데이터 분석을 통한

서울시 공공자전거 효율적인 재배치 방안 제안



Part 1

따릉이 현황 분석을 통한 연구 방향 및 목표 수립

01.Business Understanding

02.Data Understanding

03.Data Preparation

04.Modeling

05.Appendix

Business Understanding

이용 및 운영 현황 440대로 시작하여 누적 도입 대수 4만 3500대, 한 해 이용건수 약 4100만 건 (2022년 기준) 성공적인 서울시 공공 정책의 대표 사례

1. 성공적인 공공자전거 서비스 "따름이"

- 2022년 총 따름이 이용량: 4094만 8900건
- 3년 연속 서울시 우수 정책 1위 (2017년 ~ 2019년)
- 도입 대수(누적) 5600대(2016년)에서 4만3500대(2022년) 로 증가

2. 이용할수록 늘어나는 적자

- 2022년 적자 규모: 94억원
- 15년 째 동결 중인 이용 요금 (시간 당 천 원)
- 도입 대수가 늘어남에 따른 유지보수 비용 및 인건비 증가로 인한 지속적인 운영 비용 증가

3. 지속 가능한 서비스인가?

- 대중교통으로써 공공성 vs 일부 연령층의 시민이 주로 이용
- 이용 요금 인상 및 광고 입찰 등 수익성 개선을 위한 다양한 방법을 고려 중이지만 실행 가능 여부 재고 필요



(출처: news.mt.co.kr/mtview.php?no=2024013009201542610)

- 운영비 감당이 어려운 구조로 인해 **적자의 폭이 증가**하는 추세 > 운영 **효율 개선**이 필요

Business Understanding

서비스 적자 분석 운영 관리 비용 중 가장 큰 비중을 차지하는 항목은 **자전거 재배치 비용**(단일 항목 기준)

1. 자전거 재배치 효율화 => 적자 감소

- 공공 자전거 사업의 총 운영비 중 약 30% 지출 추정
- 단일 항목 지출 비용 중 가장 큰 비중

자료 출처: 한국 교통연구원 – 공공자전거 효과 분석 및 발전 방안 보고서

2. 서울시 따릉이 재배치 현황

- 각 자전거 관리소 149명 중 114명이 재배치 담당
- 3교대 2조로 운영: 07시 ~ 16시 / 16시 ~ 23시 / 23시 ~ 04시(약)

자료 출처: 서울시 공공시설 관리 공단 – 자전거 운영처 종합현황 보고서

| 주요 공공자전거 적자 현황 ^{적자규모(원)} ^{운용수(대)} 누비자(창원) | 따름이(서울) 103억 3만7500 타슈(대전) |
|---|----------------------------|
| 3900 | 36° 2700 |
| 타랑께(광주) | 어울링(세종) |
| 10 200 | 6 9 3 200 |

| < · · · · · · · · · · · · · · · · · · · | 유지과리비 | 추정격과 | 종합(창원시) |
|---|-------|------|---------|
| | | | |

| | 금액 | | | | | |
|------------------------|----------|-----|--|--|--|--|
| 인건비 | 관리자 | 45 | | | | |
| | 관제 | 140 | | | | |
| | 상담 | 180 | | | | |
| | 시무관리 | 18 | | | | |
| | 자전거수리 | 375 | | | | |
| | 현장/분배 | 500 | | | | |
| | 자전거수리비 | 470 | | | | |
| 자전거 및 시설 유지관리비, 부품비 | 시설물수리비 | 351 | | | | |
| TB91 | 시설물전기통신료 | 184 | | | | |
| | 차량 | 200 | | | | |
| 정비 및 수리 센터 운영비 | 공구구매 | 15 | | | | |
| | 기타 | 18 | | | | |
| 센터 운영비 기타 | 전기/통신료 | 19 | | | | |
| (일반관리 또는 이윤) | 기타 | 31 | | | | |
| | <u>а</u> | | | | | |

(출처: www.joongang.co.kr/article/25099146)

(신희철, 김동준, 정성엽, 공공자전거 효과 분석 및 발전 방안, 한국교통연구원, P.150)

- **자전거 재배치 비용을 줄이는 것**이 만성적인 서비스 적자 감소의 핵심 아이디어

Business Understanding

연구 방향





반납량

대여량 = 반납량 (평형 상태)

대여량 > 반납량 (비평형 상태 - 부족)

대여량 < 반납량 (비평형 상태 - 과잉)

