



발표 로딩중..

bada

바다톤 최종발표

바다의 공주들
이세은 김윤아 정성희

©YOUNG TOYS, INC. All Rights Reserved.



Contents



1. Business Problem 정의
2. 데이터 분석 문제 정의
3. 분석 결과
4. 비즈니스적 인사이트 제안
5. 한계점

Business Problem 정의

택시요금 인상
[주간요금: 오전 4시 - 오후 10시]

	기존	인상 후
기본요금	3,800원	4,800원
기본거리	2km	1.6km
거리당요금	132m/100원	131m/100원
시간당요금	31초/100원	30초/100원

심야요금 할증
[오후 10시- 다음날 오전4시]

피크타임 [오후 11시 - 오전 2시]
: 할증률 40%

그외시간 [오후 10-11시, 오전 2-4시]
: 할증률 20%

Business Problem 정의

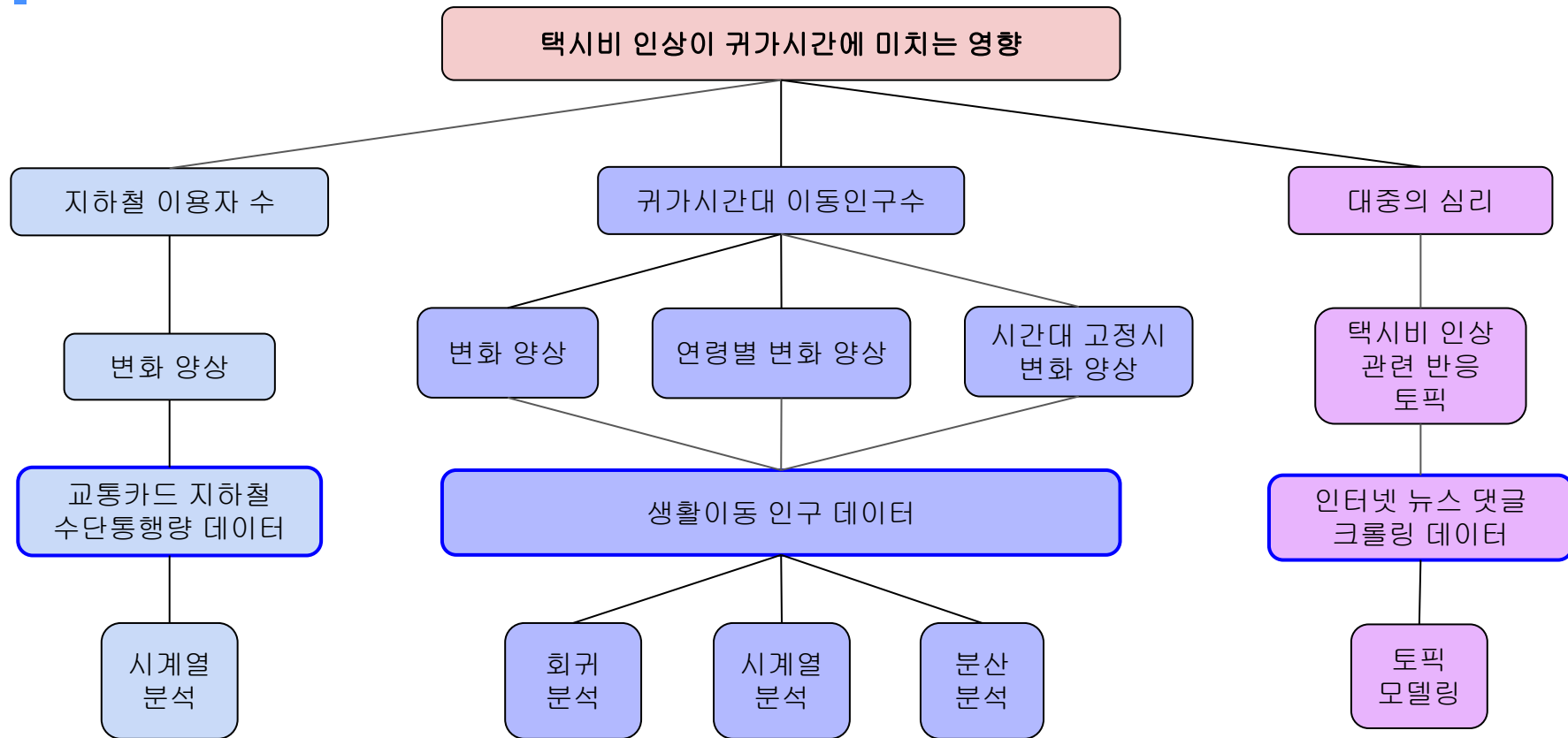


데이터 분석 문제 정의

택시비 인상이 귀가시간에 미치는 영향

- 1) 택시비 인상 전후로 귀가시간대(18시~익일 5시)에 이동인구는 어떠한 양상으로 변화했는가
- 2) 택시비 인상 전후로 지하철 이용자수는 어떻게 변화했는가
- 3) 택시비 인상에 대한 대중의 심리는 어떠하며, 특히 어느 키워드에 가장 민감하게 반응하는가

도식화



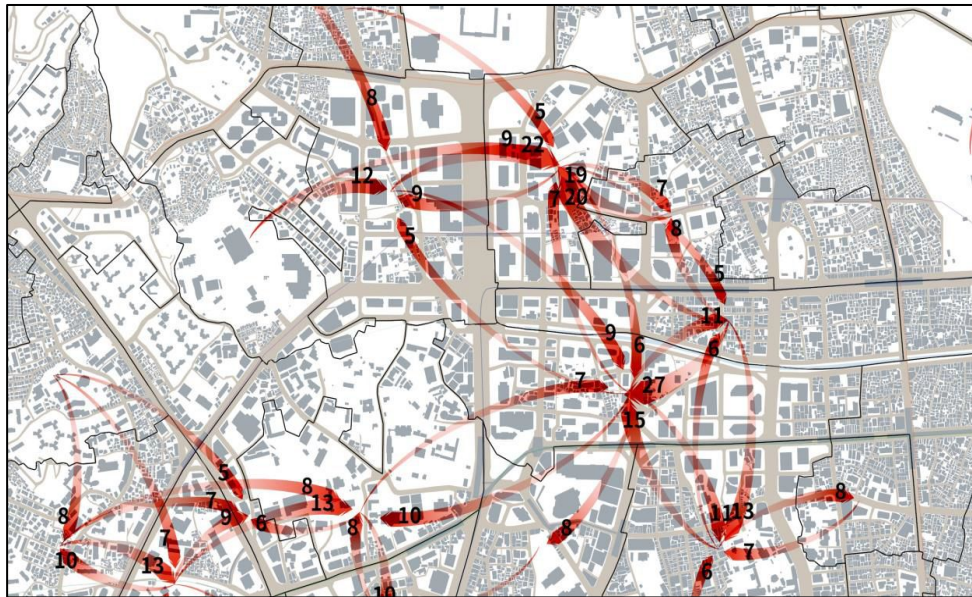
1. 택시비 인상이 귀가시간대 인구 이동수에 미치는 영향

사용 데이터

[서울생활이동 인구 데이터]

-> 생활이동: 특정 시점(일별, 시간대별)에 서울 안에서 이동하거나 서울 외부에서 서울로 오고 간 사람들의 이동

-> KT 휴대폰을 사용하는 사람들 중 LTE와 5G사용자의 신호를 바탕으로 KT의 시장점유율 등을 고려한 여러 가지 보정 과정을 거쳐 전체 인구의 이동으로 확장 보정해 산출한 데이터



데이터 사용 범위

1. 출발 지역구: 서울특별시 내 모든 지역구
2. 도착 지역구: 송파구, 강서구, 강남구, 노원구, 관악구

-선정기준: 서울특별시 내 주민등록세대수 1위 ~ 5위

-사용 데이터: 행정구역(시군구)별 주민등록세대수

3. 시간: **2022-12, 2023-01, 2023-02, 2023-03** 데이터셋 중

18시, 19시, 20시, 21시, 22시, 23시, 익일 00시, 01시, 02시, 03시, 04시

4. 이동구분: **EH, HH, WH**

***W: 주간상주지(근무지 혹은 학교), H: 야간상주지, E: 기타 (주간상주지와 야간상주지를 제외한 지역)

데이터 전처리 - 생활이동 데이터

	대상연월	요일	도착시간	출발 시군구 코드	도착 시군구 코드	성별	나이	이동유형	평균 이동 시간(분)	이동인구(합)
0	202212	일	0	11010	11010	F	0	EE	27	83.82
1	202212	일	0	11010	11010	F	10	EH	75	17.68
2	202212	일	0	11010	11010	F	10	HH	59	11.8
3	202212	일	0	11010	11010	F	10	WW	66	6
4	202212	일	0	11010	11010	F	10	EE	96	11.92

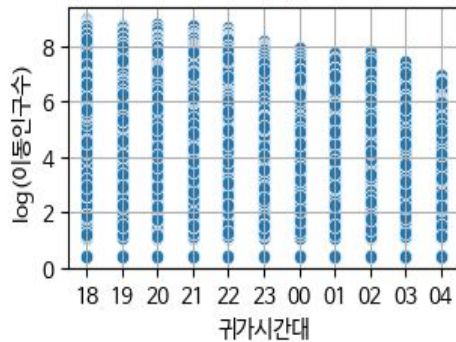


	대상연월	요일	출발 시군구 코드	도착 시군구 코드	성별	나이대	이동유형	이동인구수
귀가시간대								
00:00:00	2022-12-01	금	11010	11110	F	20대	EH	225.18
00:00:00	2022-12-01	금	11010	11110	F	20대	HH	1.50
00:00:00	2022-12-01	금	11010	11110	F	20대	WH	175.66
00:00:00	2022-12-01	금	11010	11110	F	30대	EH	91.37
00:00:00	2022-12-01	금	11010	11110	F	30대	WH	79.61
...

