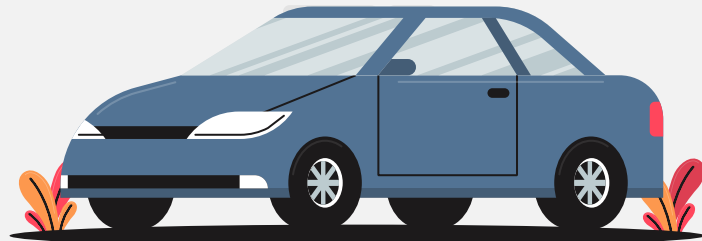


2024학년도 2학기 시계열자료분석팀 주제분석

P-SAT X 투루카 기업연계 프로젝트

김나현 강철석 이승아 김재원 이신영



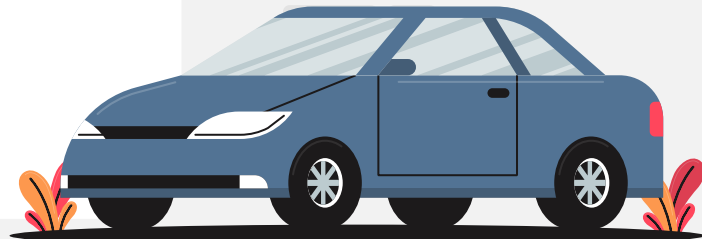
목차

01 프로젝트 소개

02 과제 1
DR-GPS 효과 테스트

03 과제 2
대리기사 탐지 모델링

04 과제 3
스팸 수요 요인 분석



01

프로젝트 소개

분석 도메인 선정

공유차량 업계에 대한 관심



현업에서 사용되는 데이터를
다루는 경험에 대한 니즈

공유차량 기업과의 연계 프로젝트 제안

01 프로젝트 소개

P-SAT 24-2학기
시계열자료분석팀

기업 컨택 과정

한국 공유차량 업계 내 4개 기업에 컨택 시도

The logo for SOCAR, featuring the word "SOCAR" in a bold, blue, sans-serif font.The logo for G-CAR, featuring a stylized "G" with a red and black design, followed by the word "CAR" in a bold, black, sans-serif font.The logo for Turu CAR, featuring the word "Turu" in a bold, black, sans-serif font, followed by the word "CAR" in a smaller, black, sans-serif font.The logo for towncar, featuring a green location pin icon followed by the word "towncar" in a green, sans-serif font.

01 프로젝트 소개

P-SAT 24-2학기
시계열자료분석팀

기업 컨택 과정

투루카와의 컨택에 성공하여 프로젝트 추진

SOCAR

G CAR

TURU CAR

towncar



01 프로젝트 소개

P-SAT 24-2학기
시계열자료분석팀

기업 컨택 과정

투루카와의 컨택에 성공하여 프로젝트 추진

컨택 시 작성한 프로젝트 제안서

TURU CAR X 상권대학교 통계분석학회 P-SAT 기업연계 프로젝트 제안서

사전 미팅 분석흐름 제안 PT

05. 데이터 및 산학 프로세스 관련 제안

데이터 범위

투루카 주차장별 데이터

1. 사용자 및 수도권의
리턴 브라운 및 카세이먼트 리스토

항목	내용
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목

2. 계명시(투루카)가 보유한
가용 주차장 리스토

고객 로그 데이터

1. 고객 데이터 로그 데이터

항목	내용
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목
항목	항목

TURU CAR

[투루카X피셋] 프로젝트 분석 흐름 제안서

상권대학교 통계분석학회 P-SAT 시계열자료분석팀
금년 상반기에 이루어질 예정인 P-SAT

과제(DR-GPS 효과 테스트)

1. 분석 목적

2. 분석 방법

- 프로젝트 및 연구실 분석목표에 의해 테스트 설계 과정에서 요구사항 반영
- 투루카의 실시간 DR-GPS의 유용성 검증에 의해 테스트 설계한 분석의 설계가 필요함

3. 분석 방법

- 프로젝트의 분석을 할 수 있는 상황에서 투루카로부터 제공된 데이터를 이용하여 두 집단의 평균이 차이가 있는지를 검증하는 방법

4. 분석 방법

- 프로젝트의 분석을 할 수 있는 상황에서 투루카로부터 제공된 데이터를 이용하여 두 집단의 평균이 차이가 있는지를 검증하는 방법

5. 분석 방법

- 프로젝트의 분석을 할 수 있는 상황에서 투루카로부터 제공된 데이터를 이용하여 두 집단의 평균이 차이가 있는지를 검증하는 방법

6. 분석 방법

- 프로젝트의 분석을 할 수 있는 상황에서 투루카로부터 제공된 데이터를 이용하여 두 집단의 평균이 차이가 있는지를 검증하는 방법

7. 분석 방법

- 프로젝트의 분석을 할 수 있는 상황에서 투루카로부터 제공된 데이터를 이용하여 두 집단의 평균이 차이가 있는지를 검증하는 방법

8. 분석 방법

- 프로젝트의 분석을 할 수 있는 상황에서 투루카로부터 제공된 데이터를 이용하여 두 집단의 평균이 차이가 있는지를 검증하는 방법

구체적인 분석 흐름 제안서

01 프로젝트 소개

P-SAT 24-2학기
시계열자료분석팀

투루카 서비스 개요

TURU CAR

일정 시간 경과 후 대여한 스팟에
다시 반납해야 하는 **왕복** 형태의 서비스로,
일반 렌터카의 개념과 동일

카셰어링

왕복으로 미리
예약할 땐



신규지역 오픈

리턴프리

편도로 바로
이용할 땐



배달렌트

내맘대로 불러
출발할 땐



커뮤니티

아파트, 회사
전용 카셰어링

가입하지 않음 >



대여 스팟과 반납 스팟이 불일치하는
편도 형태의 서비스

투루카가 제공하는 서비스 목록

01 프로젝트 소개

P-SAT 24-2학기
시계열자료분석팀

프로젝트 개요 | 과제 1

TURU CAR

제안 내용을 바탕으로 현재 회사 내에서 인지하고 있는
문제 혹은 개선의 여지가 있는 분석과제 제안

DR-GPS 효과 테스트

분석 기간
주제분석 1~2주차

담당 부서
서비스운영팀

분석 목적

차량에 설치된 DR-GPS
기술의 실제 효과 입증

분석 내용

통계적 가설 검정 및
통계모델 기반 분석

분석 산출물

DR-GPS 설치 여부(X)에 따른
차량의 Spot-Out 여부(Y) 확인
과정 및 결과를 담은 보고서

프로젝트 개요 | 과제 2

TURU CAR

제안 내용을 바탕으로 현재 회사 내에서 인지하고 있는
문제 혹은 개선의 여지가 있는 분석과제 제안

대리기사 탐지 모델링

분석 기간	분석 목적	분석 내용	분석 산출물
주제분석 1~2주차	투루카의 주 이용고객인 대리기사 고객 탐지를 통한 맞춤 전략 수립	대리기사 고객만의 특징을 반영할 수 있는 파생변수 생성 및 모델 구축	대리기사 탐지를 위한 머신러닝 모델

프로젝트 개요 | 과제 3

TURU CAR

제안 내용을 바탕으로 현재 회사 내에서 인지하고 있는
문제 혹은 개선의 여지가 있는 분석과제 제안

효율적인 SPOT 선정 프로세스 구축

분석 기간	분석 목적	분석 내용	분석 산출물
주제분석 3주차~2월 中	고객의 수요 요인 분석을 통한 매출 극대화 및 스팟 선정의 효율화	현재 스팟의 포괄적 특징 추출을 통해 얻은 인사이트에 기반한 스팟 판단 기준 제시	<ul style="list-style-type: none">스팟 선정 프로세스/기준수요가 높을 것으로 예상되는 미진출 지역에 대한 실제 적용 결과

02

과제 1

DR-GPS 효과 테스트

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

주제 선정 배경

투루카의 'DR-GPS' 기술 도입



비대면 기반의 카셰어링 서비스는 GPS 수신에 어려운
지하 주차장 등에서 차량의 위치 추적에 한계가 존재



차량의 회전 방향을 파악할 수 있는 센서 탑재 기술인
DR-GPS 도입

주제 선정 배경

사측 요청 내용

“DR-GPS의 도입으로 Spot-out* 관련 VOC* 건수가 줄었는지”를
통계적으로 검정

Spot-out*: GPS 신호가 끊겨 잘못된 장소에 차량이 반납된 상황

VOC* : Voice of Customer, 해당 프로젝트에서는 스팟 아웃 관련 고객 민원을 의미

⋮

즉, DR-GPS 기술이 민원 감소에 유의미한 영향을 주었는지를 확인하고자 함

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

데이터 구조 파악

WK	차량번호	DR 구분	주간 차량 이용건수	VOC 건수	VOC 건수/주간 차량 이용건수	반납스팟
1	aa1	설치	7	1	1/7	b12
2	aa2	미설치	5	2	2/5	b18
5	aa2	미설치	6	1	1/6	b33
2	aa3	설치	12	4	4/12	b21
...

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

데이터 전처리

동일 주간, 동일 차량번호 처리

WK	차량번호	DR 구분	주간 차량 이용건수	VOC 건수	반납스팟
4	aa4	설치	8	1	b55
4	aa4	설치	8	2	b58



WK	차량번호	DR 구분	주간 차량 이용건수	VOC 건수
4	aa4	설치	8	1

반납 스팟에 대한 고려 X → WK와 차량번호가 동일한 경우 VOC 건수 SUM

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

데이터 전처리

동일한 차량번호의 WK 처리

차량번호	DR 구분	주간 차량 이용 건수	VOC 건수
aa9	설치	2	1
aa9	설치	3	2
aa9	설치	4	1



차량번호	DR 구분	주간 차량 이용 건수	VOC 건수
aa9	설치	3	2

WK의 영향을 없애야 함 → 동일한 차량번호 중 하나의 행을 랜덤으로 추출



데이터 전처리 Week의 차이를 없애는 전처리가 가능한 근거

동일한 차량번호의 WK 처리



차량번호

aa9

DR 구분

설치

주간 차량 이용 건수

2

VOC 건수

1

제공받은 데이터의 경우 공유차량 시장에서의 성수기 시기이므로,
해당 기간 내에서의 이용 내역의 차이가 크게 존재하지 않는다는

사측의 의견 반영

설치

4

1



차량 번호 이외의 모든 변수들이 통제된 데이터셋이 구축되어야 함

→ 평균, 합 등은 적용 불가

aa9

설치

주간 차량 이용 건수

3

VOC 건수

2

: 현재의 데이터셋으로 유의성 검정을 진행할 수 있는 최선의 전처리라고 판단

WK의 영향을 없애야 함 → 동일한 시장년도 중 하나의 일일 데이터를 추출

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

최종 데이터셋

모두 중복되지 않는
다른 차량으로 구성

차량번호	DR 구분	주간 차량 이용 건수	VOC 건수
aa3	설치	2	1
aa9	설치	10	1
aa17	설치	7	3
...

독립적인 차량에 대해 DR-GPS와 일반 GPS 간의 차이를 검정할 수 있는
test를 위한 데이터셋 완성!

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

분석방법론 요약

Proportion Test #1
: Normal
Approximation

DR-
GPS

VS

일반
GPS

$$\frac{\text{그룹별 총 VOC 건수}}{\text{그룹별 전체 이용 건수}} = \text{민원확률}$$

Proportion Test #2
: Wilson Score
Confidence Interval

DR-
GPS

VS

일반
GPS

민원확률의 신뢰구간
비교를 통한 두 집단 비교

Chi-square test
: Independence
Test

	VOC 발생 O	VOC 발생 X
DR- GPS		
일반 GPS		

자세한 내용은 P-SAT 네이버 카페를 참고해주세요!

