ISSN 1229-1889(Print)
ISSN 2287-9005(Online)
*J Korean Soc Qual Manag* Vol. 51, No.4: 497-514, December 2023
https://dx.doi.org/10.7469/JKSQM.2023.51.4.497

# 서울시 공영주차장 군집화 및 수요 예측

황정준* · 신영현** · 심효섭** · 김도현* · 김동근**†

* 명지대학교 산업경영공학과
** 서울연구원

# Clustering of Seoul Public Parking Lots and Demand Prediction

Jeongjoon Hwang* · Young-Hyun Shin** · Hyo-Sub Sim** · Dohyun Kim* · Dong-Guen Kim**†

* Department of Industrial and Management Engineering, Myongji University
** The Seoul Institute

## ABSTRACT

**Purpose:** This study aims to estimate the demand for various public parking lots in Seoul by clustering similar demand types of parking lots and predicting the demand for new public parking lots.

**Methods:** We examined real-time parking information data and used time series clustering analysis to cluster public parking lots with similar demand patterns. We also performed various regression analyses of parking demand based on diverse heterogeneous data that affect parking demand and proposed a parking demand prediction model.

**Results:** As a result of cluster analysis, 68 public parking lots in Seoul were clustered into four types with similar demand patterns. We also identified key variables impacting parking demand and obtained a precise model for predicting parking demands.

**Conclusion:** The proposed prediction model can be used to improve the efficiency and publicity of public parking lots in Seoul, and can be used as a basis for constructing new public parking lots that meet the actual demand. Future research could include studies on demand estimation models for each type of parking lot, and studies on the impact of parking lot usage patterns on demand.

**Key Words:** Parking Demand Prediction, Time-series Clustering, Quantile Regression

● Received 18 September 2023, 1st revised 6 October 2023, accepted 10 October 2023
† Corresponding Author(dgkim@si.re.kr)
ⓒ 2023, Korean Society for Quality Management
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
* 이 논문은 2022년도 서울연구원 연구과제 "빅데이터를 활용한 공영주차장 유형 분류 및 수요추정 방법론 연구[서울연 2021-BR-43]"를 바탕으로 작성되었음.

# 1. 서 론

## 1.1 연구의 배경 및 목적

주차문제는 많은 도시에서 겪고 있는 교통문제 중 하나이다. 인구가 밀집된 대도시에서는 차량 증가에 의한 주차 공간 부족 문제가 빈번히 발생하여 많은 문제를 야기한다(Son et al., 2020). 이러한 주차문제를 해결하기 위해서는 적절한 위치에 충분한 주차장이 공급되어야 하지만 건설비용, 부지확보 등을 고려할 때 쉬운 일은 아니다 (Shin and Lee, 2021). 서울시는 주차문제를 해결하기 위하여 다양한 정책과 신규 주차장 건립을 추진해 왔다. 주차장에 대한 설치기준을 정비하고 부설주차장 설치를 의무화하는 등 주차장 공급을 관리하는 정책들(Kim, 2016)과 더불어 노상, 노외, 시설 등 다양한 주차 시설을 건립하여 운영하고 있지만, 주차공급의 부족으로 인한 불법 주차는 줄어들지 않고 있다. 주차장의 수요는 지역 거주자, 상업시설 방문객, 대중교통 환승 이용자 등 이용 목적에 따라 주차 패턴의 유형이 상이하며, 계절, 주중/주말, 주변 상권/주택가 현황, 인구 등에 따라 다르게 나타난다. 특히, 서울시는 시간대와 지역별로 유·출입 교통량에 큰 차이가 발생(Park et al., 2017)하기 때문에 합리적인 주차장 공급을 위해서는 다양한 유형을 고려한 수요 분석이 필요하다 (Kim et al., 2017). 그러나 현재 주차장 공급은 주변 지역의 인구, 토지이용 및 교통계획 특성, 교통 수요 등과 같이 주차장 건설에 고려되어야 할 요인들이 제대로 반영되지 못하고 있는 실정이다(Kim et al., 2017). 따라서 본 연구에서는 서울시 공영주차장을 대상으로 정확한 수요를 예측하기 위해 주차장들의 수요패턴을 유형별로 군집화하고, 군집화 결과와 함께 주차수요와 관련된 다양한 요인의 데이터를 활용하여 공영주차장의 주차 수요 예측 모델을 만들고자 한다.

논문의 흐름은 다음과 같다. 먼저 주차수요 및 군집분석을 활용한 선행연구와 방법론에 대한 검토를 수행한다. 다음으로 공영주차장의 주차실적을 확인할 수 있는 자료인 '실시간 입·출차 기록' 자료와 '실시간 주차가능면 수' 자료를 수집하고 두 데이터 간의 정합성을 확인한다. 다음으로 현장조사를 통해 데이터의 신뢰성을 판단한 후, 분석 대상 주차장을 선정하고 선정된 공영주차장에 대한 10분 단위 실시간 주차점유율 자료를 작성한다. 이를 바탕으로 시계열 군집분석을 수행한 후 영향권 내 용도지역 면적과 비교하여 군집분석의 결과를 검증한다. 검증된 군집분석과 함께 주차수요와 관련된 주차실적 데이터, 상권분석 데이터, 용도지역별 면적 데이터, 기타 영향권 데이터를 함께 활용하여 기계학습 기반의 회귀분석을 통해 공영주차장의 수요 예측을 수행한다. 마지막으로 해당 결과를 바탕으로 본 연구의 결론과 시사점을 제시하고자 한다.

# 2. 선행연구

연구의 목적을 고려하여 먼저 '주차 수요추정에 관한 연구'를 살펴보고자 한다. Park and Ha(2004)는 공동주택단지의 주차문제를 해결하기 위하여 아파트 규모를 세 가지로 구분하고 규모에 따른 주차수요를 분석하였으며, 주차수요와 주차장 공급 기준의 불균형이 주차문제를 발생시킨다는 것을 밝혀냈다. Shin and Lee(2021)의 연구에서는 토지용도별(주거지역, 상업지역, 녹지지역, 공공시설) 면적을 고려하여 주차수요를 분석하였고, 상업지역 면적이 주차수요와 유의한 상관관계를 가지며 주거지역 면적과의 상관계수는 주간과 비교해 야간에 증가한다고 밝혀냈다. Choi and Kim(2017)의 연구에서는 공영주차장 주차 회전율에 영향을 줄 것으로 예상되는 주요 변수들을 독립변수로 다중회귀분석 모형을 시행하여 영향권의 대중교통 편의지수 및 불법 주·정차지수, 중심상업지역의 토지이용 특성이 주

차 회전율과 높은 영향 관계가 있는 것을 밝혀내고 수요예측모형을 제시하였다. 또한, Cho and Cho(2020)에서는 종합병원의 주차장 입·출차 데이터를 기반으로 주차특성지수 분석 모델과 병원 특성을 고려한 주차수요 추정 모델을 제시하였다. Shin et al.(2020)에서는 주차수요의 시·공간상 분포를 살펴보고, 특정 지역에 대한 시간대별 주차수요 패턴을 세 가지로 유형화하여 주차수요 패턴별 분산 가능성을 제시하였다. 그러나 본 연구처럼 주차 수요 패턴 뿐만 아니라 다양한 주차수요와 관련된 다양한 요인의 빅데이터를 동시에 활용하여 주차 수요예측모델을 개발한 사례 및 연구는 거의 찾아보기 힘들다.

본 연구에서 다루고자 하는 예측을 위해 군집분석을 활용한 연구는 경제·금융 분야, 기상관측 분야, 의학 분야 그리고 제조 분야 등 많은 분야에서 활발히 진행되고 있다(Kim and Park, 2018; Lee and Kim, 2022). Kim and Kim (2021)에서는 많은 양의 전력 데이터를 전력수요 패턴별로 군집화한 후 수요예측모델을 제시하였으며, Park et al.(2022)에서는 광양만권 내 17개 도시 대기측정소의 관측자료와 2개 광화학 측정소의 자료를 이용하여 오존 경보제 기간 동안의 일 평균농도, 일 최대 농도를 기준으로 군집화하여 군집별 원인을 분석하고 군집별 저감방안을 제시하였다. 그러나 앞서 살펴본 바와 같이 선행연구에서는 특정 용도지역이나 시설, 전기, 대기 등 다양한 분야에서 빅데이터를 활용한 수요추정 및 군집분석에 관한 연구(Kim and Sohn, 2020)가 진행되었으나 주차장 유형을 고려한 수요패턴 및 군집분석에 관한 연구는 부재한 상황이다. 따라서 본 연구에서는 서울시에서 관리하는 '실시간 주차가능면 수' 입·출차 데이터 활용하여 주차점유율에 대한 패턴을 확인하고 시계열 군집분석을 통해 주차장 유형을 구분한다. 이 때 활용되는 주차점유율은 시간대별 주차대수/주차용량(=총 주차면 수)을 나타내는 것으로 주차정보를 분석하는 데 있어 고려해야 할 주차 특성 중 하나이다(Lee et al., 2008).

