

## 뮤지컬 예매점유율별

# 주요 취소위험요인과 예매생존확률 비교분석

---

예매 후 취소 발생 시기와 변수 영향력 분석을 통한 취소 위험 지표 제시  
실관객수 예측 및 취소위험군집 이탈 방지 전략 제안

# ... Index ...

1

주제 선정

3

분석 방법

5

기대효과

2

데이터

4

분석 결과

# 1

## 뮤지컬 공연의 예매 취소 시기

# 2

## 취소에 영향을 미치는 위험인자

공연 및 공연시설별 특성, 예매 정보, 날씨, 교통, 출연진 인지도 등 **종합적 고려**  
생존분석을 이용해 예매 이후 **취소가 발생하는 시기** 예측  
여러 설명변수 중 위험인자 색출해 **취소에 대한 영향력** 확인

투자자 입장

공연업계의 선투자 문화 문제점



생존분석을 통해 각 변수의 위험률 계산 후  
취소 위험 지표 제시함으로써 정보비대칭현상 완화

제작자 입장

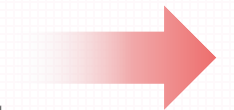
고객의 초기 투자금이 드는 뮤지컬



각 예매데이터 기반 기간별 생존 확률 예측을  
통한 실관객수 예측 가능 & 정확한 예산 수립

관객 입장

불완전 대체재를 고려한  
관객 이탈 분석의 중요성



이탈 확률 높은 관객군 파악을 통한  
군집별 효율적 마케팅 전략 제시

## 생존분석 방법의 차별성

## 1. 이진분류모델과의 비교

생존분석은 취소 여부 뿐만 아니라  
각 고객의 특성에 따라 다르게 나타나는  
취소까지의 과정에 초점을 맞추어 분석 가능

2. 예매와 취소를 모두 고려한  
순수요 예측 가능

기존의 수요 예측 모델은 총예매자수와  
같이 예매 행위에 초점 BUT  
구매한 후 취소하는 행위는 불확실성에  
근거한 행동이므로 분석 어려움

## 분석결과의 독창성

## 1. 시점에 따른 취소확률 예측

생존분석을 통해 예매부터 취소까지의  
시점별로 취소가 발생할 확률 정보 제공

## 2. 위험인자의 수치화

각 변수별로 취소에 영향을 미치는  
위험비(Hazard Ratio)로 추정 가능

분석기간 2022.04.18 ~ 2022.12.31

코로나 이펙트에 의해 편향된 분석 결과가 나오지 않도록 통제하기 위해  
거리두기 정책이 해지된 이후 데이터 사용

분석지역 서울 전역

2022년 기준 전국 공연 건수의 44%가 개최되고 전체 공연 수요의 67.3%  
티켓 판매액의 75.8%가 이뤄지는 서울

분석장르 뮤지컬

2022년 티켓판매수 581만 매와 티켓판매액 3262억원을 기록  
공연시장 총 티켓 판매량과 판매액의 각각 56%, 78%를 차지하며  
가장 큰 수요를 가지는 동시에 유료 취소율이 가장 높은 장르

## 뮤지컬 장르 내 공연 간 양극화 현상 ↑

같은 뮤지컬 장르 내에서도 공연별 분석 결과 다를 것



장르 내 세부 분류를 위해 공연의 인기를 판단하는 기준으로 예매점유율 사용

$$\text{예매점유율}^* = \frac{\text{해당 공연 순예매건수}}{\text{전체 뮤지컬 공연 순예매건수}}$$

Log Rank Test를 통해 분류된 세 가지 그룹의 생존함수에 차이가 있다는 것 확인

예매점유율	test_statistic	p-value	결과
상 vs 중	4653.51	<0.005	귀무가설 기각
상 vs 하	2793.93	<0.005	귀무가설 기각
중 vs 하	251.74	<0.005	귀무가설 기각

\* 공연별 예매점유율 기준 그룹

상 : 예매점유율 상위 33% 공연에 대한 예매 데이터 (인기도 상)

중 : 예매점유율 상위 34%~66% 공연에 대한 예매 데이터 (인기도 중)

하 : 예매점유율 상위 67%~100% 공연에 대한 예매 데이터 (인기도 하)

반응변수		
변수명	type	변수 설명
delta	Category	예매생존여부 [0: 예매, 1: 취소]
y	Numeric	예매생존기간 [예매: 공연일-예매일+1, 취소: 취소일-예매일+1]

각 그룹 내의 delta count			
예매점유율	상	중	하
전체	1161738	1218911	1191146
delta = 0	799276	810576	824274
delta = 1	362462	408335	366872

예매를 출생으로, 취소를 죽음으로 설정

예매에서 취소까지의 시간은 관객의 수명으로 해석하여 생존분석을 진행



- Delta = 1 (예매를 취소한 관객)
- Delta = 0 (예매를 취소하지 않은 관객)