데이터 재요청

- 한 차량이 여러 번 사용되는 것은 당연하기 때문에, 전처리 과정에서 **Week** 변수를 반영해주어야 함
- 기존에 제공받은 데이터셋의 경우 VOC 건수가 1 이상인, 즉 VOC가 발생한 경우의 데이터만 존재
즉, 전체 차량 중 VOC가 발생하지 않은 건에 대한 정보는 존재하지 않음



해당 데이터셋으로는 완결성 있는 분석이 진행 불가하므로 데이터 재요청



02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

데이터 재요청



추가 요청 내역

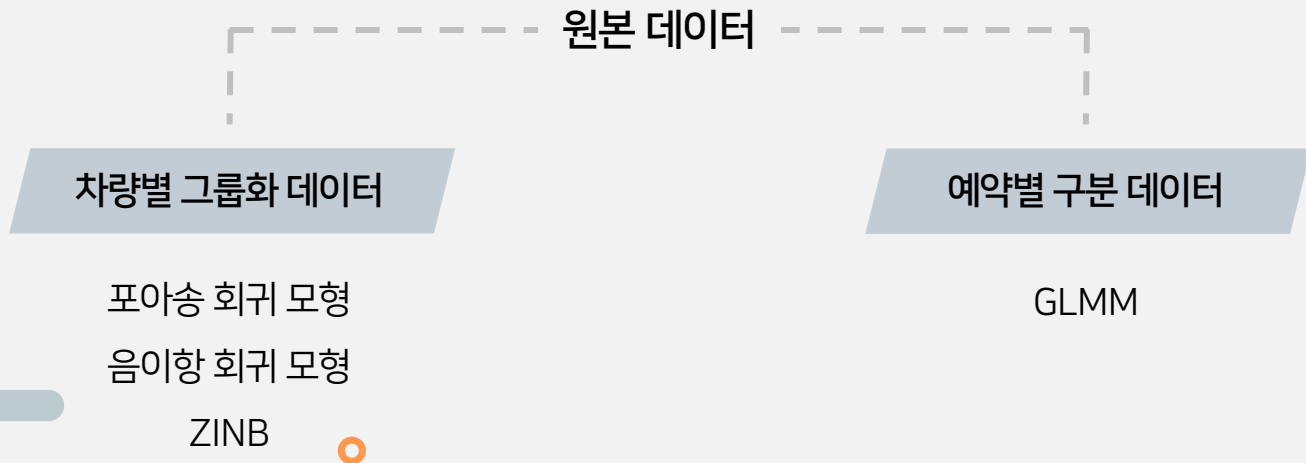
- ✓ 민원 수집 과정에서의 오류를 최대한 배제하고자
VOC 건수가 아닌 **실제 Spot-out 건수** 데이터
- ✓ 제공받은 **Week 내 모든 차량**의 이용 건수 및
Spot out 건수 데이터

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

원본 데이터 전처리

추후 사용할 모델에 맞춰 두 가지 데이터 셋으로 전처리 진행



02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

원본 데이터 전처리 | ① 차량별 그룹화

리턴프리 차량별 그룹화 결과

Car_ID	Total_count	Spot_out	DR-GPS
1_840	56	0	1
1_736	113	2	0
1_840	56	0	1
...

왕복 차량별 그룹화 결과

Car_ID	Total_count	Spot_out	DR-GPS
2_736	240	9	1
2_606	37	2	0
2_111	40	1	0
...

Car_ID : 차량 ID

Total_Count : 차량별 총 운행 건수

Spot_out : 차량별 총 스팟 아웃 발생 건수

DR_GPS : DR GPS 설치 여부 (Factor)

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

원본 데이터 전처리 | ② 예약별 구분

리턴프리 예약별 그룹화 결과 (왕복도 동일)

Car_ID	Start_spot_ID	End_spot_ID	Start_date	End_date	DR-GPS	Spot_out
1_840	149	220	2024-05-18 12:10:12.000	2024-05-18 15:16:15.090	1	0
1_736	482	138	2024-06-04 15:36:16.000	2024-06-04 18:30:19.620	0	1
...

Car_ID : 차량 ID

Start(End)_spot_ID : 출발(도착) 스팟 번호

Start(End)_date : 운행 시작(종료) 시점

DR_GPS : DR GPS 설치 여부 (설치=1)

Spot_out : 스팟 아웃 발생 여부 (발생=1)

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

분석방법론 요약

포아송 회귀 #1

Poisson Regression

스팟아웃 ~ Total Count + DR GPS 여부

음이항 회귀 #2

Negative Binomial Regression

스팟아웃 ~ Total Count + DR GPS 여부

$$Var(\text{스팟아웃}) = \mu_{\text{스팟아웃}} + D\mu_{\text{스팟아웃}}$$

영과잉 음이항 회귀 #3

Zero Inflated Negative Binomial Regression

스팟아웃 ~ Total Count + DR GPS 여부

Checking for Sampling Zero

$$Var(\text{스팟아웃}) = \mu_{\text{스팟아웃}} + D\mu_{\text{스팟아웃}}$$

스팟아웃 건수가 0일 확률

~ Total Count + DR GPS 여부

Checking for Structural Zero

GLMM #4

General Linear Mixed Model

스팟아웃 발생 확률

~ 차량별 DR GPS 설치 여부 + 시간

Fixed Effect

Random Effect

원본 데이터 분석내용 | ① 포아송 회귀모형

포아송 회귀 모형

사건의 발생 횟수를 예측할 때 사용하는 회귀 방법으로,
종속 변수가 0 이상의 정수로 이루어진 카운트 데이터일 때 적합한 모델

리턴프리: $E(Y_i) = 1.7316, \text{Var}(Y_i) = 4.28$

왕복: $E(Y_i) = 0.494, \text{Var}(Y_i) = 1.783$



분산이 평균보다 큰 **과산포(overdispersion)** 발생
평균과 분산이 같은 포아송 분포의 성질을 만족하지 못하므로

○ 이를 해결해줄 수 있는 **음이항 회귀모형** 사용

원본 데이터 분석내용 | ② 음이항 회귀모형

음이항 회귀 모형

종속 변수 Y가 과대 분산을 가진 이산형 데이터일때 사용하는 GLM의 한 종류로,
포아송 회귀 모형의 분산 가정을 완화한 보다 유연한 모델



$$\ln E(Y_i | \text{Count}_i, X_i) = \ln \text{Count}_i + \beta_0 + \beta_1 X_i$$

$$\text{where } \text{Var}(Y_i) = \mu_i + D\mu_i^2$$

$$\text{Count}_i = \text{Total_count}, Y_i = \text{Spot_out}, X_i = \text{DR_GPS}$$



차량별 그룹화 데이터 사용

원본 데이터 분석내용 | ② 음이항 회귀모형

적합 결과

리턴프리

$$\widehat{\beta}_0 = -2.55, p\text{-value} < 0.001$$

$$\widehat{\beta}_1 = -1.50, p\text{-value} < 0.001$$

Residual deviance: 835.32

(degree of freedom: 814)

AIC: 2631.2

왕복

$$\widehat{\beta}_0 = -4.24, p\text{-value} < 0.001$$

$$\widehat{\beta}_1 = -0.53, p\text{-value} < 0.001$$

Residual deviance: 2056.6

(degree of freedom: 3166)

AIC: 5733.6

리턴프리와 왕복 모두 DR-GPS 설치 여부가

스팟 아웃 발생에 통계적으로 유의하게 음의 영향을 미치는 것을 확인



원본 데이터 분석내용을 수용한 회귀모형 영과잉(Zero Inflation) 현상 발생

데이터에서 0의 비중이 높게 관측되는 것을 일컫는 현상
음이항 모형에서 추정하는 확률보다 0이 과도하게 관찰되어

$\hat{\beta}_0 = -2.55, p\text{-value} < 0.001$	리턴프리	$\hat{\beta}_1 = -0.53, p\text{-value} < 0.001$
$\hat{\beta}_1 = -1.50, p\text{-value} < 0.001$	왕복	
Residual deviance: 835.32		Residual deviance: 2056.6
(degree of freedom: 314)		(degree of freedom: 3166)
Spot_out 값이 0인 데이터의 비율	30% (252건)	73% (2328건)
(스팟아웃이 발생하지 않을 확률)		

➔ 데이터의 영과잉 현상을 고려하기 위해 **영과잉 음이항 모형** 사용
스팟 아웃 발생에 통계적으로 유의하게 음의 영향을 미치는 것을 확인

원본 데이터 분석내용 | ③ 영과잉 음이항 회귀모형

영과잉 음이항 회귀 모형

종속 변수 Y에 0이 비정상적으로 많이 포함된
과대 분산 이산형 데이터에 대해 사용하는 모델



$$\begin{aligned}\text{logit}(P(y_i = 0 | \text{Count}_i, X_i)) &= \ln \text{Count}_i + \gamma_0 + \gamma_1 X_i \\ \ln E(Y_i | \text{Count}_i, X_i) &= \ln \text{Count}_i + \beta_0 + \beta_1 X_i \quad \text{where } \text{Var}(Y_i) = \mu_i + D\mu_i \\ \text{Count}_i &= \text{Total_count}, Y_i = \text{Spot_out}, X_i = \text{DR-GPS}\end{aligned}$$



차량별 그룹화 데이터 사용

원본 데이터 분석내용 | ③ 영과잉 음이향 회귀모형

적합 결과

리턴프리

$$\widehat{\beta}_0 = -2.55, p - value < 0.001$$

$$\widehat{\beta}_1 = -1.50, p - value < 0.001$$

$$\widehat{\gamma}_0 = -2.11, p - value = 0.03$$

$$\widehat{\gamma}_1 = -10.20, p - value = 0.84$$

AIC: 2633.3

왕복

$$\widehat{\beta}_0 = -4.24, p - value < 0.001$$

$$\widehat{\beta}_1 = -0.53, p - value < 0.001$$

$$\widehat{\gamma}_0 = -2.97, p - value = 0.12$$

$$\widehat{\gamma}_1 = -7.77, p - value = 0.86$$

AIC: 5737.3

리턴프리와 왕복 모두 DR-GPS 설치 여부가

스팟 아웃 발생에 통계적으로 유의하게 음의 영향을 미치는 것을 확인

원본 데이터 분석내용 | ④ GLMM

일반화선형혼합모델(GLMM, Generalized Linea Mixed Model)

GLM에 랜덤 효과 (Random Effects)를 추가하여 보다 복잡한 데이터에 대해서 설명할 수 있도록 확장된 모델로, 고정 효과와 랜덤 효과를 모두 고려하기에 **반복 측정 데이터** 및 계층, 군집 구조가 있는 데이터에 대해 사용하기 적합



$$\text{logit}\left(P(y_i|X_{i1}, b_{Date_i})\right) = \beta_0 + \beta_1 X_{i1} + b_{i0} + b_{Date_i}$$
$$\begin{pmatrix} b_{i0} \\ b_{Date_i} \end{pmatrix} \sim MVN(0, \Sigma)$$

Where i = 차량번호, X_{i1} = 차량별 DR GPS 설치 여부, $b_{Date_i} = 1, 2, \dots, n_i$

원본 데이터 분석내용 | ④ GLMM

적합 결과

리턴프리

$$\begin{aligned}\widehat{\beta}_0 &= 0.08, p\text{-value} < 0.001 \\ \widehat{\beta}_1 &= -0.06, p\text{-value} < 0.001 \\ AIC &: -91693.8 \\ BIC &: -91656.78\end{aligned}$$

왕복

$$\begin{aligned}\widehat{\beta}_0 &= 0.012, p\text{-value} < 0.001 \\ \widehat{\beta}_1 &= -0.005, p\text{-value} < 0.001 \\ AIC &: -360082.82 \\ BIC &: -3640041.7\end{aligned}$$

리턴프리와 왕복 모두 DR-GPS 설치 여부가

스팟 아웃 발생에 통계적으로 유의하게 음의 영향을 미치는 것을 확인

02 과제 1: DR-GPS 효과 테스트

P-SAT 24-2학기
시계열자료분석팀

원본 데이터 분석내용 | ④ GLMM

리턴프리

적합 결과

왕복

리턴프리와 왕복에서 DR-GPS 유의성에 차이가 존재했던 기존 분석과 달리,
새로운 데이터와 분석에서는 두 서비스 모두에서 DR-GPS의 유의성 입증 가능함

$$\hat{\beta}_1 = -0.06, p\text{-value} < 0.001$$

AIC: -91693.8

BIC

$$\hat{\beta}_1 = -0.005, p\text{-value} < 0.001$$

AIC: -360082.82

BIC



BACK TO BASIC...

피셋 여러분 다들 기본 가정에 주의하세요~!

리턴프리와 왕복 모두 DR-GPS 실시 여부가

스팟 아웃 발생에 통계적으로 유의하게 음의 영향을 미치는 것을 확인



03

과제 2

대리기사 탐지 모델링

03 과제 2: 대리기사 탐지 모델링


P-SAT 24-2학기
시계열자료분석팀

주제 선정 배경

경쟁업체 대비 투루카의 장점

- ✓ 거리 기반이 아닌 시간 기반 요금 책정 → 상대적으로 **저렴한 비용**
- ✓ 이용 시작 지점 외 다른 장소에 차량을 유동적으로 반납할 수 있는 **리턴프리존**

교통 체증이 적은 밤 시간대에
빠른 속력으로 운전할 확률이 높으며
고객의 위치에 따라
다양한 곳을 방문해야 할 확률이 높은
대리기사에게 투루카의 서비스는
경제적/시간적 측면에서 이득



대리기사와 대리기사가 아닌 고객의
마케팅 전략을 달리하여
기존 대리기사 고객 유지와 **새로운 잠재고객 유치**라는
두 가지 목표를 달성하고자 **대리기사 탐지 모델링** 고안

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

분석 Task | ① 대리기사 분류

사용 데이터 : 2023년 3월 17일 ~ 2023년 3월 21일까지 투루카에서 자체적으로 실시한 설문조사 데이터

1. 어떤 직무에 종사하나요?	2. 주중에는 어떤 목적으로 가장 많이 이용하시나요?	고객 ID
사무직	여가용	411
각워커(대리기사, 배달원, ...)	업무용	412
...

