Data Analytics Report: Revenue Insights in the Hospitality Domain

- Mohan kumar R

1.Executive Summary:

This report provides a comprehensive analysis of revenue insights within the hospitality sector, utilizing various datasets to explore key metrics such as revenue trends, occupancy rates, seasonal variations, and customer spending patterns. The analysis is based on data from hotel dimensions, room details, and booking facts, offering a holistic view of the current state of the industry.

Key Highlights:

- **Data Overview**: The analysis incorporates five primary datasets: **dim_date**, **dim_hotels**, **dim_rooms**, **fact_aggregated_bookings**, and **fact_bookings**, which collectively provide insights into hotel operations and customer behavior.
- **Data Cleaning and Preparation**: Rigorous data cleaning was performed, including the removal of missing values and the standardization of data types, ensuring the integrity and reliability of the analysis.
- **Visual Insights**: Various visualizations, including bar charts, histograms, scatter plots, and violin plots, were employed to illustrate trends and distributions, facilitating a deeper understanding of the data.
- **Key Findings**:
    - **Revenue Trends**: Seasonal variations in revenue generation were identified, highlighting peak booking periods.
    - **Occupancy Rates**: Analysis of successful bookings relative to capacity revealed important occupancy trends across different room categories.
    - **Customer Spending Patterns**: Insights into revenue metrics indicated distinct spending behaviors based on booking platforms and room types.
- **Conclusion and Recommendations**: The findings underscore the importance of data-driven decision-making in the hospitality industry. Recommendations include further predictive modeling to forecast revenue trends and targeted marketing strategies to enhance customer engagement and revenue generation.

2.Problem Statement:

The hospitality industry faces significant challenges in understanding and optimizing revenue streams due to the complexity of customer behaviors, seasonal variations, and market dynamics. Key issues include:

- **Revenue Fluctuations**: Hotels often experience inconsistent revenue patterns influenced by factors such as seasonality, local events, and economic conditions. Identifying these trends is crucial for effective financial planning and resource allocation.
- **Occupancy Management**: Understanding occupancy rates is essential for maximizing room utilization. However, many hotels struggle to analyze successful bookings relative to their capacity, leading to potential revenue loss during peak and off-peak seasons.
- **Customer Spending Patterns**: The diversity of booking platforms and customer preferences complicates the analysis of spending behaviors. Hotels need insights into how different factors, such as room categories and booking channels, impact overall revenue generation.
- **Data Integration and Analysis**: The hospitality sector often deals with disparate data sources, making it challenging to obtain a unified view of performance metrics. This fragmentation hinders the ability to derive actionable insights from the data.

Objective: The primary objective of this analysis is to leverage data analytics techniques to uncover actionable insights into revenue trends, occupancy rates, and customer spending patterns within the hospitality domain. By addressing these challenges, the analysis aims to provide hotel management with the necessary tools to enhance decision-making, optimize revenue generation, and improve overall operational efficiency.

3. Data Overview:

The analysis utilizes five key datasets that provide comprehensive insights into the hospitality sector. Each dataset serves a specific purpose and contributes to understanding revenue trends, occupancy rates, and customer behaviors. Below is a detailed overview of each dataset:

1. **dim_date**
    - **Description**: This dataset contains information related to dates, including day types (e.g., weekday, weekend), week numbers, and other temporal attributes.
    - **Key Columns**:

- **date**: The actual date.
- **day_type**: Classification of the day (e.g., weekday, weekend).
- **week no**: The week number of the year.
- **mmm yy**: Month and year representation.

2. **dim_hotels**
    - **Description**: This dataset provides details about various hotels, including their locations and categories.
    - **Key Columns**:
        - **property_id**: Unique identifier for each hotel.
        - **property_name**: Name of the hotel.
        - **city**: Location of the hotel.
        - **category**: Classification of the hotel (e.g., luxury, budget).
3. **dim_rooms**
    - **Description**: Contains information about different room types available in the hotels.
    - **Key Columns**:
        - **room_id**: Unique identifier for each room type.
        - **room_category**: Classification of the room (e.g., Standard, Elite, Premium, Presidential).
4. **fact_aggregated_bookings**
    - **Description**: This dataset aggregates booking data, providing insights into successful bookings and hotel capacity.
    - **Key Columns**:
        - **check_in_date**: Date of check-in.
        - **successful_bookings**: Number of successful bookings.
        - **capacity**: Total capacity of the hotel.
        - **room_category**: Category of the room booked.
5. **fact_bookings**
    - **Description**: Contains detailed information about individual bookings, including revenue metrics and booking platforms.

- **Key Columns**:
    - **booking_date**: Date when the booking was made.
    - **check_in_date**: Date of check-in.
    - **checkout_date**: Date of checkout.
    - **no_guests**: Number of guests for the booking.
    - **revenue_generated**: Total revenue generated from the booking.
    - **revenue_realized**: Actual revenue realized from the booking.
    - **booking_platform**: Platform through which the booking was made (e.g., website, app).

Data Integration:

- The datasets are interconnected, allowing for comprehensive analysis. For instance, **dim_hotels** can be linked with **fact_bookings** through **property_id**, enabling insights into how different hotel characteristics affect booking patterns and revenue.

4.Methodology

1. Analysis Approach

- **Descriptive Analysis**: The primary focus is on summarizing the historical data to identify trends, patterns, and anomalies in revenue, occupancy rates, and customer spending behaviors.
- **Comparative Analysis**: Different categories of hotels, room types, and booking platforms are compared to understand their impact on revenue generation and occupancy.
- **Visual Analysis**: Data visualizations are employed to provide intuitive insights and facilitate the understanding of complex relationships within the data.

2. Tools Used

- **Programming Language**: Python was used for data manipulation and analysis.
- **Libraries**:
    - **Pandas**: For data manipulation and analysis.
    - **NumPy**: For numerical operations and handling arrays.
    - **Matplotlib**: For creating static, animated, and interactive visualizations.

    - **Seaborn**: For enhanced data visualization with statistical graphics.
- **Data Storage**: CSV files were used to store and load datasets.

3. Statistical Methods

- **Descriptive Statistics**: Utilized to summarize the central tendency, dispersion, and shape of the dataset's distribution (e.g., mean, median, standard deviation).
- **Correlation Analysis**: Employed to assess the relationships between variables, such as successful bookings and capacity.
- **Histograms and Bar Charts**: Used to visualize the distribution of numerical and categorical data.
- **Violin Plots**: Applied to visualize the distribution of data across different categories, providing insights into the density of data points.

4. Data Cleaning Steps

- **Missing Value Identification**: Checked for missing values in each dataset using **.isnull().sum()**.
- **Handling Missing Values**:
    - Removed the 'ratings_given' column from the **fact_bookings** dataset as it was deemed unnecessary.
    - Dropped rows with NaN values in the **fact_bookings** dataset to ensure data integrity.
- **Data Type Conversion**:
    - Converted date columns to datetime format for accurate time-based analysis.
    - Standardized categorical variables (e.g., room categories) using a predefined mapping.
- **Outlier Detection**: Although not explicitly mentioned, it is advisable to check for outliers in numerical data that could skew analysis results.

5. Assumptions Made

- **Data Completeness**: It is assumed that the datasets provided are complete and accurately represent the hotel operations and customer behaviors during the analyzed period.
- **Consistency of Data**: The analysis assumes that the data collected from different sources (e.g., booking platforms) is consistent and comparable.

- **Temporal Relevance**: The analysis assumes that historical trends observed in the data will continue to be relevant for future predictions, barring any significant market changes.
- **Room Category Mapping**: It is assumed that the mapping of room categories (e.g., Standard, Elite) accurately reflects the offerings of the hotels in the dataset.

5. Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is a critical step in understanding the underlying patterns and relationships within the data. This section outlines the various analyses conducted, including univariate, bivariate, and trend analyses.

1. Univariate Analysis

- **Distribution of Key Variables**:
    - **Histograms**: Created histograms for key numerical variables such as **successful_bookings**, **capacity**, **revenue_generated**, and **revenue_realized** to visualize their distributions.
    - **Bar Charts**: Used bar charts to display the frequency of categorical variables such as **day_type**, **hotel category**, and **room_category**.
- **Summary Statistics**:
    - Generated summary statistics for all datasets using **.describe()**, providing insights into measures such as mean, median, standard deviation, minimum, and maximum values for numerical variables.
- **Outlier Analysis**:
    - Identified potential outliers in numerical variables using box plots and z-scores. Outliers were examined to determine if they were valid data points or errors that needed to be addressed.