관측시작시점  
(2022.04.18)

관측종료시점  
(2022.12.31)

B관객  
예매시점



취소발생시점

Time

A관객  
예매시점

공연관람시점

$T = t_0$

$T = t$

설명변수	
공연시설 및 공연장 관련 변수 (4개)	공연 관련 변수 (21개)
공연시설인기*, 좌석수, 시설합*, 주차시설	소요시간, 팔로워합, 수상실적개수, 매출액, 좌석등급개수, 세부장르명, 단독판매여부, 할인종류개수, 공연별전송사업자코드개수*, 공연_요일*, 공연_평일여부*, 공연_월*, 공연_오전오후*, 공연횟수*, 순예매건수, 좌석점유율, 아동공연여부, 축제여부, 내한공연 여부, 취소여유기간*, 예매가능공연개수*
외부변수 (5개)	예매 및 취소 관련 변수 (10개)
강수, 미세먼지등급, 불쾌지수*, 대중교통개수*, 일일확진자	성별, 예매금액, 할인율, 예매_요일*, 예매_평일여부*, 예매_월*, 결제수단코드, 예매방식코드, 할인종류코드, 취소여유기간

\*로 표시된 변수는 본 팀이 새로 추가한 파생변수임

## Cox PHM

(Cox Proportional-Hazards Model)

1. 생존 분석 및 위험 회귀 분석을 위한

전통적 통계 모델

2. 사건 발생과 시간 간격 사이의 관계를

조사하고 이해하는 데 사용

→ 특정 변수가 생존 시간에 미치는 영향을

평가하는 데 유용

vs.

## DeepSurv

(Deep Learning for Survival Analysis)

1. 비선형 관계 또는 복잡한 구조를 가진

데이터에 적절

2. 신경망 아키텍처를 이용해

생존 분석을 수행하는 모델

→ 다양한 입력 특성을 고려하여

생존 시간 예측을 개선하는 데 도움

DeepSurv 환경표

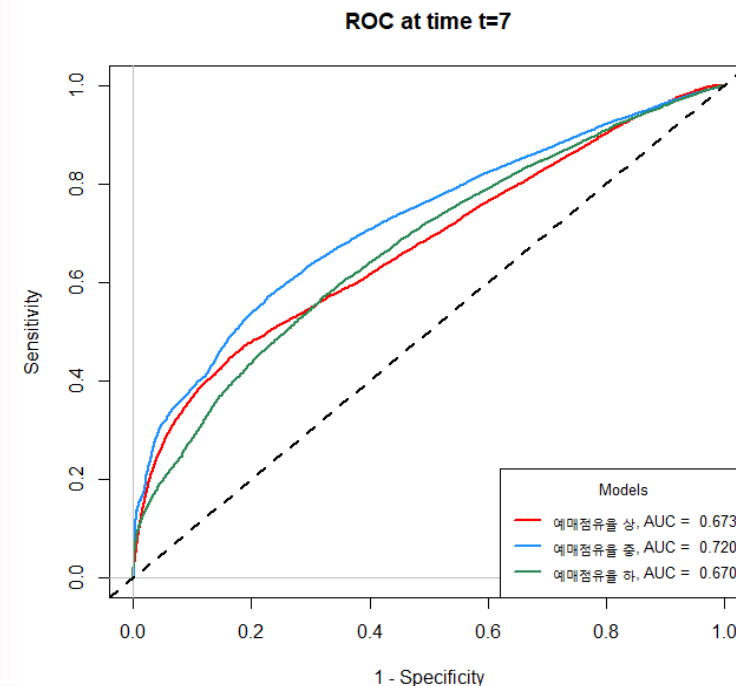
환경	anaconda python 3.7.13 pytorch 1.13.1 pycox 0.2.3 scikit-survival 0.21.0 NVIDIA GeForce RTX 3080		
	예매점유율 상	예매점유율 중	예매점유율 하
node 개수	[1024, 1024]	[1024, 1024]	[1024, 1024]
dropout	0.1	0.1	0.1
batch size	256	256	128
learning rate	0.0001	0.0001	0.0001
epoch	512	256	256

1에 가까울수록  
Best

0에 가까울수록  
Best

1에 가까울수록  
Best

예매점유율	Model	C-index	BS (t=7)	Time-dependent AUC (t=7)
상	Cox PHM	0.6539	0.2483	0.6737
	DeepSurv	0.7942	0.1097	0.8200
중	Cox PHM	0.6839	0.3814	0.7207
	DeepSurv	0.8063	0.1075	0.8357
하	Cox PHM	0.6640	0.3438	0.6702
	DeepSurv	0.8031	0.1010	0.8170