Part 2

원본 데이터 수집 및 이해를 통한 분석 목표 설정

01.Business Understanding

02.Data Understanding

03.Data Preparation

04.Modeling

05.Appendix

데이터 출처

서울시 열린 데이터 광장 실데이터 활용

서울시 따릉이 대여소별 대여/반납 승객수 정보 >> tpss_bcycl_od_statnhm_yyyymmdd.csv

데이터 정보

특징: 따릉이 이용 기록을 실시간으로 기록하는 로그 데이터 형태

기간: 2023년 4월 1일 ~ 2024년 3월 31일

구성: 86,434,238개의 data, 10개의 Columns

주요 정보: 기준_날짜, 집계_기준, 시작_대여소_ID, 종료_대여소_ID(문자형 데이터), 기준_시간대, 전체_건수(숫자형 데이터)

| | 기준_날짜 | 집계_기준 | 기준_시간대 | 시작_대여소_ID | 시작_대여소명 | 종료_대여소_ID | 종료_대여소명 | 전체_건수 | 전체_이용_분 | 전체_이용_거리 |
|----------|----------|-------|--------|-----------|------------|-----------|------------|-------|---------|----------|
| 84599450 | 20240312 | 도착시간 | 2355 | ST-963 | 논현2동_018_1 | ST-787 | 논현1동_043_1 | 1 | 12 | 2390 |
| 84599451 | 20240312 | 도착시간 | 2355 | ST-965 | 낙성대동_007_2 | ST-1270 | 신림동_030_1 | 2 | 29 | 4970 |
| 84599452 | 20240312 | 도착시간 | 2355 | ST-983 | 중곡1동_022_1 | ST-1469 | 창2동_036_1 | 1 | 57 | 15738 |
| 84599453 | 20240312 | 도착시간 | 2355 | ST-984 | 자양1동_008_1 | ST-2891 | 자양2동_012_2 | 1 | 6 | 1225 |
| 84599454 | 20240312 | 도착시간 | 2355 | ST-985 | 중곡3동_030_2 | ST-985 | 중곡3동_030_2 | 1 | 90 | 19601 |

데이터 출처

서울시 열린 데이터 광장 실데이터 활용

서울시 따름이 대여소 마스터 정보 >> 서울시 따름이 대여소 마스터 정보.csv

데이터 정보

특징: 따릉이 대여소 정보가 담긴 데이터

구성: 3,286개의 data, 5개의 Columns

주요 정보: 대여소_ID, 주소 1, 주소 2(문자형 데이터), 위도, 경도(숫자형 데이터)

| | 대여소_ID | 주소1 | 주소2 | 위도 | 경도 |
|---|--------|-------------------------|---------------|-----------|------------|
| 0 | ST-999 | 서울특별시 양천구 목동서로 280 | 목동아파트 8단지 상가동 | 0.000000 | 0.000000 |
| 1 | ST-998 | 서울특별시 양천구 목동서로 130 | 목동아파트 4단지 상가동 | 0.000000 | 0.000000 |
| 2 | ST-997 | 서울특별시 양천구 목동중앙로 49 | 목동3단지 시내버스정류장 | 37.534390 | 126.869598 |
| 3 | ST-996 | 서울특별시 양천구 남부순환로88길5-16 | 양강중학교앞 교차로 | 37.524334 | 126.850548 |
| 4 | ST-995 | 서울특별시 양천구 중앙로 153 공중화장실 | NaN | 37.510597 | 126.857323 |
| 5 | ST-994 | 서울특별시 양천구 목동서로161 | SBS방송국 | 37.529163 | 126.872749 |

데이터 출처

기상청 자료 개방 포털 실데이터 활용

데이터 정보

특징: 기온, 강수량, 대기질 정보를 일별로 기록하는 로그 데이터 형태

기간: 2023년 4월 1일 ~ 2024년 3월 31일

구성: 367개의 data, 7개의 Columns

주요 정보: 기준_날짜, 지점명(문자형 데이터), 지점, 미세먼지, 초미세먼지, 평균기온(°C), 일강수량(mm)(숫자형 데이터)

| | 기준_날짜 | 미세먼지 | 초미세먼지 | 지점 | 지점명 | 평균기온(°C) | 일강수량(mm) |
|---|------------|-------|-------|-------|-----|----------|----------|
| 0 | 2023-04-01 | 111.0 | 56.0 | 108.0 | 서울 | 17.5 | NaN |
| 1 | 2023-04-02 | 60.0 | 32.0 | 108.0 | 서울 | 18.8 | NaN |
| 2 | 2023-04-03 | 58.0 | 32.0 | 108.0 | 서울 | 19.0 | NaN |
| 3 | 2023-04-04 | 38.0 | 18.0 | 108.0 | 서울 | 17.6 | 6.9 |
| 4 | 2023-04-05 | 20.0 | 10.0 | 108.0 | 서울 | 11.9 | 50.3 |
| 5 | 2023-04-06 | 73.0 | 55.0 | 108.0 | 서울 | 8.8 | 2.7 |

데이터 접근 방향

HOW?

