Application of Statistical and Spatial Analysis Techniques for Optimization of Electric Vehicle Charging Station Installation

Seongjin Park
University of Wisconsin-Madison
Computer Science & Data Science
(seongjinpark99@gmail.com)

1. Introduction

With the recent rapid increase in the adoption of electric vehicles worldwide, the efficient establishment of electric vehicle charging infrastructure has emerged as a critical socio-economic challenge. Rather than simply increasing the number of installations, it is essential to ensure user accessibility, maximize operational efficiency, and maintain an appropriate distribution in line with predicted future demand. Accordingly, active research is being conducted to systematically analyze existing charging station distributions and regional demand, and to identify areas where additional installations are necessary.

In this study, we combined data on the installation status of electric vehicle charging stations from Korea Electric Power Corporation (KEPCO), the 2021 estimated traffic volume data (Total Annual Daily Traffic, ALL_AADT) from the National Transportation Data Open Market, and the electric vehicle registration status data as of February 28, 2025. This allowed us to quantify the nationwide distribution of KEPCO's electric vehicle charging infrastructure. The traffic volume and electric vehicle registration numbers for each region were normalized to calculate a demand index, and the total number of charging stations was proportionally allocated to derive an ideal distribution. The gap (supply-demand gap) between the ideal allocation and the actual installations was then computed. For regions with a positive gap, a K-Means clustering technique incorporating candidate coordinates based on traffic volume data and demand weights was applied to derive candidate locations for additional installations, thereby exploring solutions to infrastructure imbalances and supporting policy measures.

2. Related Research

With the widespread adoption of electric vehicles, the efficient placement of charging infrastructure has emerged as an important research topic. Many scholars have defined the problem of selecting locations for electric vehicle charging stations as a multi-objective optimization problem and have proposed various models and algorithms.

2.1 Location Optimization Models
p-Median and p-Center Models:

The p-median model aims to minimize the total travel distance between demand points and charging stations, whereas the p-center model seeks to minimize the maximum service distance. Hodgson et al. (1996) applied these models to transportation infrastructure planning to balance efficiency and equity, and Felipe et al. (2014) along with Chung &

Kwon (2015) proposed extended models that consider the driving range of electric vehicles.

Maximal Covering Location Problem (MCLP):
This model selects charging station locations to cover the maximum demand within a given budget, and it was utilized in studies by Chung & Kwon (2015) and Capar et al. (2013).

2.2 Data-Driven Clustering Techniques
K-Means and Hierarchical Clustering:
Chen et al. (2013) derived candidate sites for charging stations using K-Means clustering by combining traffic volume, population density, and land use data, while Wi et al. (2015) proposed a location selection model that reflects uncertainty through weighted fuzzy clustering. In this study, a demand-based K-Means clustering approach is likewise used to derive candidates for additional installations in regions with a positive supply-demand gap.

2.3 GIS and Integrated Approaches
Use of GIS:
Lee et al. (2014) combined Seoul's traffic flow data with GIS to calculate the optimal density of charging stations, while Xiong et al. (2018) developed a dynamic demand forecasting model using real-time traffic data.

Multi-objective Optimization:
Zhang et al. (2017) and Park et al. (2020) applied multi-objective optimization models that simultaneously consider installation costs, environmental impacts, and user convenience.

3. Proposed Method and Application

This study conducted the following procedures, based on nationwide data, for the optimization of KEPCO's electric vehicle charging stations.

3.1 Calculation of the Demand Index

3.1.1. Normalization of Traffic Data
For each administrative district (at the town, township, and neighborhood level), the "Total Annual Daily Traffic" (ALL_AADT) provided by the National Transportation Data Open Market was divided by the maximum value in the dataset to yield a normalized value between 0 and 1.
Normalized Traffic Indicator = (Total Annual Daily Traffic (ALL_AADT) for the district) / (Maximum Total Annual Daily Traffic (ALL_AADT) among all districts)

3.1.2. Normalization of Electric Vehicle Registration Data
The number of electric vehicle registrations as of February 28, 2025, was normalized by dividing by the maximum registration count in the dataset.