생존예측모델 성능 평가 지표 C-index, Brier Score, Time-dependent AUC 기준

DeepSurv가 Cox PHM에 비해 예측 성능이 높음

# 예매점유율 그룹별 위험인자 도출 (Cox PHM)

예매점유율 상, 중, 하 그룹 Cox PHM exp(coef) table							
Variable	상	중	하	Variable	상	중	하
groupB	1.243	1.09	1.143	세부장르명뮤지컬	0.008		1.126
groupC	1.114	1.125	1.184	세부장르명악극			0.927
groupD	0.91	1.114	1.024	소요시간	0.781	0.997	1.001
groupE	0.728	1.088	0.795	수상실적개수		1.054	
강수	1.005	0.995	1.001	순예매건수		1.0	1.0
결제수단코드2	0.603	0.713	0.803	시설합		1.016	0.995
결제수단코드4	0.93	1.102	1.028	아동공연.여부True		0.654	0.77
결제수단코드5	0.609	1.081	1.297	예매_요일Mon	1.027	0.906	1.021
결제수단코드99	0.664	0.802	0.863	예매_요일Tues	0.9	0.912	1.019
...(생략) ...							
공연시설인기F			1.074	좌석수			1.0
공연횟수	0.919	0.999	0.998	좌석점유율		0.396	0.452
교통개수	1.367	1.009	1.004	주차시설		0.775	1.185
내한공연.여부True			0.546	축제.여부True			1.206
단독판매여부True	6.098	1.24	0.997	팔로워합		1.0	1.0
매출액		1.0		할인율	0.001	0.256	0.13
미세먼지등급1	0.964	1.017	1.009	할인종류개수	1.02	1.004	1.004
미세먼지등급2	1.115	0.957	0.947	할인종류코드4		1.013	
미세먼지등급3		1.33	0.895	할인종류코드99		0.975	1.183
불쾌지수	1.004	0.993	1.0	취소여유기간	0.994	0.983	0.986
성별1	0.906	1.032	1.412				
성별2	0.988	1.06	1.604	p<0.05에서 유의하지 않은 변수 회색 음영으로 표시			

## 예매점유율 그룹별 위험인자 도출 (Cox PHM)

예매점유율 상, 중, 하 그룹 Cox PHM exp(coef) table	
Variable	상
groupB	1.243
groupC	1.114
groupD	0.91
groupE	0.728
강수	1.005
결제수단코드2	0.603
결제수단코드4	0.93
결제수단코드5	0.609
결제수단코드99	0.664
...(생략) ...	
공연시설인기F	0.074
공연횟수	0.919
교통개수	1.367
내한공연여부True	0.000
단독판매여부True	6.098
매출액	0.000
미세먼지등급1	0.964
미세먼지등급2	1.115
미세먼지등급3	0.000
불쾌지수	1.004
성별1	0.906
성별2	0.988