시간대별 반납량 / 대여량 패턴 파악 WHAT?

따릉이 수요 예측



Part 3 예측 분석을 위한 데이터 선별 및 분류 **01.Business Understanding**

02.Data Understanding

03.Data Preparation

04.Modeling

05.Appendix

기상 데이터

따릉이 재배치의 **정지 조건으로써 기상 데이터 활용** 가능 여부를 확인하기 위한 변수 간의 관계 분석

1. 독립 변수 간의 관계에 대한 가설 수립

- 야외에서 이용하는 따름이의 특성 상 기상 상황에 많은 영향을 받음
- 각 독립 변수의 특정 값에서 따름이 이용량이 큰 영향을 받을 것이라는 가설 수립

2. 각 기상 데이터와 이용량 상관 계수

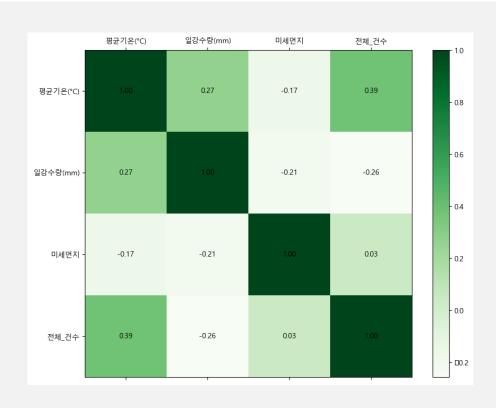
- 미세먼지: 0.062857 (초미세먼지: 0.040196)

- 기온: 0.603032

- 4월 ~10월 강수량: -0.83772

3. 대기질과 따름이 이용량 관계 분석

- 대기질 수치가 좋지 않을 때, 따름이 이용량이 감소하는 가설 수립
- 대기질과 따름이 이용량을 일별 비교한 결과 대기질 수 치에 따른 이용량 변화는 없음



- 대기질과 따릉이 이용량은 두 변수 사이에 유의미한 관계를 찾을 수 없음
- 기온, 강수량 데이터와 따릉이 이용량 간의 유의미한 상관 관계 확인

