

타슈 공공자전거 수요 예측 및 패턴 분석에 관한 연구

송현호, 장하린, 이상금*

*국립한밭대학교

20227132@edu.hanbat.ac.kr, hrjang713@gmail.com *sangkeum@hanbat.ac.kr

A Study on Predicting and Analyzing Demand Patterns for Tashu Public Bicycles

Hyunho Song, Harin Jang, Sangkeum Lee*

*Hanbat National University

요약

본 연구는 대전광역시 공공자전거 타슈의 동적 수요 패턴을 규명하고 효율적인 운영 정책 수립을 지원하고자 2024년 8월부터 2025년 11월까지 약 16개월간의 대여 이력 데이터를 분석했다. 데이터 정제 및 특징 공학을 통해 시간 단위 시계열 데이터를 구축하고 기술통계 분석, 독립표본 T-검정, 다중 선형 회귀 분석(OLS)을 수행했다. 분석 결과, 타슈의 이용 패턴은 주중과 주말 간에 통계적으로 유의미한 차이를 보였다. 특히 기준의 통념과 달리 주중 일평균 이용량이 주말보다 유의미하게 높게 나타나 타슈가 단순한 여가 수단을 넘어 시민의 핵심적인 생활 교통수단으로 정착했음을 입증했다. 또한 OLS 회귀분석 결과, 주말 여부는 시간당 대여량에 음의 영향을 미치는 것으로 나타났다. 시간대별 패턴 분석에서는 주중의 명확한 쌍봉형 패턴이 관찰되었으며, 이는 타슈가 대중교통의 보완 수단으로서 역할을 수행함을 확인하고 출퇴근 시간대 거점 중심의 운영 전략의 필요성을 제언한다. 향후 연구에서는 OLS 가 통제하지 못한 공간적 자기상관성 문제를 극복하기 위해 공간 계량 모델 및 시계열 특화 모델을 적용할 것을 제안한다.

I. 서론

최근 기후 변화와 에너지 고갈 문제에 대한 전 세계적인 관심이 높아짐에 따라, 자전거 이용은 친환경 교통수단이자 대중교통을 보완하는 핵심적인 수단으로 주목받고 있다. 대전광역시는 2008년부터 공공자전거 타슈를 운영하고 있으며, 2024년 기준 약 400개 대여소와 4,000여 대의 자전거를 보유하고 있다. 연간 이용 건수는 600만 건을 초과하며, 지속적인 이용 수요 증대에 따라 자전거 부족 또는 과잉 문제를 해소하기 위한 데이터 기반의 과학적인 운영 전략이 필수적으로 요구된다.

기존의 공공자전거 관련 연구들은 주로 공간 계량 분석을 적용하여 청년층 비중, 대중교통 접근성 등 정적인 도시 특성이 이용에 미치는 영향을 규명하거나[1], 마르코프 체인 및 시뮬레이션 기법을 활용하여 고객 불만족도를 최소화하는 정적 재배치 목표 재고 수량 결정 방안을 제시하는 데 주력했다[2]. 또한, 타슈가 여가 교통수단으로 전락했으며, 대중교통 연계성을 강화하고 시민 참여를 통한 재배치가 필요하다는 정책적 시사점을 제시했다[3].

그러나 이러한 선행 연구들은 최근 이용 트렌드인 '생활 교통수단화' 경향을 반영하지 못하거나 시간의 흐름에 따른 동적인 수요 변화를 분석하는 데 한계가 있다. 따라서 본 연구는 타슈의 동적 수요 패턴을 정량적으로 규명하고 정책 결정에 활용 가능한 구체적인 통계적 증거를 도출하는 것을 목표로 한다.

