HOTEL BOOKING ANALYSIS PROJECT

USING PYTHON

By Akanksha Kumari

EXECUTIVE SUMMARY

- **Project Goal:** Analyse hotel booking data to identify key drivers of cancellations and suggest strategies to reduce them.
- **Data Used:** 119,000+ records from two hotels (City and Resort) between 2015–2017.
- **Key Findings:**
  - ~37% of bookings are cancelled.
  - City hotels have higher cancellation rates than resort hotels.
  - Cancellations are more frequent with higher ADRs and through OTAs.

BUSINESS PROBLEM

- In recent years, City Hotel and Resort Hotel have seen high cancellation rates. Each hotel is now dealing with several issues as a result, including lower revenues and less-than-ideal hotel room use. Consequently, lowering cancellation rates is both the hotels' primary goal to increase their efficiency in generating revenue and for us to offer business advice to address this problem.

The analysis of hotel booking cancellations, as well as other factors that have no bearing on their business and yearly revenue generation, is the main topic of this report.

PROJECT OBJECTIVE

- Analyse patterns in hotel bookings.
- Identify causes of cancellations.
- Provide data-driven suggestions to reduce cancellations and increase revenue.

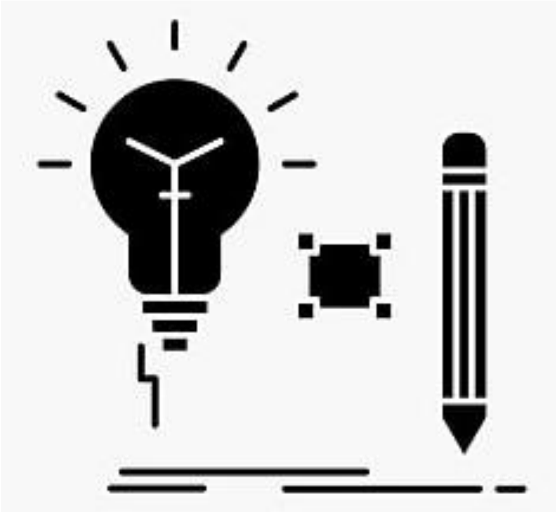
ASSUMPTIONS

- No unusual occurrences between 2015 and 2017 will have a substantial impact on the data used.
- The information is still current and can be used to analyse a hotel's possible plans efficiently.
- There are no unanticipated negatives to the hotel employing any advised technique.
- The hotels are not currently using any of the suggested solutions.
- The biggest factor affecting the effectiveness of earning income is booking cancellations.
- Cancellations result in vacant rooms for the booked length of time.
- Clients make hotel reservations the same year they make cancellations.

RESEARCH QUESTION

1. What are the variables that affect hotel reservation cancellations?
2. How can we make hotel reservation cancellations better?
3. How will hotels be assisted in making pricing and promotional decisions?

HYPOTHESES:



1. More cancellations occur when prices are higher.
2. When there is a longer waiting list, customers tend to cancel more frequently.
3. The majority of clients are coming from offline travel agents to make their reservations.