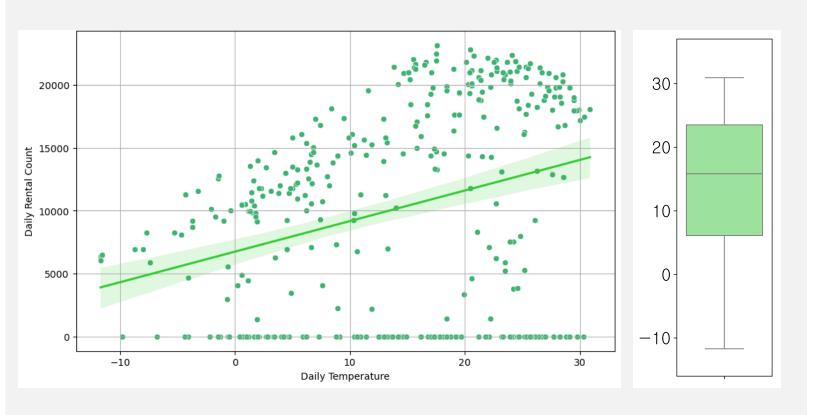
기온과 이용량

1. 가설 수립

- 따릉이 이용량이 급감하는 온도가 있을 것으로 추측
- ARIMA 모델 시계열 분석으로 유의미한 온도 값 탐색

2. 기온과 따름이 이용량 관계 분석

- 기울기가 가장 가파른 온도 값 = 8도
- 이용량이 많은 4월~10월 사이 정지 조건 적용 불가



- 특정 온도 값을 재배치 정지 조건으로 활용 불가
- 온도의 계절성으로 인해 따릉이 이용량과 함께 독립 변수로 사용할 수 없음

강수량과 이용량

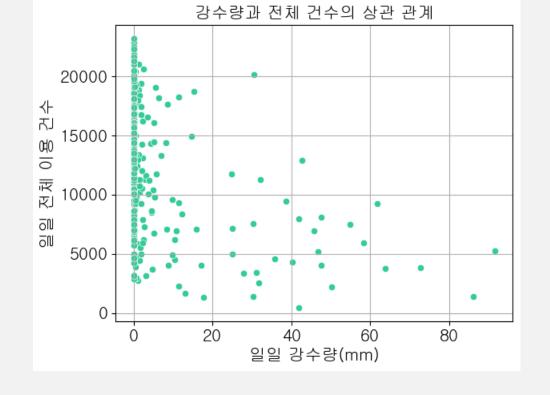
1. 가설 수립 및 전처리

(Ex. 92mm에도 5287건 이용)

- 비가 오면 따름이 이용량이 급감할 것이라는 가설 수립
- 특정 강수량 값 => 재배치 정지 조건
- 비가 오지 않은 날 중 결측치 값을 0mm로 처리

2. 강수량과 따름이 이용량 관계 분석

- 강수량이 적은 날 => 따릉이 이용량이 높음
- 하지만 강수량이 높은 날에도 따릉이를 이용

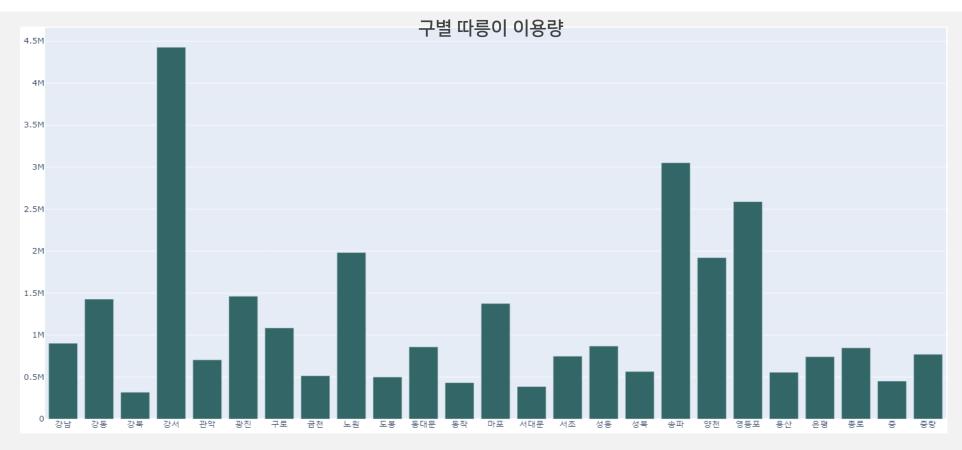


- 서울 전 지역의 일평균 강수량과 강서구 따름이의 시간대별 이용량이라는 원본 데이터의 한계
- '강수량이 4mm 이상일 때 따릉이 대여량이 10분의 1 감소'

(김유신, "공유자전거 따름이 재배치를 위한 실시간 수요예측 모델 연구", 한국 인터넷 정보학회, 24(5), 2023, P.112)

지역별 이용량

각 지역별 특징이 뚜렷한 서울 특성상, 전체 데이터를 일반화할 수 없음(전체데이터약 8,600만개) (23년 4월 ~ 24년 3월 기준)



- 서울시 25개 구 중 이용량이 가장 많은 강서구 데이터 선별
- 시계열 데이터 특성상 분석 정확도를 높이기 위해 가장 최신 데이터를 선별(23.04-24.03)

시간별 이용량

1. 따름이 주 이용 연령대

- 전체 이용자 중 70.4%가 20대(45.5%)와 30대(24.9%)

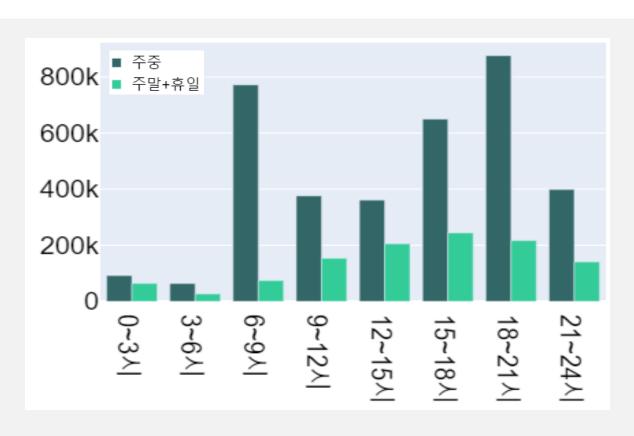
자료 출처: 서울연구원 – 서울시 '따릉이', 누가 언제 이용하나?

