

MIXTURE FINAL PROJECT

User Response Data to Netflix New Ads Plan with LCA

Narin Park | annie9166@g.skku.edu

29, November 2022

Statistics , SungKyunKwan University

Contents

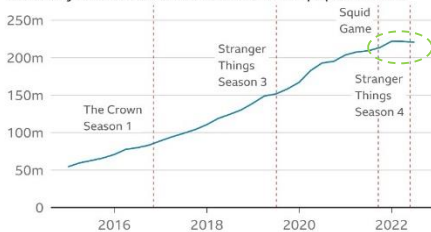
1. Introduction
2. Data
3. Methodology
4. Latent Class Analysis
5. Conclusion

INTRODUCTION

Motivation

Netflix subscribers 2015 to 2022

Quarterly subscribers and release dates of popular series



Source: Netflix, July 2022

B B C

넷플릭스

\$285.54

↑ 52.84% +98.72 5년

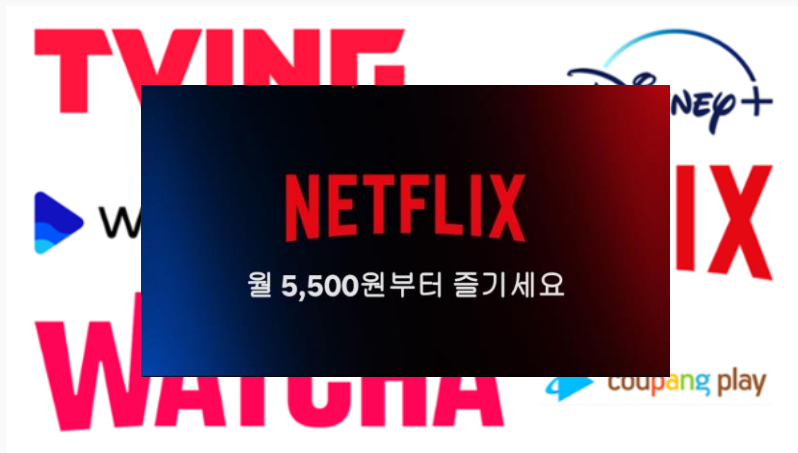
시간외: \$285.04 (↓ 0.18%) -0.50

마감일: 11월 25일, 오후 4시 53분 35초 UTC-5 · USD · NASDAQ · 현액초장

1일 5일 1개월 6개월 YTD 1년 5년 최대



Motivation



Purpose

- 넷플릭스의 광고 요금제 도입에 대한 구독자 반응 별 잠재 변수 탐지
- Use LCA(Latent Class Analysis) model

DATA

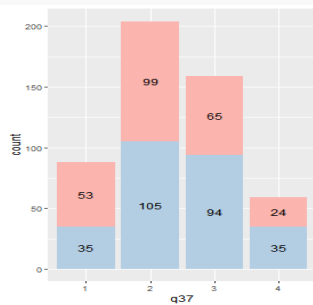
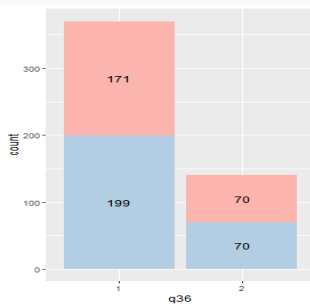
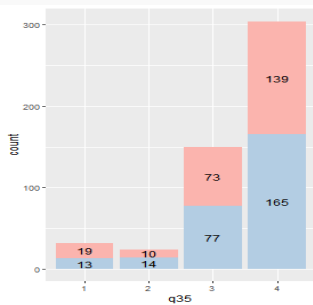
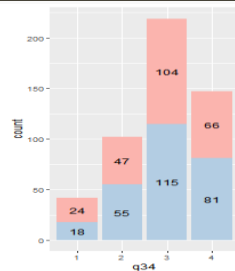
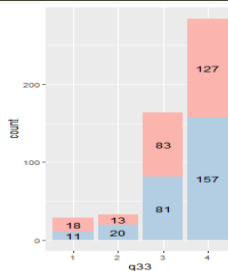
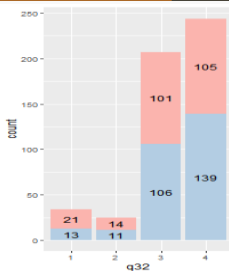
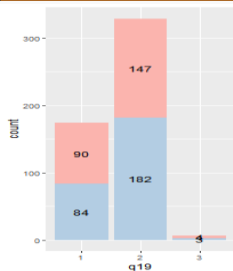
'광고요금제'도입을 앞둔 넷플릭스에 대한 인식 및 이용조사

