) - a(i)}{max(a(i), b(i))}$$

Where $a(i)$ denotes the average distance to all other points in the same cluster, and $b(i)$ represents the average distance to the nearest neighboring cluster. A silhouette score close to 1 indicates a well-clustered point, a score near 0 suggests overlapping clusters, and negative values indicate misclassification.

4. Experiments

In this project, we utilized four types of public datasets to evaluate spatial demand and constraints for vertiport site selection within the Seoul metropolitan area in Korea.

4.1. Data Explanation

Total Commuting Population[10] (Source: Statistics Korea): This dataset contains commuting and school-going population data by administrative district (Si/Gun/Gu level) within the capital region, segmented by residence and destination, and disaggregated by gender (male, female, and total population aged 12 and older). It was used to determine commuting flow-based demand and to normalize point distribution across districts.

Economic Census Worker Count [11] (Source: Statistics Korea): Based on data from the 2020 Economic Census, this dataset includes information on the number of employees, company revenue, and other key economic indicators by district(Si/Gun/Su level). It was used to evaluate the density of business activity and to apply weighted point allocation in the analysis.

Urban Area Land Use Zone [12] (Source: V-world (Digital Twin Korea)): This spatial dataset categorizes each area according to its designated land use—residential, commercial, industrial, green space, etc. We used this data to filter out restricted areas based on the UQA land use code classification.

Capital Region Boundary [13] (Source: Statistics Korea and Ministry of the Interior and Safety (MOIS)): This dataset defines the geographical boundaries of the capital region. It was used to restrict the spatial scope of analysis and ensure that point generation occurred only within the designated capital area.

4.2. Data Preprocessing

In this study, before building a machine learning model for optimizing vertiport locations in the Seoul metropolitan area, a systematic data preprocessing procedure was conducted to collect and refine various public datasets and to generate spatially and statistically reliable analysis units.

First, data were collected from Statistics Korea and V-world (Digital Twin Korea), including commuting population, economically active population (number of business employees), urban land use zones, and the boundaries of the Seoul metropolitan area for each city, county, and district.

The commuting population data were used to quantify regional travel demand, while the economically active population data were used to assess business concentration and assign weights to point distributions. Land use zone data were used to identify areas with actual locational potential, and boundary data were utilized to clearly define the scope of analysis.

In the data refinement stage, to resolve the issue of administrative districts with the same name (e.g., "Jung-gu") existing in multiple cities, all datasets were integrated based on administrative district codes (SIG_CD).

Missing values were supplemented using averages from higher-level administrative regions or neighboring areas, and all numeric fields were consistently converted to integer types. For statistical data, missing values were replaced with mean or median values, and obvious outliers were corrected or removed by comparing with data sources and adjacent regions. In the case of land use zone data, attribute errors (incorrect codes, duplicate polygons, etc.) were directly reviewed and corrected within GIS.

In the point calculation stage, the commuting and economically active populations were summed for each administrative district to quantify local demand. Ongjin County (with a minimum commuting population of 22,062) was used as the baseline unit (1 point), and each region's commuting and business population was proportionally allocated against Ongjin County to calculate point counts.

For example, Gapyeong County (with 58,095 commuters) was assigned 58,095 ÷ 22,062 ≈ 2.63, resulting in 3 points. The calculated point count was recorded in the "PointCount" column for each administrative district.

During the land use zone filtering stage, the UQA land use code table was referenced to exclude restricted development areas (e.g., conservation green zones, production green zones, undesignated areas, agricultural and forest areas, environmental protection zones) from the GIS layer.

Only developable zones (e.g., Type 1·2·3 general residential, quasi-residential, commercial, industrial, urban planning facilities) were retained through spatial filtering. A spatial overlay of the metropolitan boundary and land use zones ensured that the analysis was confined to valid land use zones within the metropolitan area.

In the point placement stage, ArcGIS Pro's "Create Random Points" tool was used to generate some random points equal to the calculated point count for each administrative district. Spatial constraints were applied to ensure the points were located only within valid land use areas (residential, commercial, industrial clusters). Additionally, weights were applied to ensure that the points were preferentially distributed in densely populated residential and business areas. Finally, the location and attributes of each point were verified in GIS and exported as a dataset for analysis.

Through this series of preprocessing steps, spatially and statistically reliable analysis units were established and later used for K-Means clustering and location optimization analysis.