2. 평일과 주말, 그리고 공휴일 별 분류

- 평일, 휴일 여부에 따른 이용 양상의 차이(이용 시간, 거리 등)
- 주말 이용량 < 평일 이용량

3. 재배치 최적 시점

- 평일/주말 첨두 시간 이전 = 따릉이 재배치 최적 시점
- 첨두 시간을 기준으로 일 2회 재배치



- 평일의 첨두 시간인 출근 시간과 퇴근 시간의 쏠림 현상을 해결하는 것이 우선 과제

지역 특성 정의

1. 쏠림 현상에 따른 대여소 특성 정의

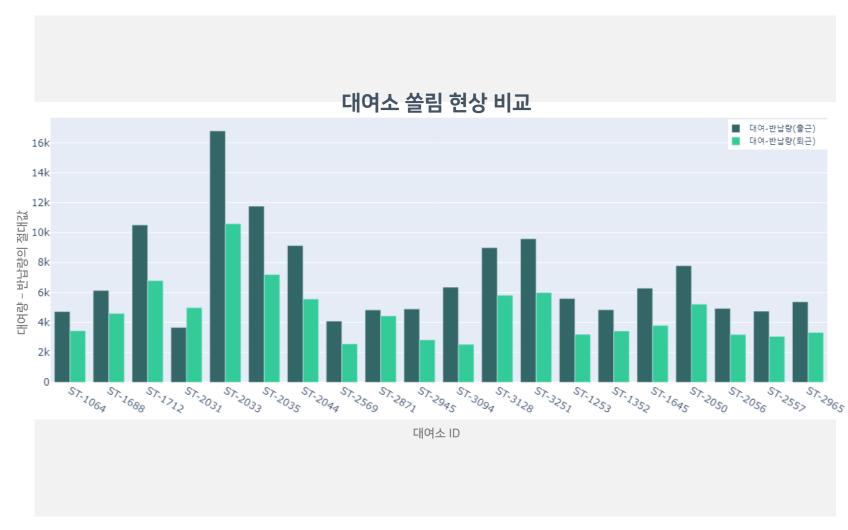
- 출근 (부족): 주거 지구 / 출근 (여유): 상업 지구
- 퇴근 (부족): 상업 지구 / 퇴근 (여유): 주거 지구
- 결과 일치: 특성 정의 / 결과 불일치: 상업 지구로 간주

2. 지적 편집도와의 비교

- 이음지도 크롤링 후 지적 편집도의 지역 특성과 정의된 지역 특성을 비교

3. 재배치가 필요한 대여소 선별

- |(대여량 반납량)| 그래프 내 제일 큰 기울기 값 기준
- 쏠림 현상이 심한 상위 20개 대여소 선별



지역 특성 정의

1. 쏠림 현상에 따른 대여소 특성 정의

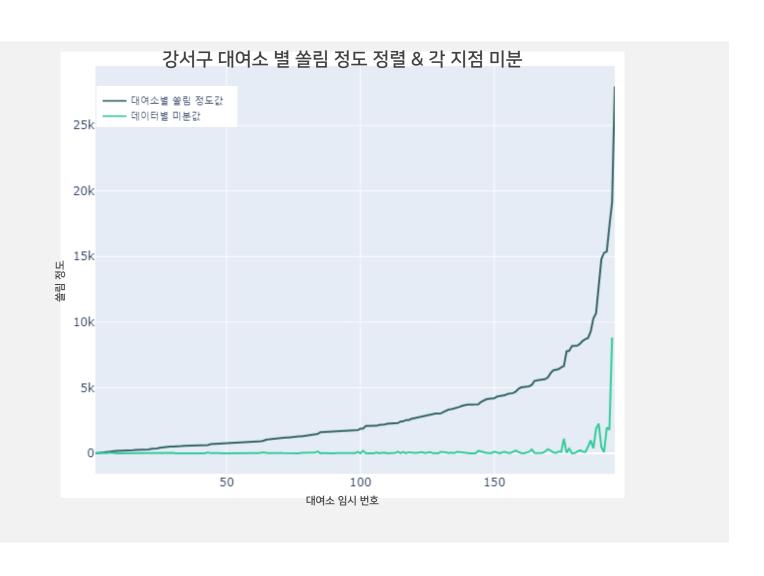
- 출근 (부족): 주거 지구 / 출근 (여유): 상업 지구
- 퇴근 (부족): 상업 지구 / 퇴근 (여유): 주거 지구
- 결과 일치: 특성 정의 / 결과 불일치: 상업 지구로 간주

