

Hotel Booking Demand

호텔 매출 증가를 위한 마케팅 전략



CONTENTS

01. 배경분석

02. 전처리

03. 일변량, 이변량 분석

3.1 일변량 분석

3.2 이변량 분석

- 비유럽권 국가

04. 마켓 세그먼트 분석

4.1 포르투갈

4.2 City Hotel

4.3 Resort Hotel

05. 추가 분석 및 결론



01. 배경 분석

Hotel Booking Demand Dataset & About Portugal



데이터 배경 – City Hotel

데이터 셋 전반

포르투갈 지역의 두 호텔의 호텔 예약 정보

- City Hotel: Lisbon 지역
- Resort Hotel: Algarve 지역
- 크기: 119390x 32
- 결측치 및 계절적인 특징 존재
-> 분석 시, 계절성을 반영



City Hotel in Lisbon

비수기 (10월~2월)

- 일교차가 매우 크다 (낮 평균 15도, 밤 평균 7도)
 - 관광하기 좋지 않음
- 리스본 사람들에게겐 Algarve 여행가는 시기

성수기 (6월 ~ 8월)

- 포르투갈의 주요 축제가 이 시기에 열림
 - Festa de Lisboa, Carnivals 등
 - 6월 13일은 은행도 쉬는 큰 축제일

준성수기 (3~5월, 9~10월)

- 관광에 적합한 날씨 지속
- 호텔 Last Minute Deal 성사



데이터 배경 – Resort Hotel

데이터 셋 전반

포르투갈 지역의 두 호텔의 호텔 예약 정보

- City Hotel: Lisbon 지역
- Resort Hotel: Algarve 지역

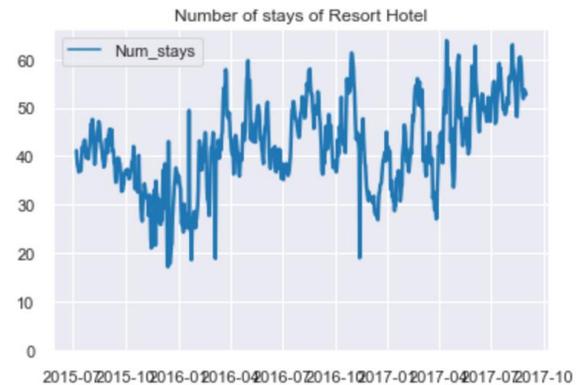
(포르투갈의 남부지역)



Resort Hotel in Algarve

포르투갈 최대 휴양지

- 휴양지 특성 상 겨울이 비수기
 - 11월부터 2월까지 비가 많이 온다.
 - 시계열 그래프를 통해, 11월~2월 사이 투숙객 감소함을 알 수 있음
- 나머지 기간은 연간 투숙객 일정하게 많은 편



ALGARVE WEATHER CHART

	MIN °C	MAX °C	RAIN (mm)
JAN	9	15	100
FEB	10	16	77
MAR	11	17	57
APR	11	18	45
MAY	13	20	30
JUN	16	23	16
JUL	17	25	2
AUG	17	25	2
SEP	17	25	16
OCT	15	22	58
NOV	12	18	90
DEC	10	16	102

상황 가정

고객

2015년 7월 포르투갈 리스본, 알가르브에 호텔을 새로 설립한 사장

고객의 니즈

호텔 매출 증진

City Hotel, Resort Hotel의 효율적인 운영

Data 분석 목적

영업 시작 이후 2015년 7월-2017년 8월 간의 Data를 바탕으로 호텔의 매출을 증진시키기 위한 마케팅 방안 구축

02. 데이터 전처리

Data Preprocessing For Analysis



1. 결측치 및 중복값 처리

결측치 처리

- Column: children, country, agent, company
- 처리 방법:
 - 'country', 'agent', 'company' 의 경우 'Undefined' 로 대체
 - 'children' 의 경우 행 삭제

중복값 처리

- 중복된 행이 31,994개 존재
- 중복된 데이터는 분석의 결과 왜곡할 수 있으므로 삭제
- 단, 중복된 데이터의 수가 꽤 많으므로, 추후 별도의 분석 실시
- df.shape: (87392, 32)

```
In [3]: df.isnull().sum()
Out[3]: hotel                0
         is_canceled         0
         lead_time           0
         arrival_date_year   0
         arrival_date_month  0
         arrival_date_week_number 0
         arrival_date_day_of_month 0
         stays_in_weekend_nights 0
         stays_in_week_nights 0
         adults              0
         children           4
         babies             0
         meal               0
         country            488
         market_segment     0
         distribution_channel 0
         is_repeated_guest   0
         previous_cancellations 0
         previous_bookings_not_canceled 0
         reserved_room_type  0
         assigned_room_type  0
         booking_changes     0
         deposit_type        0
         agent              16340
         company            112593
         days_in_waiting_list 0
         customer_type       0
         adr                0
         required_car_parking_spaces 0
         total_of_special_requests 0
         reservation_status  0
         reservation_status_date 0
         dtype: int64
```

```
In [44]: df[df.duplicated()].shape
Out[44]: (31994, 32)
```



