

TEAM.4 하이파이브

PROJECT 1

# 호텔 예약 데이터 분석

-매출 안정화 및 리스크 관리 전략 수립-

팀장: 이준수

팀원: 문은아 | 이승윤 | 이지영 | 허대권



# Index >>

01

## Intro

분석 배경  
사용한 데이터 셋 소개  
전처리 과정 소개

02

## Analysis

리드 타임 및 특별 요청 분석  
고객 집단 및 유통 채널 분석  
'환불 불가' 옵션의 높은 취소율 추적  
분석 결과 종합 > 리스크 경고

03

## Conclusion

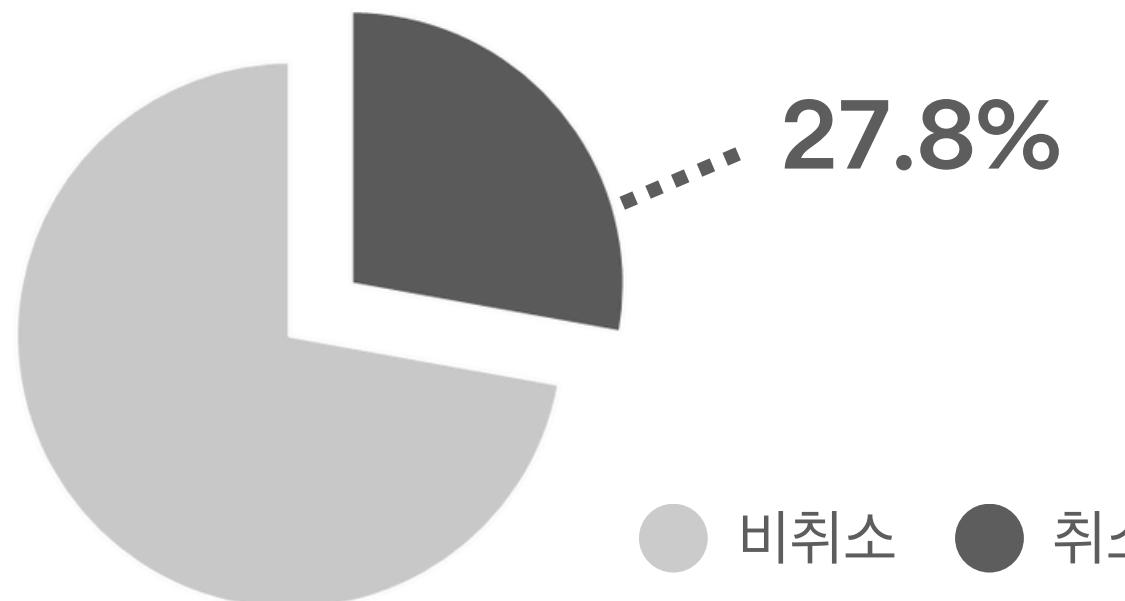
전략적 제안  
분석의 한계점 및 기대효과

# “ 왜 ‘취소’ 데이터에 주목해야 하는가? ”

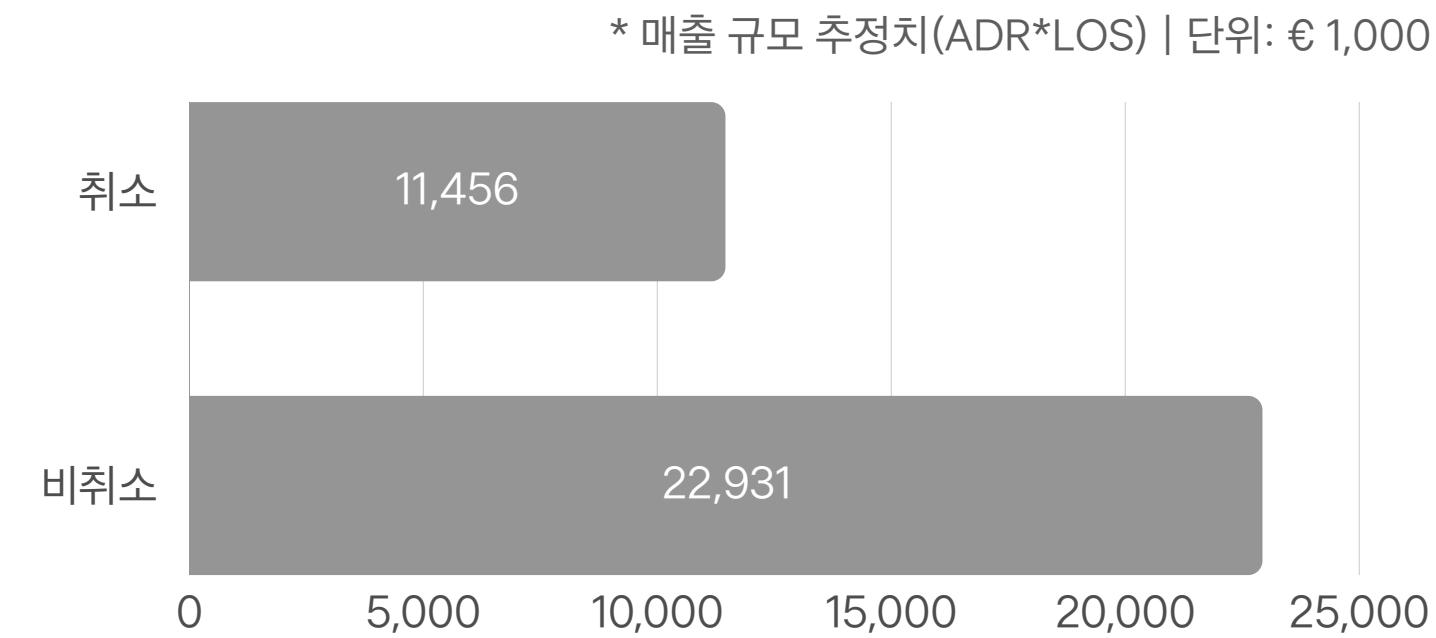
**27.8%** 의 취소율

**196억원** 의 예상 손실액

4명 중 1명 꼴로 예약 취소