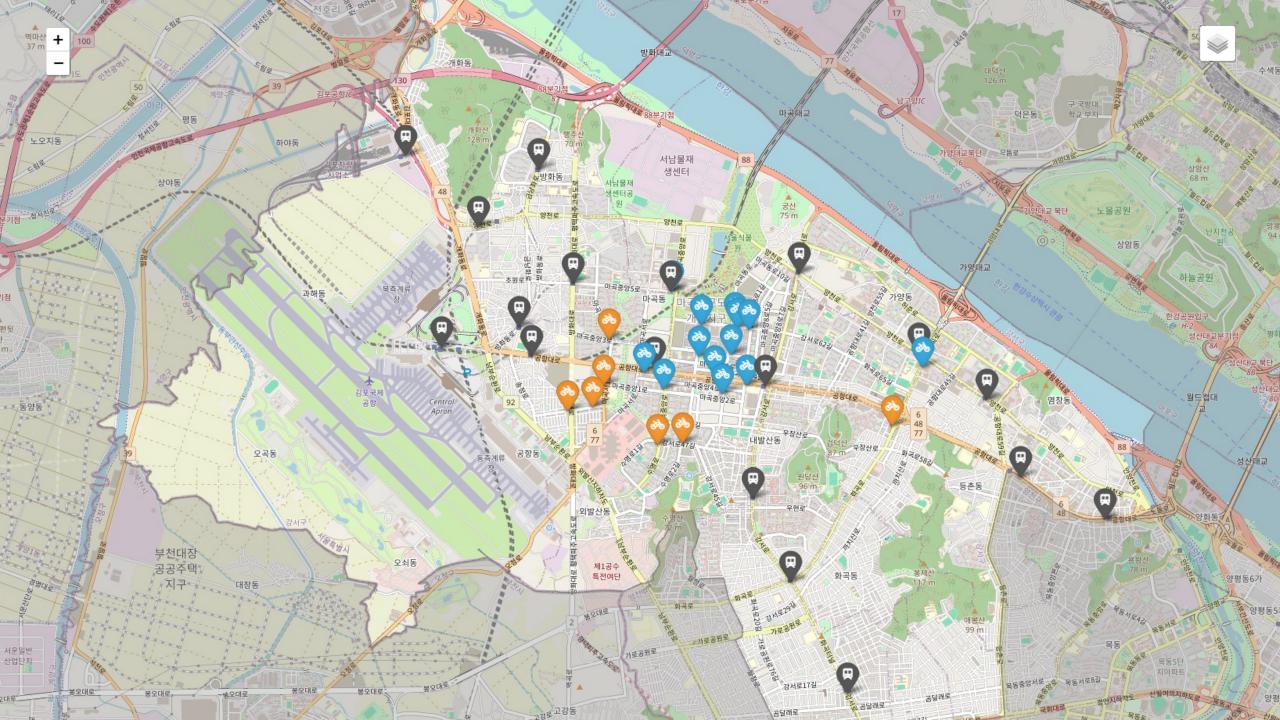
2. 지적 편집도와의 비교

- 이음지도 크롤링 후 지적 편집도의 지역 특성과 정의된 지역 특성을 비교

3. 재배치가 필요한 대여소 선별

- |(대여량 반납량)| 그래프 내 제일 큰 기울기 값 기준
- 쏠림 현상이 심한 상위 20개 대여소 선별





변수 간의 관계

독립 변수 X



종속 변수 Y

시간_범주형 변수 일별 이용량_연속형 변수 지역 정의_이진 변수 강수량_연속형 변수

각 대여소의 시간대별 대여랑

Part 4 시계열 분석을 중심으로 한 이용량 예측

01.Business Understanding

02.Data Understanding

03.Data Preparation

04.Modeling

05.Appendix

Modeling

Time Series Forecast Modeling

목적



강서구 내 선별한 대여소 별



출근 및 퇴근 시간의

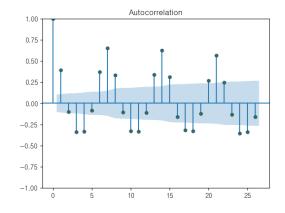


예측한다.