TOOLS & DATASET

- **Programming language:** Python
- **Tool:** Jupyter Notebook
- **GitHub:** HOTEL-BOOKING ANALYSIS PROJECT USING PYTHON
- **Libraries**: Pandas, Matplotlib, Seaborn
- **Dataset**: hotel_bookings.csv
- **Data Cleaning:**
  - Dropped columns: company, agent
  - Removed outliers (ADR > 5000)
  - Converted date fields

```
Index: 118897 entries, 0 to 119389
Data columns (total 31 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           118897 non-null  object
 1   is_canceled                     118897 non-null  int64
 2   lead_time                       118897 non-null  int64
 3   arrival_date_year               118897 non-null  int64
 4   arrival_date_month              118897 non-null  object
 5   arrival_date_week_number        118897 non-null  int64
 6   arrival_date_day_of_month       118897 non-null  int64
 7   stays_in_weekend_nights         118897 non-null  int64
 8   stays_in_week_nights            118897 non-null  int64
 9   adults                          118897 non-null  int64
 10  children                        118897 non-null  float64
 11  babies                          118897 non-null  int64
 12  meal                            118897 non-null  object
 13  country                         118897 non-null  object
 14  market_segment                  118897 non-null  object
 15  distribution_channel            118897 non-null  object
 16  is_repeated_guest               118897 non-null  int64
 17  previous_cancellations          118897 non-null  int64
 18  previous_bookings_not_canceled  118897 non-null  int64
 19  reserved_room_type              118897 non-null  object
 20  assigned_room_type              118897 non-null  object
 21  booking_changes                 118897 non-null  int64
 22  deposit_type                    118897 non-null  object
 23  days_in_waiting_list            118897 non-null  int64
 24  customer_type                   118897 non-null  object
 25  adr                             118897 non-null  float64
 26  required_car_parking_spaces     118897 non-null  int64
 27  total_of_special_requests       118897 non-null  int64
 28  reservation_status              118897 non-null  object
 29  reservation_status_date         118897 non-null  datetime64[ns]
 30  month                           118897 non-null  int32
dtypes: datetime64[ns](1), float64(2), int32(1), int64(16), object(11)
```

METHODOLOGY

1 Data Collection

•Used hotel_bookings.csv dataset (119k+ records)

2 Data Cleaning

•Dropped irrelevant columns: company, agent
•Handled null values and removed ADR outliers (> 5000)

3 Exploratory Data Analysis (EDA)

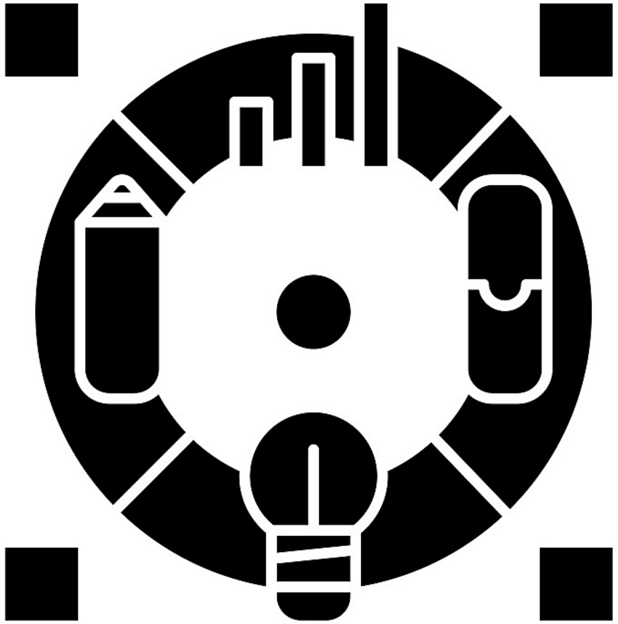
•Visualized trends: cancellation rates, ADR, seasonality, booking channels