Normalized Electric Vehicle Registrations = (Electric vehicle registration count in the region) / (Maximum electric vehicle registration count among all regions)

3.1.3. Calculation of the Demand Index
Both the normalized traffic indicator and the normalized electric vehicle registrations were assigned equal weights (1.0 each) and simply summed to calculate a relative demand index for each region.

Demand Index = (Normalized Total Annual Daily Traffic (ALL_AADT)) + (Normalized Electric Vehicle Registrations)

Through the normalization process, the scales were unified, allowing the simple summation of the two indicators to effectively reflect the differences in demand across regions.

3.2 Ideal Allocation of Charging Stations and Calculation of Supply-Demand Gap

3.2.1. Calculation of the Ideal Number of Charging Stations
The total number of installed charging stations (approximately 4,614) was aggregated from KEPCO's electric vehicle charging station installation data. The ideal number of charging stations was derived by allocating the total number in proportion to each region's share of the total demand index.

$$\text{Ideal Count}_{\text{지역}} = \text{Total Charging Stations} \times \frac{\text{Demand Index}_{\text{지역}}}{\sum \text{Demand Index}}$$

3.2.2. Aggregation of Actual Charging Stations and Gap Calculation

Using the address information from the installation data, the actual number of installed charging stations in each region was tallied. The supply-demand gap was then computed as:

Supply-Demand Gap (Gap) = Ideal Number of Charging Stations − Actual Number of Charging Stations

A positive gap indicates that a region requires additional installations.

3.3 Derivation of Candidate Locations for Additional Installations: Application of the K-Means Clustering Technique

3.3.1 Selection of Candidate Coordinates

The traffic volume data's "city, county, town, and neighborhood" addresses were concatenated to form a full address. Using Nominatim (a geocoding API), the latitude and longitude for each address were obtained. These coordinates were then used as candidate data, with the normalized "Total Annual Daily Traffic" (ALL_AADT) value assigned as the demand weight for each coordinate.

3.3.2 Demand-Based K-Means Clustering

For each region with a positive supply-demand gap, the number of additional charging stations needed (rounded to the nearest integer) was set as the number of clusters, and K-Means clustering was performed on the candidate coordinates within that region. In clustering, the normalized ALL_AADT value for each coordinate was applied as a weight so that locations with higher demand had a greater influence on the calculation of cluster centers. The resulting cluster centers were designated as candidate locations for additional installations.

3.3.3 Handling Insufficient Candidate Data

If the number of candidate coordinates in a given region is less than the number required for additional installations, a representative candidate location is derived by calculating a weighted average of the candidate coordinates in that region based on their demand weights.

4. Experiment

In this section, the application process of the proposed methodology is explained in stages: data preprocessing, statistical and clustering analysis, and result visualization.

4.1 Data Preprocessing

First, the data based on electric vehicle charging station installation status, traffic volume, and electric vehicle registrations were processed. The charging station data was standardized to the EPSG:4326 (latitude/longitude) coordinate system, and regional information at the province/city level was extracted using the address information.

```
[1] EV 충전소 GeoDataFrame (gdf)
----------------------------------------
형태: (4614, 12)
컬럼: ['field1', 'field2', 'field3', 'field4', 'field5', 'field6', '_GC_TYPE', '_CLEANADDR', 'geometry', 'region', 'lat', 'lon']
샘플 데이터 (처음 3개행):
  field1    field2                field3 field4 field5           field6  \
0     전체     가평소방서  경기도 가평군 가평읍 대곡리 213-5      1      0        SM3 Z.E,테슬라
1     전체   가평종합운동장    경기도 가평군 가평읍 대곡리 316      2      0  BMW i3,스파크EV,볼트EV
2     전체  가평하수도사업소    경기도 가평군 가평읍 달전리 364      1      0        SM3 Z.E,테슬라

  _GC_TYPE           _CLEANADDR                  geometry region  \
0        정  경기도 가평군 가평읍 대곡리 213-5  POINT (127.51669 37.82366)    경기도
1        정    경기도 가평군 가평읍 대곡리 316  POINT (127.50722 37.82503)    경기도
2        정    경기도 가평군 가평읍 달전리 364  POINT (127.52045 37.81625)    경기도

        lat         lon
0  37.823665  127.516690
1  37.825032  127.507218
2  37.816248  127.520453
```