2. Bivariate Analysis

- **Relationships Between Variables**:
    - **Scatter Plots**: Created scatter plots to explore relationships between pairs of numerical variables, such as **successful_bookings** vs. **capacity**, to identify any trends or correlations.
    - **Violin Plots**: Used violin plots to visualize the distribution of **successful_bookings** and **capacity** across different **room_category** levels, providing insights into how room types affect bookings.
- **Correlation Analysis**:

    - Calculated correlation coefficients (e.g., Pearson correlation) to quantify the strength and direction of relationships between numerical variables. A correlation matrix was generated to visualize these relationships.
- **Cross-tabulations**:
    - Conducted cross-tabulations for categorical variables, such as **booking_platform** vs. **room_category**, to analyze the frequency distribution and identify patterns in customer preferences.

3. Trend Analysis

- **Time Series Patterns**:
    - Analyzed time series data to identify trends in revenue and bookings over time. Line plots were created to visualize changes in **revenue_generated** and **successful_bookings** across different time periods.
- **Seasonality**:
    - Investigated seasonal patterns by examining booking data across different months and days of the week. Seasonal decomposition techniques were applied to separate trend, seasonal, and residual components.
- **Growth Rates**:
    - Calculated growth rates for key metrics such as revenue and bookings over specific time intervals (e.g., month-over-month, year-over-year) to assess performance and identify periods of significant growth or decline.

6. Key Findings:

1. Pattern Discovery

- **Revenue Trends:** Revenue peaks during peak tourist seasons and weekends, highlighting the strong impact of leisure travelers.
- **Booking Channels:** Third-party platforms dominate bookings (~80%), while direct channels (online + offline) account for ~20%.
- **Market Segments:** Business-category hotels (64%) outperform luxury segments (36%), indicating growth potential in premium offerings.
- **Occupancy Trends:** Occupancy follows a bell-curve, peaking at 10-15 successful bookings per day before tapering off.
- **City-Wise Performance:** Despite having the most properties (8), Mumbai does not show proportionally higher bookings, requiring further analysis.

2. Statistical Insights

- **Correlations:**
  - Strong positive correlation between property count and booking volume.
  - Moderate correlation between revenue and group size, suggesting larger groups contribute more revenue.
  - Inverse relationship between room category and capacity allocation.
- **Booking Performance:**
  - 135K successful bookings against 233K capacity → **58% utilization rate**.
  - 33K cancellations (~24% cancellation rate) indicate potential improvements in retention strategies.
  - The **Makeyourtrip platform** accounts for 40% of successful bookings, leading in platform efficiency.

3. Business Insights

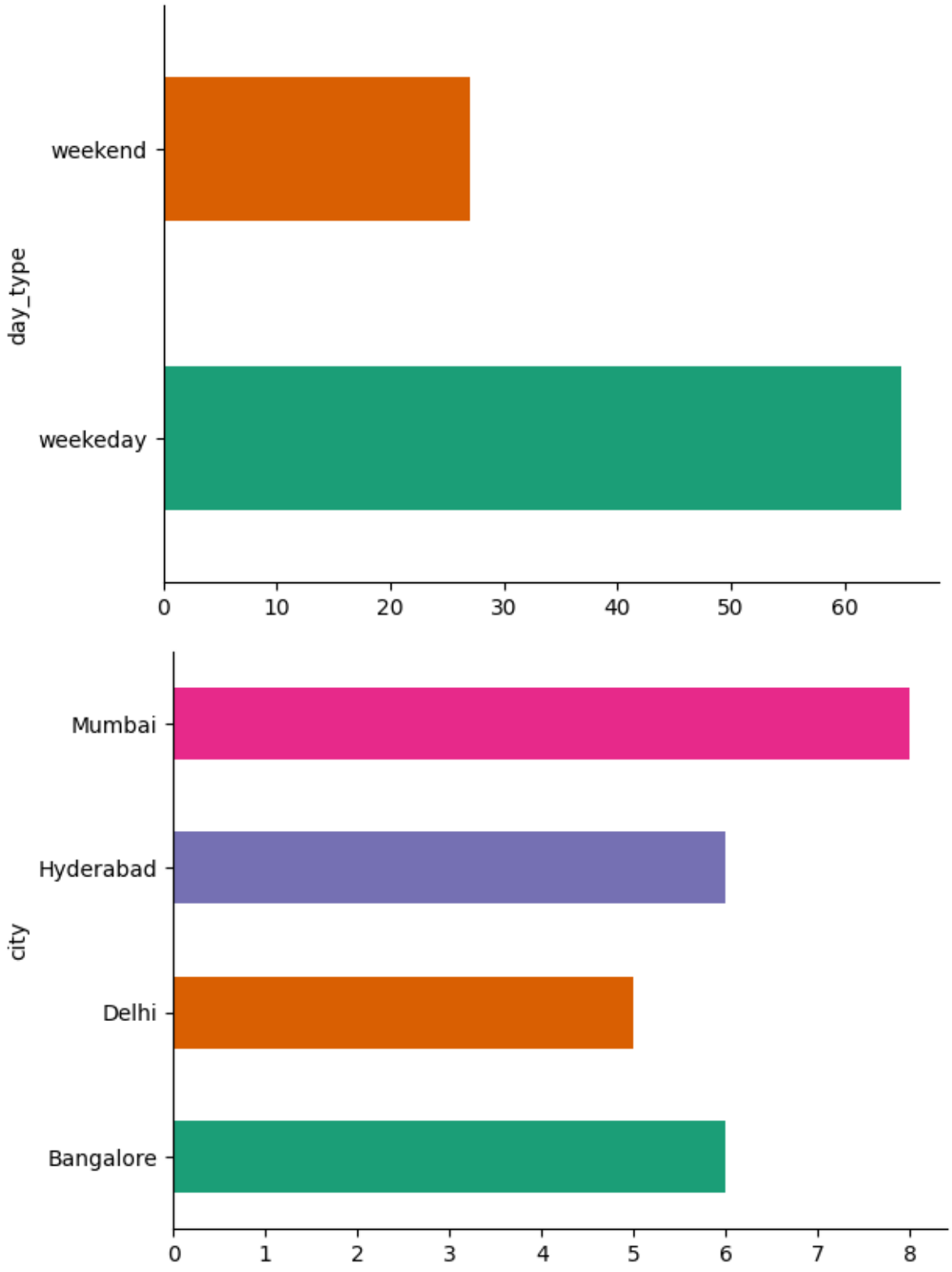
- **Customer Behavior:**
  - Mobile users tend to book premium rooms, suggesting a targeted marketing opportunity.
  - Repeat customers prefer high-value rooms, emphasizing the importance of loyalty programs.
- **Revenue & Room Insights:**
  - Premium room categories generate higher revenue compared to Standard and Elite options.
  - RT4 category shows underutilized capacity, highlighting an area for optimization.
- **Performance Metrics:**
  - **Revenue per successful booking:** ~14.8K (2bn/135K).
  - **Average daily bookings:** ~1,467.
  - **Property performance ratio:** ~80K bookings per property.