4 Insights Extraction

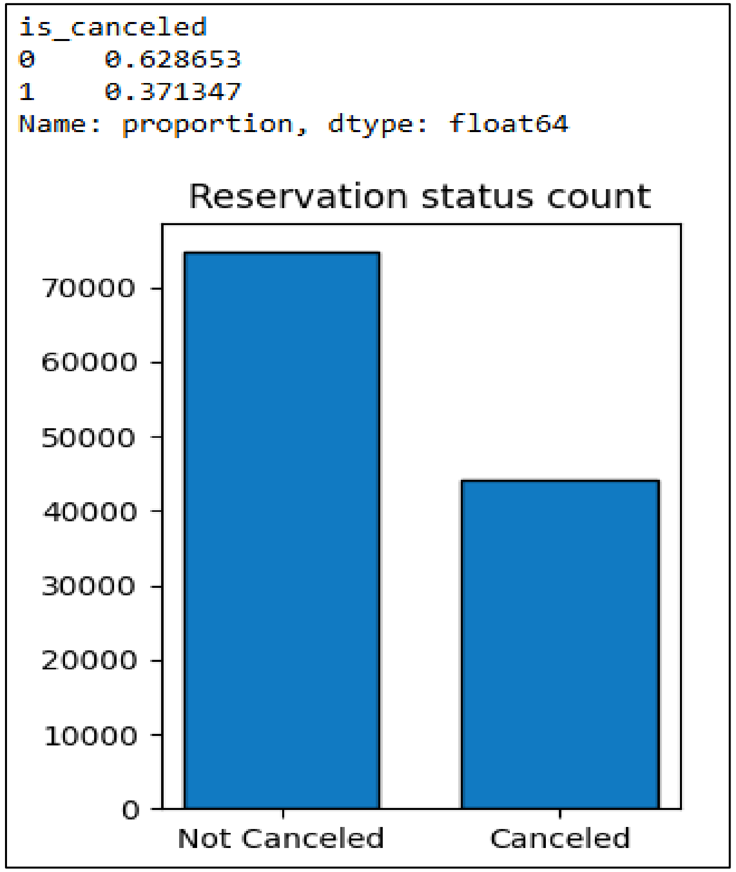
•Identified patterns and variables strongly linked to cancellations

