

공공 데이터 분석을 통한 서울시 전동 공공자전거의 최적의 입지 선정에 대한 연구

전준현*, 이우진**

A study on the selection of the best location for electric shared bicycles in Seoul through open data analysis

Jun-Heon Jeon*, Woo-Jin Lee**

요 약

본 논문은 서울시 공공 데이터를 분석하여 서울시 시범사업인 전동 공공자전거의 최적의 입지에 대한 선정을 제안하고자 한다. 서울시는 2014년 시범 운영을 시작으로 2015년 10월부터 본격적으로 공공자전거 대여 서비스를 시작하였으며, 2020년 시범 사업으로 전동 공공자전거 대여 서비스 시작하고 있다. 데이터 분석 결과 초기의 입지 선정이 중요하다. 그래서 본 논문에서는 서울시 공공 데이터를 분석하여 사용자가 가장 많이 사용할 수 있는 기준으로 최적의 입지 자치구를 선정하였다. 본 논문은 데이터 분석을 위하여 jupyter notebook 툴을 사용하였으며, 모듈은 pandas, matplotlib, seaborn, forlium를 사용하여 분석 및 시각화를 하였다.

ABSTRACT

In this paper, we proposed the selection of the optimal location for electric shared bicycles, a pilot project of Seoul, by analyzing Seoul's open data. The Seoul Metropolitan Gaverment started a shared bicycle rental service in October 2015, starting with the pilot operation in 2014, and is preparing for an electric shared bicycle rental service as a pilot project in 2020. As a result of data analysis, it is important to select an initial location. So, in this paper, we analyzed the public data of Seoul and selected the optimal location autonomous region as the criterion that users could use the most. This paper used the jupyter notebook tool for data analysis, and the module was analyzed and visualized using pandas, matplotlib, seaborn, and forlium.

키워드 : 공공 데이터, 전동 공공자전거, 공공자전거, 데이터 분석, 시각화

key Words : public data, electric shared bicycles, shared bicycles, data analysis, visualization

I. 서 론

지구온난화의 원인 중에 하나로 환경오염에 대한 관심이 최근 들어 커지고 있다. 이러한 시기에 맞추어 서울시는 2014년 시범 운영을 시작으로 2015년 10월부터 본격적으로 무인 공공자전거 “따릉이”의 대여 서비스를 시작했다. 특히 공공자전거(Shared bicycle)의 시작은 네덜란드에서 시작되어 친환경 교통수단으로 주목 받아 오고 있다[1].

환경오염 방지와 도심 속 교통체증을 줄이기 위

해 각 시도에서 운영 중에 있으며, 전 세계에서 많은 도시에서 운용되고 있다. 국내는 2016년 3월 1일 기준으로 12개 도시에서 19,421대의 공공자전거가 운영되고 있으며[2], 전 세계에서 2010년 기준 125개 도시에서 약 139,300대의 공공자전거가 친환경 교통수단으로 운영되고 있다[3,4,5]. 또한 정보통신기술의 발달로 자전거와 스마트폰을 연동하는 서비스를 제공함으로써 관리가 더욱 용이해 지고 있다. 서울시는 한발 더 나아가 자전거 전용도로 하이웨이(Cycle Rapid Transportation)를 구상 중에

* 상명대학교 계당교양교육원(junheon@smu.ac.kr), ** (교신저자) 상명대학교 계당교양교육원(woojin@smu.ac.kr)

§ 논문번호 : 211604, 접수일자 : 2021년 01월 04일, 수정일자 : 2021년 02월 01일, 심사완료일자 : 2021년 03월 03일

있으며, 이에 대한 일환으로 2020년 전기 배터리를 사용한 전동 공공자전거(전동따릉이) 1000대 도입을 진행 하고 있다.

(주)벨로스타는 2019년 11월 서울시설관리공단과 서울시 공공자전거 따릉이 후속버전인 전동따릉이 500대 납품을 확정하였으며, 서울시는 20여억원의

예산을 반영하여 시범운영 후 서울시 전역으로 확대할 예정이다.

