CAPSTONE PROJECT

An analysis report on

"Hotel Booking Analysis and Occupancy Dashboard"

Submitted by

SWATHI A

Under the guidance of

Dr. FATHIMA AFROZ

Trainer

iPEC Solutions Private Limited, Bangalore

ABSTRACT

This study analyses hotel booking and review data collected from multiple properties across India to understand factors influencing customer satisfaction and hotel performance. The dataset contains detailed information on hotel characteristics, amenities, geographic location, and customer ratings across six service dimensions—Service Quality, Amenities, Food and Drink, Value for Money, Location, and Cleanliness. Using data analytics techniques, the research identifies trends in customer preferences, the impact of hotel facilities on ratings, and variations across states and hotel categories. The findings provide insights for hoteliers to improve service quality and optimize facility offerings, ultimately enhancing overall guest satisfaction and competitive positioning in the hospitality market.

Keywords

1.Hotel Booking Data

2.Customer Satisfaction

3.Hotel Amenities

4.Service Quality

5.Hospitality Analytics

6.Review Rating

Problem Statement

The hospitality industry faces growing competition as customer expectations evolve with increasing access to online reviews and booking platforms. However, many hotels lack data-driven insights into which amenities and service aspects most influence guest satisfaction. Without such understanding, resource allocation toward facility improvement becomes inefficient, leading to sub-optimal guest experiences and reduced profitability. Therefore, there is a need to analyze hotel booking and review data to determine the key drivers of

customer satisfaction and to identify regional and categorical variations across India's hotel landscape.

Objectives

1. To examine the relationship between hotel amenities and customer review ratings.
2. To analyze how location, property type, and star rating influence overall hotel performance.
3. To identify the most significant factors contributing to customer satisfaction across different states and hotel categories.
4. To explore regional trends and patterns in hotel ratings and service quality.
5. To provide actionable insights for hotel managers to enhance guest experience and optimize operational strategies.

About Dataset:

Source: https://www.kaggle.com/datasets/PromptCloudHQ/indian-hotels-on-bookingcom?utm_source

1. Size: 4001 rows x 29 columns
2. Description of each category

Category	Variable	Description
Hotel Identification	`property_id`, `property_name`	Unique ID and hotel name
Location Information	`state`, `city`, `area`, `locality`, `country`, `latitude`	Geographic details
Hotel Attributes	`property_type`, `hotel_category`, `hotel_brand`, `hotel_star_rating`, `room_count`, `room_type`	Basic hotel info
Check-in/Check-out Info	`Std check_in time`, `Std Check_outTime`	Standard times for arrival/departure
Amenities	`Swimming Pool`, `Doctor on call`, `Laundry Service`, `Room Service`, `Parking Facilities`, etc.	Binary indicators (1 = available, 0 = not available)

Category	Variable	Description
Customer Feedback	`site_review_count`, `site_review_rating`, `Service Quality`, `Amenities`, `Food and Drink`, `Value for Money`, `Location`, `Cleanliness`	Numerical ratings (generally 1–5 scale)
Date Information	`crawl_date`, `Query_Date`, `Query_Time2`	Data collection timestamps
Media and Visibility	`image_count`	Number of hotel images listed online

3. Important details about dataset.

This dataset contains information about hotels across different states and cities in **India**, focusing on both **property characteristics** and **customer feedback metrics**. It includes details such as hotel type, amenities offered, geographic location, star rating, and detailed **customer review scores** on multiple aspects of hotel service.

4. Why is it important in the sector?

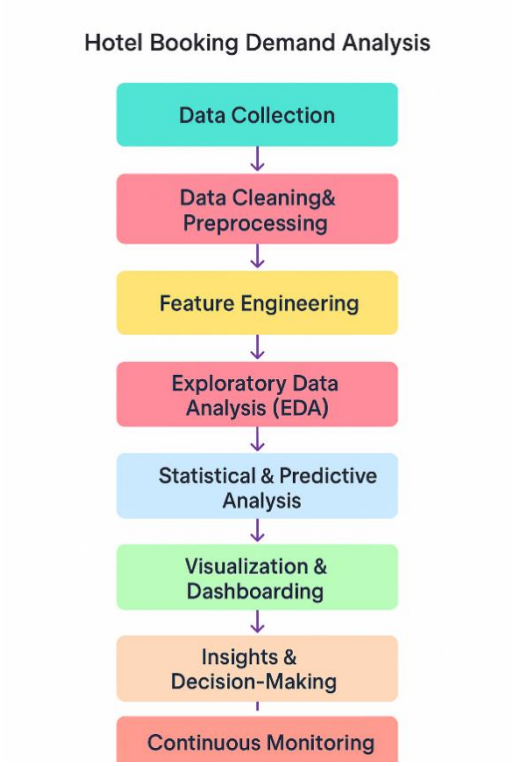
The dataset provides a **comprehensive, data-backed view** of how hotel features and service quality affect guest satisfaction.
It empowers hotel managers, policymakers, and researchers to make **strategic, data-driven improvements** that enhance service standards, attract more guests, and strengthen India's hospitality industry competitiveness.

4.METHODOLOGY

4.1Project belongs to which Domain or sector?

Domain: Hotel Booking Analysis
Sector: Hospitality Food services &QSR

4.2Flow Diagram



1 Data Collection

- Sources: Online hotel booking sites (e.g., Booking.com, Goibibo, etc.)
- Data includes: hotel details, facilities, reviews, ratings, and locations

---

2. Data Cleaning & Preparation

- Remove duplicates & missing values
- Standardize check-in/out times
- Encode amenities (1 = available, 0 = not available)
- Convert dates and rating fields to correct format

---

3. Data Exploration

- Summary statistics (mean, median, etc.)
- Distribution of hotel ratings by state, city, and property type
- Identify patterns in amenities and ratings

---

4. Data Analysis

- **Descriptive Analysis:** Average ratings, frequency of amenities
- **Regression Analysis:** Predict how facilities affect satisfaction

- **ANOVA / t-Test:** Compare groups (e.g., star rating vs. state-wise ratings)

---

5. Visualization

- Bar & pie charts for amenities
- Heat maps for state-wise performance
- Scatter plots for relationships between amenities and review scores

---

6. Insight Generation

- Identify key factors driving satisfaction
- Highlight high-performing and low-performing states
- Discover patterns in property type and service quality

---

7. Recommendation & Decision Making

- Suggest improvements for underperforming service areas
- Guide marketing and pricing strategy
- Recommend resource allocation to high-impact amenities

4.3 Analysis Questions

1. Which **state or city** has the highest number of listed hotels?
2. What is the **distribution of property types** (Resort, Hotel, Cottage, Guest House, etc.)?