II. 본론

1. 데이터 구축 및 전처리

본 연구는 대전교통공사에서 제공하는 2024년 8월부터 2025년 11월까지의 타슈 대여 이력 데이터를 핵심 자료로 활용했다. 데이터의 주요 변수는 대여소 ID, 대여 일시, 반납 일시, 이용 시간, 이용 거리 등으로 구성된다. 원본 데이터는 약 770만 건에 달하며, 분석 기간 동안 일평균 대여 건수는 약 16,900건으로 나타났다. 분석의 정확도를 높이기 위해 시스템 장애로 인해 데이터 소실이 발생한 2025년 2월 26일부터 2025년 2월 28일과 2025년 3월 1일부터 2025년 3월 3일 동안의 기간은 전처리 과정에서 노이즈로 간주하여 제거했다. 또한, 대여 시간이 1분 미만이거나 24시간을 초과하는 비정상적인 기록은 시스템 오류 또는 미반납으로 판단하여 분석에서 제외했다.

2. 특징 공학 및 분석 모델 설정

시계열 특성을 반영하기 위해 대여일시 데이터로부터 Hour, Weekday, Is_weekend 등의 파생 변수를 생성했다. Hour 변수는 0부터 23까지의 정수값을 가지며, Is_weekend 변수는 토요일과 일요일을 1, 평일을 0으로 코딩한 이진 변수이다. 데이터 분석을 위해 본 연구에서는 A. Field[4]가 제시한 방법론에 따라 독립표본 T-검정과 다중 선형 회귀 분석(OLS)을 수행했다. 독립표본 T-검정은 주중과 주말이라는 두 집단 간 일일 이용량의 평균 차이가 통계적으로 유의한지 검증하기 위해 사용되었다. 또한, OLS 회귀분석은 시간대, 요일 등 다수의 독립변수가 대여 건수에 미치는 상대적 영향력을 정량적으로 추정하기 위해 활용되었다.

III. 결과

시간대별 이용 패턴 시작화 결과, 이용 목적에 따라 뚜렷한 이중적 특성이 관찰되었다.

- 평일: 오전 8~9 시와 오후 6~8 시에 두 개의 뚜렷한 피크가 나타나는 쌍봉형 패턴이 관찰된다. 이는 타슈가 시민들의 통근 및 통학 수단으로 활발히 이용되고 있음을 강력하게 시사한다.
- 주말: 오후 시간대(14~18 시)에 완만하게 상승하여 최대치에 도달하는 단봉형 패턴이 나타나 여가 및 레저 활동 목적의 이용 특성을 보인다.

독립표본 T-검정과 OLS 통계 검정 결과 본 연구에서 가장 주목할 만한 발견은 주중과 주말의 이용량 비교에서 나타난다. 독립표본 T-검정 결과 주중의 일평균 대여 건수는 약 17,605 회로 주말의 약 15,703 회보다 약 1.12 배 높게 나타났다($t = 2.39$, $p = 0.017$). 이는 기존에 타슈를 단순 여가 수단으로 인식하던 경향과 달리 타슈가 대전 시민의 필수적인 평일 이동 수단으로 자리 잡았음을 통계적으로 입증한다.