# 3. 분석방법론

군집분석은 전체 데이터를 몇 개의 집단으로 그룹화한 후 집단의 성격을 파악하여 데이터 전체의 구조에 대한 이해를 돕는 분석법이다. 군집분석은 각 객체 간의 유사도 또는 비유사도를 기반으로 서로 유사한 객체를 같은 그룹에 할당하여 분석하며, 모집단 또는 범주에 대한 사전 정보가 없는 경우 주어진 개체 사이의 거리나 유사성을 이용하여 분석한다(Kim et al., 2021). 군집분석은 크게 계층적 군집분석과 비계층적(분할적) 군집분석으로 나뉘며, 계층적 군집분석은 처음에 n개의 군집으로부터 시작하여 점차 군집의 개수를 줄여나가는 방법이고, 비계층적 군집분석은 군집의 계층을 고려하지 않고 평면적으로 군집을 형성하는 방법이다.

본 연구에서는 실시간 주차대수 자료가 시계열 자료인 점을 고려하여 비계층적 군집분석 중 시계열 군집분석에 주로 활용되는 K-means와 K-shape 시계열 군집분석방법을 사용하였다. K-means 시계열 군집분석은 미리 군집의 수(K)를 정하고 각 군집의 중심과 군집 내의 시계열 데이터들과의 거리의 제곱합을 최소화하는 방법론으로 관측치의 유사도를 측정하기 위해 다양한 거리 척도를 활용하는데 Euclidean distance(유클리드 거리), Dynamic Time Warping distance(DTW, 동적 시간 와핑 거리), soft-DTW 등이 있다. K-shape 시계열 군집분석은 K-means 시계열 군집분석과 유사하지만, 형태에 기반을 둔 거리측정 방식(shape-based distance measure)과 중심 계산법을 사용한다는 측면에서 차이가 있다.

그리고 주차수요 예측을 위해서는 공영주차장의 실제 주차수요 값을 목표변수로 활용하는 지도학습 기반의 회귀분석을 수행하였다. 시간대별 공영주차장의 수요의 편차가 크기 때문에 평균 수요를 예측하는 일반적인 회귀분석이 아닌 주차수요의 각 분위수를 독립변수로 추가하여 특정 분위수의 수요를 예측하는 회귀분석을 수행하였다. 회귀분석을 수행하기 위해 대표적인 회귀 알고리즘인 선형회귀(Linear Regression), LASSO(Least Absolute Shrinkage

and Selection Operator), 랜덤포레스트(Random Forest), GBR(Gradient Boosting Regressor), OMP(Orthogonal Matching Pursuit)를 사용하였다.

## 3.1 시계열 군집분석

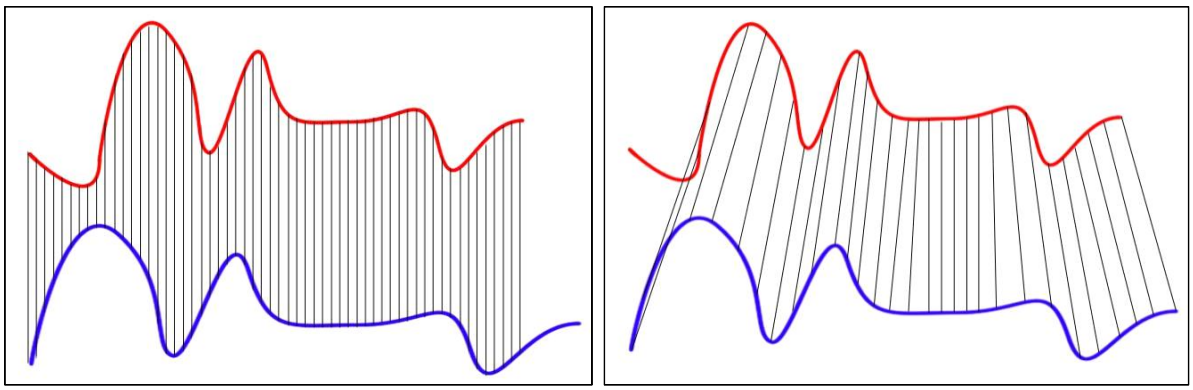
### (a) Euclidean K-means clustering

Euclidean K-means clustering은 K-means clustering 분석의 가장 대표적인 방법으로 각 군집의 중심점을 임의로 설정한 후 중심점 좌표와 데이터 사이의 유클리드 거리를 계산하여 가장 가까운 거리를 도출하여 군집에 재할당하고 이를 다시 반복 계산하여 모든 군집의 중심좌표가 사전에 정해진 오차 범위 내에서 수렴하거나 미리 정해진 반복 횟수를 만족하면 최종적으로 군집화 수행과정이 완료되는 분석법이다(Kim, 2017). 두 지점이 $(x_1, y_1)$, $(x_2, y_2)$일 때, Euclidean 거리는 식 (1)과 같이 계산할 수 있다.

$$d_{ij} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{1}$$

### (b) DTW(Dynamic Time Warping) K-means clustering

DTW(Dynamic Time Warping)는 주로 속도가 다른 유사 모양의 시계열 패턴의 유사성을 비교할 때 사용된다. 같은 시점의 거리를 계산하는 유클리드 거리는 속도를 반영할 수 없어 같은 패턴이어도 속도가 다르면 시작점을 다르게 인식하는 단점이 있다. 이를 보완하기 위한 것이 DTW이며, 그림 1과 같이 거리가 가장 짧은 시점(유사도가 높은 시작점)을 찾아 임의의 두 시점에 대한 최소거리를 계산하는 방법이다. 또한, 모든 시간대를 고려하기 때문에 관계없는 시간대까지 알고리즘 결과에 반영할 수 있다(Yang et al., 2017).



(a) Euclidean Matching (b) Dynamic Time Warping Matching

**Figure 1.** Comparison of Euclidean and DTW distance (Yang et al., 2017)

### (c) Soft-DTW K-means clustering

Soft Dynamic Time Warping은 최적 경로(optimal path)만을 고려하고 미분할 수 없는 일반적인 DTW와는 다르게, 미분이 가능하도록 변형하면서 최적 경로 이외의 다른 경로들도 함께 고려하여 gradient 기반의 최적 경로를 탐

색하는 방법이다. DTW는 시계열 자료를 비교하는 과정에서 거리가 가장 짧은 임의의 두 시점에 대한 최소거리를 계산하는 반면, Soft-DTW는 다른 경로도 함께 고려하기 때문에 최적 경로 이외의 다른 경로도 분석할 수 있다.

(d) K-Shape clustering

K-shape 군집분석은 K-means 군집분석과 유사하게 반복적으로 중심(centroid)을 개선해나가는 알고리즘을 사용하면서도, K-means 군집분석처럼 유클리드 거리 기반의 거리측정 방식을 사용하지 않는다는 점이 다르다. K-shape 군집분석은 시계열 데이터에 적합하도록 두 시계열 간의 Time lag에 따른 내적값을 계산하는 Cross-correlation을 통해 시계열 형태에 기반을 둔 거리측정 방식(shape-based distance measure; SBD)을 사용한다.

언급된 거리측정 방식에 따른 세 가지 K-means 시계열 군집분석 방법과 K-shape 시계열 군집분석 방법, 총 네 가지 방법을 사용하여 주차장 점유율의 시계열 패턴에 대해 군집분석을 수행하고 이를 비교하고자 한다.