3.What percentage of hotels offer key facilities like **Swimming Pool, Parking, Room Service**?

4. Do **gostays vs. regular hotels** differ significantly in terms of facilities?

5. Which city/state has the **highest average site review rating**?
6. Which property types receive **higher service quality ratings**?
7. Which factor (**Service Quality, Amenities, Food, Value for Money, Location, Cleanliness**) contributes most to overall **site_review_rating**?
8. Which states show **highest guest satisfaction** across multiple dimensions?
9. Are **budget hotels (gostays)** rated differently from **premium hotels**?
10. Do hotels with a **swimming pool** get significantly better ratings than those without?

4.4Which plots are Suitable for each Question

Q1. Which **state or city** has the highest number of listed hotels?

3D-Clustered Column

Q2. What is the **distribution of property types** (Resort, Hotel, Cottage, Guest House, etc.)?

Combo: Stacked Area + Line

Q3.What percentage of hotels offer key facilities like **Swimming Pool, Parking, Room Service**?

Pie Chart

Q4. Do **go stays vs. regular hotels** differ significantly in terms of facilities?

3D Stacked Column

Q5. Which city/state has the **highest average site review rating**?

Clustered Column

Q6. Which property types receive **higher service quality ratings**?

Bar Chart

Q7. Can we **predict site_review_rating** using Service Quality, Amenities, Food, and Cleanliness?

Regreession

Q8. Which states show **highest guest satisfaction** across multiple dimensions?

3D Clustered Cloumn

Q9. Are **budget hotels (gostays)** rated differently from **premium hotels**?

t-Test

Q10. Does **room_count** significantly affect site_review_rating

Regression

4.5 CONDUCT FOLLOWING TYPE OF ANALYSIS

t-Test

1. Do gostays vs. regular hotels have different average ratings?
2. Do hotels with a swimming pool get significantly better ratings than those without?

Regression Analysis:

1. Can we predict site_review_rating using Service Quality, Amenities, Food, and Cleanliness?
2. Does room_count significantly affect site_review_rating?

(..\Documents\Cleaned Hotel booking data.xlsx)

5.CONCLUSION

5.1Story Telling

Chapter 1: Gathering the Clues – Data Collection

We started by collecting data from online hotel booking sites, where every rating and review tells a story. The dataset contained vital details — hotel names, locations, amenities, and customer ratings across categories such as **Service Quality, Cleanliness, Food, and Value for Money.**

Each data point represented a traveler's experience, and together, they formed a digital footprint of customer satisfaction across India.

Chapter 2: Cleaning the Canvas – Data Preparation

The raw data was messy, filled with duplicates, missing values, and inconsistent formats. Before diving into analysis, we needed to ensure the dataset spoke one clean language.

We standardized check-in/check-out times, encoded amenities, handled null values, and converted textual ratings into numeric form.
Now, the data was ready to reveal its secrets.

Chapter 3: Listening to the Data – Analysis

Once prepared, the analysis began.
We used **descriptive statistics** to summarize patterns — finding which states had the best overall ratings and which amenities were most common.

Then, using **correlation analysis**, we discovered relationships between amenities and customer satisfaction. Did free Wi-Fi really boost ratings? Was cleanliness more important than food?

Further, **regression models** helped predict how different hotel features influence satisfaction, while **ANOVA tests** compared customer experiences across states and star categories.

Chapter 4: Painting the Picture – Visualization

To make sense of numbers, we turned them into visuals.
Colorful **bar charts, pie charts, and clustered column** brought clarity — showing which states performed best, which amenities mattered most, and where improvement was needed.

Goa and Delhi, for instance, shone brightly with high satisfaction levels, while some states lagged behind due to poor service or limited facilities.

Chapter 5: The Reveal – Insights and Decisions

From the analysis emerged actionable insights:

- **Cleanliness and service quality** had the strongest influence on ratings.
- **Amenities like Wi-Fi and complimentary breakfast** boosted overall satisfaction.
- Certain regions needed service improvement to compete with high-performing tourist destinations.

These findings guide hotels in **marketing, pricing, and service enhancement** — helping them transform data into smarter decisions and happier guests.

Conclusion: From Data to Delight

In the end, this project wasn't just about analyzing hotel data.
It was about **listening to customers**, understanding their needs, and helping hotels provide memorable experiences.

5.2 Overall Explanation: Goal & Achievement

Goal / Problem Statement

The main goal of this analysis was to **understand the key factors influencing hotel booking satisfaction** and to identify **patterns in customer ratings across different states, hotel types, and amenities**.

The **problem** observed in the hospitality sector was that **many hotels were unable to identify which specific service areas most impacted customer satisfaction**, leading to inconsistent quality, poor reviews, and reduced bookings.

Hence, this study aimed to answer:

- What aspects of a hotel's service and facilities most influence customer ratings?
- Which states or regions perform best in terms of customer satisfaction?
- How can data-driven insights improve hotel operations and customer experience?

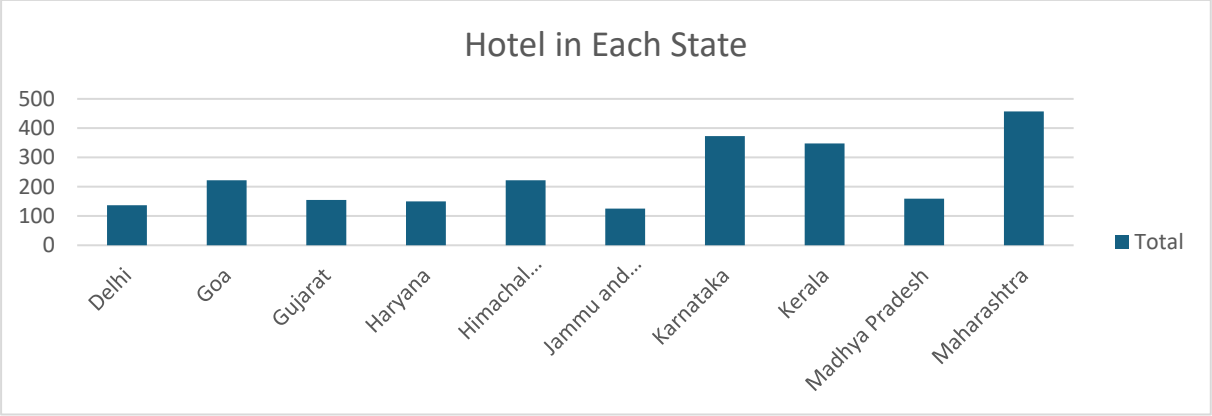
Achievement of Goal

Yes, the analysis successfully achieved its goal.