### 1. 해석 (Category)

i) 변수의 위험률이 1보다 높은 경우

기준 카테고리에 비해 해당 카테고리일 때 위험률이  $(HR-1)*100\%$  로 증가

ii) 변수의 위험률이 1보다 낮은 경우

기준 카테고리에 비해 해당 카테고리일 때 위험률이  $(HR-1)*100\%$  로 감소

### 2. 해석 (Numeric)

i) 변수의 위험률이 1보다 높은 경우

해당 변수를 한 단위 추가했을 때 위험률이  $(HR-1)*100\%$  로 증가

ii) 변수의 위험률이 1보다 낮은 경우

해당 변수를 한 단위 추가했을 때 위험률이  $(HR-1)*100\%$  로 감소

→ 위험률이 1에서 많이 떨어진 변수를 변동성이 큰 위험인자로 판단

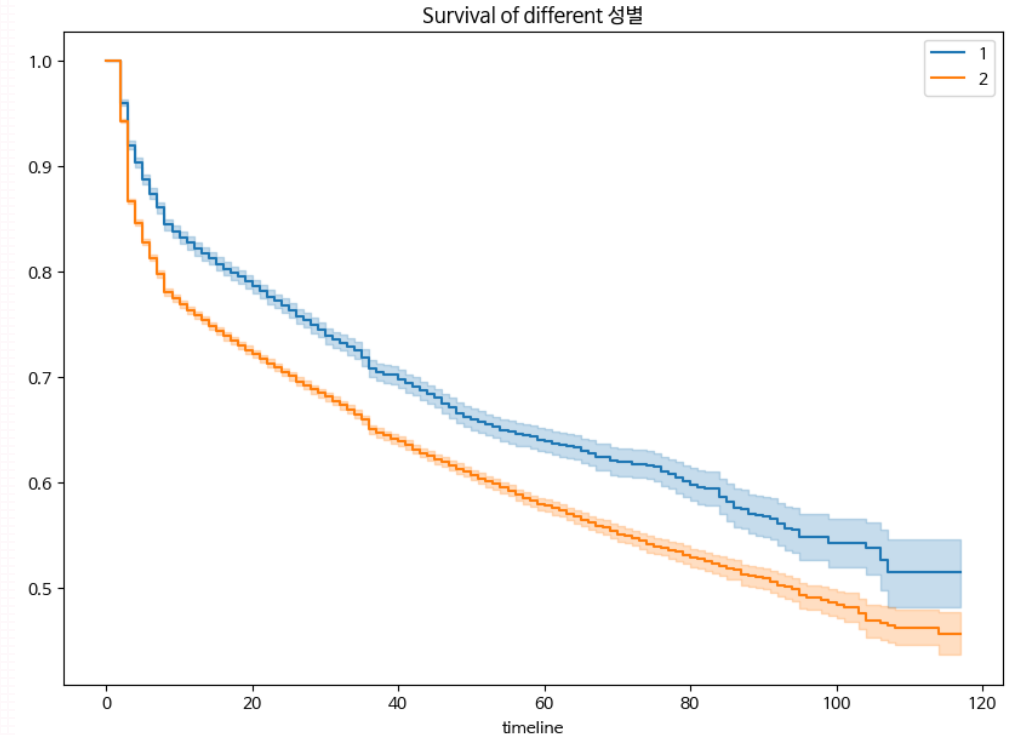
## 변수별 위험률 해석 (전체 그룹, 성별)

### Hazard Ratio

예매점유율 상, 중, 하 그룹 Cox PHM exp(coef) table			
Variable	상	중	하
성별1	0.906	1.032	1.412
성별2	0.988	1.06	1.604

기준이 되는 성별 0은 불명이기 때문에 제외하고 비교 시  
 위험비가 세가지 그룹 모두에서 더 높은 것을 확인할 수 있음  
 → 남성에 비해 여성이 예매했을 시 취소 위험이 높다고 해석 가능

→ 여성 예매자의 초기 이탈 주의 요망



모든 생존기간에 걸쳐 남성(1)의 생존율이  
 여성(2)의 생존율보다 높음  
 → 남성이 여성에 비해 취소를 덜 하는 경향 0

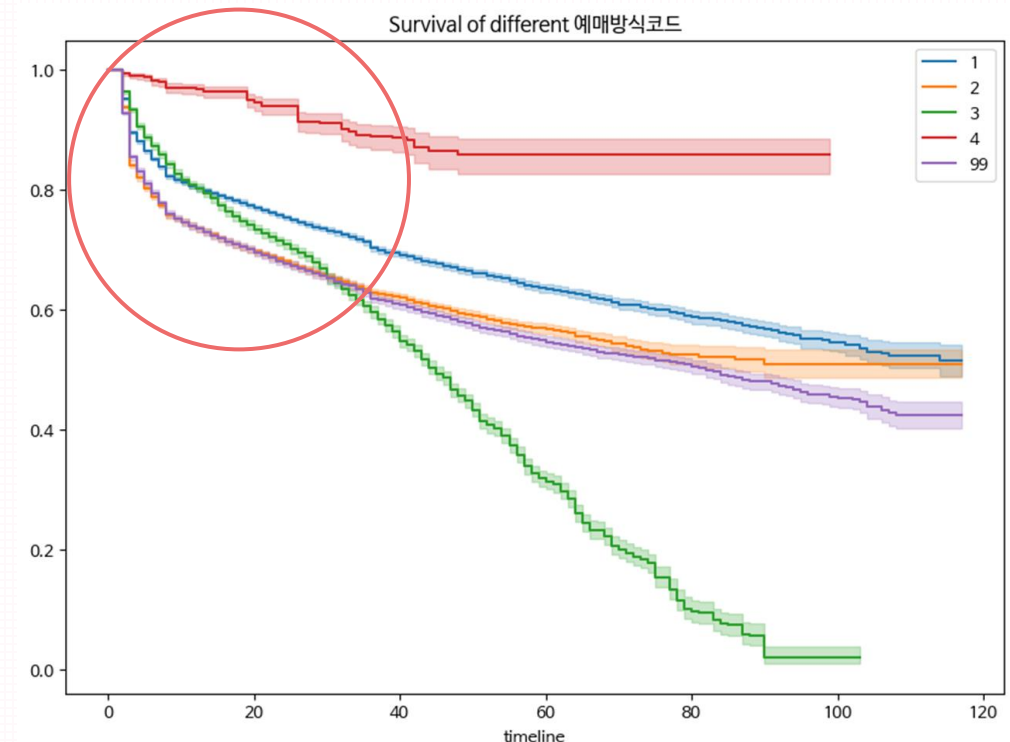


## 변수별 위험률 해석 (상 그룹, 예매방식코드)

### Hazard Ratio

예매접유율 상 그룹 Cox PHM exp(coef) table					
		예매방식코드			
Variable	1(웹)	2(모바일)	3(현장)	4(전화)	99(기타)
Exp(coef)	Baseline	1.257	0.959	0.469	1.366

