

# 넷플릭스 광고 요금제 도입에 따른 유저 반응에 대한 잠재범주분석

: 혼합모형 기말 프로젝트

(Latent class analysis of user response data to launch Basic-Ads plan by Netflix)

: Mixture model final project

박나린 (Park, Narin)

2022711906

성균관대학교 대학원 통계학과 석사 2 기

## I. 개요

---

지난 2 년간 코로나 19 의 확산과 OTT<sup>1</sup> 서비스의 발전이 맞물려 전 세계적으로 OTT 이용량이 급증했다. 이 경향은 한국에서도 찾아볼 수 있었는데, 실제로 2019 년 한국의 OTT 서비스 이용률이 41.4%에서 2021 년 81.7%로 2 배 이상 증가했음을 알 수 있다 (정보 통신정책연구원, 2022-07). 이렇게 방대해진 OTT 시장에서 큰 점유율을 차지하는 플랫폼 중 하나인 넷플릭스(Netflix) 또한 2 년동안 가파른 성장을 하였지만, 2022 년 상반기에 처음으로 주춤하게 되었다. 바로 구독자 수가 2011 년 이후로 처음으로 감소 했기 때문이다. 넷플릭스에게 구독자 수는 바로 수입의 원천이기에 구독자 수 감소 현상이 장기적으로 가속화 된다면 회사의 큰 위험이 될 수 있다. 실제로 상반기 실적 발표 후 700\$에 달하던 넷플릭스 주식이 200\$ 초반대로 까지 떨어지는 하락세를

---

<sup>1</sup> Over-the-Top: 셋톱박스과 같은 단말기에 국한되지 않고 인터넷과 같은 모든 방식을 통해 제공되는 다양한 유형의 스트리밍 미디어 콘텐츠

보이기도 했다. 구독자 수의 감소 원인으로는 크게 3 가지를 볼 수 있다. 먼저, 단기적으로, 러시아 우크라이나 침공으로 넷플릭스가 러시아 서비스를 중지하게 되어 70 만명이 급감하게 되었다. 두번째, 코로나 19 팬데믹(Pandemic)이 장기화 되면서 엔데믹(Endemic; 풍토병)화 되었고, 사람들이 집에서 보내는 시간이 줄어 들었다. 마지막으로 다른 OTT 서비스들이 우후죽순(雨後竹筍)으로 생겨나고 성장하는 바람에 OTT 시장의 경쟁이 가속화 되었다.

넷플릭스는 이러한 위기를 탈피하고자 2022 년 11 월부터 12 개 국가 (한국, 독일, 미국, 멕시코, 브라질, 스페인, 영국, 이탈리아, 일본, 캐나다, 프랑스, 호주)에서 '광고형 베이식 요금제(Basic Ads plan)'을 새로 출시하게 되었다. '광고형 베이식 요금제'란 라이선스 제한으로 인한 일부컨텐츠를 제외하고 기존 베이식 요금제와 동일한 내용에 시간당 시작 전이나 중간에 4~5 분 정도되는 광고를 영상에 붙이는 요금제 이다. 이 요금제는 광고가 붙는 대신 원래 베이식 요금제의 가격인 9500 에서 5500 으로 반 이상의 할인된 금액으로 구독자들에게 제공된다. 넷플릭스는 이 저렴한 광고 지원 요금제로 더 많은 구독자를 유치하고 기존 구독자들을 지켜내서 수익성을 개선하는 방안으로 다시 시장에서 우위를 차지하고자 한다.

프로젝트 진행자는, 과연 광고요금제가 생겨도 구독자들이 실제로 광고요금제를 구독하는 것을 고려할지, 광고에 대한 거부감이 없는지에 대한 의문이 들었다. 따라서 이번 프로젝트는 넷플릭스 광고요금제에 대한 구독자들에 대한 반응에 대한 잠재 요인을 탐지하는 것을 목적으로 진행하였다.

## II. 자료 해설

### 1. 데이터 출처 및 소개

프로젝트에 사용된 데이터는 '광고요금제' 도입을 앞둔 넷플릭스에 대한 인식 및 이용에 관한 패널 데이터<sup>2</sup>로 한국언론진흥재단 미디어 연구 센터에서 실시한 온라인 조사로, 20

<sup>2</sup> 2022 미디어이슈 5 호 <'광고요금제' 도입을 앞둔 넷플릭스에 대한 인식 및 이용조사> 설문조사 데이터 :

[https://www.kpf.or.kr/front/board/boardContentsView.do?board\\_id=292&contents\\_id=90f2e08ef50](https://www.kpf.or.kr/front/board/boardContentsView.do?board_id=292&contents_id=90f2e08ef50)

~ 50 대 1 천명을 대상으로 했다. 성별, 연령대 및 거주지역을 기준으로 할당하여 모집했으며, 2022 년 9 월 16 ~ 20 일에 실시 되었다. 15,625 명에게 조사 안내 이메일을 발송하여, 응답한 1,366 명 중 데이터 클리닝 과정을 거쳐 총 1,000 명을 뽑아냈으며 이 데이터에 대한 표본오차는 95% 신뢰수준에서  $\pm 3.0\%$  포인트이다. 결국 이 데이터는 42 개의 질문지 문항과 1000 개의 응답자로 이루어져 있다고 볼 수 있다.