기회비용을 제외한 예상 손실액



# Intro >> 데이터셋 소개

Data set

<https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand>

## • Hotel Booking Demand

: 포르투갈의 실제 호텔 데이터 기반 | 2015 ~ 2017 년의 호텔 예약 데이터 수록

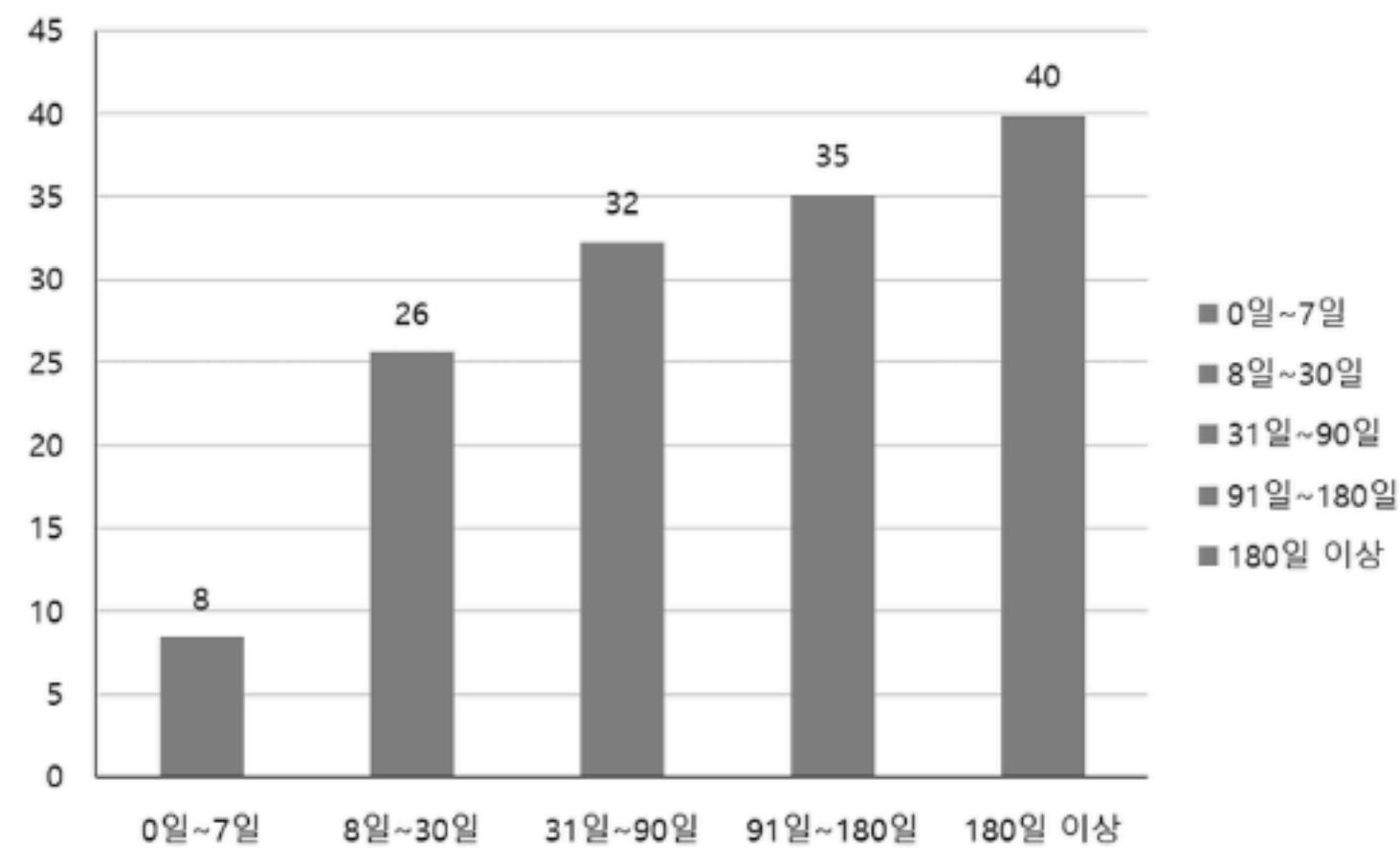
### 데이터 전처리 과정

- $adr \leq 0 \rightarrow$  삭제
- $adults = 0 \rightarrow$  삭제
- $stays\_in\_weekend\_nights$   
+  $stays\_in\_week\_nights = 0 \rightarrow$  삭제
- $children, babies = NULL \rightarrow 0$ 으로 보정

