

Portfolio

정지원(JUNG JI WON)

010-9253-6036

jwjung323@naver.com

About Me

데이터 속에서 가치를 찾는 신입 데이터 분석가 정지원입니다.

- 학력
 - 2021.3 백석대학교 컴퓨터공학부 인공지능학과 입학
 - 2025.2 백석대학교 컴퓨터공학부 인공지능학과 졸업

인공지능을 전공하며 자연스럽게 데이터를 다루는 것에 흥미를 가지게 되었습니다.
학교 수업과 다양한 프로젝트를 통해 데이터를 수집·분석·시각화하며 여러 실전 경험을 쌓아왔습니다.

정지원 (JI WON JUNG)
2002.03.23
010-9253-6036
jwjung323@naver.com
경기도 오산시 거주

- 스킬
 - Python
 - SQL(MySQL)
 - Tableau

Skill

언어

- Python
 - (pandas, Numpy, matplotlib, sklearn)
 - 데이터 전처리 및 분석을 위해 적절한 코드를 작성할 수 있으며, 데이터에 대한 기초 통계 및 시각화를 진행할 수 있습니다.
 - 분석 목적에 맞게 예측 모델링을 구현할 수 있습니다.

SQL

- SQL(MySQL)
 - 데이터베이스 구조를 이해하고, 필요한 데이터를 추출할 수 있습니다.
 - 추출한 데이터를 Python과 연계하여 가공, 분석할 수 있습니다.

BI 도구

- Tableau
 - 데이터 비교 및 발표 자료용 시각화를 제작해 본 경험이 있습니다.

기타

- ERDCloud
 - 데이터베이스에 저장된 데이터의 관계를 시각적으로 설계하여 데이터 구조를 쉽게 이해하고 관리할 수 있도록 활용했습니다.
- Exel
 - 기본 함수 사용이 가능하며, 데이터 간의 관계 분석을 할 때 데이터 분석 툴을 활용할 수 있습니다.

구독서비스 데이터 A/B테스트 분석

2025.03.27 ~ 2025.04.01
개인 프로젝트

구독서비스 데이터 A/B테스트 분석

기간: 2025.03.27 ~ 2025.04.01

분석 툴: Python, Tableau, Exel(구글 스프레드시트)

- 개요
 - A/B테스트를 실제 적용해보기 위한 목적으로 진행한 개인프로젝트로 A군과 B군으로 나누어진 데이터를 사용해 진행하였다.
- 목표
 - A안(1개월 및 12개월 멤버십 가격을 단순 비교하여 보여줌)과 B안(12개월 멤버십의 월간 가격을 강조하여, 더 경제적인 선택이라는 점을 부각)으로 나누어진 구독 페이지 이용자들에 대해 A/B테스트를 적용하여 비교분석을 진행
- 데이터
 - 구독 서비스 일자별 데이터(30일 분량)(2023.09.15~2023.10.14)

페이지 방문자 수	
구독 전환율	페이지 방문자 중 실제 구독을 완료한 비율
구독자 수	1개월 및 12개월 구독자 수
수익	각실험안에서 발생한 총 수익

구독서비스 데이터 A/B테스트 분석

- 분석 과정
 - EDA
 - A그룹과 B그룹으로 데이터 프레임을 나누어 진행

- 결과

비교군	A안(기존안)	B안(개선안)
1개월 구독자 비율(%)	82.77	80.62
12개월 구독자 비율(%)	17.23	19.38
총 구독 전환율(%)	0.049	0.068
총 수익	59157	88354
12개월 구독자 수익	37632	59094
총 수익대비 12개월 구독자 수익 비율	0.636	0.669

- 결과 활용 방안
 - 통계적 분석또한 진행한 결과 P값이 유의미함을 확인하였으므로 개선안으로 변경을 추진하는것이 옳음.

- 코드 링크

HTTPS://GITHUB.COM/JIWON-JUNG323/GIT_REPO/BLOB/MAIN/PROJECT1/PROJECT1.IPYNB

자전거 대여 시스템 데이터를 이용한 자전거 이용 패턴 분석

2025.05.09 ~ 2025.05.20
개인 프로젝트

자전거 대여 시스템 데이터를 이용한 자전거 이용 패턴 분석

기간: 2025.05.09 ~ 2025.05.20

분석 툴: Python(pandas, numpy, matplotlib, sklearn)

- 개요
 - 예측모델링 설계에 대한 실습용 프로젝트로 자전거 대여 패턴을 분석하여 자전거 배치 및 운영 전략을 최적화하고 대여 수요를 정확히 예측하고자 함.
- 목표
 - 자전거 대여 패턴을 분석하여 대여 수요를 예측하는 모델을 설계.
- 분석 과정
 - 데이터 전처리
 - 결측치 제거
 - 데이터의 형태가 올바르지 않은 데이터들에 대해 형변환 실행(DATETIME)

- 데이터(12 컬럼으로 구성, 학습군과 테스트군으로 나누어져 있으나 테스트군에는 COUNT컬럼 미포함)

컬럼명	설명
datetime	자전거 대여 기록의 날짜 및 시간
season	계절(1:봄, 2:여름, 3:가을, 4:겨울)
holiday	공휴일 여부(0: 평일, 1: 공휴일)
workingday	근무일 여부(0: 주말/공휴일, 1: 근무일)
weather	날씨 상황(1: 맑음, 2: 구름짐/안개, 3: 약간의 비/눈, 4: 폭우/폭설)
temp	실측 온도(섭씨)