- 한국언론진흥재단 미디어연구센터에서 20-50대 1천명을 대상으로 온라인 설문조사
- 성별, 연령대 및 거주지역을 기준으로 할당해 모집 (2022.09.16 ~ 20)
- 13565 이메일 랜덤 발송-> 1366 응답 -> 클렌징 후 1000명
- N : 1000 , P: 42(질문지 문항)

DATA :: INTRODUCTION TO DATA

번호명	질문내용	class	#	결측치	번호명	질문내용	class	#	결측치
Q1	성별	범주형	2	0	Q22	OTT서비스 중 넷플릭스만 사용하는 이유 (중복문항)	범주형	8	818
Q2	연령	연속형		0	Q23	문1-10 로직 1번	범주형	3	0
Q3	연령	범주형	4	0	Q24	문1-10 로직 2번	범주형	3	0
Q4	지역	범주형	17	0	Q25	문1-10	범주형	3	0
Q5	넷플릭스	범주형	3	0	Q26	OTT서비스를 하나만 이용하는 이유 (중복문항)	범주형	8	455
Q6	티빙	범주형	3	0	Q27	OTT서비스를 2개이상 이용하는 이유 (중복문항)	범주형	8	817
Q7	웨이브	범주형	3	0	Q28	광고요금제 알았는지 여부	범주형	2	0
Q8	쿠팡플레이	범주형	3	0	Q29	광고요금제 출시에 대한 입장	범주형	4	0
Q9	디즈니 플러스	범주형	3	0	Q30	광고요금제 도입으로 인한 넷플릭스 전망	범주형	5	0
Q10	애플티비	범주형	3	0	Q31	광고요금제 출시 하면 서비스 이용 의향	범주형	3	1.이용X 278 2.광고의양 보고 결정 689 3.무조건 이용 33
Q11	왓차	범주형	3	0	Q32	광고요금제 이용시 광고개수 및 시간	범주형	4	311(이용X+ 무조건 이용)
Q12	시즌	범주형	3	0	Q33	광고요금제 중간 광고 유무	범주형	4	311(이용X+ 무조건 이용)
Q13	OTT 지불비용	연속형		435	Q34	광고가 인기 콘텐츠 위주로만 광고가 붙는가, 아니면 모든 콘텐츠에 다 동일하게 붙는가	범주형	4	311(이용X+ 무조건 이용)
Q14	OTT 지불비용 적절여부	범주형	3	435	Q35	광고를 보는 대신 기존이용료 보다 얼마나 저렴	범주형	4	311(이용X+ 무조건 이용)
Q15	과거 넷플릭스 이용 경험	범주형	3	655	Q36	광고요금제 구성 선호도	범주형	2	278(이용X)
Q16	넷플릭스 중단이유 (중복문항)	범주형	8	845	Q37	광고 수준 별 이용 의향	범주형	4	278(이용X)
Q17	넷플릭스 콘텐츠 선호유형	범주형	3	190	Q38	직업	범주형	13	0
Q18	이용중인 넷플릭스 요금제	범주형	5	345	Q39	학력	범주형	5	0
Q19	넷플릭스 이용요금 적절성	범주형	3	345	Q40	가정 한달 수입	범주형	8	0
Q20	문제1에서 2~8까지 3 고른사람	범주형	2	1. 365 2. 635	Q41	사회계층	범주형	5	0
Q21	넷플릭스 요금 수준	범주형	3	424	Q42	정치적 성향	범주형	5	0

DATA :: EDA (EXPLORATORY DATA ANALYSIS)



0. Problems in Proposal

- 多 변수 (41개)
- 중복응답 변수 처리
- 결측치 존재
- Y변수가 존재한다면 다중로지스틱으로 분석 가능

1. Create Variables

- Q32~Q35 척도 변수에 무조건 이용(Q31)하는 경우 광고요금제 요인 질문 결측치 해당 -> 전혀 중요하지 않음에 일괄 적용 → Q31 변수 이용 X but 결측치x
- ott 변수 생성 : 현재 사용하고 있는 OTT

1 : ONLY Netflix , 2:Netflix+Others , 3:Others , 4:None

(182)

(473)

(162)

(179)

DATA PROCESSING #2

· cnt 변수 생성 : # 사용 OTT

1: 0개 (183) 2: 1개(172) 3: 2개 (223) 4: 3개 이상(322)