### KPI 설명

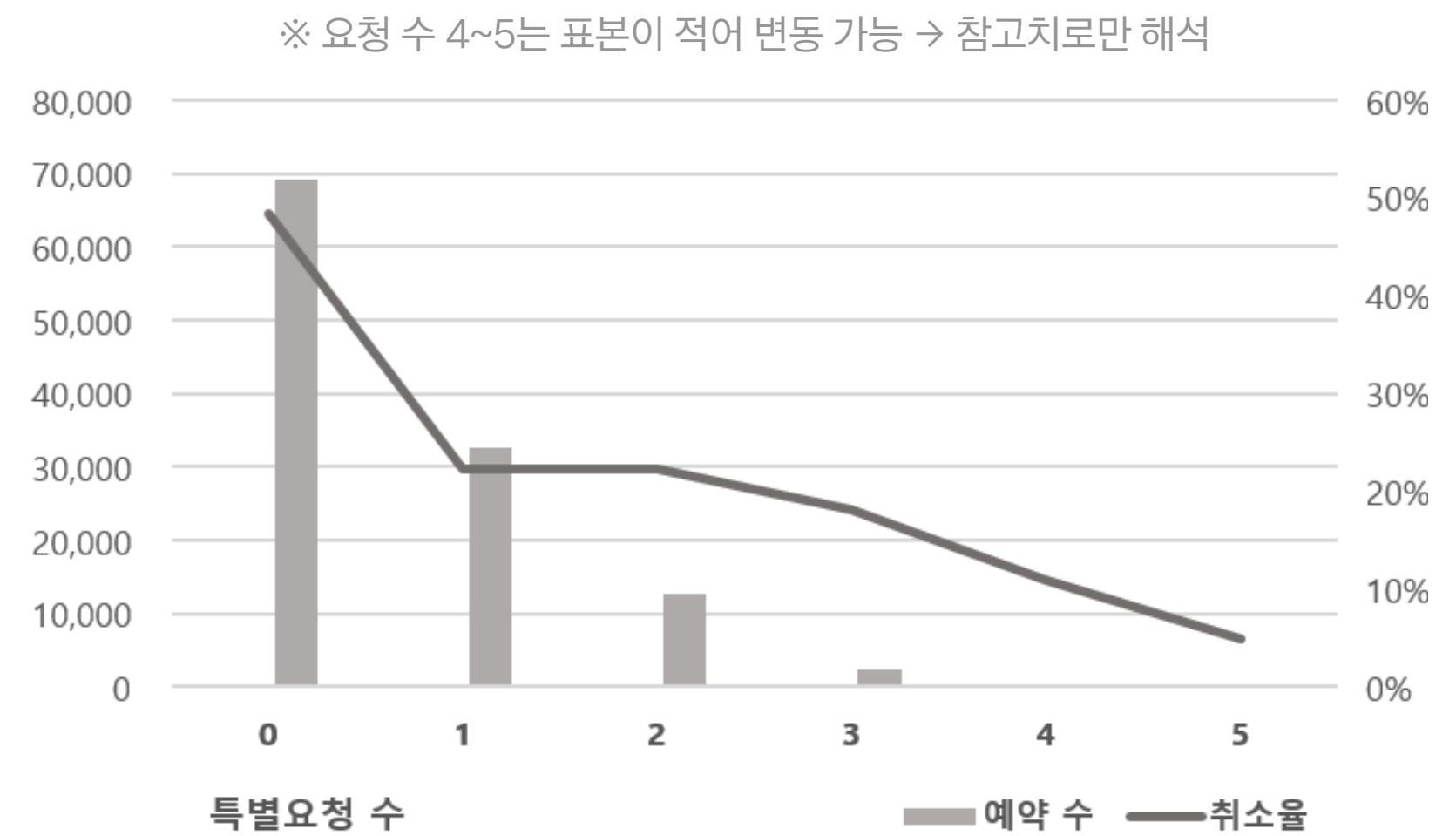
- 예약수 = COUNT(\*)
- 취소건수 = SUM(is\_canceled)
- 취소율(%) = AVG(is\_canceled) × 100
- 평균 객실 단가(ADR) = AVG(adr)
- 숙박일수(LOS) = stays\_in\_weekend\_nights + stays\_in\_week\_nights