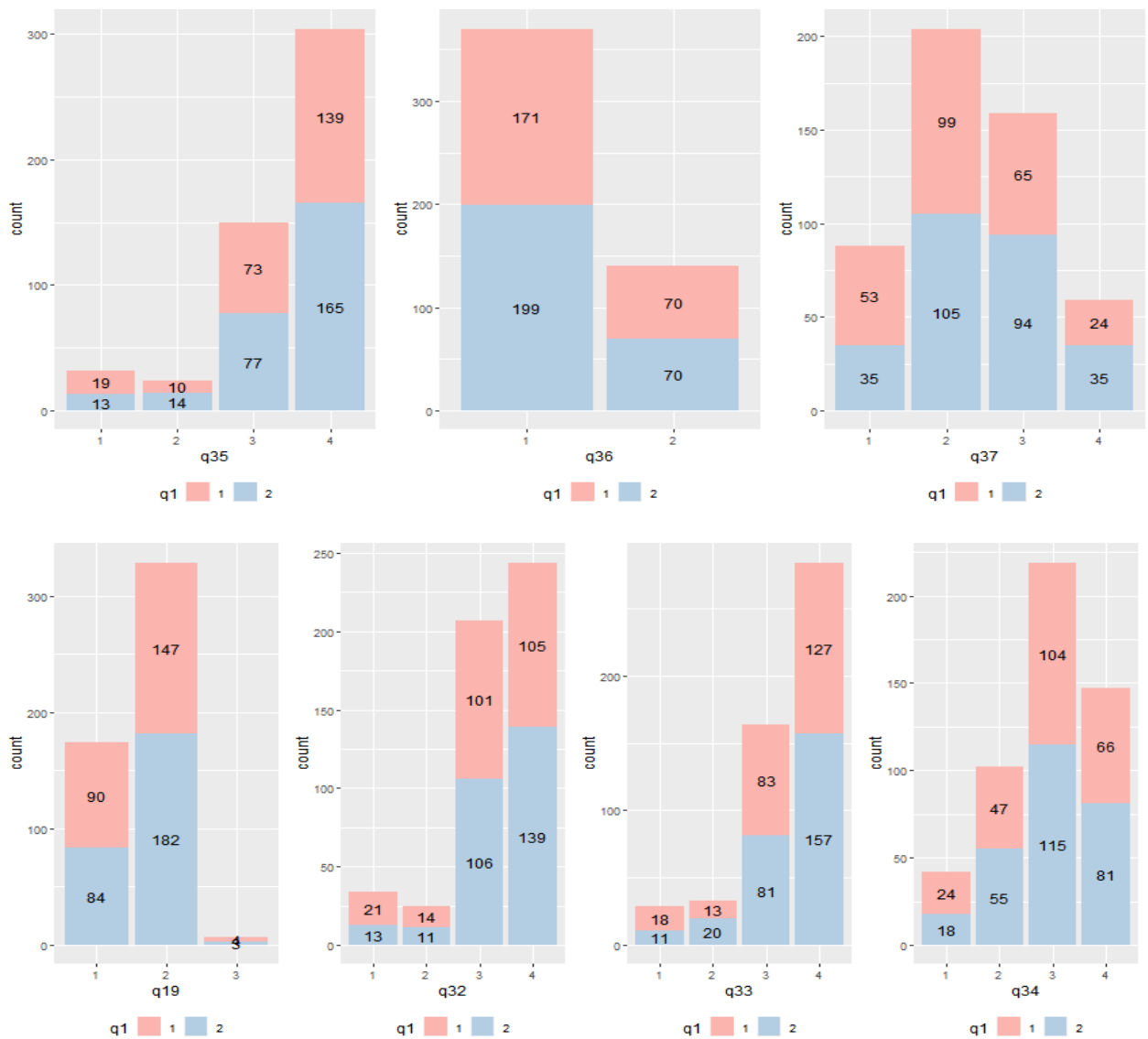
## 2. EDA (EXPLORATORY DATA ANALYSIS)

전반적인 데이터의 구성을 확인해 보았을 때 이 데이터는 전형적인 설문조사 데이터로, 연령과 OTT 지불비용 외에는 모두 범주형 변수로 이루어져 있다. 42 개의 변수 중 19 개의 변수에서 결측치가 나타났으며, 대부분의 결측치는 연결 문항에 해당 되는 경우였다. 즉, 상위 질문에서 응답한 대답이 아래 항목에 대답하지 않아도 되어서 결측치가 생기는 경우로 이를 제외하고 실제 설문지 응답자가 미 응답(non-response)하여 결측치가 생긴 문항은 Q13, Q17, Q19 이다. 또한 Q5 부터 Q12 까지는 사용 OTT 별 로 동일 질문을 하고 있어 이 점을 활용한다면 변수를 줄일 수 있을 걸로 확인된다.

