Data Analysis Report: Hotel Booking Cancellation Patterns

Introduction

Cancellations are an inevitable reality in the hospitality industry, impacting revenue and guest satisfaction. But what if you could understand why cancellations happen and identify patterns to minimize them? This analysis delves into hotel booking cancellation data, exploring guest demographics, room preferences, hotel types, and seasonal trends to uncover valuable insights.

Objective

1. Cancellation Dynamics:
   * Explore factors influencing hotel booking cancellations.
   * Uncover patterns to minimize cancellations and enhance guest satisfaction.
2. Data Cleaning and Exploration:
   * Ensure dataset integrity through cleaning.
   * Conduct Exploratory Data Analysis (EDA) for insights.
3. Guest Demographics Analysis:
   * Analyze cancellation trends among families, singles, and couples.
   * Provide strategic insights based on demographics.
4. Room Assignment Impact:
   * Investigate the link between room assignments and cancellations.
   * Assess cancellation rates for desired and undesired room types.
5. City vs. Resort Hotel Comparison:
   * Compare guest ratios and cancellation rates.
   * Recommend tailored strategies for each hotel type.
6. Time Series Analysis:
   * Uncover monthly and seasonal cancellation trends.
   * Provide insights into influencing factors.
7. Limitations Recognition:
   * Acknowledge analysis limitations and propose areas for improvement.
8. Recommendation Formulation:
   * Develop a practical cancellation action plan.
   * Tailor recommendations for different scenarios.
9. Conclusion and Takeaways:
   * Summarize key findings.
   * Emphasize data-driven strategies for revenue and guest satisfaction.

Methodology

Data:

* CSV file containing hotel booking cancellation data.

Tools:

* Power Query (Excel) for data cleaning.
* Excel for data analysis (pivot tables, descriptive statistics, visual graphs).

Approach:

1. Data Cleaning:
   * Removed rows with missing values.
   * Imputed missing "children" counts to zero.
2. Exploratory Data Analysis (EDA):
   * Used pivot tables and graphs to identify patterns and potential cancellation factors.
3. Segmentation Analysis:
   * Compared cancellation rates across guest segments (families, singles, couples).
   * Analyzed differences between city and resort hotels.
4. Time Series Analysis:
   * Examined monthly and seasonal cancellation trends.
5. Limitations:
   * Acknowledged potential data quality and contextual limitations.

Priorities:

* Clear, concise presentation of findings.
* Actionable insights for business decisions.

Data Overview

Data Source:

* For this project, we are using Hotel Booking dataset from Kaggle.
* Dataset Link: <https://www.kaggle.com/datasets/mojtaba142/hotel-booking>

Timeframe:

* 2015 to 2017

Observations:

* 119,390 records

Variables:

* 32 variables covering various booking details.

Key Variables:

* + hotel:
    - Type: Categorical (Resort Hotel or City Hotel)
    - Explanation: Indicates the type of hotel where the booking was made.
  + is\_canceled:
    - Type: Binary (0 or 1)
    - Explanation: Binary variable indicating whether the booking was canceled (1) or not canceled (0).
  + arrival\_date\_year:
    - Type: Numeric (Year)
    - Explanation: The year of arrival for the booking.
  + arrival\_date\_month:
    - Type: Categorical (Month)
    - Explanation: The month of arrival for the booking.
  + adults:
    - Type: Numeric
    - Explanation: The number of adults included in the booking.
  + children:
    - Type: Numeric
    - Explanation: The number of children included in the booking.
  + babies:
    - Type: Numeric
    - Explanation: The number of babies (infants) included in the booking.
  + reserved\_room\_type:
    - Type: Categorical
    - Explanation: The type of room reserved by the guest.
  + assigned\_room\_type:
    - Type: Categorical
    - Explanation: The type of room assigned to the guest upon arrival.
  + reservation\_status:
    - Type: Categorical
    - Explanation: The reservation status. (e.g., Check-Out, Canceled)
  + Room Status:
    - Type: Categorical
    - Explanation: Indicates whether the room status is correct or incorrect.
  + Guest Type:
    - Type: Categorical
    - Explanation: The type of guest (e.g., Single, Two Adults, Family).