Figure 1: Dataset Structure – Electric Vehicle Charging Station Installation Status (first 3 rows)

For the traffic volume data, a full address was generated by combining the city, county, town, and neighborhood information. Nominatim geocoding was then used to obtain the latitude and longitude for each address; missing values were removed to secure candidate coordinates.

```
[3] 교통량 데이터 (traffic_merged)
----------------------------------------
형태: (3427, 18)
컬럼: ['sido_code', 'sigungu_code', 'emd_code', 'week_type', 'ALL_AADT', 'PSCR_AADT', 'BUS_AADT', 'FGCR_AADT', 'area_code', '시도', '중분류', '시군구', '소분류', '읍면동', 'full_address', 'coords', 'lat', 'lon']
샘플 데이터 (처음 3개행):
   sido_code  sigungu_code  emd_code week_type  ALL_AADT  PSCR_AADT  BUS_AADT  \
0      11000         11010   1101053   weekday     13404      11442       576
1      11000         11010   1101054   weekday      8751       7861       247
2      11000         11010   1101055   weekday      9707       8477       402

   FGCR_AADT  area_code     시도      중분류   시군구         소분류   읍면동    full_address  \
0       1385   11010530  서울특별시  11010.0  종로구  11010530.0  사직동  서울특별시 종로구 사직동
1        644   11010540  서울특별시  11010.0  종로구  11010540.0  삼청동  서울특별시 종로구 삼청동
2        828   11010550  서울특별시  11010.0  종로구  11010550.0  부암동  서울특별시 종로구 부암동

                          coords        lat         lon
0  (37.5738317, 126.97030410681309)  37.573832  126.970304
1            (37.58964, 126.98112)  37.589640  126.981120
2        (37.5949602, 126.9657272)  37.594960  126.965727
```

Figure 2: Dataset Structure – Traffic Volume Data (first 3 rows)

The electric vehicle registration data was processed to reflect the registration numbers for each region as of February 28, 2025.

```
[2] EV 등록 데이터 (ev_reg_dict)
----------------------------------------
형태: 딕셔너리, 항목 수: 17
항목 예시 (처음 5개):
  서울: 84172
  인천: 55919
  경기: 155103
  강원: 21287
  충북: 25932
```

<Figure 3: Dataset Structure – Electric Vehicle Registration Status Data (first 5 rows)

4.2 Preprocessing Process

The EV_Charging_geo data was loaded using GeoPandas, the coordinate system was standardized to EPSG:4326 (latitude/longitude), and the province/city information was extracted from the address to create a regional variable.

The electric vehicle registration status data was converted into a dictionary for calculating regional demand.

The traffic volume data was merged with area_code.csv, a full address in the format “city, county, town, neighborhood” was generated, and Nominatim geocoding was used to derive the latitude and longitude for use as candidate data.

```
----------------------------------------
충전소 지역별 개수 (상위 5개):
region
경기도       859
서울특별시    586
경상남도      418
경상북도      375
전라남도      266
Name: count, dtype: int64
```

```
[1] 교통량 평균(mean) 기준 상위 5개 지역
----------------------------------------
                mean      sum
region
서울     15011.835714  6304971
인천     11637.953846  1512934
경기     10891.197125  5304013
대구      9376.576923  1218955
부산      8680.375635  1710034

[2] EV 등록 데이터 상위 5개 지역
----------------------------------------
경기    155103
서울     84172
인천     55919
제주     49690
경남     48510
dtype: int64
```

Figure 3: Dataset Structure – Regional Statistics (first 5 rows)

4.3 Application of Statistical and Clustering Analysis

Calculation of the Demand Index:
The traffic volume and electric vehicle registration data were normalized and summed with equal weights to compute the demand index for each region. Based on this, the ideal number of charging stations was calculated by proportionally allocating the total number of 4,614 charging stations.

Calculation of the Supply-Demand Gap:
The difference between the ideal number of chargin g stations and the actual installations for each regio n was computed. Regions with a positive gap were identified as targets for additional installations.