기준예매방식코드1에 비해 3일 때 95.9%, 4일 때 46.9%로  
취소 위험이 낮아지고, 2일 때 125.7%, 99일 때 146.6%로  
취소 위험이 높아짐



전화(4)로 예매한 경우 생존율이 높은 상태로  
유지되는 경향이 있으며, 현장(3)에서 예매한 경우  
35일 이후 생존율이 급락하는 경향 0

→공연일 전에 현장 예매를 오픈한 경우 35일을 넘지 않도록 하는 것이 취소 예방에 도움 0

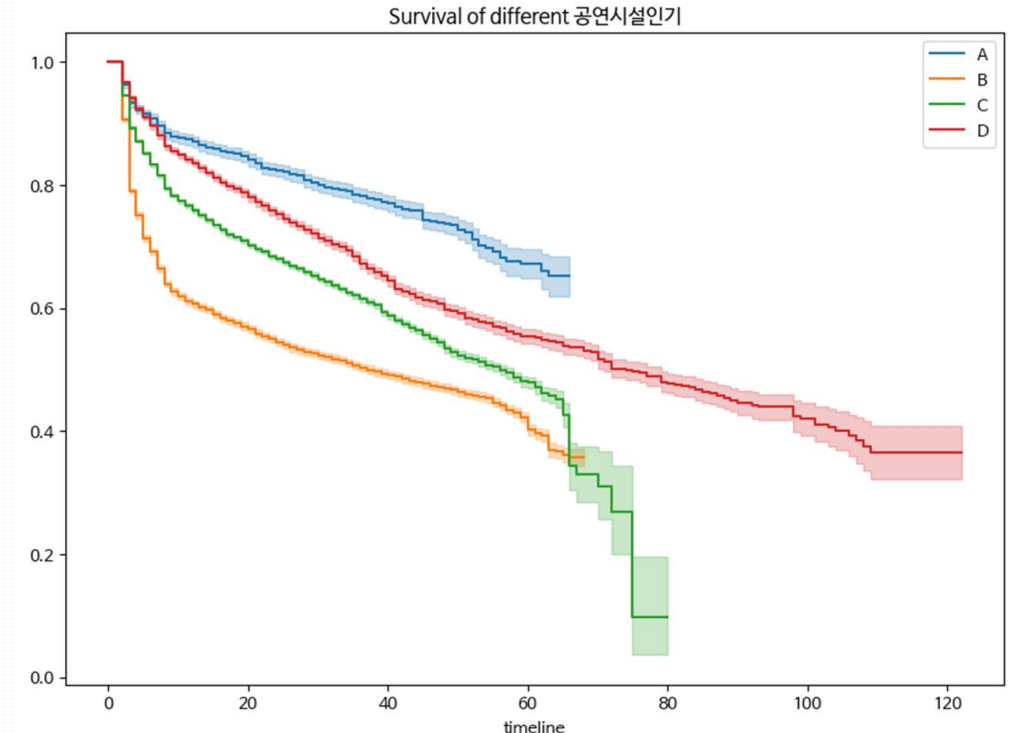
## 변수별 위험률 해석 (중 그룹, 공연시설인기)

### Hazard Ratio

예매점유율 중 그룹 Cox PHM exp(coef) table				
	공연시설인기			
Variable	A	B	C	D
Exp(coef)	Baseline	3.448	2.066	1.615

기준 공연시설인기 A에 비해 B,C,D는 최소 위험이 높음 →  
공연시설인기 B, C, D, A 순으로 더 높은 최소 위험률을  
가진다고 해석할 수 0

→ 공연시설인기 B그룹의 경우 가장 최소의 위험이 높으므로 해당 그룹에 포함되는 공연시설의 특징을 파악할 필요 0



가장 인기있는 공연시설(A)의 생존함수가 가장 높게 위치하며,  
그 다음으로 가장 인기도가 낮은 공연시설(D), 공연시설인기 C,  
B그룹 순으로 생존함수가 위치