· exp 변수 생성 : Netflix 이용 경험

1 : 현재 이용자(655) 2:과거에만 이용(155) 3: Never(190)

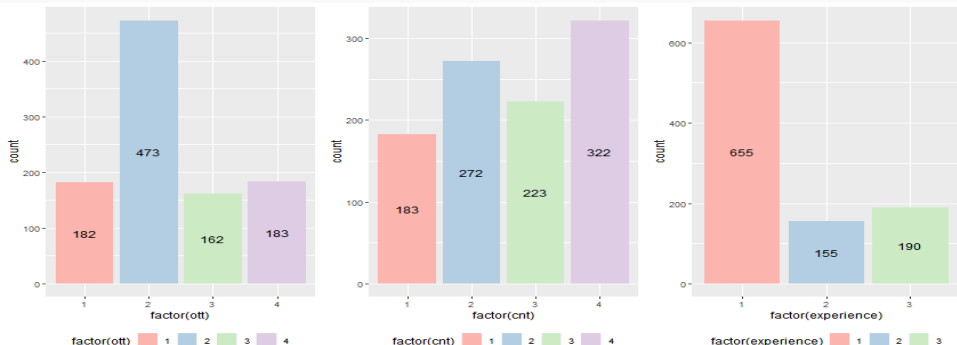


Figure : Histogram of ott , cnt , exp

2. Remove Variables

- 중복 문항 -> REMOVE

(특정 이용자들에게만 해당되는 질문들의 경우이기에 삭제 처리함.)

- 중복 문항 및 사용하지 못하는 문항 삭제 -> 변수 21개
- 넷플릭스 광고 요금제 전혀 이용할 생각x+ 넷플릭스 유저x 삭제
-> +NEFLIX only + NETFLIX with Others

⇒ ‘넷플릭스 사용자 중 광고 요금제 서비스 이용 의향이 있는 사람 대상
(510)

DATA PROCESSING #4

변수명	질문내용	class	#	결측치
Q1	성별 {1: 241(여), 2: 269(남) }	범주형	2	0
Q3	연령 {1:142, 2:137, 3:111, 4:120}	범주형	4	0
ott	현재 사용 하고 있는 ott 1 : Netflix only 139 2: Netflix + others 371	범주형	4	0
cnt	# of 사용 OTT 2: 1 (139); 3: 2(146), 4: 3 (325)	범주형	4	0
exp	Netflix 경험 1: 현재 사용(510)	범주형	4	0
Q17	넷플릭스 콘텐츠 선호 유형	범주형	3	0
Q18	이용중인 넷플릭스 요금제	범주형	5	0
Q19	넷플릭스 이용요금 적절성	범주형	3	0
Q28	광고요금제 알았는지 여부	범주형	2	0
Q29	광고요금제 출시에 대한 입장	범주형	4	0
Q30	광고 요금제 도입으로 인한 넷플릭스 전망	범주형	5	0
Q31	광고 요금제 출시 하면 서비스 이용 의향	범주형	3	0
Q32	광고 요금제 이용시 광고 개수 및 시간	범주형	4	0
Q33	광고 요금제 중간 광고 유무	범주형	4	0
Q34	광고가 인기 콘텐츠 위주로만 광고가 붙는가, 아니면 모든 콘텐츠에 다 동일하게 붙는가	범주형	4	0
Q35	광고를 보는 대신 기존 이용료 보다 얼마나 저렴	범주형	4	0
Q36	광고 요금제 구성 선호도	범주형	2	0
Q37	광고 수준 별 이용 의향	범주형	4	0
Q38	직업	범주형	13	0
Q39	학력	범주형	5	0
Q40	가정 한달 수입	범주형	8	0
Q41	사회계층	범주형	5	0

Figure : processed data table

METHODOLOGY

LCA (Latent Class Analysis ; 잠재범주분석)

- 유사한 특성을 지닌 개인들의 잠재집단을 찾아내고, 통계적으로 다양한 적합지표를 제시함으로써 잠재집단 도출의 객관적 집단 분류가 가능
 - 자료의 정규성이나 선형성, 분산의 동질성과 같은 통계적 선행요건을 충족시키지 않아도 된다는 점에서 군집분석과 달리 자료의 활용도가 높음
- ⇒ LCA Model seems to fit my data