2. Outlier 처리 및 날짜 데이터 통합

Outlier 처리

- 투숙객 기준
 - total guest 변수 설정: Adult, children, babies 열의 값을 더함
 - 총 투숙객 수 0명: 180개 삭제
 - Children 2명+ baby 1명: 3개 삭제
 - Adult 1명+ baby 9명: 1개 삭제
 - 주차 공간 기준
 - 총 투숙객 5명 이하+ 주차 공간 8대 필요: 2개 삭제
 - ADR 기준
 - ADR 음수: 1개 삭제
 - ADR 0원: 1643개 삭제 -> 별도의 데이터 프레임 구성 후 추후 분석
 - ADR 평균의 50배 이상: 1개 삭제
- ```
In [59]: df_free_of_charge.shape
Out[59]: (1643, 32)
```

### 날짜 데이터 처리

Arrival date의 year, month, day의 정보가 별도의 column으로 분리되어 있음

‘arrival\_date\_month’ 변수: 문자열

| arrival_date_year | arrival_date_month | arrival_date_week_number | arrival_date_day_of_month | s |
|-------------------|--------------------|--------------------------|---------------------------|---|
| 2015              | July               | 27                       | 1                         |   |
| 2015              | July               | 27                       | 1                         |   |
| 2015              | July               | 27                       | 1                         |   |

| reservation_status_date | arrival_date |
|-------------------------|--------------|
| 2015-07-02              | 2015-07-01   |
| 2015-07-02              | 2015-07-01   |
| 2015-07-03              | 2015-07-01   |

### 데이터 통합

- ‘arrival\_date’ 변수 생성
- 모든 날짜 데이터는 pandas의 datetime 자료형 변환

### 3. 고객 '국적' 관련 column 변환 및 추가

#### Country 열 변환

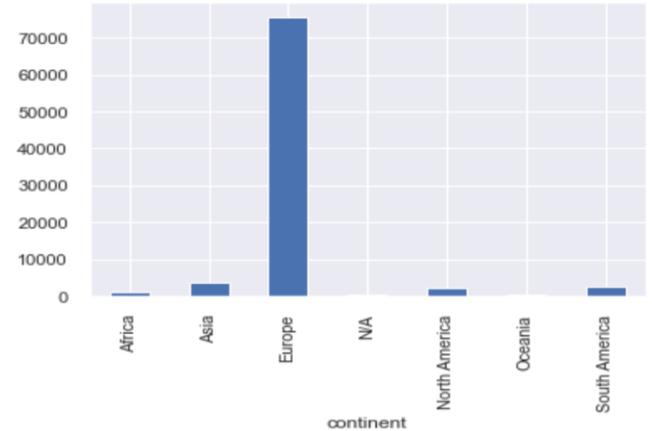
- 투숙객의 국적이 국가코드로 인코딩
  - 알아보기 쉽게 국가명으로 디코딩

#### Continent 열 추가

- 대륙별 추이 알아보기 위해 투숙객의 국적이 속한 대륙 column 추가
- 데이터 셋에 포함된 두 호텔의 투숙객들의 절대 다수가 유럽인
  - 분석을 위해 데이터 셋을 투숙객의 국적 바탕으로 유럽/비유럽권으로 분리

#### Continent\_label 열 추가

- 유럽 내 고객 세그먼트 분류 위해 유럽권 지리적 위치 바탕으로 세분화
- 분류 기준: 포르투갈, 스페인, 서유럽, 영국/아일랜드, 기타 유럽 국가



| country_decoded | continent | continent_label |
|-----------------|-----------|-----------------|
| United Kingdom  | Europe    | GB              |
| United Kingdom  | Europe    | GB              |
| United Kingdom  | Europe    | GB              |
| Portugal        | Europe    | Portugal        |
| Portugal        | Europe    | Portugal        |



# 03. 일변량 및 이변량 분석

---

Univariate & Bivariate Data Analysis



# 01. 일변량 분석

```
for i in range(8):
 for j in range(4):
 axs[i, j].hist(df[col_names[temp])
 axs[i, j].set_title(col_names[temp])
 temp += 1
```



- 1 hotel : Resort Hotel/City Hotel -> 범주형(2)
- 2 is\_canceled : 1(canceled)/0(not) -> 범주형(2)
- 3 lead\_time : Number of days that elapsed between the entering date of the booking into the PMS and the arrival date -> 연속형
- 4 arrival\_date\_year : Year of arrival date -> 연속형? 범주형?
- 5 arrival\_date\_month : Month of arrival date -> 날짜
- 6 arrival\_date\_week\_number : Week number of year for arrival date -> 날짜
- 7 arrival\_date\_day\_of\_month : Day of arrival date -> 날짜
- 8 stays\_in\_weekend\_nights : Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel -> 연속형
- 9 stays\_in\_week\_nights : Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel -> 연속형
- 10 adults : Number of adults -> 연속형
- 11 children : Number of children -> 연속형
- 12 babies : Number of babies -> 연속형
- 13 meal : SC/BB/HB/FB/Undefined -> 범주형(5)
- 14 country : Country of origin. Categories are represented in the ISO 3155-3:2013 format -> 범주형(178)
- 15 market\_segment : Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators" -> 범주형(8)
- 16 distribution\_channel : Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators" -> 범주형(5)
- 17 is\_repeated\_guest : repeated(1)/not(0) -> 범주형(2)
- 18 previous\_cancellations : Number of previous bookings that were cancelled by the customer prior to the current booking -> 연속형
- 19 previous\_bookings\_not\_canceled : Number of previous bookings not cancelled by the customer prior to the current booking -> 연속형
- 20 reserved\_room\_type : Code of room type reserved -> 범주형(10)
- 21 assigned\_room\_type : Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. -> 범주형(12)
- 22 booking\_changes : Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation -> 연속형
- 23 deposit\_type : No Deposit/Non Refund/Refundable -> 범주형(3)
- 24 agent : ID of the travel agency that made the booking -> 범주형(334)
- 25 company : ID of the company/entity that made the booking or responsible for paying the booking -> 범주형(353)
- 26 days\_in\_waiting\_list : Number of days the booking was in the waiting list before it was confirmed to the customer -> 연속형?
- 27 customer\_type : Contract/Group/Transient/Transient-Party -> 범주형(4)
- 28 adr : Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights -> 연속형
- 29 required\_car\_parking\_spaces : Number of car parking spaces required by the customer -> 연속형
- 30 total\_of\_special\_requests : Number of special requests made by the customer (e.g. twin bed or high floor) -> 연속형
- 31 reservation\_status : Canceled/Check-Out/No-Show -> 범주형(3)
- 32 reservation\_status\_date : Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel -> 날짜