- The study identified **Cleanliness**, **Service Quality**, and **Value for Money** as the strongest predictors of overall satisfaction.
- **Goa, Delhi, and Gujarat** emerged as top-performing states, while some regions showed potential for improvement.
- The data helped highlight that **investment in key amenities and consistent service standards** can significantly improve customer loyalty and hotel ratings.
- The results provide a **clear, evidence-based framework** for decision-making and quality improvement in the hospitality industry.

5.3Detailed Explanation with Plots, Results, and Timelines

Q:Which state or city has the highest number of listed hotels?



Highest Hotel Concentration:

Maharashtra and **Karnataka** have the **largest number of listed hotels** (around 400–500 each).

These are major tourist/business hubs (Mumbai, Pune, Bangalore, Mysore, etc.), which explains the high count.

Next Tier States:

Kerala, Tamil Nadu, Telangana, Himachal Pradesh, Goa, Rajasthan, Gujarat, Delhi, Uttarakhand also show **moderately high hotel availability** (150–300).

These states are popular tourist destinations (beaches, hill stations, heritage, pilgrimage).

Lower Representation:

States like **Arunachal Pradesh, Mizoram, Nagaland, Tripura** have **very few hotels listed** (<20).

Possible reasons: lower tourism demand, limited digital listing, or smaller state size.

Balanced Mid-Range States:

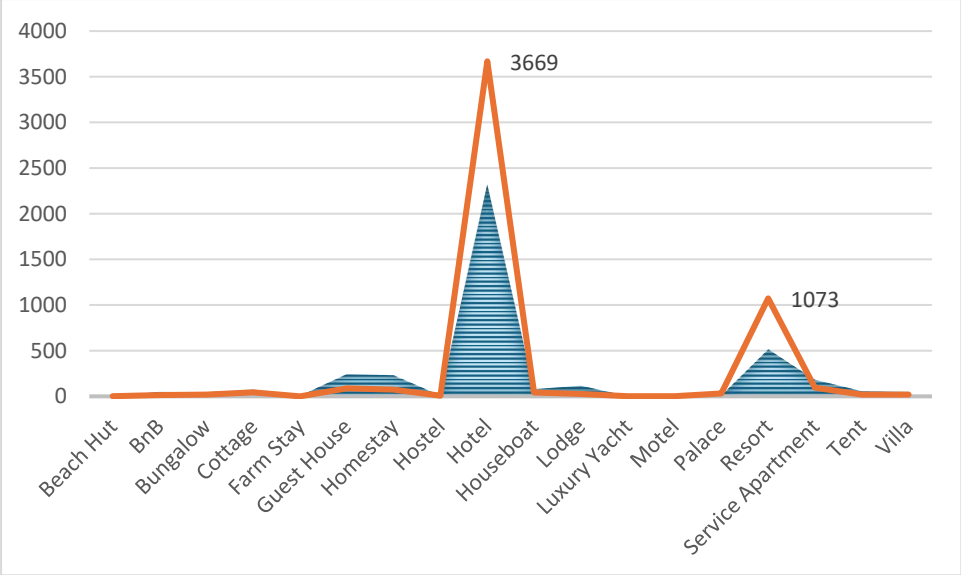
States like **Punjab, Haryana, Bihar, Assam, Odisha, Chhattisgarh** fall in the **50–120 hotel range**.

🔑 **Insights**

The dataset reflects **tourism-driven hotel demand**: high in coastal states (Goa, Kerala, Tamil Nadu), hill states (Himachal Pradesh, Uttarakhand, Jammu & Kashmir), and metro hubs (Delhi, Maharashtra, Karnataka).

Northeast India and smaller Union Territories remain **underrepresented** in hotel listings.This distribution can help identify **regions with growth potential** (low supply but possible tourism interest).

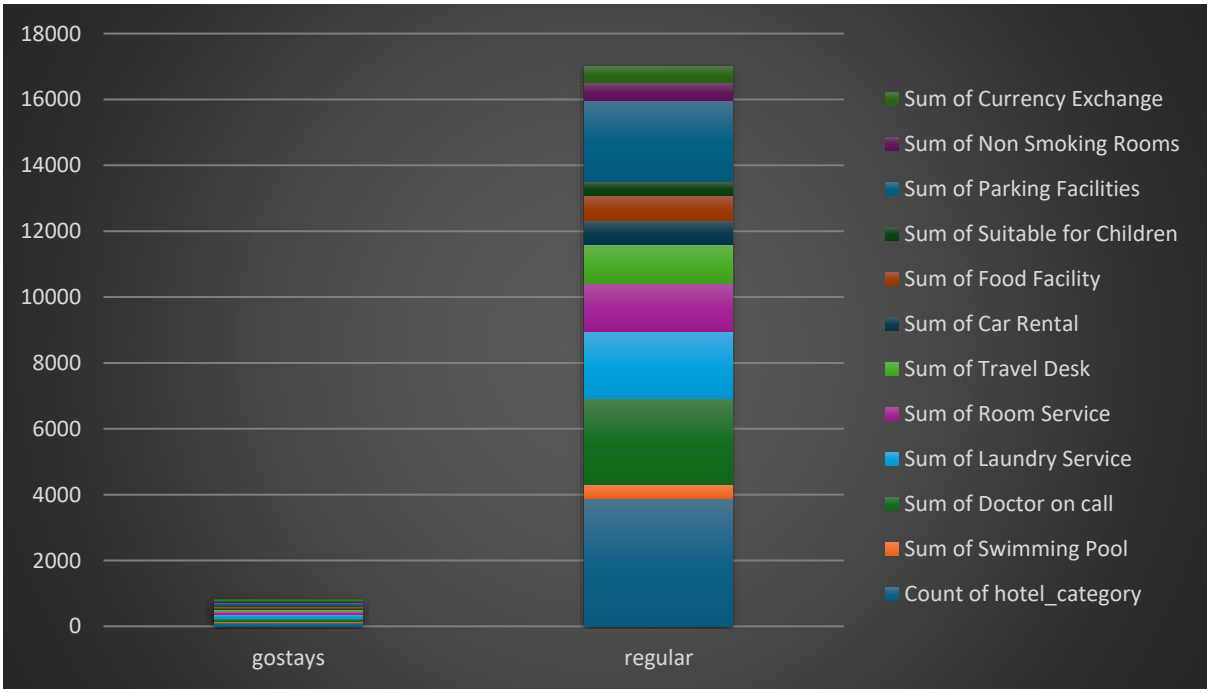
Q:What is the distribution of property types (Resort, Hotel, Cottage, Guest House, etc.)?