5 Actionable Recommendations

•Proposed strategies to reduce cancellations and improve revenue

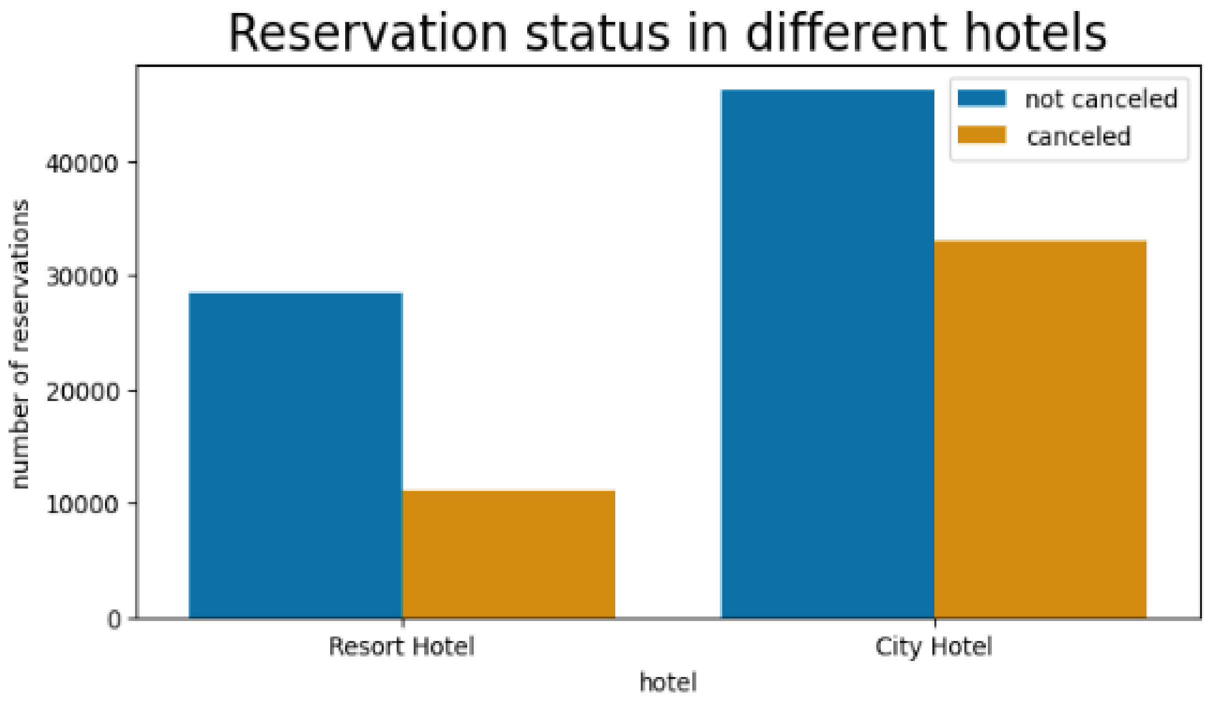


ANALYSIS AND FINDINGS - 1



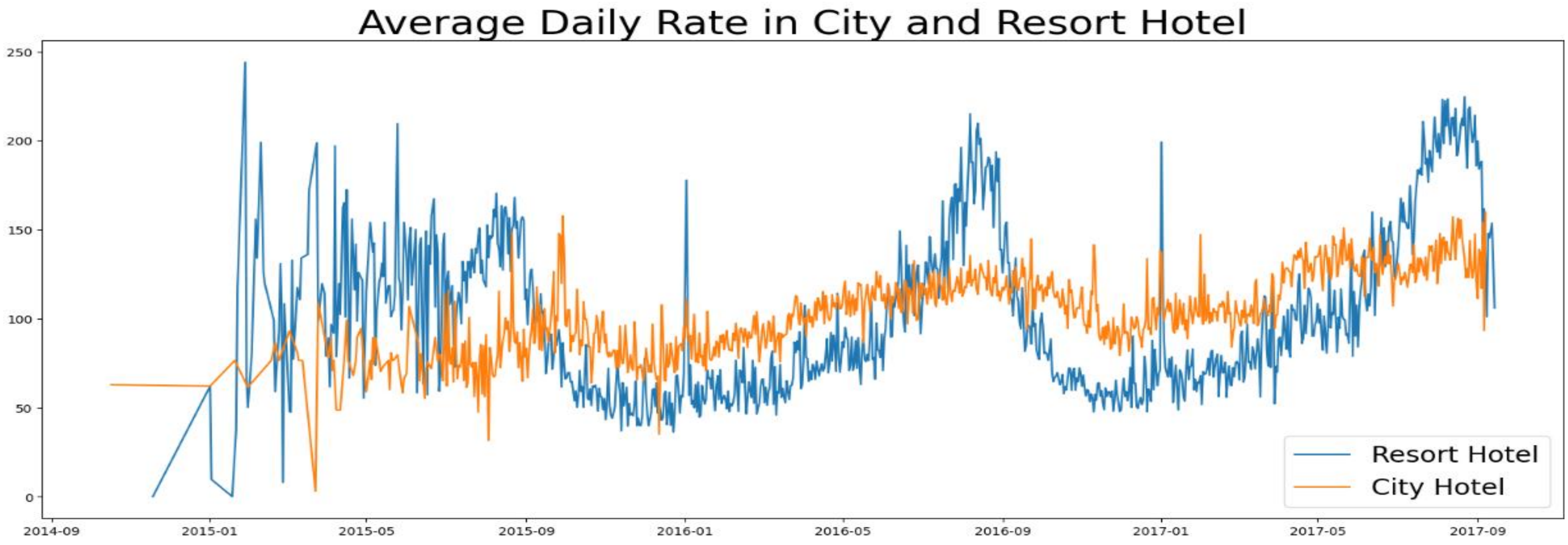
- The bar chart reveals that approximately **37% of hotel bookings were cancelled**, while **63% were fulfilled**. This high cancellation rate presents a significant challenge for hotel operations and revenue management. Each cancelled booking represents a potential loss in income and contributes to inefficiencies in room utilisation. Addressing this issue is crucial for improving profitability and operational planning.

