

**Optimizing Hospitality: Data-Driven Insights into India's Hotel Economy**



HOTALS IN INDIA

Presented by: -

R. ANIL

**ACKNOWLEDGMENT**

I want to extend my sincere gratitude to the entire team at “Analytics Space” for granting me the opportunity to work on the “Hospitality Analysis on Indian Hotels” project. I would also like to express my heartfelt thanks to my academic team, 'Data Trained,' for providing me with this enriching experience. Working on this project not only allowed me to enhance my analytical skills but also exposed me to a plethora of new insights. The invaluable suggestions and guidance from both teams played a pivotal role in the successful completion of this project, and I am truly grateful for the learning and growth it has brought to me.

Chapter 1:

INTRODUCTION

* 1. **Project Objective**

The objective of the “Hospitality Analysis on Indian hotels” project is to conduct an in-depth analysis of the pricing variations among different hotels in the India. To analyze hotel data across India to uncover customer preferences, optimize pricing strategies, and provide actionable market insights that will help improve hotel operations, marketing, and overall competitiveness in the Indian hospitality industry."

The data is taken from “Booking.com” website by using some python web scrapping to extract the data in the form of datasets

* **Overview of Project**

Discover the main objectives and scope of Project for web scraping and Data cleaning, Data Modelling, Data preprocessing In Python and data visualization in Power Bi.

**1.2 Problem Statement**

Despite the availability of extensive data, many hotels in India struggle to harness its full potential. The lack of comprehensive, data-driven insights often results in missed opportunities for optimizing pricing strategies, tailoring marketing efforts, and enhancing service offerings. Without a clear understanding of customer preferences and market trends, hotel operators may find it difficult to stay competitive, leading to reduced customer satisfaction and lower occupancy rates

**Key challenges include:**

**Sentiment Analysis**: Extracting meaningful insights from unstructured and subjective customer reviews.

**Price Optimization**: Identifying trends and correlations between pricing and customer satisfaction across various segments.

**Dynamic Market Trends**: Keeping insights relevant amidst rapidly changing customer preferences and industry conditions.

#### **1.3 Project Objectives**

The primary objective of this project is to analyze a dataset comprising hotel names, reviews, pricing information, and other relevant variables from various cities across India. Through this analysis, the project aims to achieve the following specific objectives:

1. **Understanding Customer Preferences:**

By analyzing customer reviews, the project seeks to identify the key factors that influence customer satisfaction and dissatisfaction across different regions in India. Understanding these preferences will enable hotel operators to tailor their offerings to meet customer expectations better.

1. **Price Optimization:** The project will explore pricing structures among various hotels to identify trends and competitive pricing strategies. By examining the correlation between prices and customer satisfaction, the project aims to provide insights that can help hotels optimize their pricing models for different market segments.
2. **Market Insights:** The project will analyze regional trends and demand patterns across different types of hotels in India. These insights will support strategic decision-making, enabling hotels to target specific market segments more effectively and allocate resources where they are most needed

#### **1.4 Scope**

The scope of this project encompasses data collected from the Booking.com website, focusing on hotels located in tier-1 cities such as Bangalore, Chennai, Mumbai, Pune and Delhi. The analysis will cover various aspects of hotel operations, including customer sentiment, pricing strategies, and market segmentation. The insights derived from this analysis will be relevant to hotel operators, marketers, and other stakeholders in the hospitality industry who are looking to enhance their competitiveness and improve customer satisfaction.

By leveraging data analytics, this project aims to provide actionable recommendations that can help hotels in India navigate the challenges of a dynamic and competitive market. The findings will not only contribute to the optimization of individual hotel operations but also have the potential to impact the broader hospitality industry in India.

Chapter 2:

DATA COLLECTION AND DATA SOURCES

#### **Data Source**

The primary data source for this project is the Booking.com website, one of India's leading online travel platforms. Booking.com offers a comprehensive range of travel-related services, including hotel bookings, flight reservations, holiday packages, and more. The website features an extensive database of hotels across India, catering to various customer preferences and budget ranges. For this project, the focus was specifically on collecting data related to hotels in tier-1 cities such as Bangalore, Chennai, Mumbai, Delhi and Pune.