## 3.2 수요 예측을 위한 회귀모형

주차장 수요 예측모델 개발을 위해서 본 논문에서는 앞서 제시된 군집분석을 통해 도출된 주차장 유형정보와 주차수요와 관련된 다양한 변수들을 활용하여 특정 분위수의 주차장 수요를 예측하는 회귀분석 방법을 사용하고자 한다. 주차수요와 관련된 다양한 변수들에 대한 자세한 설명은 4.1절에서 다루고자 한다. 특정 분위수의 수요를 예측하기 위해, 최소제곱법을 사용하여 수요와 관련된 독립변수에 따른 종속변수의 조건부 평균을 추정하는 기존의 회귀분석과는 다르게, 분위수를 독립변수로 추가하여 회귀분석을 수행하였다. 분위수를 고려한 회귀분석을 사용한 이유는 다음과 같다. 첫째, 특정 분위수에 초점을 맞출 수 있으며, 이상치에 둔감하다. 평균 수요를 예측하는 일반적인 회귀분석의 경우, 이상치가 예측에 큰 영향을 주는 반면, 분위수를 고려한 회귀분석은 데이터의 특정 분위수에 초점을 맞추어서 모델링을 수행하기 때문에 극단값에 덜 민감한 모델을 개발할 수 있다. 두 번째, 데이터가 정규 분포를 따르지 않거나 비대칭적인 분포를 가지는 경우에 더 정확한 결과를 얻을 수 있다 (Ha et al., 2021). 따라서 수요의 편차가 큰 시간대별 공영주차장 데이터의 경우에 분위수를 고려한 회귀분석이 적합하다. 이러한 이유로 주차장 수요예측을 위해서 분위수를 고려한 회귀분석을 수행하였다. 본 논문에서는 대표적인 회귀 알고리즘인 즉, 선형회귀, LASSO, 랜덤포레스트, GBR, OMP을 이용하여 회귀분석을 수행하였다.

(a) 선형회귀(Linear Regression)

선형회귀(Linear Regression) 알고리즘은 하나의 종속 변수와 하나 이상의 독립 변수 간의 선형 관계를 모델링하는 회귀 분석 기법으로, 단순 선형 회귀(한 개의 독립 변수)와 다중 선형 회귀(둘 이상의 독립 변수)로 나눌 수 있다. 이 기법은 주로 예측 문제를 해결하는 데 사용되며, 종속 변수와 독립 변수 간의 선형 관계를 가정한다. 선형회귀는 선형 예측 함수를 사용하여 회귀 모델을 만들게 되며, 모델 파라미터는 데이터로부터 추정된다. 일반적으로 최소제곱법을 사용하여 관찰된 데이터와 회귀 모델 간의 오차를 최소화하도록 모델을 구성합니다. 이로써 관측 데이터와 가장 잘 맞는 선형 회귀 모델을 생성하게 된다.

(b) LASSO(Least Absolute Shrinkage and Selection Operator)

LASSO(Least Absolute Shrinkage and Selection Operator) 알고리즘은 기본적인 선형회귀 모델에서 평균제곱

오차(MSE)가 최소가 되는 회귀계수를 찾으면서 회귀계수의 절댓값의 합이 최소가 되도록 제약조건을 추가한 알고리즘이다. LASSO 알고리즘은 회귀계수에 대한 제약조건을 추가함으로써 전체 변수 중에서 중요한 변수를 선별하여 더욱 일반화된 모델을 제안하고, 모델에 사용된 변수들의 중요도를 파악하여 중요한 변수들을 선별해주는 모델 해석력을 제공하는 특징이 있다.

(c) 랜덤포레스트(Random Forest)

랜덤포레스트(Random Forest) 알고리즘은 여러 개의 의사결정나무(Decision Tree)를 활용하여 예측모델을 생성하는 알고리즘이다. 이 알고리즘은 분류 및 예측 문제 양쪽에서 활용 가능하며, 예측을 위해서는 다수의 트리의 예측 결과를 평균화하여 안정적인 최종 예측을 수행한다, 이를 통해 예측의 노이즈를 줄이고 신뢰성 있는 결과를 얻을 수 있다. 또한 랜덤포레스트는 독립 변수와 종속 변수 간의 비선형 관계나 복잡한 패턴을 모델링할 수 있으며, 데이터의 부트스트랩 샘플링과 독립 변수의 랜덤 선택을 통해 각 트리의 다양성을 증가시켜 과적합을 줄이고 일반화 성능을 향상시키게 된다.

(d) GBR(Gradient Boosting Regressor)

GBR(Gradient Boosting Regressor) 알고리즘은 앙상블 기법 중 부스팅(Boosting) 알고리즘의 한 가지 방법이다. 부스팅 알고리즘은 여러 개의 약한 학습기를 차례로 학습하여 이전 학습기의 예측 오차를 보완하는 방식으로 약한 학습기들을 결합하여 모델의 성능을 향상시킨다. GBR 알고리즘은 단계별로 예측 오차를 추정하여 추정된 오차의 값을 점진적으로 줄여나가는 알고리즘으로 예측값과 실제값의 차이를 단계적으로 개선하기 때문에 예측의 정확도가 매우 높은 알고리즘으로 다양한 분야의 데이터 분석에서 많이 활용되고 있다.

(e) OMP(Orthogonal Matching Pursuit)

OMP(Orthogonal Matching Pursuit) 알고리즘은 최적 해를 찾기 위해 각 단계에서 최적의 선택을 반복적으로 수행하는 greedy 알고리즘의 한 가지 방법으로, 최적 해를 찾기 위해 상관도가 높은 변수부터 단계적으로 반복하여 해를 찾아가는 방법이다. 반복을 통하여 중요 변수를 탐색하면서 사전에 설정한 오차 범위 내에 도달하게 되면 찾아낸 중요 변수를 모두 업데이트한 뒤 알고리즘을 종료한다.

# 4. 데이터 수집 및 실험결과

## 4.1 데이터 수집 및 정제

본 연구의 분석 대상 범위는 서울시 공영주차장의 노상주차장, 노외주차장, 시설주차장 중 주차데이터를 시스템상으로 관리하는 노외주차장과 시설주차장으로 설정하였다. 서울시 공영주차장의 주차실적을 확인할 수 있는 데이터는 '실시간 입·출차 기록' 자료와 '실시간 주차가능면 수' 자료가 있다. '실시간 입·출차 기록' 자료는 주차장 출입구에서 확인되는 차량의 입·출차 시간, 결제 금액, 이용형태(정기권, 시간권) 등을 저장한 자료로 공영주차장을 운영·관리하는 서울시설공단과 25개 자치구별 시설관리공단(도시관리공단 등)이 보유하고 있으며, 이 중 '실시간 입·출차 기록'은 주차장명, 차량번호, 주차구분, 입차일시, 출차일시, 주차시간, 과금금액, 결제금액, 출차구분으로 구성되어 있

다. 반면, '실시간 주차가능면 수' 자료는 해당 주차장의 규모, 실시간 점유면 수를 기록·저장한 자료로 서울시 주차계획과가 서울특별시 주차정보안내시스템에서 제공하고 있다. 해당 시스템에서 수집되고 있는 모든 주차장에 대해 주차장 정보와 전체 주차면 수를 제공하고 있으며, 실시간 연계서비스가 제공되는 주차장에 대해서는 실시간 주차 가능 면 수가 별도로 제공되고 있다. 그밖에도 전기차 충전이 가능한 주차장 정보 및 전체 주차면 수, 실시간 주차 가능 면 수에 대해서도 제공하고 있다. 해당 자료를 활용하기 위해서 각각 시설관리공단(Seoul Facilities Corporation)의 공영주차장 관리 담당자와 서울시 주차계획과의 주차정보안내시스템(Seoul Parking Information Guidance System) 관리 담당자와의 협의를 통해 자료를 제공받았다. 본 연구에서는 많은 주차데이터를 분석에 활용하기 위해 '실시간 입·출차 기록' 자료와 '실시간 주차가능면 수' 두 자료를 모두 수집하여 활용가능성에 대해 살펴보았다.

우선 두 데이터 간의 정합성을 살펴보았는데, '실시간 입·출차 기록'을 기준으로 주차장 내에 체류 중인 차량 수를 10분 단위로 구분하여 시간대별 점유면 수 데이터 형식으로 변환하고, 두 가지 주차실적 자료의 형식을 일치시킨 후 동일한 요일 및 시간대의 점유면 수를 비교하였다. 데이터 형식 변환을 통해 각 주차장의 시간대별 점유면 수를 비교한 결과, 그림 2와 같이 대부분의 주차장에서 '실시간 주차가능면 수' 자료의 주차점유면 수와 '실시간 입·출차 기록'의 주차점유면 수 차이가 크게 나타났다. '실시간 주차가능면 수' 자료를 이용한 주차점유면 수가 '실시간 입·출차 기록'의 점유면 수보다 높게 나타나 두 자료 간 정합성은 낮은 것으로 나타났다.