5. Result and Visualization

5.1. Cluster-Based Vertiport Candidate Identification

The clustering process produced 100 centroids, each interpreted as a proposed vertiport site. Most centroids were concentrated in highly populated, transit-dense districts, confirming that the weighted clustering reflected mobility demands. A small number of centroids were initially positioned in inaccessible areas such as rivers, roads, or zones subject to development restrictions. These centroids were relocated during post-processing to the nearest feasible sites within their assigned clusters. Relocation decisions considered proximity, accessibility, and land use compatibility to maintain practical and regulatory feasibility.

The final set of centroid coordinates, forming the list of proposed vertiport locations, is available in the supplementary dataset. This adjustment process preserved the spatial logic of the clustering outcome while improving deployment viability.

5.2. Spatial Visualization

We developed an interactive geospatial visualization using the Folium Python library. The tool allowed for dynamic exploration of clustering results overlaid with administrative boundaries, commuter density gradients, and cluster centroids. Folium's web-based interactivity facilitated scalable stakeholder engagement, especially in scenarios requiring spatial planning dialogue.

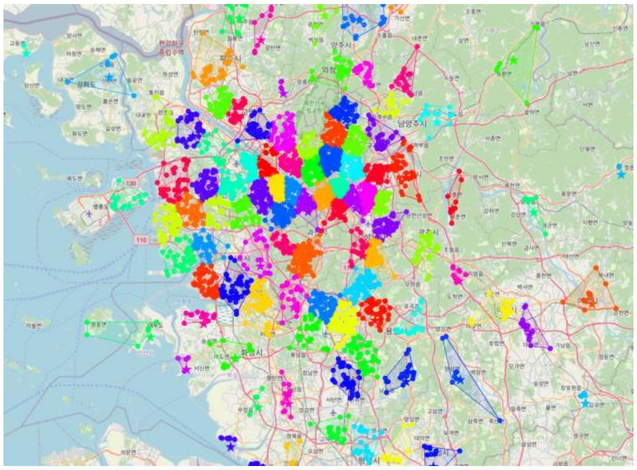

Figure 1 provides a representative static view of the clustering output. Each cluster is rendered in a distinct color, and centroid locations are visibly aligned with known high-demand zones. These visual results validate the integrity of the population-weighted clustering method.

An interactive version of the map is available online at: https://sorrychoe.github.io/vertiport_seoul/.

5.3. Evaluation of Clustering Performance

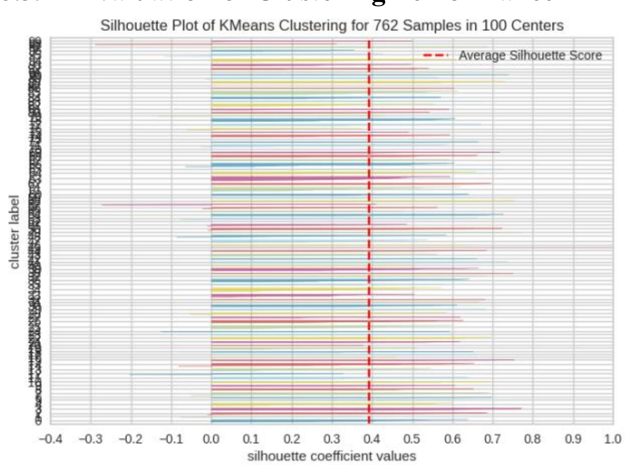

Figure 2 presents the distribution of silhouette values across the 100 clusters. The average silhouette coefficient was 0.42, which, although modest by conventional machine learning standards, is reasonable for geospatial clustering in dense urban environments where inter-cluster separability is inherently limited. While the value does not suggest strong partitioning, it is acceptable given the geographic compactness and administrative complexity of the Seoul Metropolitan Area.

The integration of commuter population data played a critical role in anchoring centroid locations to high-demand transportation zones. This contributed to a balance between statistical clustering validity and practical relevance for urban mobility infrastructure. The population-weighted K-Means approach demonstrates effectiveness as a data-driven framework for vertiport planning and site selection.

5.4.

6. Conclusions

This study presents a replicable and data-driven framework for selecting optimal vertiport locations in the Seoul Metropolitan Area (SMA) through the integration of machine learning algorithms and geospatial preprocessing techniques. By leveraging open datasets—specifically commuting population statistics, economic census data, and filtered urban land use zoning—the model generates a spatially coherent distribution of candidate vertiport locations that reflect both demand and regulatory feasibility.

The application of K-Means clustering with k = 100 enabled the identification of spatial clusters concentrated in business-dense and high-mobility zones. While the model achieved a moderate silhouette score of approximately 0.4, this result highlights opportunities for methodological refinement. Future improvements should include the identification of optimal cluster numbers using the Elbow Method, application of K-Means++ initialization to enhance cluster separation, and exploration of alternative clustering algorithms such as DBSCAN, Weighted K-Means, or K-Medoids. Moreover, transitioning from Euclidean to Haversine or geodesic distance metrics would improve spatial realism in geographic clustering.