## ● 리드 타임과 예약 취소율 간의 상관관계



>> 리드 타임 ↑ 취소율 ↑

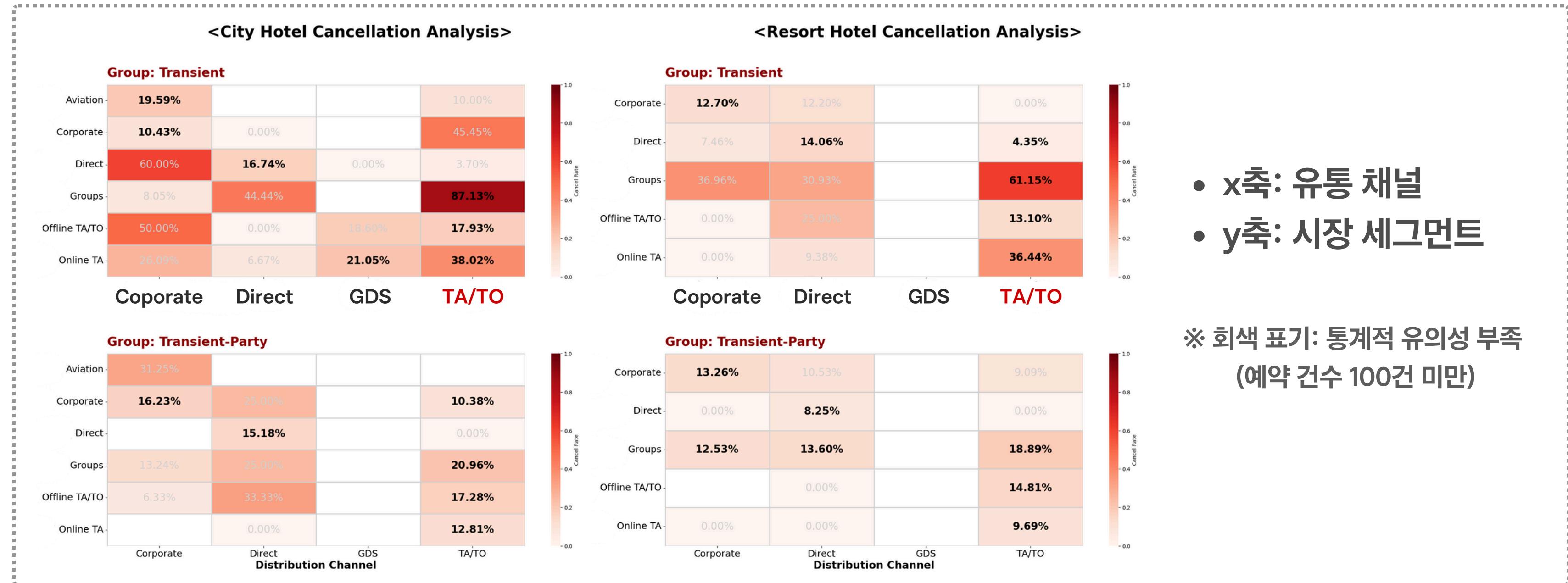
## ● 특별요청과 예약 취소율 간의 상관관계



>> 특별 요청 사항 ↑ 취소율 ↓

# Analysis >> 기본 분석

## ● 고객집단 \* 유통채널 별 취소율 히트맵



>> 개별 고객(Transient)이 