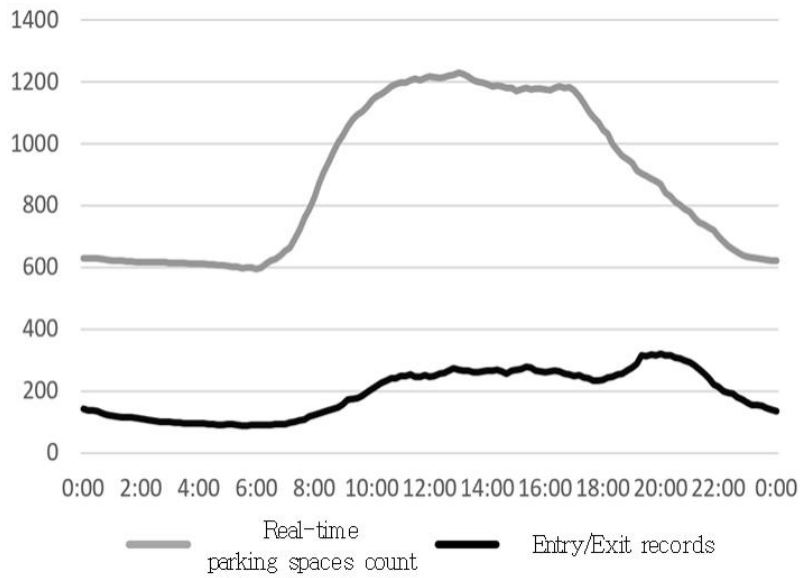



**Figure 2.** Number of parking spaces in ○○ public parking lot by time of day

이에 '실시간 입·출차 기록' 자료와 '실시간 주차가능면 수' 자료 중 실제값에 가까운 자료를 확인하기 위해 2022년 2월 21일(월)부터 2022년 2월 25일(금)까지 총 5일간 현장조사를 진행하였다. 기계식 주차장과 노상주차장을 제외한 총 131개 공영주차장을 대상으로 주차장별 조사시간과 실제 운영 중인 주차면 수, 당시 실제 주차점유면 수를 기록하고 동시에 주차정보안내시스템이 제공하는 '실시간 주차가능면 수' 자료를 수집하는 방식으로 진행하였다(표 1 참조). 앞서 정합성 검증에서 사용했던 방법과 마찬가지로 동일한 요일의 동일 시간대 기준으로 현장조사 결과와 기존 자료를 비교한 결과, '실시간 주차가능면 수 자료'가 '실시간 입·출차 기록 자료'보다 오차가 적은 것으로 조사되었다.

**Table 1.** Verification of reliability of parking performance data through field survey

| No. | Parking lot name | Number of parking spaces in information system (spaces) | Number of parking spaces in field survey (spaces) | Number of parking spaces (cars) | | | Difference in number of parking spaces from field survey (spaces) | |
|---|---|---|---|---|---|---|---|---|
| | | | | Entry/ Exit records | Parking information System | Field survey | Entry/ Exit records | Parking information System |
| 1 | ○○ | 65 | 61 | 15 | 31 | 34 | 19 | 3 |
| 2 | △△ | 41 | 41 | 20 | 32 | 33 | 13 | 1 |
| 3 | □□ | 79 | 79 | 36 | 51 | 52 | 16 | 1 |
| 4 | ◇◇ | 199 | 199 | 128 | 162 | 164 | 36 | 2 |
| 5 | ... | 93 | 93 | 42 | 48 | 48 | 6 | 0 |
| 6 | ... | 128 | 130 | 35 | 42 | 42 | 7 | 0 |
| 7 | ... | 237 | 324 | 53 | 76 | 76 | 23 | 0 |
| 8 | ... | 137 | 137 | 87 | 130 | 130 | 43 | 0 |
| 9 | ... | 172 | 172 | 92 | 109 | 109 | 17 | 0 |
| 10 | ... | 328 | 337 | 178 | 225 | 224 | 46 | 1 |
| . . | . . | . . | . . | . . | . . | . . | . . | . . |
| 131 | ... | ... | ... | ... | ... | ... | ... | ... |

데이터의 신뢰도가 상대적으로 높다고 판단된 '실시간 주차가능면 수 자료'에도 실제 값과 큰 오차를 나타내는 주차장이 일부 있는 것으로 확인되었다. 예를 들어, 현장조사 당시 주차장 일부가 코로나 선별진료소나 따릉이(서울시 공공자전거) 보관소로 이용되는 등 실제 주차현황과 데이터 간의 오차가 큰 공영주차장이 존재하였으며, '실시간 주차가능면 수'가 원활히 업데이트되지 않거나 특정 기간에 시설 유지보수 등의 사유로 주차점유면 수 데이터가 큰 폭의 변화를 보이는 자료도 존재함을 확인하였다. 따라서 현장조사를 통해 실시간 주차점유면 수를 확인한 131개 공영주차장 중 시스템 오류 및 시설 점검 등의 사유로 자료의 일관성이 부족한 공영주차장 자료 20개는 제외하고, 나머지 공영주차장 중 전체 주차면 수 대비 오차가 20% 이하인 주차장 68개를 분석 대상으로 선정하였다. 68개 주차장의 목록은 표 2와 같다.

**Table 2.** List of 68 parking lots

| Category | Parking lot name |
|---|---|
| District (29 locations) | Nam-bu women's development center, Daerim parking lot (Outdoor), Dongjak bridge, Sadang parking lot (Outdoor), Shinbanghwa station, Sinseol-dong, Cheonho station, Gaehwa station, Guro digital complex station, Gupabal station, Dobongsan station, Dongdaemun, Boramae commercial area, Bokjeong station, Bonghwasan station south, Seoul global center, Suraksan station, Suseo station south, Suseo station north, Sincheon Yusuji, Yeongdeungpo-gu office station, Jamsil station, Jongmyo parking lot, Changdong station west, Cheonwang station, Hak Yeoul, Hangangjin station, Hwarangdae station, Training institute park, |
| City (39 locations) | Gaepo-dong park, Nonhyeon elementary school, Daerim 1-dong, Daerim 2-dong, Daerim children's park, Daeji market street, Daecheong station, Daechi 2 cultural center, Dogok-ro 21-gil 7, Dorim elementary school, Dongsan village, Macheon-dong 2, Macheon-dong 3, Banpo-dong noise barrier hill type, Bangbae 1-dong, Samsung 1 cultural center, Songjung-dong, Singil 4-dong 1 village park, Sinwol 3-dong Yeoul, Yeoksam 1 cultural center, Yeongdeungpo Main-dong 2, Yeongdeungpo girls' high school, Yeonghui elementary school, Cheonma, Tanchun 2nd street, Daewang elementary school, Dogok elementary school, Mullae neighborhood park, Mullae-dong, Banpo 4-dong community center, Bammogae-ro 21-gil, Seosomun behind Samsung life, Songpa-dong, asia park, Yangpyeong 2-dong office, Eonnam cultural sports center, Eonbuk elementary school, Yeongdeungpo-dong 3, Poi elementary school |

분석 대상으로 선정된 68개 공영주차장에 대해 6개월간의 시간대별(10분 단위, 26,064개) 주차점유면 수를 그래프로 작성한 후 시간대별 주차점유면 수 패턴을 확인하였다. 작성된 그래프를 검토하여 국경일, 명절 연휴, 주차장 공사 등 일반적이지 않은 패턴이 포함된 기간을 제외한 시간대별 주차점유면 수 패턴을 확인하고 해당 주차장을 대표하는 대표기간(1주일, 1,008개)를 선정하였다. 주차장별로 선정한 1주일간의 10분 단위 주차점유면 수 데이터는 1,008개씩이고 68개 주차장을 대상으로 총 68,544개의 데이터를 추출하였다.

주차장 수요 예측을 위해 추가적으로 사용한 데이터로는 상권분석 데이터, 용도지역별 면적 데이터, 기타 영향권 데이터가 있다. 공영주차장 주변 지역의 상권분석 데이터는 서울신용보증재단의 전문가 상권분석시스템의 '정책 활용 서비스'와 가상화 공간데이터를 이용하여 수집하였다. '정책 활용 서비스' 중에서는 상권에 대한 인구 현황, 사업자현황, 업종매출분석, 상권특성분석 등 종합정보를 제공하는 상권 영향력 평가의 데이터를 활용하였다. 반면, 가상화 공간은 사용자가 목적(대상, 범위, 기간 등)에 따라 빅데이터를 직접 도출하고 분석하는 서비스이다. 본 분석을 위해서는 자치구, 행정동, 블록 등 공간 범위를 설정할 필요가 있기 때문에, 가장 작은 단위인 블록 단위에서 주차장 영향권을 반경 300m로 설정하고, 반경내 상주인구, 가구소득, 유동인구, 카드매출 등의 정보를 활용하였다.