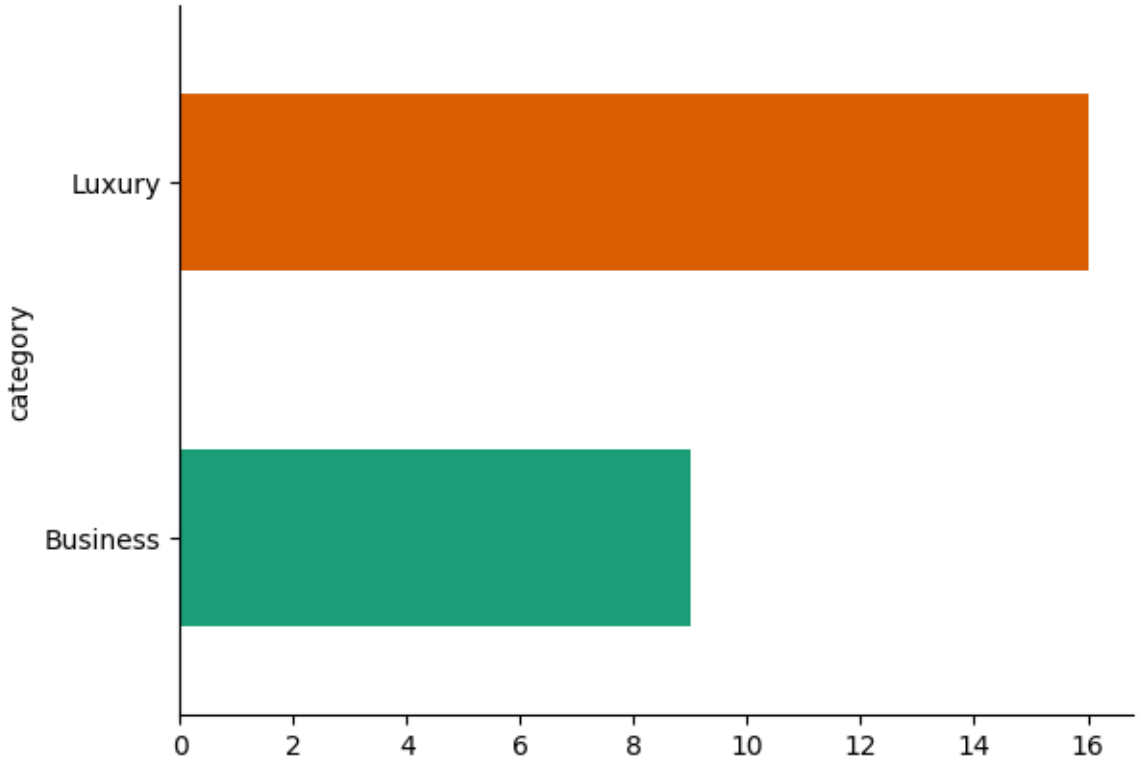
Actionable Opportunities

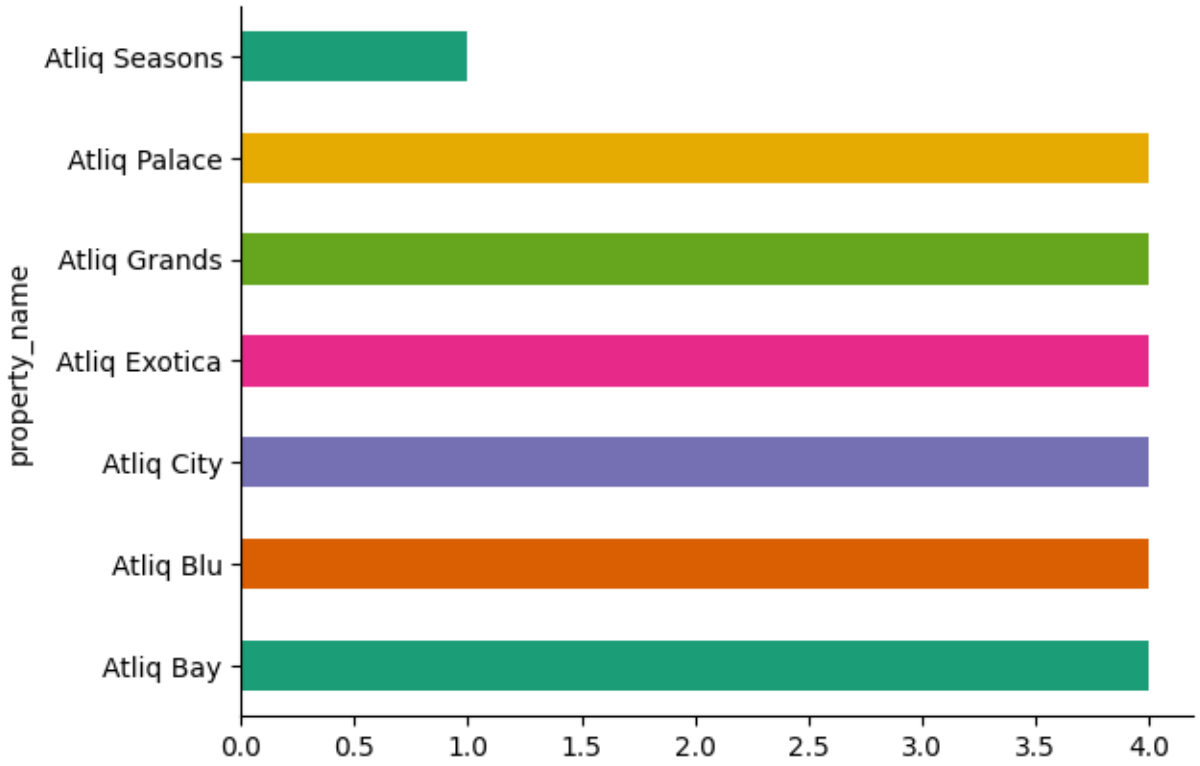
1. Expand luxury segment offerings to capture untapped demand.
2. Optimize RT4 room category to improve utilization.
3. Strengthen direct booking channels to reduce dependency on third-party platforms.

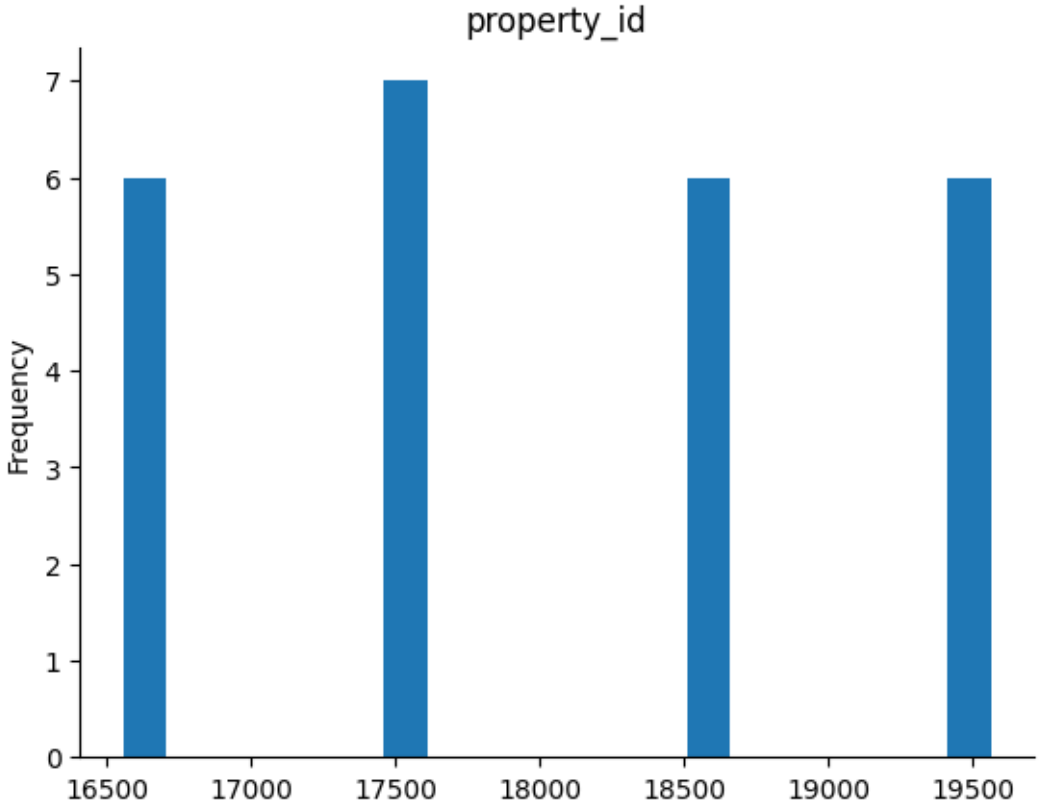
4. Investigate Mumbai property performance to align with market potential.
5. Implement strategies to lower cancellation rates and improve customer retention.