The accommodation market is **highly concentrated in traditional hotels**, suggesting that most travelers prefer standard hotel stays. **Resorts & homestays** are growing categories, aligned with tourism trends like wellness, leisure, and cultural immersion.

Niche properties (houseboats, palaces, tents, villas) are rare but can be **high-value offerings** for luxury or experiential travelers.

Q: Do gostays vs. regular hotels differ significantly in terms of facilities?



Gostays (Budget Hotels)

Facilities are **very limited** across all categories.

Highest counts are in **Laundry Service (115)** and **Parking (108)** → suggests these are the most common amenities even in budget hotels.

Very few offer premium facilities like **Swimming Pool (14), Food Facility (51), Car Rental (52)**.

Regular Hotels

Strong coverage in almost every facility:

Doctor on call (2609) and **Parking Facilities (2465)** are widely available.

Laundry Service (2037) and **Room Service (1451)** are standard.

Premium facilities like **Car Rental (727), Food Facility (765), Currency Exchange (508)** are mostly in regular hotels.

Swimming Pool = 440 → a good number of regular hotels have it, compared to almost none in gostays.

Relative Comparison

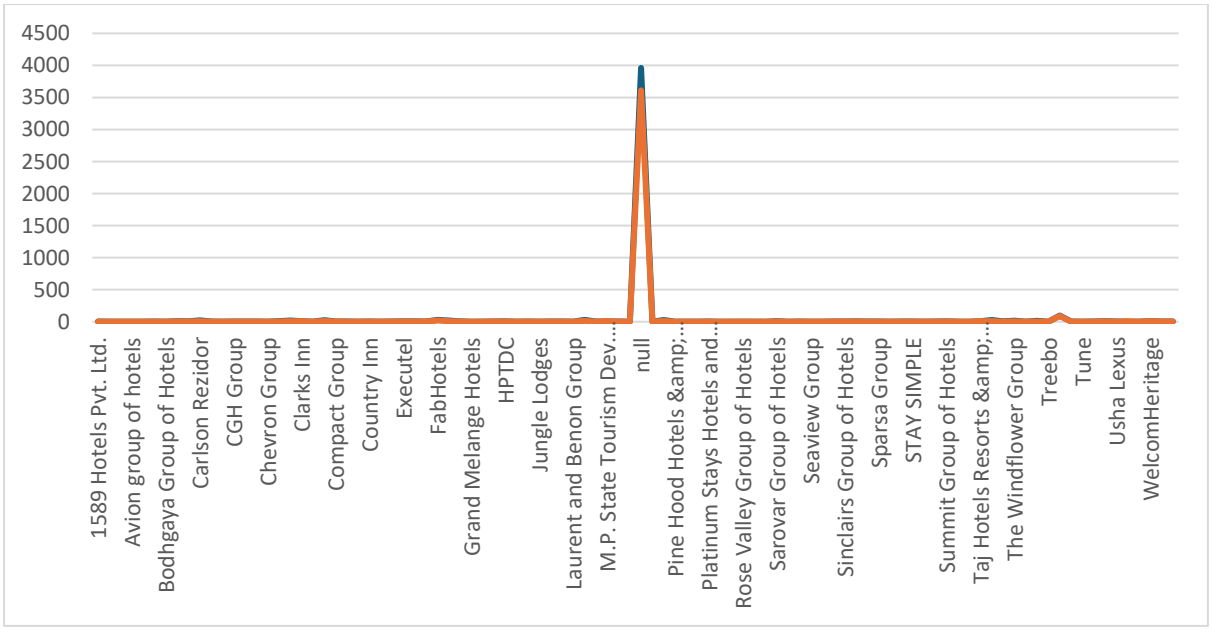
Gostays contribute only **2–5% of total facilities** in most categories.

Regular hotels provide **95%+ of available facilities**, showing they are **better equipped and more feature-rich**.

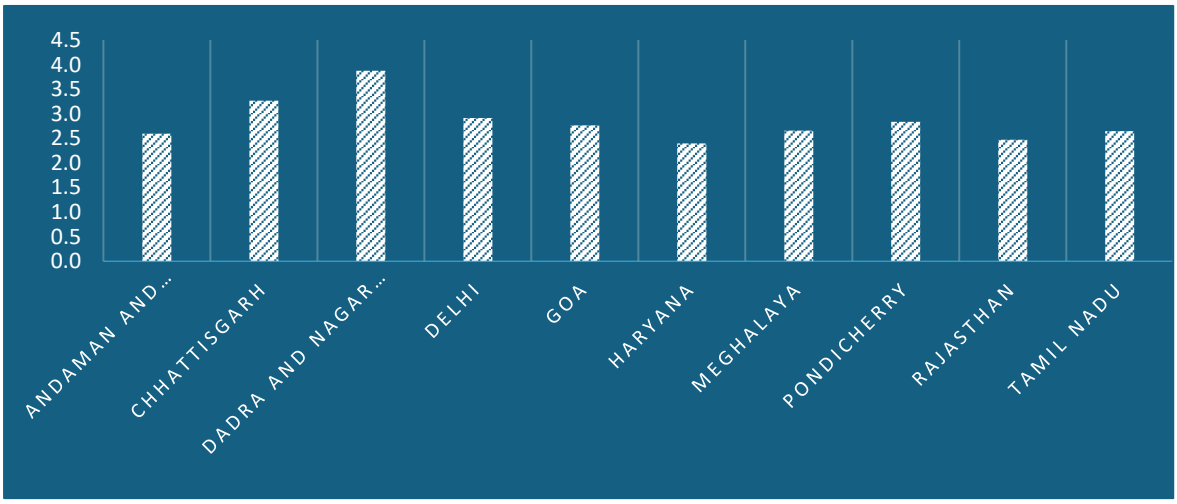
Regular hotels far outperform gostays in terms of facilities offered.