또한, 공영주차장 주변 지역의 영향권 내 토지의 용도지역별 면적 데이터는 국토교통부의 토지 이음에서 제공하는 용도지역 데이터를 기준으로 수집하고 대상 주차장을 중심으로 반경 300m 내 토지의 용도지역별 면적 데이터를 사용하였다. 수집한 결과 전체 공영주차장의 반경 300m 내 용도지역별 면적 분포 현황은 주거지역(66.7%), 상업지역(13.8%), 녹지지역(12.9%), 공업지역(6.6%) 순으로 면적비율이 나타났다.

앞서 수집한 데이터 이외에 공영주차장 주차수요에 영향을 미칠 수 있는 기타 영향권 데이터를 추가로 수집하였다. 기타 영향권 데이터로는 주차장 별 대중교통 시설 분포 현황, 주차장별 주변의 경쟁주차장 분포 현황, 주차장별 주차요금 현황을 추가로 수집하였다. 대중교통 시설 분포 현황은 주차장별 영향권 300m 반경을 기준으로 포털 사이트에서 제공하는 지도를 활용하여 수집하였고, 경쟁주차장 분포 현황은 부설주차장을 제외한 공영 또는 민영 주차장

을 대상으로 집계하고, 반경 600m 내의 주차장 개소 수와 주차면 수를 집계하였다. 주차장 수요 예측을 위해 회귀분석에 활용한 독립변수에 대한 상세한 정보는 표 3과 같다.

**Table 3.** Collected total independent variables

| Category | | Independent Variables |
|---|---|---|
| Parking lot data | | Parking lot size |
| Commercial analysis data | Virtual space | BL area, Total worker, Male worker, Female worker, Day floating population, Night floating population, Total households, Total population, Male population, Female population, BC affiliate, BC sale number, BC sales, BC individual sale number, BC individual sales, BC corporation sale number, BC corporation sales, BC male sale number, BC male sales, BC female sale number, BC female sales, SH affiliate, SH sale number, SH sales, SH individual sale number, SH individual sales, SH corporation sale number, SH corporation sales, SH male sale number, SH male sales, SH female sale number, SH female sales, Total household income, Male household income, Female household income |
| | Policy utilization services | Population policy, Household policy, Street floating population policy, Building floating population policy, Worker policy, Tradi-market policy, Tradi-market area policy, Big mart policy, Department policy, Shopping center policy, Shop number policy, Fran shop(restaurant) number policy, Normal shop(restaurant) number policy, Fran shop(service) number policy, Normal shop(service) number policy, Fran shop(retail) number policy, Normal shop(retail) number policy, Open shop(restaurant) number policy, Close shop(restaurant) number policy, Open shop(service) number policy, Close shop(service) number policy, Open shop(retail) number policy, Close shop(retail) number policy, Open/Close shop number policy, Sale(restaurant) number policy, Sale(retail) number policy, Sales(restaurant) policy, Sales(service) policy, Sales(retail) policy, Stations policy, Bus stops policy |
| Land use zone data | | Housing area, Commercial area, Industrial area, Green area, Type 1 exclusive housing area, Type 2 exclusive housing area, Type 1 general housing area, Type 2 general housing area, Type 3 general housing area, Semi housing area, Central commercial area, General commercial area, Neighborhood commercial area, Distribution commercial area, Exclusive industrial area, General industrial area, Semi industrial area, Preserved green area, Agricultural green area, Natural green area |
| Cluster analysis data | | Cluster Info by Euclidean K-means (K=4) |
| Quantile data | | Quantile |

## 4.2 시계열 군집분석 결과 및 검증

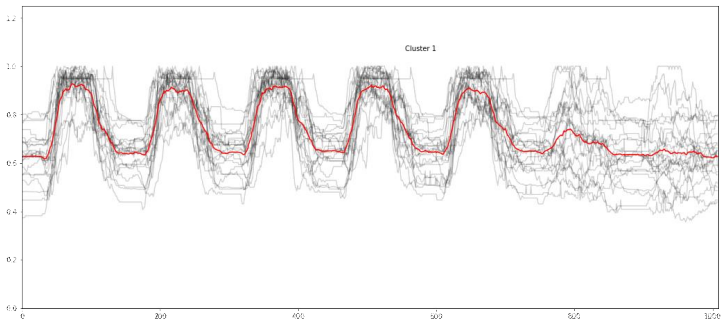
시계열 군집분석을 위해 먼저 각 주차장의 시간대별 주차실적 데이터에서 68개 주차장별 규모 차이를 보정하고자 피크 대비 점유율로 표준화(Normalization)를 진행하였다. 군집의 개수를 2개, 3개, 4개로 설정하여 각각 네 가지 분석방법론별로 총 12번의 군집분석을 수행하였고 군집분석의 수행결과는 표 4와 같다. 시계열 군집분석을 위해서는 파이썬 시계열 분석 패키지인 ts-learn 패키지를 활용하였다.

**Table 4.** Time series clustering results

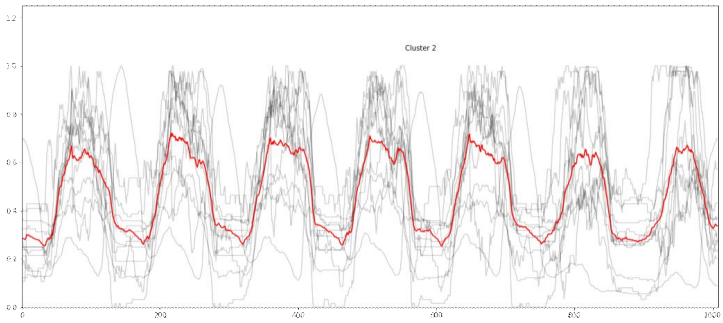
(Unit: Count)

| Category | | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Total |
|---|---|---|---|---|---|---|
| Clustering Method | Cluster Number(K) | | | | | |
| Euclidean K-means | 2 | 32 | 36 | - | - | 68 |
| | 3 | 22 | 34 | 12 | - | 68 |
| | 4 | 22 | 12 | 12 | 22 | 68 |
| DTW K-means | 2 | 33 | 35 | - | - | 68 |
| | 3 | 19 | 35 | 14 | - | 68 |
| | 4 | 19 | 21 | 14 | 14 | 68 |
| soft DTW K-means | 2 | 37 | 31 | - | - | 68 |
| | 3 | 31 | 4 | 33 | - | 68 |
| | 4 | 17 | 4 | 26 | 21 | 68 |
| K-shape | 2 | 32 | 36 | - | - | 68 |
| | 3 | 12 | 31 | 25 | - | 68 |
| | 4 | 23 | 26 | 2 | 17 | 68 |

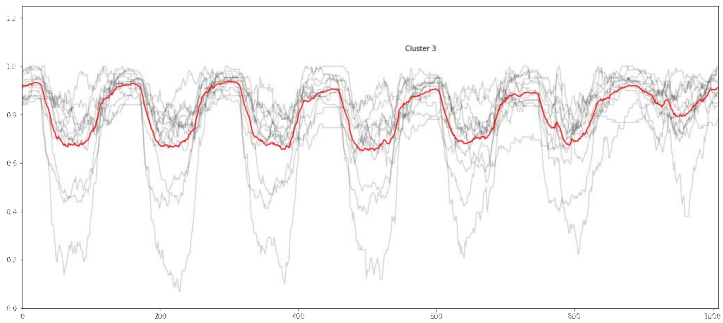
군집별 주차장 개수를 살펴보면, 군집 개수가 하나 늘어날 때 기존 군집 중 하나가 둘로 분화되는 경향을 나타냈다. 방법론별로 군집 수가 2개일 때와 3개일 때를 비교하면, 군집 수가 2개일 때의 군집 중 하나의 군집이 두 개로 분화되어 3번째 군집을 형성하는 경향을 보였으며, 유사하게 군집 수가 3개일 때의 군집 중 하나의 군집이 두 개로 분화되어 4번째 군집을 형성하는 것으로 나타났다. 이러한 군집 수 증가에 따른 변화 양상은 다른 군집분석 방법론에서도 유사한 형식의 변화를 나타냈다. 최적의 클러스터 개수를 탐색하기 위해 SSE(Sum Squared Error)값이 안정화된 순간의 군집 수를 정하는 Elbow method를 이용하였으며, 이를 통해 최적 군집 수를 4개로 결정하였다. 군집 수를 4로 설정하였을 때 Euclidean K-means 분석을 제외한 나머지 군집분석에서는 동일 군집 내에 피크 시간대가 다른 주차장이 섞여 있거나, 주중과 주말의 차이에서 유사성이 낮은 그래프가 섞여 있는 등의 모습이 관찰되었다. 그래프의 진폭과 주중·주말의 차이, 피크 시간대 등을 고려한 군집 내 유사성과 군집 간 차별성 등과 SSE 값을 종합적으로 검토한 결과, 군집 개수가 4인 Euclidean K-means 방법론의 군집분석 결과가 가장 설명력이 우수한 것으로 나타났다.