따릉이 **대여량** 및 **반납량**을

사용 모델

SARIMA



모델 선택 이유

- 계절성을 갖는 데이터 특성을 고려
- 일반 ARIMA에서 계절성 패턴을 추가한 SARIMA 모형 선정

Modeling

SARIMA

1. SARIMA 파라미터

- SARIMA 모델 파라미터 : (p, d, q)(P, D, Q, m)

2. SARIMA 파라미터 최적화 과정

- 계절성 주기를 나타내는 m은 데이터 기준(평일)이 5일 주 기로 반복되므로 5로 설정
- 데이터의 계절성 유무에 따라 결정되는 D는 1로 설정
- 이외 파라미터는 pmdatima 라이브러리 내장 함수 auto arima를 사용하여 최적의 하이퍼 파라미터 선별

p: AR(Auto-Regressive) 파라미터

d : 차분(Differencing) 차수

q: MA(Moving Average) 파라미터

P: 계절성 AR 파라미터

D: 계절성 차분 차수

Q: 계절성 MA 파라미터

m : 계절성 주기

예측 값과 실제 값 비교

| | 대여소 ID | all de est | | | | |
|----|---------|------------|----|-------------|----|------|
| | -11-110 | 예측값 | 지구 | 시간대_구간 | 출퇴 | 실제 값 |
| 0 | ST-1253 | 56.293843 | 주거 | 06:00~10:00 | 대여 | 32 |
| 1 | ST-1352 | 39.957055 | 주거 | 06:00~10:00 | 대여 | 56 |
| 2 | ST-1645 | 61.012862 | 주거 | 06:00~10:00 | 대여 | 54 |
| 3 | ST-2050 | 58.811373 | 주거 | 06:00~10:00 | 대여 | 52 |
| 4 | ST-2056 | 33.379616 | 주거 | 06:00~10:00 | 대여 | 34 |
| | | | | | | |
| 75 | ST-2871 | 18.121891 | 상업 | 17:00~21:00 | 반납 | 53 |
| 76 | ST-2945 | 18.161601 | 상업 | 17:00~21:00 | 반납 | 22 |
| 77 | ST-3094 | 22.000821 | 상업 | 17:00~21:00 | 반납 | 51 |
| 78 | ST-3128 | 2.572052 | 상업 | 17:00~21:00 | 반납 | 5 |
| 79 | ST-3251 | 8.364749 | 상업 | 17:00~21:00 | 반납 | 8 |

Modeling

모델링 결과 활용

1. 활용 방안

- 모델링 결과로 나온 시간대의 대여소의 반납/대여 예측 량을 실시간 거치 수량에 반영 후 과잉대여소에서 부족 대여소로 재고 조정을 실시
- Ex. 출근 시간에는 주거 지구의 수요량이 증가하므로 출근 시간 전 상업지구의 재고를 주거 지구로 재배치