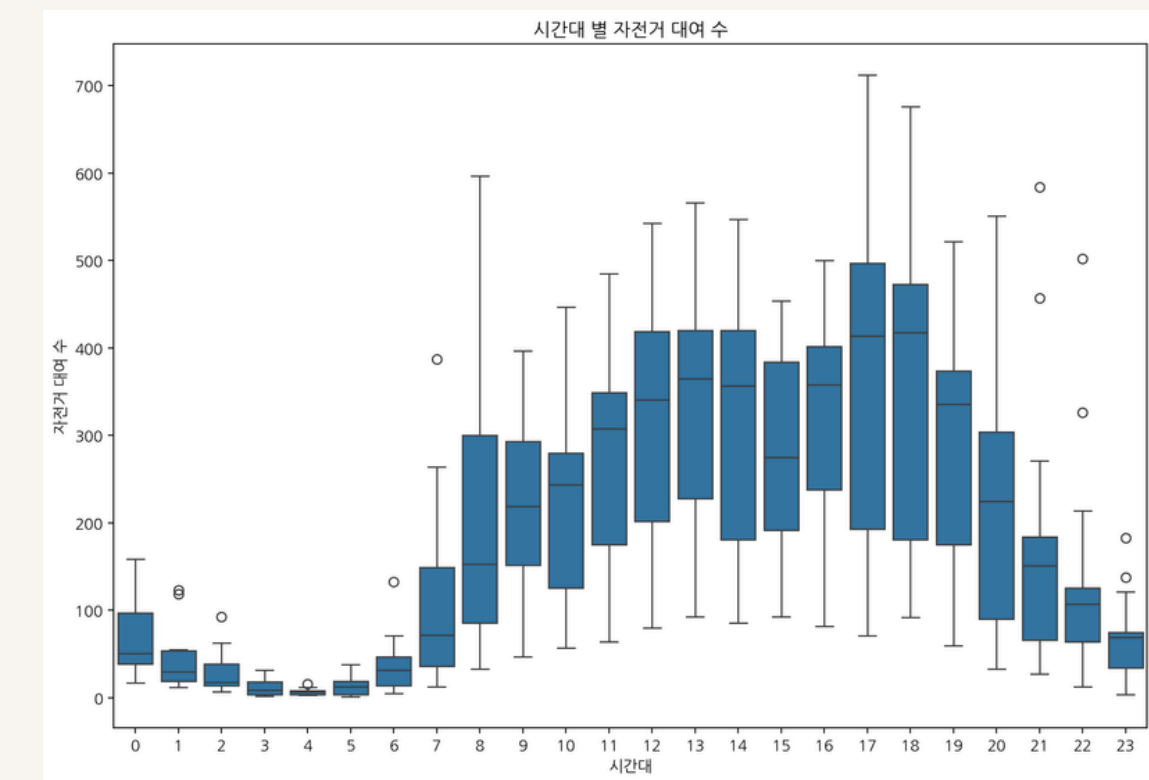
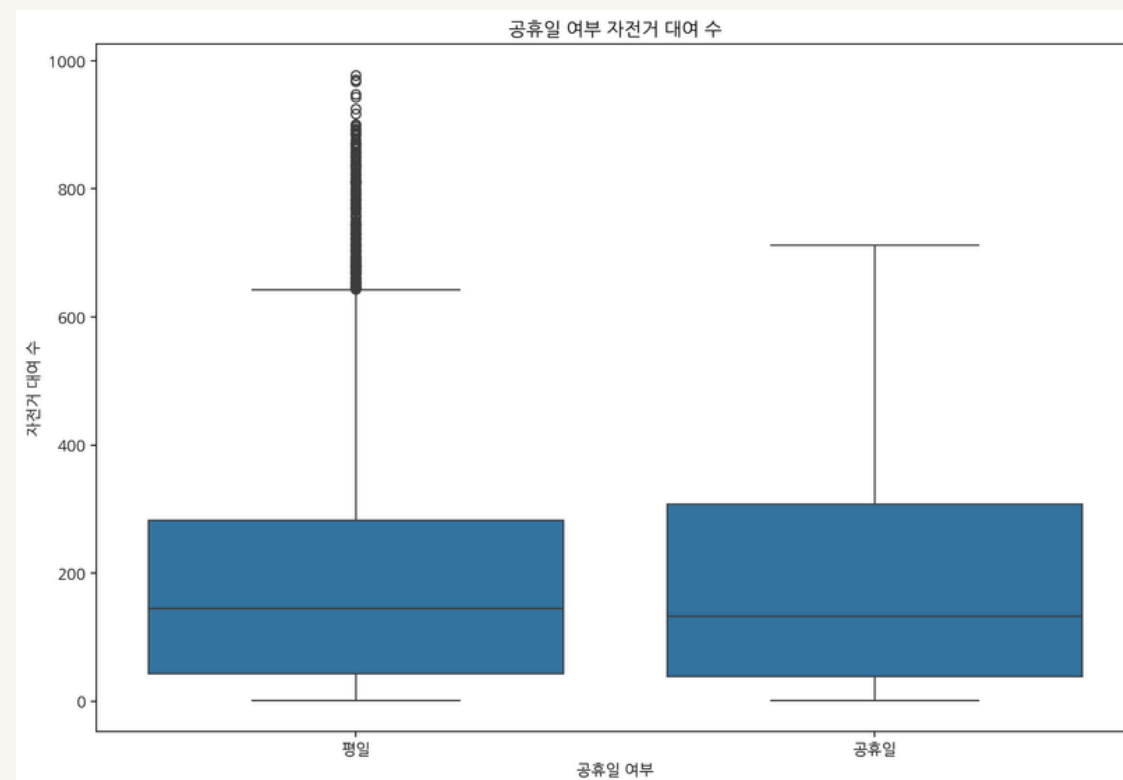
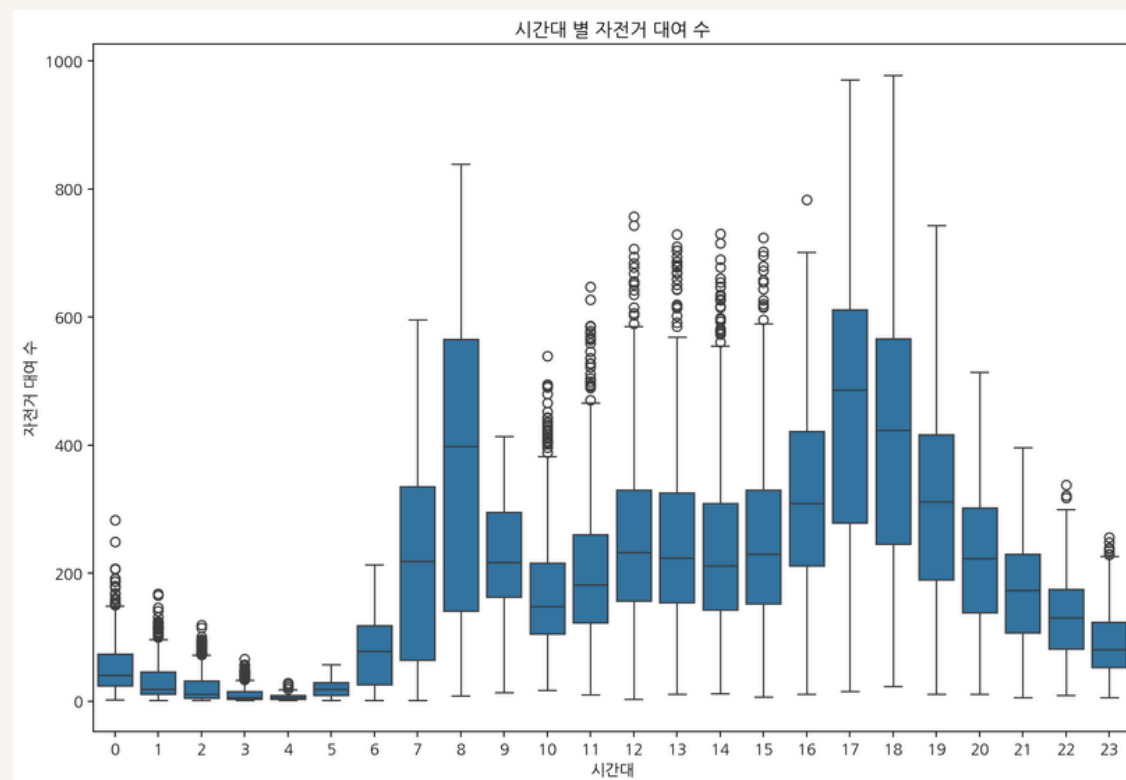
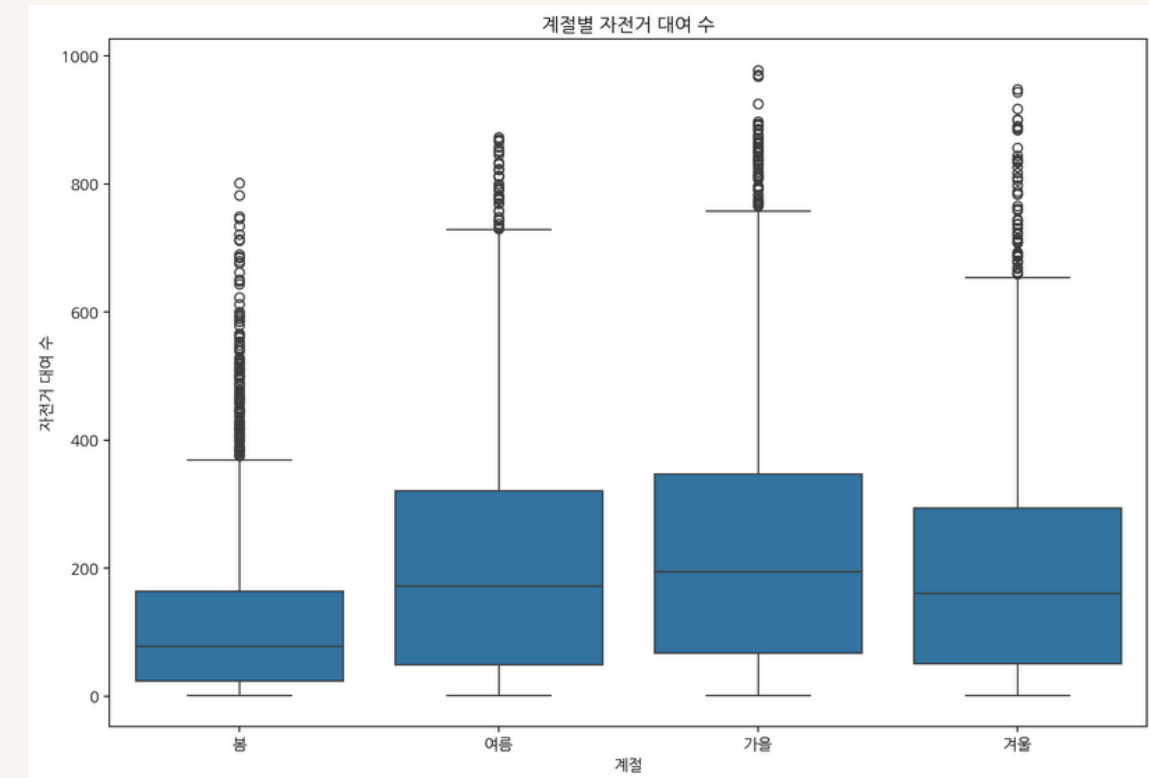
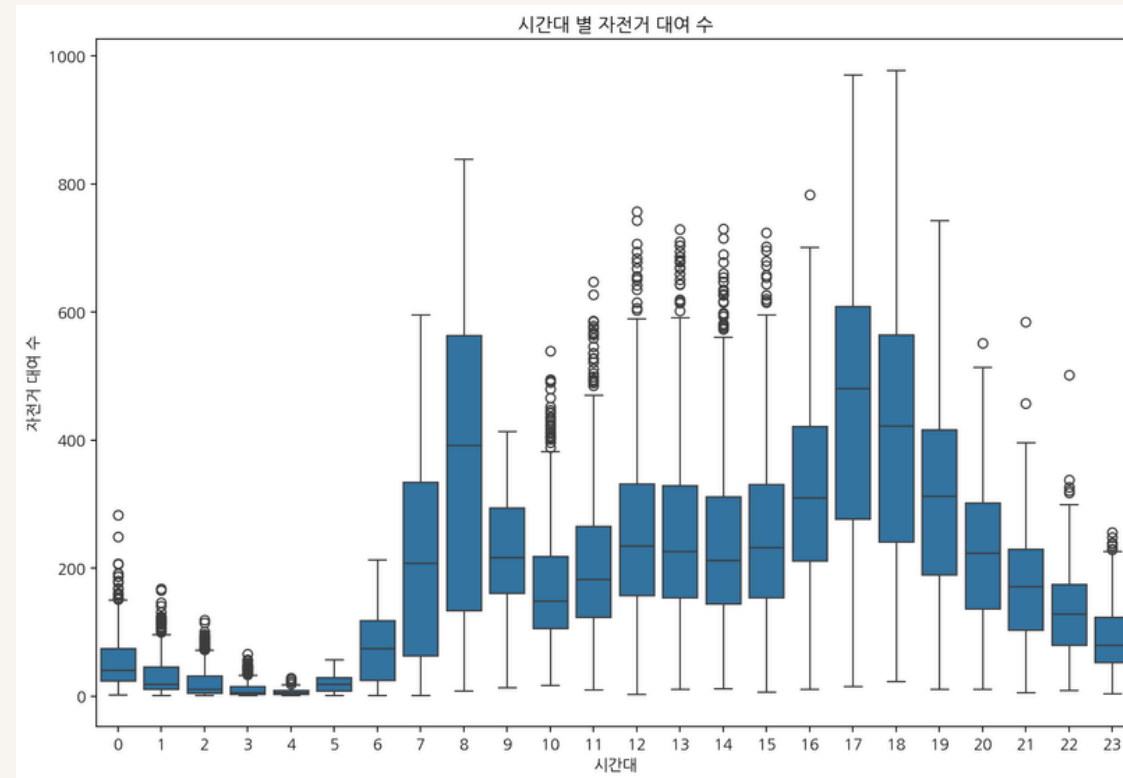
컬럼명	설명
atemp	체감 온도(섭씨)
humidity	습도(%)
windspeed	풍속(m/s)
casual	등록되지 않은 사용자의 대여 수
registered	등록된 사용자의 대여 수
count	총 대여수 (종속변수)

자전거 대여 시스템 데이터를 이용한 자전거 이용 패턴 분석

- 분석 과정

- EDA

- 시간대 별 자전거 대여 패턴 분석
 - 계절별 자전거 대여 패턴의 차이 분석
 - 공휴일 여부 자전거 대여 수 분석
 - 시간대 별 자전거 대여수(공휴일)



자전거 대여 시스템 데이터를 이용한 자전거 이용 패턴 분석

- 주요 결과 & 시각화

- 히트맵 분석결과

- 시간(HOUR)과 대여 건수(COUNT) 간 상관관계가 가장 높은것을 확인
 - 특정 시간대 별로 수요 패턴이 뚜렷함을 의미함.
 - 계절(SEASON) 또한 대여 건수와 양의 상관관계인 것을 확인
 - 계절별 수요 차이가 존재함을 의미함
 - 공휴일 여부(HOLIDAY)와는 약한 음의 상관관계인 것을 확인
 - 공휴일에는 대여 패턴이 일정하지 않음을 의미함.