ANALYSIS AND FINDINGS - 2



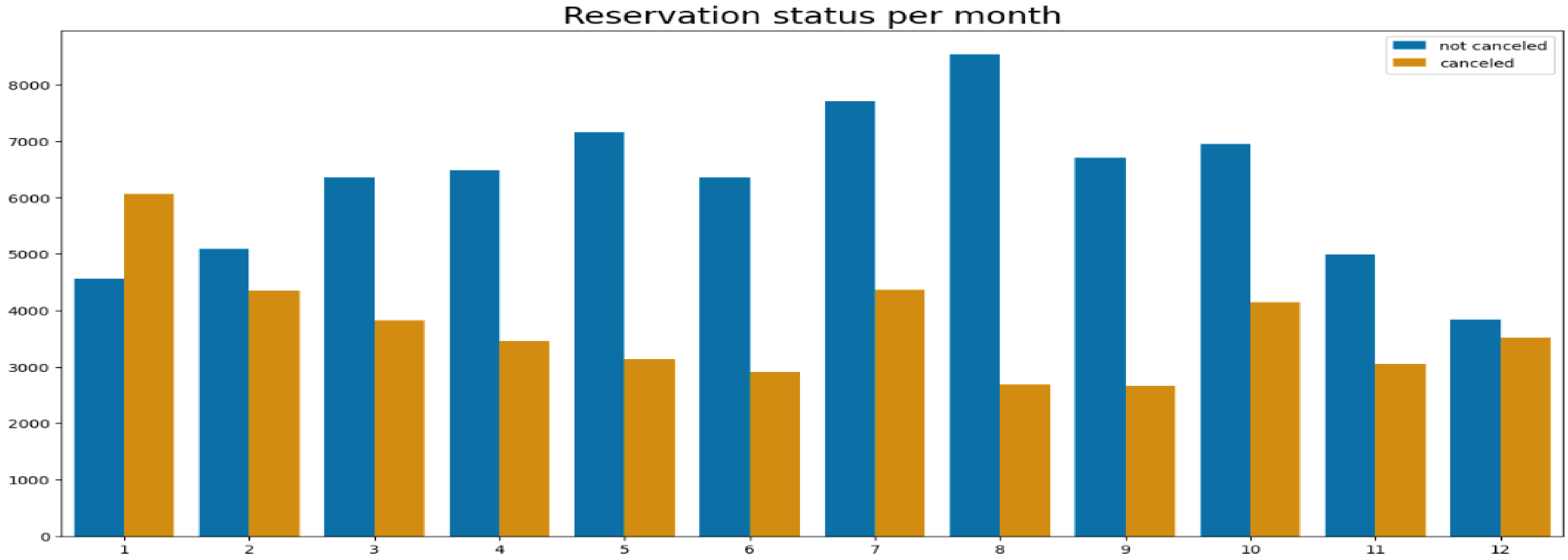
- The bar chart compares reservation outcomes between resort and city hotels. **City hotels receive more bookings overall**, but also face a **higher absolute number of cancellations** compared to resort hotels. While **resort hotels have fewer cancellations**, they also attract fewer total bookings. This suggests that **city hotels are more popular but more prone to cancellations**, possibly due to easier accessibility, lower booking commitment, or more flexible cancellation policies.