## 02. 이변량 (비유럽권 세그먼트) 분석

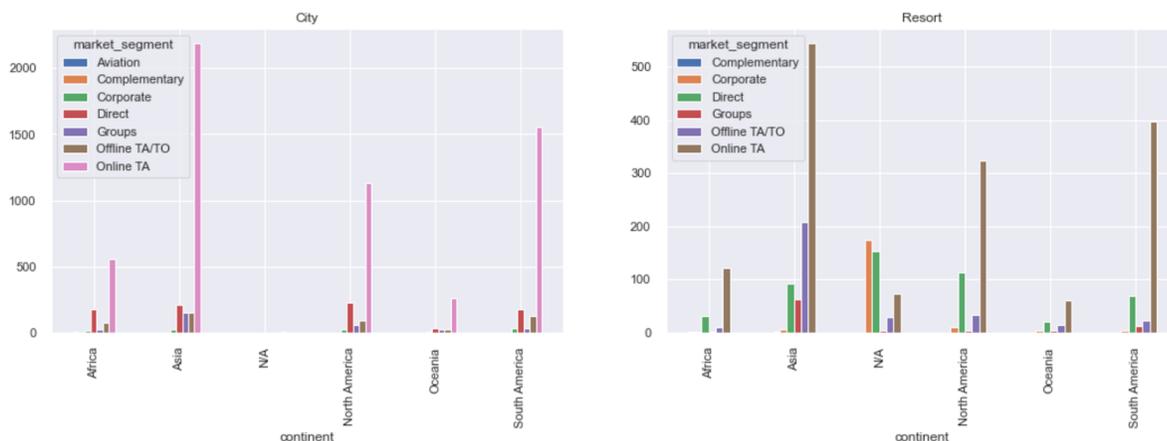
### - 분석 개요

```
print("전체 투숙객에서 비유럽권 손님이 차지하는 비율: {} %".format(round(len(non_europe_df)/len(df),2)*100))
```

전체 투숙객에서 비유럽권 손님이 차지하는 비율: 12.0 %

- 포르투갈 내 위치한 호텔인만큼, 전체에서 비유럽권 고객 차지 비중은 12%
- **분석 전략:** 유럽권 고객에 더 집중하되, 비유럽권 고객의 특징만 파악

### - 예약 채널 비교



- City, Resort 모두 Online TA/TO 비중이 높음
- **MKT 전략:**  
온라인 여행사를 통한 퍼포먼스 마케팅 증가

# 04. 유럽권 마켓 세그먼트 분석

---

유럽 내 국가별 마켓 세그먼트 설정 및 마케팅 전략 구축



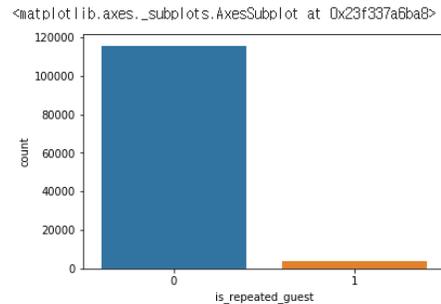
# 00. 국가 별 호텔 이용 현황?



# 01. 포르투갈 – Repeated Guest

## Repeated Guest 분석

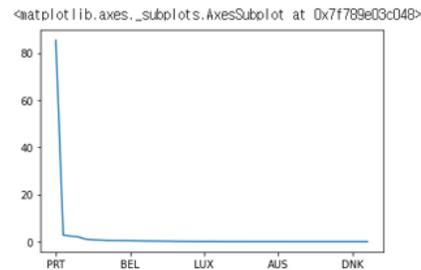
- 전체 Dataset 내 Repeated guest의 비율은 적음.



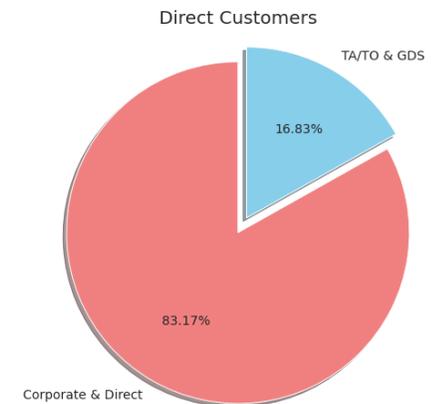
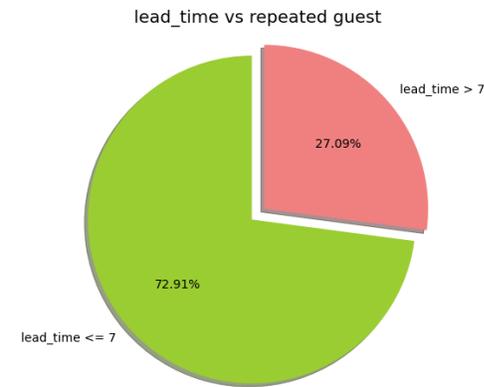
|          | resort   | city     |
|----------|----------|----------|
| repeated |          |          |
| 0        | 0.245031 | 0.306949 |
| 1        | 0.042254 | 0.113458 |