Data Cleaning:

Import the data file in Excel and transform it in Power Query. We are going to delete unnecessary features then handle the missing values.

Some steps which are used to clean data are:

* Remove Unwanted columns and missing values
* Check and impute null values.

Let’s clean our Hotel booking cancellation dataset.

Step 1: Import the data and Feature selection in Power Query.

* + Open Excel, navigate to the "Data" tab.
  + Choose "From Text/CSV" and select the dataset from the local system.
  + Import the data and initiate the transformation process in Power Query.
  + Delete unnecessary columns to retain key variables.
  + Verify and set the data types for all columns.
  + Close and Load the cleaned dataset.

Steps 2: Handle the missing values.

Navigate to the ‘Home’ tab -> Find & Select -> Go to Special -> select ‘Blanks’ -> click ‘ok’

Observation:- 4 rows are having blank in column 'children'

Step 3 : Calculate the mean or median of the 'children' column

Following Formulas in Excel :

=AVERAGEIF(hotel\_bookings[children],"<>""")

=MEDIAN(IF(hotel\_bookings[children]<>"",hotel\_bookings[children]))

Below are the findings:

* Total record count : 119,390
* Average : 0.103889903
* Median : 0

Observation: The low average and median indicate that the majority of values in the 'children' column are likely zero.

Step 4: Impute Missing Values with Zero

* Since the average and median are close to zero, it's reasonable to assume that most guests have zero children. We impute the missing values with zero.

Exploratory Data Analysis

**Guest Demographics Distribution:**

Pivot Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Total Guest** | **Cancel Booking** | **Cancel Rate** | **Population Rate** |
| Family | 15253 | 5245 | 34% | 13% |
| Single | 22577 | 6555 | 29% | 19% |
| Two Adults | 81560 | 32424 | 40% | 68% |
| **Grand Total** | **119390** | **44224** | **37%** | **100%** |

Visualization: -

Observations:

* The guest distribution is as follows: The majority of guests, around 68%, consist of pairs such as couples or friends, falling under the category of "Two Adults." Solo travelers, referred to as Singles, make up 19% of the guests, while Families, consisting of groups with children, represent the smallest proportion at 13%.
* Families: despite being the smallest segment, are most likely to cancel. (34% of families cancel).
* Two Adults: This group has the highest cancellation rate among the three categories (40% of couples cancel).
* Single: The lowest cancellation rate is in the "Single" category, with 29% of bookings being cancelled, suggesting a relatively more stable booking pattern among singles.

Interpretations:

* Most Cancels: Families (highest rate for fewer bookings). While Families cancel the most individually, they are so few that their total cancellations are the least.
* Frequent Cancels: Couples (moderately high rate for LOTS of bookings). Two Adults, despite being frequent cancellers, are a large customer base, so understanding their cancellation reasons and offering targeted deals could be key.
* Stable Bookers: Solo travelers (moderate rate for medium bookings). They seem the most reliable bunch, with only 29% of solo trips getting canceled. Offering them incentives for longer stays or group bookings could further attract them.

**Room Assignment Impact on Cancellation Trends**

Pivot Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Total Guest** | **Cancel Booking** | **Cancel Rate** | **Room Assign Rate** |
| Desired | 104473 | 43422 | 42% | 88% |
| Un-desired | 14917 | 802 | 5% | 12% |
| **Grand Total** | **119390** | **44224** |  | **100%** |

Visualization:

Observations:

* Most guests (88%) received their desired room type. This is a positive sign, indicating that the hotels are generally able to meet guest preferences.
* However, the cancellation rate for this group is relatively high at 42%, suggesting that even when guests get their desired room, a substantial number still choose to cancel.
* A smaller proportion of guests (12%) received a different room type than originally reserved ('Un-desired').
* The cancellation rate for this group is notably lower at 5%, indicating that guests who receive an undesired room are less likely to cancel their bookings.