- 선형회귀분석을 사용한 예측모델 구현

- 결과 활용 방안

- 수요 예측을 통해 자전거 사전 배치

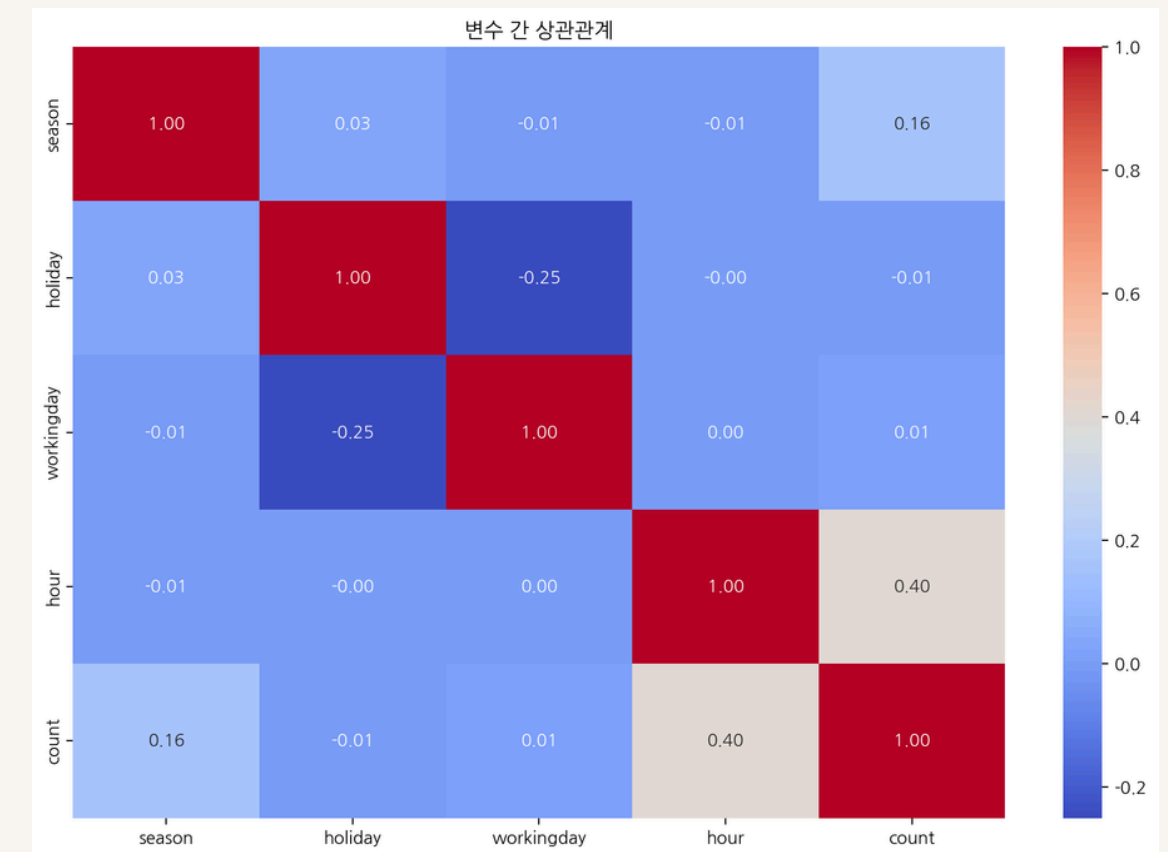
- 특정 시간대나 계절에 맞춘 공급을 계획하여 운영 비용 절감 및 이용자 만족도를 높일 수 있다.

- 마케팅/프로모션 전략 수립

- 수요가 낮은 시간대/계절에 할인이나 이벤트를 계획하여 이를 보완한다.
 - 수요가 많은 시간대에는 적절한 요금 정책으로 수익을 극대화시킬 수 있다.

- 코드 링크

- [HTTPS://GITHUB.COM/JIWON-JUNG323/GIT_REPO/BLOB/MAIN/PROJECT2/PROJECT2.IPYNB](https://github.com/JIWON-JUNG323/GIT_REPO/blob/main/project2/project2.ipynb)



- ✓ 검증 데이터 성능

RMSE: 164.01

R²: 0.19

회귀계수: [10.5544872 26.93974052]

절편: 2.147308497257285

- ✓ 검증셋 예측값 샘플

	hour	season	true_count	predicted_count
3133	11	3	127	199.065889
5786	6	1	13	92.413972
5224	18	4	163	299.887040
8953	10	3	233	188.511402
8054	23	2	222	298.779995

- ✓ 제출용 데이터 예측값 샘플

	datetime	count
0	2011-01-20 00:00:00	29.087049
1	2011-01-20 01:00:00	39.641536
2	2011-01-20 02:00:00	50.196023
3	2011-01-20 03:00:00	60.750511
4	2011-01-20 04:00:00	71.304998

서울시 자치구별 따릉이 이용량 차이

2025.02.19 ~ 2025.02.27
협업프로젝트(3인)

서울시 자치구별 따릉이 이용량 차이

기간: 2025.02.19 ~ 2025.02.27

분석 툴: Python, Tableau, Excel(구글 스프레드시트)

팀 내 역할: 팀장, 프로젝트 설계, 주요 코드 설계

- 개요

- 대중교통 서비스관련 데이터들을 살펴보던 중 서울시 공공 자전거 대여 서비스인 '따릉이'의 이용량이 자치구별로 큰 차이가 발생함을 확인하여 이에 대한 의문을 가지게 되었습니다.

- 목표

- **이용량 차이의 발생 원인을 파악하는 것을 목표로 하였습니다.**

- 데|이|터

- 서울특별시 공공자전거 이용정보
- 서울특별시 공공자전거 대여소 정보
- 서울시 자동차도로면적
- 서울시 자전거도로 현황
- 서울교통공사 자치구별지하철역정보
- 서울교통공사 역주소
- 서울시 지하철호선별 역별 승하차 인원 정보
- 서울시 사업체정보

- 분석 과정

- 데이터 전처리
 - 주말과 평일로 나누어 보고싶은 가설들이 있어 이에대해 전처리 진행
 - 결측치 제거
 - 데이터 형변환(DATETIME)



서울시 자치구별 따릉이 이용량 차이

- EDA

- 가설1. 대여소 개수가 많을수록 해당 지역 자전거 이용률이 높을 것이다.

- 평일 73%, 주말 68%로 신뢰성이 높게 나타났습니다.
- 모델이 유의미함을 확인하였습니다.

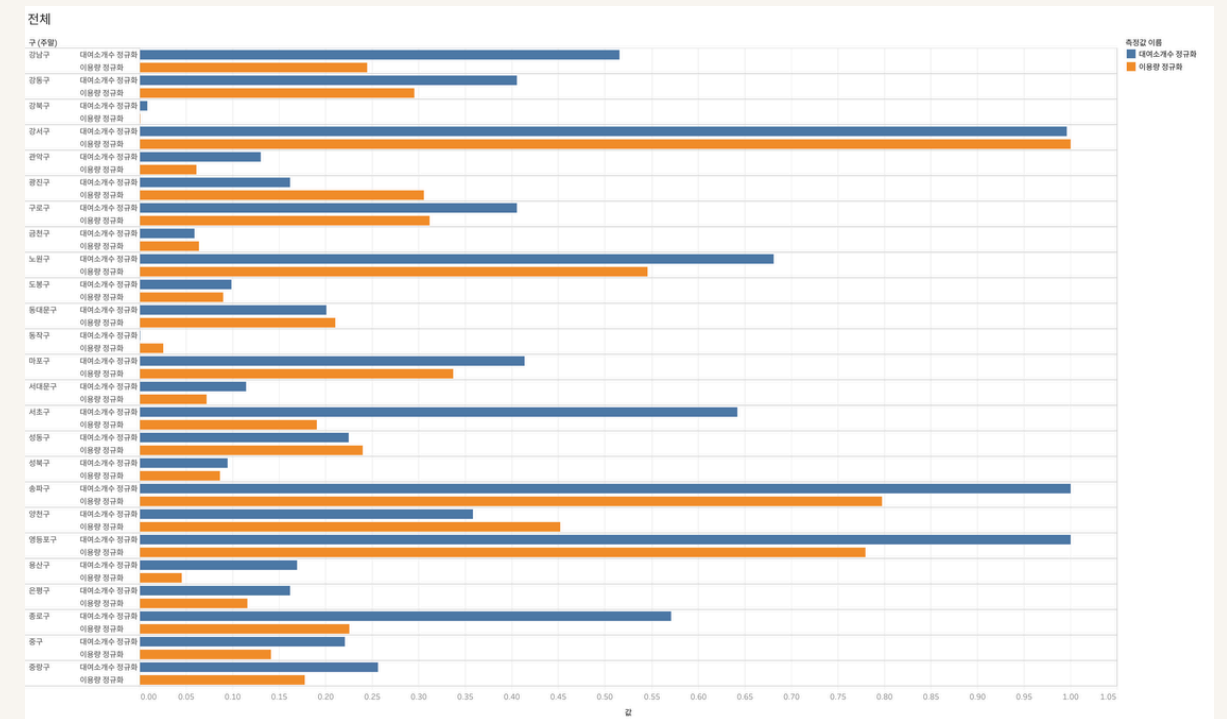
○ P값

- **평일:** 0.00000005310498852
- **주말:** 0.0000003617452922

- 대여소개수가 1 단위 증가할 때 마다, 이용량은 평균적으로 0.8548 만큼 증가했습니다.(평일)
- 대여소개수가 1 단위 증가할 때 마다, 이용량은 평균적으로 0.8262 만큼 증가했습니다.(주말)

- 가설2. 도로면적대비 자전거 도로 면적이 높을수록 자전거 이용률이 높을것이다.