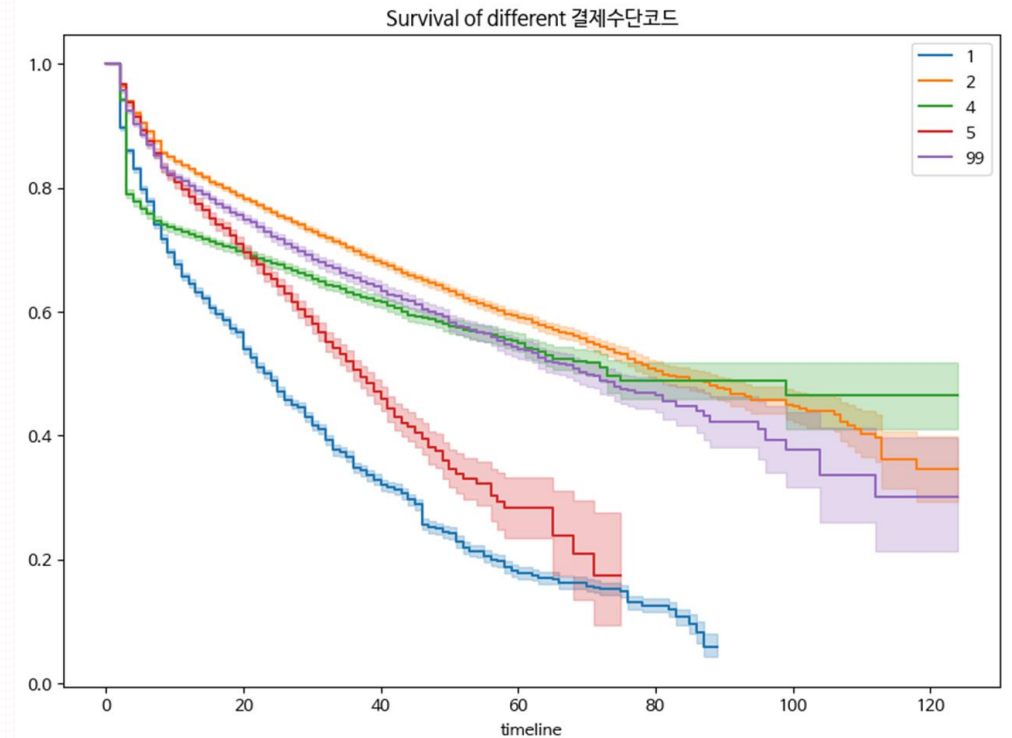
## 변수별 위험률 해석 (하 그룹, 결제수단코드)

### Hazard Ratio

예매점유율 중 그룹 Cox PHM exp(coef) table					
	결제수단코드				
Variable	1 (현금)	2 (카드)	4 (무통장)	5 (다중결제)	99 (기타)
Exp(coef)	Baseline	0.803	1.028	1.297	0.863

결제수단코드 1에 비해 취소 위험이 결제수단코드 2는  
19.7%, 결제수단코드5는 20%가 더 높음  
→ 다중결제, 현금, 카드 순으로 취소 위험이 높다는 의미

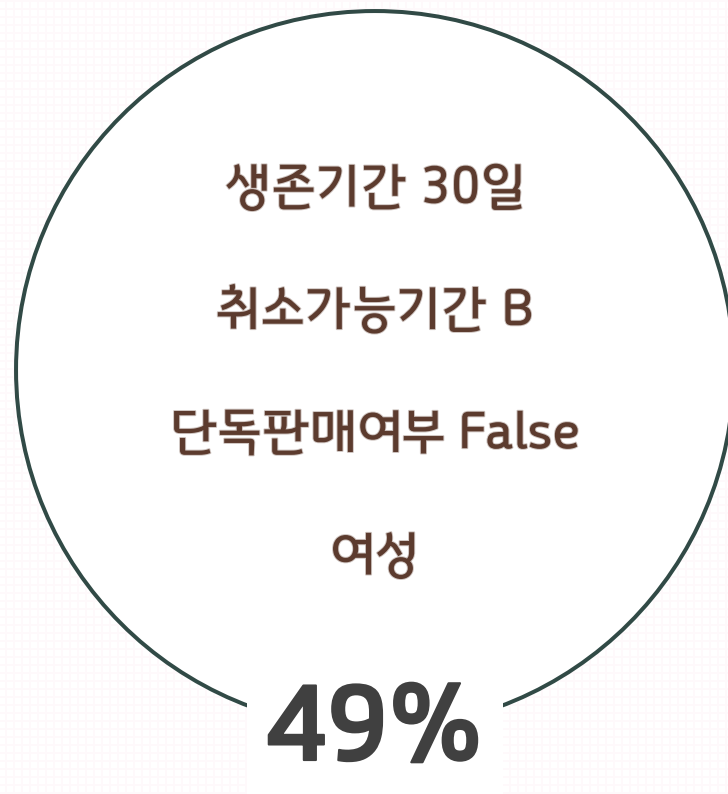
→ 무통장입금 방식으로 결제한 고객이 3일 이내에 취소하지 않았다면  
해당 고객은 예매 상태를 유지할 가능성이 높으므로 이탈 위험 관리 고객에서 제외



무통장입금(4)의 경우 예매 직후 생존율이 급락 →  
그 이후에는 생존율을 유지하는 경향 0

## 페르소나별 특징 및 생존함수 추론 (DeepSurv)

### 최소**위험**군



vs.