번호	질문내용	NA	번호	질문내용	NA
Q1	성별	0	Q22	OTT 중 넷플릭스만 사용하는 이유	818
Q2	연령	0	Q23	문 1-10 로직 1 번	0
Q3	연령	0	Q24	문 1-10 로직 2 번	0
Q4	지역	0	Q25	문 1-10	0
Q5	넷플릭스	0	Q26	OTT 하나만 이용하는 이유	455
Q6	티빙	0	Q27	OTT 중복으로 이용하는 이유	817
Q7	웨이브	0	Q28	광고요금제 알았는지 여부	0
Q8	쿠팡플레이	0	Q29	광고요금제 출시에 대한 입장	0
Q9	디즈니플러스	0	Q30	광고요금제 도입으로 인한 넷플릭스 전망	0
Q10	애플티비	0	Q31	광고요금제 출시 하면 서비스 이용 의향	311
Q11	왓차	0	Q32	광고요금제 중간 광고 유무	311
Q12	시즌	0	Q33	광고가 인기콘텐츠 위주로 광고가 붙는가, 모든 콘텐츠에 동일하게 붙나	311
Q13	OTT 지불비용	436	Q34	광고요금제 구성 선호도	311

Q14	OTT 지불비용 적절여부	435	Q35	광고를 보는 대신 기존 이용료보다 얼마나 저렴	311
Q15	과거 넷플릭스 이용 경험	655	Q36	광고요금제 구성 선호도	278
Q16	넷플릭스 중단 이유	845	Q37	광고 수준 별 이용 의향	278
Q17	넷플릭스 콘텐츠 선호유형	190	Q38	직업	0
Q18	이용중인 넷플릭스 요금제	345	Q39	학력	0
Q19	넷플릭스 이용요금 적절성	345	Q40	가정 한달 수입	0
Q20	문제 1 에서 2~8 까지 3 고른사람	0	Q41	사회계층	0
Q21	넷플릭스 요금수준	424	Q42	정치적 성향	0

[표 1] 변수



[그림 1] 주요 질문지 문항 별 응답 히스토그램(1 :여성 2:남성)

또한 성별로 변수들의 분포가 차이가 있는지 주요 변수들을 기준으로 확인해 보았을 때, 대부분의 분포가 크게 차이는 없음을 알 수 있다. 하지만 문항 대답을 비교 해보았을 때 성별에 따라 좀 더 많은 인원이 대답하는 문항이 있었고 이에 프로젝트 진행자는 성별로 따로 잠재범주모형을 적용해 볼 필요성을 느끼게 되었다. 20 명정도 차이가 1000 명의 응답자 속에서는 유의미한 차이가 있을 것으로 생각 되었기 때문이다.

결측치와 더불어 이 데이터에는 중복응답 변수가 존재한다. Q16, Q22, Q26, Q27 4 가지 질문이 모두 중복응답 질문인데, 중복 응답 질문인 동시에 연결 문항에 따른 결측치인 것으로 나타나서 아마 제거가 필요한 문항들로 보인다.

### 3. 데이터 전처리

#### 3.1 변수 생성 (Create Variables)

먼저 Q32~Q35(광고 요금제 요인 질문 문항) 척도변수에서 Q31)'광고 요금제 출시하면 서비스 이용 의향이 있는지'에 대한 질문에서 '*무조건 이용하겠다*' 라고 응답한 응답자들의 광고 요금제 요인 질문지에 결측치로 설정 되어있다. 따라서 이 부분에서 무조건 이용하겠다고 한 사람들은 요금제 요인에 상관없이 사용할 것이므로 결측치를 *1.전혀 중요하지 않음*에 일괄 적용 하였다. 따라서 Q31 변수에 따라 일괄 부여 했으므로 Q31 변수를 사용하지 않게 되었다.

두번째로 OTT 변수를 생성하였다. OTT 변수는 현재 사용하고 있는 OTT 변수로 Q5 부터 Q12 변수를 활용하여 하나의 변수로 줄이게 되었다. 그렇게 만든 변수의 범주 별 개수는 [표 2]와 같다.

1 : Netflix only	2 : Netflix + Others	3 : Others	4 : None
182	473	162	179

[표 2] OTT 변수

세번째로 CNT 변수를 생성하였다. CNT 변수는 사용자당 OTT 사용개수를 범주화 한 것으로 OTT 변수와 마찬가지로 Q5 부터 Q12 변수를 활용하였다.