Interpretations:

* While room preferences are fulfilled for most guests, there may be factors beyond room type influencing cancellations, such as changes in travel plans or personal preferences.
* Guests seem more flexible when assigned a room different from their reservation, potentially indicating a higher tolerance for room variations.

**Guest Trends: City vs. Resort Hotels:**

A Deep Dive into Booking and Cancellation Dynamics between City and Resort Hotels

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Total Guest** | **Guest Ratio** |
| City Hotel | 79330 | 66.45% |
| Resort Hotel | 40060 | 33.55% |
| **Grand Total** | **119390** | **100.00%** |

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Total cancellation** | **Cancel Ratio** |
| City Hotel | 33102 | 74.85% |
| Resort Hotel | 11122 | 25.15% |
| **Grand Total** | **44224** | **100.00%** |

Observations:

* Booking Ratio: City hotels host a larger proportion of guests (66.45%) compared to resort hotels (33.55%).
* Cancellation Ratio: City hotels also experience a higher cancellation rate (74.85%) compared to resort hotels (25.15%).

Interpretations:

* City Hotels:
  + Reasons for higher cancellations can be Business travel, competition, seasonality.
  + Solutions to reduce cancellations: Flexible policies, business-friendly perks, dynamic pricing.
* Resort Hotels:
  + Reasons for lower cancellation rates can be Vacation plans, unique offerings, relaxation focus.
  + Solutions to capitalize on their lower cancellation rates by: Highlight features, personalized service, early booking incentives.

**Time Series Analysis**

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Total Guest** | **Cancel Booking** |
| January | 5929 | 1807 |
| February | 8068 | 2696 |
| March | 9794 | 3149 |
| April | 11089 | 4524 |
| May | 11791 | 4677 |
| June | 10939 | 4535 |
| July | 12661 | 4742 |
| August | 13877 | 5239 |
| September | 10508 | 4116 |
| October | 11160 | 4246 |
| November | 6794 | 2122 |
| December | 6780 | 2371 |
| **Grand Total** | **119390** | **44224** |

Visualization:

Observations:

1. Overall Cancellation Rate:
   * The overall cancellation rate across all months is 37%.
2. Monthly Cancellation Rates:
   * Highest Cancellation Rates: April, May, and June have the highest cancellation rates at 41%.
   * Lowest Cancellation Rate: January has the lowest cancellation rate at 30%.
   * Stable Cancellation Rates: November (31%), December (35%), and February (33%) show relatively stable rates.
3. Seasonal Trends:
   * Winter Months (Nov-Dec-Jan): Generally, exhibit lower cancellation rates.
   * Spring and Summer Months (Apr-May-Jun-Jul-Aug): Show higher cancellation rates, peaking in April, May, and June.
4. Variability in Total Guests:
   * August has the highest total guests (13,877), but its cancellation rate (38%) is lower than in some other high-traffic months like June and May.

Interpretations:

1. Possible reasons for seasonal variations in total guest bookings:
   * Summer months are typically peak season for tourism in many parts of the world, due to warmer weather and school holidays.
   * The city hotel might be particularly popular with business travelers who attend conferences or events held throughout the year.
2. Possible reasons for seasonal variations in cancellation rates:
   * Guests who book during the peak season might be more likely to cancel if their plans change or if they find a better deal elsewhere.
   * Winter weather conditions could lead to cancellations, especially for guests who are traveling from warmer climates.

Limitations:

(Acknowledging limitations or constraints in the analysis.)