Derivation of Additional Candidate Locations via K-Means Clustering:
For candidate coordinates from the traffic volume data in regions with a positive gap, K-Means clustering was performed using the normalized ALL_AADT value as a weight. The derived cluster centers were designated as candidate locations for additional installations.

4.4 Result Visualization and Analysis

Bar Chart Visualization:
Bar charts were used to provide a side-by-side comparison of the actual number of installed charging stations, the ideally allocated number, and the

supply-demand gap (Gap) for each region. These charts grouped the categories "Actual," "Ideal," and "Gap" to intuitively illustrate regional imbalances and clearly demonstrate the need for additional installations in regions such as Seoul, Gyeonggi, and Busan where the gap is significant.

Figure 4: Number of Charging Stations by Region – Actual vs. Ideal (Supply-Demand Gap) bar graph

Map Visualization:
Using Plotly Express and Plotly Graph Objects, existi ng charging stations across the nation (blue dots) a nd the proposed additional installation candidate lo cations (red stars) were displayed on a map.

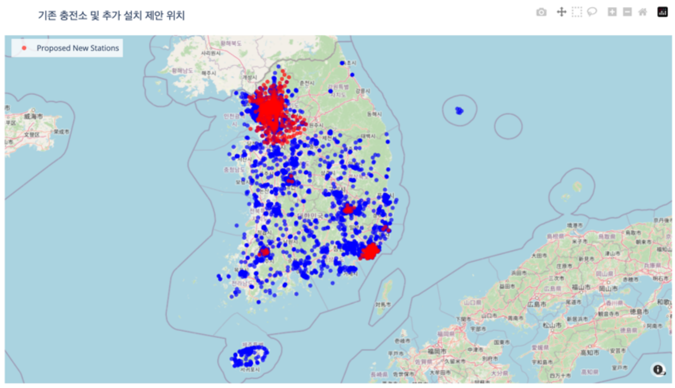

Figure 5: Existing Charging Stations and Proposed Additional Installation Locations (Red: Proposed Additional Installation Locations, Blue: Existing Charging Stations)

Summary of Experimental Results:
A significant supply-demand gap was observed in m ajor urban areas (Seoul, Gyeonggi, Busan), and a tot al of 703 additional candidate locations were derive d for regions with a positive gap. These candidate l ocations, effectively determined in areas of concentr ated demand, reflect the urgency for additional inst allations.

```
==================================================
Supply-Demand Gap Analysis (수치):
               Ideal  Actual         Gap
강원              NaN     NaN         NaN
강원특별자치도        NaN   261.0         NaN
경기      1081.118274   859.0  222.118274
경남       269.977662   418.0 -148.022338
경북       221.798000   375.0 -153.202000
광주       114.643513   108.0    6.643513
대구       247.822454   237.0   10.822454
대전       142.456049   139.0    3.456049
부산       330.681517   195.0  135.681517
서울       905.814390   586.0  319.814390
세종        30.782564    72.0  -41.217436
울산        65.431535    62.0    3.431535
인천       352.586562   190.0  162.586562
전남       164.106548   266.0 -101.893452
전북       157.418549     NaN         NaN
전북특별자치도        NaN   205.0         NaN
제주       205.407299   215.0   -9.592701
충남       185.247902   213.0  -27.752098
충북       138.707181   213.0  -74.292819
```

Figure 6: Number of Charging Stations by Region – Actual vs. Ideal (Supply-Demand Gap) numerical values

5. Experimental Evaluation

5.1 Quantitative Evaluation

1. Evaluation of Regional Supply-Demand Gap Figures:

Seoul: The ideal number of charging stations was a pproximately 905.81; however, only 586 stations wer e actually installed, resulting in a shortfall of 319.81.