...

설문조사 결과를 통해 고객별 직무 및 주중 이용 목적에 대한 392개의 표본 확보
설문 데이터의 내용을 기반으로 모델에서 종속변수로 사용할 고객별 라벨 부여

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

분석 Task | ② 해석 제공

단순 분류 예측과 더불어 어떤 변수가 대리기사 여부를
결정함에 있어 큰 기여를 했는지에 대한 인사이트를 사측에 제공하고자 함

회귀 모델 적합 결과, 트리 모델의 Feature Importance,
SHAP Value 등을 종합적으로 활용 예정

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

분석 Task | ③ 다양한 파생변수 생성

	User_ID	HJD(행정동)	이용시작일시	...	유동인구
1번 유저 (2건)	1	명륜 1동	2022-05-31 00:00:00	...	123
	1	사직동	2022-05-31 12:00:00	...	323
3번 유저 (1건)	2	이화동	2022-08-05 17:00:00	...	124
	3	서초동	2022-05-31 18:00:00	...	356

이용 건수가 사용자 별로 상이함을 확인

고객별로 각기 다른 이용 건수를 합쳐
고객별로 그룹화한 데이터셋을 구성해야 함



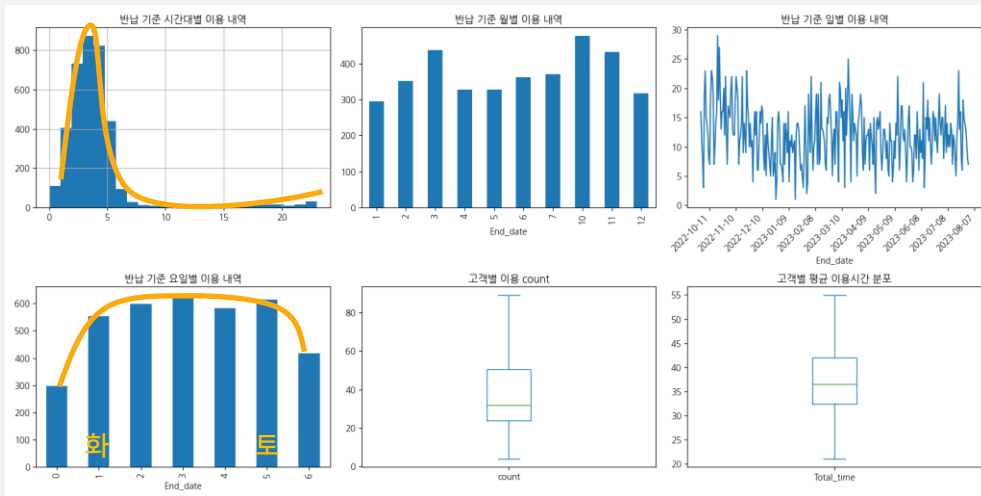
고객 간의 이용 건수 차이가 영향을 주지 않도록
개별 고객 자체의 이용 패턴을 반영할 수 있는
파생변수를 제작하자!

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 초기 EDA

가정 ① “대리기사는 야간에만 활동할 것이다”



대리기사의 투루카 이용 내역은
반납 기준 화요일~토요일, 0~5시
사이에 높게 분포

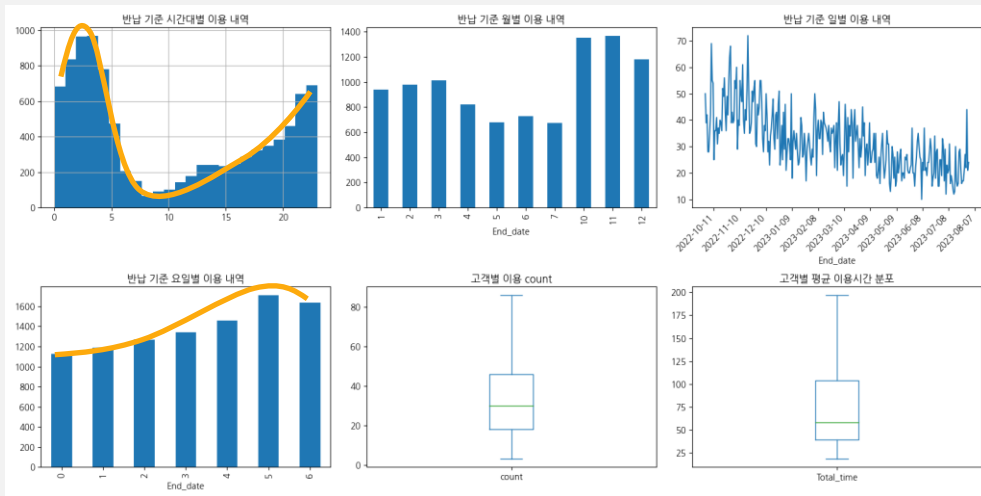
평균 이용 시간은 20~55분에 분포

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 초기 EDA

가정 ① “대리기사는 야간에만 활동할 것이다”



일반사용자의 투루카 이용 내역은
반납 기준 주말로 갈수록 증가,
8시부터 증가해 3시에
최고점을 찍으며 급격히 감소

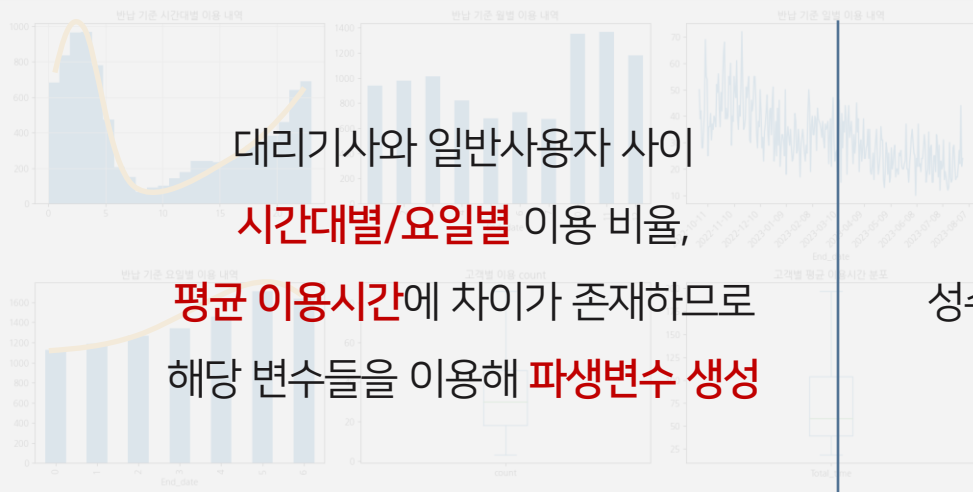
평균 이용 시간은 25~200분에 분포

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 초기 EDA

가정 ① “대리기사는 야간에만 활동할 것이다”



대리기사와 일반사용자 사이
시간대별/요일별 이용 비율,

평균 이용시간에 차이가 존재하므로
해당 변수들을 이용해 파생변수 생성

일반사용자의 투루카 이용 내역은
두 집단이 월별 이용량에서
상이한 분포를 보이지만,

성수기인 8~9월 데이터를 제공받지 못해

월별 이용 내역에 대한 활용은 보류

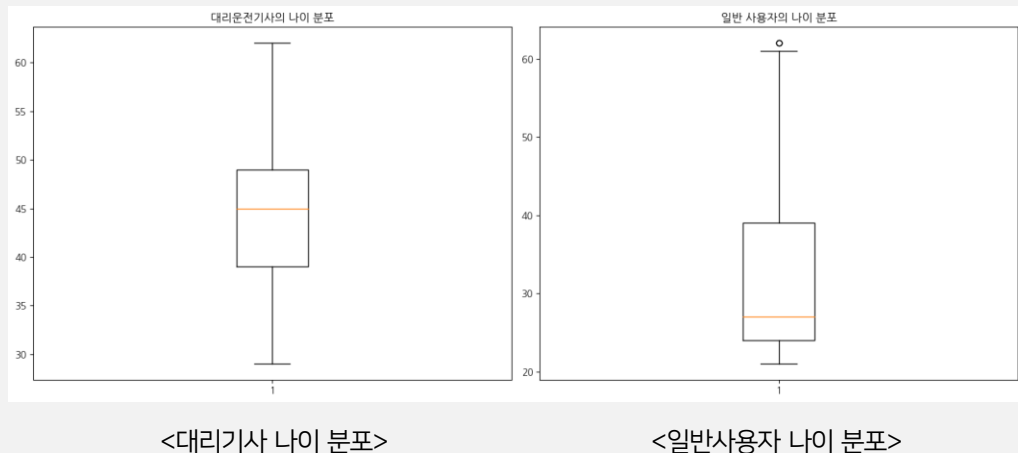
평균 이용 시간은 25~200분에 분포

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 초기 EDA

가정 ② “대리기사의 연령대가 일반사용자보다 높을 것이다”



대리기사와 일반사용자의 나이 분포는
대리기사가 더 높게 나타남

나이대가 높다면

대리기사일 것이라 추정 가능

03 과제 2: 대리기사 탐지 모델링

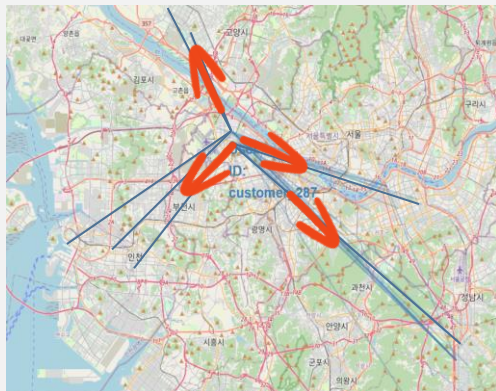
P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 초기 EDA

가정 ③ “대리기사는 일반사용자보다 더 넓은 이동반경을 기록할 것이다”



<일반사용자 customer_204>



<대리기사 customer_287>

일반사용자는 이동경로가 유사하나
대리기사의 이동경로는 산발적으로 분포

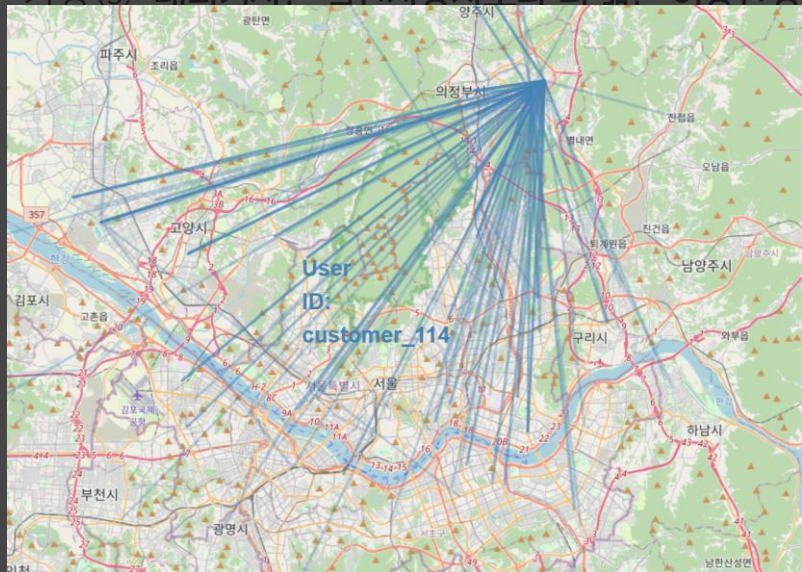
이동경로가 다양하면 높은 확률로
대리기사일 것이라 추정 가능

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 초기가정에 맞지 않는 EDA 결과 포착

가정 ③ “대리기사는 일반사용자보다 더 넓은 이동반경을 기록할 것이다”