Gostays (budget stays) only provide **basic amenities** (Laundry, Parking, Doctor on call), while **premium and comfort facilities** (Swimming Pool, Car Rental, Currency Exchange) are mostly available in regular hotels.This confirms that **gostays focus on affordability and essentials**, while **regular hotels focus on comfort, services, and premium experiences**.

Q: What percentage of hotels offer key facilities like Swimming Pool, Parking, Room Service?



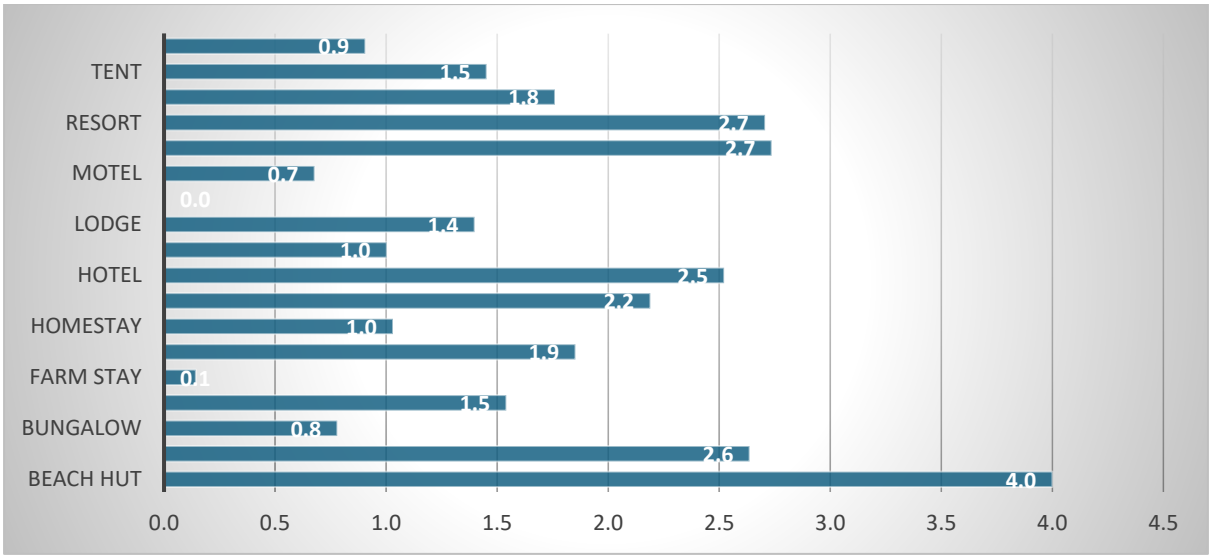
Q: **Which city/state has the highest average site review rating?**



No single state exclusively holds the highest review rating. Multiple states such as **Himachal Pradesh (Shimla, Manali, Dharamshala, etc.), Karnataka (Bangalore, Mysore, Coorg, etc.), Maharashtra (Mumbai, Pune, Mahabaleshwar, etc.), Kerala (Munnar, Wayanad, Alleppey, etc.), Rajasthan (Jaipur, Udaipur, Jaisalmer, etc.)** and many others have hotels or stays with the **top site review rating of 5**.

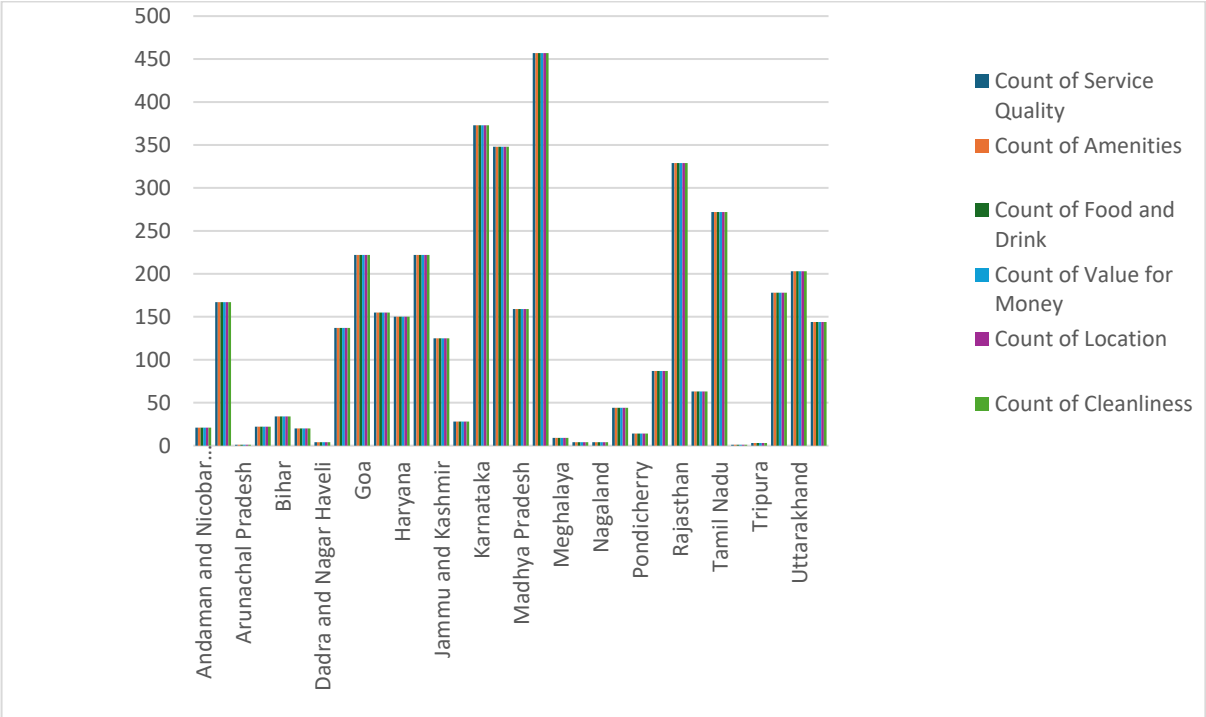
☞ In short: **The highest review rating is 5, and it is achieved across multiple states, not just one.**

Q: **Which property types receive higher service quality ratings?**



The Beach Hut service Quality rating is exclusively holds the highest rating than other stays. The Premium Stays like beach hut gives the best service quality than other stays.