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

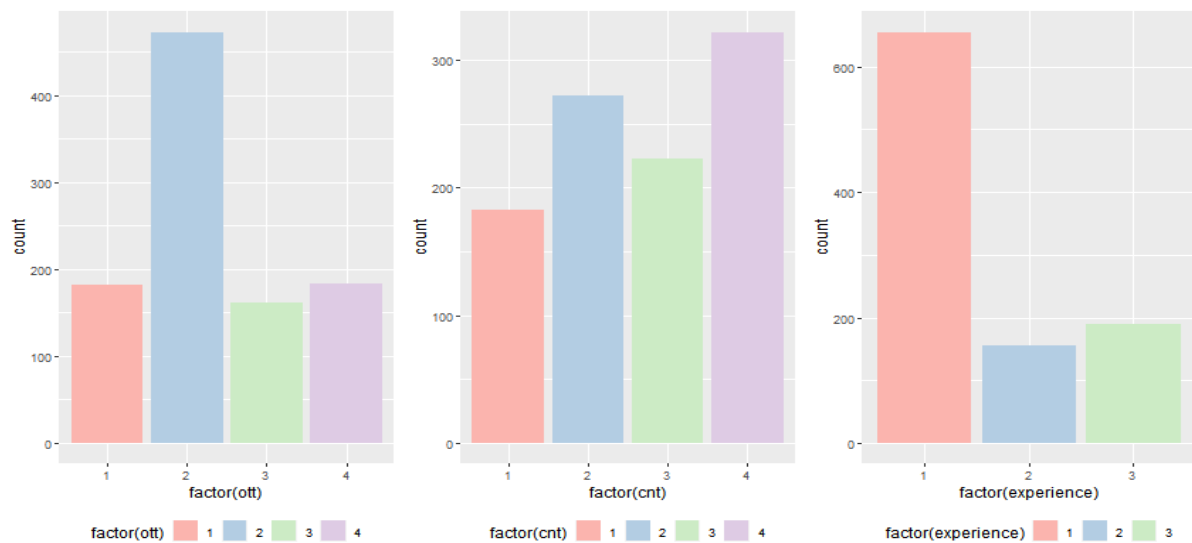
[표 3] CNT 변수

마지막으로 EXP 변수이다. EXP 변수는 Netflix 이용 경험(Experience) 여부를 묻는 변수로 OTT 변수와 Q15 변수를 이용하여 생성해냈다.

1 : 현재 이용자	2 : 과거에만 이용	3: NEVER
655	155	190

[표 4] EXP 변수

생성된 변수들은 cnt 변수를 제외하고는 조금씩 치우쳐 있는 분포의 형상을 띤다. 3 가지 생성 변수의 히스토그램은 아래 [그림 2]와 같다.



[그림 2] Histogram of ott ,cnt, exp

### 3.2 변수 및 응답자 삭제

먼저 4 개의 중복 문항 모두 연결문항으로 인한 결측값이 존재하며 그 결측값이 매우 크다. 또한, 문항 자체가 응답자들의 광고요금제에 대한 해석과는 관련이 적은 문항이기에 모두 삭제 했다.

두번째로, 로직변수들과 변수 생성과정에서 사용하지 못하게 된 Q31 문항을 삭제 했다. 로직변수들은 데이터 자체에서 가지고 있던 변수로 유의미하게 사용할 수 없고 Q31 은 변수생성에서 언급했던 이유로 삭제 했다.

마지막으로 넷플릭스 광고요금제를 전혀 이용할 생각이 없고 넷플릭스 유저가 아닌 사람들을 삭제 했다. 그 이유는 Q32~Q37 변수는 광고요금제를 전혀 이용할 생각이 없는 사람들은 답변하지 않았지만 프로젝트 목적상 이 변수들이 중요변수로 작용하기 때문에

결측값을 없애 주기 위하여 일괄 삭제 하였다. 또 넷플릭스 유저가 아닌 사람들을 삭제한 이유는 Q17~Q19 문항이 넷플릭스 유저가 아닌 사람들은 답변하지 않은 문항이지만, 넷플릭스 사용 유형을 알 수 있는 문항들이기에 결측치를 최대한 줄이고 사용하고 싶어서 진행자는 넷플릭스 유저가 아닌 사람들을 삭제하기로 했다. 결국 이 데이터에는 넷플릭스 사용자중 광고 요금제 서비스 이용 의향이 있는 사람을 대상으로 한 데이터셋이 되었으며, 변수 생성 과정 및 변수 삭제 과정을 거친 데이터는 21 개의 변수와 510 명의 응답자들이 남게 되었다.