<일반사용자 customer_114>

대리기사와 일반사용자로만 구분하여 EDA 진행 시
일반사용자 중에서도 이동경로가 산발적인 경우 존재
대리기사의 이동경로는 산발적으로 분포



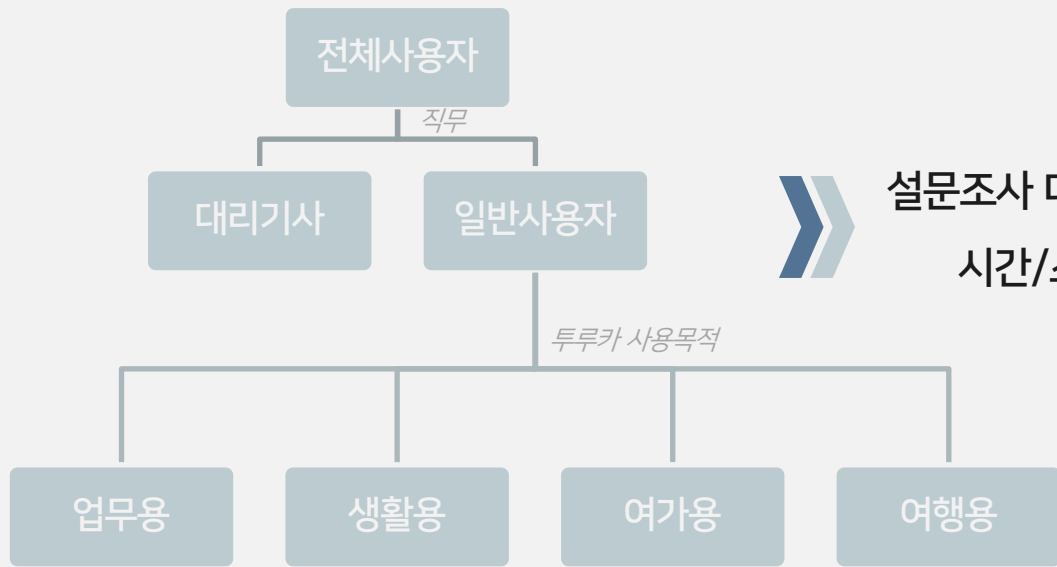
일반사용자를 세분화 한 EDA 필요성
이동경로가 다양하면 높은 확률로
대리기사일 것이라 추정 가능

customer_287>

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 일반사용자 세분화



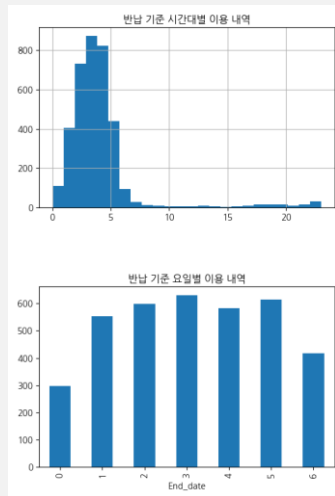
설문조사 데이터 기반 일반 사용자를 세분화 후
시간/스팟/사용자 관련 파생변수 생성

03 과제 2: 대리기사 탐지 모델링

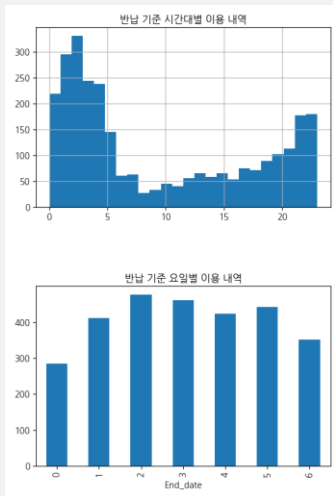
P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 시간 EDA - 요일/시간

설문조사 데이터 기반 일반 사용자를 세분화



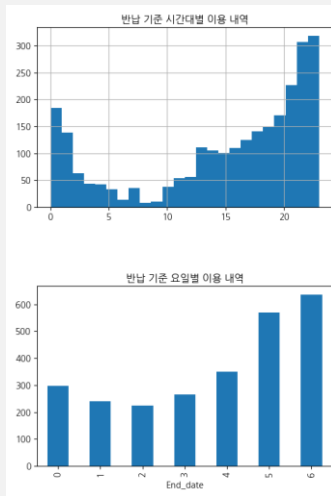
대리기사



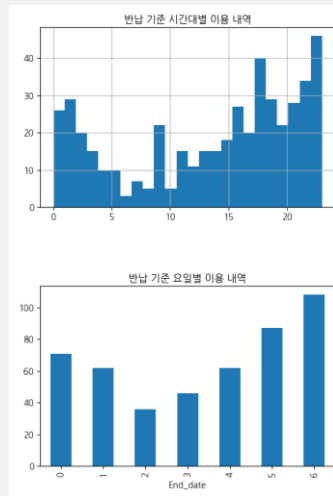
업무용 사용자



생활용 사용자



여가용 사용자



여행용 사용자

03 과제 2: 대리기사 탐지 모델링

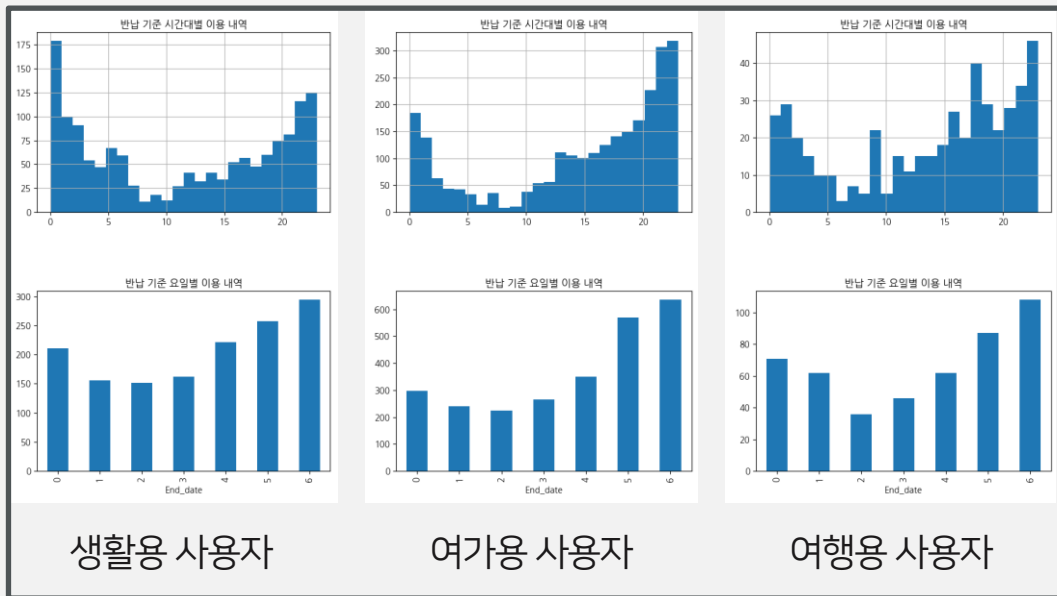
P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 시간 EDA - 요일/시간



대리기사

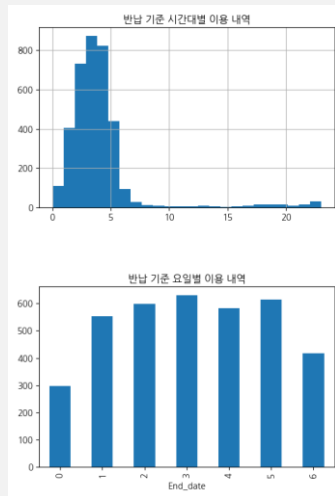
업무용 사용자



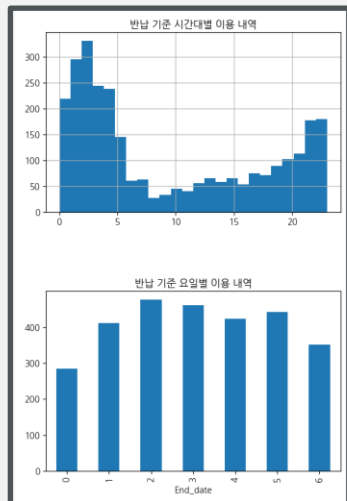
03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

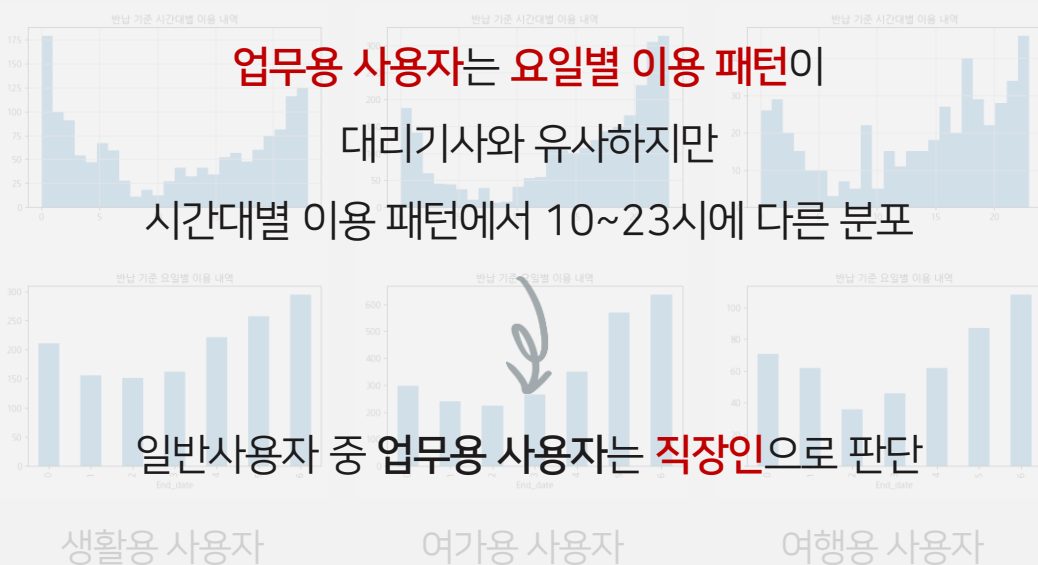
EDA 및 파생변수 | 시간 EDA - 요일/시간



대리기사



업무용 사용자



03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 시간 파생변수 - 요일별 이용 비율

User_ID	이용종료일시
12	2023-01-01 00:13:00
12	2023-01-02 00:13:00
12	2023-01-03 00:14:00
12	2023-01-05 00:15:00
...	...

User_ID	월요일 이용 비율	...	일요일 이용 비율
12	0.05	...	0.4
14	0.2	...	0.05
15	0.03	...	0.5
27	0.12	...	0.21
...

$$\text{월요일 이용 비율} = \frac{\text{월 이용 건수}}{\text{전체이용건수}}$$

⋮

$$\text{일요일 이용 비율} = \frac{\text{일 이용 건수}}{\text{전체이용건수}}$$

한 고객의 요일별 이용 비율 변수 생성

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 시간 파생변수 - 시간대별 이용 비율

User_ID	이용종료일시
12	2023-01-01 00:13:00
12	2023-01-02 00:13:00
12	2023-01-03 00:14:00
12	2023-01-05 00:15:00
...	...

User_ID	0~5시 이용 비율	10~23시 이용 비율
12	0.25	0.75
14	0.2	0.05
15	0.03	0.5
27	0.12	0.21
...

$$0\sim5\text{시 이용 비율} = \frac{0\sim5\text{시 이용 건수}}{\text{전체이용건수}}$$

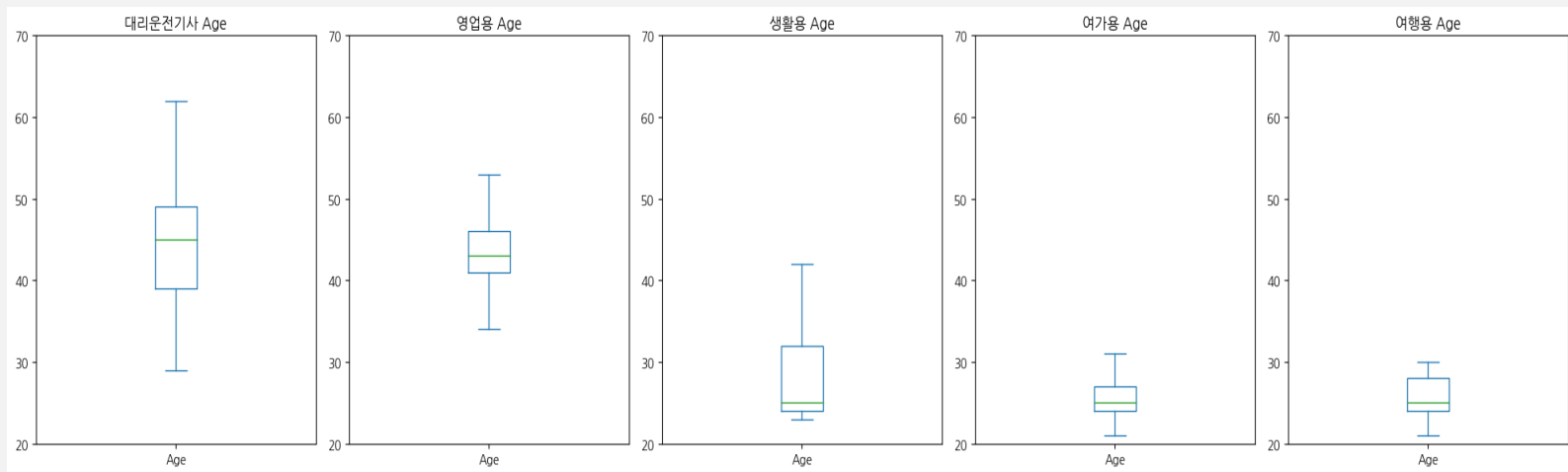
$$10\sim23\text{시 이용 비율} = \frac{10\sim23\text{시 이용 건수}}{\text{전체이용건수}}$$

차이가 존재하는 시간대에서의 한 고객의 이용 비율 변수 생성

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 사용자 - 나이



대리기사의 나이는 일반사용자의 나이 분포와 다르게 나타남

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 사용자 - 스팟 다양성

User_ID	출발스팟ID	도착스팟ID
8994	12	54
8994	21	39
8994	12	129
8810	22	22
...