- Repeated guest의 취소율은 첫 방문 고객에 비해 Resort 호텔 4%, City Hotel 11%로 매우 낮은 편

- Repeated Guest의 85%는 포르투갈 자국민



- Repeated Guest의 70% 이상이 사용일로부터 7일 이내 예약
- Repeated Guest의 80% 정도가 Agent를 통하지 않은 직접적인 방법으로 예매



- 자국민의 급박한 호텔 예약 니즈 발생 시 자사 호텔이 고객에게 ToM(Top of Mind)이 될 수 있는 MKT 전략 필요
- Direct distribution channel을 통해 예약 시 추가 할인 서비스 제공

호텔의 취소를 하락, Agent 수수료 비용 감소, loyal customer 확보를 위해 Repeated Guest를 늘릴 수 있는 MKT 전략 구축이 핵심

# 01. 포르투갈 – Travel Agency

## Travel Agency

- Agent 변수: Distribution Channel에서 TA/TO로의 예약 진행(약 70%) 시 고객이 이용한 Agent의 고유 번호

## City Hotel

- 자국민이 가장 많이 사용한 Travel Agency는 9번 여행사
- 자국민이 사용하는 여행사는 약 180개 정도

## Resort

- 자국민이 가장 많이 사용한 Travel Agency는 240번 여행사 (약 30%)
- 자국민이 Resort hotel 예약 시 사용하는 Agent는 약 165개 정도

## 인사이트

- City hotel 예약 시 사용하는 Agent != Resort hotel 예약 시 사용하는 Agent가 다름
  - 9번 agent를 통한 Resort 예약 건수 5건/ 240번 agent를 통한 City 예약 건수 2건

|           | agent |           | agent |
|-----------|-------|-----------|-------|
| Undefined | 3549  | 240       | 4694  |
| 9         | 3277  | Undefined | 4268  |
| 1         | 772   | 250       | 1296  |
| 7         | 759   | 241       | 453   |
| 14        | 643   | 242       | 349   |
| 8         | 460   | 314       | 304   |
| 6         | 258   | 6         | 189   |
| 28        | 252   | 196       | 185   |
| 83        | 226   | 5         | 163   |

- 가장 많이 예약 건이 이루어지는 9번, 240번 Agent는 대형 여행사일 확률 높음
- 2-5위의 비교적 높은 실적을 내는 Agent와 파트너십을 맺어 Distribution Channel 최적화 하여 운영 효율화를 위한 전략 구축

An aerial photograph of a city at sunset. The sun is low on the horizon, casting a warm orange glow over the city. The foreground is dominated by the dark, silhouetted branches of a pine tree. The city below is densely packed with buildings, and the sky is a mix of orange and blue.

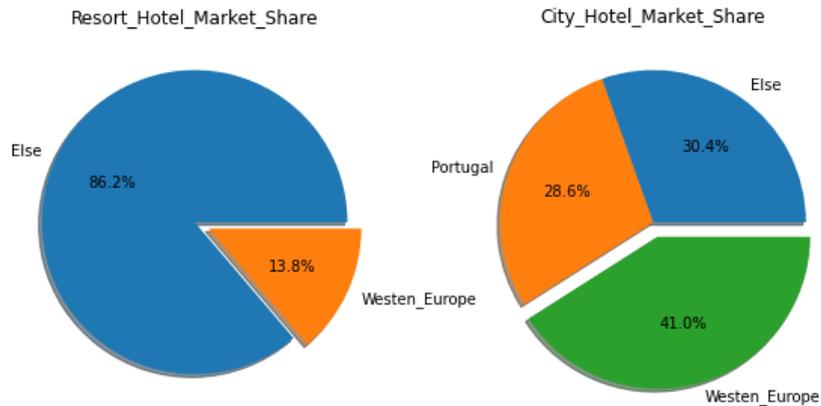
# City Hotel 매출 증진 전략

# 02. 서유럽-City Hotel 전략

## 서유럽 개요

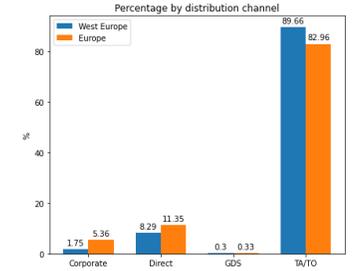
- 프랑스, 독일, 이탈리아, 벨기에, 네덜란드, 스위스
- Resort Hotel의 서유럽 market share은 상대적으로 낮음
- City Hotel의 서유럽 market share는 국내 이용률보다 높음을 확인

>> City Hotel의 서유럽 고객 유치를 통한 매출 증진 전략 구축 필요성



## Distribution Channel & Agents

- TA/T0를 통해 예약을 하지 않는 비율이 유럽 평균에 비해 40% 낮음
- 포르투갈과 마찬가지로 9번 Agent가 예약 점유율 63% 차지
- 서유럽 고객의 호텔 예약 시 사용하는 상위 agent 중 Online TA/T0에 속하는 agent는 7번 agent



```

9 Online TA 11795
Name: market_segment, dtype: int64
14 Direct 1163
Name: market_segment, dtype: int64
7 Online TA 1069
Name: market_segment, dtype: int64
28 Offline TA/T0 1010
Name: market_segment, dtype: int64

