

Project: Hotel Booking Cancellation Analysis

Data Source

The dataset is sourced from **Kaggle's Hotel Booking Dataset**, which includes hotel reservations data from **2015 to 2019**. This dataset provides a comprehensive view of various attributes related to guests, bookings, and hotel characteristics.

Data Quality Metrics

Accuracy:

- **Missing Values:** The country column initially had missing values, which were filled using the mode. The children column had missing values filled with the median.
- **Incorrect Values:** Outliers were detected in the ADR (Average Daily Rate) column, with some values being excessively high; these were replaced with the median.
- **Inconsistencies:** 'Undefined' values were present in the market segment and meal columns and were replaced with the mode.

Completeness:

- **Missing Data:** The dataset contained some missing values in the country and children columns, which were addressed during data cleaning.
- **Incomplete Data:** Not applicable, as all critical columns for analysis were retained.

Consistency:

- **Duplicate Values:** No duplicates were detected in essential columns after initial data inspection.
- **Conflicting Data:** None detected in the dataset after data cleaning.

Timeliness:

- **Data Latency:** The dataset covers a comprehensive period from 2015 to 2019, providing a good timeframe for analyzing booking trends over the years.

Data Quality Issues

- **Missing Guest Information:** Missing data in the country and children columns was addressed by filling in with the mode and median, respectively.
- **Outliers in ADR:** Extremely high values in the ADR column were replaced with the median to maintain realistic daily rates.
- **Undefined Market Segment and Meal Data:** Some bookings had 'Undefined' values, which could hinder understanding of guest preferences. These were replaced with the most common values.

Data Cleaning Steps Involved

1.Handling Missing Values:

- The country column had missing values filled with the mode.
- The children column had missing values filled with the median.

2.Data Type Conversion:

- Converted reservation_status_date from object type to datetime format for accurate date-based analysis.

3.Column Removal:

- Removed any irrelevant columns that do not contribute to the analysis.

4.Replacing Undefined Values:

- Replaced 'Undefined' values in the market segment and meal columns with their respective modes.

5.Handling Outliers:

- In the ADR column, outliers were adjusted by replacing unusually high values with the median.

Recommendations

Cancellation rates rise as the price dose, in order to prevent cancellations of reservations , hotels could work on their pricing strategies and try to lower the rates for specific hotels on locations. They can also provide some discounts to the consumers.

In the month of Jan , hotels can start campaigns or marketing with a reasonable amount to increase their revenue as the cancellation is the highest in this month.

They can also increase the quality of their hotels and their services mainly in portugal to reduce the cancellation rate.

Evaluate BB Offerings: Investigate the reasons behind the high cancellation rate for the BB meal plan. This could involve enhancing breakfast options or offering more appealing pricing packages.

Promote Alternative Meal Plans: Highlight the advantages of alternative meal plans (HB, SC) in marketing campaigns, emphasizing their perceived value and how they may reduce the likelihood of cancellations.

Set Minimum Lead Times for Different Market Segments: Consider implementing minimum lead times for specific market segments. For example, Group segments show a significantly higher average lead time and subsequent cancellations, you can enforce a longer lead time for those bookings to ensure better planning and commitment from guests.

Conclusion

The analysis of hotel bookings and cancellations reveals several key patterns. Both city and resort hotels show high cancellations during periods of higher Average Daily Rates (ADR), indicating that increased prices are a major factor in cancellations. Specifically, January sees the highest cancellations in city hotels, while June and July have the most cancellations for resort hotels.

Market segment analysis shows that cancellations are particularly high for the "Groups" segment across both hotel types, with the "Online TA" segment also contributing significantly. Additionally, in PRT (Portugal), almost 80% of cancellations occur, severely affecting hotel revenues. The "Groups" market segment has the highest cancellation rates in PRT, and the Bed + Breakfast (BB) meal plan is often linked to these cancellations.

Furthermore, the analysis suggests that longer lead times are associated with higher cancellation rates, especially in the "Groups" market segment. This highlights the importance of adjusting pricing strategies, market targeting, and booking policies to reduce cancellations.

Overall, the project concludes that high ADR, specific market segments (especially Groups), meal plans, and longer lead times are key contributors to hotel cancellations. By addressing these factors, hotels can reduce cancellations and improve revenue stability.