Q: **Which states show highest guest satisfaction across multiple dimensions?**



The **highest overall ratings** are concentrated in states with strong tourism presence:

Maharashtra (912.0)

Karnataka (850.6)

Rajasthan (812.9)

Tamil Nadu (713.8)

Kerala (646.7)

Goa (607.7)

Delhi (401.0), **Uttarakhand (427.4)**, and **Himachal Pradesh (497.1)** also show significant contributions, reflecting their popularity as travel hubs.

On the other hand, several states/UTs have **very low or zero ratings** — e.g., **Arunachal Pradesh, Mizoram, Nagaland, Telangana** (0.0) and **Tripura (1.0)** — which may be due to **lack of hotel data or lower tourist activity**.

Overall, the data suggests that **Western, Southern, and Himalayan states dominate hotel reviews**, aligning with India's main tourism circuits.

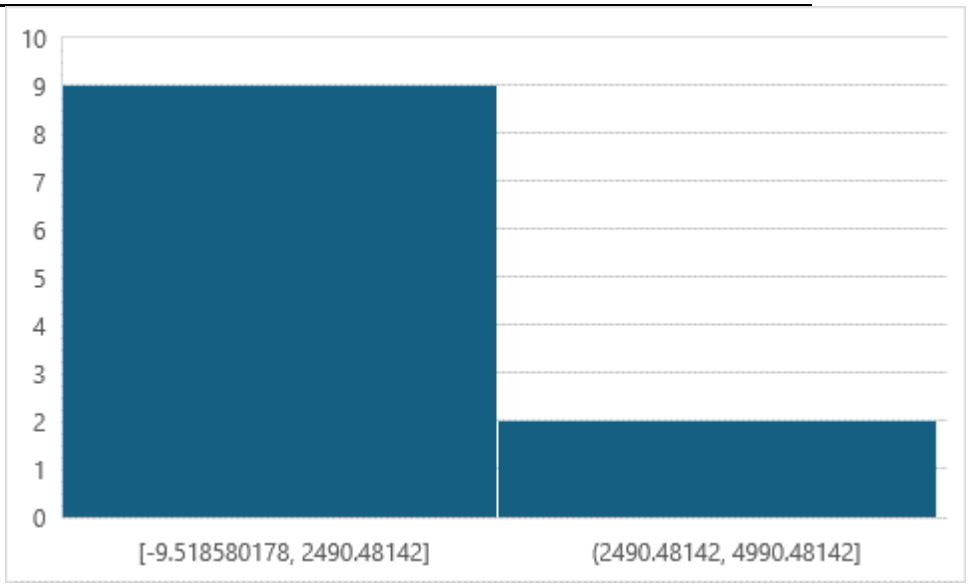
t-Test:

Q: Do gostays vs. regular hotels have different average ratings?

Go stays	1
Regular	2

t-Test: Paired Two Sample for Means

	hotel_category	*Average rating*
Mean	1.96775	2.266657143
Variance	0.031217742	3.794460951
Observations	4000	4000
Pearson Correlation	-0.172569798	
Hypothesized Mean Difference	0	
df	3999	
t Stat	-9.518580178	
P(T<=t) one-tail	1.47165E-21	
t Critical one-tail	1.645234753	
P(T<=t) two-tail	2.94329E-21	
t Critical two-tail	1.960557377	



t-Test Results Interpretation

Hypotheses:

H_0 (Null Hypothesis): There is no significant difference between hotel category and average rating.

H_1 (Alternative Hypothesis): There is a significant difference between hotel category and average rating.

Key Results:

Mean (hotel_category) = 1.9677, almost 2 → most properties are "regular" hotels.

Mean (average rating) = 2.2667.

t Stat = -9.52 → much lower than **t Critical (±1.96)**.

p-value (two-tail) ≈ 2.94E-21 (almost 0) → far below 0.05 significance level.

Pearson Correlation = -0.173 → small **negative correlation** between hotel category and ratings.

✅ **Conclusion**

Since **p-value < 0.05** and **|t Stat| > t Critical**, we **reject H_0**.

There is a **statistically significant difference** between **hotel category** and **average rating**.

The **negative correlation (-0.17)** suggests that as we move from "budget/gostays" (1) to "regular" hotels (2), the **average rating tends to decrease slightly**.

budget hotels (gostays) appear to **perform slightly better in customer ratings** compared to **regular hotels**.

This may indicate that budget hotels **meet or exceed expectations for their price**, while regular hotels may **fall short of customer expectations**.

Q: Do hotels with a swimming pool get significantly better ratings than those without?

Property Type	P-code
Resort	1
Guest House	2
Cottage	3
Hotel	4
Homestay	5
Villa	6
Palace	7
Lodge	8
Houseboat	9
Service Apartment	10

BnB	11
Farm Stay	12
Hostel	13
Beach Hut	14
Palace	15
Luxury Yacht	16
Motel	17
Tent	18
Bungalow	19

t-Test: Paired Two Sample for Means

	swimming Pool	*Average rating*
Mean	0.113698973	2.266430539
Variance	0.10079676	3.793233725
Observations	3993	3993
Pearson Correlation	0.202158654	
Hypothesized Mean Difference	0	
df	3992	
t Stat	-71.26050415	
P(T<=t) one-tail	0	
t Critical one-tail	1.645235422	
P(T<=t) two-tail	0	
t Critical two-tail	1.960558418	

t-Test Results Interpretation

Null Hypothesis (H_0): There is no significant difference between Swimming Pool availability and Average Rating.

Alternative Hypothesis (H_1): There is a significant difference.

Key Results:

t Stat = -71.26 (very large in magnitude).

t Critical (two-tail) = 1.96 → Since $|t\ Stat| \gg t\ Critical$, we reject H_0.

p-value (two-tail) = 0.000… (≈0) → This is far below 0.05, confirming significance.

Pearson Correlation = 0.20 → Indicates a weak but positive correlation between having a swimming pool and average rating.

Conclusion:

There is a statistically significant difference between hotels with and without swimming pools in terms of their average ratings.