ANALYSIS AND FINDINGS - 3



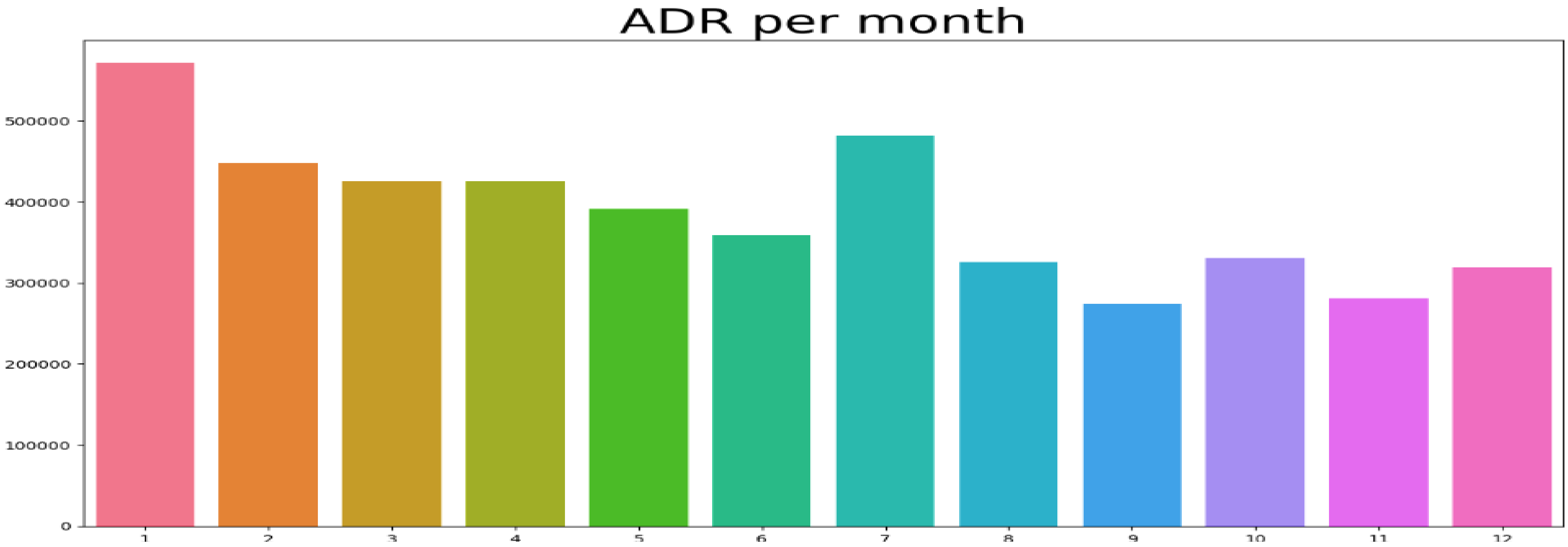
- City hotels show steady growth in ADR with less volatility, indicating stable business demand. In contrast, resort hotels exhibit strong seasonal fluctuations and more volatile pricing, likely tied to tourism patterns. Both segments reflect an overall upward trend in pricing from 2015 to 2017

ANALYSIS AND FINDINGS - 4



- We have developed the grouped bar graph to analyse the months with the highest and lowest reservation levels according to reservation status. As can be seen, both the number of confirmed reservations and the number of cancelled reservations are largest in the month of August, whereas January is the month with the most cancelled reservations.