```

| hotel     |       |
|-----------|-------|
| agent     |       |
| 9         | 62.82 |
| 14        | 6.19  |
| 7         | 5.69  |
| 28        | 5.38  |
| Undefined | 4.16  |
| 8         | 1.31  |
| 27        | 1.29  |
| 6         | 1.10  |
| 1         | 0.98  |
| 85        | 0.93  |

7번 Agent와의 파트너십 구축 및 Agent를 통한 적극적 프로모션을 통한 서유럽 고객 유치 MKT 전략





# Resort Hotel 매출 증진 전략

# 03. 스페인

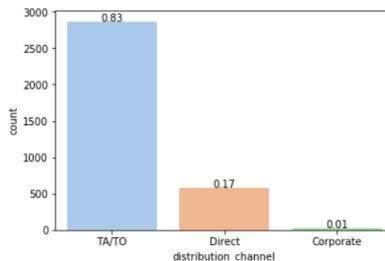
## Resort Hotel 주요 고객

- 스페인, 영국, 아일랜드 고객은 Resort Hotel 점유율 높음



## Distribution Channel & Agents

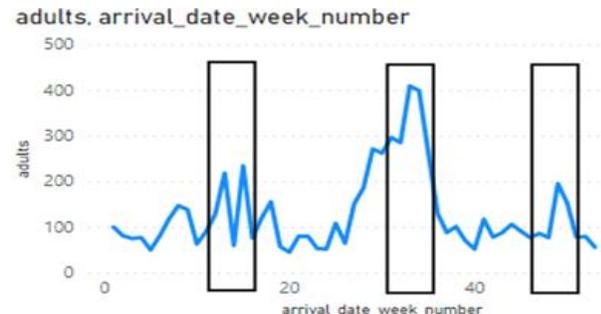
- 포르투갈 제외 타 유럽국가와 비교하여 **Direct Distribution Channel**을 통한 예약 비율이 높음
  - 이름 스페인, 포르투갈이 같은 언어권이기 때문으로 유추 가능



## 예약 시기 경향성

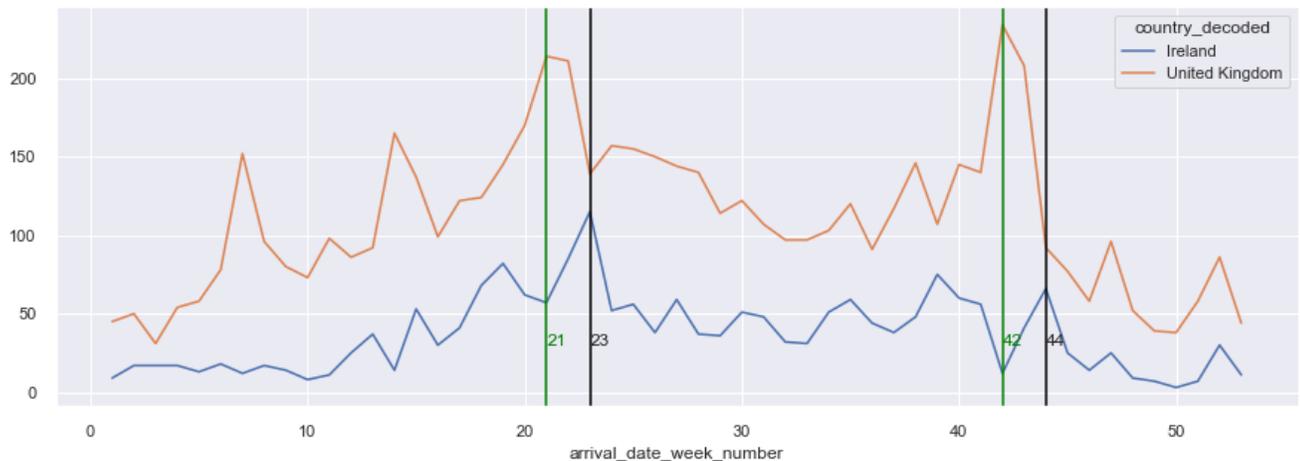
스페인 및 포르투갈의 경우 천주교가 국교  
-> 스페인 고객의 포르투갈 방문 시기는 **천주교 공휴일**과 일치

- 3월 말~ 4월초: 부활절 행사
- 33-37주: 극 성수기, 성모승천일
- 46-50주: 만성절



- 다른 국가에 비해 Direct Distribution Channel을 통한 예약을 유도하는 프로모션 진행
- Algarve 랜드마크 중 하나인 봉안당 성당 등의 투어 패키지 만들어 천주교 공휴일에 스페인 관광객 유치

# 04. 영국 & 아일랜드 - 예약 시기



## 예약 시기 경향성

영국: 22주(5월 말)와 42주(10월 초) 근처  
 나머지 국가: 33주(8월 중순) 근처  
 예약 건수가 많은 경향 존재

## 공휴일

### 영국의 공휴일

*England Spring Bank Holiday*: 매년 5월 마지막 주 월요일  
*England Autumn Break*: 10월 초~말, England, Wales, Scotland, Northern Ireland의 가을방학