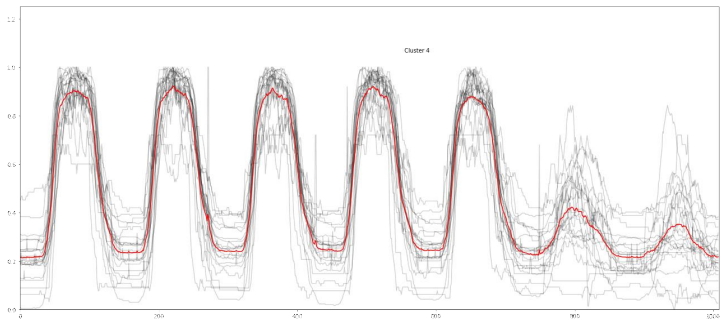

(a) Cluster 0 (Neighborhood type)




(b) Cluster 1 (Commercial type)




(c) Cluster 2 (Residential type)




(d) Custer 3 (Transfer type)

**Figure 3.** Euclidean K-means clustering results

주차장마다 시계열 패턴 간에 약간의 시차가 있을 것으로 판단되어 다양한 군집분석 방법을 제시하였으나, 실제 데이터에서 주차장 패턴 간의 시차가 거의 없어서 Euclidean K-means 군집분석의 결과가 가장 적합한 것으로 판단된다. 그림 3에서 Euclidean K-means(K=4) 군집분석으로 분류된 4개 군집의 그래프와 패턴을 살펴보면 군집 0은 점유율의 진폭이 작고, 상대적으로 주말의 점유율이 낮은 특성을 나타내 근린 유형의 주차장으로 판단된다. 군집 1은 점유율의 진폭이 상대적으로 큰 편이고, 주중과 주말의 점유율이 비슷하게 나타나 상업 유형의 주차장으로 분석된다. 또한, 군집 2는 밤 시간대에 점유율이 높고, 주중과 주말의 차이가 적어 주택가 유형의 주차장으로 판단되며, 군집 3은 점유율의 진폭이 크고 주중에 비해 주말에 점유율이 낮은 특성이 있어 환승 유형의 주차장으로 분석된다.

군집분석으로 도출한 네 가지 주차장 유형의 적합성을 검증하기 위해 해당 주차장의 입지특성을 나타낼 수 있는 용도지역과 비교하였다. Lee(2008)에서는 용도지역이 주차 특성의 차이를 나타내는 요인 중 하나이며 용도지역별

주차원단위도 차이가 있음을 밝혔다. 또한, Oh(2016)에서는 용도와 지역적 특성에 따라 공영주차장의 유형을 세 가지로 구분하여 각 유형에 대한 전략과 계획의 차별성에 의한 도심 활성화를 도모해야 한다고 제시하였다. 따라서 본 연구에서는 군집분석을 통해 분류된 주차장 유형이 주차장 입지특성 중 하나인 용도지역과 상관성이 있을 것으로 판단하여, 표 5와 같이 해당 주차장 영향권의 용도지역별 면적 분포를 통하여 네 가지 군집 유형을 검증하였다. 각 주차장의 영향권은 「주차장법」 제19조 제4항과 「동법 시행령」 제7조 제2항을 참고하여 주차장 반경 300m로 규정하였다. 68개 공영주차장의 영향권 내 용도지역별 면적 비율은 주거지역이 가장 높은 비율을 차지하고 있었으며, 상업지역, 녹지지역, 공업지역 순으로 나타났다. 특히, 주거지역과 상업지역은 전체의 3분의 2 이상을 차지하는 것으로 나타나 상세한 용도지역별 면적비율을 확인하여 유형에 대한 특성을 살펴보았다.

군집 0(근린 유형)은 제3종 일반주거지역과 근린상업지역 비율이 군집 중 가장 높고, 제2종 일반주거지역 비율도 비교적 높은 것으로 나타났다. 상대적으로 용적률이 낮은 제1종일반주거지역의 비율이 낮고 근린생활시설 설치가 가능한 제2종과 제3종 일반주거지역 비율이 높아 주거지와 근린상업지역이 많이 분포된 입지에 위치한 특성을 보이는 것으로 나타났다. 군집 1(상업 유형)은 상업지역과 공업지역 비율이 모두 평균 대비 높게 나타났으며, 영향권 내 상업지역은 모두 일반상업지역으로 상권이 발달한 지역 인근에 위치한 주차장의 특성을 보이는 것으로 나타났다. 군집 2(주택 유형)은 영향권 내 주거지역 비율이 87.5%로 가장 높았으며 특히 제1종일반주거지역과 제2종 일반주거지역의 비율이 다른 용도지역보다 상대적으로 높아 단독주택이나 소규모 아파트 등 주거지 인근에 위치한 주택 유형의 주차장 특성을 보이는 것으로 나타났다. 군집 3(환승 유형)는 녹지와 상업지역 비율이 높으며 주거지역 비율이 가장 낮은 것으로 나타났으며, 녹지지역 비율이 평균 대비 두 배가량 높으며, 주거지역 비율은 45.4%로 가장 낮은 것으로 나타나 주로 시 외곽이나 교통 요충지 인근에 위치한 특성을 보이는 것으로 확인되었다. 4개 군집에 대한 영향권 내 용도지역별 면적(비율)은 표 5와 같다. 군집분석으로 분류한 주차장 유형을 용도지역 면적비율과 비교·검토한 결과 상관성이 있는 것으로 확인되었다.

**Table 5**. Parking lot influence area land use area by type (Ratio)

| Category | Residential area | | | | Commercial area | | | | Industrial area | Green area | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Type 1 general | Type 2 general | Type 3 general | Other residential | Core commercial | General commercial | Neighborhood commercial | Retail commercial | | | |
| Cluster 0 (Neighborhood type) | 16,276 (5.8%) | 109,747 (38.8%) | 87,619 (31.0%) | 12,931 (4.6%) | - | 23,303 (8.2%) | 1,830 (0.6%) | - | 12,846 (4.5%) | 18,202 (6.4%) | 282,753 (100%) |
| Cluster 1 (Commercial type) | 37,410 (13.2%) | 67,995 (24.1%) | 43,314 (15.3%) | 21,199 (7.5%) | - | 55,959 (19.8%) | - | - | 36,806 (13.0%) | 20,088 (7.1%) | 282,770 (100%) |
| Cluster 2 (Residential type) | 37,227 (13.2%) | 153,719 (54.5%) | 42,124 (14.9%) | 13,866 (4.9%) | - | 3,457.4 (1.2%) | 1,347.9 (0.5%) | - | 1,088 (0.4%) | 29,504 (10.5%) | 282,333 (100%) |
| Cluster 3 (Transfer type) | 20,366 (7.2%) | 61,976 (22.0%) | 38,559 (13.7%) | 6,119 (2.2%) | - | 61,717 (21.9%) | - | 174 (0.1%) | 24,117 (8.6%) | 67,631 (24.0%) | 281,252 (100%) |

주: 각 영향권의 면적은 국토교통부의 토지 이음에서 제공하는 용도지역 데이터를 기준으로 수집하였으며, 일부 개발계획이 진행되는 면적은 누락된 부분이 있어 용도지역 전체 면적 합계가 상이함

## 4.3 회귀분석 결과

공영주차장 수요 예측을 위해 독립변수로 공영주차장 주변 지역의 상권분석 데이터, 공영주차장 주변의 용도지역별 면적 데이터 그리고 공영주차장 수요패턴의 군집 정보를 활용하였다. 사용한 회귀분석 알고리즘은 가장 기본적인 선형회귀모델, 선형회귀모델을 기반으로 모델의 단순화를 수행하는 LASSO, 의사결정나무 기반의 랜덤포레스트, 부스팅 기반의 앙상블 모델인 GBR, 변수의 수를 제어하여 모델의 단순화와 일반화에 초점을 두는 OMP 알고리즘이다. 언급한 다섯 가지 모델에 대해 RMSE(Root Mean Square Error)와 R2(R-squared)라는 두 가지 평가지표로 회귀분석을 수행하였다. 또한, 객관적인 데이터 비교를 위해 주차장을 기준으로 학습데이터와 평가데이터를 나누어 평가를 진행하였다. 모든 알고리즘에 대한 실험환경은 동일한 조건에서 실험이 진행되었으며, 주차수요의 10 분위수 전체(total)와 중앙값(median)만을 예측하는 두 가지 모델에 대해 학습을 수행하였다. 학습이 완료된 각 회귀모델의 평가데이터에 대한 성능평가는 표 6과 같다.