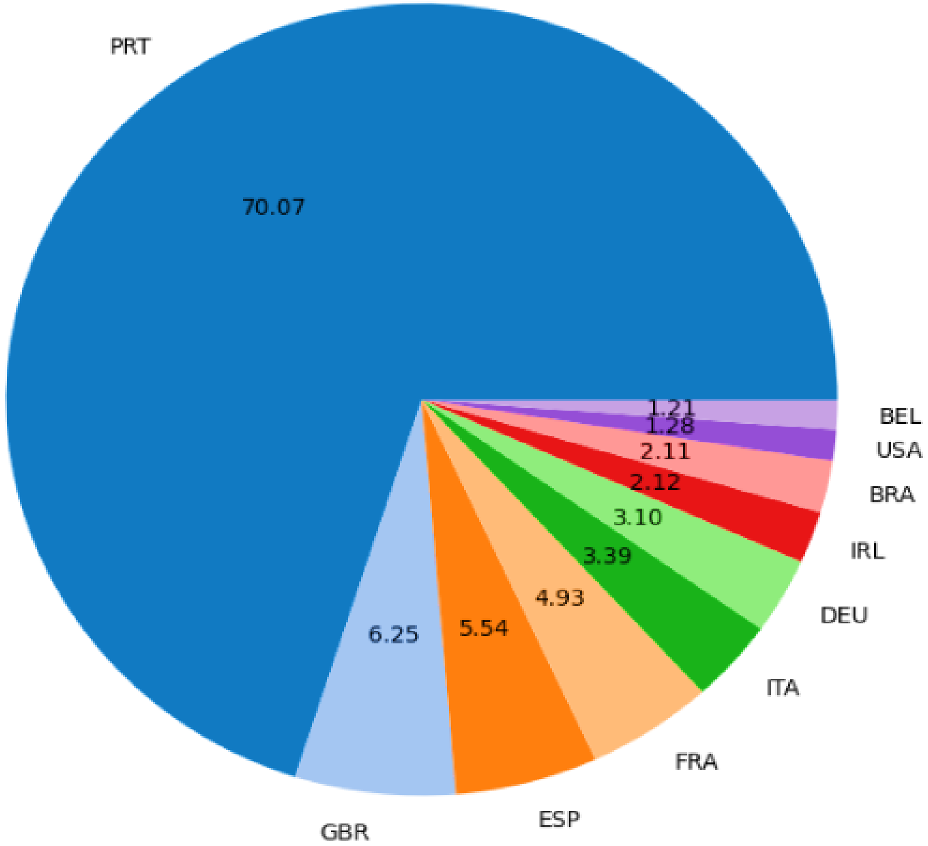
ANALYSIS AND FINDINGS - 5



- The bar graph demonstrates that cancellations are most common when prices are greatest and are least common when they are lowest. Therefore, the cost of the accommodation is solely responsible for the cancellation.

ANALYSIS AND FINDINGS - 6

Top 10 Countries by Reservation Cancellations



- The pie chart illustrates the countries with the highest number of reservation cancellations, with Portugal leading, followed by the United Kingdom, Spain, France, and others.

ANALYSIS AND FINDINGS - 7

```
df['market_segment'].value_counts(normalize = True)
```

```
market_segment
Online TA        0.474377
Offline TA/TO    0.203193
Groups           0.166581
Direct           0.104696
Corporate        0.042987
Complementary    0.006173
Aviation         0.001993
```

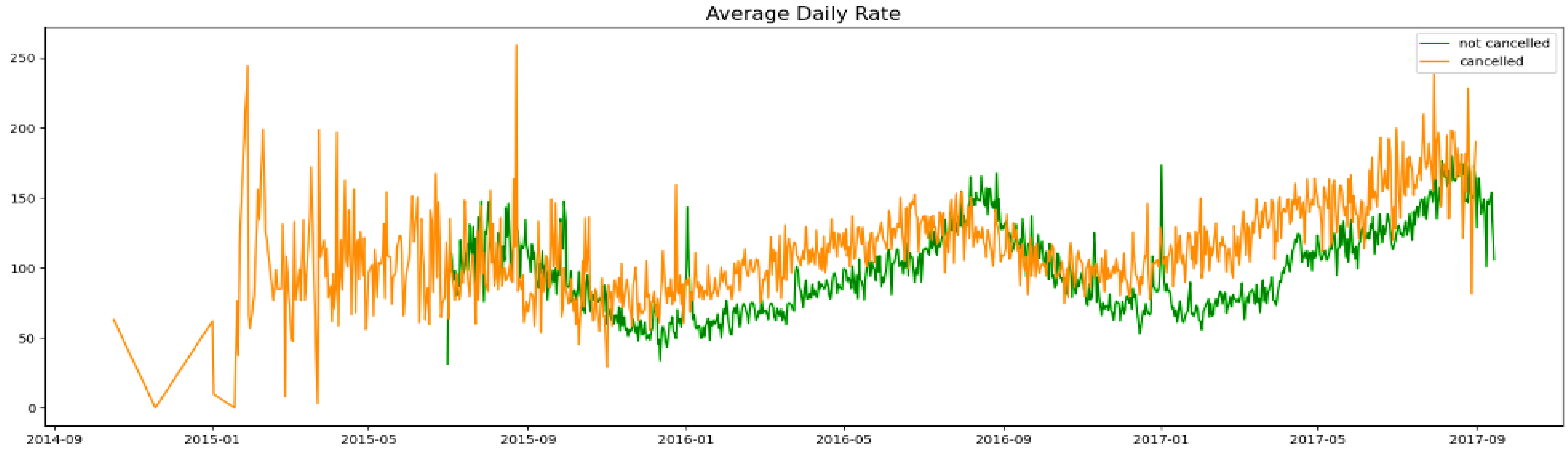
- Most hotel bookings come through **Online Travel Agencies (47%)**, followed by **Group bookings (27%)**. Only about **4%** of guests book directly with the hotel, highlighting the hotels’ strong dependency on third-party channels.

```
cancelled_data['market_segment'].value_counts(normalize = True)
```

```
market_segment
Online TA        0.469696
Groups           0.273985
Offline TA/TO    0.187466
Direct           0.043486
Corporate        0.022151
Complementary    0.002038
Aviation         0.001178
```