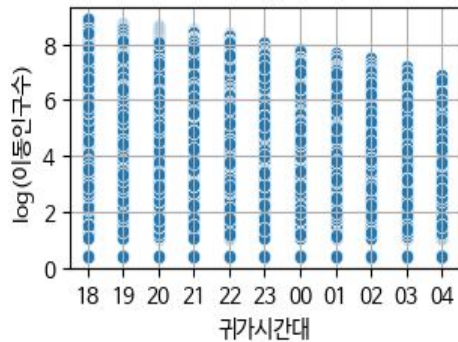
1. 데이터 병합(달 + 도착시간 별로 분리된 파일)
2. 이동유형이 **EH, WH, HH**인 행만 남기기
3. 필요없는 칼럼(평균 이동시간) 삭제
4. 도착시간 -> 귀가시간대 추출
5. **출발 시군구 코드 1**로 시작하는 칼럼만 남기기(1: 서울지역)
6. **연령대** 묶기(10, 15 -> 10대)
7. 데이터프레임에 **귀가시간대** 추가하고 **index**로 지정
8. 결측치인 2월 21시, 22시 데이터 각각 20시, 23시 데이터로 채우기
9. 이동인구 **결측치**(3미만인 값) 처리-> 1,2 중 하나이므로 평균인 1.5로 대체

회귀분석

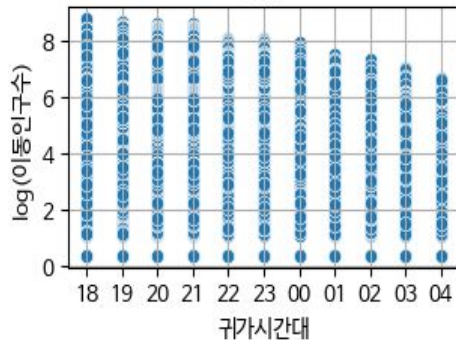
12월



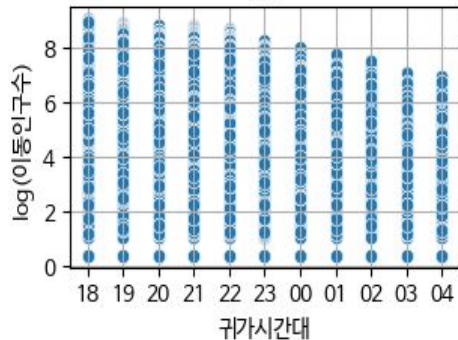
1월



2월



3월



-귀가시간대(x축), 이동인구수(y축)

-12, 1, 2, 3월 모두 귀가시간대가 늦어질수록 이동인구수가 감소하는 경향 땀.

- 가설

“귀가시간대에 따라 이동인구수가 감소하는 기울기를 비교하면, 택시비 인상 전(12월, 1월)에 비해 택시비 인상 후(2월, 3월)에 기울기 감소 정도가 더 가파를 것이다.”

- 분석방법: 단순선형회귀 적합 이후 회귀계수 비교

회귀분석

-단순선형회귀모델: 이동인구수 = $\alpha + \beta$ * 귀가시간대

-적합 결과 (도착 시군구 구분 없음)

$\alpha=0.05$ 하, 회귀계수, 회귀모형 모두 유의

월	회귀계수
12월	-20.379
1월	-19.947
2월	-20.666
3월	-23.757

=> 회귀계수 절대값 비교:

3월 > 2월 > 12월 > 1월 => 택시비 인상 후 > 택시비 인상 전

회귀분석

-적합 결과 (도착 시군구 구분)

$\alpha=0.05$ 하, 회귀계수, 회귀모형 모두 유의

지역구	월	회귀계수
노원구	12월	-18.332
	1월	-17.579
	2월	-18.429
	3월	-21.328
관악구	12월	-19.200
	1월	-18.683
	2월	-19.458
	3월	-22.277

지역구	월	회귀계수
송파구	12월	-25.414
	1월	-25.136
	2월	-26.000
	3월	-29.935
강서구	12월	-20.118
	1월	-19.731
	2월	-20.582
	3월	-23.296
강남구	12월	-18.902
	1월	-18.683
	2월	-19.458
	3월	-22.277

=> 앞 슬라이드와 동일한 결과

회귀분석

결과 해석

- 택시비 인상 전(12월, 1월)에 비해 택시비 인상 후(2월, 3월)에 귀가시간대가 늦어질수록 이동인구수가 감소하는 기울기 정도가 더 커졌다. 즉, 이동인구수가 감소하는 경향이 더 뚜렷해졌다.



원인 분석

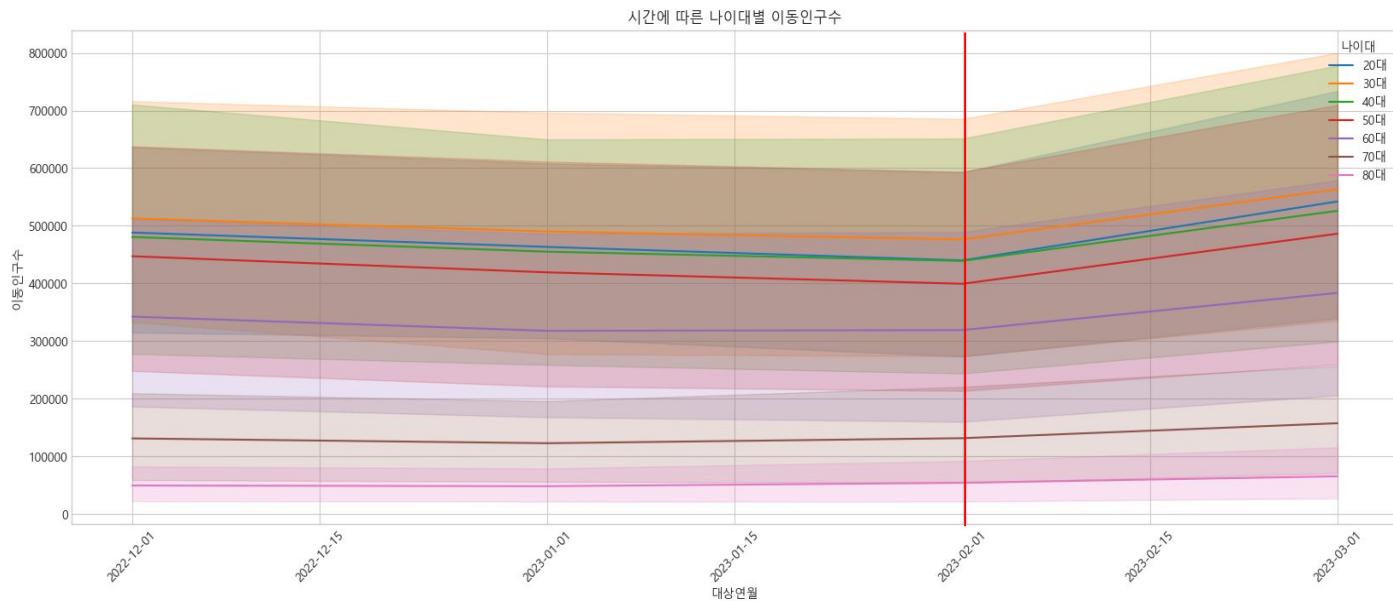
- 1) 심야시간에 택시를 타고 귀가하는 것을 꺼려 18시쯤의 이른 시간에 귀가해서 해당 시간대의 이동인구수가 증가하였을 것이다.
- 2) 지하철, 버스 막차가 끝난 시간대(1시~5시)에 택시를 타고 귀가하지 않고, 밖에서 첫차를 기다리는 사람들이 많아져 해당 시간대의 이동인구수가 감소하였을 것이다.