The dataset collected from the Bookin.com website includes the following key variables:

1. **Hotel Names:** The names of the hotels listed on the platform, which provide a basic identification of the properties being analyzed.
2. **Addresses:** The geographical locations of the hotels, including city, state, and specific locality. This information is crucial for analyzing regional trends and market segmentation.
3. **Prices:** The room rates offered by each hotel. These prices vary depending on factors such as room type, time of booking, and additional services included. The price data is essential for conducting a pricing analysis and understanding competitive pricing strategies across different regions.
4. **Reviews:** Customer reviews provide qualitative insights into the experiences of hotel guests. These reviews include both positive and negative feedback, offering a rich source of data for sentiment analysis and understanding customer satisfaction.
5. **Ratings:** Along with written reviews, customers also provide numerical ratings, typically on a scale of 1 to 5. These ratings offer a quantitative measure of customer satisfaction, allowing for the identification of trends and correlations between ratings and other variables such as price and location.

The combination of these data points allows for a comprehensive analysis of the factors influencing hotel performance and customer preferences in the Indian hospitality market.

#### **Web Scraping Process**

Given the need for large-scale data collection, the web scraping process was chosen as the most efficient method to extract information from the Booking.com website. Web scraping involves automatically extracting data from websites using specialized software or scripts. For this project, Python was the primary programming language used, due to its robust libraries and ease of use for data scraping tasks.

The key Python libraries utilized in the web scraping process were:

1. **BeautifulSoup:** This library was used to parse the HTML content of the Booking.com website. BeautifulSoup provides a simple way to navigate and search through the HTML structure, making it easier to locate and extract specific pieces of data, such as hotel names, prices, and reviews.
2. **Requests:** The Requests library was employed to send HTTP requests to the Booking.com website and retrieve the HTML content of the pages. By using Requests, we could access multiple pages of the website programmatically, allowing for the extraction of data from a large number of hotels.
3. **Pandas:** After extracting the data, Pandas was used to organize it into a structured format, typically a DataFrame. This allowed for easier manipulation and analysis of the data, including cleaning, filtering, and sorting.

The web scraping process involved several steps:

1. **Sending Requests:** The first step was to send HTTP requests to the Booking.com website to access the pages containing hotel listings. This was done for multiple cities, ensuring a diverse dataset covering different regions of India.
2. **Parsing HTML Content:** Once the HTML content was retrieved, BeautifulSoup was used to parse the data. This involved locating the specific HTML tags and attributes that contained the desired information, such as hotel names, prices, and reviews.
3. **Extracting Data:** The parsed data was then extracted and stored in a structured format using Pandas. This step involved cleaning the data to remove any irrelevant information or duplicates and ensuring that all variables were properly formatted.
4. **Saving Data:** Finally, the cleaned and structured data was saved to a CSV file, which could then be imported into Power BI for further analysis and visualization.

#### **Challenges**

While the web scraping process was effective in collecting a large volume of data, several challenges were encountered:

1. **Handling Large Volumes of Data:** Given the vast number of hotels listed on Booking.com, the volume of data was substantial. Managing and processing this data efficiently required careful consideration of system resources and processing time. To address this, data collection was performed in batches, and efficient data structures were used to minimize memory usage.
2. **Website Limitations:** Websites often implement measures to prevent automated scraping, such as CAPTCHAs, rate limiting, and dynamic content loading. These limitations posed significant challenges, particularly when trying to scrape data at scale. To mitigate these issues, strategies such as introducing delays between requests, randomizing user agents, and handling exceptions were employed.
3. **Data Quality:** Another challenge was ensuring the accuracy and consistency of the data collected. Some hotel listings had incomplete information or inconsistent formatting, which required additional data cleaning and preprocessing steps to ensure the dataset was usable for analysis.

Despite these challenges, the web scraping process successfully resulted in a rich dataset that provided valuable insights into the Indian hospitality industry. The next step involved analyzing this data to uncover trends, optimize pricing strategies, and better understand customer preferences.

Chapter 3:

DATA PREPROCESSING

Effective data preprocessing is a critical step in ensuring the quality and reliability of the analysis. This stage involves cleaning the raw data, transforming it into a suitable format for analysis, and importing it into Power BI for visualization. Below is a detailed explanation of the steps involved in the data preprocessing phase of this project.