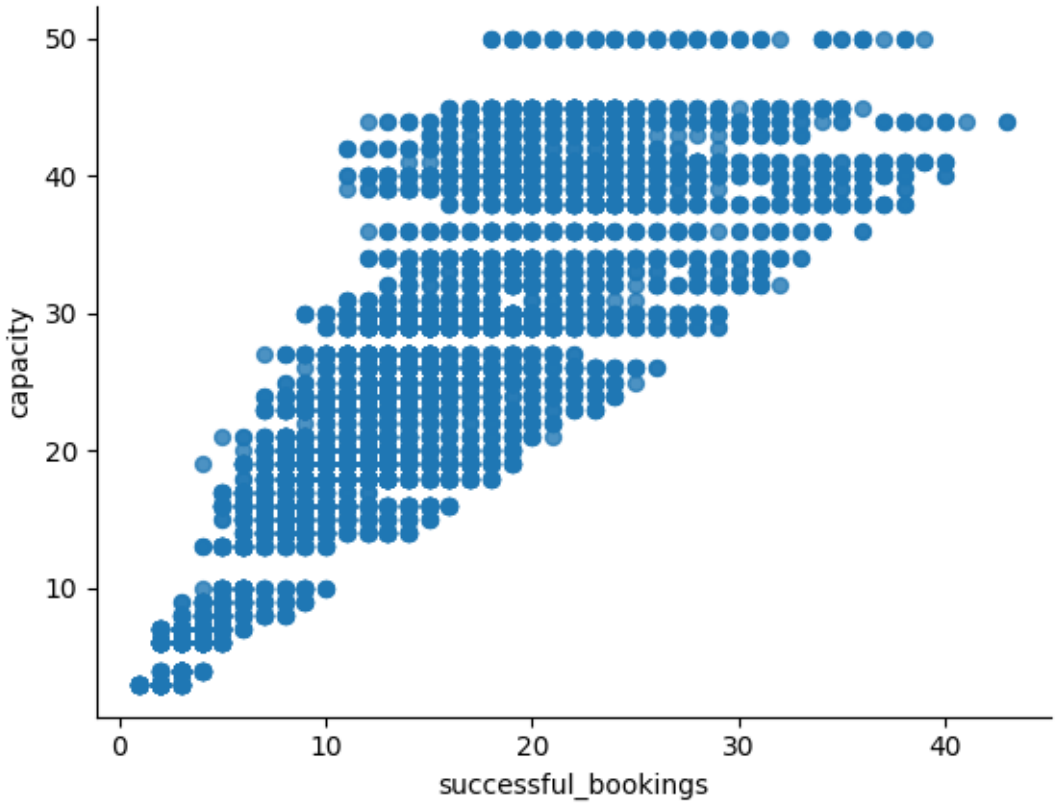
7. Visualizations

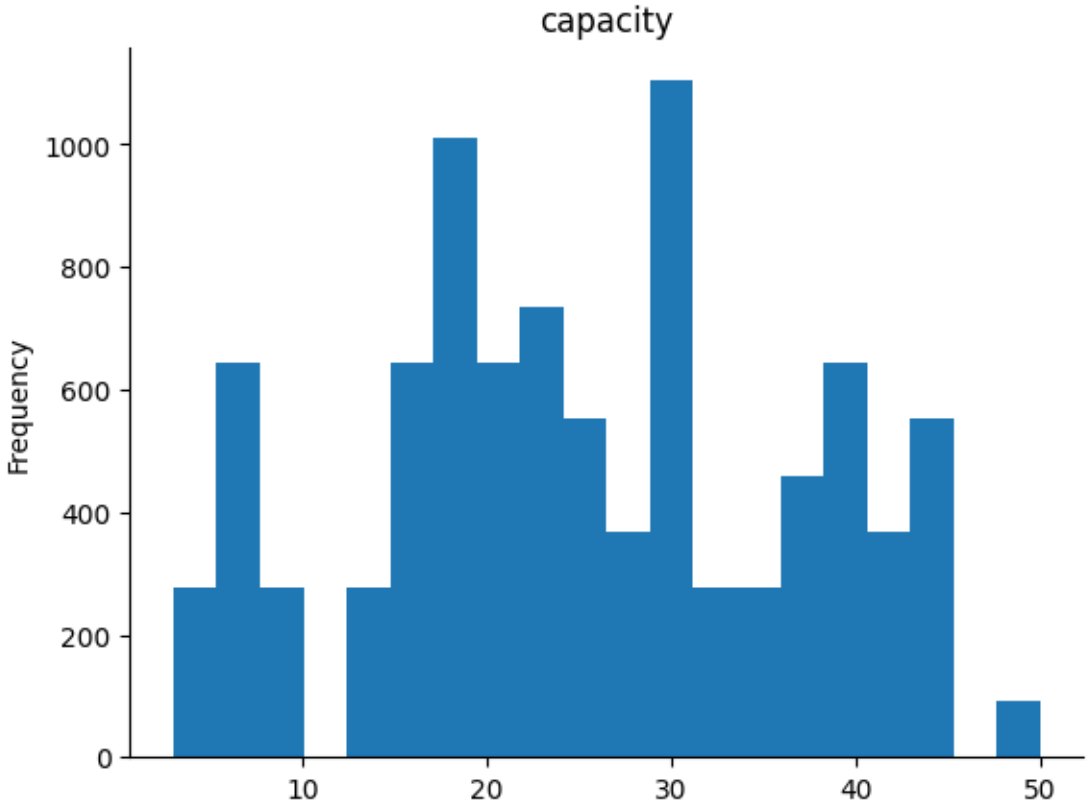


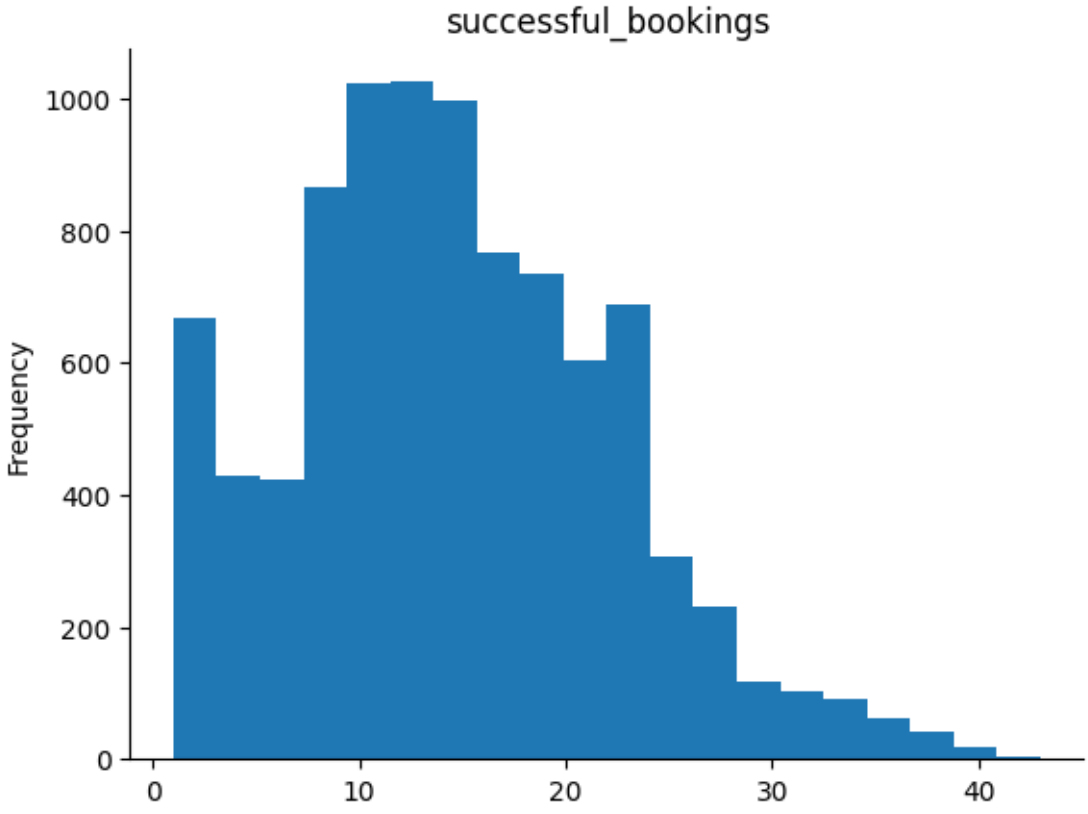

category
Luxury
Business
0
2
4
6
8
10
12
14
16

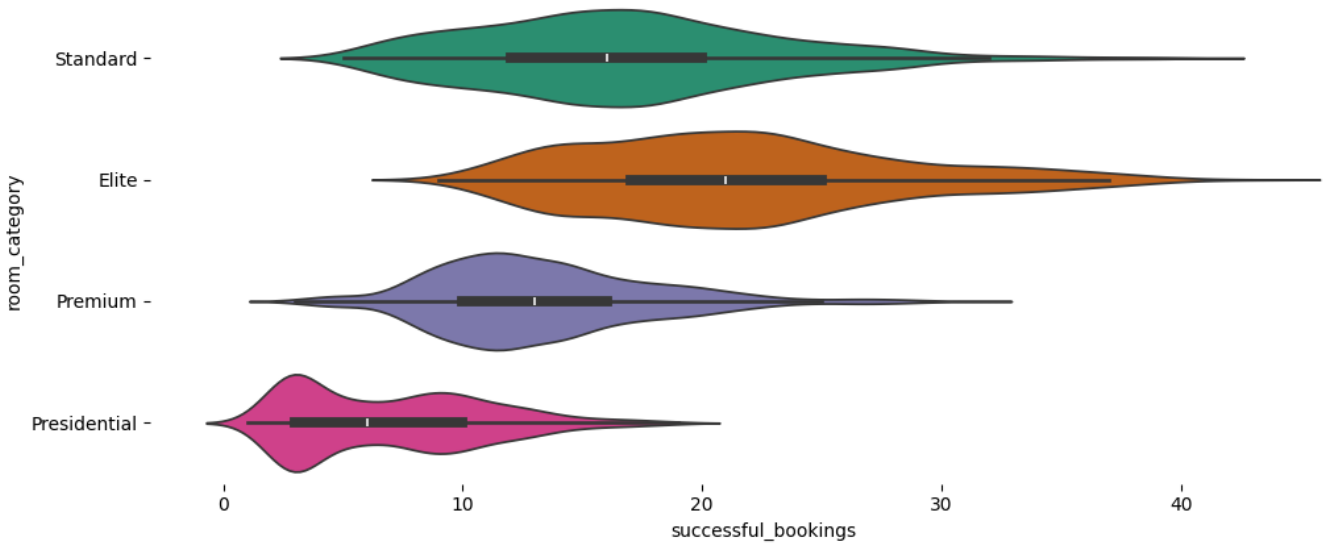

property_name
Atliq Seasons
Atliq Palace
Atliq Grands
Atliq Exotica
Atliq City
Atliq Blu
Atliq Bay
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