시계열

- 가설: 나이대(20대~80대)별 시간에 따른 이동인구수가 상이한 양상으로 변화하였을 것이다.
- 분석방법: 나이대(20대~80대)별 **ARIMA** 모델 적합 후,
택시비 인상 시점 **2023-02-01**을 기준으로 전후 이상치의 개수와 **stdresid** 수치 분석

시계열

전체적인 이동인구수 분포: 전 나이대 **2023-02-01** 이후로 증가하는 경향성을 띠



시계열

나이대별 가장 적합한 ARIMA 모델 (p, d, q 값) 도출

Best ARIMA(3, 1, 3) model for Age Group 20대 > AIC: 1180.189941027617

SARIMAX Results

Dep. Variable: 이동인구수 No. Observations: 44
Model: ARIMA(3, 1, 3) Log Likelihood: -583.095
Date: Tue, 15 Aug 2023 AIC: 1180.190
Time: 14:50:19 BIC: 1192.518
Sample: 0 HQIC: 1184.736
Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.7305	0.314	2.327	0.020	0.115	1.346
ar.L2	0.4553	0.347	1.312	0.190	-0.225	1.136
ar.L3	-0.6207	0.275	-2.257	0.024	-1.160	-0.082
ma.L1	-0.9623	1.587	-0.606	0.544	-4.073	2.149
ma.L2	-0.9849	1.550	-0.635	0.525	-4.023	2.053
ma.L3	0.9772	0.211	4.627	0.000	0.563	1.391
sigma2	3.184e+10	1.45e-10	2.2e+20	0.000	3.18e+10	3.18e+10

Ljung-Box (L1) (Q): 0.24 Jarque-Bera (JB): 12.26
Prob(Q): 0.62 Prob(JB): 0.00
Heteroskedasticity (H): 1.09 Skew: 1.18
Prob(H) (two-sided): 0.87 Kurtosis: 4.12

Best ARIMA(3, 1, 3) model for Age Group 30대 > AIC: 1201.0599691134537

SARIMAX Results

Dep. Variable: 이동인구수 No. Observations: 44
Model: ARIMA(3, 1, 3) Log Likelihood: -593.530
Date: Tue, 15 Aug 2023 AIC: 1201.060
Time: 14:50:24 BIC: 1213.388
Sample: 0 HQIC: 1205.606
Covariance Type: opg

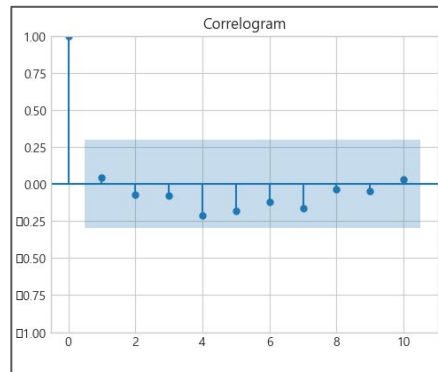
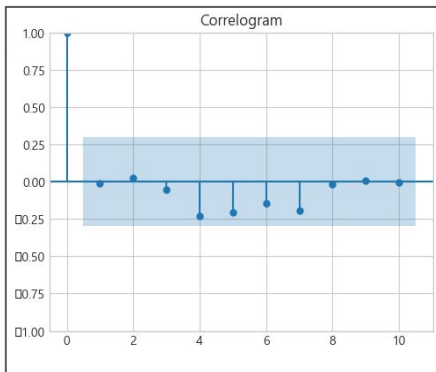
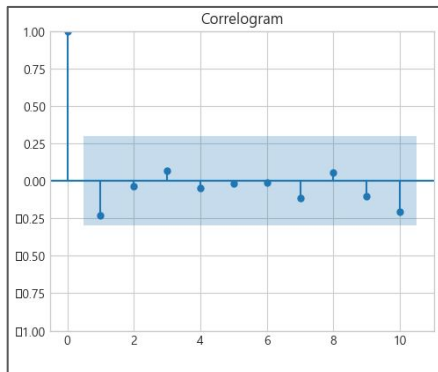
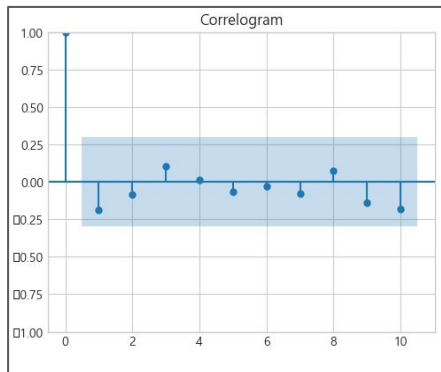
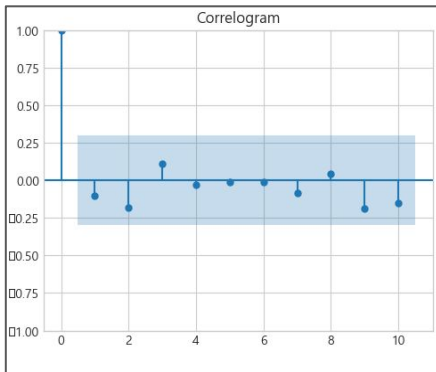
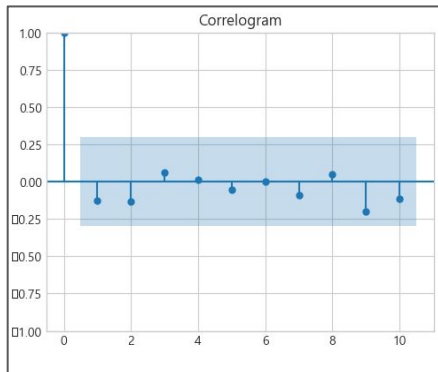
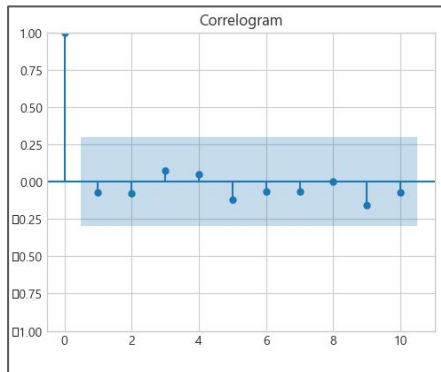
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.6378	0.364	1.754	0.079	-0.075	1.350
ar.L2	0.5274	0.367	1.438	0.150	-0.191	1.246
ar.L3	-0.7006	0.366	-1.916	0.055	-1.417	0.016
ma.L1	-0.9616	1.806	-0.532	0.594	-4.501	2.578
ma.L2	-0.9801	1.845	-0.531	0.595	-4.596	2.636
ma.L3	0.9738	0.234	4.154	0.000	0.514	1.433
sigma2	5.242e+10	1.05e-10	5.01e+20	0.000	5.24e+10	5.24e+10

Ljung-Box (L1) (Q): 0.73 Jarque-Bera (JB): 22.05
Prob(Q): 0.39 Prob(JB): 0.00
Heteroskedasticity (H): 1.12 Skew: 1.43
Prob(H) (two-sided): 0.83 Kurtosis: 5.03

(같은 방식으로
나머지 진행)

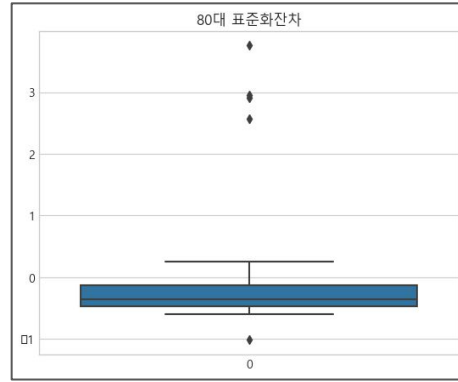
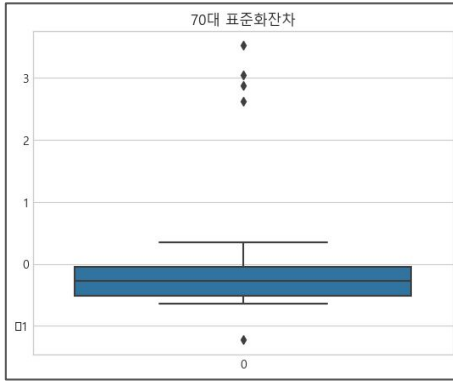
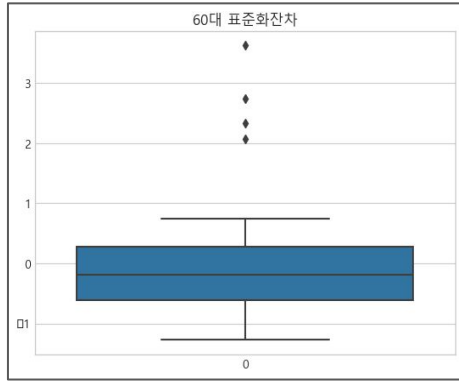
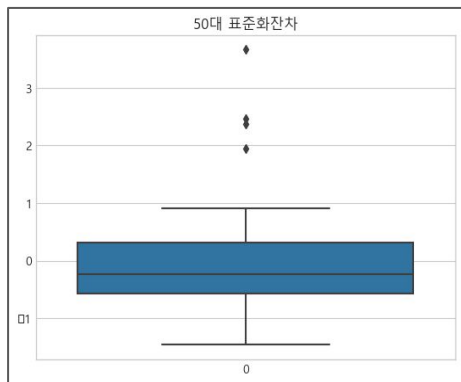
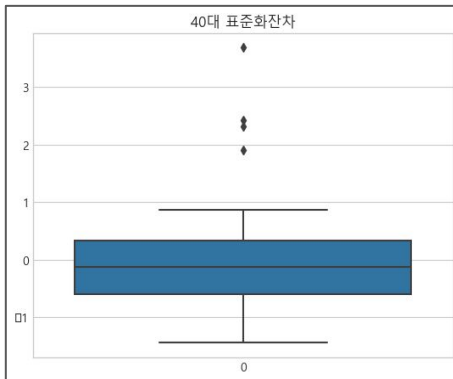
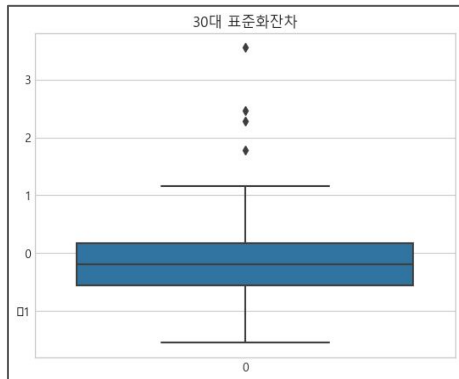
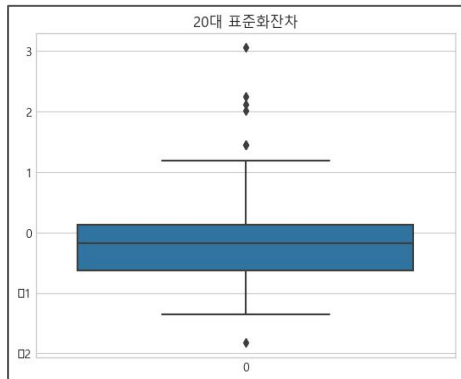
시계열