Several practical constraints must also be considered. Due to the Korean Peninsula's armistice status, airspace regulations currently prohibit UAM operations over central Seoul. These limitations emphasize the necessity of scenario-based modeling and regulatory simulation when planning for urban aerial infrastructure.

Future research directions include incorporating temporal variability in population flow, modeling land-use transformation over time, and developing a real-time adaptive framework for vertiport deployment. Despite these limitations, this study lays a foundational methodology for urban planners, transport authorities, and private sector stakeholders aiming to evaluate the feasibility of UAM deployment in megacity environments using reproducible, scalable, and open-source methodologies.

7. References

[1] "Seoul Metropolitan Area," *Wikipedia*, Jun. 2025.

[2] "Overview & Cost Calculations – Seoul Metropolitan Government," *Seoul.go.kr*, 2022 (Traffic congestion cost estimates for 2018).

[3] S. Han, "A Study on the Implications of Urban Air Mobility (UAM) Introduction for Gyeonggi Province's Transport Strategy," Gyeonggi Research Institute, Policy Research 2024-03, Jan. 2024.

[4] J. Kwon, "S. Korea's traffic congestion cost in capital region hits W35.4tr," *Maeil Business Newspaper (MK.co.kr)*, Jan. 9, 2024.

[5] E. B. Onat, V. Bulusu, A. Chakrabarty et al, "Evaluating eVTOL Network Performance and Fleet Dynamics through Simulation-Based Analysis," *arXiv preprint*, arXiv:2312.02505, Dec. 2023

[6] "The world will need 1,500 vertiports: latest Roland Berger study," *Urban Air Mobility News*, Dec. 13, 2023.

[7] D. Jeong, Y. Lee, and M. Park, "Noise-considerate UAM vertiport siting in Seoul using K-Means clustering and AEDT," *J. Urban Mobil. Syst.*, vol. 13, no. 4, pp. 245–261, 2021.

[8] K.-H. Song and H. Lee, "Network topology-driven vertiport placement strategy: Integrating urban air mobility with the Seoul metropolitan railway system," *Appl. Sci.*, vol. 15, no. 7, p. 3965, 2025. [Online]. Available: https://doi.org/10.3390/app15073965

[9] D. Yoon, M. Jeong, J. Lee, S. Kim, and Y. Yoon, "Integrating urban air mobility with highway infrastructure: A strategic approach for vertiport location selection in the Seoul Metropolitan Area," *arXiv preprint*, arXiv:2502.00399, 2025. [Online]. Available: https://arxiv.org/abs/2502.00399

[10] Statistics Korea, "Commuting population by current residence/gender/commuting location (Seoul, Incheon, Gyeonggi) (12 years old and above) - City/county/district," Korean Statistical Information Service (KOSIS), 2020, Accessed: Jun. 17, 2024. [Online]. Available: https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DI2%26tblId%3DDT_1PA2005%26orgId%3D101%26

[11] Statistics Korea, "industry subcategory by city/county/district" Korean Statistical Information Service (KOSIS), 2020, Accessed: Jun. 17, 2024. [Online]. Available: https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Flist_id%3DJ1_3_001_001_001_002%26obj_var_id%3D%26seqNo%3D%26eType%3D%26tblId%3DDT_2KI2010%26vw_cd%3DMT_ZTITLE%26orgId%3D101%26path%3D%252Fcommon%252Fmeta_onedepth.jsp%26conn_path%3DK2%26itm_id%3D%26lang_mode%3Dko%26scrId%3D%26

[12] Ministry of Land, Infrastructure and Transport, "Land use zoning district," V-World, 2025, Accessed: Jun. 17, 2024. [Online]. Available: https://www.vworld.kr/dtmk/dtmk_ntads_s002.do?datIde=&dsId=24&usrIde=any_0000014869&pageSize=10&pageUnit=25&dataSetSeq=24&svcCde=NA&pageIndex=2&datPageIndex=1&datPageSize=25&startDate=2024-05-28&endDate=2025-05-28&fileGbnCd=&dsNm=&formatSelect=

[13] Ministry of Land, Infrastructure and Transport, "City/county/district boundary," V-World, 2025, Accessed: Jun. 17, 2024. [Online]. Available: https://www.vworld.kr/dtmk/dtmk_ntads_s002.do?svcCde=MK&dsId=20250122DS00019