- 분석을 진행할 때 자치구별로 면적의 차이가 있음을 고려하여 전체 면적 대비 자전거도로 비율로 진행
- 약 34%의 설명력을 가지고있습니다.
- 모델이 유의미함을 확인했습니다.(P = 0.002)
- 자전거도로비율이 1이 늘어날 때마다 이용량이 0.5832 증가했습니다.



서울시 자치구별 따릉이 이용량 차이

- EDA

- 가설3-1. 해당 구에 지하철역이 많을수록 자전거 이용률이 높을 것이다.
 - 약 11%의 설명력을 가지고 있습니다.
 - 모델이 유의미하지 않습니다.(P = 0.1)
- 가설 3-2. 해당 구에 지하철 하차총승객수가 많을수록 자전거 이용률이 높을 것이다.
 - 약 5%의 설명력을 가지고 있습니다.
 - 모델이 유의미하지 않습니다.(P = 0.2)
- 가설 3-3. 해당 구에 지하철 승차총승객수가 많을수록 자전거 이용률이 높을 것이다.
 - 약 5%의 설명력을 가지고 있습니다.
 - 모델이 유의미하지 않습니다.(P = 0.2)

- 프로젝트를 마무리하고 가설3번에대해 좀 더 살펴보던 도중 승차와 하차에 대해 분석을 진행할때 승차승객수와 반납한 이용데이터를 분석하고 하차승객수와 대여한 이용데이터를 분석해야 했는데 자전거 이용에 대해 구분없이 진행하여 오류가 발생한 것을 발견. 이에대해 추후에 프로젝트를 진행하고자함.

서울시 자치구별 따릉이 이용량 차이

- EDA

- 가설4. 해당 구에 사업체 개수가 많을수록 자전거 이용률이 높을 것이다.
 - 신뢰성이 비교적 높게 나타났습니다.
 - 전체 : 약 83.7%
 - 평일 : 약 83%
 - 주말 : 약 84%
 - 사업체 수와 이용량 간의 상관관계가 무의미함을 확인했습니다.($P>0.05$)
 - 종사자 수와 이용량 간의 상관관계가 유의미함을 확인했습니다. ($P<0.05$)
 - 종사자 수가 1 증가할 때마다 이용량이 0.9010 감소했습니다.(전체)
 - 종사자 수가 1 증가할 때마다 이용량이 0.8763가 감소했습니다.(평일)
 - 종사자 수가 1 증가할 때마다 이용량이 0.9754가 감소했습니다.(주말)

서울시 자치구별 따릉이 이용량 차이

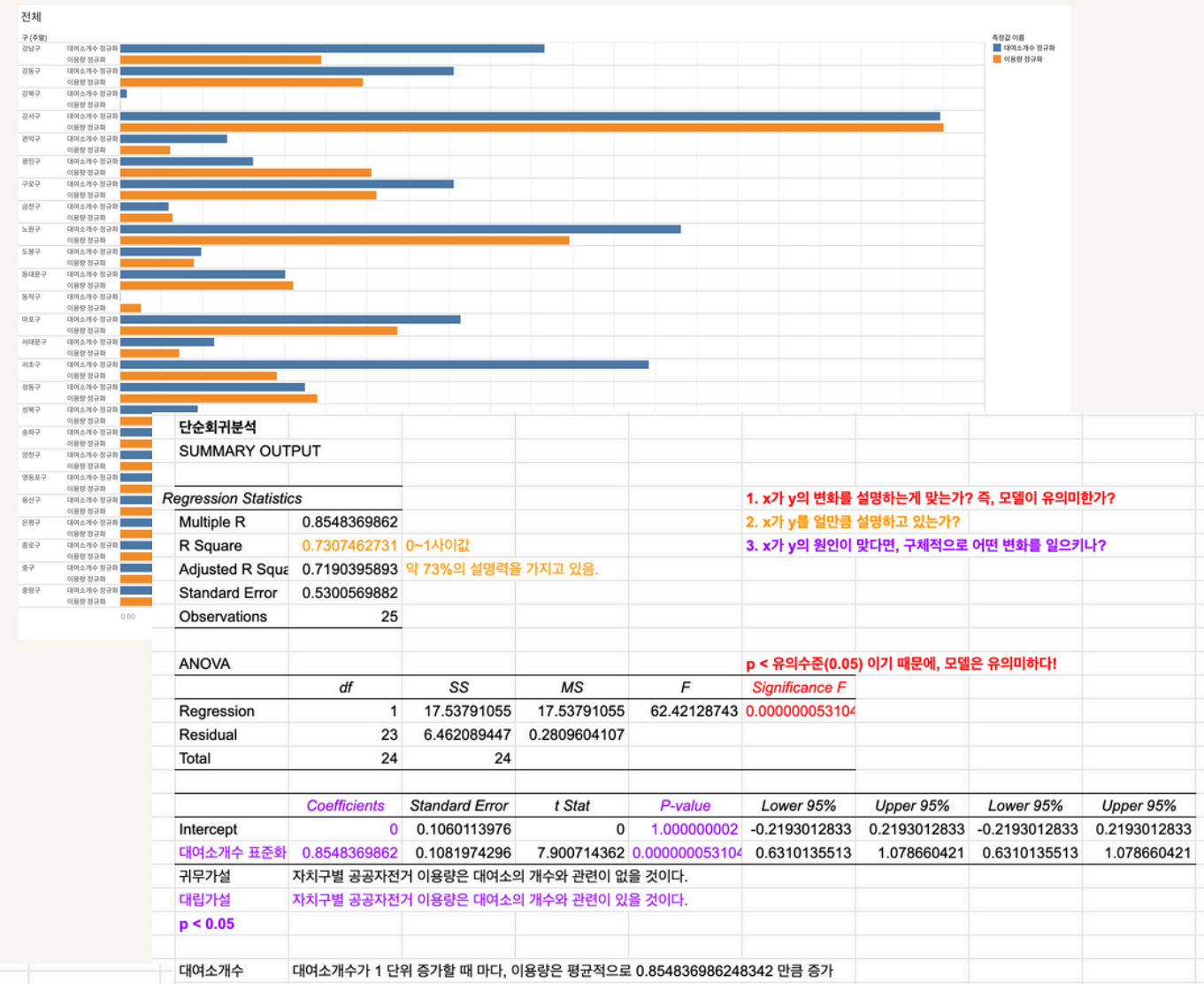
• 주요 결과 & 시각화

- 가설1. 대여소 개수와 따릉이 이용량 관계

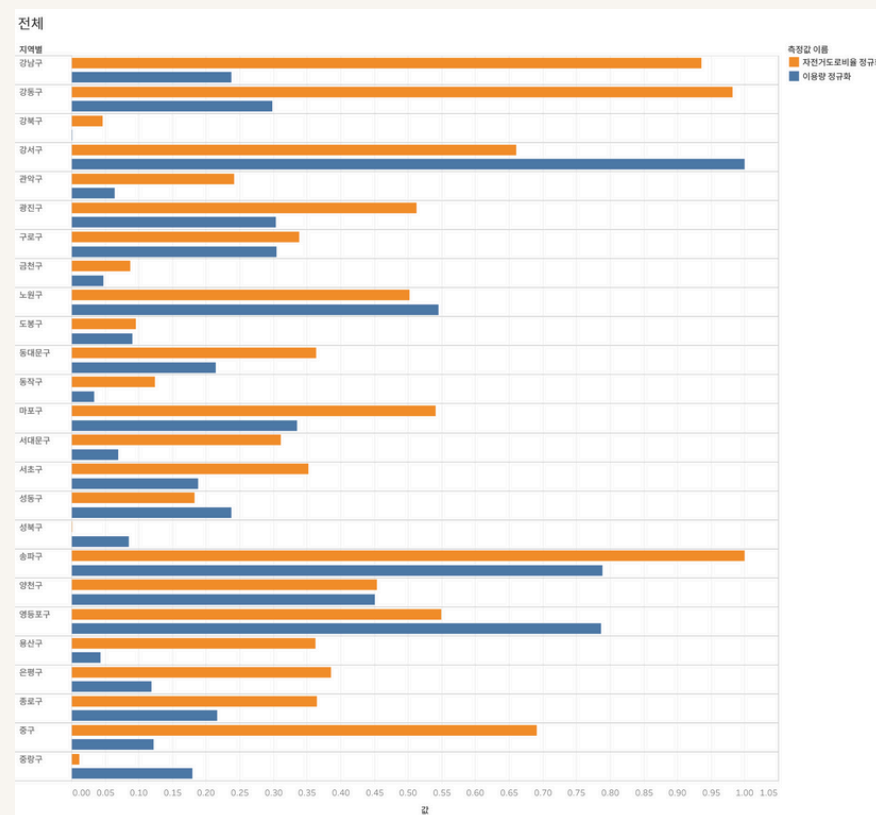
- 평일 73%, 주말 68%로 신뢰성이 높게 나타났다.(P값 유의미함.)
- 이후 가설들을 진행함에 있어 지금까지 설치된 대여소가 필요에 맞게 설치된 것인지 확인하는 절차는 필요했다 판단하여 세운 가설로 유의미함에 이후 가설들을 추가 설정할 수 있었다.

- 가설2. 자전거 도로와 따릉이 이용량의 관계

- 34%의 설명력을 가지고 있음.(P값 유의미함)
- 가설이 어느정도 영향이 있음을 확인하였다.