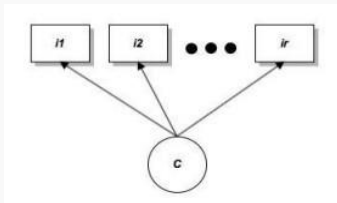


Figure : Diagram of a general latent class model

LCA WITH PANEL DATA

Process of Analysis

1. 변수를 조합해가며 class를 잘 나눠주는 변수들 조합 찾기.
(class별 특징이 유의미해야함)
2. 유의해보이는 변수들 조합들로 class 1~6 모델 비교
3. 가장 BIC가 낮은 class를 가지는 모델을 선택.
4. 분석결과 해석

Latent Class Model

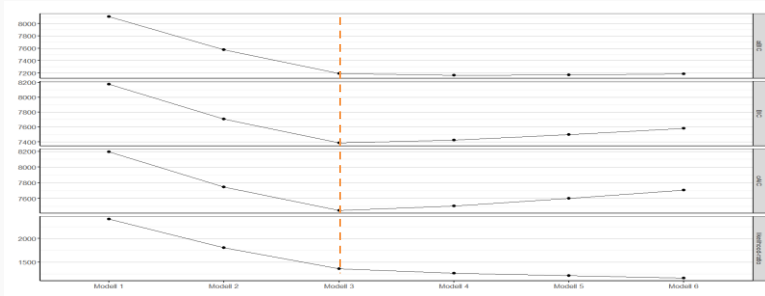


Figure : Scree Plots

Model	Log-Likelihood	Resid.df	BIC	aBIC	cAIC	Likelihood-ratio	Entropy
Model1	-4025.831	490	8176.351	8112.868	8196.351	2411.031	7.893786
Model2	-3724.59	469	7704.79	7574.65	7745.79	1808.547	7.305024
Model3	-3499.71	448	7385.953	7189.156	7447.953	1358.788	6.876606
Model4	3452.752	427	7422.96	7159.506	7505.96	1264.872	6.722338
Model5	-3425.102	406	7498.583	7168.472	7602.583	1209.572	6.722338
Model6	-3400.938	385	7581.178	7184.41	7706.178	1161.245	6.678113