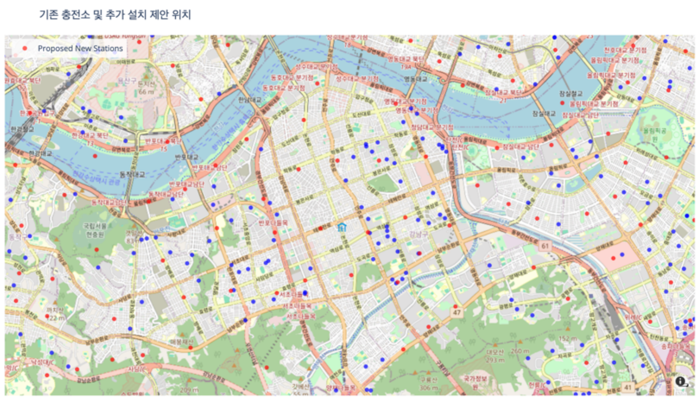

Figure 6: Existing Charging Stations and Proposed Additional Installation Locations in the Gangnam Area (Red: Proposed Additional Installation Locations, Blue: Existing Charging Stations)

Gyeonggi Province: Although the ideal number was about 1,081.12, only 859 charging stations were installed, producing a gap of 222.12.

Busan: The ideal figure was 330.68 compared to an actual 195, with a gap of approximately 135.68.

In some regions (e.g., Gyeongnam, Gyeongbuk, and Jeonnam), the actual installations exceeded the ideal numbers, resulting in negative gaps; this may indicate over-installation or issues in regional classification. Overall, the aggregated data confirmed that the nationwide total of KEPCO's electric vehicle charging stations was 4,614 and that the ideal allocation corresponded exactly to the total number of stations.

2. Number of Additional Candidate Locations Derived:

For regions with a positive supply-demand gap, the application of the K-Means clustering technique yielded a total of 703 candidate locations for additional installations. The number of charging stations needed additionally in each region (e.g., approximately 320 in Seoul, 222 in Gyeonggi, etc.) was calculated as rounded integer values, which then dictated the derived candidate locations.

3. Effectiveness of Weight Incorporation and Clustering:

The demand weights applied in the K-Means clustering were based on the normalized "Total Annual Daily Traffic" (ALL_AADT) values of each candidate coordinate. This led to the derivation of numerous additional candidate locations in high-demand areas such as Seoul and Gyeonggi, indicating the effectiveness of the weight-based clustering approach.

5.2 Qualitative Evaluation

1. Evaluation of Map and Visualization:

The map visualization clearly distinguishes existing electric vehicle charging stations (blue dots) from candidate additional locations (red stars). Particularly, the concentration of candidate locations in Seoul and Gyeonggi visually confirms the need for additional installations in regions with a high supply-demand gap.

2. Practicality and Limitations of Candidate Locations:

Since the proposed candidate locations were derived based on coordinates at the town/township/neighborhood level using traffic volume data, the detailed spatial distribution may be insufficient in some areas. Furthermore, in certain regions (e.g., Gangwon and Jeonbuk), NaN values occurred due to limitations in data mapping, indicating that improvements in regional classification are necessary.

3. Model and Data Improvement Suggestions:

Based on the quantitative results and visual evaluations, it was determined that the resolution of the candidate data influenced the derivation of additional candidate locations. In the future, more realistic candidate locations for additional installations should be derived by expanding and refining candidate coordinates using grid-based methods or Point of Interest (POI) data.

5.3 Overall Evaluation

The experimental evaluation of this study can be summarized as follows:

Quantitatively, the supply-demand gap figures and the number of candidate locations derived through K-Means clustering effectively captured the overall imbalance in charging station installations.

Qualitatively, the map visualizations and bar charts clearly identified key regions—such as Seoul, Gyeonggi, and Busan—where additional installations are urgently needed, confirming the effectiveness of the model's approach to incorporating demand.

Nonetheless, limitations such as the resolution of candidate data and data mapping errors in some regions were also identified as areas needing improvement. Overall, this study demonstrates the validity of the statistical and spatial analysis methodology for optimizing KEPCO's nationwide electric vehicle charging stations, suggesting that the proposed method can provide a substantial basis for policy decision s upport and infrastructure expansion.

6. Conclusion and Future Research

6.1 Conclusion

To address the challenges of building charging infrastructure in line with the expanded adoption of electric vehicles, this study utilized KEPCO's electric vehicle charging station installation data, the 2021 traffic volume data (Total Annual Daily Traffic, ALL_AADT), and the electric vehicle registration status data from 2025 to conduct a demand analysis and optimal allocation of electric vehicle charging infrastructure nationwide.