회귀분석의 각 결과를 살펴보면, OMP가 가장 좋은 성능을 보여주었으며, LASSO가 그다음으로 우수한 성능을 보여주었다. 두 알고리즘 모두 모델의 단순화 및 일반화 성능이 뛰어나고, 변수의 중요도를 파악하여 모델의 해석력이 뛰어난 알고리즘들이다. 본 연구에서 사용한 공영주차장의 데이터의 경우 주차장별 특징에 따른 주차수요의 편차가 매우 심하고 독립변수의 수도 96개로 분석을 수행할 때 유의미한 독립변수를 파악하고 종속변수와의 관계를 고려하기가 어려울 것으로 판단된다. 따라서 OMP, LASSO 알고리즘은 불필요한 변수를 고려하지 않고, 유의미한 변수만을 사용하여 예측을 수행하였기 때문에 높은 일반화 성능을 나타낸 것으로 판단된다.

**Table 6.** Regression results

| Category | | RMSE | R2 |
|---|---|---|---|
| Regression method | Quantiles | | |
| Linear Regression | median | $18.64 \times 10^{13}$ | - 0.8873 × 1024 |
| | total | $18.64 \times 10^{13}$ | - 0.6486 × 1024 |
| LASSO | median | 75.82 | 0.8532 |
| | total | 116.92 | 0.7449 |
| Random Forest | median | 163.02 | 0.3214 |
| | total | 163.04 | 0.5039 |
| GBR | median | 155.86 | 0.3798 |
| | total | 161.62 | 0.5125 |
| OMP | median | **45.82** | **0.9464** |
| | total | **94.40** | **0.8337** |

비선형 회귀분석에서 주로 활용되는 앙상블 기반의 알고리즘인 GBR과 Random forest 알고리즘은 학습데이터에 대한 성능 면에서 OMP, LASSO 알고리즘보다 우수하게 측정되었으나, 평가데이터에 대한 성능은 상대적으로 낮게 측정되었다. GBR 알고리즘은 추정오차를 점진적으로 감소시키며 학습을 수행하는 직렬구조의 알고리즘이다. 이로 인해 학습데이터를 과도하게 학습함으로써 평가데이터의 성능이 좋지 않은 과적합 현상이 자주 발생하는 단점이 있

다. Random forest 알고리즘은 서로 독립적인 여러 모델을 생성하여 각각 학습한 뒤, 여러 모델의 평균값으로 예측을 수행하는 병렬구조의 배깅(Bagging) 알고리즘이다. 역시 학습데이터에 지나치게 의존한 학습으로 인해, 평가데이터에서 새롭게 나타나는 주차장의 주차수요를 예측할 때 유의미한 변수를 선택하지 못하여 일반화된 모델을 구축하지 못하였고, 이로 인해 성능이 저조하게 측정되었다고 판단된다. 표 6에서 보면 주차수요의 10 분위수를 모두 예측하는 것은 주차장별 편차가 크게 존재하여 실제 분위수 값이 매우 크거나 낮은 경우 예측값과 실제값의 차이가 크게 발생하였다. 반면, 주차장별 중앙값만을 예측하는 경우 상대적으로 예측값과 실제값의 차이가 크게 감소하였으며 모델의 설명력 역시 증가하여 좋은 성능을 나타내었다.

공영주차장의 주차수요 예측에 영향을 미치는 주요 변수를 탐색하기 위해 예측성능이 가장 우수했던 OMP 알고리즘의 회귀분석 결과를 분석하였으며, 결과는 그림 4와 같다. 분석결과를 통해 'size'(주차장의 크기) 변수가 가장 중요한 변수임을 확인할 수 있으며, 이로부터 공영주차장의 크기와 주차수요 간에 매우 밀접한 상관관계가 있음을 알 수 있다. 또한, 상위의 주요 변수들을 살펴보면 군집분석의 결과로 도출된 'Cluster_0'(근린 유형)과 'Cluster_2'(주택 유형) 두 가지 군집유형의 중요도가 높게 측정된 것을 확인할 수 있다. 이를 통해 군집분석을 통해 얻어진 주차장 유형이 주차수요에 중요한 영향을 미치는 것을 알 수 있다. 그 외에는 상권분석 데이터의 정책 활용서비스를 활용한 해당 공영주차장 주변의 'Total households'(총 가구 수), 'Open shop(restaurant) number policy'(창업한 외식업 점포 수), 'Normal shop(retail) number policy'(일반 소매업 점포 수), 'Fran shop(retail) number policy'(프랜차이즈 소매업 점포 수), 'building_floating_population_policy'(건물단위 상존인구)가 주차수요를 예측하는데 주요 변수로 탐색 되었다.

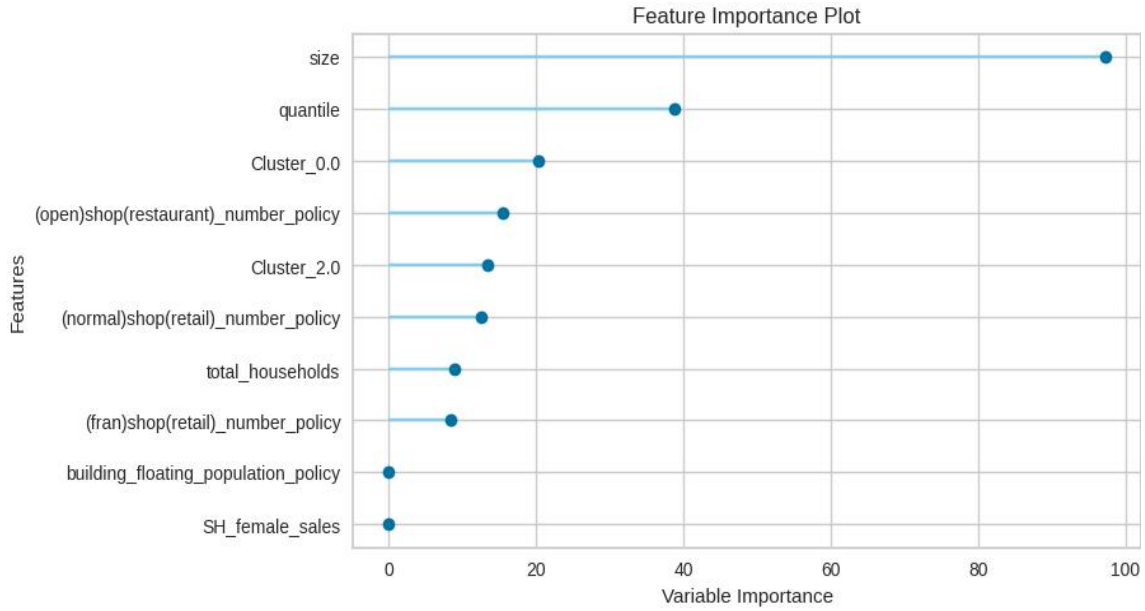



**Figure 4.** Feature importance of OMP models

# 5. 결론 및 시사점

서울시는 주차문제를 해소하기 위해 신규 공영주차장 건설을 계획하고 있다. 그런데 주차수요에 적합한 주차장을 공급하기 위해서는 주차수요에 대한 정확한 예측이 필요하다. 이러한 신규 공영주차장에 대한 정확한 수요 예측을 위해 본 논문에서는 기존 서울시 공영주차장을 수요 유형에 맞게 군집화하고, 해당 군집화 정보와 주차장의 다양한 주변 정보를 바탕으로 회귀분석을 수행하여 정확한 주차수요 예측모델을 생성하였다.

본 연구가 가지는 의의는 실시간 데이터를 활용하여 군집분석을 수행하고 실제 영향권 내 용도지역 면적을 이용하여 군집분석 결과를 검증했다는 점이다. 즉 '실시간 주차가능면 수' 데이터를 활용하여 주차점유면 수에 대한 패턴을 확인하고, 시계열 군집분석을 통해 공영주차장을 4가지 유형(근린 유형, 상업 유형, 주택 유형, 환승 유형)으로 분류하였다. 그리고 군집분석 결과를 영향권 내 용도지역 면적과의 비교 검토를 통해 검증하였다는 점에서 선행연구들과의 차별성을 가진다. 또 다른 의의로는 군집분석 결과와 함께 주차수요에 영향을 줄 수 있는 영향권 내 상권분석데이터, 용도지역별 면적 데이터와 기타 영향권 데이터 등의 다양한 이종 데이터를 바탕으로 각 주차장의 주차수요 분위수에 대한 회귀분석을 수행하여 정확한 주차수요 예측모델을 생성하였다는 점이다. 본 연구의 결과를 이용하여 공영주차장 유형별로 환승 정기권 발행이나 주차장 활용 및 요금 정책 수립 등 공영주차장의 효율성과 공공성을 제고할 수 있으며, 실제 수요에 맞는 신규 공영주차장을 건설하여 과대 또는 과소 수요 산정으로 인한 경제적 낭비를 최소화할 수 있을 것으로 기대된다.