INTERPRETATION

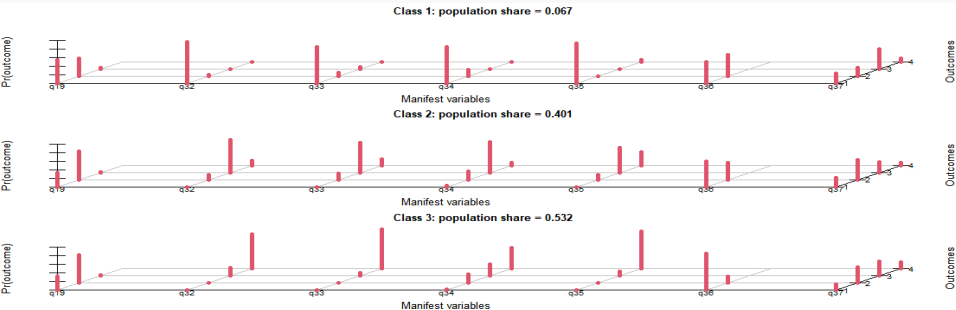


Figure : polCA plots with 3 class

Interpretation

1. Class 1 (6.7%) :요금 보다 광고 유형에 민감
2. Class 2 (40.1%) :요금에 조금 부담 느끼고 광고 방식에도 어느정도 민감
3. Class 3 (53.2%) :광고 유형 보다 요금에 민감

		class1	class2	class3
문항		요금보다 광고 민감 (0.06)	광고 + 요금 중요 (0.40)	광고보다 요금 중요 (0.53)
넷플릭스 이용요금	적절	55.93%	33.17%	32.10%
	다소 비쌌	41.13%	66.32%	66.07%
	오히려 저렴	2.94%	0.51%	1.83%
콘텐츠당 붙는 광고 개수 및 시간	전혀 중요하지않다	97.06%	0.00%	0.379%
	별로 중요하지않다	2.94%	11.89%	0.00%
	약간 중요	0.00%	76.30%	18.78%
	매우 중요	0.00%	11.96%	80.84%
콘텐츠 중간중간 광고 유무	전혀 중요하지않다	85.37%	0.00%	0.00%
	별로 중요하지않다	8.83%	14.67%	0.00%
	약간 중요	5.80%	70.18%	6.80%
	매우 중요	0.00%	15.14%	93.17%
인기 or 모든 콘텐츠 위주	전혀 중요하지않다	85.28%	2.51%	2.91%
	별로 중요하지않다	14.71%	19.06%	21.36%
	약간 중요	0.00%	71.44%	26.85%
	매우 중요	0.00%	6.98%	48.87%
기존 이용료 보다 얼마나 저렴	전혀 중요하지않다	94.20%	0.00%	0.00%
	별로 중요하지않다	0.00%	10.46%	0.96%
	약간 중요	0.00%	58.37%	11.29%
	매우 중요	5.80%	31.18%	87.75%
선호하는 광고 요금제 선택지	콘텐츠 시작 전에만 광고가 붙고 기존 요금제에서 덜 깎아주는 광고 요금제	49.96%	59.82%	84.960%
	콘텐츠 시작 전 + 중간중간 광고 붙고 기존 요금제보다 더 많이 깎아주는 요금제	50.04%	40.18%	15.040%
광고 요금제 수준	광고 多 + 이용료 완전 무료(거의 무료)	23.46%	21.22%	13.49%
	광고 적당 + 이용료 절반 수준	20.61%	46.65%	37.42%
	광고 조금 + 이용료 30% 저렴	47.10%	25.17%	33.71%
	광고를 일부 콘텐츠에만 제한적 사용 +지금보다 10~20%저렴	8.83%	6.96%	15.38%

LCA model with covariate (gender)

```
Estimated class population shares
0.0646 0.4027 0.5328
```

```
Predicted class memberships (by modal posterior prob.)
0.0647 0.3941 0.5412
```

```
=====
Fit for 3 latent classes:
=====
```

```
2 / 1
```

	Coefficient	Std. error	t value	Pr(> t)
(Intercept)	1.61946	0.26459	6.121	0.000
q12	0.46916	0.44364	1.058	0.291

```
=====
3 / 1
```

	Coefficient	Std. error	t value	Pr(> t)
(Intercept)	1.79170	0.25890	6.920	0.000
q12	0.67113	0.43511	1.542	0.124

```
=====
number of observations: 510
number of estimated parameters: 58
residual degrees of freedom: 452
maximum log-likelihood: -2992.547
```

```
AIC(3): 6101.093
BIC(3): 6346.689
X^2(3): 6837.856 (Chi-square goodness of fit)
```