변수명	질문내용		class	#	NA
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		범주형	4	0
cnt	# of 사용 OTT		범주형	4	0
exp	Neflix 경험		범주형	4	0
	1: 현재 사용(510)				
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
Q32	광고 요금제 이용 시 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

[표 5] 최종 변수 목록

### III. 방법론

#### 1. 잠재범주분석 (LATENT CLASS ANALYSIS)

이번 프로젝트에 사용될 방법론은 잠재범주분석(Latent Class Analysis)이다. LCA 모형은 mixture model 의 하나로 각 범주형 변수들에 대한 응답들 안에서 잠재범주(Latent Class)에 따라 집단을 분류하는 분석이다. 이 모형에서 범주형 변수들은 서로 독립임을 가정하며 아래의 식과 같은 모형을 가진다.

$$\begin{aligned}
 P(Y = y) &= P(Y_1 = y_1, \dots, Y_M = y_M) \\
 &= P(Y_1 = y_1, \dots, Y_M = y_M | L = 1)P(L = 1) + \dots + P(Y_1 = y_1, \dots, Y_M = y_M | L = C)P(L = C) \\
 &= \sum_{l=1}^C \pi_l P(Y_1 = y_1 | L = l) \times \dots \times P(Y_M = y_M | L = l) \\
 &= \sum_{l=1}^C \pi_l \prod_{m=1}^M P(Y_m = y_m | L = l) = \sum_{l=1}^C \pi_l \prod_{m=1}^M \prod_{k=1}^r \{P(Y_m = k | L = l)\}^{I(Y_m=k)} \\
 &= \sum_{l=1}^C \pi_l \prod_{m=1}^M \prod_{k=1}^r \rho_{mk|l}^{I(Y_m=k)}
 \end{aligned}$$

$Y$ 는  $i$  번째 응답자의 다분 문항 응답이며,  $L=1, \dots, C$  은 응답자가 속하는 잠재 계층,  $\pi_l$ 은  $l$  번째 그룹의 비율이고,  $\rho_{mk|l}$ 는  $l$  번째 잠재범주에  $Y_m = k$  가 있을 확률을 말한다. 여기서  $k=1, \dots, r$  는 질문지 문항의 개수가 된다.

적절한 잠재범주의 수를 가지는 최적의 모형을 선택하기 위해서 평가 기준으로 (Model Selection Criteria)는 우도비 검정 (Likelihood-Ratio Test), BIC(Bayesian Information), AIC(Akaike Information), 등을 생각해볼 수 있다.

### IV. 연구방법

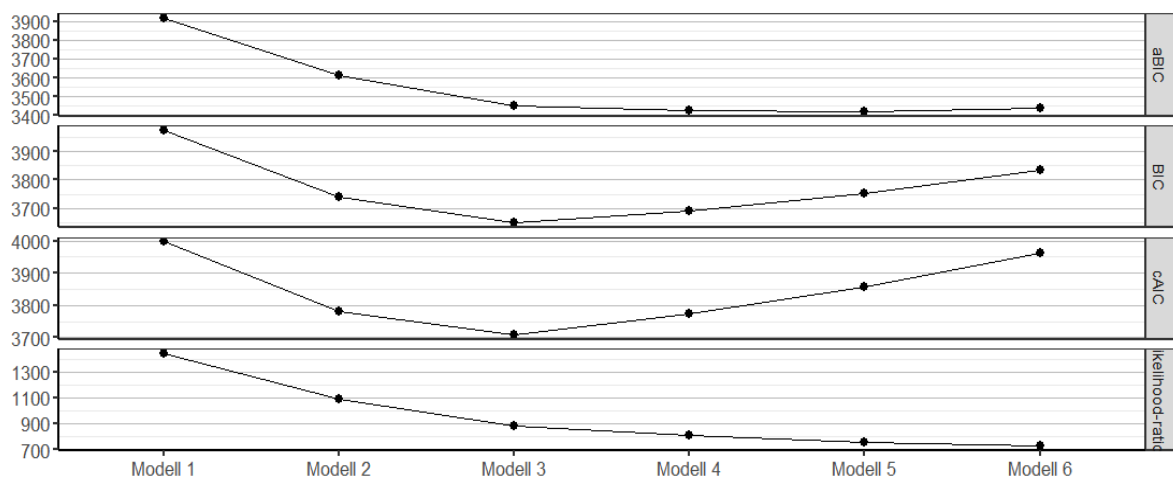
#### 1. 변수 선택

먼저 잠재범주분석에 사용될 변수들을 정하기 위해 유의미한 해석을 줄 수 있는 변수들 중 BIC 값을 가장 낮게 가지는 변수들을 선택했다. 이 프로젝트의 목적은 넷플릭스 광고요금제에 대한 구독자들에 대한 반응에 대한 잠재 요인 탐지 이므로, 광고요금제에 대한 응답자들의 생각을 알 수 있는 변수들인 Q32~Q37 를 우선 선택하였고, 이외의 변수들을 이 변수들과 조합하여 BIC 들을 서로 비교 하였다. Q38)직업 변수는 너무



범주가 다양하고 재범주화 하기도 어려워 사용 하기가 쉽지 않았고, 나머지 학력 수입 계층 연령변수 모두 포함하였을 때 결과 해석이 쉽지 않았다. 따라서 Q17 ~Q19 변수를 넣었을 때 해석이 유의미하고 가장 낮은 BIC 를 가지는 집단이 Q19, Q32, Q33, Q34, Q35, Q36, Q37 를 변수로 가지는 잠재 범주 모형이었다. Q17 ~Q19 의 여러가지 조합이 있었지만 변수를 많이 채택할수록 해석이 더 복잡해질 것으로 우려 되어 가장 낮은 BIC 를 가지면서 변수 7 개를 가진 모형을 택하게 되었다.