여행사(TA/TO) 유통 채널을 통해 예약했을 경우 취소율 ↑↑

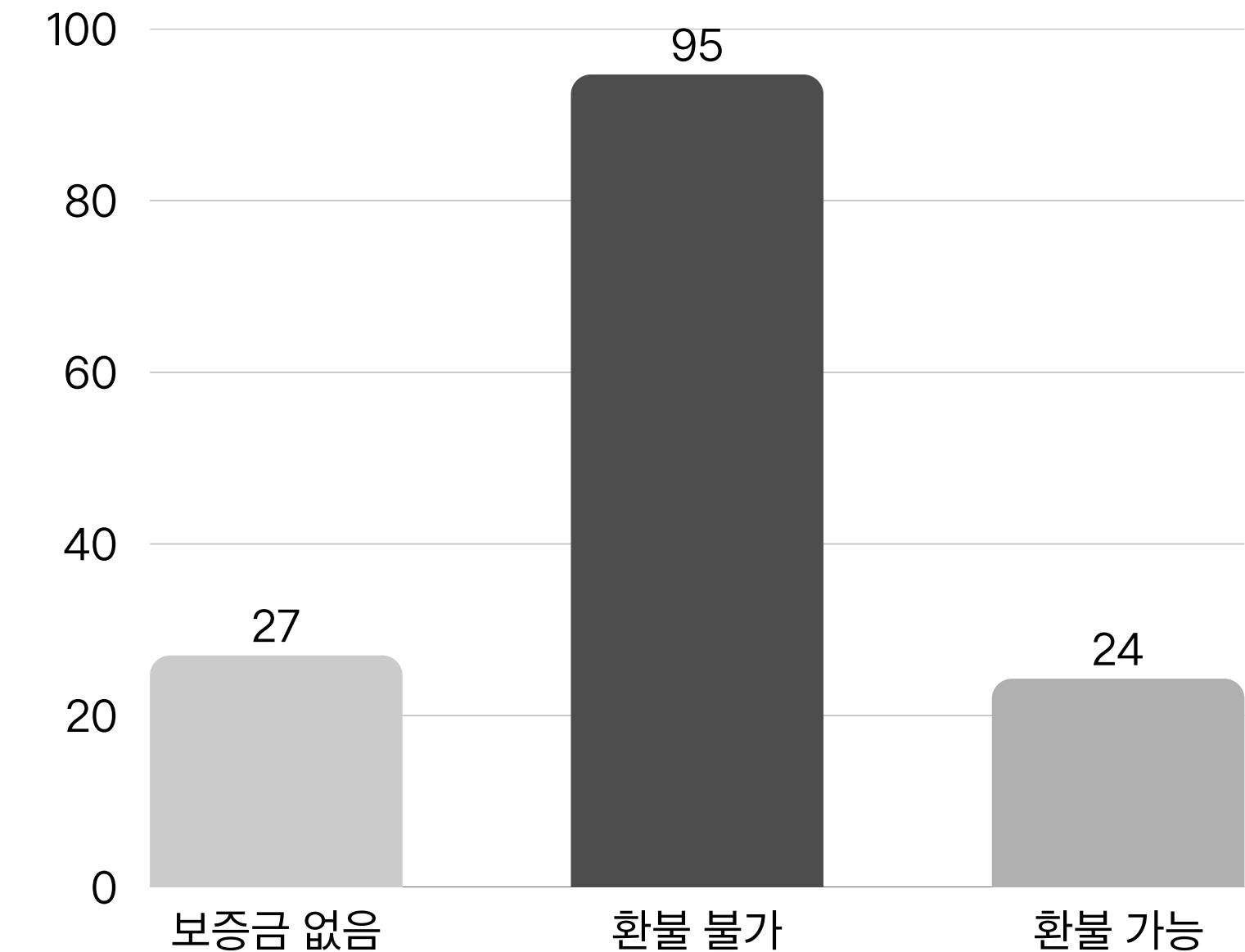
## Analysis >> 발견

Finding

“  
환불 불가의 역설  
”

환불 불가 옵션이 확정 수익을 보장해주지 않음

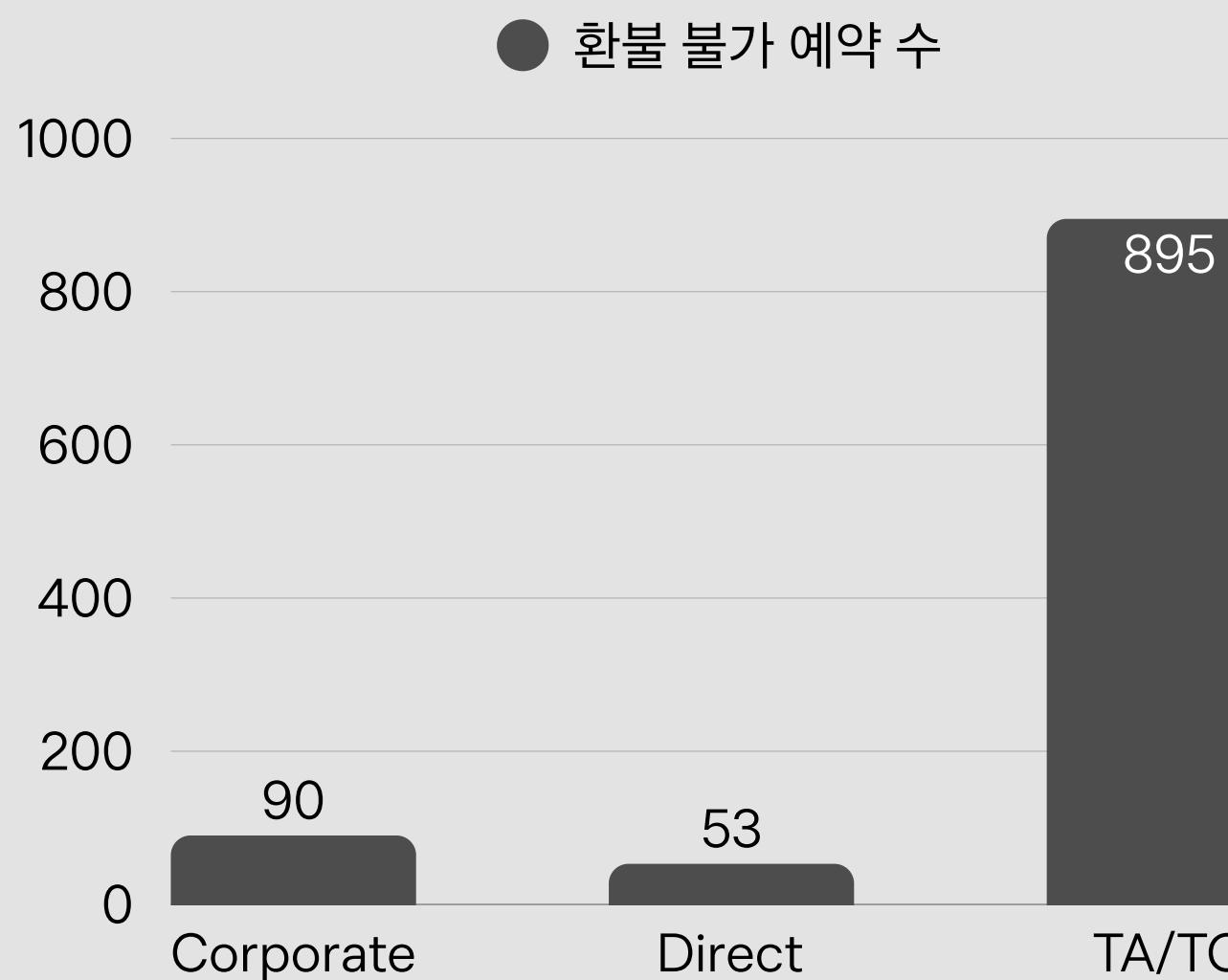
- 보증금 유형별 취소율(%)