- Nearly **47%** of cancellations come from **Online Travel Agencies**, followed by **Groups (27%)** and **Offline TA/TO (19%)**. Other segments, including **Direct** and **Corporate**, contribute minimally, showing lower cancellation tendencies.

ANALYSIS AND FINDINGS - 8



- The graph shows that cancelled bookings usually have higher average daily rates than those that are not cancelled. This suggests that higher prices often lead to more cancellations, likely due to price sensitivity or last-minute changes.

SUGGESTIONS

- 💡 **Tiered Cancellation Policies :** Adjust rules based on booking value, lead time, or season.
- 🎯 **Boost Direct Bookings :** Offer loyalty points or discounts to reduce OTA dependence.
- 📈 **Adopt Dynamic Pricing :** Lower high ADRs during low-demand periods to reduce last-minute cancellations.
- 📢 **Target Off-Peak Months :** Promote bookings during low-demand periods (e.g., January).
- 🔗 **Reassess OTA Agreements :** Minimize revenue loss by revisiting terms with third-party channels.
- 🌍 **Focus on High-Risk Regions :** Improve service and flexibility in areas like Portugal with high cancellation rates.
- 🏖️ **Resort Hotel Discounts :** Offer weekend/holiday deals to maintain low cancellation rates year-round.
- 🛎️ **Boost Direct Bookings :** Use discounts or perks to encourage booking via hotel website.
- 🤝 **Review OTA Partnerships :** Analyze OTA impact on revenue vs. cancellations and adjust strategies accordingly.

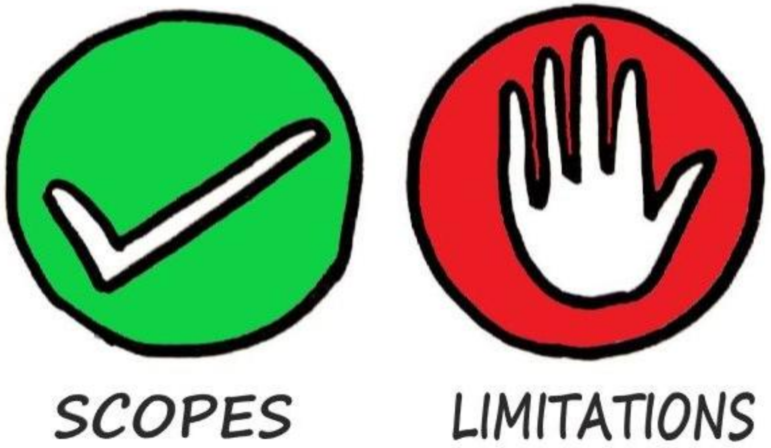
LIMITATIONS & FUTURE SCOPE

Limitations

- Data range limited to 2015–2017 (pre-COVID).
- External factors (e.g., weather, local events) not included.
- No customer review or feedback data was analyzed.

Future Scope

- Include recent data for post-pandemic analysis.
- Apply machine learning models for cancellation prediction.
- Add text-based data (e.g., customer reviews) for deeper insights.



THANK YOU

I appreciate your time and attention.