### 아일랜드의 공휴일

*Ireland June Bank Holiday*: 매년 6월 첫 주 월요일  
*Ireland October Holiday*: 매년 10월 마지막 주 월요일  
*Ireland October Holidays*: 10월 마지막 한 주, 가을방학

\* Bank Holiday: 영국/아일랜드의 빨간 날  
 \* October Holiday: 학생들의 방학

영국과 아일랜드는 여름보다 오히려 공휴일이 포함된 주에 리조트 이용률이 높으며,  
 영국과 아일랜드의 공휴일 시기 차이에 따른 이용률 차이 존재

# 04. 영국 & 아일랜드 - 예약 시기

## 5월

| country_decoded | arrival_date_week_number | count | mean     | std      | min | 25%  | 50% | 75%  | max  |
|-----------------|--------------------------|-------|----------|----------|-----|------|-----|------|------|
| Ireland         | 20                       | 62.0  | 6.629032 | 2.234116 | 1.0 | 6.25 | 7.0 | 7.0  | 14.0 |
|                 | 21                       | 57.0  | 6.298246 | 2.367676 | 1.0 | 5.00 | 7.0 | 8.0  | 11.0 |
|                 | 22                       | 85.0  | 6.717647 | 2.169111 | 1.0 | 7.00 | 7.0 | 7.0  | 14.0 |
|                 | 23                       | 115.0 | 6.573913 | 3.198302 | 1.0 | 5.00 | 7.0 | 7.0  | 21.0 |
|                 | 24                       | 52.0  | 7.423077 | 2.644611 | 1.0 | 7.00 | 7.0 | 9.0  | 14.0 |
| United Kingdom  | 20                       | 170.0 | 5.835294 | 3.027867 | 1.0 | 3.00 | 7.0 | 7.0  | 14.0 |
|                 | 21                       | 214.0 | 5.752336 | 3.005374 | 1.0 | 3.00 | 7.0 | 7.0  | 14.0 |
|                 | 22                       | 211.0 | 5.834123 | 3.155641 | 1.0 | 4.00 | 6.0 | 7.0  | 14.0 |
|                 | 23                       | 139.0 | 8.050360 | 3.628238 | 1.0 | 7.00 | 7.0 | 10.0 | 15.0 |
|                 | 24                       | 157.0 | 8.114650 | 3.324285 | 1.0 | 7.00 | 7.0 | 10.0 | 14.0 |

### 50% quantile 이상

아일랜드와 영국 모두 대부분 일주일 이상 머무는 것으로 보임

### 25% quantile

아일랜드의 경우: 5일 이상으로 길게 머무름  
 영국의 경우: 머무는 기간 3-4일 정도 (짧은 휴가)  
 영국은 공휴일 전주 목요일의 예약 건수가 많음.

## 10월

| country_decoded | arrival_date_week_number | count | mean     | std      | min | 25%  | 50% | 75% | max  |
|-----------------|--------------------------|-------|----------|----------|-----|------|-----|-----|------|
| Ireland         | 41                       | 56.0  | 5.464286 | 2.350739 | 1.0 | 4.00 | 5.0 | 7.0 | 14.0 |
|                 | 42                       | 12.0  | 5.583333 | 1.621354 | 3.0 | 4.75 | 6.0 | 7.0 | 7.0  |
|                 | 43                       | 41.0  | 5.536585 | 2.110658 | 1.0 | 4.00 | 5.0 | 7.0 | 10.0 |
|                 | 44                       | 66.0  | 5.969697 | 1.839479 | 1.0 | 4.25 | 7.0 | 7.0 | 10.0 |
|                 | 45                       | 25.0  | 5.840000 | 2.374868 | 2.0 | 4.00 | 6.0 | 7.0 | 14.0 |
| United Kingdom  | 41                       | 140.0 | 5.678571 | 3.243739 | 1.0 | 3.00 | 4.0 | 7.0 | 14.0 |
|                 | 42                       | 234.0 | 4.615385 | 2.506117 | 1.0 | 3.00 | 4.0 | 7.0 | 14.0 |
|                 | 43                       | 208.0 | 5.519231 | 2.706160 | 1.0 | 4.00 | 5.0 | 7.0 | 17.0 |
|                 | 44                       | 92.0  | 5.010870 | 2.051505 | 1.0 | 4.00 | 5.0 | 7.0 | 10.0 |
|                 | 45                       | 77.0  | 5.948052 | 3.235885 | 1.0 | 4.00 | 6.0 | 7.0 | 21.0 |

### 25%와 75% 차이

아일랜드: 3일

영국: 5일

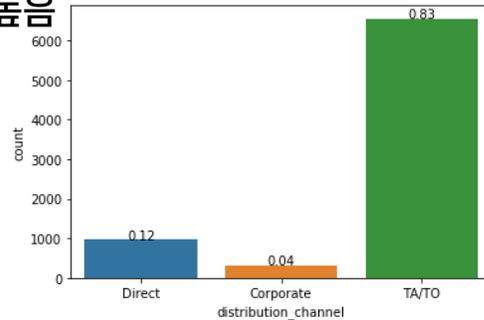
영국에 비해 아일랜드 고객의 리조트 이용 기간이 1-2일 더 김

-> 영국은 아일랜드에 비해 지역에 따라 다르지만 가을 방학이 짧은 경우가 존재하기 때문

# 04. 영국 & 아일랜드 - 예약 채널 및 MKT 전략

## Distribution Channel 및 Agent 분석

- 영국의 Direct 비율은 12.4061%로 스페인에 비해서 낮은 편.
- TA/T0의 경우 83.4671%로 높음



## 유럽 9개국(영국 포함) 및 영국의 agent 분포 비교

| agent     | count      | agent     | count      |
|-----------|------------|-----------|------------|
| 240       | 6380       | 240       | 1489       |
| 250       | 1160       | <b>40</b> | <b>814</b> |
| <b>40</b> | <b>837</b> | 241       | 467        |
| 241       | 774        | 243       | 429        |