✓ '환불 불가'의 보증금 유형에서 가장 높은 취소율 발견

# Analysis >> a. 유통 채널 분석

- 보증금 환불 불가 옵션의 분포



- 보증금 유형 \* 유통 채널 별 취소율

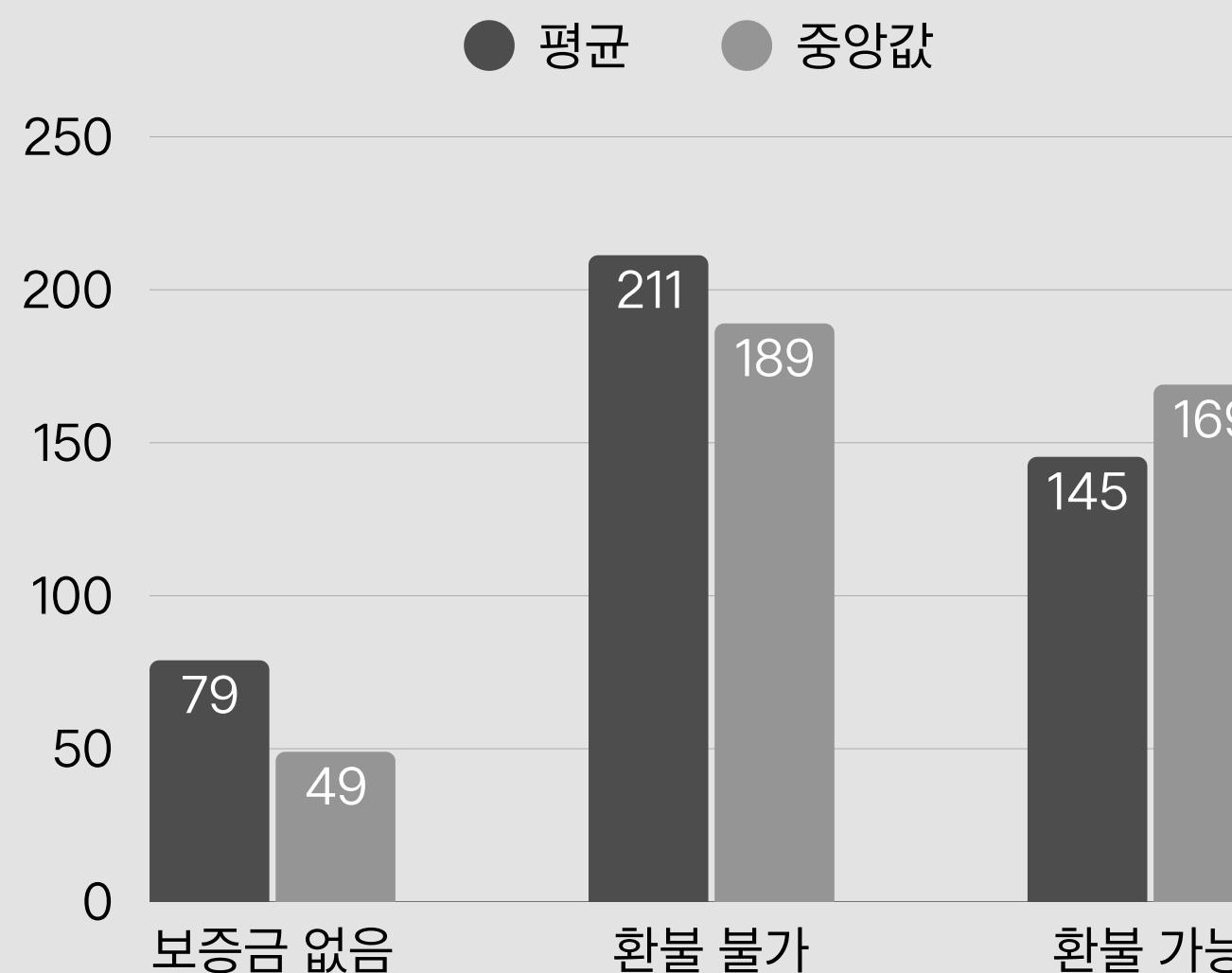
	보증금 없음	환불 불가	환불 가능	총합계
Corporate	11.68%	72.22%	33.33%	12.81%
Direct	14.73%	81.13%	6.56%	14.98%
GDS	20.34%			20.34%
TA/TO	30.33%	97.77%	48.84%	31.23%
총합계	27.01%	94.70%	24.30%	27.83%

\* 회색 표기: 통계적 유의성 부족 (예약 건수 100건 미만)

✓ '환불 불가' 예약의 대부분(86%)은 여행사(TA/TO) 채널에서 발생

## Analysis >> b. 리드 타임 분석

- 보증금 유형 별 리드타임(일)