초기의 입지 선정의 중요성을 파악하기 위하여 2015년 9월부터 2019년 3월까지 공공자전거 데이터를 분석하였다. 그림1은 공공자전거 초기 입지에 대한 분석코드 및 도표를 보여주고 있다.

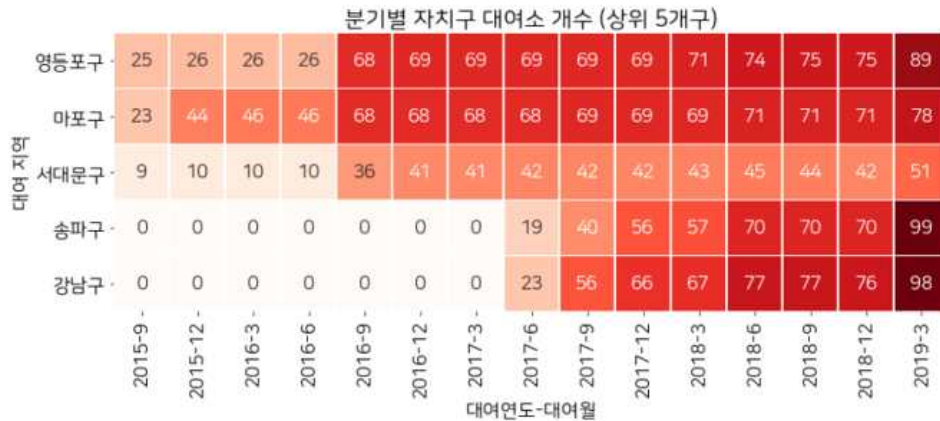


그림 1. 2015년 ~ 2019년 대여소 초기 입지 분석

그림1은 공공 데이터 중에서 상위 5개의 자치구에서 공공자전거의 대여 횟수를 추출한 것으로, 2015년 9월 자료를 보면 영등포구와 마포구는 대여소가 설치 되어있고, 송파구와 강남구는 2017년 6월부터 대여소가 설치되었음을 알 수 있다. 또한 송파구와 강남구는 2019년 3월 기준으로 가장 많은 대여소 개수를 가지고 있다.



그림 2. 대여소 사용량 분석

그림2는 공공 데이터에서 대여소 사용량을 추출하여 시각화한 자료이다. 가장 많은 대여소를 가지고 있는 송파구와 강남구 중 강남구는 대여소 사용

량 상위 5개구에 속하지 못하였고, 송파구는 상대적으로 대여소 수가 적은 마포구와 영등포구에 비해 사용량이 적음을 알 수 있다. 2019년 3월 기준으로 강남구와 송파구가 가장 많은 대여소를 보유하고 있지만, 사용량을 보았을 때 초기 대여소를 설치한 영등포구와 마포구보다 사용량이 떨어진다. 이는 대여소의 초기 입지 선정이 향후 사용량에 많은 영향을 미친다는 것을 보여준다.

이에 본 논문에서는 공공 데이터를 파이썬(Python) 프로그램을 사용하여 효율적으로 분석하고, 여러 모듈을 이용하여 시각화(visualization)하여 사용자가 가장 많이 사용할 수 있는 공공자전거 최적의 초기 입지 지역을 제안하고자 한다.

II. 공공 데이터 수집 및 분석

공공 데이터(open data)는 공공기관이 업무수행 결과물로 생성 또는 취득한 다양한 형태의 자료로서 이용자가 활용할 수 있도록 제공되는 자료이다. 그러나 공공 데이터는 이용자 편의로 정리 되어 있지 않기 때문에 이를 분석 하려면 어려운 점이 많이 있다. 그러나 공공 데이터는 객관적인 데이터를 바탕으로 만들어진 자료로서 주요 OECD 국가들은

정부 정책 자료로 공공 데이터를 분석한 자료를 많이 사용하고 있다[6]. 본 논문에서는 이런 공공 데이터를 효율적으로 수집, 분석하여 최적의 초기 입지 지역을 선정하여 제안하는 것으로, 그림 3은 본 논문의 분석 목표를 도식화 것이다.