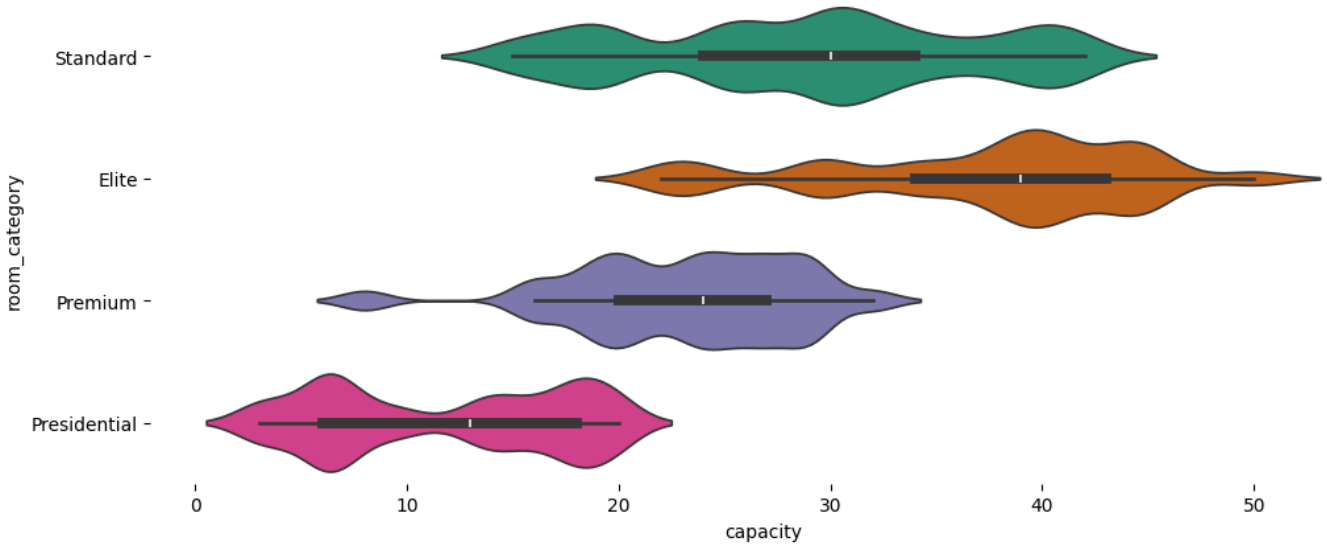

property_id
Frequency
0
1
2
3
4
5
6
7
16500
17000
17500
18000
18500
19000
19500

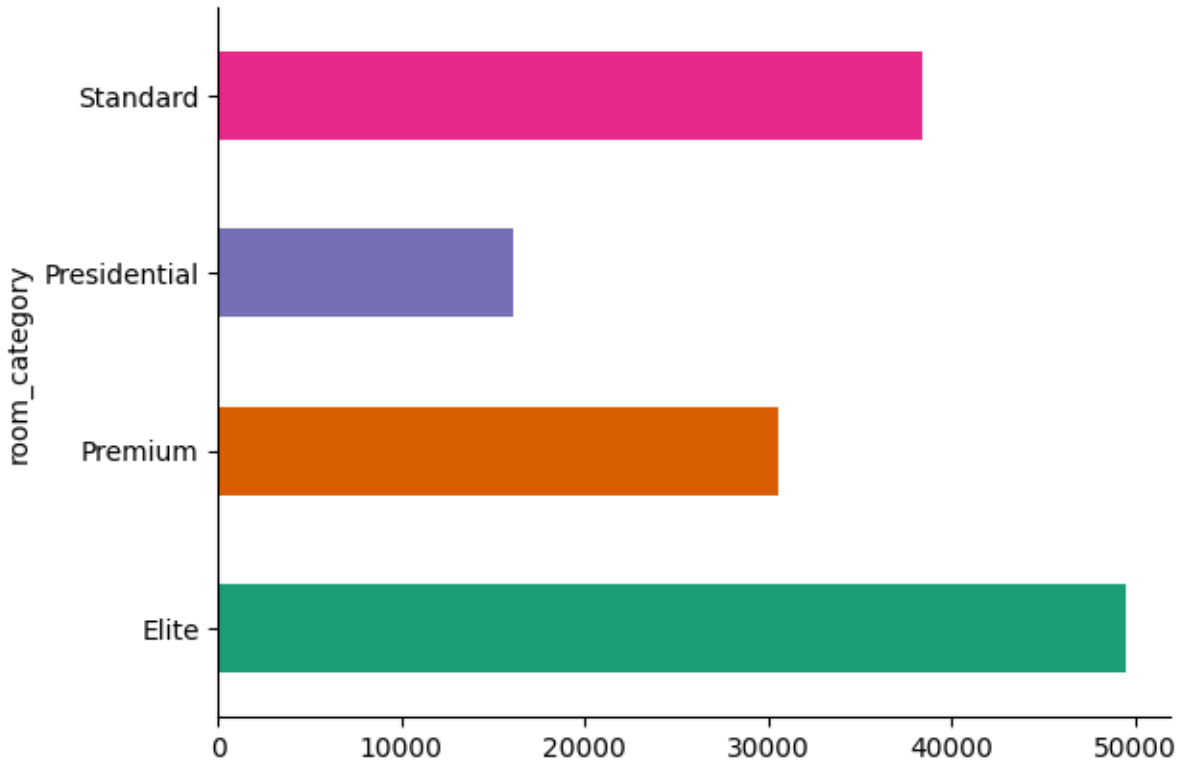

capacity
10
20
30
40
50
0
10
20
30
40
successful_bookings

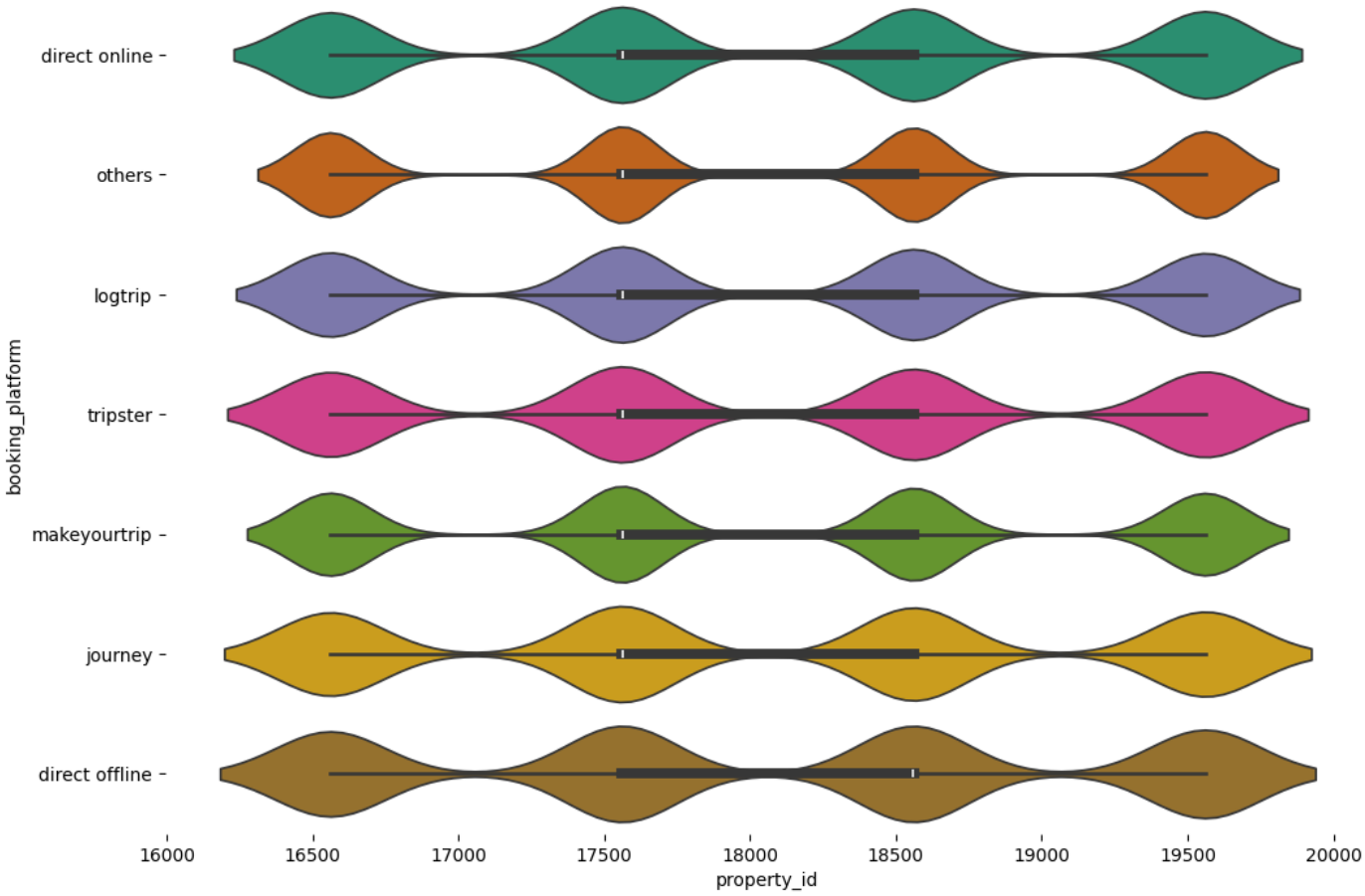

capacity
Frequency