“**‘환불 불가’ 옵션의 ‘리드 타임 중위값’**  
**: 189일 (대략 6개월)**”

	전세계 호텔 평균	분석 대상 호텔
리드 타임	32 ~ 73일	<u>80 ~ 211일</u>

\* 출처: <https://www.siteminder.com/news/siteminder-hotel-booking-trends-2025/>

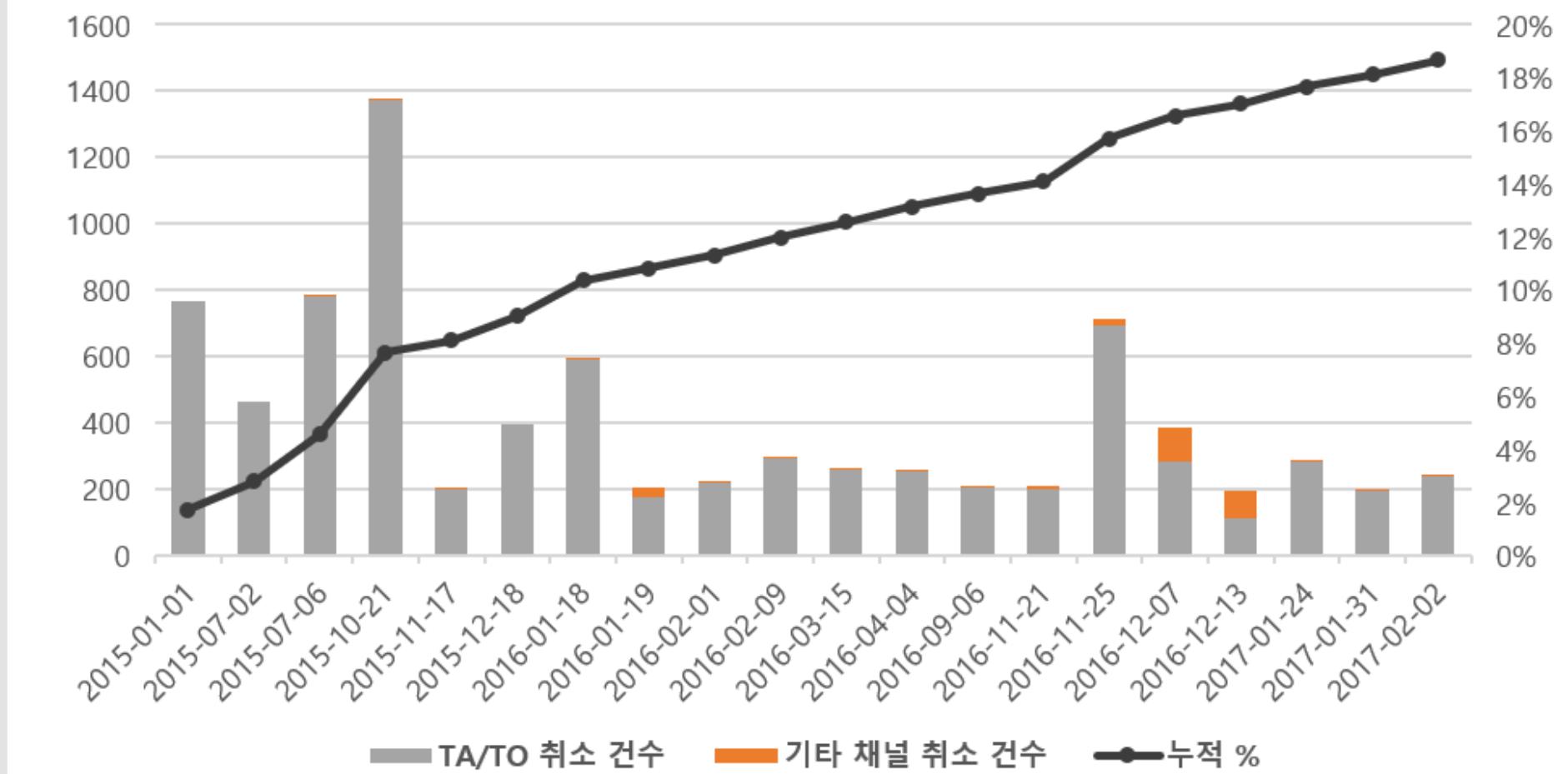
✓ 여행사(TA/TO) 채널에서 5~7개월 전에 객실을 선점했을 가능성 포착

## Analysis >> c. 상태 변경일 분석

# 단 **2.7%**의 기간에  
전체 취소의 **18.6%**가 집중 발생

# 해당 기간 취소량의 **96.4%**가  
여행사(TA/TO)채널에서 발생

- 유통 채널별 일괄 취소 패턴 분석



- x축: 예약 취소 발생량 상위 20개 일자
- y축(좌/막대): 일별 예약 취소 건수
- y축(우/선): 전체 취소 대비 누적 점유율

✓ 개별 고객의 산발적 취소가 아닌, 여행사 채널 주도 하의 일괄/정책성 취소 패턴 포착

“**호텔 측의 유통주권 실질 상태**”

- **예상 시나리오** : 여행사(TA/TO) 채널에서 5 ~ 7개월 이후의 객실 대량 선점  
→ 모객 실패 → 일괄 취소 / 호텔-TA 간 release 기간\* 도래

즉, 여행사(TA/TO)가 호텔의 'Allotment(판매 배정분)'를 일종의 '무상 옵션'처럼 활용하여 리스크를 호텔에 전가한 것.