2. 활용 방안을 통한 기대 이점

- 재배치에 들어가는 비용 절감 효과
- 쏠림 현상 완화로 인한 이용률 향상
- 소비자 경험 개선을 통한 브랜드 이미지 제고



대여소 별 반납 예측량 대여소 별 대여 예측량





실시간 거치 수량





재배치 필요 대수 할당

의미 있는 점

쏠림 현상을 정의하고 해당 기준을 통해 상업 지구와 주거 지구 분류

- 재배치 대상 대여소를 분별
- 이를 통해 유의미한 재배치 기준점 도출

'기상 데이터와 이용량 간의 상관관계가 존재한다.'는 귀무가설 검정을 통해 귀무 가설 기각 여부 확인

- '기온이 높으면 이용량이 줄어들 것이다.' 등의 귀무 가설이 기각될 수 있는 것을 데이터를 통해 확인

기존 일 3회 재배치 >> 최적의 재배치 시점 일 2회

- 비용 절감 효과 기대 가능

아쉬운 점

모델의 예측값과 실제값의 차이가 큼

- 사용자 이용 패턴의 복잡성으로 인해 결론을 도출하는데 어려움

수집 데이터 간의 정합성 부족으로 인해 사용 불가능한 데이터가 발생

Part 5

EDA 및 전처리, 예측 모델링 아카이브

01.Business Understanding

02.Data Understanding

03.Data Preparation

04.Modeling

05.Appendix

APPENDIX

디자인 아이덴티티

Font

나눔스퀘어 OTF

Idetity Color

336666 / R: 51, G: 102, B: 102

33cc99 / R: 51, G: 204, B: 153



339966 / R: 51, G: 153, B: 102



Team 3 Semi Project_멀티잇 데이터 분석&서비스개발(Python) 8회차

28

Appendix

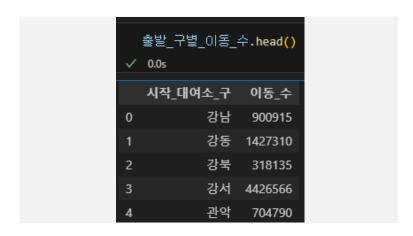
데이터 전처리

서울시 1년간 데이터 통합 > 강서구, 평일(공휴일 제외) 데이터 범위 한정을 위한 전처리

| | 기준_날짜 | 기준_시간대 | 시작_대여소_ID | 종료_대여소_ID | 전체_건수 | 전체_이용_분 | 전체_이용_거리 | 요일 | 주말 | 휴일 |
|---|------------|--------|-----------|-----------|-------|---------|----------|----|----|----|
| 0 | 2023-05-08 | 200 | ST-2037 | ST-2037 | 1 | 1 | 30 | 0 | 0 | 0 |
| 1 | 2023-05-08 | 1155 | ST-824 | ST-824 | 1 | 1 | 80 | 0 | 0 | 0 |
| 2 | 2023-05-08 | 1230 | ST-829 | ST-829 | 1 | 1 | 111 | 0 | 0 | 0 |
| 3 | 2023-05-08 | 1340 | ST-2611 | ST-2611 | 1 | 1 | 10 | 0 | 0 | 0 |
| 4 | 2023-05-08 | 1635 | ST-2031 | ST-2031 | 1 | 1 | 170 | 0 | 0 | 0 |

데이터 전처리

시작 대여소 위치를 기반으로 "구" 단위의 전처리 > 이동 수 EDA



Team 3 Semi Project_멀티잇 데이터 분석&서비스개발(Python) 8회차

Appendix

데이터 전처리

최근 1년 간 일별 평균 기상 데이터 수집







데이터 전처리

출퇴근 시간 별 |대여량 - 반납량|의 값을 대여소_ID를 기준 합쳐 전처리

| | Unnamed: 0 | 대여 소_ID | 출근대 여량 | 출근반 납량 | 대여 - 반 납량_출 근 | 차이절 대값_출 근 | 총이용 량_출근 | 퇴근대 여량 | 퇴근반 납량 | 대여-반 납량_퇴 근 | 차이절 대값_퇴 근 | 총이용 량_퇴근 | 출퇴근_ 총이용량 | 최 종 용 도 |
|---|---------------|-------------|-----------|-----------|----------------------------|------------------|-------------|-----------|-----------|-------------------|------------------|-------------|--------------|------------------|
| 0 | 0 | ST- 1062 | 3743 | 2281 | 1462 | 1462 | 6024 | 3947 | 5456 | -1509 | 1509 | 9403 | 15427 | 주 거 |
| 1 | 1 | ST- 1063 | 5029 | 8839 | -3810 | 3810 | 13868 | 5795 | 3170 | 2625 | 2625 | 8965 | 22833 | 상 업 |
| 2 | 2 | ST- 1064 | 3764 | 8489 | -4725 | 4725 | 12253 | 9005 | 5562 | 3443 | 3443 | 14567 | 26820 | 상 업 |
| 3 | 3 | ST- 1065 | 5860 | 5507 | 353 | 353 | 11367 | 14431 | 16569 | -2138 | 2138 | 31000 | 42367 | 주 거 |
| 4 | 4 | ST- 1067 | 3394 | 2906 | 488 | 488 | 6300 | 2666 | 3654 | -988 | 988 | 6320 | 12620 | 주 거 |

Team 3 Semi Project_멀티잇 데이터 분석&서비스개발(Python) 8회차

Appendix

데이터 전처리

기울기 값이 가장 높은 값을 기준으로, "쏠림 현상 " 으로 인한 재배치를 해야하는 대여소 선별



데이터 전처리

Folium 시각화를 위한 대여소 ID 별 주소 값 추가



31

Appendix

데이터 전처리

기울기 값이 가장 높은 값을 기준으로, "쏠림 현상 " 으로 인한 재배치를 해야하는 대여소 선별



데이터 전처리

Folium 시각화를 위한 대여소 ID 별 주소 값 추가