Even though the correlation is weak (0.20), hotels with swimming pools tend to receive higher ratings on average.Practically, this suggests that having a swimming pool adds to customer satisfaction, though it is not the only factor driving

Regression:

Q: Can we **predict site_review_rating** using Service Quality, Amenities, Food, and Cleanliness?

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.995731184
R Square	0.991480591
Adjusted R Square	0.991472061
Standard Error	0.179902706
Observations	4000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	15047.59	3761.897	116233.6	0
Residual	3995	129.2981	0.032365		
Total	3999	15176.89			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*
Intercept	0.018394806	0.004354	4.224879	2.44E-05	0.009859	0.026931	0.009859
Service Quality	0.422524947	0.014552	29.03515	2.4E-168	0.393995	0.451055	0.393995
Amenities	0.385243124	0.014722	26.16801	1.9E-139	0.35638	0.414106	0.35638
Food and Drink	0.060628146	0.010917	5.553448	2.98E-08	0.039224	0.082032	0.039224
Cleanliness	0.136318412	0.011663	11.6879	4.68E-31	0.113452	0.159185	0.113452

RESIDUAL OUTPUT			
Observation	*Predicted site_review_rating*	*Residuals*	*Standard Residuals*

1	3.880934102	0.119066	0.662166
2	4.729934146	-0.22993	-1.27874
3	2.234761542	0.265238	1.475082
4	5.041967951	-0.04197	-0.2334
5	2.666842522	0.133157	0.740534
6	4.042182952	-0.14218	-0.79073
7	0.018394806	-0.01839	-0.1023
8	4.018423666	-0.01842	-0.10246
9	3.069820516	-0.06982	-0.3883
10	0.018394806	-0.01839	-0.1023
11	4.367324943	-0.06732	-0.37442
12	0.018394806	-0.01839	-0.1023
13	0.018394806	-0.01839	-0.1023
14	0.018394806	-0.01839	-0.1023
15	4.603063999	-0.00306	-0.01704
16	3.646777342	-0.04678	-0.26014
17	1.023109435	-0.02311	-0.12852
18	0.018394806	-0.01839	-0.1023
19	4.17997728	0.020023	0.111353
20	0.018394806	-0.01839	-0.1023

Regression Analysis Interpretation

1. Model Fit

R Square = 0.9915 → The model explains ~99.1% of the variation in ratings.

Adjusted R Square = 0.99147 → Model remains very strong, with minimal overfitting.

F-statistic = 116,233.55, $p < 0.000$ → The model is highly significant overall.

2. Coefficients

Intercept = 0.018 ($p < 0.001$) → Baseline rating when all predictors = 0.

Service Quality ($\beta = 0.423$, $p < 0.000$) → Strongest positive effect. A 1-point increase in Service Quality raises the overall rating by ~0.42 points.

Amenities ($\beta = 0.385$, $p < 0.000$) → Second most influential predictor. A 1-point increase improves rating by ~0.39 points.

Food and Drink ($\beta = 0.061$, $p < 0.001$) → Positive but weaker effect. A 1-point increase improves rating by ~0.06 points.

Cleanliness ($\beta = 0.136$, $p < 0.000$) → Significant effect. A 1-point increase raises rating by ~0.14 points.

✅ Conclusion

The regression model is statistically significant and explains almost all variation in ratings.

Service Quality and Amenities are the two strongest drivers of customer ratings.

Cleanliness has a moderate but important effect.

Food and Drink, while significant, has the smallest impact compared to the others.

🔎 Managerial Insights

Hotels should prioritize Service Quality (staff, responsiveness, guest experience) and Amenities (comfort, facilities) to maximize customer ratings.

Cleanliness, while not the top driver, remains crucial for guest satisfaction.

Investment in Food & Drink improves ratings, but less strongly than service and amenities.

Interpretation of the Results

Most Residuals Are Small

Example: Obs 4 → Predicted = 5.042, Residual = -0.042 → **very close prediction.**

Most residuals are between -0.2 and +0.3, which means the model **is highly accurate.**

No Large Outliers in Standard Residuals

Standard residuals are mostly between -1.3 and +1.5.

In regression, residuals beyond ±2 (or ±3) usually indicate outliers.

Here, none of the shown cases exceed that range → no strong outliers.

Model Accuracy is Very High .Since the residuals are small and centered around zero, the model fits the data extremely well.This matches the earlier $R^2 = 0.991$, meaning predictions almost perfectly match actual ratings.

Example Observations

Obs 2 → Predicted = 4.73, but Actual was lower (Residual = -0.23). The model slightly overestimated.

Obs 3 → Predicted = 2.23, Actual was higher (Residual = +0.27). The model slightly underestimated.

Obs 7, 10, 12–14, 18 → Predicted ≈ 0.018, Residual ≈ -0.018. These seem like boundary or default values, maybe properties with missing/zero inputs.

✅ Conclusion

The residuals are small and evenly distributed, showing that the regression model provides an excellent fit.here are no extreme outliers in this subset.Predictions closely match actual ratings, confirming the model is statistically valid and practically reliable.

Regression

Q: Does **room_count** significantly affect site_review_rating?

SUMMARY OUTPUT

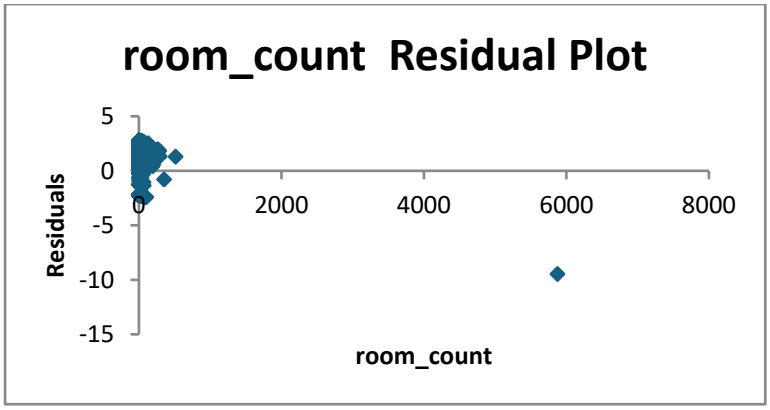
Regression Statistics	
Multiple R	0.093651684
R Square	0.008770638
Adjusted R Square	0.008522707
Standard Error	1.939800327
Observations	4000