\* release 기간: 여행사가 선점(블로킹)해둔 객실 중 예약이 차지 않은 방을  
호텔이 회수하여 직접 판매할 수 있도록 넘겨주는 '반납 시점'

# Analysis >> 발생 가능 리스크

Probable risk



## 기회 비용 손실

- 여행사 측의 장기 객실 점유  
→ 우량/실구매 고객 거절  
→ 판매 골든타임 상실



## 가격 책정 시스템 마비

- 허수 예약 기반의 고가 정책
- 임박 취소 시 덤플링 판매 가능성  
→ 수익성 하락  
→ 브랜드 가치 하락



## 자원 배분 왜곡

- 허수 수요에 기반한  
과잉 운용 비용 발생

# Conclusion >> 전략적 제안

suggestion

01

## 채널별 계약 조건 최적화

- Release 기간 단축  
→ 실판매 기간 확보
- 취소 시점별 단계적 위약금 제도 도입
- 취소 패널티 구조 재확인 및 변경  
(여행사와 호텔 간의 특약으로 인해 환불 불가여도 취소가 높았을 가능성)

02

## 예측 모델 기반 Overbooking

- 여행사(TA/TO) 채널의 예약에 대해 정교한 취소율 예측 모델을 만든 후 공격적인 '오버부킹' 시행

03

## 예약 보증(Deposit)의 실질화

- 예약 시점에 단순 카드 정보가 아닌, 선결제 비중을 높여 예약 책임감 강화
- 일정 비율 이상의 취소 발생 시 실질적인 위약금 부과

# Conclusion >> 분석의 한계점 및 기대효과

## 1

### 외부 변수 데이터 부재

: 항공편 지연/결항, 현지 축제, 기상 악화 등 호텔 내부 데이터 외에 예약 취소에 직접적인 영향을 주는 외부 환경 데이터가 분석에 포함되지 않음

## 2

### TA/TO 계약 조건의 상세 정보 부족

: '환불 불가'임에도 대량 취소가 발생하는 원인으로 'Release 기간' 정책을 추정했으나, 실제 호텔과 여행사 간의 세부 계약 조건 데이터가 없어 가설 검증에 한계가 있음

## 3

### 취소 사유 데이터 부재

: 고객의 단순 변심인지, 시스템 오류인지, 혹은 여행사의 모객 실패인지에 대한 구체적인 취소 사유 필드가 없어 패턴 기반의 추정만 가능

>> 향후 외부 변수 확장과 시계열 모델 고도화 등을 통해 실시간 취소 방지 예측 모델로 발전시킬 수 있을 것

# Next Step >> 예측 모델 확장 배경

01

## 채널별 계약 조건의 한계

- TA/TO 채널에서 환불 불가 조건임에도 높은 취소율 지속
- 동일한 계약 조건 하에서도 예약별 취소 행태에 큰 차이 존재
- 계약 조건만으로는 취소 리스크 사전 판단에 한계

02

## Overbooking 전략의 한계

- 기존 Overbooking은 평균 취소율 기반 일괄 적용  
-> 과도한 리스크 또는 기회 손실 발생
- 예약 단위의 취소 위험 예측 필요성 증가

03

## 예약 보증 정책의 한계

- 단순 카드 정보 입력만으로는 취소 억제 효과 제한적
- 고객·예약별로 취소 행태 차이 명확
- 취소 위험도 기반 차등 정책 필요

# Modeling >> 예약 시점 기반 취소 위험 예측 모델

modeling

## 모델 성능 요약

- ROC-AUC  $\approx 0.82$
- 취소 예약 Recall  $\approx 73\%$   
(임계값 0.3 기준)

```
t=0.2 precision=0.423 recall=0.903 f1=0.576
t=0.3 precision=0.520 recall=0.735 f1=0.609
t=0.4 precision=0.627 recall=0.548 f1=0.585
t=0.5 precision=0.715 recall=0.341 f1=0.462
```

## 모델 개요

- 목표: 예약 취소 여부 사전 예측
- 적용 시점: 예약 접수 시점
- 입력 정보: 리드타임, 예약 채널, 고객 유형,  
과거 취소 이력 등
- 모델: Logistic Regression  
(해석 가능성 중심)

“예약 시점에서 취소 위험을 미리 판단해, 운영 정책을 선제적으로 적용하기 위한 예측 모델입니다.”

## 주요 취소 위험 요인

- OTA 채널 (Online TA)
- 긴 리드타임
- 개별 고객(Transient)
- 과거 취소 이력
- 높은 ADR

## 운영 적용 방안

- 고위험 예약  
→ 보증금 / 선결제 적용
- 중위험 예약  
→ 취소 수수료 강화
- 저위험 예약  
→ 기존 정책 유지



감사합니다.

---

Thank you