나이대별 Auto Correlation Function(ACF)



시계열

나이대별 표준화잔차(stdresid)



시계열

분석 1) 20~60대의 04:00:00에서의 이상치

		이동인구수	resid	stdresid
대상연월	귀가시간대			
2022-12-01	00:00:00	445273.25	445273.250000	2.009887
	01:00:00	235751.75	-289031.488571	-1.824839
	18:00:00	722673.23	645611.890536	3.056107
	22:00:00	791476.01	287463.164968	1.185763
2023-01-01	18:00:00	694203.71	465649.397258	2.116297
2023-02-01	03:00:00	92214.71	-157904.047751	-1.140059
	18:00:00	680612.91	338449.428120	1.452026
2023-03-01	01:00:00	242239.73	-135396.314012	-1.022517
	03:00:00	104937.47	-151603.592155	-1.107156
	04:00:00	72790.25	-196854.539478	-1.353912
	18:00:00	821580.56	491707.970093	2.252381
	22:00:00	899710.20	337553.131264	1.447345

		이동인구수	resid	stdresid
대상연월	귀가시간대			
2022-12-01	00:00:00	340778.68	340778.680000	1.162398
	01:00:00	192274.18	-217266.101282	-1.188017
	18:00:00	953925.67	907950.225181	3.551253
	22:00:00	915599.21	605584.229668	2.277725
2023-01-01	18:00:00	915599.21	605584.229668	2.277725
	03:00:00	71180.13	-196187.450988	-1.099236
2023-02-01	04:00:00	51886.17	-218284.139093	-1.192305
	18:00:00	909699.68	488035.332047	1.782624
2023-03-01	03:00:00	81306.45	-185576.484956	-1.054544
	04:00:00	59302.38	-303057.233298	-1.549358
	18:00:00	1055023.59	649588.918095	2.463067
	22:00:00	803224.47	313182.285666	1.046165

		이동인구수	resid	stdresid
대상연월	귀가시간대			
2022-12-01	18:00:00	975328.32	952035.940219	3.685936
2023-01-01	04:00:00	45842.40	-185956.158106	-1.011394
	18:00:00	917604.58	621406.649175	2.321186
2023-02-01	04:00:00	45759.97	-261463.586360	-1.323068
	18:00:00	910222.11	520612.246112	1.905133
	22:00:00	382096.44	-190047.327599	-1.028281
2023-03-01	03:00:00	64240.35	-212492.206642	-1.120927
	04:00:00	52576.83	-290569.820987	-1.443211
	18:00:00	1065981.11	646938.273650	2.426573

		이동인구수	resid	stdresid
대상연월	귀가시간대			
2022-12-01	18:00:00	925586.73	885549.274102	3.663503
2023-01-01	18:00:00	886525.03	592624.828221	2.378969
2023-02-01	03:00:00	48838.67	-183650.271722	-1.025160
	04:00:00	46240.43	-212062.145892	-1.149752
	18:00:00	876127.82	492420.382375	1.939551
2023-03-01	03:00:00	55704.01	-196003.701912	-1.079332
	04:00:00	52643.79	-282472.207834	-1.458515
	18:00:00	1007713.08	612887.173899	2.467823

		이동인구수	resid	stdresid
대상연월	귀가시간대			
2022-12-01	18:00:00	816794.12	767493.954830	3.626437
2023-01-01	18:00:00	777933.91	508266.413521	2.327320
2023-02-01	03:00:00	42481.21	-172477.205324	-1.084221
	04:00:00	54177.94	-193001.946215	-1.187080
	18:00:00	800032.07	456052.614835	2.065651
	22:00:00	210973.90	-163203.531762	-1.037746
2023-03-01	03:00:00	48850.11	-160073.679809	-1.022061
	04:00:00	63600.47	-210005.894858	-1.272295
	18:00:00	926646.29	590812.111653	2.740997

-> 택시비 인상(2023-02-01) 이전:
04:00:00에서의 이상치가 없거나 stdresid 수치의
절댓값이 작음.



-> 택시비 인상(2023-02-01) 이후:
모든 나이대에서 04:00:00가 새로 추가되거나
stdresid 수치가 더 작아짐.

시계열

분석2) 70/80대의 18:00:00에서의 이상치

대상연월 귀가시간대		이동인구수		resid	stdresid
2022-12-01	18:00:00	400157.29	377867.484379	3.043125	
2023-01-01	18:00:00	380564.87	328185.200950	2.623824	
2023-02-01	18:00:00	416856.31	357528.286295	2.871469	
	22:00:00	60952.92	-127887.076332	-1.225267	
2023-03-01	18:00:00	500702.94	434145.784502	3.518094	

대상연월 귀가시간대		이동인구수		resid	stdresid
2022-12-01	18:00:00	172291.01	164613.381166	2.904285	
2023-01-01	18:00:00	166114.12	146155.378431	2.562821	
2023-02-01	18:00:00	190010.27	167124.136455	2.950733	
	22:00:00	23167.51	-47135.097745	-1.012953	
2023-03-01	18:00:00	236976.57	210612.548223	3.755246	

-> 70/80대는 수입원이 적어 인상된 택시비에 상대적으로 더 큰 부담을 느낌.
이로 인해 2023-03 18:00:00의 stdresid 수치가 다른 시간대에 비해 특히 높게 관측됨.

시계열

분석3) 20/30대의 22:00:00에서의 이상치

대상연월 귀가시간대		이동인구수	resid	stdresid
2022-12-01	00:00:00	445273.25	445273.250000	2.009887
	01:00:00	235751.75	-289031.488571	-1.824839
	18:00:00	722673.23	645611.890536	3.056107
	22:00:00	791476.01	287463.164968	1.185763
2023-01-01	18:00:00	694203.71	465649.397258	2.116297
2023-02-01	03:00:00	92214.71	-157904.047751	-1.140059
	18:00:00	680812.91	338449.428120	1.452026
2023-03-01	01:00:00	242239.73	-135396.314012	-1.022517
	03:00:00	104937.47	-151603.592155	-1.107156
	04:00:00	72790.25	-198854.539478	-1.353912
	18:00:00	821580.56	491707.970093	2.252381
	22:00:00	899710.20	337553.131264	1.447345

대상연월 귀가시간대		이동인구수	resid	stdresid
2022-12-01	00:00:00	340778.68	340778.680000	1.162398
	01:00:00	192274.18	-217266.101282	-1.188017
	18:00:00	953925.67	907950.225181	3.551253
2023-01-01	18:00:00	915599.21	605584.229668	2.277725
2023-02-01	03:00:00	71180.13	-196187.450988	-1.099236
	04:00:00	51886.17	-218284.139093	-1.192305
	18:00:00	909699.68	488035.332047	1.782624
2023-03-01	03:00:00	81306.45	-185576.484956	-1.054544
	04:00:00	59302.38	-303057.233298	-1.549358
	18:00:00	1055023.59	649588.918095	2.463067
	22:00:00	803224.47	313182.285666	1.046165

-> 20/30대는 택시 활동에 특히 부담을 많이 느낌.

이에, 아직 대중교통으로 귀가가 가능한 22:00:00에 이동 인구가 많아진 것으로 해석할 수 있음.

분산분석

회귀분석과 시계열 분석 결과를 바탕으로,
각 시간대를 고정하였을 때에도 월별로 이동인구수 간 차이가
유의한지를 살펴볼 필요가 있다고 생각하였다.

- 가설

“각 시간대를 고정했을 때, 각 월별 이동인구수 평균 간에는
유의한 차이가 존재할 것이다.”

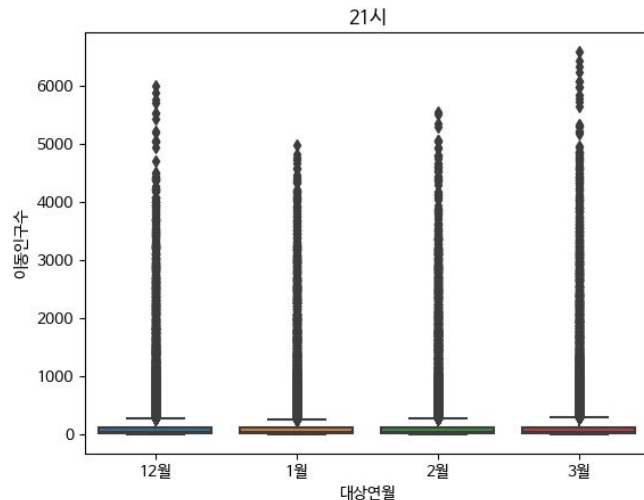
- 분산분석

H0: 이동인구수 평균 간 차이는 유의하지 않다.

H1: 이동인구수 평균 간 차이는 유의하다.