다만, 본 연구는 몇 가지 한계점을 가진다. 먼저 분석을 위해 사용한 주차실적 데이터의 수와 주차장 유형을 탐색하기 위해 데이터를 수집한 기간이 매우 한정적이라는 점이다. 추가적으로 신뢰도 높은 주차실적 데이터를 위한 데이터베이스를 구축하고, 이를 통해 공영주차장의 정확한 수요패턴을 분석할 필요가 있다. 또한, 분석에 활용한 데이터가 코로나19 기간의 데이터여서 다소 왜곡이 있을 수 있으므로 코로나19 이후 데이터를 활용하여 분석결과를 검증해 볼 필요가 있다. 향후 연구로는 주차장 유형별 수요추정 모델에 관한 연구와 주차장 이용 패턴이 수요에 미치는 영향에 관한 추가 연구가 필요할 것으로 판단된다.

# REFERENCES

Cho, Gyeongmi and Cho, Youyoung. 2020. Analysis on Parking Characteristics & Demand of General Hospitals Based on Parking Data - A Hospital Case. Journal of the Korea Convergence Society 11(9):193-202.

Choi, Younghoon and Kim, Eungcheol. 2017. A Study on Factors Influencing Turnover of Public Parking Lots in Incheon Metropolitan City. Incheon National University Urban Science Research Institute 6(2):29-34.

Friggstad Z., Khodamoradi K., and Salavatipour M. R. 2019. Exact Algorithms and Lower Bounds for Stable Instances of Euclidean K-means. In Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithm 2958-2972.

Ha, Jiyoung et al. 2021. The Analysis of the Management Efficiency and Impact Factors of Smart Greenhouse Business Entities Focusing on the Business Entities of Strawberry Cultivation in Jeolla-do. Journal of Korean Society for Quality Management 49(2):213-231.

Heo, Gyeongyong, Kim, Seonghoon, and Woo, Youngwoon. 2010. A Non-linear Variant of Global Clustering Using Kernel Methods. Journal of the Korea Society of Computer and Information 15(4):11-18.

Kim, Changlim and Yoon, Hanseong. 2020. Cluster Analysis of Attribute-Based Data Using Social Network Analysis: Case of Personality Characteristics Data. Journal of Industrial Innovation 36(2):93-110.

Kim, Donghyun. 2016. Dynamic Demand Matches Dynamic Supply. Planning and Policy 415:80-85.

Kim, Hyoungjun and Sohn, Soyoung. 2020. CUSUM Chart Applied to Monitoring Areal Population Mobility. Journal of Korean Society for Quality Management 48(2):241-256.

Kim, Inhee and Kim, Jaehee. 2021. Multivariate Time Series Clustering of Electricity Consumption Data. Journal of the Korean Data And Information Science Society 32(3):569-584.

Kim, Kyeongseop. 2017. Assessment of Premature Ventricular Contraction Arrhythmia by K-means Clustering Algorithm. Journal of the Korea Society of Computer and Information 22(5):65-72.

Kim, Sehoon, Choi, Hyungil, Rhee, Yangwon, and Jang, Seokwoo. 2011. Efficient Dynamic Time Warping Using 2nd Derivative Operator. Journal of the Korea Society of Computer and Information 16(2):61-69.

Kim, Seongtae and Park, Mansik. 2018. A Study on Time-series Clustering Analysis based on Dynamic Time Warping. Journal of The Korean Data Analysis Society 20(5):2319-2332.

Kim, Taegyun, Byun, Wanhee, and Lee, Yunsang. 2017. The Study on Parking Lots Management According to Off-Street Parking Lots Characteristics : Focusing on Residential Development District. LHI Journal of Land, Housing, and Urban Affairs 8(3):131-143.

Lee, Euieun, Lee, Junkyung, and Kim, Juyoung. 2008. Analysis of Basic Characteristics for Providing Parking Information. Journal of the Korean Society of Civil Engineers D 28(5D):639-647.

Lee, Seunghoon and Kim, Yongsoo. 2022. A Pre-processing Process Using TadGAN-based Time-series Anomaly Detection. Journal of Korean Society for Quality Management 50(3):459-471.

Lee, Youngwoo. 2008. A Study for Communication Mode of Bus Information System. Journal of The Korean Society of Industry Convergence 11(3):113-120.

Oh, Joongul. 2016. A Study on District Respectable Public Parking Planning for Urban Revitalization-Comparison of Santa Barbara Public Parking System-. Journal of the Korea Academia-Industrial Cooperation Society 17(3):684-691.

Park, Chandon and Ha, Jaemyung. 2004. A Study on the Analysis of the Parking Demands According to the Housing Unit Size in Apartment Complex-Focused on Apartment Complexes in Daegu City-. Journal of the Architectural Institute of Korea 24(1):550-553.

Park, Geoncheol et al. 2018. Research on Key Priority Areas for Smart City through Citizen Demand Analysis Based on Big Data, Seoul Digital Foundation Report.

Park, Hyunsoo et al. 2022. Characteristics of High Ozone Concentration in Gwangyang Bay Area-Based on Cluster Analysis-. Journal of Korean Society for Atmospheric Environment 38(2):188-202.

Park, Ohsung, Shon, Euiyoung, and Yu, Jeongbok. 2017. Establishing O/D Matrices by Time and by Direction to Improve Demand Forecasting Reliability Through. The Korea Spatial Planning Review 93-104.

Senin P. 2008. Dynamic Time Warping Algorithm Review. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA 855:40.

Seoul Facilities Corporation. [Internet]. https://www.sisul.or.kr/open_content/parking/

Seoul Parking Information Guidance System. [Internet]. http://parking.seoul.go.kr/

Shin, Seungwoo and Lee, Youngwoo. 2021. Establishment of a Model for Calculation the On-Street Parking and Off-Street Parking Demand Considering the Land Use Area in the District. Journal of Korean Society of Transportation 39(6):711-720.

Shin, Woojae, Kim, Gunwoo, and Kim, Jeongmin. 2020. Study on Improving Parking Efficiency for Solving Parking Issues in Seoul, Seoul Digital Foundation Report, 2020-4.

Son, Hyeonseo, Hwang, Jaeyoung, Lee, Insong, and Song Jaeseung. 2020. A Study on the Development of Integrated Control System and the Problem of Parking Space Using Deep Learning in the Internet of Things Environment. Proceedings of Symposium of the Korean Institute of Communications and Information Sciences 939-940.

Yang J. J., Ning C., Deb C., Zhang F., Cheong D., Lee S. E., Sekhar C., and Tham K. W. 2017. K-shape Clustering Algorithm for Building Energy Usage Patterns Analysis and Forecasting Model Accuracy Improvement. Energy and Buildings 146:27-37.

## 저자소개

**황정준** 명지대학교 산업경영공학과를 졸업하고, 현재 동대학원 데이터사이언스 연구실에서 석사 과정 재학 중이다. 관심분야는 기계학습, 딥러닝 등이다.

**신영현** 한양대학교 도시공학 석사를 졸업하고 동 대학원의 박사를 수료하였다. 현재 서울연구원 연구원으로 재직 중이며, 주요 관심분야는 공공투자, 타당성분석 등이다.

**심효섭** 중앙대학교 도시공학과를 졸업하고, 홍익대학교 건축도시대학원에서 석사학위를 받았으며, 고려대학교에서 박사 과정을 수료하였다. 현재 서울연구원에 재직 중이며, 주요 관심분야는 타당성조사, 민자 적격성조사 등이다.

**김도현** KAIST 산업공학과를 졸업하고, 동 대학원에서 석사, 박사학위를 받았으며, 현재 명지대학교 산업경영공학과 교수로 재직하고 있다. 주요 관심분야는 통계적 학습, 기계학습 등이다

**김동근** 연세대학교 산업시스템공학을 졸업하고, KAIST 산업공학과에서 석/박사 학위를 취득하였다. 한국과학기술기획평가원에서 근무하였으며, 현재 서울연구원 연구위원으로 재직 중이다. 주요 관심분야는 타당성조사, 최적화 등이다.