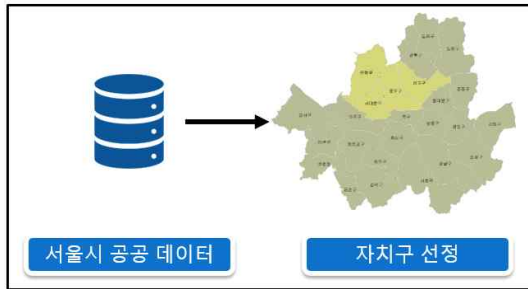


그림 3. 공공 데이터 분석 목표

본 논문에서 수집 및 분석한 서울시 공공 데이터는 서울시 공공자전거 사용 데이터, 공공자전거 대여소 개방 현황, 공공자전거 대여소 폐쇄 현황, 공공자전거 대여소 정보, 월별 신규가입자 정보로서 이를 효율적으로 분석하기 위하여 파이썬을 이용하여 데이터 수집, 데이터 가공, 데이터 분석을 하였으며, 이를 통하여 가장 많은 사용자가 이용할 수 있는 최적의 입지에 적합한 대표 자치구를 제안한다. 본 논문에서 가정한 사항은 전동 자전거는 일반 자전거와는 성격상 다르기 때문에 가까운 거리보다는 먼 거리의 이용이 가능한 특성을 고려하여 대표 자치구를 선정하고 인접한 자치구를 선정하는 방식으로 전동 자전거 대여소의 최적의 입지 지역을 선정하였다. 그림 4는 데이터 수집 과정을 보여준다.



그림 4. 공공 데이터 수집

그림 5는 공공 데이터 수집을 위한 함수의 정의이며, 그림 6은 함수의 호출로 2015년 공공 데이터를 수집한 형태를 보여준다. 데이터 수집은 서울시 공공 데이터 포털을 통해 약 4GB의 로우 데이터를 수집하였다. 수집된 로우 데이터에서 필요한 정보로 가공하기 위하여 pandas 모듈을 사용하였다.

```

all_bikeuse = []
all_bikeuse2 = []
all_bikeuse3 = []
path_bikeuse = path + '/data/2015_201905/'
path_bikeuse2 = path + '/data/inout/'
def load_file(name):
    fname, ftype = name.split('.')
    if ftype == 'csv':
        df = pd.read_csv(path_bikeuse+fname, encoding='cp949')
    elif ftype == 'xlsx':
        df = pd.read_excel(path_bikeuse+fname)
    else:
        print(f'load_file({fname}): An input file type is not acceptable!')
        return None
    df.columns = ["자전거번호", "대여 대여소 번호", "대여 대여소명", "대여 거치대번호",
                  "반납일시", "반납 대여소 번호", "반납 대여소명", "반납 거치대번호",
                  "이용시간", "이용거리"]
    return df
def load_file2(name):
    fname, ftype = name.split('.')
    if ftype == 'csv':
        df = pd.read_csv(path_bikeuse2+fname, encoding='cp949')
    elif ftype == 'xlsx':
        df = pd.read_excel(path_bikeuse2+fname)
    else:
        print(f'load_file({fname}): An input file type is not acceptable!')
        return None
    return df
    
```

그림 5. 공공 데이터를 가져오기 위한 함수 정의

```
# 2015
all_bikeuse.append(load_file('서울특별시 공공자전거 대여이력 정보_2015년.csv')); all_bikeuse[0].head()
```

	자전거번호	대여일시	대여 대여소 번호	대여 대여소명	대여 거치대번호	반납일시	반납 대여소 번호	반납 대여소명	반납 거치대번호	이용시간	이용거리
0	SPB-00186	2015-09-19 00:59:00	115	사투비야 빌딩 앞	5	2015-09-19 02:39:00	115	사투비야 빌딩 앞	5	98	15240
1	SPB-00403	2015-09-19 01:22:00	112	국동방송국 앞	3	2015-09-19 06:09:00	102	망원역 1번출구 앞	16	286	10910
2	SPB-00485	2015-09-19 01:32:00	112	국동방송국 앞	1	2015-09-19 06:09:00	102	망원역 1번출구 앞	15	276	11000
3	SPB-00292	2015-09-19 07:52:00	221	여의도조교 앞	2	2015-09-19 08:56:00	221	여의도조교 앞	1	63	16500
4	SPB-00270	2015-09-19 08:56:00	221	여의도조교 앞	3	2015-09-19 09:05:00	210	IFC몰	2	8	1870