30일 후  
평균 생존 확률

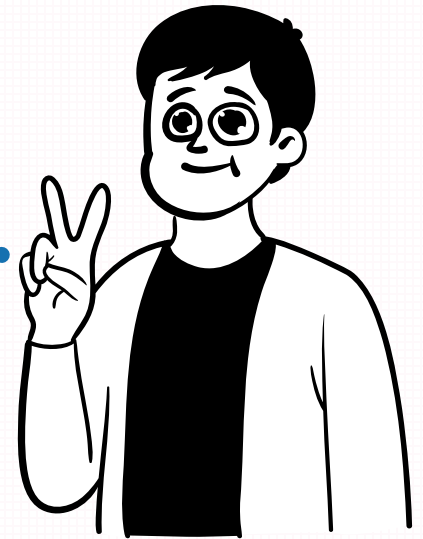
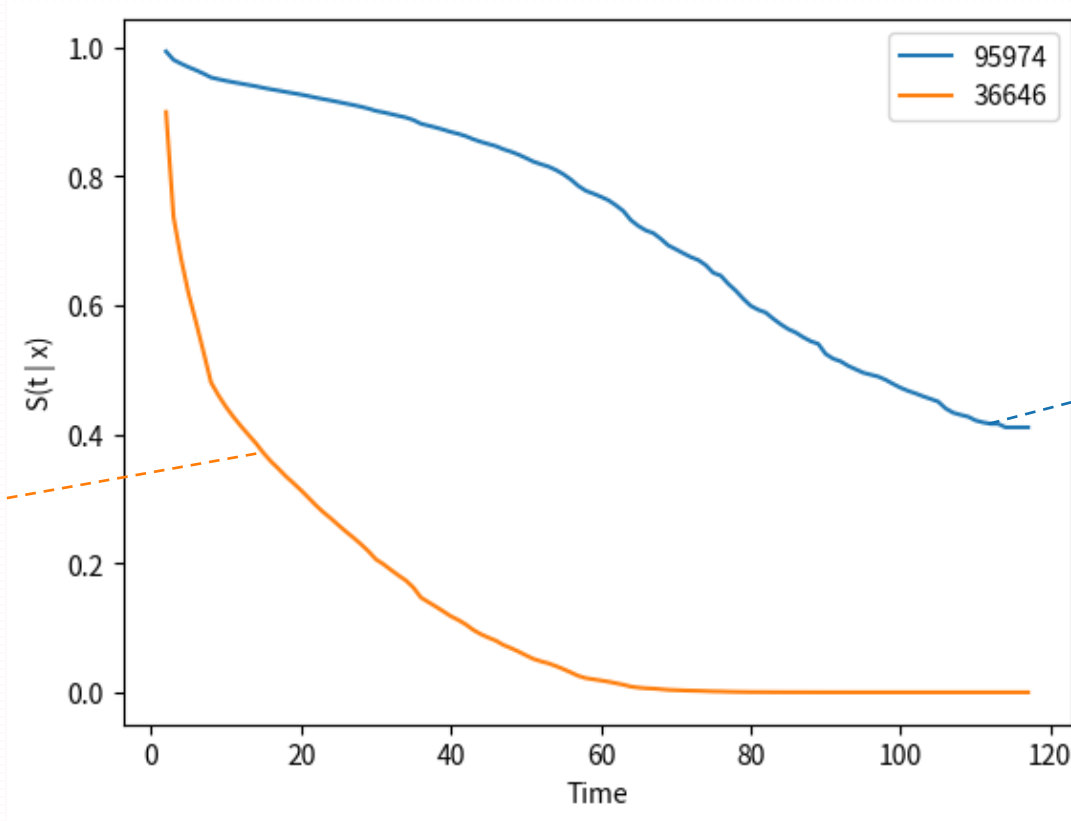
### 최소**안전**군



## 페르소나별 특징 및 생존함수 추론 (DeepSurv)



이화연 (가명, 27)



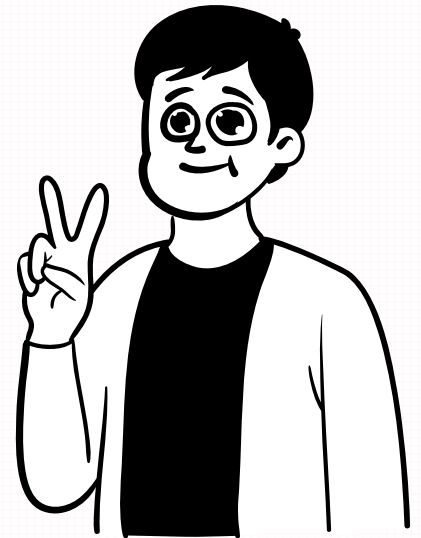
채지평 (가명, 39)