By calculating a demand index using normalized traffic volume and electric vehicle registration data for each region, allocating the ideal number of charging stations, and computing the supply-demand gap against the actual installations, a shortage of charging stations was identified in densely populated areas (Seoul, Gyeonggi, Busan). In regions with a positive gap, a total of 703 additional candidate locations were derived through K-Means clustering that incorporated candidate coordinates based on traffic volume data and demand weights. This result offers substantial information for addressing overall charging infrastructure imbalances and supporting policy decisions.

6.2 Future Research

While this study produced meaningful results, the fo llowing improvements and expansions are necessary:

1. Improvement of Candidate Data Resolution:

To overcome the limitations of candidate coordinates provided at the town/township/neighborhood level, it is necessary to subdivide candidate coordinates by incorporating grid-based methods or road network/Point of Interest (POI) data.

2. Development of a Composite Demand Weight Model:

Beyond traffic volume and electric vehicle registratio n data, a multi-factor demand model that integrates additional factors—such as population density, com mercial activity, and mobility patterns—should be int roduced to more accurately forecast charging statio n installation demand.

3. Application of Advanced Optimization Techniques:

In addition to simple K-Means clustering, advanced optimization models such as p-Median, p-Center, and MCLP could be integrated to develop a multi-objective optimization model that simultaneously considers installation costs, accessibility, and the impact on the power grid.

References

1. Capar, I., Kuby, M., Leon, V. J., & Tsai, Y. (2013). An arc cover-path-cover formulation and strategic analysis of alternative-fuel station locations. *European Journal of Operational Research, 227*(1), 142–151. https://doi.org/10.1016/j.ejor.2012.11.042
2. Chen, T. D., Kockelman, K. M., & Khan, M. (2013). The electric vehicle charging station location problem: A parking-based assignment method for Seattle. *Transportation Research Record, 2385*(1), 28–36. https://doi.org/10.3141/2385-04
3. Chung, S. H., & Kwon, C. (2015). Multi-period planning for electric car charging station locations: A case of Korean Expressways. *European Journal of Operational Research, 242*(2), 677–687. https://doi.org/10.1016/j.ejor.2014.10.029

4. Felipe, A., Ortuño, M. T., Righini, G., & Tirado, G. (2014). A heuristic approach for the green vehicle routing problem with multiple technologies and partial recharges. *Transportation Research Part E: Logistics and Transportation Review, 71*, 111–128. https://doi.org/10.1016/j.tre.2014.09.003
5. Hodgson, M. J., Rosing, K. E., & Zhang, J. (1996). Locating vehicle inspection stations to protect a transportation network. *Geographical Analysis, 28*(4), 299–314. https://doi.org/10.1111/j.1538-4632.1996.tb00935.x
6. Lee, C., Han, J., & Choi, K. (2014). GIS-based optimal site selection for electric vehicle charging stations: A case of Seoul. *Journal of the Korean Urban Geographical Society, 17*(3), 1–12.
7. Park, S., Lee, H., & Kim, J. (2020). Multi-objective optimization of electric vehicle charging stations considering power grid constraints. *Energies, 13*(15), 3916. https://doi.org/10.3390/en13153916
8. Wi, Y. M., Lee, J. U., & Joo, S. K. (2015). Electric vehicle charging station placement using fuzzy clustering. *Journal of Electrical Engineering & Technology, 10*(3), 985–992. https://doi.org/10.5370/JEET.2015.10.3.985
9. Xiong, Y., Wang, B., Chu, C. C., & Gadh, R. (2018). Electric vehicle driver clustering using statistical model and machine learning. *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. https://doi.org/10.1109/PESGM.2018.8586554
10. Zhang, L., Shao, C., & Yang, F. (2017). A bi-level optimization model for charging station placement with EVs' spatial charging flexibility. *IEEE Transactions on Smart Grid, 8*(6), 2595–2605. https://doi.org/10.1109/TSG.2016.2521178