그림 6. 2015년 공공 데이터 수집

그림 7은 데이터 수집, 가공, 분석 및 시각화를 위하여 사용된 모듈을 보여준다.

```
import json
import pandas as pd
import numpy as np

import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm
import seaborn as sns
import folium
from folium import plugins
from folium import *
import missingno as mns

import warnings
warnings.filterwarnings('ignore')
```

그림 7. 데이터 처리를 위해 사용된 모듈

그림 8은 2017년부터 2019년까지 연령별 가입자 수를 분석하여 시각화한 자료이다. 그림 8에서 가장 많은 가입자는 10대이며, 가장 적은 가입자는 70대임을 알 수 있다. 처음 연령별 가입자 수를 분석한 이유는 최적의 입지 선정에 위한 분석 연령을 구하기 위함이다. 그러나 가입자가 많은 연령이 사용량이 많은지는 더 분석이 필요하다고 판단하여 공공자전거 이용 증가율을 분석하였다.

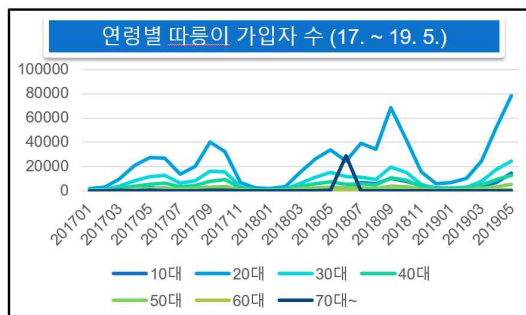


그림 8. 공공자전거 연령별 가입자 수 분석

그림 9는 2017년 대비 2018년도 공공자전거 이용 증가율을 분석하여 백분율로 나타내도록 시각화한 자료이다. 그림 9에서 가장 많은 증가율을 보인 것은 50대, 가장 적은 증가율을 보인 것은 10대로 분석되었다. 이를 기반으로 본 논문에서는 가정 적은 증가율을 보인 10대를 제외하고, 청년층(20~30대)과 중년층(40~50대)로 나누어 분석을 하였다.

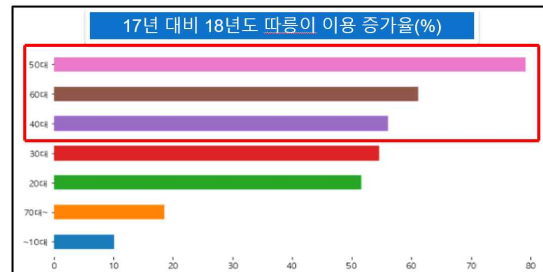


그림 9. 공공자전거 이용 증가율(%)

그림 10은 공공자전거에 대한 자치구별 평균 대여량을 분석하여 시각화한 자료이다. 공공자전거의 평균 대여량을 분석할 경우 광진구, 마포구 순으로 나타나고 있다. 좀 더 정확한 분석을 위하여 앞서 정한 청년층과 장년층을 대상을 한정하였고, 시간도 출근과 퇴근 시간으로 한정하였다. 이는 공공자전거 전체 사용시간을 분석해 보면 출퇴근 시간에 38% 이상이 집중되고 있다는 분석을 참고하였[7].



그림 10. 공공자전거 평균 대여량

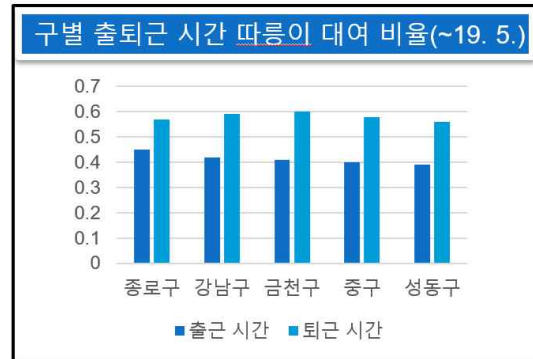


그림 11. 구별 공공자전거 대여 비율

그림 11은 공공자전거의 평균 대여량과 같은 조건을 기준으로 구별 공공자전거 대여 비율을 분석하여 시각화한 자료이다. 공공데이터의 청년층과 장년층을 대상, 출근과 퇴근 시간을 기준으로 분석한 결과 가장 많은 사용자가 사용할 수 있는 자치구는 종로구임을 알 수 있다.