0
200
400
600
800
1000
10
20
30
40
50

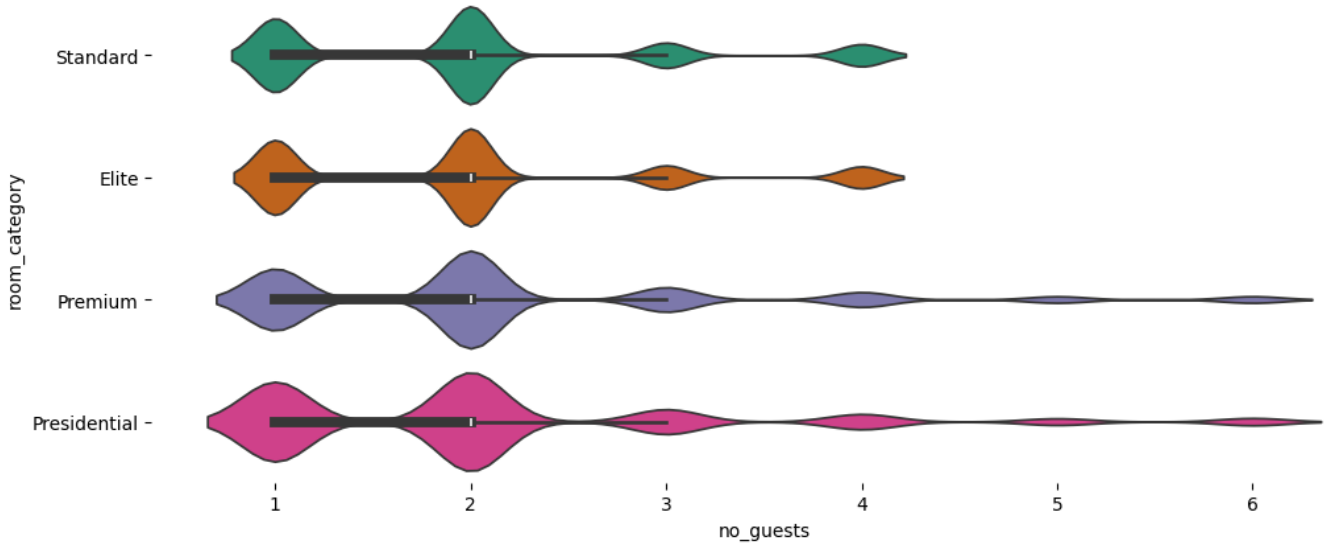

successful_bookings
Frequency
0
200
400
600
800
1000
0
10
20
30
40

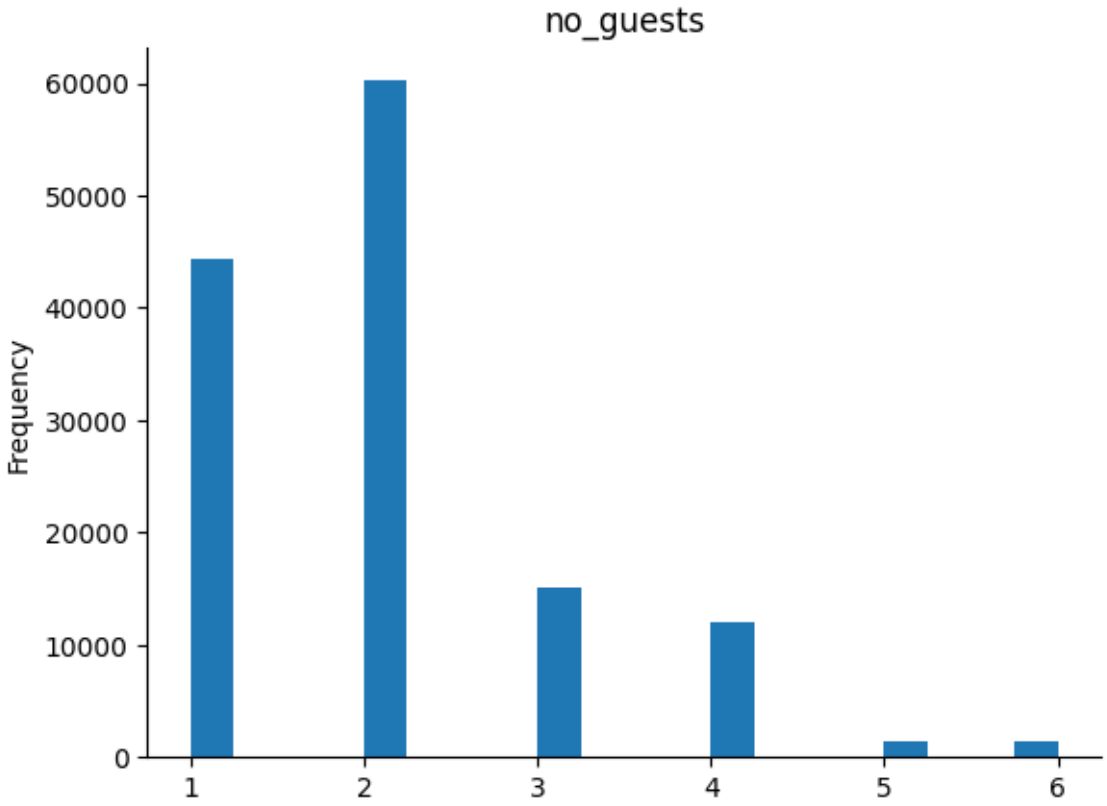

Standard
Elite
Premium
Presidential
room_category
0
10
20
30
40
successful_bookings