가설1 시각화 및 분석결과



SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.5832935318								
R Square	0.3402313443	0~1사이값							
Adjusted R Square	0.3115457505	약 34%의 설명력을 가지고 있음.							
Standard Error	0.8297314321								
Observations	25								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	1	8.165552263	8.165552263	11.86070428	0.002209900629				
Residual	23	15.83444774	0.6884542495						
Total	24	24							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%	
Intercept	0	0.1659462864	0	0.9999999981	-0.3432860464	0.3432860464	-0.3432860464	0.3432860464	
이용량	0.5832935318	0.1693682194	3.443937322	0.002209900636	0.2329286779	0.9336583858	0.2329286779	0.9336583858	
귀무가설	자치구별 자전거도로 비율은 공공자전거 이용량과 관련이 없을 것이다.								
대립가설	자치구별 자전거도로 비율은 공공자전거 이용량과 관련이 있을 것이다.								
p < 0.05									

가설2 시각화 및 분석결과

서울시 자치구별 따릉이 이용량 차이

- 주요 결과 & 시각화
 - 가설3-1. 지하철 역의 개수와 따릉이 이용량의 관계
 - 11%의 설명력을 가지고 있으나 모델이 유의미하지 않음.(P = 0.1)
 - 자치구별 따릉이 이용량은 지하철역 개수와 관련이 없을 것이라 판단.
 - 가설3-2. 지하철 역의 하차총승객수와 따릉이 이용량의 관계
 - 5%의 설명력을 가지고 있으나 모델이 유의미하지 않음(P = 0.2)
 - 자치구별 따릉이 이용량은 지하철 하차총승객수와 관련이 없을 것이라 판단.
 - 가설3-3. 지하철 역의 승차총승객수와 따릉이 이용량의 관계
 - 5%의 설명력을 가지고 있으나 모델이 유의미하지 않음(P = 0.2)
 - 자치구별 따릉이 이용량은 지하철 승차총승객수와 관련이 없을 것이라고 판단.

단순회귀분석 SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.3364238652							
R Square	0.1131810171	0~1사이값						
Adjusted R Square	0.07462366998	약 11%의 설명력을 가지고 있음.						
Standard Error	0.9619648278							
Observations	25							
ANOVA					p > 유의수준(0.05) 이기 때문에, 모델은 유의미하지 않다!			
	df	SS	MS	F	Significance F			
Regression	1	2.71634441	2.71634441	2.935394305	0.1001068114			
Residual	23	21.28365559	0.92537633					
Total	24	24						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	0	0.1923929656	0	1.000000002	-0.3979951702	0.3979951702	-0.3979951702	0.3979951702
역개수 표준화	0.3364238652	0.1963602482	1.713299246	0.1001068203	-0.06977825438	0.7426259848	-0.06977825438	0.7426259848
귀무가설	자치구별 공공자전거 이용량은 지하철 이용량과 관련이 없을 것이다.							
대립가설	자치구별 공공자전거 이용량은 지하철 이용량과 관련이 있을 것이다.							
p > 0.05								

가설3-1 분석결과

SUMMARY OUTPUT								
Regression Statistics					1. x가 y의 변화를 설명하는게 맞는가? 즉, 모델이 유의미한가?			
Multiple R	0.2379005089				2. x가 y를 얼마나 설명하고 있는가?			
R Square	0.05659665214	0~1사이값			3. x가 y의 원인이 맞다면, 구체적으로 어떤 변화를 일으키나?			
Adjusted R Square	0.01557911527	약 5%의 설명력을 가지고 있음.						
Standard Error	0.9921798651							
Observations	25							
ANOVA					p > 유의수준(0.05) 이기 때문에, 모델은 무의미하다!			
	df	SS	MS	F	Significance F			
Regression	1	1.358319651	1.358319651	1.379815963	0.2521512499			
Residual	23	22.64168035	0.9844208847					
Total	24	24						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	0	0.198435973	0	1.000000002	-0.4104960834	0.4104960834	-0.4104960834	0.4104960834
승차총승객수 표준	0.2379005089	0.2025278669	1.174655679	0.2521512444	-0.1810603019	0.6568613197	-0.1810603019	0.6568613197
귀무가설	자치구별 공공자전거 이용량은 지하철 승차총승객수와 관련이 없을 것이다.							
대립가설	자치구별 공공자전거 이용량은 지하철 승차총승객수와 관련이 있을 것이다.							
p > 0.05								

가설3-3 분석결과

SUMMARY OUTPUT								
Regression Statistics					1. x가 y의 변화를 설명하는게 맞는가? 즉, 모델이 유의미한가?			
Multiple R	0.2333414665				2. x가 y를 얼마나 설명하고 있는가?			
R Square	0.05444823997	0~1사이값			3. x가 y의 원인이 맞다면, 구체적으로 어떤 변화를 일으키나?			
Adjusted R Square	0.01333729389	약 5%의 설명력을 가지고 있음.						
Standard Error	0.9933089681							
Observations	25							
ANOVA					p > 유의수준(0.05) 이기 때문에, 모델은 무의미하다!			
	df	SS	MS	F	Significance F			
Regression	1	1.306757759	1.306757759	1.324421964	0.2616244509			
Residual	23	22.69324224	0.9866627061					
Total	24	24						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	0	0.1986617936	0	1.000000002	-0.4109632289	0.4109632289	-0.4109632289	0.4109632289
하차총승객수 표준	0.2333414665	0.2027583441	1.150835333	0.2616244462	-0.1860961227	0.6527790556	-0.1860961227	0.6527790556
귀무가설	자치구별 공공자전거 이용량은 지하철 이용량과 관련이 없을 것이다.							
대립가설	자치구별 공공자전거 이용량은 지하철 이용량과 관련이 있을 것이다.							
p > 0.05								

가설3-2 분석결과

서울시 자치구별 따릉이 이용량 차이

• 주요 결과 & 시각화

- 가설4. 사업체 수, 종사자 수와 따릉이 이용량의 관계

- 신뢰성이 높게 나타남

- 전체: 약 83.7%

- **평일: 약 83%**

- 주말: 약 84%

- 사업체 수와 따릉이 이용량 간의 상관관계는 무의미함.($P > 0.05$)

- 종사자 수와 이용량 간의 상관관계는 유의미함을 확인($P < 0.05$)

- 가설이 어느정도 영향이 있음을 확인하였음.

	SUMMARY OUTPUT								
	Regression Statistics								
	Multiple R	0.9151653162							
	R Square	0.837527556	83.7%의 설명력!						
	Adjusted R Square	0.8143172069							
	Standard Error	0.4309092632							
	Observations	25							
	ANOVA								
		df	SS	MS	F	Significance F			
	Regression	3	20.10066134	6.700220448	36.08422911	0.000000018046			
	Residual	21	3.899338655	0.1856827931					
	Total	24	24						
		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
	Intercept	0	0.08618185264	0	0.9999999995	-0.1792249733	0.1792249733	-0.1792249733	0.1792249733
기각	사업체수	0.5606744967	0.4315942268	1.299077842	0.2080000784	-0.3368748292	1.458223823	-0.3368748292	1.458223823
채택	종사자수	-0.9010948404	0.3803987878	-2.368816278	0.0275115068	-1.692177423	-0.1100122577	-1.692177423	-0.1100122577
채택	대여소개수	1.030275268	0.1454962442	7.081112457	0.000000549526	0.7276992648	1.33285127	0.7276992648	1.33285127
	p < 0.05	대립가설	자차구별 공공자전거 이용량은 대여소개수와 관련이 있을 것이다 자차구별 공공자전거 이용량은 종사자수와 관련이 있을 것이다						
	RESIDUAL OUTPUT PROBABILITY C 자차구별 공공자전거 이용량은 사업체수와 관련이 없을 것이다								

가설4 분석결과

- **결과 활용 방안**

- 분석 결과를 통해 영향을 주는 가설들의 데이터들을 이용해 예측모델링을 구현하여 자전거의 공급에 대한 전략수립을 통해 비용절감 및 이용자 만족도 증가를 기대할 수 있다.

- 코드 링크

- [HTTPS://GITHUB.COM/JIWON-](https://github.com/jiwon-)

JUNG323/GIT_REPO/BLOB/MAIN/TEAMPROJECT1/TEAMPROJECT_1.IPYNB

공유 오피스 이벤트 이용자의 유료 전환 가능성 예측 및 전략 제안

2025.06.09 ~ 2025.06.19
협업 프로젝트(3인)

공유 오피스 이벤트 이용자의 유료 전환 가능성 예측 및 전략 제안

기간: 2025-06-09 ~ 2025-06-18

분석 툴: Python(pandas, numpy, matplotlib, sklearn)

팀 내 역할: 팀장, 아이디어 제안, 데이터 시각화(방문지점), 예측모델링 설계

- 개요

- 공유오피스의 3일 무료 체험 이벤트에 관한 데이터들을 이용해 유료전환 예측 모델을 만드는 것을 목표로 진행한 프로젝트로 주어진 데이터를 이용해 유저들의 행동 패턴을 모델링에 알맞은 형태로 정리하는 것을 중점으로 진행한 프로젝트이다.