출발 스팟의 다양성과

도착 스팟의 다양성 관련 변수 생성

: 0~1 사이 값으로 1에 가까울 수록

다양한 스팟을 이용함을 시사

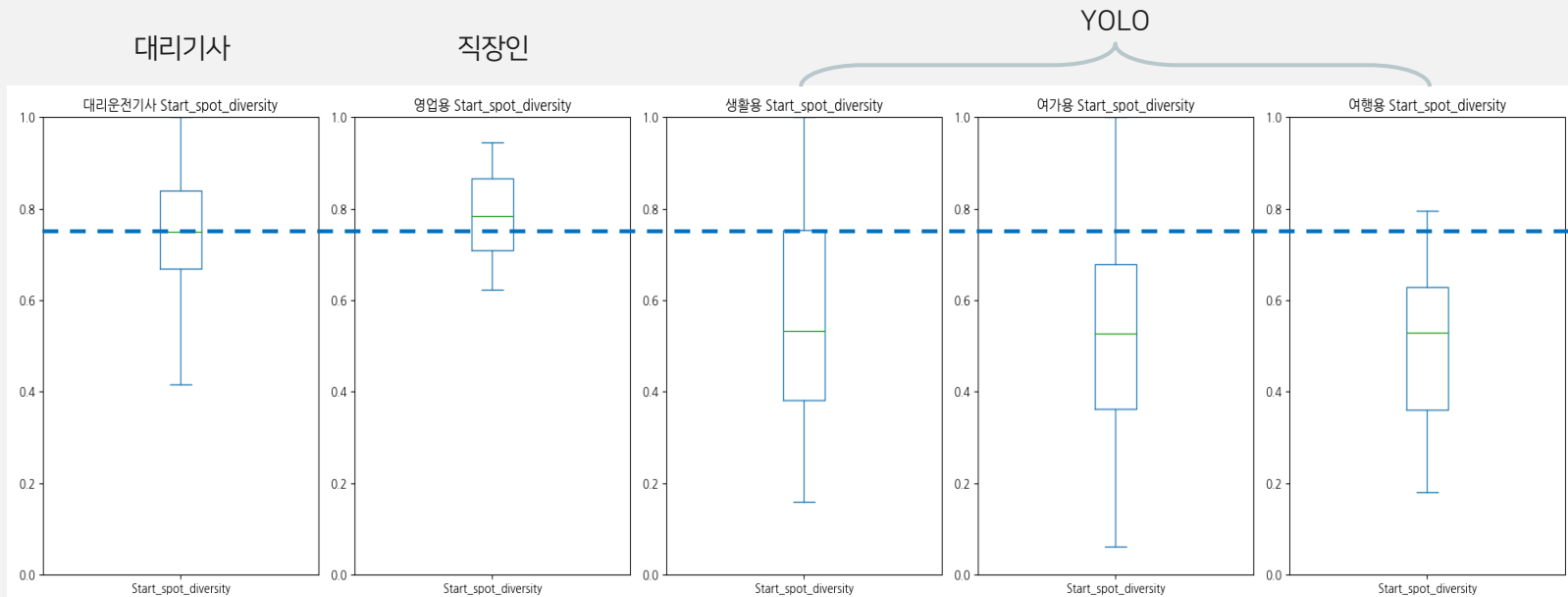
$$\text{Start_spot_diversity} = \frac{\text{Unique}(\text{출발 스팟})}{\text{전체이용건수}}$$

$$\text{End_spot_diversity} = \frac{\text{Unique}(\text{도착 스팟})}{\text{전체이용건수}}$$

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 사용자 - 스팟 다양성



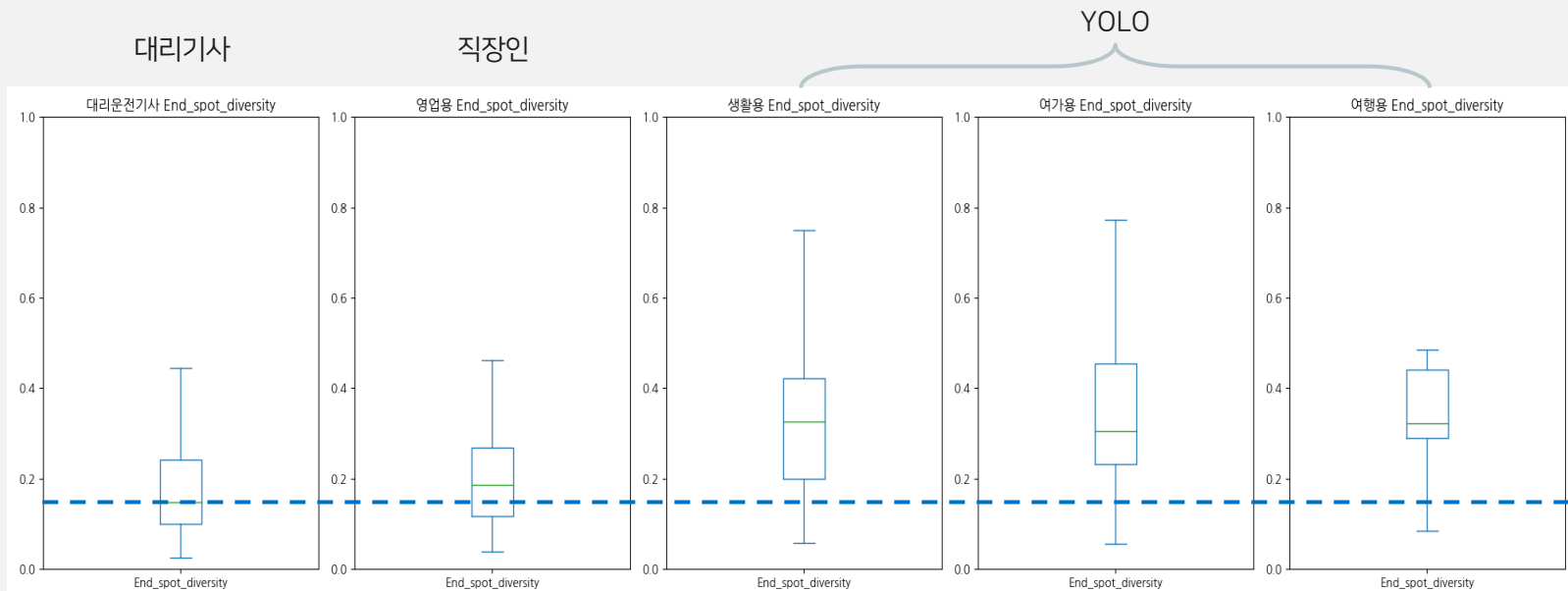
사용자 구분 별 출발 스팟 다양성

출발 스팟 다양성은 **대리기사가 YOLO 보다 높게** 나타남

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 사용자 - 스팟 다양성



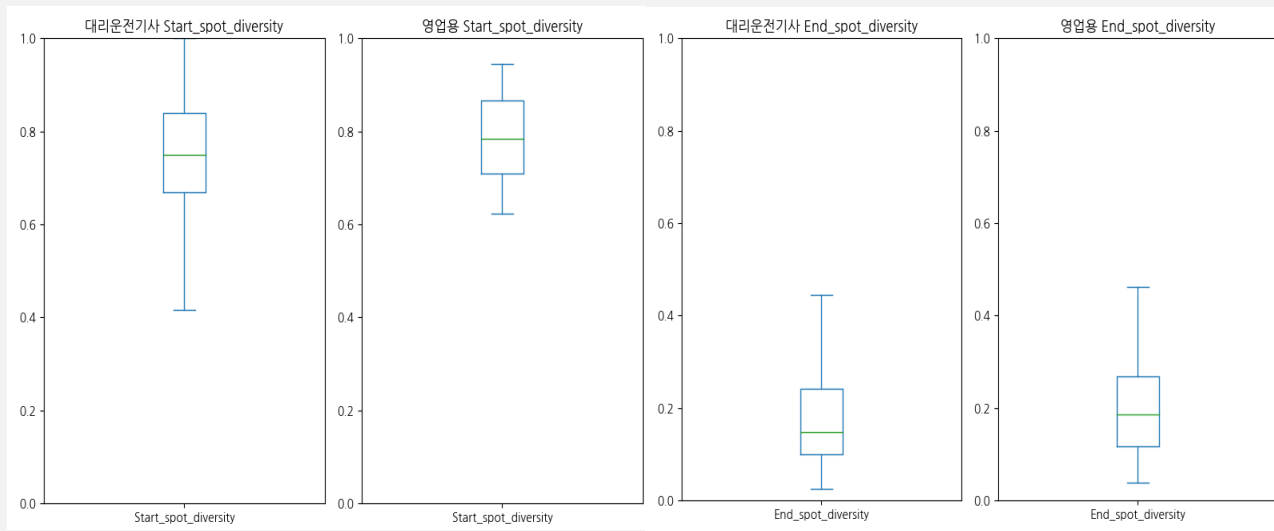
사용자 구분 별 도착 스팟 다양성

도착 스팟 다양성은 **대리기사가 YOLO 보다 낮게** 나타남

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 사용자 - 스팟 다양성



대리기사와 직장인의 출발 / 도착 스팟 다양성

출발과 도착 스팟 다양성 모두 대리기사와 직장인에서 유사한 분포

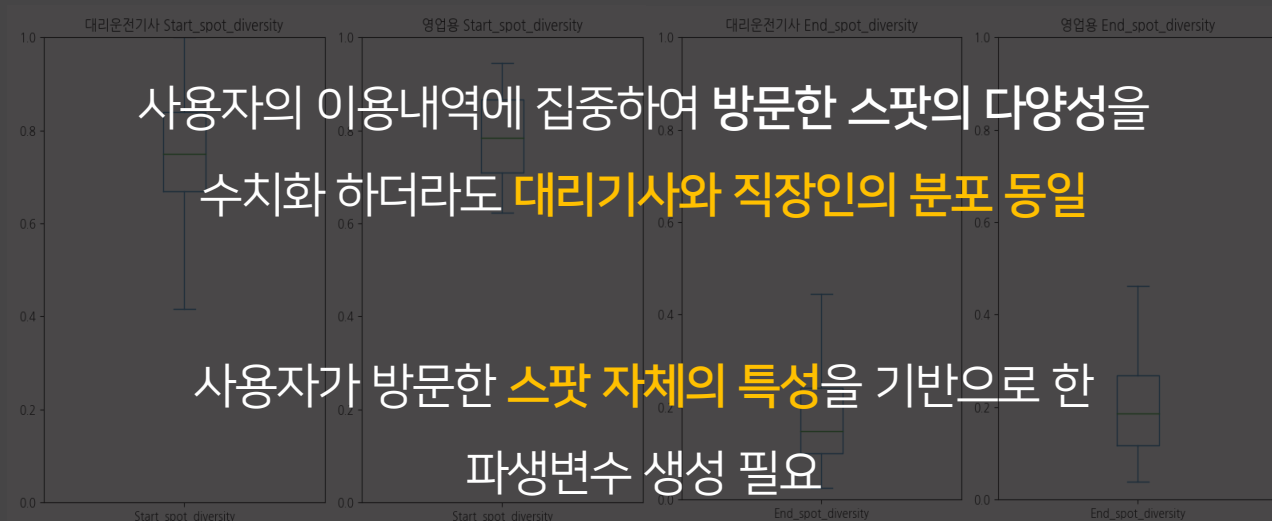
03 과제 2: 대리기사 탐지 모델

P-SAT 24-2학기
시계열자료분석팀



EDA 및 파생변수 | 사용자 - 스팟 다양성

스팟 다양성의 한계



대리기사와 직장인의 출발 / 도착 스팟 다양성

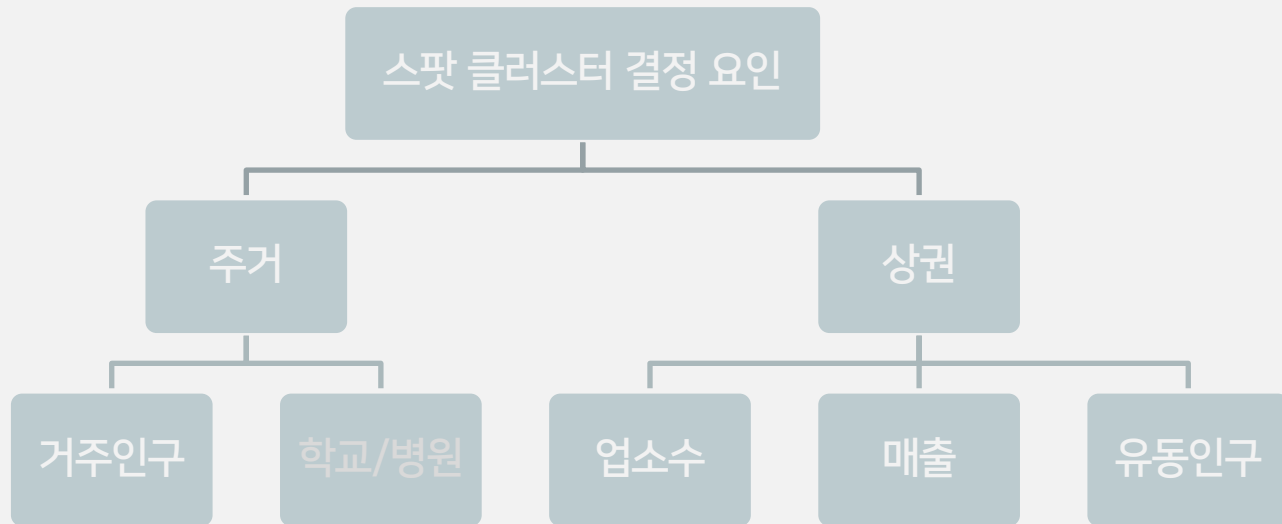
출발과 도착 스팟 다양성 모두 대리기사와 직장인에서 유사한 분포

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링

투루카 측에서 고려한 스팟 결정 요인 : 주거와 상권



주거 요인과 상권 요인을
혼재하는 변수라 판단하여 제거

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 변수 수집

데이터	출처
행정동별 거주인구	통계청
행정동별 업소수	소상공인 마당 상권정보 분석 시스템의 행정동별 업소수 크롤링
행정동별 매출	소상공인 마당 상권정보 분석 시스템의 행정동별 매출액 크롤링
행정동별 유동인구	소상공인 마당 상권정보 분석 시스템의 행정동별 유동인구 크롤링