- 9개국 내에서 agent 40 예약 건 수 총 837건, 그 중 814건이 영국에서 발생
- Agent 40을 통한 영국 고객 예약 건수가 2위

## MKT 전략

### 5월 프로모션 전략: 직장인을 공략

- 영국과 아일랜드 고객 대상으로 5-6월 사이 1-2주 패키지 프로모션
- 영국 고객 대상으로 Spring Bank Holiday 3일 단기 리조트 패키지 프로모션

### 10월 프로모션 전략: 가족을 공략

- 영국 가족 고객 대상으로 10월 내 5일 compact 리조트 패키지 프로모션
- 아일랜드 학생들을 대상으로 October Holiday 일주일 리조트 패키지

### Agent와의 프로모션 진행

영국&아일랜드 고객 대상 프로모션 진행 시,  
40번 agent와 파트너십을 맺어 진행



# 05. 마케팅 전략 정리

대륙 및 국가별 데이터 분석을 통해 국가마다 자주 이용하는 호텔의 **지역, 기간, 유입경로가 다른 경향**을 보임을 확인 -> 지역별로 다른 마케팅 전략을 실행하여 매출 증가를 기대

## 포르투갈

- Loyal Customer 확보 위해 Repeated Guest를 늘릴 수 있는 전략
- 중소규모 Travel Agent와의 파트너십을 통한 Distribution Channel 운영 최적화

## 서유럽

- City Hotel에 주력
- 7번 Travel Agent와 파트너십 구축과 적극적 프로모션 진행

## 스페인

- Resort Hotel에 주력
- 천주교 공휴일을 공략한 투어 패키지 통한 관광객 유치

## 영국& 아일랜드

- Resort Hotel에 주력
- 5월, 10월 공휴일을 공략한 직장인/가족/학생 패키지
- 40번 Travel Agent와의 파트너십을 통한 영국 관광객 유치



# 05. 추가 분석 및 결론

---

이외 추가 분석을 통한 인사이트 도출 및 한계, 느낀점



# 1. 중복된 자료들에 대한 분석

## Distribution Channel

- 중복된 데이터의 경우 'TA/TO'의 비중이 타 데이터에 비해 높음
- 중복된 데이터에 비해 직접 예약한 비율이 2배 가량 낮음

### 중복 데이터

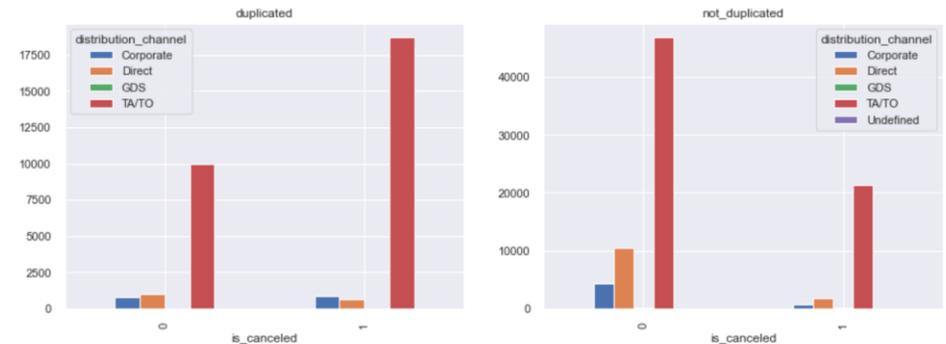
| distribution_channel | Corporate | Direct | GDS   | TA/TO  |
|----------------------|-----------|--------|-------|--------|
| is_canceled          |           |        |       |        |
| 0                    | 6.53%     | 8.69%  | 0.09% | 84.69% |
| 1                    | 4.09%     | 3.13%  | 0.0%  | 92.78% |

### 중복되지 않은 데이터

| distribution_channel | Corporate | Direct | GDS   | TA/TO  | Undefined |
|----------------------|-----------|--------|-------|--------|-----------|
| is_canceled          |           |        |       |        |           |
| 0                    | 6.92%     | 16.82% | 0.23% | 76.03% | 0.0%      |
| 1                    | 2.63%     | 7.69%  | 0.15% | 89.53% | nan%      |

## 예약 취소율

- 중복 데이터와 중복되지 않은 데이터의 canceled의 상대적 비율이 **반대양상**을 보임
- 중복된 데이터에서 Canceled된 예약 더 많음