- 분석방법

일원배치 분산분석 적용 -> 귀무가설 기각 시 Tukey HSD 사후검정



분산 분석

결과 분석 (지역 구분 없음 / 일원배치분산분석, 사후검정 모두 $\alpha=0.05$ 하에서 유의한 결과만 작성하였음)

귀가시간대	이동인원수 비교
18시	3월 > 12월, 1월, 2월
19시	3월 > 12월, 1월, 2월
20시	3월 > 12월, 1월, 2월
21시	3월 > 12월, 1월 2월 > 1월
23시	12월 > 1월, 2월 3월 > 1월, 2월
0시	12월 > 1월, 2월 3월 > 1월, 2월

	df	sum_sq	mean_sq	F	PR(>F)
대상연월	1.0	1.329030e+05	132903.029635	3.988646	0.045811
Residual	76212.0	2.539410e+09	33320.337826	NaN	NaN

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-6.2998	0.004	-11.0866	-1.5131	True
0	2	-4.8323	0.0487	-9.6459	-0.0188	True
0	3	3.4307	0.2523	-1.3465	8.208	False
1	2	1.4675	0.8634	-3.3635	6.2985	False
1	3	9.7306	0.0	4.9357	14.5254	True
2	3	8.2631	0.0001	3.4415	13.0847	True

분산 분석

결과 분석 (지역 구분 있음 / 일원배치분산분석, 사후검정 모두 $\alpha=0.05$ 하에서 유의한 결과만 작성하였음)

노원구

18시	3월 > 12월, 1월, 2월
19시	3월 > 12월, 1월, 2월
20시	3월 > 12월, 1월, 2월
21시	3월 > 1월

강서구

18시	3월 > 1월
19시	3월 > 12월, 1월
20시	3월 > 12월, 1월
21시	x

관악구

18시	3월 > 12월, 1월, 2월
19시	3월 > 12월, 1월, 2월
20시	3월 > 12월, 1월, 2월
21시	3월 > 1월

분산 분석

결과 분석 (지역 구분 있음 / 일원배치분산분석, 사후검정 모두 $\alpha=0.05$ 하에서 유의한 결과만 작성하였음)

강남구

18시	3월 > 1월, 2월
19시	X
20시	3월 > 1월
21시	3월 > 1월

송파구

18시	3월 > 1월
19시	3월 > 12월, 1월, 2월
20시	3월 > 1월
21시	3월 > 1월

분산 분석

결과 해석

- 18시 ~ 22시에서 3월이 12월, 1월보다 이동인구수가 유의하게 더 많은 경우가 대부분.
- 2월의 이동인구수가 12월, 1월보다 유의하게 낮다고 나온 결과가 아예 없음.
- 1시 이후에는 유의한 결과 없었음.

이전 분석들과의 종합적인 해석

- 18 ~ 22시(지하철, 버스 막차 끊기기 전)에 귀가하는 사람들이 택시비 인상 이후에 비교적 증가했다.
- 3월이 12, 1, 2월보다 절대적인 인구 이동 수가 많을 것이라는 예상과는 달리, 자정 이후에는 12월, 1월에 비해 인구이동이 증가하지 않았던 결과로 보아, 심야 시간에 택시를 타고 귀가하는 사람들이 줄어들었음을 예측할 수 있다.

2. 택시비 인상이 지하철 이용에 미치는 영향

사용 데이터

1. 서울시 지하철호선별 역별 승하차 인원 정보
2. 서울교통공사 역별 일별 시간대별 승하차 인원 정보
3. 서울교통공사 9호선 2~3단계 역별 일별 시간대별 승하차인원

지역구	1번 데이터셋	2, 3번 데이터셋
송파구	'잠실나루', '잠실(송파구청)', '잠실새내', '종합운동장', '가락시장', '경찰병원', '오금', '올림픽공원 (한국체대)', '방이', '개롱', '거여', '마천', '몽촌토성 (평화의문)', '석촌', '송파', '문정', '장지', '북정', '삼전', '석촌고분', '송파나루', '한성백제'	'잠실나루', '잠실(송파구청)', '잠실새내', '종합운동장', '가락시장', '경찰병원', '오금', '올림픽공원 (한국체대)', '방이', '개롱', '거여', '마천', '몽촌토성 (평화의문)', '석촌', '송파', '문정', '장지', '북정', '삼전', '석촌고분', '송파나루', '한성백제'
강서구	'방화', '개화산', '김포공항', '송정', '마곡', '발산', '우장산', '화곡', '까치산', '개화', '공향시장', '신방화', '마곡나루 (서울식물원)', '양천향교', '가양', '증미', '등촌'	'방화', '개화산', '김포공항', '송정', '마곡', '발산', '우장산', '화곡', '까치산'
강남구	'삼성(무역센터)', '선릉', '역삼', '강남', '압구정', '신사', '매봉', '도곡', '대치', '학여울', '대청', '일원', '수서', '청담', '강남구청', '학동', '논현', '언주', '선정릉', '삼성중앙', '봉은사', '신논현', '압구정로데오', '한티', '구룡', '개포동', '대모산입구'	'삼성(무역센터)', '선릉', '역삼', '강남', '압구정', '신사', '매봉', '도곡', '대치', '학여울', '대청', '일원', '수서', '청담', '강남구청', '학동', '논현', '언주', '선정릉', '삼성중앙', '봉은사'
노원구	'당고개', '상계', '노원', '석계', '태릉입구', '화랑대 (서울여대입구)', '수락산', '마들', '중계', '하계', '공릉 (서울과학기술대)', '월계', '광운대'	'당고개', '상계', '노원', '석계', '태릉입구', '화랑대 (서울여대입구)', '수락산', '마들', '중계', '하계', '공릉 (서울과학기술대)'
관악구	'낙성대 (강감찬)', '서울대입구 (관악구청)', '봉천', '신림', '당곡', '서원', '서울대벤처타운', '관악산 (서울대)'	'낙성대 (강감찬)', '서울대입구 (관악구청)', '봉천', '신림'

데이터 전처리

6가지 경우의 데이터셋 생성

경우	기간	시간대	지하철역
1	2022.12.01 ~ 2023.03.31	모든 시간대	지역구 내 모든 역
2	2023.01.17 ~ 2023.02.15	모든 시간대	지역구 내 모든 역
3	2022.12.01 ~ 2023.03.31	귀가시간대 (18시 ~ 6시)	지역구 내 1~8호선, 9호선 2~3단계 역
4	2023.01.17 ~ 2023.02.15	귀가시간대 (18시 ~ 6시)	지역구 내 1~8호선, 9호선 2~3단계 역
5	2022.12.01 ~ 2023.03.31	심야시간대 (22시 ~ 3시)	지역구 내 1~8호선, 9호선 2~3단계 역
6	2023.01.17 ~ 2023.02.15	심야시간대 (22시 ~ 3시)	지역구 내 1~8호선, 9호선 2~3단계 역

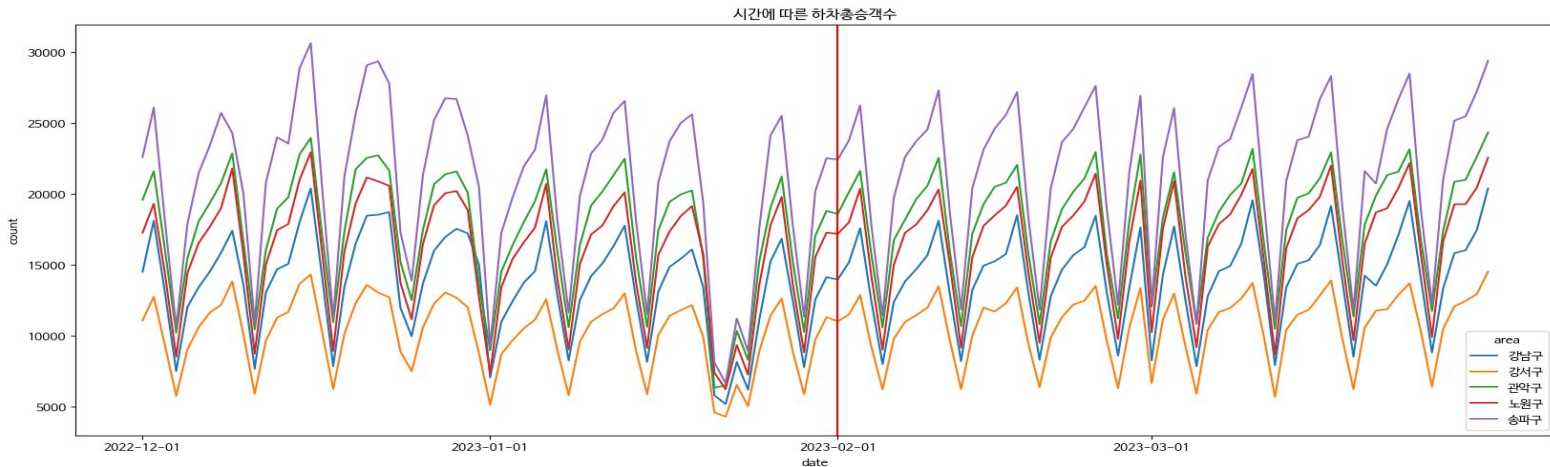
시계열 분석

택시비 인상 이후, 지하철로 귀가하는 사람들이 많아져 지하철 하차 인원수가 증가할 것이라고 생각했다.