## 2. 그룹 개수 선택



[그림 3] Scree Plots

선택한 변수들로 1~6 개의 그룹으로 LCA 를 시도하여 각각의 BIC 를 서로 비교 하였다. 비교 결과 [그림 1]에서 볼 수 있듯이 class 가 3 개일 때 가장 낮은 BIC 를 가지고 있으며 3 개보다 커지면 다시 BIC, cAIC 가 커짐을 알 수 있다. 따라서 class 가 3 인 잠재범주모형이 가장 적합해 보인다.

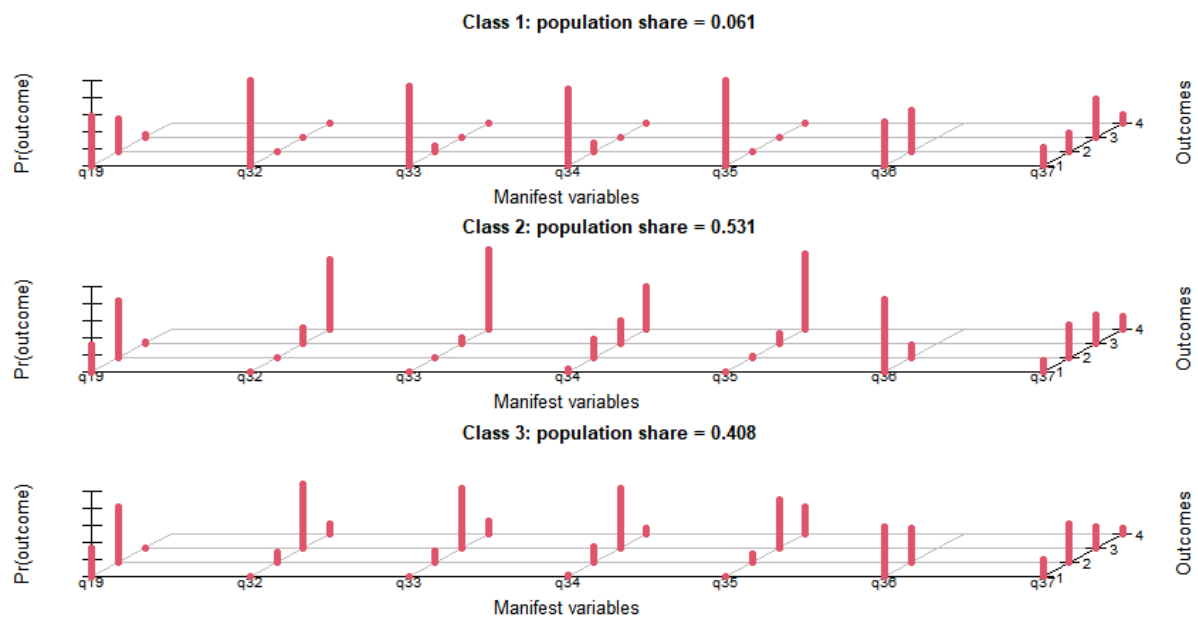
Model	Log-Likelihood	Resid.df	BIC	aBIC	cAIC	Likelihood-ratio	Entropy
Model1	-4025.831	490	8176.351	8112.868	8196.351	2411.031	7.893786
Model2	-3724.59	469	7704.79	7574.65	7745.79	1808.547	7.305024
<b>Model3</b>	<b>-3499.71</b>	<b>448</b>	<b>7385.953</b>	<b>7189.156</b>	<b>7447.953</b>	<b>1358.788</b>	<b>6.876606</b>
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

[표 6] Model Selection Criteria

### 3. 잠재범주모형 분석

R 의 poLCA 패키지를 이용하여 진행한 잠재범주 모형 분석에서 3 개의 잠재집단을 가지는 모형의 그래프는 [그림 4]와 같다. Class1 은 6.1%로 요금 보다 광고 유형에 굉장히 민감한 집단, Class2 는 53.1%로 광고 유형 보다 요금에 아주 민감한 집단 마지막으로 Class3 는 40.8%로 요금에 조금 부담 느끼고 광고 방식에도 어느정도 민감한 집단으로 해석 할 수 있다. 더 자세한 분석은 분석결과에서 해석하겠다.