- 목표

- 공유 오피스의 3일 무료 체험 이벤트 이용자 대상으로 유저들의 행동 패턴들을 분석해 유효 데이터들을 이용하여 유료전환 예측모델링 설계

- 데이터

테이블 명	테이블 설명	주요 컬럼
trial_register	3일 체험 신청	유저 id, 3일체험신청일지
trial_visit_info	3일 체험 신청자 일자별 방문 기록	유저 id, 날짜, 지점id, 최초입실시각, 최종퇴실시각, 체류 시간
trial_access_log	3일 체험 신청자 출입 기록	유저id, 지점id, timestamp, 출입 방 향
trial_payment	3일 체험 신청자 결제 여부	유저 id, 결제 여부
site_area	지점별 면적	지점id, 지점 면적

공유 오피스 이벤트 이용자의 유료 전환 가능성 예측 및 전략 제안

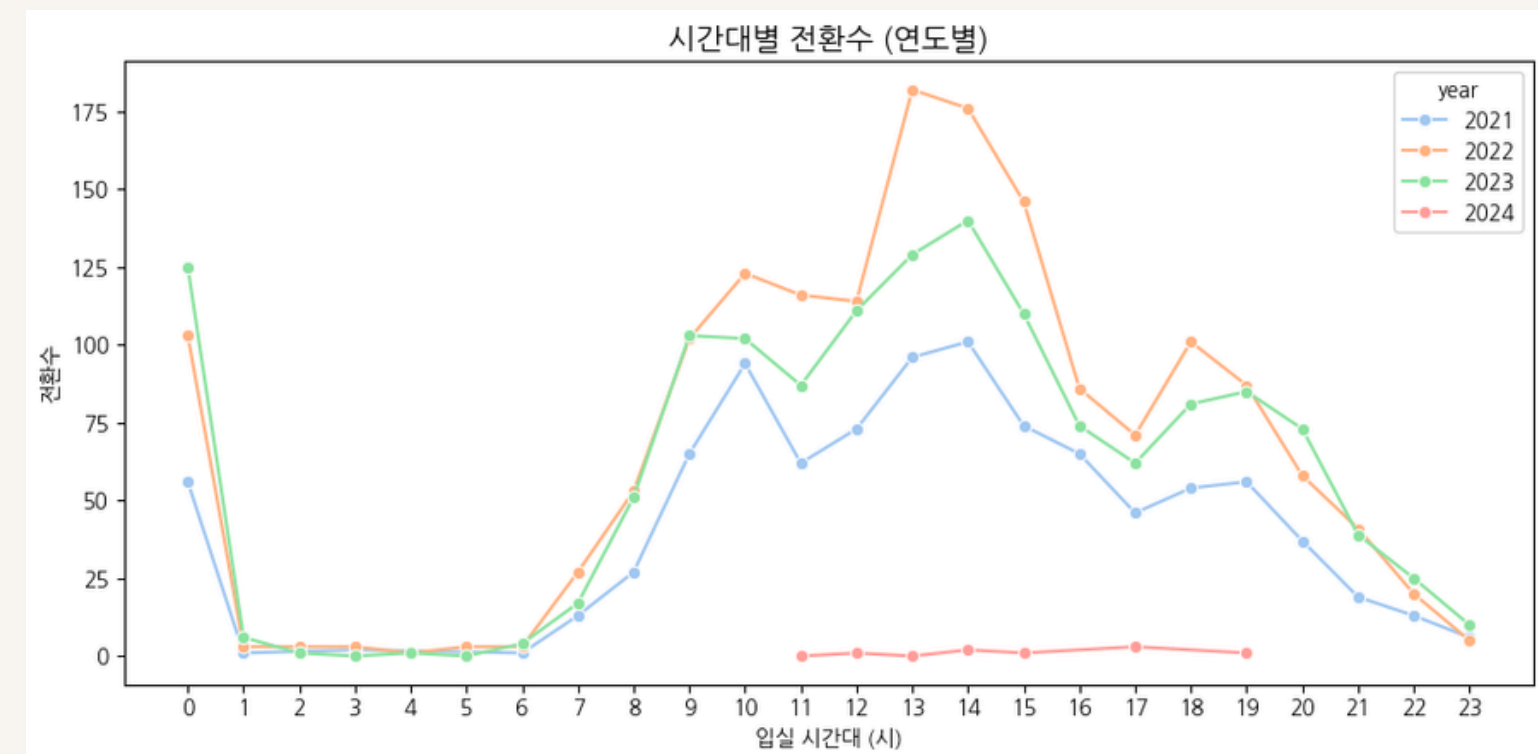
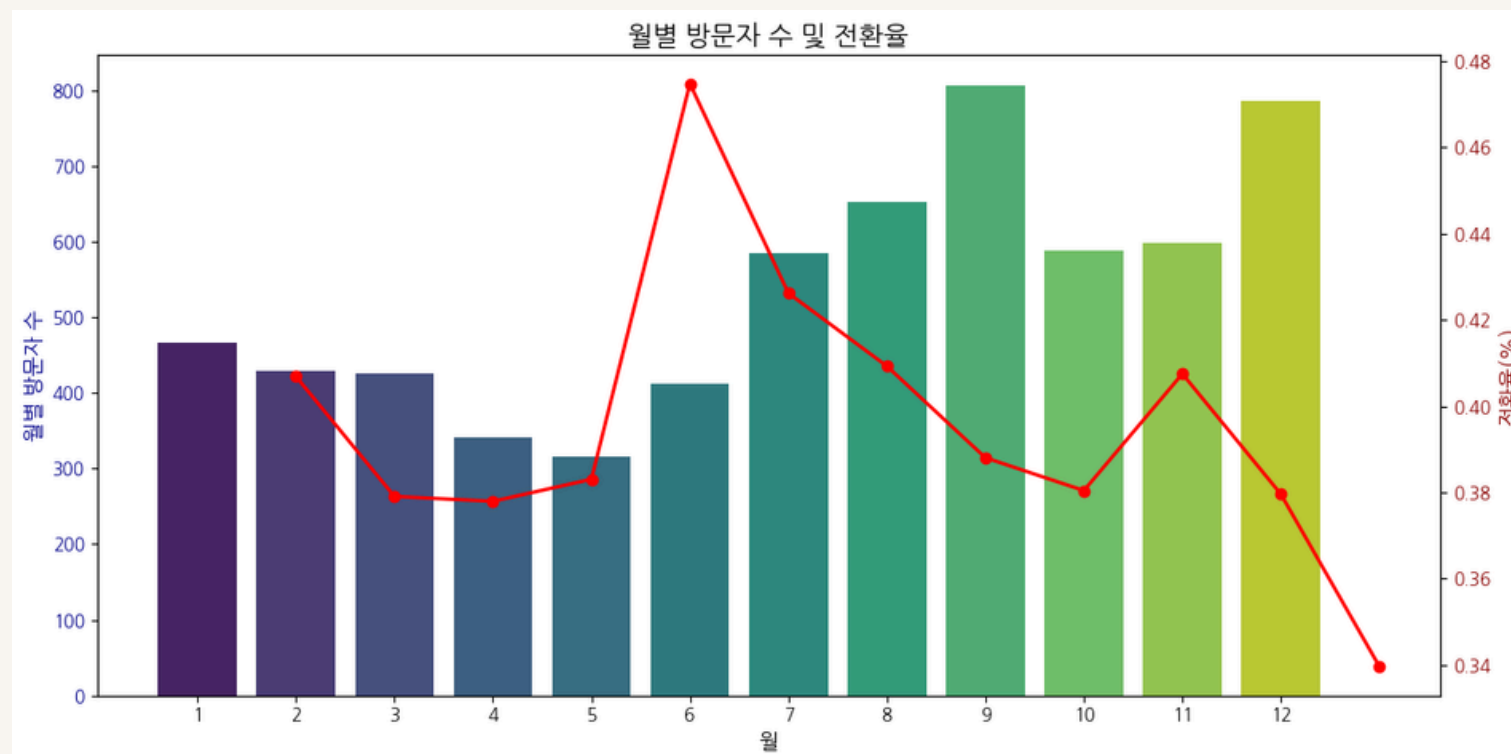
- 분석 과정

- 데이터 전처리

- 결측치 제거
- 데이터 형변환(DATETIME)
- 파생변수 생성(방문시간, 방문 달, 방문 횟수 등)
- TRIAL_ACCESS_LOG
 - 불필요한 로그(00시00분 일괄처리유저, 1분 이내 중복, 입장 없는 퇴장로그)삭제
 - 동일 유저의 연속 출입기록을 독립적인 세션으로 묶어 구성
 - 실제 프로그램을 신청하고 체험을 진행한 유저에 대해 집중적으로 분석(6026명)