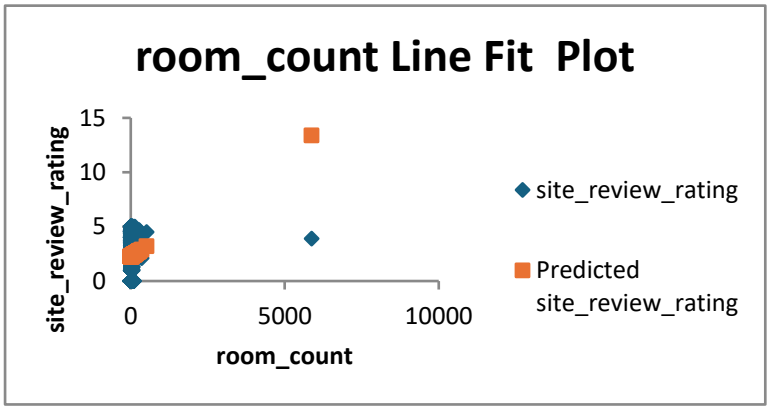
ANOVA

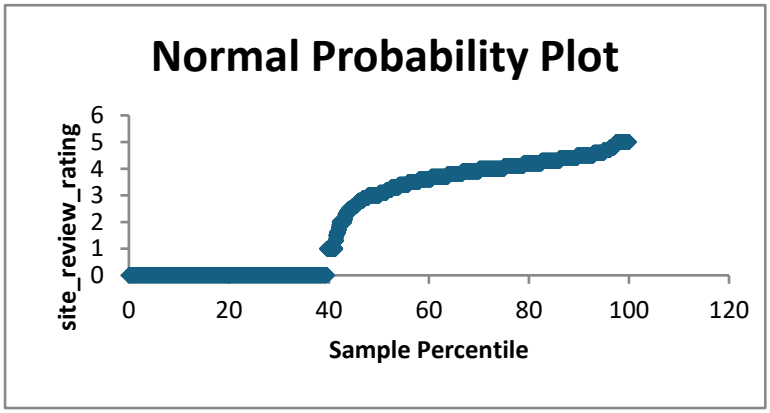
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	133.111	133.111	35.37527	2.95E-09
Residual	3998	15043.78	3.762825		
Total	3999	15176.89			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.223467183	0.031478	70.63477	0	2.161752	2.285182
room_count	0.001897853	0.000319	5.947712	2.95E-09	0.001272	0.002523

Observation	*Predicted site_review_rating*	*Residuals*	*Standard Residuals*		*Percentile*	*site_review_rating*
1	2.255730689	1.744269	0.899313		0.0125	0
2	2.257628542	2.242371	1.156125		0.0375	0
3	2.251934982	0.248065	0.127898		0.0625	0
4	2.269015661	2.730984	1.408045		0.0875	0
5	2.261424248	0.538576	0.27768		0.1125	0
6	2.251934982	1.648065	0.849712		0.1375	0
7	2.242445716	-2.24245	-1.15616		0.1625	0
8	2.269015661	1.730984	0.892463		0.1875	0
9	2.22726289	0.772737	0.398409		0.2125	0
10	2.231058596	-2.23106	-1.15029		0.2375	0







Conclusion

The regression model underfits: it predicts nearly the same rating for all hotels (around 2.2).

Residuals show large deviations (±2 to 2.7), meaning actual ratings differ substantially.

No extreme statistical outliers (since standard residuals < |2|), but practically the model is poor at prediction.

The predictor (room_count) alone is not enough to explain or predict site_review_rating.

5.4 Who Can Benefit from Your Analysis or Dashboard

Hotel Owners and Managers

- Can identify which **facilities and services** most influence guest satisfaction.
- Helps in **strategic decision-making** related to pricing, amenities, and customer experience improvements.
- Enables them to benchmark performance against other properties in similar categories or regions.

Hospitality Marketing Teams

- Understand which states or cities have the **highest customer satisfaction** to target advertising efforts effectively.
- Use insights on customer preferences to design **personalized marketing campaigns** and **loyalty programs**.
- Predict customer behavior and improve brand positioning in competitive markets.

Tourism Boards and Policy Makers

- Can use regional performance trends to identify **tourism growth opportunities**.
- Support policy planning for **infrastructure and hospitality development**.
- Evaluate the quality of hospitality services across states to promote sustainable tourism.

Data Analysts and Researchers

- Gain access to a rich dataset for applying **data science, statistical modeling, and visualization techniques**.
- Useful for academic research in **hospitality management, consumer behavior, and service quality analytics**.
- Enables experimentation with **predictive modeling and trend forecasting**.

Tour Operators and Travel Agencies

- Use analysis results to **recommend high-rated hotels** to customers.
- Strengthen partnerships with well-performing hotels and improve package design.
- Helps in building **trust and reliability** with travelers by promoting top-quality accommodations.

Customers and Travelers

- Indirectly benefit through **improved hotel experiences** driven by data-informed service enhancements.
- Better quality, cleaner environments, and more value for money — outcomes of data-backed decision-making by hotels.

5.5 Key Learnings from the Project

- **Technical Skills:**

  Core Excel for Data Analysis: Developed a solid foundation in using Excel's essential tools for data analysis. This includes:

  Data Cleaning & Preparation: Gained hands-on experience in sorting, filtering, and using **Text-to-Columns** to organize raw data. I learned to handle inconsistencies and prepare a clean dataset for analysis.

  Essential Formulas: Became proficient with fundamental Excel functions crucial for analysis, such as **IF statements, SUM, AVERAGE, COUNT, and basic logical operators** to create new metrics and categorize data.

  Data Summarization & Visualization: Mastered the use of **PivotTables** to quickly and effectively summarize large datasets, identify trends, and calculate key metrics. I learned to create a variety of charts (**Bar Charts, Pie Charts, Line Graphs**) directly from this summarized data to communicate findings visually.

- **Professional Skills:**

  Structured Problem-Solving: Enhanced my ability to tackle a complex business question by breaking it down into a logical sequence: data preparation, summarization, visualization, and interpretation.

  Attention to Detail: Cultivated a meticulous approach to data work, carefully checking calculations and ensuring that my summaries accurately reflected the underlying data.

- **Confidence:**

  Successfully completing an end-to-end project using a fundamental tool like Excel has greatly boosted my confidence. It proved that I could derive meaningful insights and create clear, valuable reports without needing advanced software. This project demonstrates my ability to learn a process, pay attention to detail, and use accessible tools to solve real-world problems.