- 가설

“시계열 모델 적합 후 표준화잔차를 비교했을 때, 택시비 인상 후보다 **택시비 인상 전에 표준화잔차가 낮게 나타날 것이다.**”

- 분석방법: 6가지 데이터셋에 대해, 각 지역구별로 **ARIMA** 모델 적합 후 표준화잔차 비교

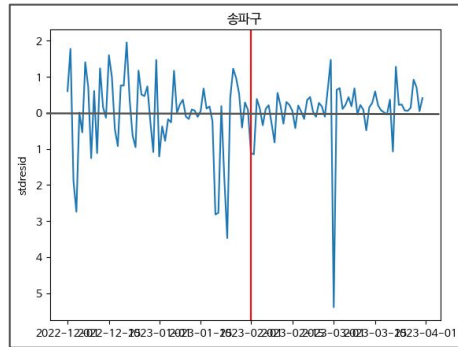
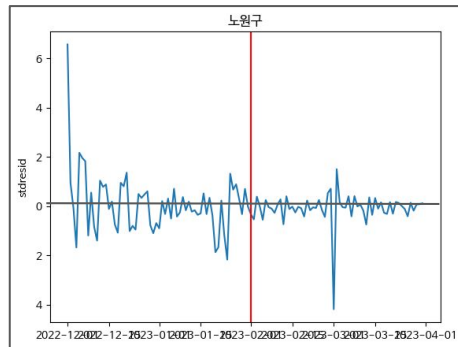
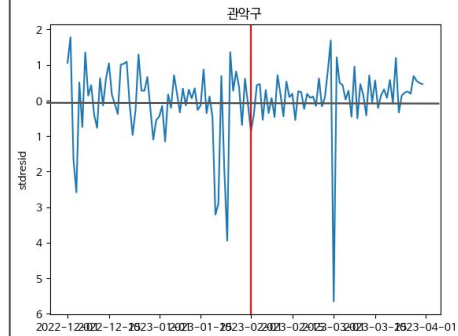
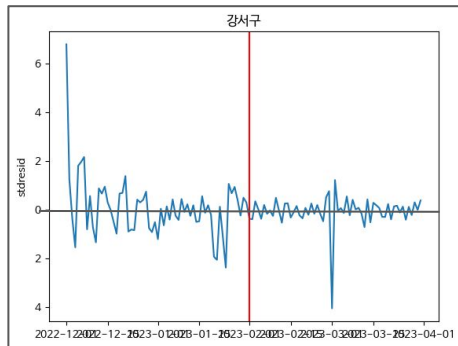
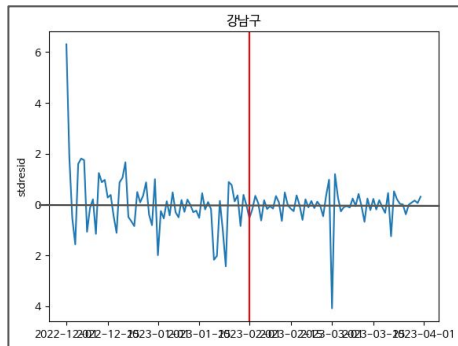


시계열 분석

결과: 5번 데이터의 모든 지역구에서,
택시비 인상 이후보다 택시비 인상 이전에
표준화잔차가 더 작게 나타났다.



결과해석: 심야시간대(22시 ~ 3시)에
지하철 이용 인원수가 택시비 인상 전보다
택시비 인상 후에 전체적으로 더 많아졌다.
귀가 여부를 구분할 수는 없으나,
12 ~ 3월이 같은 조건 하에서 비교되었음을
고려할 때, 이는 심야시간대에 지하철을
타고 귀가하는 인원이 늘었음을 시사한다.



3. 택시비 인상에 따른 대중의 반응 분석

유튜브 댓글 크롤링

* 비슷한 주제에 쏠리는 것을 방지하기 위해 주제를 선정하고, 주제별로 댓글수를 맞춰서 영상 선정



서울 택시 기본요금 인상...밤 11시 넘으면 '6700원' | 뉴스A

조회수 1.6만회 · 6개월 전

채널A 뉴스

서울 택시 기본요금 인상...밤 11시 넘으면 '6700원' 내일부터 심야 시간 서울 택시 기본요금이 6700원까지 오릅니다. 주간 기본 요금 ...



"와..6분 만에 1만원 넘네" "밤에 딱북이도 안 팔려" [뉴스.zip/MBC뉴스]

조회수 198만회 · 6개월 전

MBCNEWS

00:00 택시요금 인상 후폭풍...기사는 물론 자영업자도 울상 (2023.02.02/뉴스데스크/MBC) 02:30 택시요금 1천원 인상 첫날...예상 ...

택시요금 인상 후폭풍...기사는 물론 자영업자도 울상...

펼쳐 3

▲ 택시 기본요금 ▼ 귀가 시간



[밀착카메라] 택시비 인상의 나비효과...달라진 '밤 11시 풍경' / JTBC 뉴스룸

조회수 102만회 · 6개월 전

JTBC News

서울 택시 요금이 오른 지 2주가 지났습니다. 그 사이에, 택시기사들 뿐 아니라 시민들의 일상도 달라졌다고 하는데요. 기사도, 승객도 ...



[발품뉴스] 불금에도 빈 택시 수두룩...기사·승객 모두 불만에 대안은? / JTBC 뉴스룸

조회수 99만회 · 6개월 전

JTBC News

이슈가 있는 현장을 직접 찾아가는 윤정식 기자의 발품뉴스 시간입니다. 조금 전 영상에서 지난밤 강남역 상황을 보셨습니다. 정부가 ...



[자막뉴스] 급한 불 끄려다 '직격탄'...도심 한복판 결국 이렇게 변했다 / JTBC News

조회수 547만회 · 6개월 전

JTBC News

이슈가 있는 현장을 직접 찾아가는 윤정식 기자의 발품뉴스 시간입니다. 조금 전 영상에서 지난밤 강남역 상황을 보셨습니다. 정부가 ...

▲ 대중교통 이용 ▼ 심야 할증



'빈 차' 택시 늘었지만, 할증 요금에 부담 '썩'... '겁나서 못 탔다' | 뉴스A

조회수 6.6만회 · 6개월 전

채널A 뉴스

'빈 차' 택시 늘었지만, 할증 요금에 부담 '썩'... '겁나서 못 탔다' 난방비 폭탄에 이어 택시비도 폭탄이라는 말이 나옵니다. 서울 택시가 ...



[자막뉴스] 그렇게 안잡히더니 12시 넘자 우르르 쏟아진 택시, '심야할증' 노리고 양떼

운행? / KBS 2022.10.13.

조회수 124만회 · 10개월 전

KBS News

자정이 되자 2분 전, 대전역 앞 택시 승강장입니다. 택시를 기다리는 줄이 백미터 넘게 늘어 있습니다. 그런데, 멀리 택시 승강장 ...

전처리

	Comment
0	버스 탈려고 해도 문제 예전에 30분 걸리던 거리가 요즘은 1시간 이상 걸림 중앙 ...
1	분명한 사실은 서울시 시의원과 구의원들 택시 요금 인상 빌미로 의원들 운영비 의원 ...
2	택시는 대중교통 아니고 고급이용 수단입니다 택시비 없으면 대중교통이용하세요 물가에 ...
3	아직 우리에게겐 심야버스가 있다
4	이제 점점 심야 문화 자체가 없어질꺼같다
...	...

1. 특수문자 및 숫자 제거
2. 불용어 제거 (불용어 사전 및 데이터에 맞는 불용어 '택시' 제거)
3. 데이터 치환 (같은 의미로 사용된 여러 데이터를 특정 단어로 통일화)
4. 한글자 키워드 제거 (대부분 불용어에 해당)

*한국어 형태소 분석을 위해 **Konlpy/mecab** 사용

100



유튜브 댓글의 전체적인 내용 파악

1. 택시 대체: 대중교통, 버스, 지하철, 우버
2. 요금 관련: 할증, 가격, 만원, 심야
3. 택시 서비스 관련: 기사, 서비스, 카카오
4. 경제적 어려움 관련: 월급, 시민

-> 택시비 인상에 대해 대중의 반응을 구체적으로 살펴보기 위해 연관성이 높은 토픽별로 묶어보자

LDA

```
num_topics = 5
chunksize = 2000
passes = 20
iterations = 400
eval_every = None

temp = dictionary[0]
id2word = dictionary.id2token

model = LdaModel(
    corpus=corpus,
    id2word=id2word,
    chunksize=chunksize,
    alpha='auto',
    eta='auto',
    iterations=iterations,
    num_topics=num_topics,
    passes=passes,
    eval_every=eval_every
)
```

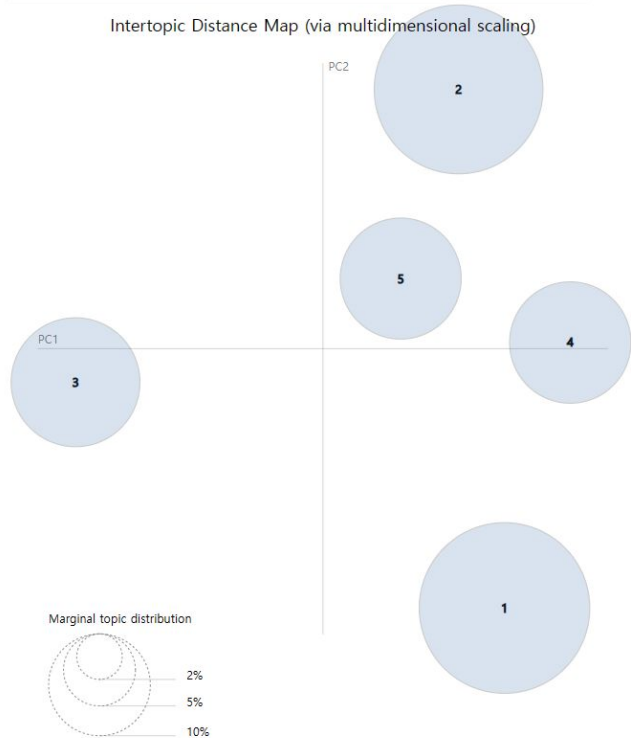
파라미터 조정

- 📌 num_topics: 생성될 토픽의 개수 = 5
- 📌 chunksize: 한번의 트레이닝에 처리될 문서의 개수 = 2000
- 📌 passes: 전체 코퍼스 트레이닝 횟수 = 20
- 📌 iterations: 문서 당 반복 횟수 = 400