#### **Data Cleaning**

The raw data collected through web scraping contained several inconsistencies and errors that needed to be addressed before proceeding with the analysis. Data cleaning is essential to remove noise and inaccuracies, ensuring that the dataset is accurate, consistent, and ready for analysis. The following steps were taken to clean the data:

1. **Removal of Duplicates:**
   * **Identification of Duplicates:** The first step in the data cleaning process was to identify and remove duplicate entries. Duplicate data can occur when the same hotel is listed multiple times, possibly due to variations in room types or slight differences in hotel name formatting.
   * **Removal Process:** Using the Pandas library in Python, duplicate entries were detected based on unique identifiers such as hotel names and addresses. These duplicates were then removed, leaving only one entry per hotel. This step ensured that each hotel was represented only once in the dataset, preventing any skewing of the analysis.
2. **Handling Missing Values:**
   * **Identifying Missing Data:** Missing values can significantly impact the results of the analysis if not handled properly. In this project, missing data was primarily found in columns related to customer reviews and ratings, where some entries lacked complete information.
3. **Correction of Errors:**
   * **Data Accuracy:** During the data collection process, some errors were identified, such as incorrect price formatting or misaligned columns. For example, some prices were mistakenly captured in different currencies or formats.
   * **Standardization:** These errors were corrected by standardizing the data. Prices were converted to a consistent currency format (Indian Rupees), and columns were realigned to ensure that all data points were correctly labeled and categorized.
4. **Outlier Detection:**
   * **Identifying Outliers:** Outliers, such as extremely high or low prices, were identified as they could distort the analysis. These outliers were detected using statistical methods like the interquartile range (IQR) and Z-score analysis.
   * **Handling Outliers:** Depending on the nature of the outlier, some were retained if they provided meaningful insights, while others were adjusted or removed if they were determined to be errors or anomalies.

Chapter 4:

DATA ANALYSIS AND VISUALIZATION

This section summarizes the analyses conducted to uncover customer preferences, pricing trends, and market segmentation in India's hospitality industry.

**Key Metrics Analyzed**

1. **Customer Satisfaction**:
   * The average hotel rating across cities ranged between 7.36 (Chennai) and 7.95 (New Delhi), with New Delhi scoring the highest.
   * Customer feedback highlighted service quality, cleanliness, and location as critical factors driving satisfaction.
2. **Pricing Trends**:
   * **Average Prices by City**: Mumbai had the highest average hotel price (₹7,000), followed by Pune and New Delhi.
   * **Top Expensive Hotels**: "SaffronStays Happy Fields, Pune" stood out with the highest average price of ₹28,010, coupled with a high customer rating (9.8).
3. **Market Segmentation**:
   * **Luxury vs. Budget Hotels**: Luxury hotels typically had higher prices and better ratings. For example, "Trident Nariman Point, Mumbai" had an average price of ₹18,500 with a high customer rating (8.7).
   * **Amenities Impact**: Specific facilities, such as executive suites and premium rooms, commanded higher prices.
4. **Geographic Distribution**:
   * The distribution of hotels highlighted high concentrations in cities like Mumbai and New Delhi, which also aligned with higher pricing and demand trends.
   * Coastal regions and popular tourist destinations showed higher customer ratings and demand.

**Visualizations and Interpretations**

1. **Price Analysis**:
   * **Price Distribution**:
     + The pricing spectrum revealed that mid-range hotels dominated the market.
     + Outliers like high-priced luxury hotels in Pune and Mumbai highlighted the demand for premium accommodations.
   * **Regional Price Trends**:
     + Higher prices in Mumbai and New Delhi reflected increased demand, likely due to their status as business and tourism hubs.
2. **Customer Sentiment Analysis**:
   * **Overall Sentiment**:
     + The majority of reviews were positive, with "FabHotel Grand Prime, Mumbai" achieving a perfect score of 10.0.
   * **Common Dissatisfaction Drivers**:
     + Locations with lower ratings often cited cleanliness and outdated amenities as areas for improvement.
3. **Geographic Analysis**:
   * **Demand Hotspots**:
     + Heatmaps indicated high demand in Mumbai, New Delhi, and Pune. Bangalore, though not as high-priced, showed strong mid-range demand.
   * **Expansion Opportunities**:
     + Areas with fewer accommodations could benefit from mid-range and budget hotel developments.
4. **Market Segmentation**:
   * **Customer Demographics**:
     + Tourists favored locations and amenities such as recreational facilities.
     + Business travelers prioritized connectivity and business centers.
   * **Hotel Preferences**:
     + Segments such as serviced apartments and boutique hotels catered to niche audiences, showing higher customer satisfaction.