[그림 4] Latent Class Graph

## V. 분석 결과 및 해석

### 1. 전체 모형

먼저 Class1 은 요금보다 광고 민감한 유형으로 전체의 6.7% 밖에 되지 않는 유형이다. 요금은 다른 그룹에 비해 적절하다는 사람이 많으며, 광고 시행 형태는 전혀 중요하지

않지만, 광고가 적게 붙는 쪽을 선택한 사람이 더 많다. 또한 다른 그룹들이 요금에 더 민감하게 반응하는 것에 비해 이 그룹은 광고 개수에 더 민감하게 반응하기에 요금보다 광고 개수에 민감한 그룹이라고 생각 할 수 있다.

두번째 집단은 전체의 40.1%로 광고와 요금 모두 중요한 집단이다. 요금이 다소 비싸다고 생각하는 사람들이 다른 집단에 비해 많으며 전반적으로 광고 개수 및 형태와 요금 모두 민감하게 반응 하고 있음을 알 수 있다.

세번째 집단은 3 개의 집단 중 가장 많이 차지 하고 있는 53.2%의 집단으로 요금에 더 민감한 집단이다. 이 집단들은 다른 집단에 비해 요금을 묻는 항목에 더 민감하게 반응 했으며, 광고 요금제 수준에서 도 다른 집단 보다 요금이 저렴한 항목에 더 답변을 많이 한 것을 볼 수 있다. 다만, 선호하는 광고 요금제 선택지에서 이 그룹의 84.96%가 콘텐츠 시작 전에만 광고가 붙고 기존 요금제에서 덜 깎아주는 광고 요금제를 선택했다. 이 질문지 문항이 선호하는 요금제에 대한 질문인데, 기본 요금제에서 덜 깎아준다는 말이 수치적으로 나타낸 것이 아니기 때문에, 응답자들이 이런 부분에서 그 정도를 생각하지 않고 선택했을수 있지않을까 라고 생각해보았다. 이 항목을 제외하고는 모든 지표들이 요금에 더 민감하게 반응 했기 때문에 프로젝트 진행자는 3 번째 집단을 요금에 좀 더 민감한 집단으로 분류 하였다.

문항		class1	class2	class3
		요금보다 광고 민감 ( 0 . 067 )	광고 + 요금 중요 ( 0 . 401 )	광고보다 요금 중요 ( 0 . 532 )
넷플릭스 이용요금	적절	<b>55.93%</b>	33.17%	32.10%
	다소 비쌌	41.13%	<b>66.32%</b>	<b>66.07%</b>
	오히려 저렴	2.94%	0.51%	1.83%
콘텐츠당 붙는 광고 개수 및 시간	전혀 중요하지않다	<b>97.06%</b>	0.00%	0.379%
	별로 중요하지않다	2.94%	11.89%	0.00%
	약간 중요	0.00%	<b>76.30%</b>	<b>18.78%</b>
	매우 중요	0.00%	11.96%	80.84%
콘텐츠 중간중간 광고 유무	전혀 중요하지않다	<b>85.37%</b>	0.00%	0.00%
	별로 중요하지않다	8.83%	14.67%	0.00%
	약간 중요	5.80%	<b>70.18%</b>	6.80%
	매우 중요	0.00%	15.14%	<b>93.17%</b>
인기 or 모든 콘텐츠 위주	전혀 중요하지않다	<b>85.28%</b>	2.51%	2.91%
	별로 중요하지않다	14.71%	19.06%	21.36%
	약간 중요	0.00%	<b>71.44%</b>	26.85%
	매우 중요	0.00%	6.98%	<b>48.87%</b>

기존이용료 보다 얼마나 저렴	전혀 중요하지않다	<b>94.20%</b>	0.00%	0.00%
	별로 중요하지않다	0.00%	10.46%	0.96%
	약간 중요	0.00%	<b>58.37%</b>	11.29%
	매우 중요	5.80%	31.18%	<b>87.75%</b>
선호하는 광고 요금제 선택지	콘텐츠 시작 전에만 광고가 불고	<b>49.96%</b>	<b>59.82%</b>	<b>84.960%</b>
	기존 요금제에서 덜 깎아주는 광고 요금제			
	콘텐츠 시작 전 + 중간중간 광고 불고	<b>50.04%</b>	40.18%	15.040%
	기존 요금제보다 더 많이 깎아주는 요금제			
광고요금제 수준	광고 多 + 이용료 완전 무료(거의 무료)	23.46%	21.22%	13.49%
	<b>광고 적당 + 이용료 절반 수준</b>	20.61%	<b>46.65%</b>	<b>37.42%</b>
	광고 조금 + 이용료 30% 저렴	<b>47.10%</b>	25.17%	33.71%
	광고를 일부 콘텐츠에만 제한적 사용	8.83%	6.96%	15.38%
	+지금보다 10~20%저렴			

[표 7] proportion for LCA

## 2. 성별 공변량 모형

R 의 poLCA 패키지의 LCA 모형은 공변량을 추가 할 수가 있는데, 성별 혹은 나이대를 공변량으로 추가하여 진행하고자 하였으나 나이는 변수 삭제후 510 개의 변수로 줄이자 4 개의 범주를 가지 나이 변수를 공변량으로 넣는 것이 작동하지 않았다. 따라서 나이를 공변량으로 넣은 모델을 확인 했는데, P 값이 0.1 보다도 다 커서 유의 하지 않다고 나왔기 때문에 해석을 할 수 가 없었다.