시각화 및 결과 해석

Selected Topic:

Intertopic Distance Map (via multidimensional scaling)

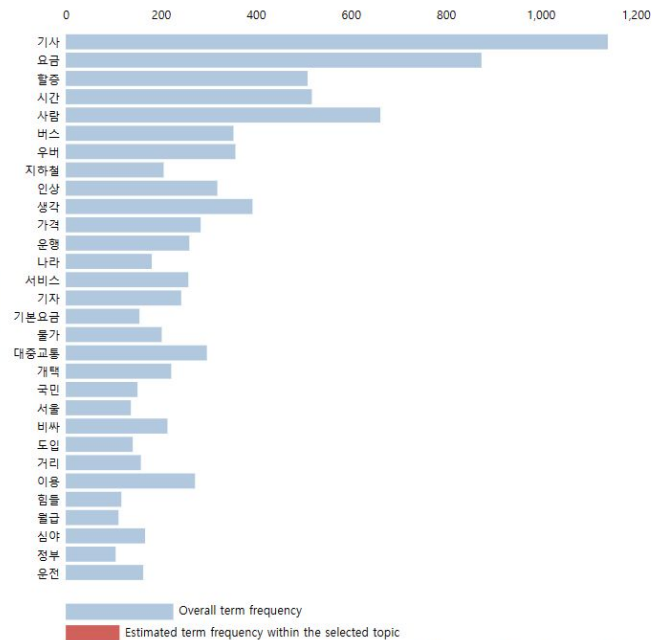


Slide to adjust relevance metric:(2)

$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

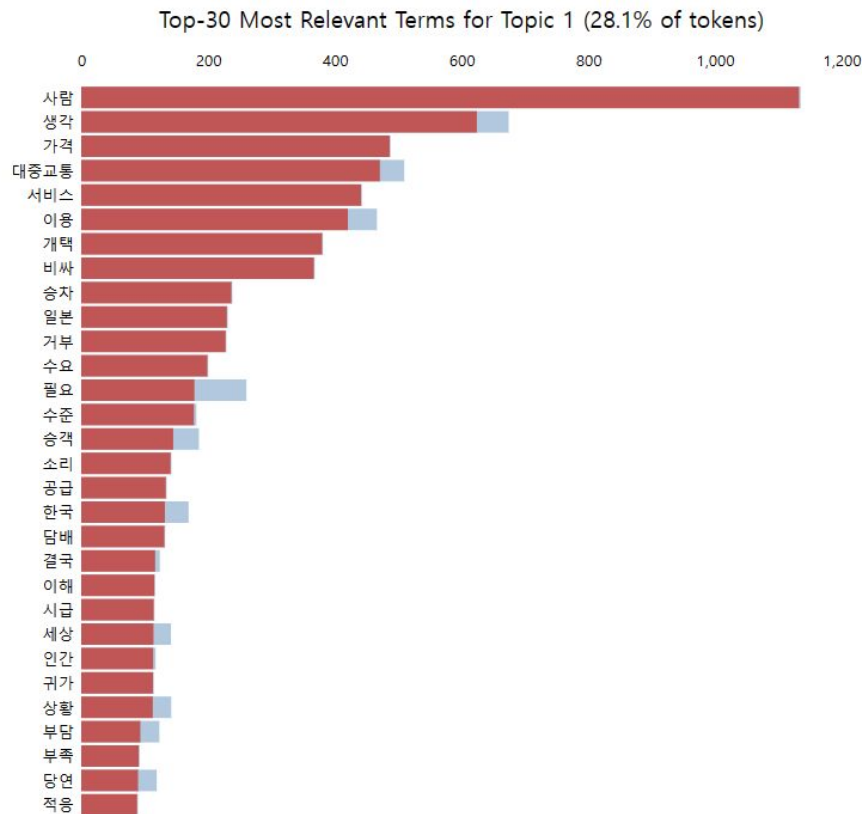
Top-30 Most Salient Terms¹



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t ; see Chuang et. al (2012)

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

시각화 및 결과 해석



Topic 1 :

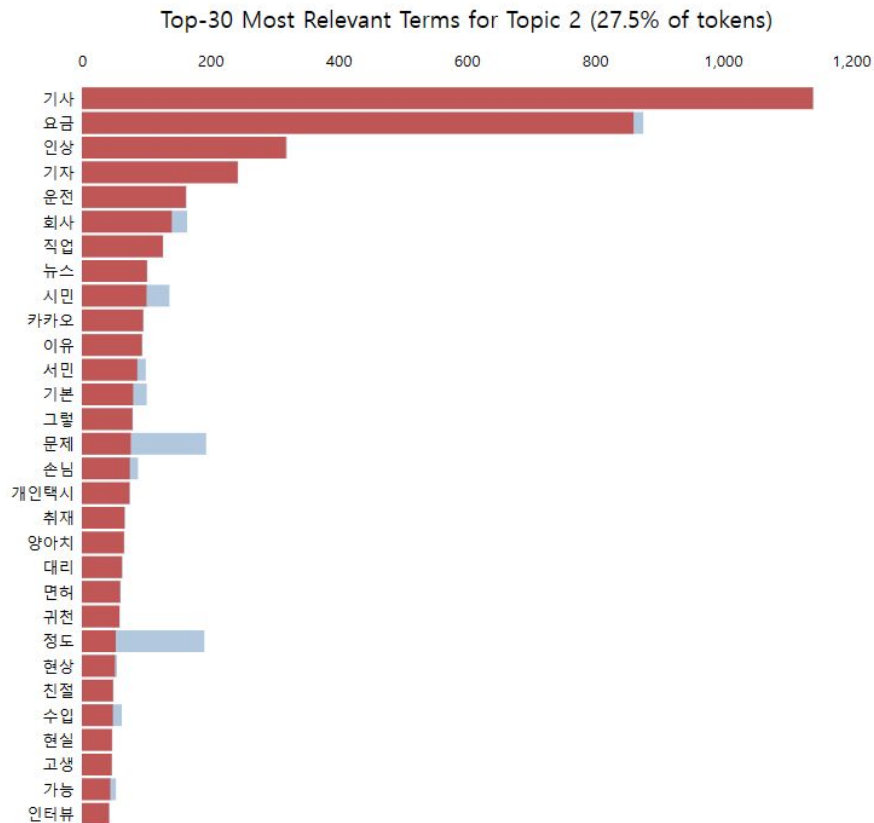
택시에 관한 부정적인 이미지

- 토픽1에는 '가격', '서비스', '개택', '승차', '거부', '승객' '부담' 등의 단어가 빈출되었다.



- 택시의 가격과 서비스에 관한 단어, 그리고 택시에 대한 불만을 나타내는 단어가 빈출된 것을 보아 대중들이 택시에 대해 부정적인 인상을 가지고 있다는 것을 알 수 있다.

시각화 및 결과 해석



Topic 2 :

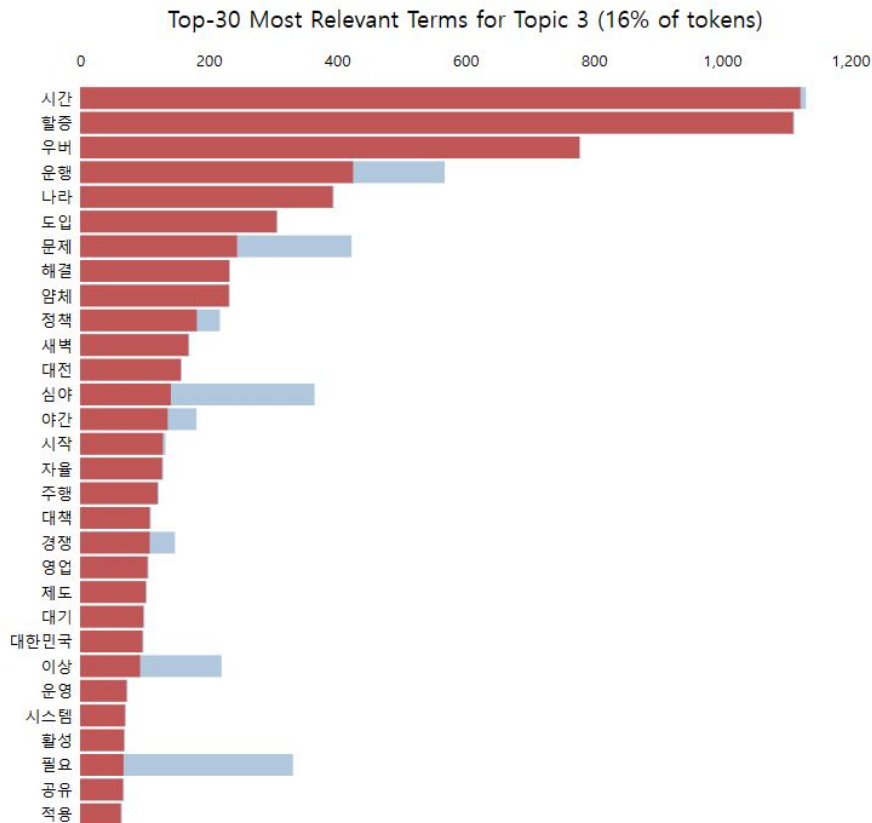
택시 서비스에 관한 불만

- 토픽2에는 '기사', '회사', '카카오', '문제', '양아치' 등의 단어가 빈출되었다.