**Interpretation of Findings**

1. **Customer Behavior**:
   * Cleanliness and quality service are non-negotiable for high satisfaction.
   * Facilities like premium rooms and recreational amenities significantly impact pricing and ratings.
2. **Pricing Strategies**:
   * Metropolitan areas justify higher pricing due to higher demand and the availability of premium services.
   * Mid-range pricing remains popular, offering a balance between affordability and quality.
3. **Market Insights**:
   * Seasonal pricing strategies should capitalize on peak demand during holidays and festivals.
   * Strategic investments in underserved regions can capture untapped demand.
4. **Actionable Recommendations**:
   * Focus on enhancing cleanliness and service quality in lower-rated hotels.
   * Expand marketing for mid-range hotels in high-demand areas.
   * Introduce dynamic pricing to maximize revenue during peak tourist seasons.

Chapter 5:

Tools and Technologies

Software Requirement:-

1)Python(Jupyter notebook)

2)Power bi

3)Excel

Main Software Libraries:-

1)Pandas

2)Numpy

3)Beautiful Soups

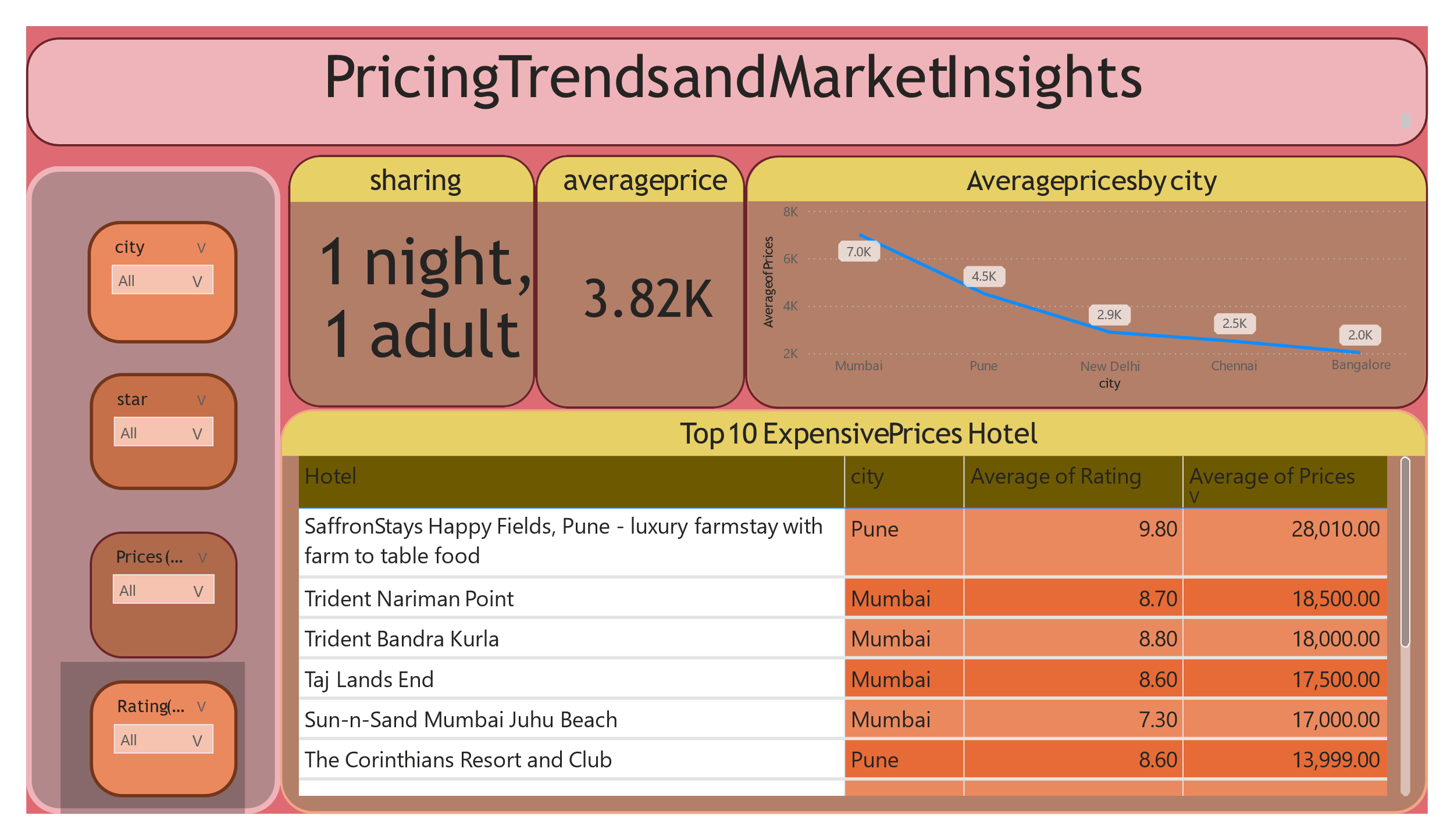
4)Requests

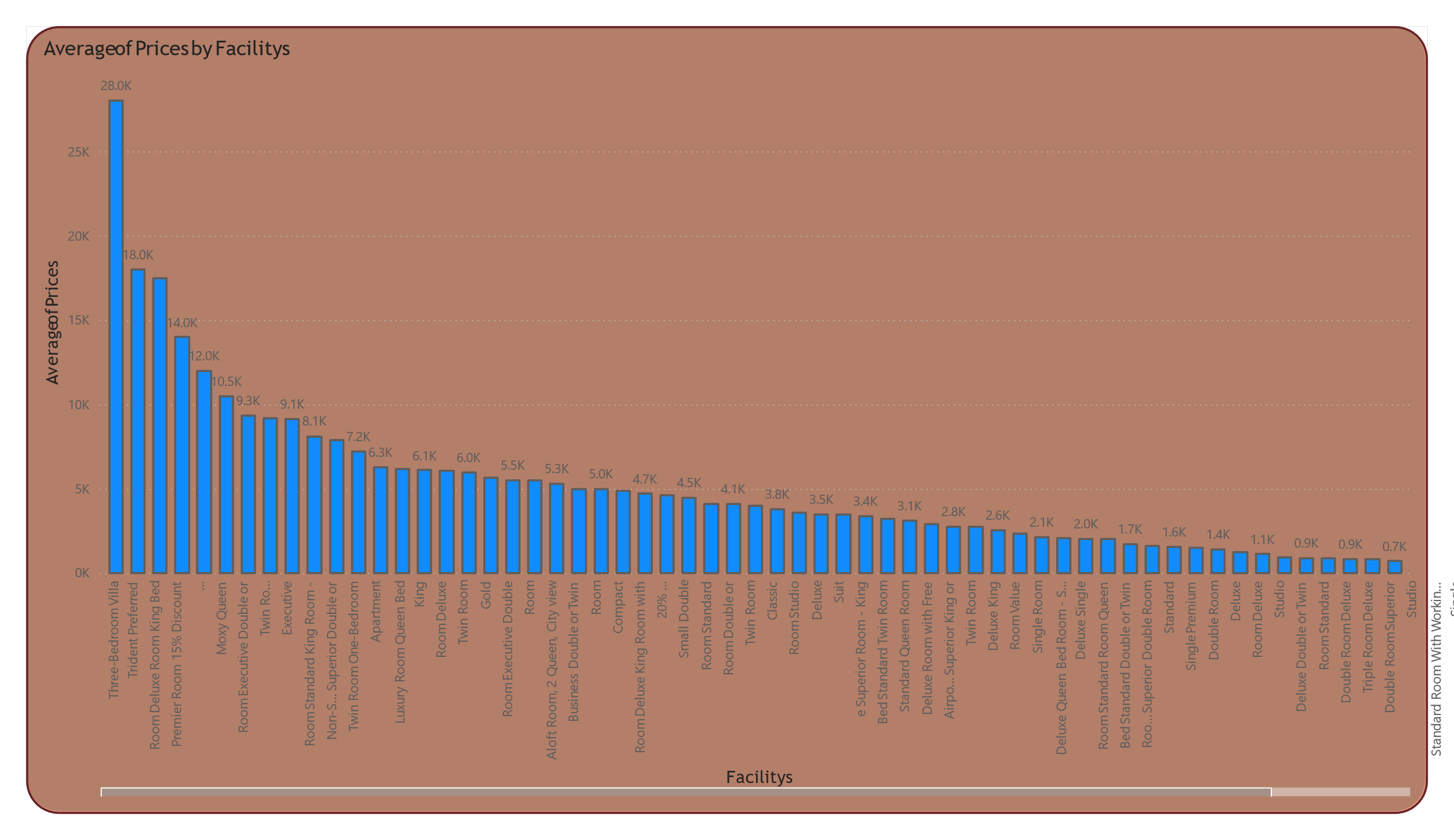
5)CSV

Chapter 6:

Dashboard







Chapter 7:

CONCLUSION

The primary objective of this project was to utilize data analytics to optimize operations, marketing strategies, and overall efficiency within the Indian hospitality industry. By analyzing comprehensive hotel data—including hotel names, addresses, prices, reviews, and ratings—the study uncovered key insights into customer preferences, pricing trends, geographic patterns, and market segmentation.