- EDA

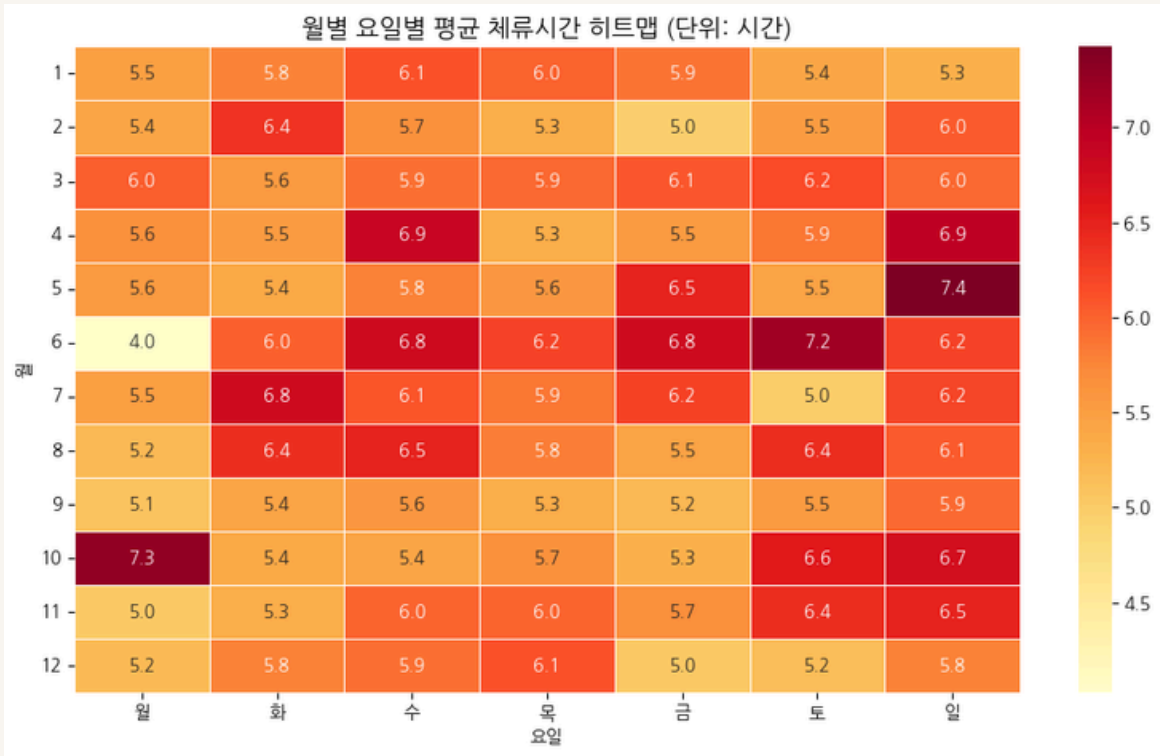
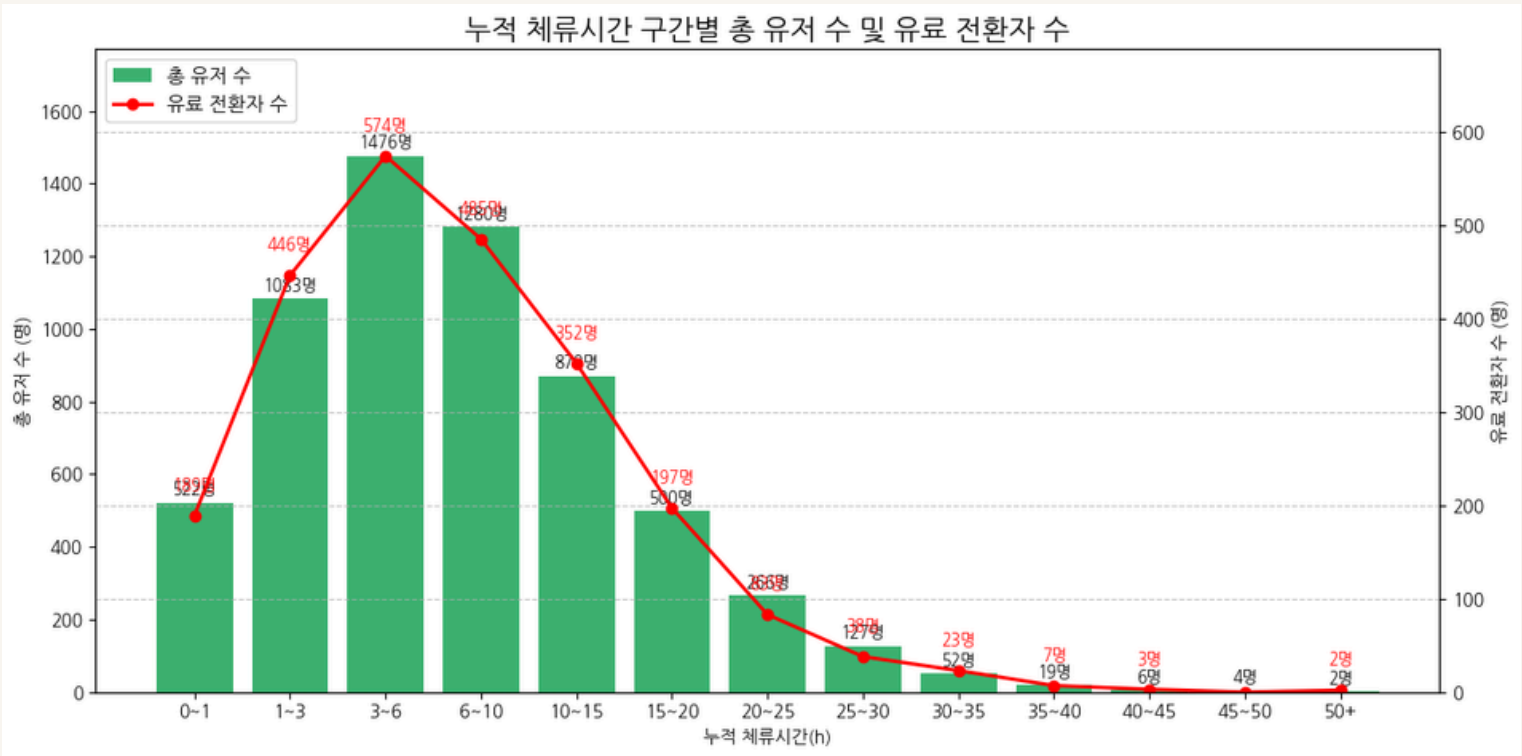
- 방문시기(월, 일)



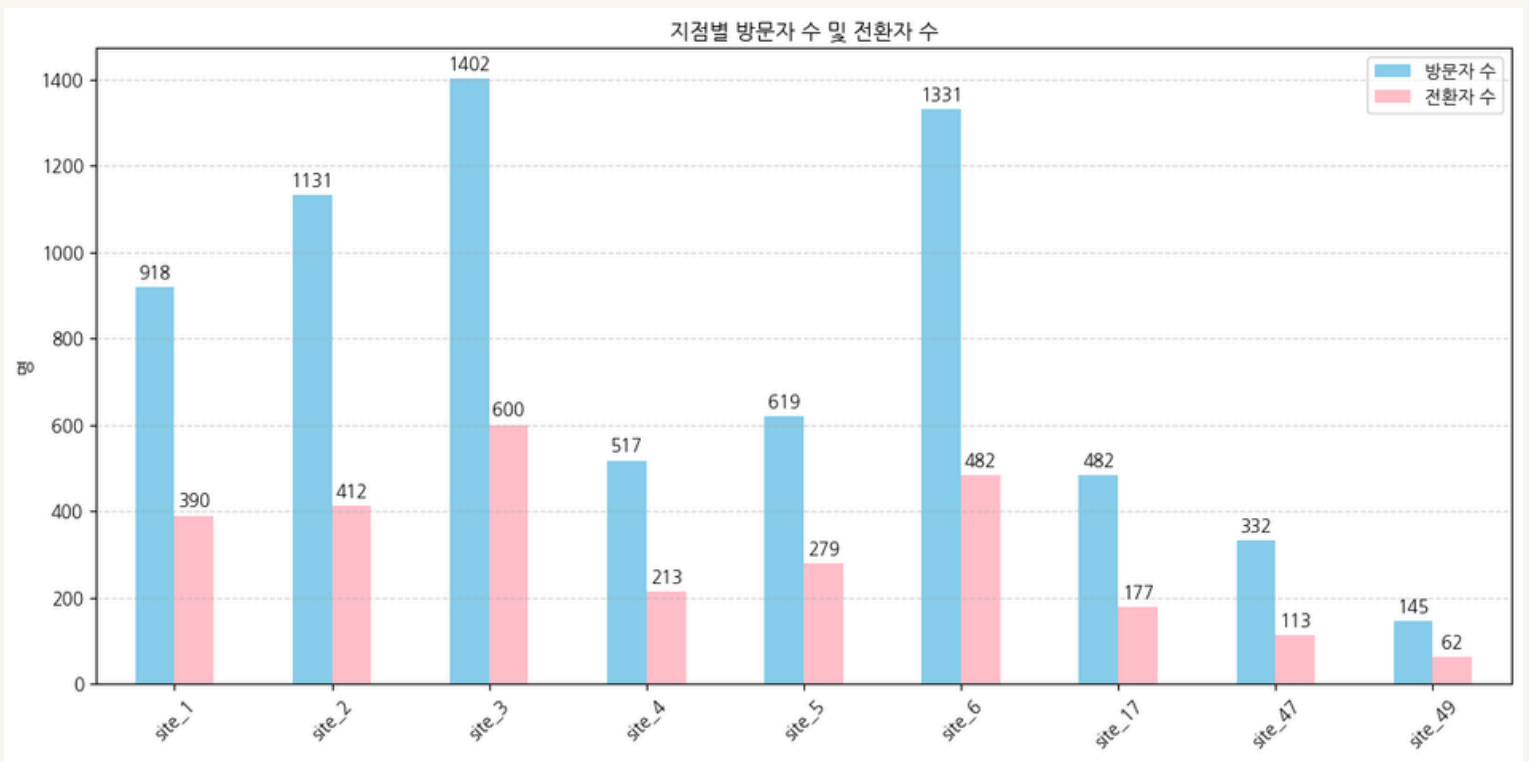
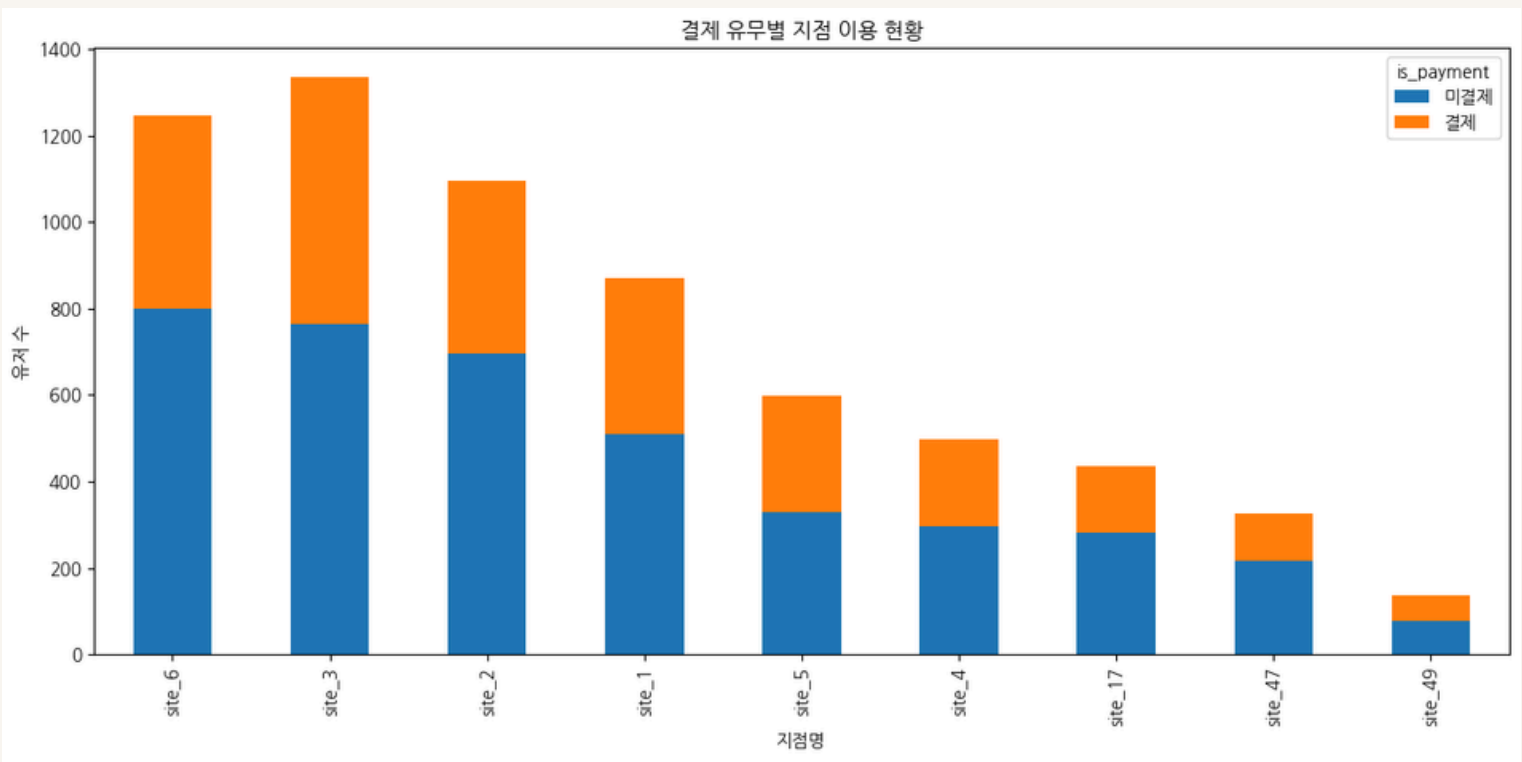
공유 오피스 이벤트 이용자의 유료 전환 가능성 예측 및 전략 제안