## 페르소나별 특징 및 생존함수 추론 (DeepSurv)



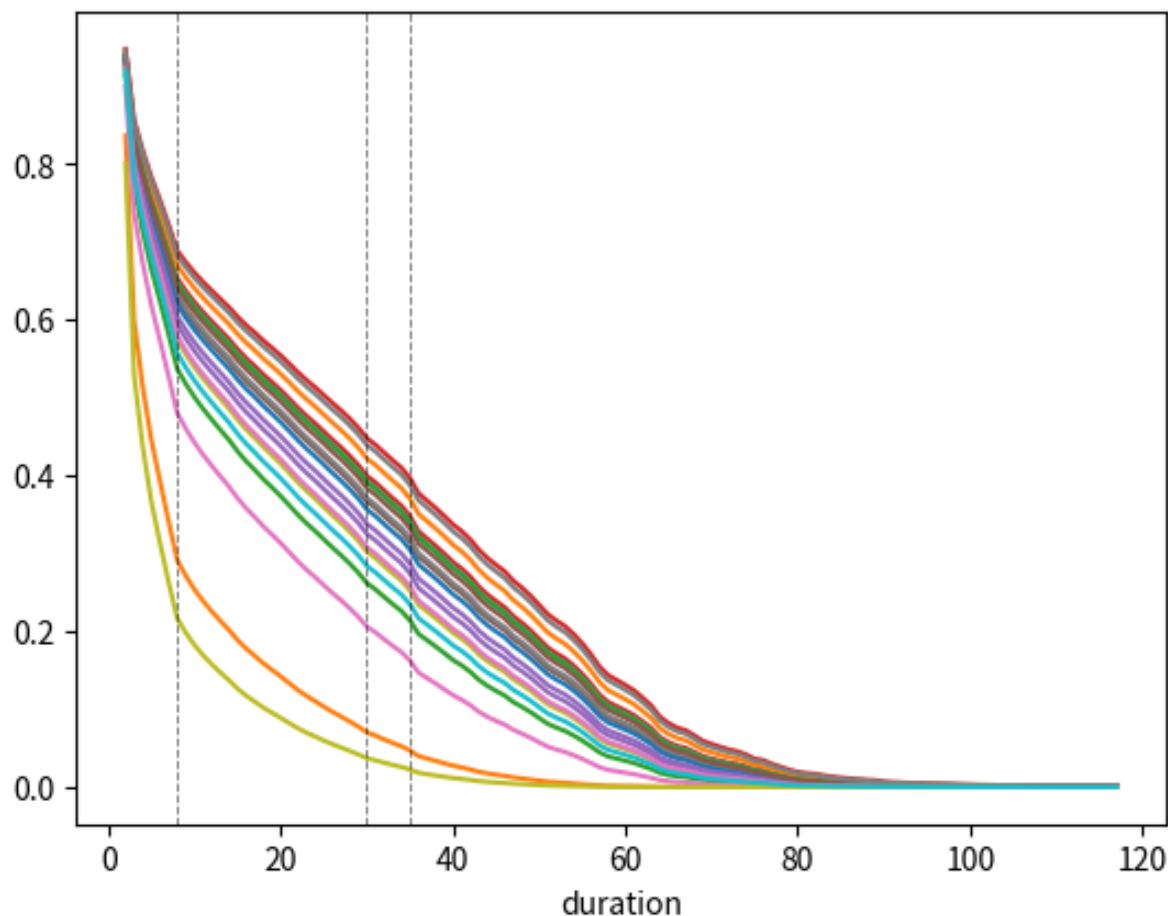
이화연 (가명, 27)

105000	-----	예매금액	-----	60000
30551	-----	팔로워합	-----	195312
C	-----	공연시설인기	-----	A
토요일	-----	공연_요일	-----	수요일
61.71	-----	불쾌지수	-----	78.11
1242	-----	좌석수	-----	3022
0.73	-----	좌석점유율	-----	0.29
10827	-----	일일확진자	-----	22993
6	-----	할인종류개수	-----	3
10	-----	시설합	-----	4



채지평 (가명, 39)

## 최소위험 군집의 생존함수 추론 (DeepSurv)



그래프의 수직선은  $t=8$ ,  $t=30$ , 그리고  $t=35$ 인 경우인데 이 시점을 기준으로 유독 생존확률이 눈에 띄게 달라지는 것을 확인할 수 있음

각 시점에 맞추어 공연별 타겟 관객군의 특징을 파악한다면 최소 확률을 더 낮출 수 있을 것으로 기대

투자자 입장

투자자와 제작자간 신뢰관계 구축 및 공연 투자 활성화



제작자 입장

최적의 예산 기획을 통한 재무 안정성 강화



관객 입장

공연별 예매자 이탈 방지로 인한 매출 증대 기대



# 경청해 주셔서 감사합니다!

---