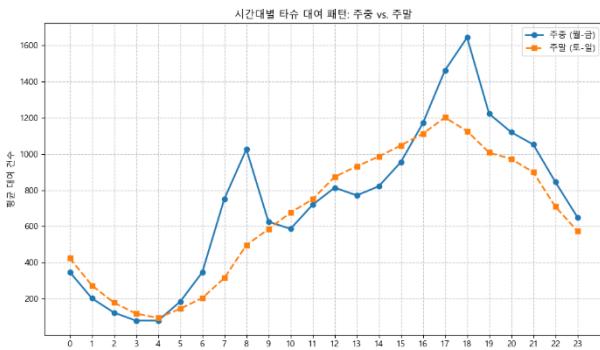


그림 1. 시간대별 타슈 대여 패턴: 주중 vs. 주말

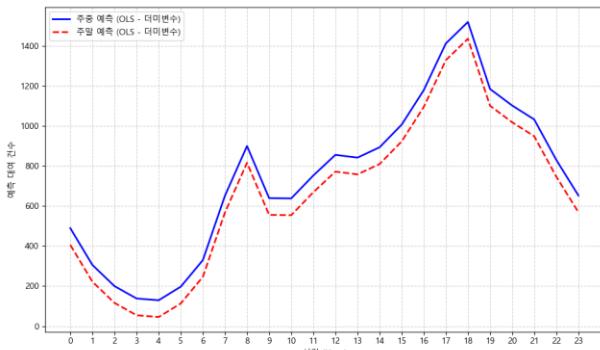


그림 2. OLS 통계 검정 결과 그래프

또한, 시간당 대여량을 종속변수로 한 OLS 회귀분석 결과에서도 주말 여부의 회귀계수는 -84.01 로 나타나($p < 0.001$) 다른 조건이 동일할 때 주말의 시간당 평균 대여량이 평일보다 낮음을 재확인했다. 반면, 시간 변수는 양의 계수를 보여 활동 시간이 늦어질수록 누적 대여 수요가 증가하는 경향을 보인다.

본 연구의 OLS 분석은 시계열적 트렌드와 요일별 차이를 설명하는 데 유효했으나, 특정 지역의 수요가 인접 지역에 영향을 미치는 공간적 자기상관성 문제를 통제하지 못했다는 한계가 있다. 선행 연구[3]에서 지적된 바와 같이 향후 연구에서는 LISA 분석[5]을 통해 핫스팟을 도출하고 공간 계량 모델을 도입하여 분석의 정교함을 높일 수 있다.

IV. 결론

본 연구는 2024년과 2025년의 대규모 데이터를 분석하여 타슈가 주말 여가용보다 평일 교통수단으로서의 기능이 더욱 강화되었음을 통계적으로 입증한다. 특히 평일 출퇴근 시간대에 집중된 이용 패턴은 타슈가 대중교통의 'First & Last Mile'을 담당하는 핵심 수단임을 보여준다.

따라서 타슈의 운영 전략은 기존의 여가 중심 관리에서 벗어나 출퇴근 시간대 거점 지역을 중심으로 한 동적 재배치 전략으로 전환되어야 한다. 이는 대중교통 연계성을 강화하고 시민의 이동 편의성을 극대화하는 가장 효율적인 정책 방향이 될 것이다[3].

향후 연구에서는 시계열 특성을 효과적으로 반영할 수 있는 XGBoost[6] 및 LSTM[7] 모델을 적용하고, 선행 연구[8]의 방법론을 벤치마킹하여 시간순 교차 검증 전략을 기반으로 한 예측 실험을 수행할 예정이다. 이를 통해 시간당 대여량 예측의 평균절대백분율오차를 20% 이내로 달성하는 것을 목표로 한다.

참고문헌

- [1] 안민수, 이창효, "공공자전거 이용에 영향을 미치는 공간 특성 분석: 대전광역시 '타슈'를 대상으로," 지역연구, vol. 38, no. 4, pp. 75-91, 2022.
- [2] 윤영채, 조병용, "대전광역시 공공자전거 '타슈' 발전 방안에 관한 연구," 사회과학연구, vol. 25, no. 2, pp. 591-608, 2014.
- [3] 송재인, "공용자전거 타슈의 정적 재배치 방안에 대한 연구," 석사학위논문, 한남대학교, 대전, 2015.
- [4] A. Field, Discovering Statistics Using IBM SPSS Statistics, 5th ed. London, UK: Sage, 2018.
- [5] L. Anselin, "Local indicators of spatial association—LISA," Geogr. Anal., vol. 27, no. 2, pp. 93-115, 1995.
- [6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., San Francisco, CA, USA, 2016, pp. 785-794.
- [7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.
- [8] S. Lee, "Cooperative decentralized peer-to-peer electricity trading of nanogrid clusters based on predictions of load demand and PV power generation using a gated recurrent unit model," Renew. Energy, vol. 210, pp. 68-80, 2023.