(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	<b>0.291</b>
(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	<b>0.124</b>

[표 8] Fit for 3 latent class

## 3. 성별 그룹 모형

남자 여자로 두 그룹을 나눈 뒤 거기서 각각 LCA 를 하였을 때 모형은 기존 전체에서 LCA 한 모형과 비교적 유사했다. 두 그룹 모두 각각 Class3 에서 가장 낮은 BIC 를

가졌기 때문에 여자 남자 모두 3 가지 범주를 가진다고 판단하고 표를 만들어 보았을 때 [표 9]와 같다. 전반적으로 집단간 비율과 집단의 유형이 크게 많이 다르지는 않지만, 여성은 낮은 요금을 매우 선호 하고 기본적으로 요금이 낮아야 한다는 생각이 전제 되어 있다. 반면 남성은, 요금이 중요하긴 하지만 요금보다는 광고 개수에 조금 민감하게 반응 하는 것을 알 수 있다. 따라서 성별 간 차이가 조금은 있긴 하지만 전체로 분석했을 때 뚜렷하게 갈렸던 반면에 성별을 나누자 조금씩 잠재 변수가 혼재 되어있어 해석이 유의하게 보이지는 않음을 알수있다.

문항	범주	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
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%
선호하는 광고 요금제 선택지	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%
---------------------------------------	--------	-------	--------	-------	--------	--------

[표 9] proportion for LCA by gender

## VI. 논의 및 결론

### 1. 최종 결론 (CONCLUSION)

전처리한 데이터 셋에서 7 개의 범주형 변수를 사용하여 3 가지 잠재집단으로 나누는 모형이 이 프로젝트에서는 가장 좋은 모델로 보인다. 성별로 나누면 더 뚜렷하게 차이가 나기를 기대 했으나 잠재 변수가 혼재되어 있는 모습이 보였다. 따라서 전체 데이터에서 결과를 보는 것이 더 잠재변수의 탐지가 쉽다. 3 가지 집단은 크게 요금과 광고 개수의 중요도에 따라 나뉘었으며 전반적으로 3 집단 모두에게 요금이 영향을 미치고 있음도 알 수 있다. 따라서 넷플릭스 현 사용자들은 광고요금제 서비스를 요금이 많이 낮아지면 어느정도 이용 의향이 있음을 알 수 있으며 일부 소수 집단에서는 광고유형을 더 중요하게 생각 하고 있다고 해석해 볼 수있다.

### 2. 한계점 (LIMITATION)

데이터 전처리 과정상, 넷플릭스 사용자를 대상으로 했기 때문에 다른 플랫폼만 쓰는 구독자나 구독하지 않는 사람들에 대해 광고요금제에 대한 입장을 알 수 없게 되었다. 또한 모든 다른 변수들의 조합으로 시도 해보지는 않았기에 다른 변수들의 조합으로 하면 결과가 더 유의하게 나왔을 수도 있다. 나이별로 나누면 유의했을 수도 있으나 데이터 개수가 전처리 후 510 개로 줄게 되어 크기가 너무 작아져서 identifiable 문제로 결과를 보기 어려웠다.

### 3. 후속 연구 (FURTHER STUDY)

실제 넷플릭스에서 시행된 광고요금제의 형태는 시작 전과 도중에 15 초 또는 30 초 길이의 광고를 재생 (시간당 평균 4~5 분)과 함께 기존의 베이식 요금제 보다 약 58% 할인된 5500 원 요금제를 내놓았다. 대부분의 이용자가 요금에 더 민감하게 반응 함에 따라 시행된 결과라고 생각된다. 따라서 실제 시행된 광고 요금제로 전환한 넷플릭스 이용자들에 대해서 분석하여 이 프로젝트에서 분석한 결과와 비교 하는 연구를 해볼 수 있겠다.

## 참고문헌

---

Finite Mixture Models Geoffrey J. McLachlan, Sharon X. Lee, and Suren I. Rathnayake Annual Review McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, 6, 355-378.

Masyn, K. E. (2013). 25 latent class analysis and finite mixture modeling. *The Oxford handbook of quantitative methods*, 2, 551.

McCutcheon, A. L. (1987). *Latent class analysis* (No. 64). Sage.

Virtanen, Iita M., et al. "Occupational and genetic risk factors associated with intervertebral disc disease." *Spine* 32.10 (2007): 1129-1134.

성미영(Sung, Mi Young);서병태(Seo, Byung Tae);서혜린(Seo, Hye Rin). (2022). COVID-19 와 유아 재난안전에 대한 사회적 이슈 분석: 유튜브 동영상 댓글을 중심으로. *아동과 권리*, 26(1), 115-131.