Figure : Fit for $f=(q19,q32,q33,q34,q35,q37)\sim q1$

LCA MODEL GROUPED BY AGE

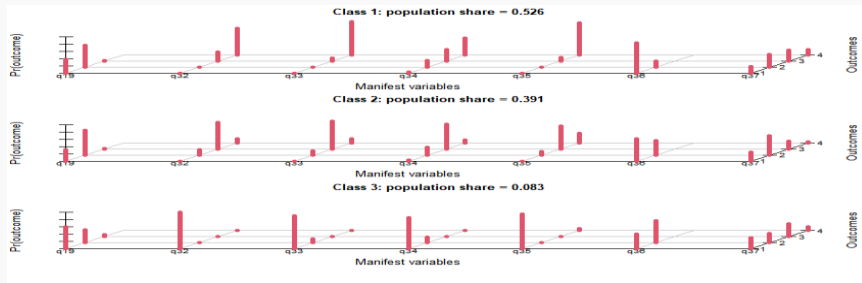


Figure : polCA plots with 3 class with Female

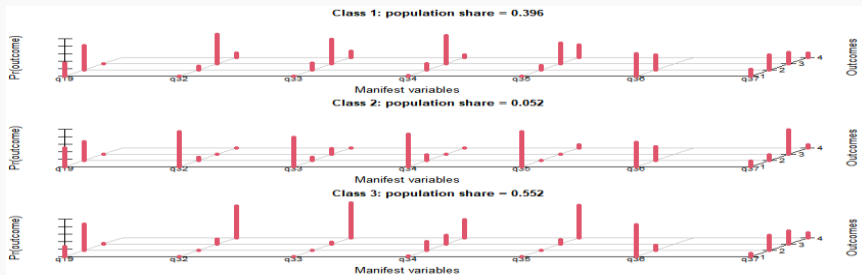


Figure : polCA plots with 3 class with Male

		Female (낮은 요금 매우 선호)			Male (요금보다는 광고 개수에 민감하게 반응)		
		class1 요금부담小 + 광고小 0.083	class2 요금부담 + 광고중요 0.397	class3 요금보다 광고중요 0.526	class1 광고개수중요 0.052	class2 광고+요금 중요 0.396	class3 요금보다 광고중요 0.552
문항	범주						
넷플릭스 이용요금	적절	60.00%	30.10%	38.50%	50.00%	34.50%	27.11%
	다소 비쌌	35.00%	67.91%	59.96%	50.00%	65.50%	70.87%
	오히려 저렴	5.00%	1.11%	1.54%	0.00%	0.00%	2.02%
콘텐츠당 붙는 광고개수및 시간	전혀 중요하지않다	100.00%	0.00%	0.79%	92.86%	0.00%	0.00%
	별로 중요하지않다	0.00%	14.85%	0.00%	7.14%	9.39%	0.00%
	약간 중요	0.00%	73.62%	24.93%	0.00%	78.87%	14.81%
	매우 중요	0.00%	11.53%	74.28%	0.00%	11.74%	85.19%
콘텐츠 중간중간 광고 유무	전혀 중요하지않다	90.00%	0.00%	0.00%	78.57%	0.00%	0.00%
	별로 중요하지않다	10.00%	11.67%	0.00%	7.14%	17.83%	0.00%
	약간 중요	0.00%	76.23%	8.79%	14.29%	65.52%	6.20%
	매우 중요	0.00%	12.11%	91.21%	0.00%	16.65%	93.80%
인기 or 모든 콘텐츠 위주	전혀 중요하지않다	85.00%	3.34%	3.04%	85.71%	1.92%	2.66%
	별로 중요하지않다	15.00%	20.86%	19.19%	14.28%	17.07%	23.45%
	약간 중요	0.00%	68.31%	31.25%	0.00%	74.45%	24.04%
	매우 중요	0.00%	7.48%	46.51%	0.00%	65.56%	49.85%
기존이용료 보다 얼마나 저렴	전혀 중요하지않다	95.00%	0.00%	0.00%	92.59%	0.00%	0.00%
	별로 중요하지않다	0.00%	9.67%	0.70%	0.00%	12.13%	0.73%
	약간 중요	0.00%	63.62%	10.28%	0.00%	54.89%	12.48%
	매우 중요	0.00%	26.71%	89.03%	7.14%	32.97%	86.80%
선호하는광고 요금제선택지	콘텐츠 시작 전에만 광고가 붙고 기존 요금제에서 덜 깎아주는 광고 요금제	40.00%	60.48%	83.63%	64.29%	59.65%	85.17%
	콘텐츠 시작 전 + 중간중간 광고 붙고 기존 요금제보다 더 많이 깎아주는 요금제	60.00%	39.52%	16.37%	35.71%	40.35%	14.80%
광고요금제 수준	광고 多 + 이용료 완전 무료(거의 무료)	30.00%	24.27%	19.03%	14.29%	17.84%	9.43%
	광고 적당 + 이용료 절반 수준	25.00%	52.68%	34.98%	14.29%	40.86%	40.06%
	광고 조금 + 이용료 30% 저렴	35.00%	20.62%	30.43%	64.29%	29.40%	36.16%
	광고를 일부 콘텐츠에만 제한적 사용 +지금보다 10~20%저렴	10.00%	2.42%	15.56%	7.14%	11.91%	14.35%

CONCLUSION

Summary

- Best LCA model has 3 classes with 7 categorical variables
- 성별로 나누게 되면 잠재 변수가 뚜렷하게 보이지 않는다 . (혼재되어있음)
- 따라서 합쳐서 결과 보는 것이 더 잠재 변수 탐지가 쉬움

Limitation

- 데이터 전처리 과정상, 넷플릭스 사용자를 대상으로 했기 때문에 다른 플랫폼만 쓰는 구독자나 구독하지 않는 사람들에 대해 광고요금제에 대한 입장을 알 수 없다.
- 성별로 다른 변수들의 조합으로 하면 결과가 유의 했을 수 도 있다.
- 나이대로 나누게 되면 n 이 너무 작아져서 결과를 보기가 어렵다.
- 여전히 많은 변수들이 있기에 더 잠재변수가 뚜렷하게 나타나는 조합이 있을수도 있음

Further Study

- 다른 변수들의 조합으로 성별 LCA 찾아보기
- 실제 시행된 광고요금제의 형태 :
시작 전과 도중에 15초 또는 30초 길이의 광고 재생(시간당 평균 4~5분의 광고)
+ 기존의 베이식 보다 약 58% 할인된 요금 (9500->5500)
- 실제 시행된 광고요금제로 전환한 넷플릭스 이용자들에 대해서 분석하여 이 결과와 비교

THANK YOU FOR YOUR ATTENTION !