그림 12는 종로구를 기준으로 사용량을 평일, 주말, 출근, 퇴근 시간으로 분석하여 Folium 모듈을 이용하여 시각화한 자료이다. 그림 13은 그림 12의 분석 및 시각화에 사용된 함수 중 일부이다.

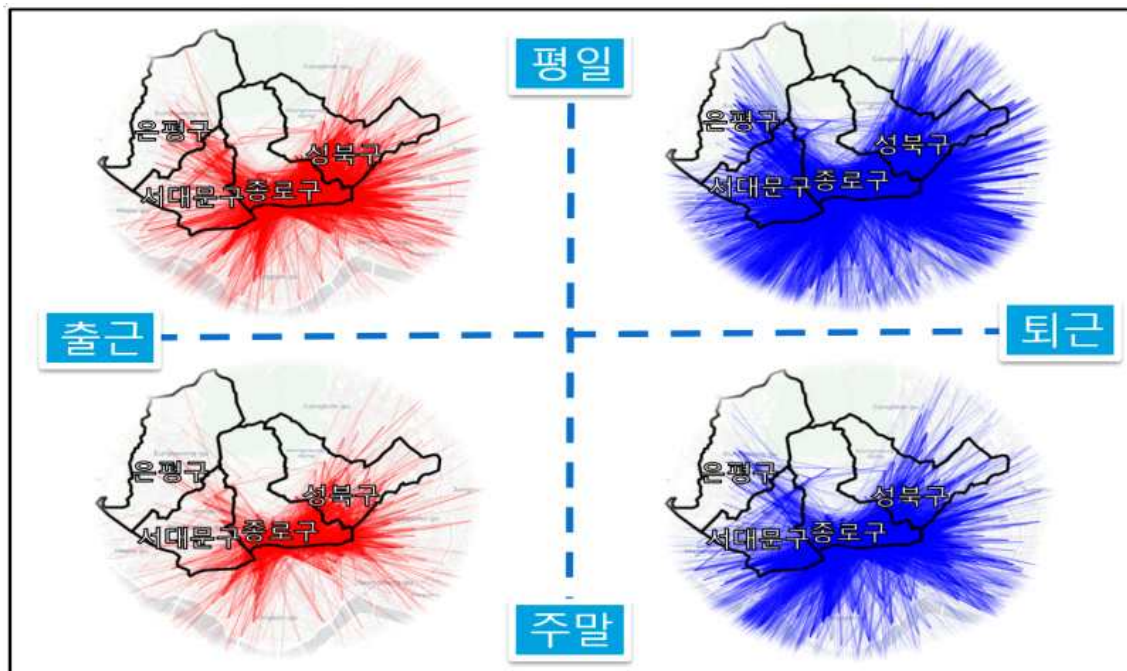


그림 12. 출퇴근 공공자전거 사용량 시각화


```
def draw_usage(df, region):
    fig, axes = plt.subplots(1, 2, figsize=(24, 5))
    for i, dayofweek in enumerate(["평일", "주말"]):
        if dayofweek == "평일":
            _dayofweek = set(range(0,5))
        elif dayofweek == "주말":
            _dayofweek = set(range(5,7))

        by_hour_rental = df[(df['대여요일'].isin(_dayofweek)) & (df['대여 지역'] == region)].groupby('대여시간').size() / len(_dayofweek)
        by_hour_return = df[(df['대여요일'].isin(_dayofweek)) & (df['반납 지역'] == region)].groupby('반납시간').size() / len(_dayofweek)
        by_hour = pd.DataFrame(data={
            "대여": by_hour_rental,
            "반납": by_hour_return
        }).plot(style='-', rot=0, title="%s, 시간대별 사용량(대여, 반납)" % dayofweek, ax=axes[i])
        plt.sca(axes[i])
        plt.xticks(range(0, 24, 1))
        plt.xlabel("시")
        plt.box(False)
        plt.legend(frameon=False)
    plt.show()
```