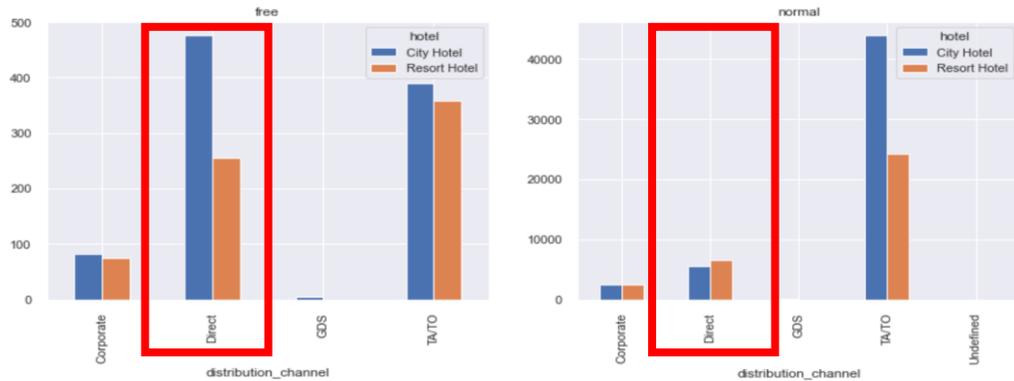
## Duplicated Data의 경우

1. 특정 여행사가 손님 모집 후 일괄적으로 여러 개의 방을 예약 했을 가능성
2. 취소된 예약들에 대해 수정을 하면서 동일 데이터 한번 더 '잘못' 추가했을 가능성

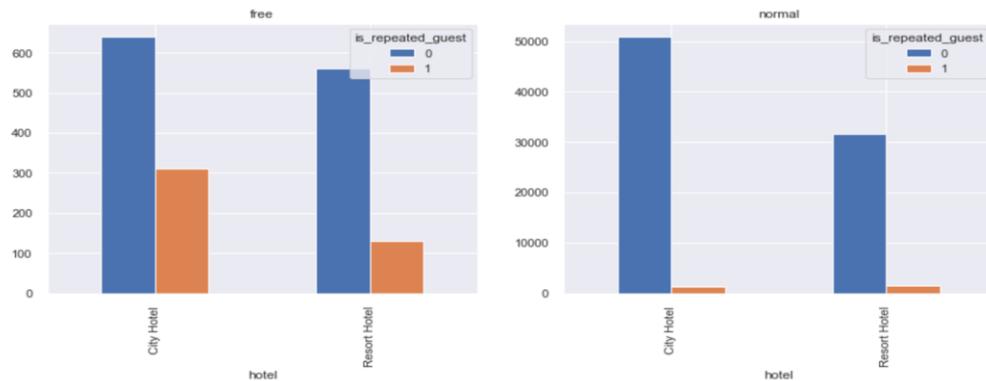
## 2. ADR=0인 data 분석

Data 크기: 1634x32

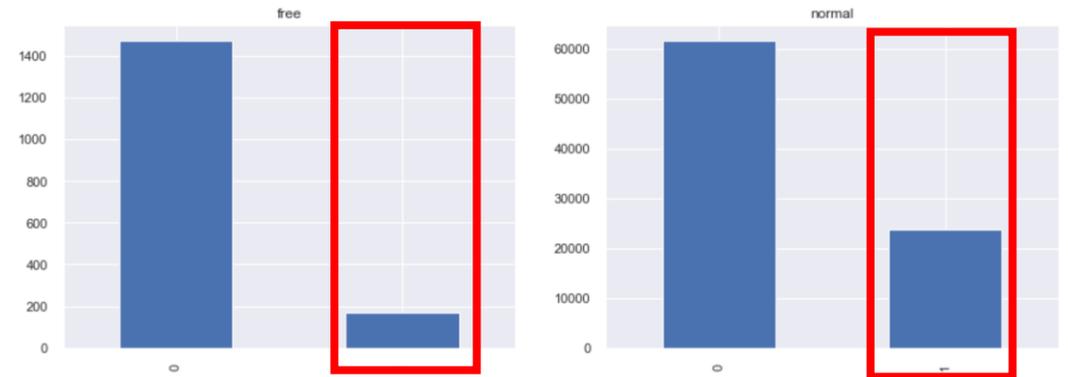
Distribution Channel: Direct 비중에서의 큰 차이 존재



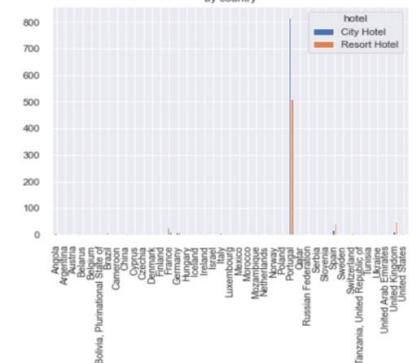
Repeated Guest: 재방문 고객 비율 높음



예약 취소율: 현저히 낮음



투숙객 국적: 대부분 포르투갈 자국민



ADR=0의 경우

단골 고객 혹은 회원 대상 이벤트 성 객실이라 추정

### 3. 한계 및 느낀 점

#### 한계

1. 비유럽권의 절대적 파이가 적음
2. 추가 분석은 데이터를 기반으로 예측한 것일 뿐 사실 여부는 알 수 없었음
3. Customer ID의 부재로 분석의 다양성의 한계

#### 느낀 점

- JW** “인사이트가 파도 파도 계속 나온다… 가끔은 너무 사소하기도 한데, 가끔은 소름 돋을 정도로 신기하기도 하다”
- DW** “다양한 정보(열) 중 어떤 것을 어떤 조합으로 엮어야 되는지가 어렵다…”
- HI** “데이터로부터 경향성을 찾았을 때 이것이 실제로 문화적 특성과 연관되어 있다고 확신을 갖기가 어려운 것 같다“
- SK** “Background 지식이 많으면 많을수록 다양한 분석이 가능하다는 것을 좀 더 확실하게 느낄 수 있었다“
- NE** “이렇게 Dataset을 deep하게 파본 것은 처음인 것 같다“



감사합니다