

SAI

pandas를 활용한

6조

육캔두잇

데이터분석프로젝트

S+

지역별 지하철 이용량 분석 및 날씨와의 상관관계 시각화

분석 대상
분석 내용

사용 데이터
사용 프로그램

지하철
이용량 분석

날씨와의
상관관계 분석

결론 및 한계점

지역별 지하철 이용량 분석 및 날씨와의 상관관계 시각화

분석 대상:

서울(종로3가역, 1호선), 인천(인천역, 경인선), 수원(수원역, 경부선),
춘천(춘천역), 부산(부산역)의 2022년 승하차량 및 기후 정보

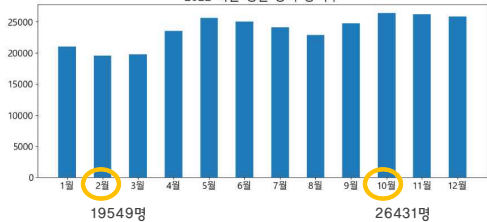
분석 내용:

- ① 5개 역별 이용량 분석
- ② 역별 이용량과 계절별 날씨의 연관성 분석
- ③ 분석 결과의 시각화

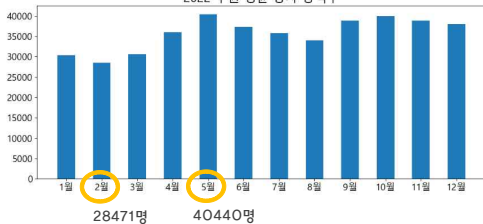
- 기상 자료 개방 포털
(<https://data.kma.go.kr/cmmn/main.do>)
- 서울 열린 데이터 광장
(<https://data.seoul.go.kr/>)
- 부산 공공데이터 포털
(<https://data.busan.go.kr>)

 pandas
NumPy matplotlib
seaborn

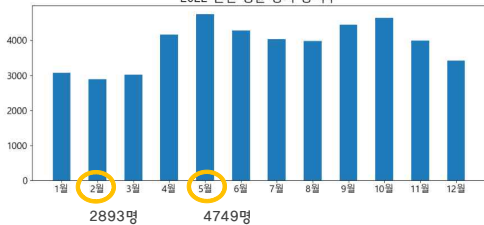
2022 서울 평균 승차 승객수



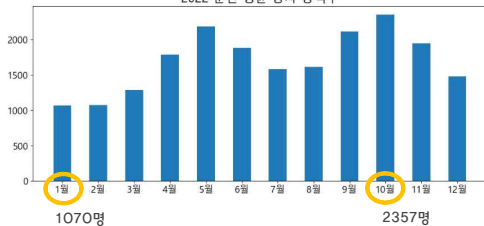
2022 수원 평균 승차 승객수

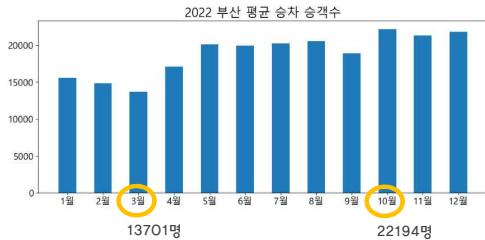


2022 인천 평균 승차 승객수

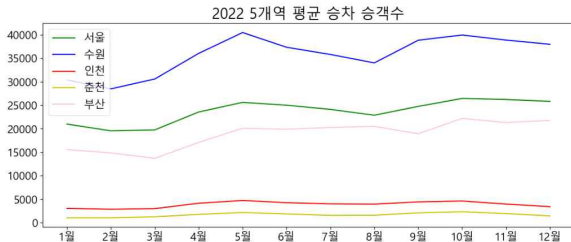


2022 춘천 평균 승차 승객수





지하철 이용량 분석



**봄**

- 4~5월
- 기온
 - 강수량
 - 미세먼지

**여름**

- 7~8월
- 기온
 - 강수량
 - 습도

**가을**

- 9~10월
- 기온
 - 미세먼지
 - 적운량

**겨울**

- 12~1월
- 기온
 - 강수량(눈 포함)
 - 풍속

서울

총 승객 수	1	0.77	-0.068	0.25
기온	0.77	1	0.12	0.21
강수량	-0.068	0.12	1	-0.078
미세먼지	0.25	0.21	-0.078	1

총 승객 수

기온

강수량

미세먼지

수원

1	0.7	-0.028	0.44
0.7	1	0.19	0.4
-0.028	0.19	1	-0.2
0.44	0.4	-0.2	1

인천

1	0.86	-0.06	-0.16
0.86	1	0.071	-0.12
-0.06	0.071	1	0.061
-0.16	-0.12	0.061	1

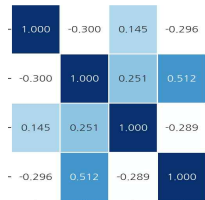
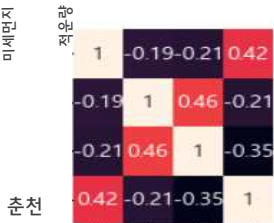
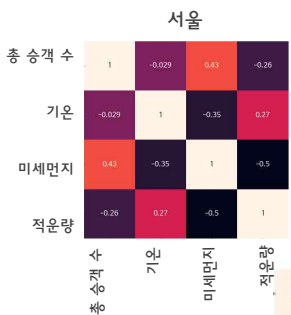
춘천

1	0.21	0.28	-0.083
0.21	1	0.58	0.016
0.28	0.58	1	0.38
-0.083	0.016	0.38	1

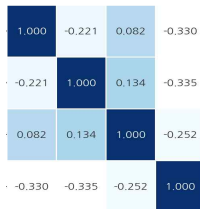
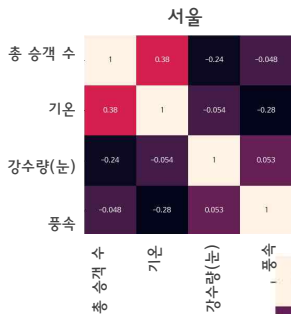
1.000	-0.040	0.383	0.375
-0.040	1.000	0.335	0.458
0.383	0.335	1.000	0.875
0.375	0.458	0.875	1.000

부산





부산



부산

-0.5419	강수량	인천	여름		0.4877	기온	수원	봄
-0.4953	적운량	서울	가을		0.4815	기온	인천	겨울
-0.4952	강수량	수원	여름		0.4350	미세먼지	수원	가을
-0.4639	습도	인천	여름		0.3960	미세먼지	춘천	가을
-0.4259	습도	수원	여름		0.3084	기온	춘천	봄
-0.3759	적운량	수원	가을		0.3053	기온	부산	봄
-0.3718	강수량	춘천	여름					
-0.3261	풍속	부산	겨울					

여름의
강수량



여름의
습도

봄의
기온



분석 결과의 활용 방안

- 배차 간격 설정에 사용
- 역 내 시설 점검 시간 결정에 사용
- 역 내 상점 운영시간 결정에 사용

- 날씨와 지하철 이용량의 상관관계 잘 보이지 않음
- 시각화를 더 해보지 못한 아쉬움
- 공부한 내용을 활용해볼 수 있는 기회였음
- 상관관계 분석, 가설 검정을 통해 결과의 신뢰성을 높이고 싶음