**Key Findings**

**1. Customer Preferences**

* **Service Quality and Amenities**: Customer satisfaction was highly influenced by the quality of service and amenities such as complimentary breakfasts and free cancellation policies. These features consistently emerged as drivers of positive reviews and ratings.
* **High-Rated Hotels**: Hotels with ratings above 8 achieved significantly higher customer satisfaction, emphasizing the importance of maintaining high service standards.

**2. Pricing Trends**

* **Wide Pricing Range**: Hotel prices across India ranged from as low as ₹150 for budget stays to over ₹46,000 for luxury accommodations, highlighting the diverse market landscape.
* **Regional Price Variations**:
  + Mumbai exhibited the highest average hotel prices (₹7,000), reflecting strong demand in a business and tourist hub.
  + Bangalore's luxury market also stood out, with premium hotels averaging ₹10,191 for 5-star properties.
* **Competitive Mid-Range Market**: 3-star and 4-star hotels demonstrated price sensitivity, with competitive pricing strategies aimed at capturing cost-conscious travelers.

**3. Market Segmentation**

* **Distinct Market Segments**:
  + **Budget Hotels (3-star)**: Cater primarily to cost-conscious customers. These properties had lower average ratings (6.3) but achieved high occupancy rates due to affordability.
  + **Luxury Hotels (5-star)**: Target affluent travelers, boasting higher average ratings (8.3) and premium pricing justified by superior amenities.
* **Regional Dynamics**:
  + Bangalore and Mumbai emerged as dominant markets for luxury accommodations.
  + Pune and Chennai displayed strong demand for mid-tier hotels, driven by business and tourism activities.

**Implications for Stakeholders**

**1. Hoteliers**

* Focus on improving customer satisfaction by investing in facilities that customers value most, such as complimentary breakfasts and free cancellation.
* Implement dynamic pricing strategies tailored to regional demand and peak seasons, particularly in metropolitan and tourist-heavy areas.

**2. Policy Makers**

* Encourage standardization of service quality across hotel categories to ensure a competitive and equitable market.
* Support initiatives for infrastructure development in regions with potential market gaps to foster growth in the hospitality sector.

**3. Marketing Teams**

* Leverage insights on regional and segment-specific preferences to craft targeted marketing campaigns.
* Highlight features like location, amenities, and value-for-money offerings to attract the desired customer demographics effectively.

**Summary**

This analysis provides actionable insights that empower stakeholders in the hospitality industry to make informed decisions and refine their strategies. By addressing customer preferences, optimizing pricing structures, and understanding market dynamics, the Indian hotel economy can achieve enhanced competitiveness and increased customer satisfaction. Through a data-driven approach, the hospitality industry in India can adapt to the dynamic market landscape, ensuring sustainable growth and success.