그림 13. 공공자전거 사용량 분석 및 시각화를 위한 함수 중 일부

그림 12에서 오른편의 파란색으로 표시된 부분은 퇴근시간에 공공자전거가 사용된 사용량을 보여주고 있으며, 왼편의 빨간색으로 표시된 부분은 출근시간에 공공자전거가 사용된 사용량을 보여주고 있다. 그림 12에서 위편은 평일(월요일 ~ 금요일) 출퇴근 시간에 공공자전거 사용량을 보여주고 있으며, 아래편은 주말(토요일, 일요일) 출퇴근 시간에 공공자전거 사용량을 보여주고 있다.

III. 결론

본 논문에서는 공공데이터를 효율적이고 객관적인 분석을 위하여 python언어를 사용하여 프로그래밍 하였으며, 분석결과를 시각화하여 보여주었다. 단순한 대여량을 비교한 것이 아니라 가장 많이 사용하는 연령대 분석 및 가장 많이 사용하는 시간 구간을 분석하여, 이를 기준으로 가장 많은 사용자가 사용할 수 있는 자치구를 선정하였다. 또한 선정된 자치구의 사용량을 분석하여 시각화하여 선정된 자치구 이외에도 추가적인 자치구를 선정할 수 있도록 하였다.

본 논문에서는 분석을 통하여 최적의 전동 공공자전거 초기 입지지역으로 종로구를 선정하였다. 분석을 통해 선정된 종로구는 많은 사용자가 사용할 수 있는 최적의 입지 기준으로 선정한 것이며 연령대, 시간 구간 등의 조건이 다를 경우 결과가 다를 수 있다. 앞으로 더 많은 데이터의 분석을 통하여 선정된 자치구를 기준으로 자치구내의 동을 선정하는 연구를 계속 진행할 계획이다.

감사의 글

본 연구는 2021학년도 상명대학교 교내연구비를 지원받아 수행하였음.

참고문헌

- [1] 임희중, 정광현, "확률적 동적 계획법을 이용한 공공자전거 정비소에서의 최적 재고관리," 한구생산관리학회지 제31권 제1호, pp. 15-28, 2020.
- [2] 장동익, "재미있는 교통통계," <https://www.ktdb.go.kr/www/selectBbsNttView.do?key=42&bbsNo=7&nttNo=628>, 국가교통DB, 2016.
- [3] Shaheen, S. A., H. Zhang, E. Martin, S. Guzman, "China's Hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing," Transportation Research Record, Vol.2247, No.1, p p. 33-41, 2011.
- [4] P. Midgley, "The role of smart bike-sharing systems in urban mobility," Journeys, Vol.2, No.1, pp. 23-31, 2009.
- [5] B. Cohen, J. Kietzmann, "Ride on! Mobility business models for the sharing economy," Organization and Environment, Vol.27, No.3, pp. 279-296, 2014.
- [6] 이재만, "빅 데이터와 공공 데이터 활용," Internet and Information security 제2권 제2호, pp.47-64, 2011.
- [7] "서울시 '따릉이' 회원 62만 돌파... 출퇴근시간 38% 집중," https://www.sisul.or.kr/open_content/main/bbs/bbsMsgDetail.do?msg_seq=421&keyfield=title&keyword=%EB%94%B0%EB%A6%89%EC%9D%B4&listsz=10&bcd=report&pgno=2, 서울시설공단, 2018.

저자소개



전 준 헌(Junheon Jeon)
2015년 상명대학교에서 이
학박사(컴퓨터과학 전공)를
취득하였다. 현재 상명대학
교에 근무하고 있으며, 주
요 관심분야는 컴퓨터네트
워크, 빅데이터 분석, 인공
지능 등이다.



이 우 진(Woojin Lee)
2007년 숭실대학교에서 공
학박사(컴퓨터공학 전공)를
취득하였다. 현재 상명대학
교에 근무하고 있으며, 주
요 관심분야는 소프트웨어
공학, 빅데이터 분석, 인공
지능 등이다.