03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 변수 수집

Spot_ID	주소
163	서울특별시 종로구 세종대로 123길 47
231	서울특별시 중구 서소문로 32
116	서울특별시 구로구 디지털로 12길 39
97	서울특별시 서대문구 통일로 81
...	...

기존 스팟 리스트

위/경도 맵핑
(Kakao API)



Spot_ID	경도	위도
163	127.070056	37.638130
231	127.068935	37.639343
116	127.072356	37.638540
97	127.022935	37.679343
...

행정동 맵핑



Spot_ID	행정동
163	사직동
231	길음1동
116	성산1동
97	가양1동
...	...

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 변수 수집

Spot_ID	행정동
163	사직동
231	길음1동
116	성산1동
97	가양1동
...	...

행정동 단위로
거주인구, 업소수,
매출, 유동인구 맵핑



Spot_ID	거주인구	업소수	매출	유동인구
163	168945	672	37272668	9006
231	194845	352	2485455	8248
116	278785	142	1969310	18248
97	157645	237	4484765	6973
...

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 최종 데이터셋

Spot_ID	행정동	거주인구	업소수	매출	유동인구	경도	위도
163	사직동	168945	672	37272668	9006	127.070056	37.638130
231	길음1동	194845	352	2485455	8248	127.068935	37.639343
116	성산1동	278785	142	1969310	18248	127.072356	37.638540
97	가양1동	157645	237	4484765	6973	127.022935	37.679343
...

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링

Kmeans Clustering

데이터를 K개의 그룹으로 나누는 알고리즘으로,
각 클러스터는 중심(centroid)에서 가장 가까운 데이터를 모아 형성

투루카측에서는 스팟에 대해 관찰할 때 상권과 거주로 분류



서울과 같은 도심지에는 상권과 거주가 명확히 구분되지 않고
혼재된 경우가 존재하기 때문에 클러스터의 개수를 3개로 두고 진행



03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 파생변수

User_ID	도착스팟ID	스팟 클러스터
8994	12	0
8994	21	0
8994	12	1
8810	22	2
...

사용자별 방문한 도착 스팟 중
각 클러스터의 비율

$$\text{Cluster 0 Ratio} = \frac{\text{Cluster 0 방문 횟수}}{\text{전체이용건수}}$$

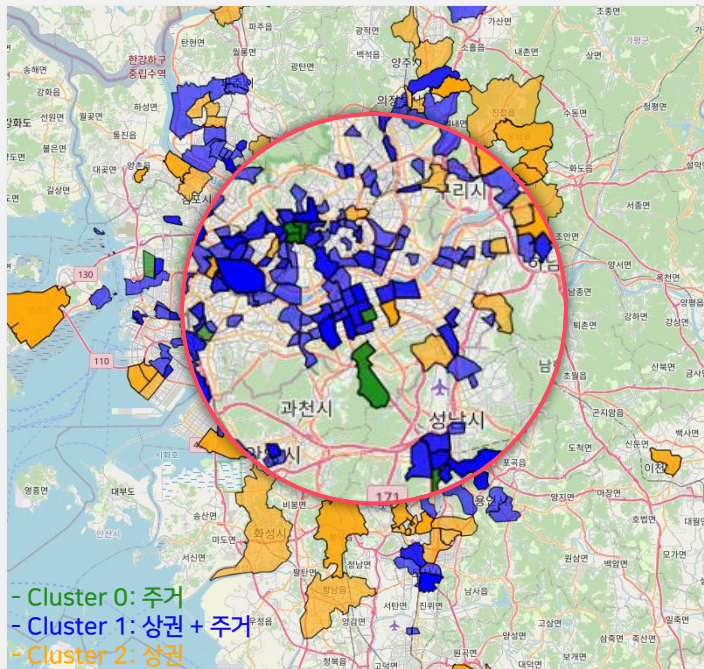
$$\text{Cluster 1 Ratio} = \frac{\text{Cluster 1 방문 횟수}}{\text{전체이용건수}}$$

$$\text{Cluster 2 Ratio} = \frac{\text{Cluster 2 방문 횟수}}{\text{전체이용건수}}$$

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 해석

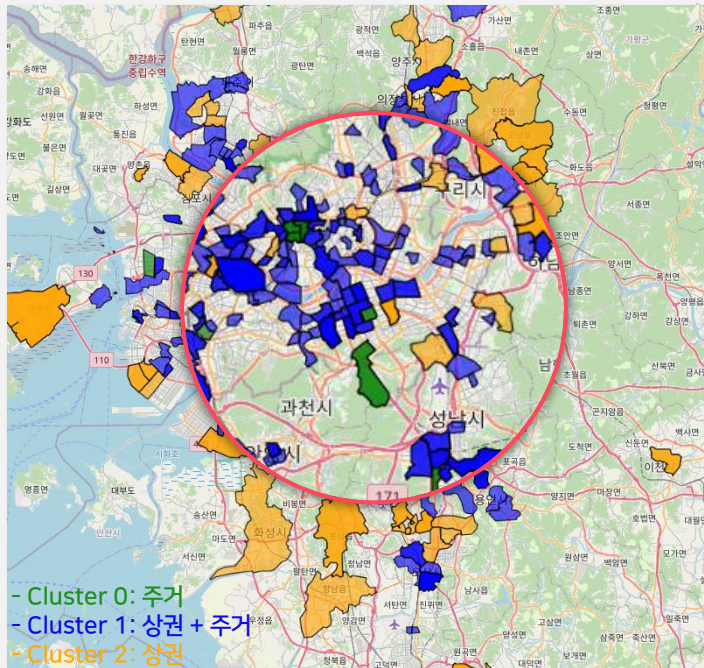


클러스터링 결과, **Cluster 0(초록)** 이
주거<상권인 상권지로 분류되어,
해당 클러스터에는 강남구 테헤란로,
중구 등이 포함되어 직관에 부합

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 클러스터링 해석



Cluster 1(파랑)의 경우 주거와 상권이 혼합된 클러스터로 서울 내의 대부분의 지역이 해당

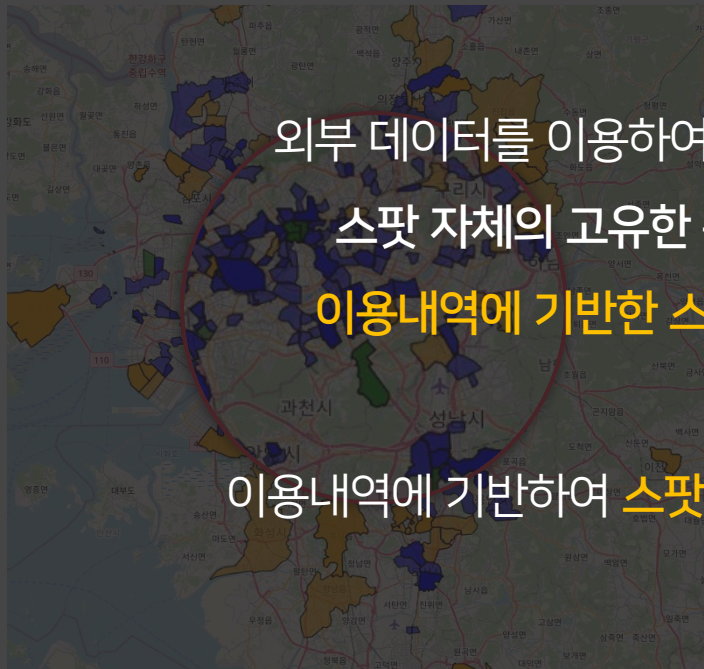
Cluster 2(주황)의 경우 주거의 비율이 높은 클러스터로 경기도와 서울 외곽에 다수 분포

03 과제 2: 대리기사 탐지 모델

P-SAT 24-2학기
시계열자료분석팀



EDA 및 파생변수 | 스팟 - 클러스터링 해석 클러스터링의 한계



외부 데이터를 이용하여 스팟을 클러스터링 한 경우

스팟 자체의 고유한 특성은 고려할 수 있으나

이용내역에 기반한 스팟 간의 관계를 고려 불가
서울 내의 대부분의 지역에 해당
주황의 경우 주거의 비율이 높은 클러스터로

경기도와 서울 외곽에 다수 분포

이용내역에 기반하여 스팟 간의 관계를 고려한 지표 필요 !

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 중심성

고유 벡터 중심성 (Eigenvector Centrality)

주변 node의 중심성을 고려하여 Network를 통해 각 node의 중요도를
무한히 전파시켰을 때, 수렴하게 되는 중요도의 양의 비율



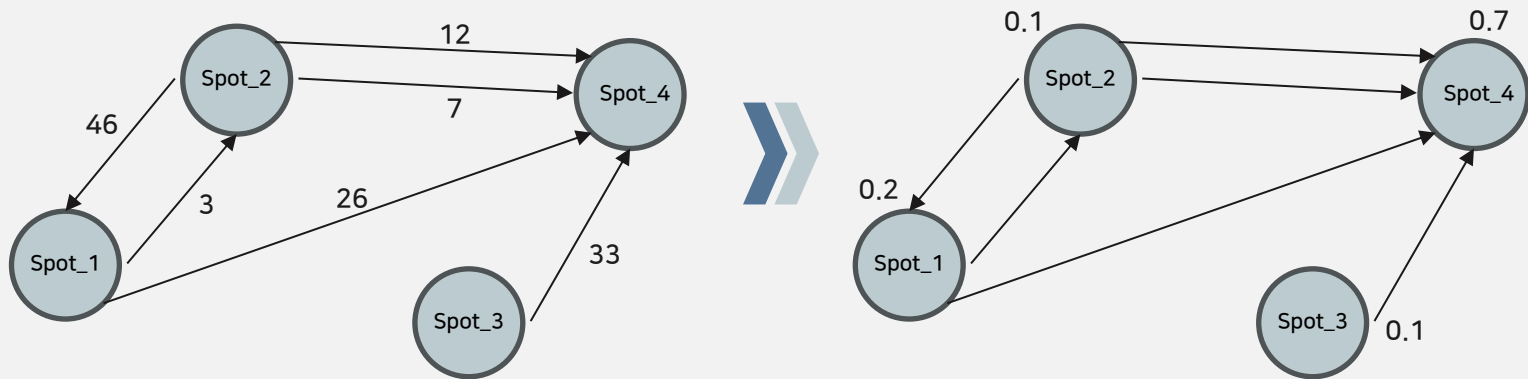
연결된 엣지의 수만 고려하는 것이 아닌
연결된 노드의 중요도도 함께 반영되기 때문에
네트워크 내 **실질적 영향력** 판단 가능