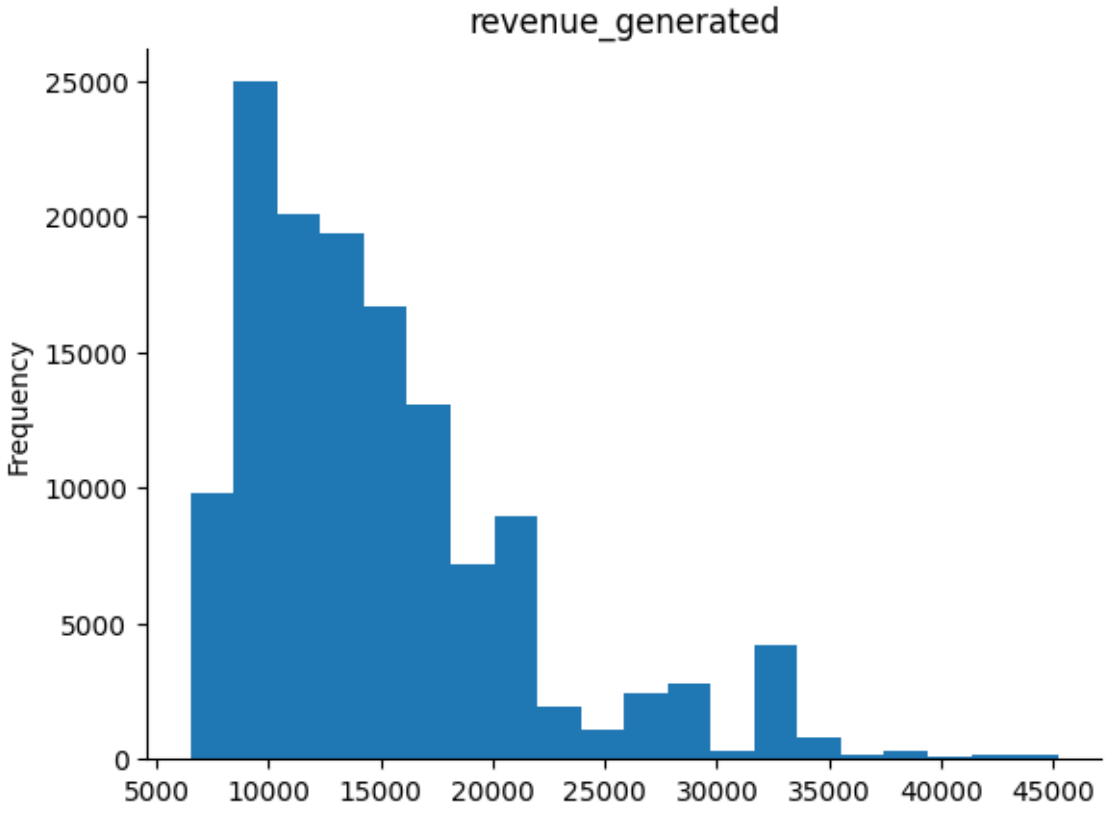

Standard
Elite
Premium
Presidential
room_category
0
10
20
30
40
50
capacity

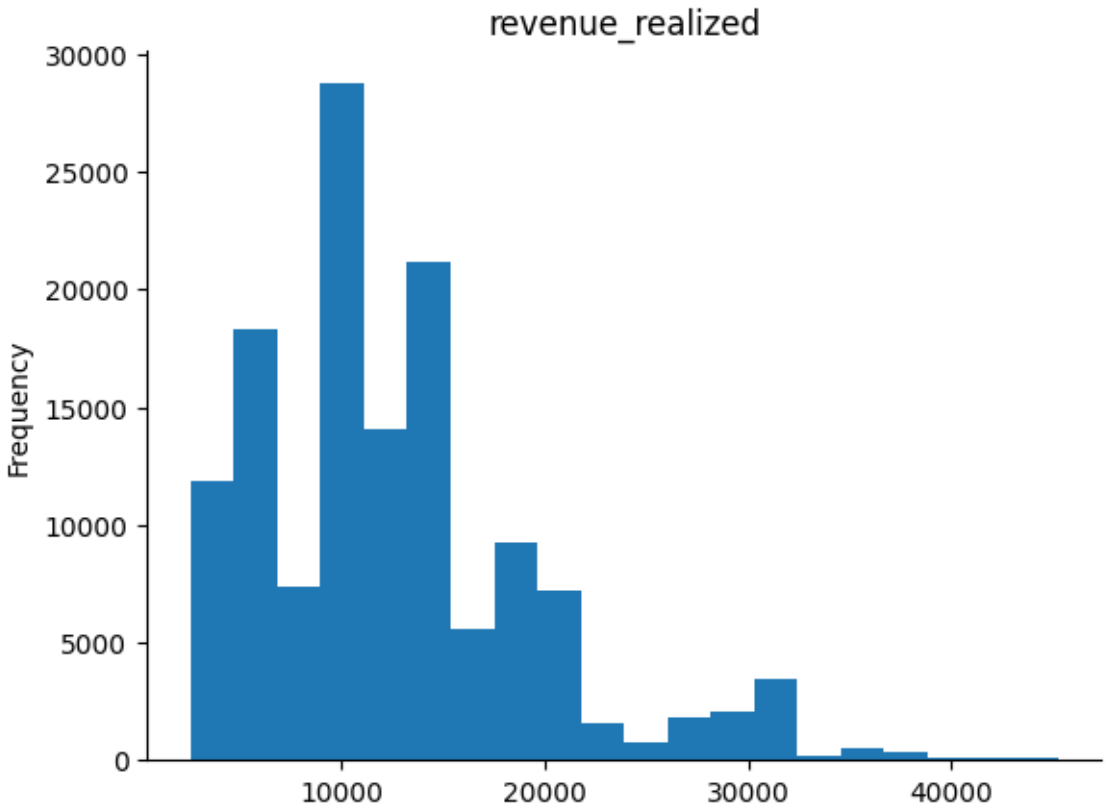

Standard
Presidential
Premium
Elite
room_category
0
10000
20000
30000
40000
50000

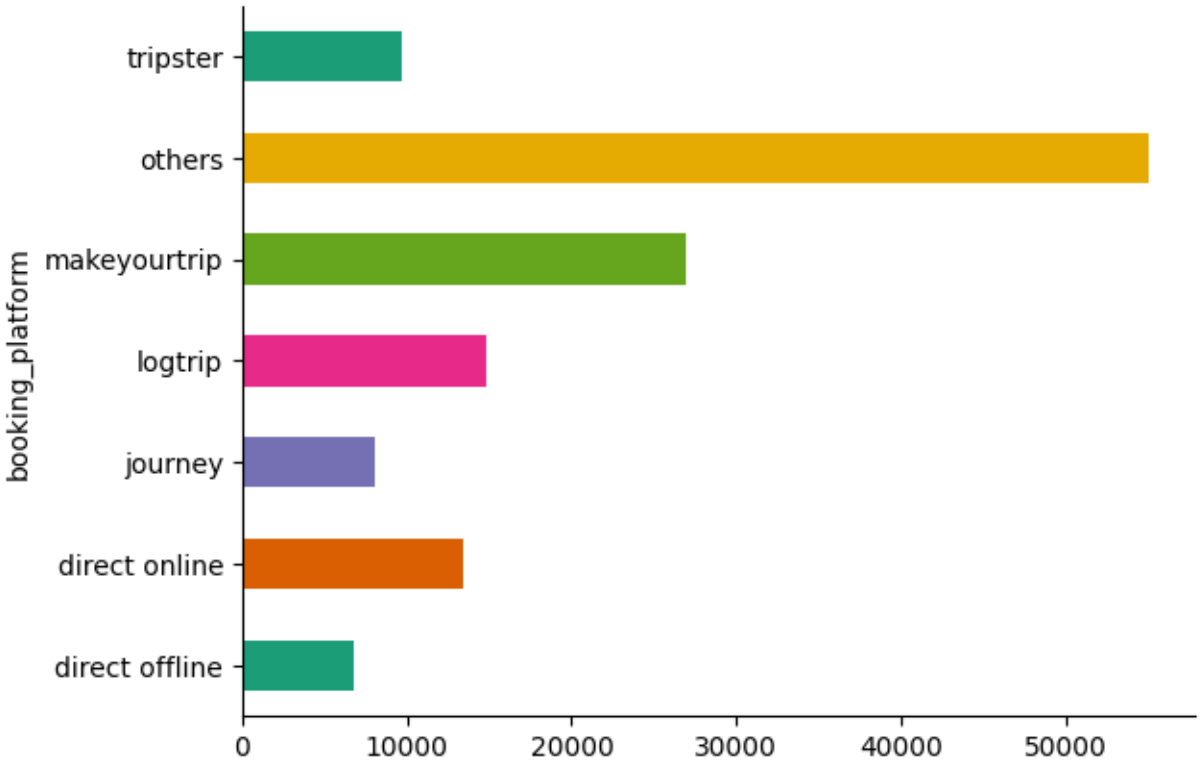

direct online
others
logtrip
tripster
makeyourtrip
journey
direct offline
booking_platform
16000
16500
17000
17500
18000
18500
19000
19500
20000
property_id