- 택시 서비스 및 택시 회사에 관한 단어가 빈출된 것을 보아 택시 서비스에 관한 불만을 가진 반응이 많았다는 것을 알 수 있다.

시각화 및 결과 해석



Topic 3 :

택시 할증률 인상

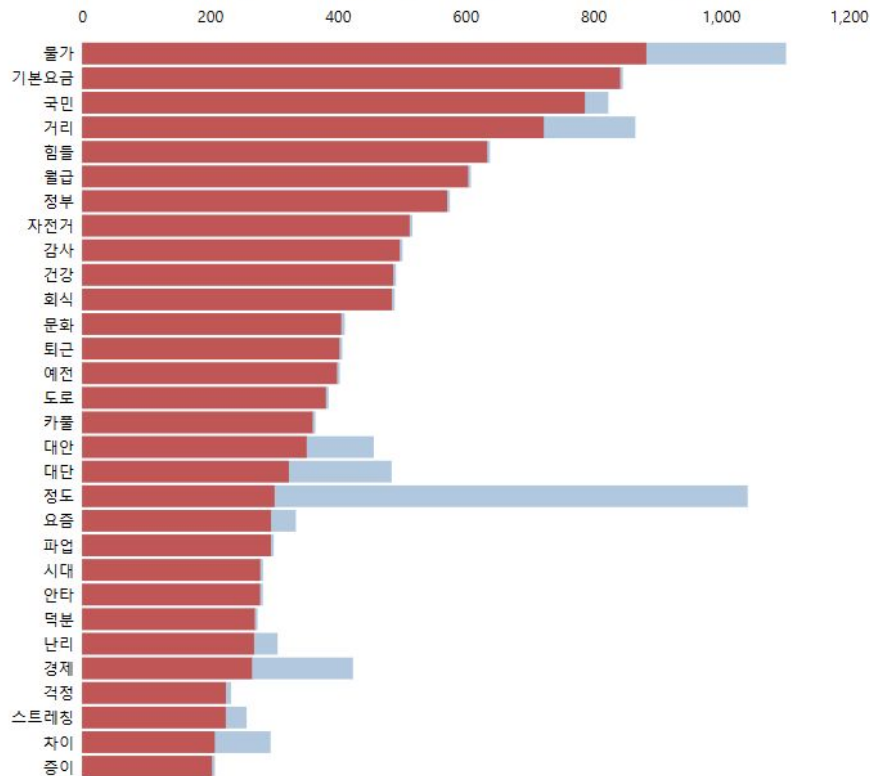
- 토픽3에는 '시간', '할증', '새벽', '심야', '야간' 등의 단어가 빈출되었다.



- 할증에 관한 단어가 빈출된 것으로 보아, 택시비 요금 인상 중에서도 특히 심야 할증이 증가한 것에 민감하게 반응함을 알 수 있다.

시각화 및 결과 해석

Top-30 Most Relevant Terms for Topic 4 (14.2% of tokens)



Topic 4 :

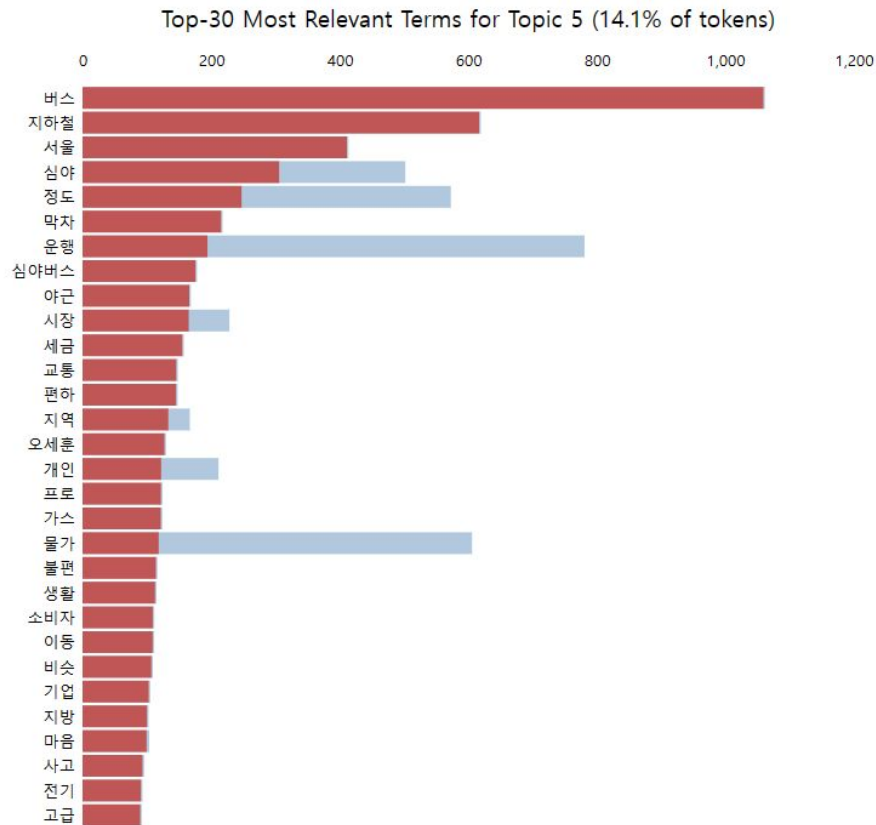
택시 기본요금 인상 및 경제적 어려움

- 토픽4에는 '물가', '기본요금', '월급', '국민', '경제' 등의 단어가 빈출되었다.



- 기본요금 인상과 물가 및 월급에 관한 단어가 빈출된 것을 보아, 기본요금 인상에 따른 경제적 어려움에 대한 불만의 반응을 보이는 것을 알 수 있다.

시각화 및 결과 해석



Topic 5 :

대중교통 이용

- 토픽 5에는 ‘버스’, ‘지하철’, ‘막차’, ‘심야버스’, ‘교통’ 등의 단어가 빈출되었다.



- 택시비 인상에 관한 반응으로 버스나 지하철, 특히 막차나 심야버스에 관한 언급이 많은 것을 알 수 있다.

비즈니스적 인사이트 제안

제안1) “택시요금 인상 권한을 갖는 지방자치단체 및 지방정부”

유튜브 뉴스의 크롤링 결과를 통해 알게 된 사실은,
대다수의 사람들이 택시 요금 인상에 부정적이지만, 그 중에서도 특히 '심야 할증'과 같은 키워드에
대해 더욱 민감하게 반응한다는 것이다.

-> 추후 택시요금 인상 관련 정책 수립 시, '심야 할증'부분의 인상을 최소화하고,
대중이 덜 부정적으로 받아들이는 **운행요금 (거리당/시간당 요금)**을 적절하게 인상 한다면,
정책의 목표를 달성하는 동시에 대중의 수용도도 높일 수 있을 것으로 보인다.

비즈니스적 인사이트 제안

제안2) “택시 회사(ex 카카오T, 아이엠택시 등)의 고객 유치 방안 제시 ”

생활 이동 데이터 분석 결과, 택시비 인상 후 각 연령층의 주요 이동 시간대가 상이하게 나타났으며, 막차가 끊기기 전에 귀가해 택시를 이용하지 않는 경향을 확인할 수 있었다.

-> 택시 회사들은 **특정 연령층의 활발한 이동시간대**에 적용 가능한 **할인쿠폰이나 적립 포인트**를 제공할 수 있다. 이러한 **연령별 타겟팅**으로 택시 회사는 대중들의 택시 거부 반응을 완화하는 동시에 고객 만족도를 높일 것으로 보인다.

ex>

나이대	제공 혜택 (예시)
20/30대	22:00-23:00에만 사용가능한 20% 할인쿠폰
70/80대	18:00-19:00에만 사용가능한 20% 할인쿠폰 + ‘대신 불러주기’ 기능
전 연령층	자정부터 사용가능한 10% 할인쿠폰, 앱 서비스 이용시 5% 추가 포인트 적립 혜택 제공

한계점

1. 지하철 하차인원수 시계열 분석: 귀가시간 및 심야시간대 분석에서, 데이터가 없어 지역구에 있는 모든 지하철역에 대해 분석하지는 못했음
2. 지하철 하차인원수 시계열 분석: 분석에 사용한 기간이 짧아 하차인원수가 특히 많거나 적은 날의 영향을 완전히 제거하지는 못했음
3. 토픽모델링: 의 최적인 파라미터를 찾는 추가적인 분석이 필요해보임
4. 토픽 모델링: 결과를 해석 할 때 분석가의 도메인 지식과 키워드에 대한 이해에 따라 주제 라벨링과 해석이 달라짐

감사합니다!