03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 중심성

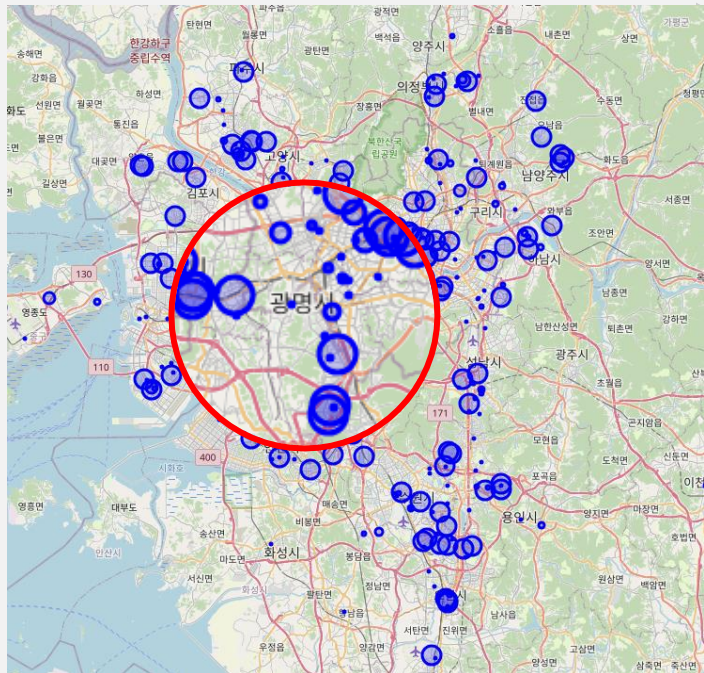


각 Spot을 Node로 간주하고 대리기사의 이용 내역을 기반으로
그래프 구축 후 고유 벡터 중심성 계산

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 중심성

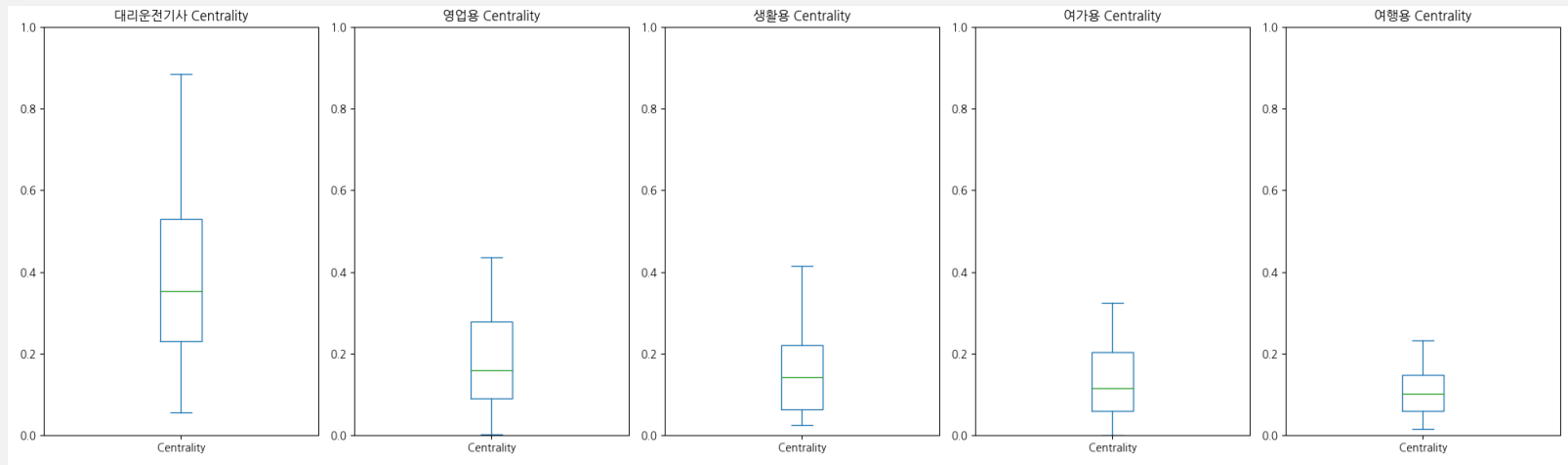


각 스팟의 중심성 시각화 결과
각 지역별 대리기사가 **자주 방문**하는 스팟인 동시에
인근 스팟과 연결된 지역에 높은 중심성 값 부여

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

EDA 및 파생변수 | 스팟 - 중심성



각 사용자별 방문한 스팟의 중심성 간 평균치로
대리기사와 직장인, YOLO를 분리할 수 있음

P-SAT 24-2학기
시계열자료분석팀

[illegible]

03 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

모델링

전처리 및 파생변수의 유의성 확인을 위해 기본적인 모델링 시도
Train - Test set을 8:2의 비율로 나눈 후 각 모델의 성능 비교

자세한 모델링 과정은 P-SAT 네이버 카페를 확인해주세요!

Logistic Regression / SVM / Gradient Boosting /
XGBoost / LightGBM / CatBoost / RandomForest Classifier

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | SHAP

SHAP Value

개별 특성이 모델의 예측에 미친 영향을 수치적으로 나타내는 값



이진 분류에서 SHAP value는 log-odds인 $f(x)$ 에 대해
주어진 특성 값이 log-odds에 기여하는 정도를 수치화



Feature Value의 범위에 따라 Shap value가 다르게 분포하면

해당 Feature가 예측에 어떻게 기여하는지 파악할 수 있음

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | SHAP

SHAP Value

개별 특성이 모델의 예측에 미친 영향을 수치적으로 나타내는 값

단, 회귀 모델과 달리 직접적인 선형관계로 해석 불가

이진 분류에서 SHAP value는 log-odds인 $f(x)$ 에 대해
주어진 특성 값이 log-odds에 기여하는 정도를 수치화

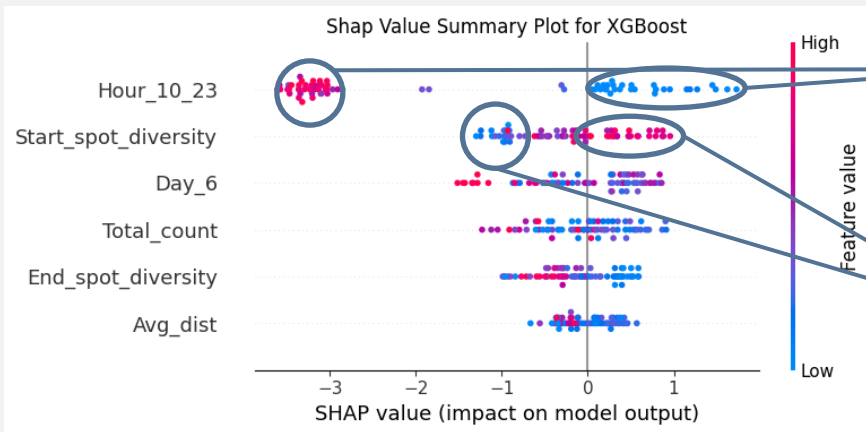


어떻게 예측을 내렸는지에 대한 상대적인 기여도를 제공하여 모델의 예측 과정에 대한
직관적인 이해는 제공할 수 있으나 변수간 상호작용과 비선형 관계를 고려해야 함

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | SHAP



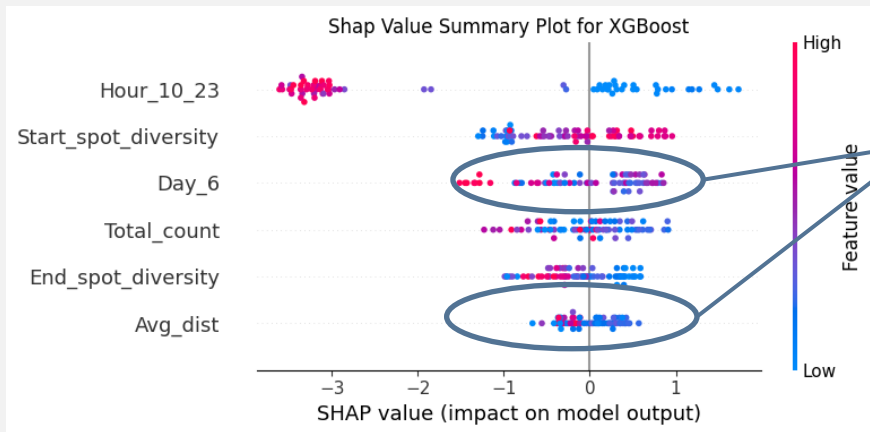
10~23시 이용시간 비율이 **높을수록**
대리기사일 확률을 **낮추는** 방향으로 기여

출발 스팟 다양성이 **클수록**
대리기사일 확률을 **높이는** 방향으로 기여

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | SHAP



일요일 이용 비율, 평균 이동거리

Feature value의 범위에 따라

SHAP value가 뚜렷하게 구분되지 않음



예측에 미치는 영향이 일정하지 않거나
복잡한 상호작용이 존재할 수 있기 때문에 해석 보류!

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | Waterfall

Waterfall Plot

각 특성이 예측값에 증가 또는 감소한 기여도를 단계적으로 나타낸 그래프



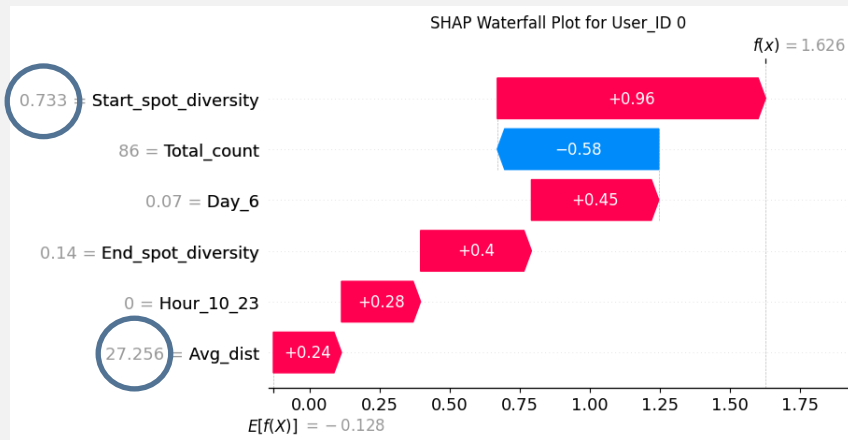
이진 분류 상에서 $f(x)$ 는 log-odds로
주어진 feature value가 log-odds에 어떤 기여를 하는지 파악 가능

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | Waterfall - 대리기사

$f(x) = 1.626$ 은 최종 확률 0.834를 의미



0.733의 출발 스팟 다양성은
log-odds 에 +0.96의 기여

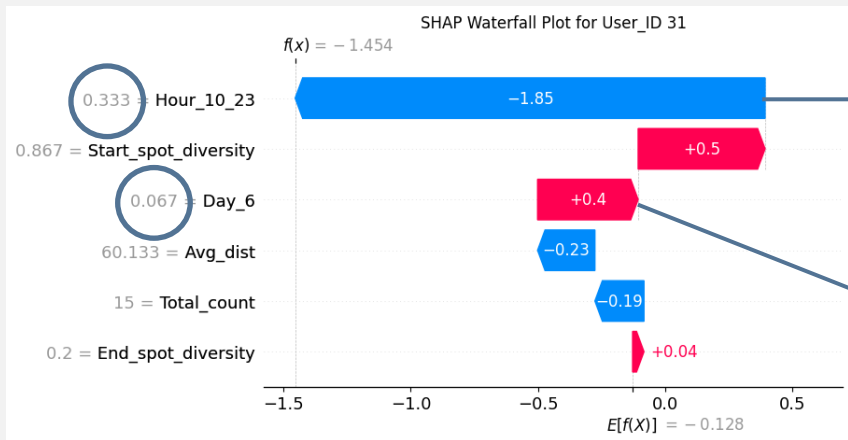
27.256의 평균 이동 거리는
log-odds 에 +0.24의 기여

출발 스팟이 다양할수록 대리기사일 확률이 높을 것이라는
앞선 해석 결과와 일맥상통

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | Waterfall - 직장인



0.333의 10~23시 이용 비율은
log-odds 에 -1.85의 기여

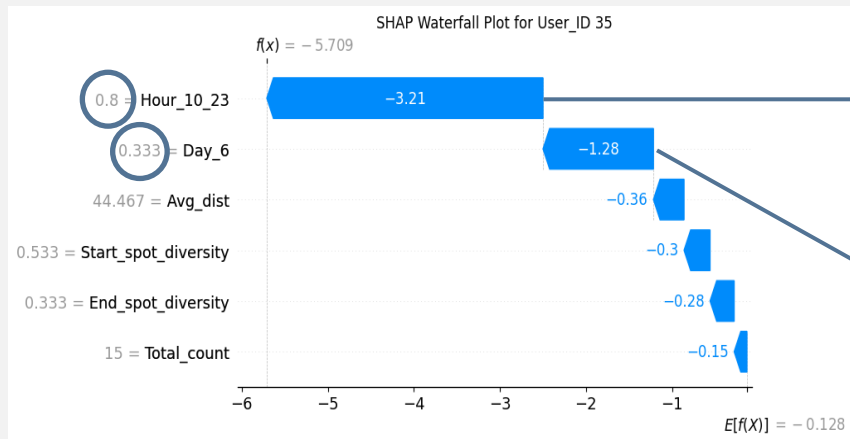
0.067의 일요일 이용 비율은
log-odds 에 +0.4의 기여

10-23시 이용 비율이 대리기사일 확률을 낮출 것이라는
앞선 해석 결과와 일맥상통

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | Waterfall - YOLO



0.333의 10~23시 이용 비율은
log-odds 에 -3.21의 기여

0.333의 일요일 이용 비율은
log-odds 에 -1.28의 기여

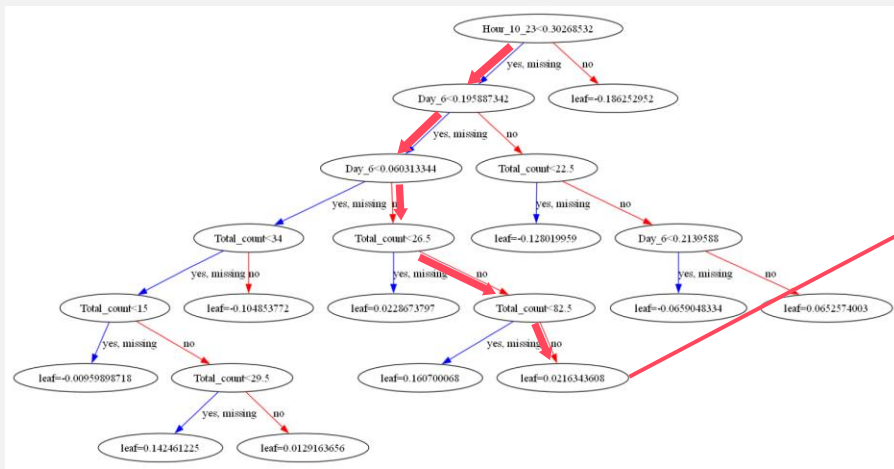
일요일 비율이 양의 영향을 미쳤던 직장인과 차이가
생긴다는 점에서도 직관과 부합