- EDA

- 체류시간

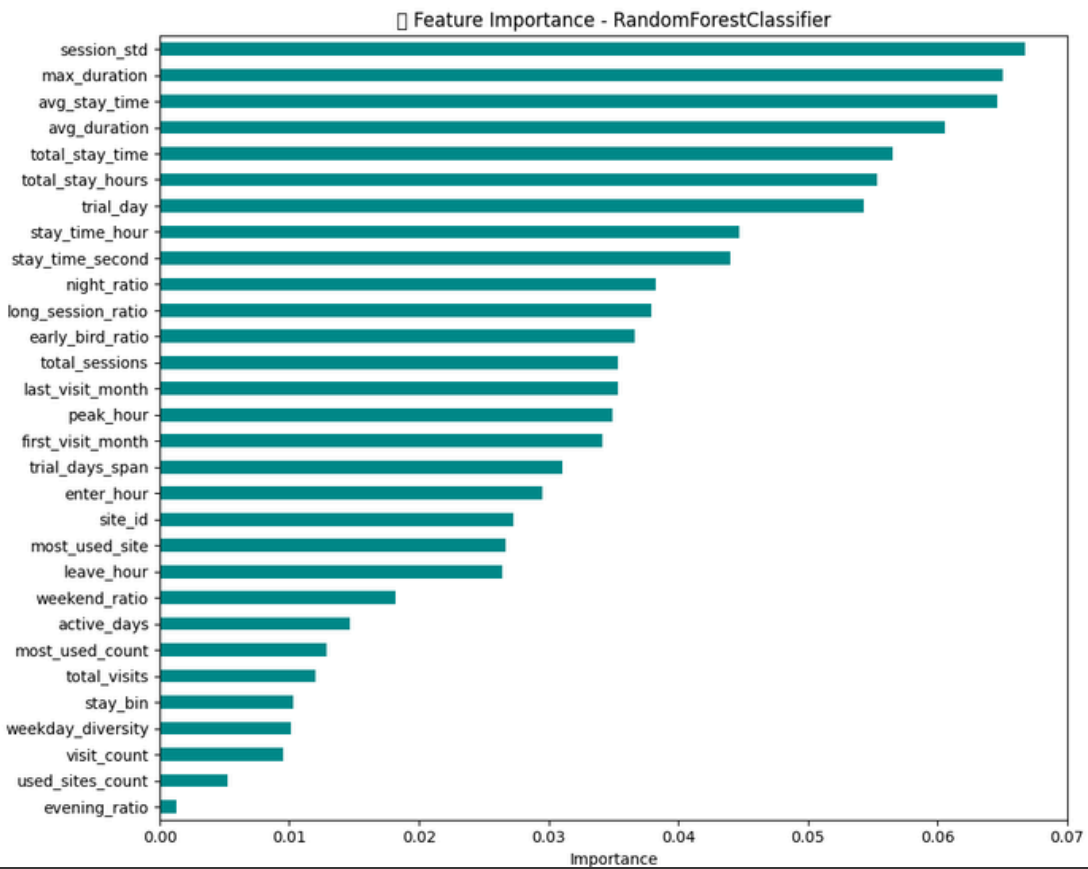


- 방문 지점



공유 오피스 이벤트 이용자의 유료 전환 가능성 예측 및 전략 제안

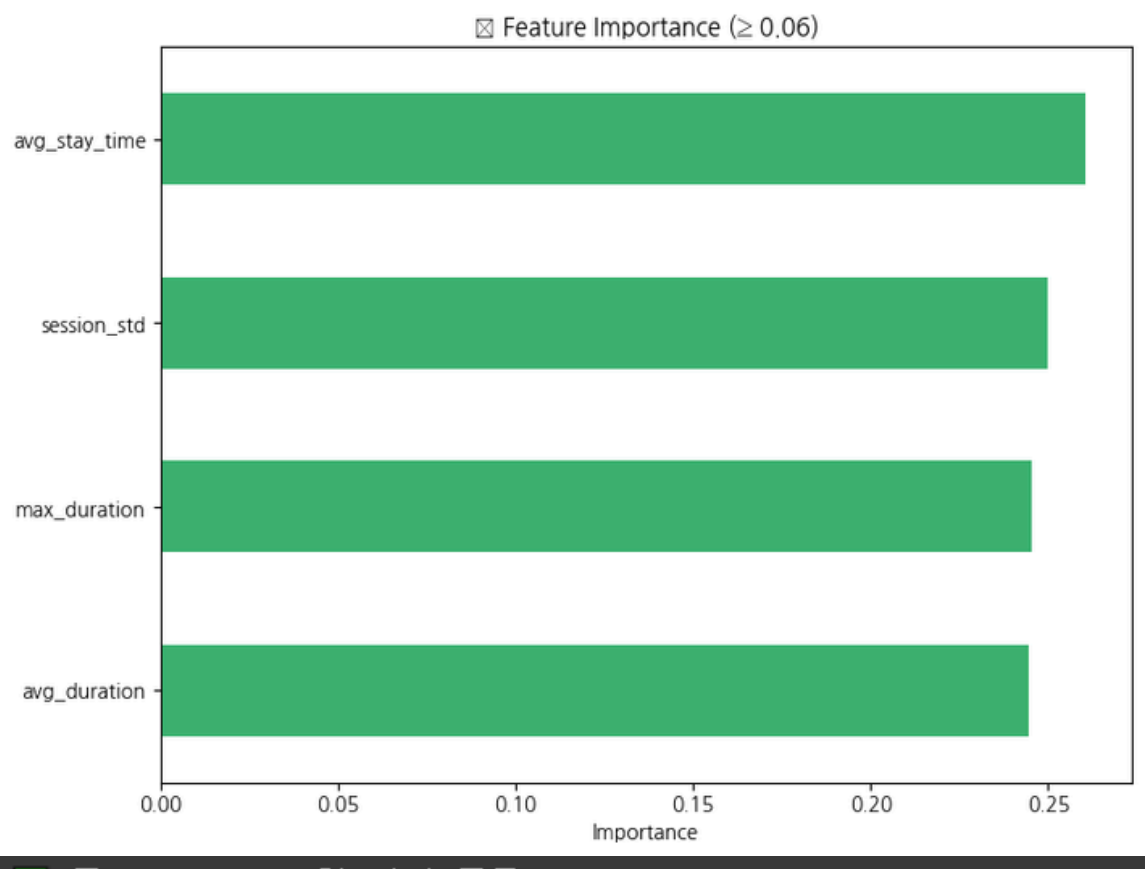
- 주요 결과(모델링)
 - 중요도 범위의 차이를 두어 비교결과 전체와 0.06이 성능이 크게 차이 나지 않아 효율성면에서 좋은 0.06모델을 최종모델로 선정



정확도: 0.8588531821045998

분류 리포트:

	precision	recall	f1-score	support
0	0.83	0.96	0.89	914
1	0.92	0.73	0.81	673
accuracy			0.86	1587
macro avg	0.88	0.84	0.85	1587
weighted avg	0.87	0.86	0.86	1587



중요도 ≥ 0.06인 피쳐 목록: ['avg_duration', 'max_duration', 'session_std', 'avg_stay_time']

정확도: 0.8771266540642723

분류 리포트:

	precision	recall	f1-score	support
0	0.87	0.92	0.90	914
1	0.89	0.81	0.85	673
accuracy			0.88	1587
macro avg	0.88	0.87	0.87	1587
weighted avg	0.88	0.88	0.88	1587

- 결과 활용 방안
 - 사용자 이탈 방지 전략
 - 체류 시간이 길고 규칙적인 유저 식별 → 맞춤형 혜택, VIP 프로그램을 제공
- 공간&서비스 최적화
 - 환경 개선 및 편의 시설 확충
 - 서비스 만족도 상승
 - 사용자 체류시간 증가
- 사용자 세그먼트별 서비스 상품 개발
 - 체류 패턴에 따른 유저 세분화
 - 맞춤형 요금제나 서비스 패키지를 개발
 - 유료 전환율을 극대화

• 코드 링크
https://github.com/jiwon-jung323/git_repo/blob/main/teamproject2/teamproject_2.ipynb

Thank you.

010-9253-6036

jwjung323@naver.com

[https://github.com/jiwon-jung323/git_repo?](https://github.com/jiwon-jung323/git_repo?tab=readme-ov-file)
[tab=readme-ov-file](#)