1. **Guest Demographics Distribution:**
   * **Data Representation:** The analysis relies on the assumption that the provided categories (Family, Single, Two Adults) accurately represent guest demographics. However, more granular categories or additional demographic information could provide a more nuanced understanding.
   * **Cancellation Reasons:** The analysis suggests possible reasons for cancellations but lacks direct insights from guest feedback or surveys. Understanding the specific reasons behind cancellations may require additional qualitative data.
2. **Room Assignment Impact on Cancellation Trends:**
   * **Causation vs. Correlation:** While the analysis highlights a correlation between room assignment and cancellations, it doesn't establish causation. Other factors, such as service quality or external events, could contribute to cancellations independently of room assignment.
   * **Limited Room Types:** The analysis assumes that "Desired" and "Un-desired" adequately capture guest preferences. However, a more detailed breakdown of room types and guest preferences could offer a more accurate assessment.
3. **Guest Trends: City vs. Resort Hotels:**
   * **Simplified Hotel Types:** The analysis categorizes hotels broadly as City and Resort without considering variations within these categories. A more detailed classification based on amenities or location could provide deeper insights.
   * **External Factors:** The analysis suggests reasons for cancellation rates in City and Resort hotels, but external factors such as local events, economic conditions, or global crises may also play a role.
4. **Time Series Analysis:**
   * **Limited Context:** The analysis identifies seasonal and monthly trends, but without external context, the reasons behind cancellations remain speculative. External factors like regional holidays or economic conditions should be considered for a comprehensive understanding.
   * **Data Granularity:** Monthly aggregates may mask variations within each month. Daily or weekly data could provide a more detailed picture of booking and cancellation patterns.

**General Limitations:**

1. **External Factors:**
   * External factors such as economic changes, public health crises, or natural disasters, not accounted for in the dataset, could influence booking and cancellation patterns.
2. **Lack of Qualitative Insights:**
   * The analysis primarily focuses on quantitative data. Qualitative insights from guest feedback, surveys, or interviews could provide a more comprehensive understanding of guest behavior.
3. **Dynamic Nature of Hospitality Industry:**
   * The hotel industry is dynamic, and trends may evolve over time. The analysis captures a specific period, and findings may not be universally applicable across different timeframes.

Recommendations:

(Offers actionable suggestions based on the analysis.)

Cancellation Action Plan: Quick Tips

1. **Guest Type**:

* Families: Flexible policies, family packages, bundled attractions.
* Singles: Long-stay discounts, solo-focused experiences.
* Couples: Off-season deals, romance/wellness retreats.

1. **Room Assignment**:

* Understand "desired" cancellations. Offer upgrades for undesired rooms.

1. **City vs. Resort**:

* City Hotels: Business perks, dynamic pricing.
* Resort Hotels: Highlight unique features, personalize service, early booking incentives.

1. **Time Trends**:

* Winter: Off-season promotions, target low-cancellation months.
* Peak Season: Address cancellation reasons with targeted offers.
* High-Volume Months: Identify and replicate success factors.

Bonus: Track cancellation reasons, personalize communication, stay competitive.

Remember, data-driven insights = happy guests & healthy revenue!

Conclusion: Key Takeaways from Booking Cancellation Analysis

Guest Demographics:

* Families cancel the most but represent a smaller guest segment.
* Couples cancel frequently though they form a large guest base.
* Singles are the most reliable bookers with the lowest cancellation rate.

Room Assignment:

* Fulfilling room preferences doesn't guarantee low cancellations.
* Guests receiving undesired rooms cancel less often, suggesting flexibility.

City vs. Resort Hotels:

* City hotels experience higher occupancy and cancellations, likely due to business travel and seasonality.
* Resort hotels enjoy lower cancellations, driven by vacation plans and unique offerings.

Time Series Analysis:

* Spring/Summer (Apr-Jun) sees peak cancellations, while Winter (Nov-Dec-Jan) shows the least.
* August has a high guest volume, but a lower cancellation rate compared to other peak months.

Actionable Insights:

* Tailor cancellation policies and deals to each guest segment (families, couples, singles).
* Improve communication and offer alternative options for undesired room assignments.
* Implement dynamic pricing and business-friendly perks for city hotels.
* Highlight unique features and personalize services for resort hotels.
* Offer early booking incentives and target off-peak seasons for both types of hotels.

By understanding these key takeaways and implementing targeted strategies, Hotel can effectively reduce cancellations, optimize guest satisfaction, and boost the revenue.

GitHub Repository

Repo Link :

<https://github.com/abhishekmishra8/Excel-Project-Hotel-Booking-Cancellation-Data-Analysis>

Thank You