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | 대리기사 분기예시

Tree Index 0



Base 확률 0.5 + Leaf Node value 0.0216
→ 트리 한 개만 이용했을 때 **0.5216**으로 예측

이후 트리의 Leaf Node 값을 합하여
최종 대리기사일 확률 0.8356 도출

대리기사 고객의 분기 기준

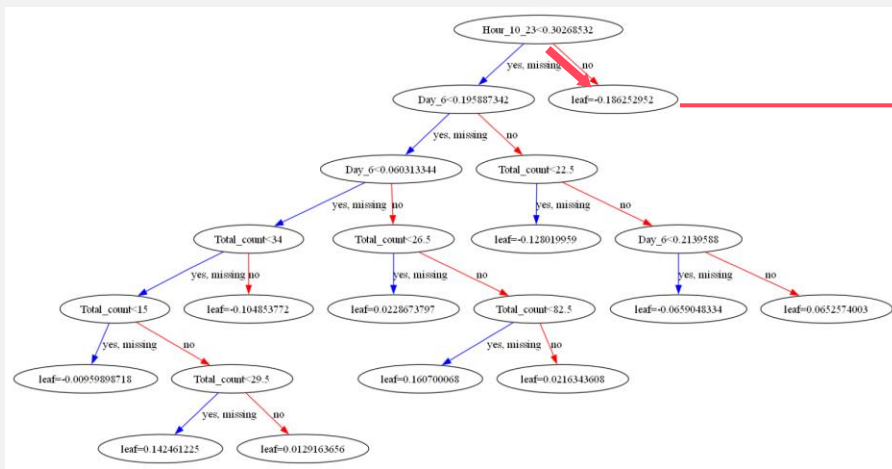
분류	Total_count	Day_6	Hour_10_23	Start_spot_diversity	End_spot_diversity	Avg_dist
대리기사	86	0.0697	0.0	0.7325	0.1395	27.2558

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

결과 해석 | 대리기사 분기예시

Tree Index 0



Base 확률 0.5 + Leaf Node value -0.1862

→ 트리 한 개만 이용했을 때 **0.3138**로 예측

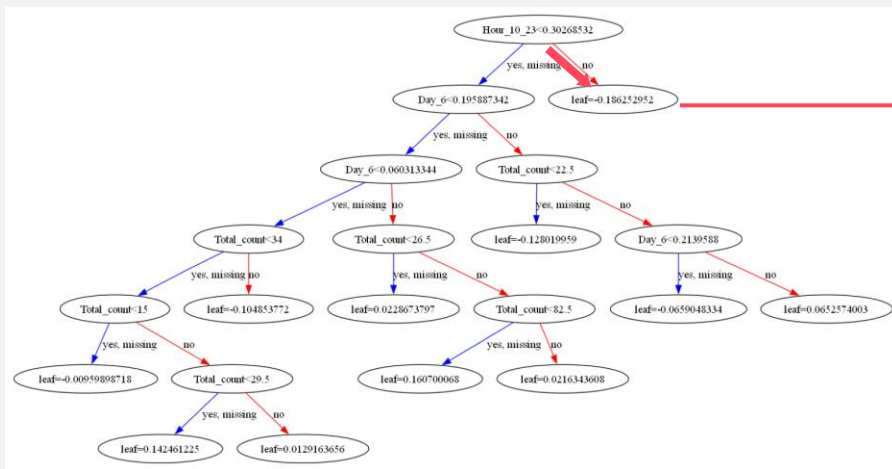
이후 트리의 Leaf Node 값을 합하여

최종 대리기사일 확률 0.1893 도출

대리기사 고객의 분기 기준

분류	Total_count	Day_6	Hour_10_23	Start_spot_diversity	End_spot_diversity	Avg_dist
직장인	15	0.0667	0.3333	0.8667	0.2	60.1333


Tree Index 0



→ 트리 한 개만 이용했을 때 0.3138로 예측

이후 트리의 Leaf Node 값을 합하여
최종 대리기사일 확률 0.0033 출력

대리기사 고객의 분기 기준

분류	Total_count	Day_6	Hour_10_23	Start_spot_diversity	End_spot_diversity	Avg_dist
YOLO	15	 0.3333	0.8	0.5333	0.3333	44.4667

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

추가 결과 해석 | 로지스틱 회귀모형

로지스틱 회귀모형 (Logistic Regression Model)

범주형 종속변수에 대해 독립 변수의 선형 결합을 통해
종속변수가 특정 범주에 속할 확률을 모델링하는 통계적 기법



자세한 내용은 P-SAT 네이버 카페를 확인해주세요!

독립 변수들이 어떻게 영향을 미치는지 확인할 수 있음

최종 결과 해석



XGBoost 모델 적합 결과, 주야간 이용 시간이 낮을수록, 출발 스팟 다양성이 높을수록 대리기사일 확률은 증가



XGBoost 분기 처리 과정에서 Day_6의 빈도가 가장 높으며 주야간 이용 시간 비율은 gain 측면에서 가장 높은 변수 중요도를 보임



Logistic 모델 적합 결과 EDA에서 얻은 인사이트와 동일한 방향의 계수를 가지며 XGBoost의 Shap value 기반 해석과도 동일한 방향의 해석이 가능함

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

추가 과제 | 소표본 분석



투루카 서비스운영팀

중간 미팅 후...

대리기사 탐지를 위해 각 고객의 이용내역 데이터가
몇 건 정도 쌓여야 할까요?

사측은 최대한 적은 이용내역으로 특정 고객의 대리기사 여부를 판단할 수 있기를 바램



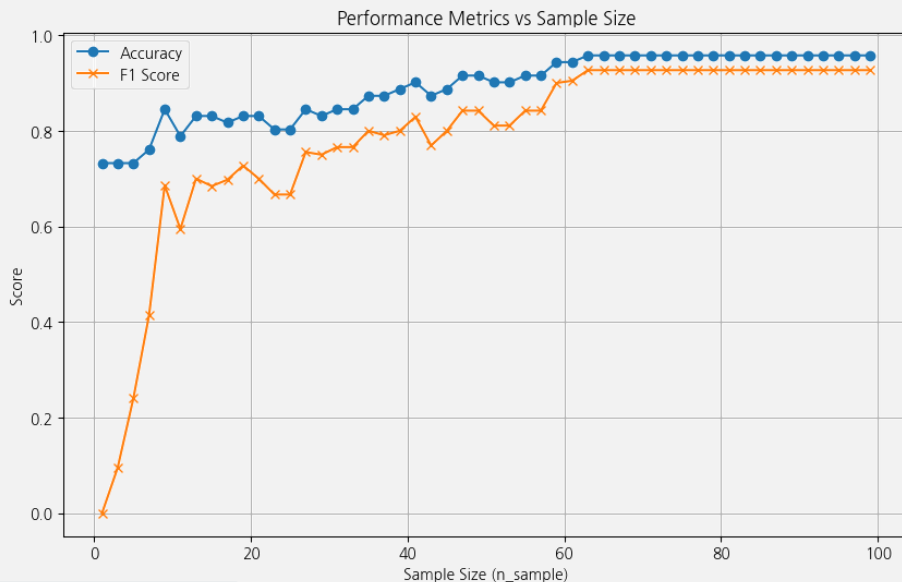
고객별 이용내역 건수(n)를 달리했을 때 모델 평가지표의 변화를 관찰하자

($n = 1, 3, 5, 10, 15, \dots$)

01 과제 2: 대리기사 탐지 모델링

P-SAT 24-2학기
시계열자료분석팀

추가 과제 | 소표본 분석



소표본 분석 결과

- Accuracy 0.8 이상 : 이용내역 **13개** 이상
- F1 0.8 이상 : 이용내역 **35개** 이상

이용내역 분포에 따라 결과가 달라질 수 있기에
테스트 결과에 절대적인 정답은 존재하지 않음
(현재 데이터 상의 추정치일 뿐)

04

과제 3

스팟 수요 요인 분석

02 과제 3: 스팟 수요요인 분석

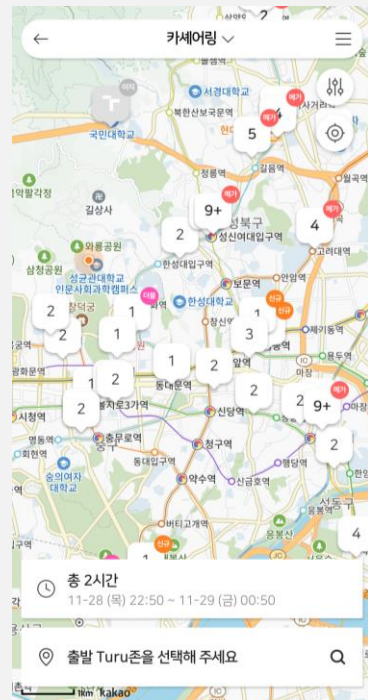
P-SAT 24-2학기
시계열자료분석팀

주제 선정 배경

‘어떤 지역에 스팟을 선정해야 할까?’

공유차량 업계에서 스팟 선정은 매출과 직결되는 요소이므로

‘수요가 높을 만한 스팟’을 선정하는 것은 매우 중요



투루카 카세어링 서비스 애플리케이션 화면

02 과제 3: 스팟 수요요인 분석

P-SAT 24-2학기
시계열자료분석팀

주제 선정 배경

기존 투루카의 스팟 선정 방식

유효고객 (20대~50대 남성) 중 메인 타겟 고객층인

특정 연령대 가 많이 있을 법한 특정 장소

NDA 이슈로 정보 공개가 불가합니다 T.T

이를 보완하여, 스팟 인근의 인구 사회학적 특징을 고려한
스팟 선정 프로세스 구축을 요구하심

이후 얻은 인사이트를 바탕으로 미진출 스팟을 선정하고,
실제로 거점존 운영을 진행하기로 결정 (1월 中)

02 과제 3: 스팟 수요요인 분석

P-SAT 24-2학기
시계열자료분석팀

주제 선정 배경

분석 최종 목적

스팟 유형별 왕복 카셰어링의 수요 요인 분석을 통한
매출 극대화 및 스팟 선정 프로세스의 효율화

과제 3의 분석 과정은 P-SAT 네이버 카페를 확인해주세요!

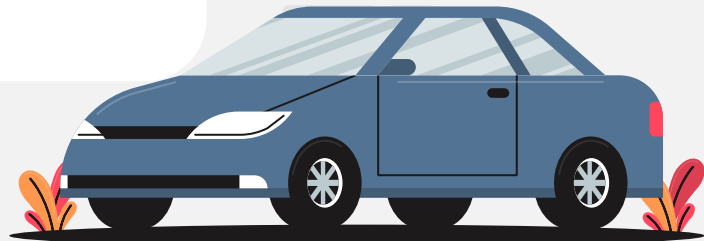
스팟 수요 요인 분석

스팟 선정 프로세스
효율화

4주차 주제 분석에서는 해당 단계 분석 과정 일부에 대해 발표할 예정!

방학 주제분석 예고

- 01 과제 3
효율적 스팟 선정 프로세스
- 02 과제 3
매출 최적화 전략 수립
- 03 최종발표
투루카 대표님과 함께
하는 세미나



감사합니다