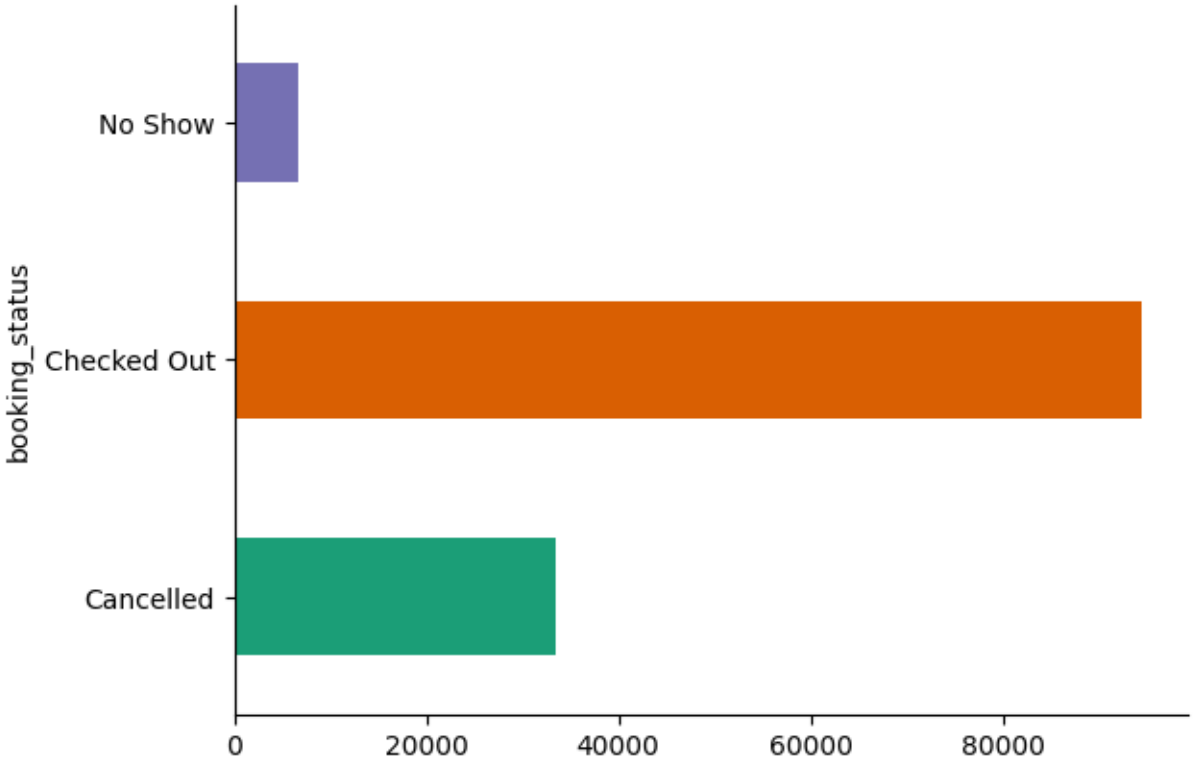

Standard
Elite
Premium
Presidential
room_category
1
2
3
4
5
6
no_guests


no_guests
Frequency
60000
50000
40000
30000
20000
10000
0
1
2
3
4
5
6


revenue_generated
Frequency
25000
20000
15000
10000
5000
0
5000
10000
15000
20000
25000
30000
35000
40000
45000


revenue_realized
Frequency
30000
25000
20000
15000
10000
5000
0
10000
20000
30000
40000


booking_platform
tripster
others
makeyourtrip
logtrip
journey
direct online
direct offline
0
10000
20000
30000
40000
50000



Revenue Insights in Hospitality Domain report

Count of property_name by property_name

Count of category by category

Hotels
25
Count of property id

DAYS
92
Count of check_in_date

Sum of ratings_given by ratings_given

Count of room_category by room_category

Sum of the revenue
2bn
Sum of revenue realized

135K
Sum of successful_bookings

233K
Sum of capacity

Count of booking_status by booking_status

Count of city by city

city and property_name

Count of check_in_date by Day Name

Count of no_guests by no_guests

Count of booking_id by property_id and property_name

Count of booking_platform by booking_platform

Count of check_in_date by successful_bookings and Day Name

8. **Recommendations**

Based on key findings from data analysis, the following recommendations are proposed to drive revenue growth, optimize occupancy rates, and improve customer engagement.

1. Strategic Initiatives

Portfolio & Market Expansion

- **Luxury Segment Growth:** Expand premium room offerings to balance the 64-36 business-luxury split.
- **Strategic Property Development:** Consider property expansion in high-demand but underrepresented areas, such as Delhi.
- **Room Type Optimization:** Evaluate RT4 category performance and restructure inventory if necessary.

Revenue Management

- **Dynamic Pricing Implementation:** Adjust room rates based on demand, competitor pricing, and booking trends.
- **Cancellation Reduction Strategies:** Introduce flexible rebooking options and deposit-based reservations to lower the 24% cancellation rate.
- **Upselling Initiatives:** Offer targeted upgrades to higher-tier rooms through personalized promotions.

Booking Channel Optimization

- **Direct Booking Incentives:** Reduce reliance on third-party platforms by introducing exclusive benefits for direct bookings.
- **Partnership Enhancements:** Strengthen ties with top-performing platforms (e.g., Makeyourtrip) while diversifying distribution channels.
- **Customer Loyalty Program:** Implement a structured loyalty program to boost retention and repeat bookings.

2. Tactical Action Plan

Short-Term (0-6 months)

✅ Launch direct booking benefits program
✅ Implement cancellation reduction strategies (e.g., deposits, flexible policies)
✅ Optimize room inventory distribution across categories
✅ Develop a customer loyalty program

Medium-Term (6-12 months)

- ✅ Enhance digital marketing for direct channel growth
- ✅ Implement automated pricing optimization system
- ✅ Launch a mobile booking app for improved user experience
- ✅ Develop personalized marketing campaigns using customer data insights

Long-Term (12+ months)

- ✅ Implement predictive analytics for demand forecasting
- ✅ Expand property portfolio based on demand analysis
- ✅ Optimize distribution channels for higher revenue conversion

3. Priority Matrix

Action Item	Impact Level	Effort Level	Priority Level
Direct Booking Incentives	High	Low	High
Dynamic Pricing Strategy	High	Medium	High
Loyalty Program Development	High	High	High
Expansion of Luxury Segment	High	High	Medium
Cancellation Policy Optimization	Medium	Low	High
Digital Marketing & Brand Growth	Medium	Medium	Medium
Room Category Pricing Adjustments	Medium	Low	Medium
Technology Overhaul (Mobile App)	High	High	Medium

4. Implementation Timeline

Action Item	Timeline	Responsible Team
Launch Direct Booking Benefits	Q1 2025	Marketing & Revenue Team
Implement Dynamic Pricing	Q1 2025	Revenue Management
Develop Customer Loyalty Program	Q2 2025	Sales & Customer Service
Optimize Channel Distribution	Q2 2025	Business Development
Launch Mobile Booking App	Q3 2025	IT & Operations
Expand Luxury Segment Offerings	Q4 2025	Executive & Development
Implement Predictive Analytics	Q4 2025	Data & Strategy Team

5. Critical Success Factors & Risk Mitigation

Key Success Factors:

1. Strong executive support for revenue initiatives.
2. Cross-departmental collaboration for seamless execution.
3. Data-driven decision-making to refine strategies over time.
4. Effective allocation of resources (budget, personnel, technology).
5. Continuous performance tracking and adaptation.

Risk Mitigation Strategies:

✅ Phased implementation to minimize disruptions.
✅ Regular feedback loops with customers and stakeholders.
✅ Contingency planning for external disruptions (e.g., economic shifts, competitor moves).
✅ Performance monitoring with KPIs to ensure strategy effectiveness.

9. Limitations and Future Work:

1. Data Limitations

- **Temporal Scope**: The dataset covers only 92 days, restricting seasonal trend analysis and preventing year-over-year comparisons.
- **Granularity Issues**: Missing detailed customer demographics, room category price points, operational cost metrics, and revenue breakdown by services (e.g., F&B, amenities).
- **Data Quality Concerns**: Lack of clarity on data collection methodology, potential missing booking status data, and no handling details for partial stays or modified bookings.
- **External Influences**: The data does not account for market conditions, competitor actions, or external economic factors that impact the hospitality industry.

2. Methodology Constraints

- **Analysis Limitations**:
    - Limited segmentation due to missing customer profiles.

    - Inability to assess pricing effectiveness or profitability without rate and cost data.
    - Restricted evaluation of marketing ROI across channels.
- **Statistical Constraints**:
    - No seasonal adjustment or long-term forecasting.
    - Limited ability to analyze customer lifetime value or predictive trends.

3. Areas for Further Research

- **Customer Behavior Analysis**:
    - Detailed study of booking patterns by traveler segments (business, leisure).
    - Analysis of cancellation behavior and contributing factors.
    - Investigation of rating patterns and customer satisfaction drivers.
    - Impact of loyalty programs on booking frequency.
- **Revenue Optimization**:
    - Effectiveness of dynamic pricing strategies.
    - Optimal distribution channel mix for maximum revenue.
    - Room category utilization and revenue contribution.
    - Expansion of ancillary revenue opportunities (F&B, spa, events).
- **Operational Efficiency**:
    - Staff-to-room ratio effectiveness.
    - Check-in/check-out process efficiency.
    - Impact of maintenance schedules on room availability.
    - Sustainability and energy efficiency metrics.

10. Conclusion

In conclusion, the insights derived from this analysis serve as a foundation for informed decision-making in the hospitality industry. By leveraging data-driven strategies, hotels can enhance their operational efficiency, improve customer satisfaction, and ultimately drive revenue growth in an increasingly competitive market. The ongoing commitment to data analysis and continuous improvement will be crucial for adapting to changing market conditions and meeting the evolving needs of customers.