The Dynamic Pricing and Human Resource Analysis for a Guest House

A Mid Term report for the BDM capstone Project

Submitted by

Name: Rahul Pathak

Roll number: 23f2000798

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai Tamil Nadu, India, 600036

Contents

Declaration Statement

I am working on a Project titled "**The Dynamic Pricing and Human Resource Analysis for a Guest House**". I extend my appreciation to **Garjiya Homestay**, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfillment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

Signature of Candidate:

Name: Rahul Pathak

Date: July 8, 2025

1 Executive Summary

Project Title: The Dynamic Pricing and Human Resource Analysis for a Guest House

Nestled in the tourist-rich region near Jim Corbett National Park, the guest house under study operates with 6 rooms (4 AC, 2 Non-AC) and caters to both leisure and business travelers. Despite consistent demand, the owner faces two core challenges: (1) inconsistent revenue due to lack of good room pricing across seasons, and (2) declining staff morale and high turnover, particularly among receptionists. These issues are limiting profitability and operational efficiency, prompting the need for a data-driven approach to optimize pricing and HR decisions.

Primary data was collected manually over a 12-month period (May 2024–April 2025) from booking records, competitor research, customer reviews, and employee feedback. The dataset captures over 15 key variables including guest check-in/out dates, room types, seasonal indicators, occupancy rates, pricing, revenue, and employee satisfaction metrics. Supporting data includes interviews, images of the service facility, and review sentiment analysis. Descriptive statistics highlighted pricing variability, high occupancy in peak months, and staff dissatisfaction linked to long hours and low pay.

Using Python and Excel, regression models were built to predict optimal final prices based on base rates, room type, occupancy, and season. The model achieved an R^2 of 0.93 with an RMSE of ₹150, indicating strong predictive performance. Simultaneously, HR analysis revealed low job satisfaction driven by poor management relations. The remaining phase will focus on refining the pricing model, benchmarking with competitors, and proposing workforce retention strategies for long-term sustainability.

2 Proof of Originality of Data

To verify the originality and authenticity of the primary data used in this project, I have included the following supporting materials:

- A signed letter from the guest house owner on official letterhead
- Images of the property
- A short video (in English) capturing a conversation with the owner
- The Manual Records photos

All of these resources have been compiled and securely shared via the following Google Drive link:
https://drive.google.com/drive/folders/1dSYrSk0cI0_bESPAbss6cO3gpWK5oxTf?usp=sharing

3 Metadata and Descriptive Statistics

3.1 Metadata

Metadata refers to data about the data — it provides detailed context about each variable collected, explaining its significance and how it contributes to solving the core problems in the project: optimizing guest house revenue through dynamic pricing and reducing employee dissatisfaction.

Source: The data was manually collected from the guest house's physical registers covering the period from **May 2024 to April 2025**, and later digitized in Excel. Employee data was collected through **interviews with the manager and owner**, along with basic documentation regarding roles, tenure, wages, and feedback.

Guest Booking Dataset Metadata

Variable	Description	Relevance
Guest_Name	Name of the guest	Used for identification (anonymized during analysis)
Customer_ID	Unique ID assigned to each guest	Helps eliminate duplicates and track repeat customers
Check_In / Check_Out	Date of arrival and departure	Used to calculate total nights stayed and daily occupancy
Room_Type	AC or Non-AC room	Key factor in pricing and room preference trends
Base_Price_Per_Night	Standard rate fixed by the owner	Serves as the reference point for evaluating dynamic pricing
Final_Price_Per_Night	Actual rate charged after discussion or adjustment	Central variable for modeling dynamic pricing
Group_Size	Total number of guests in one booking	Indicates resource usage and helps forecast demand

Visit_Purpose	Tourist / Personal	Useful for segmenting customers and understanding seasonal variations
Total_Amount	Final bill paid for entire stay	Core input for revenue and trend analysis
Bargain_or_Member	Y = Bargained, N = No Bargain, S = Paid above base	Reveals negotiation trends and perceived value by customers
Season	Peak / Mid-Peak / Off	Critical input for modeling seasonal pricing behavior
Rooms_Available / Rooms_Booked	Daily room stats	Used to derive occupancy percentage
Occupancy_%	% of booked rooms per day	Key performance indicator (KPI) for demand and utilization
Revenue	Total daily earnings	Core business metric reflecting performance
Review_text / Review_rating	Guest feedback in text and numeric form	Used for sentiment and service quality analysis
Sentiment_Label	Positive / Neutral / Negative	Derived using NLP tools for customer perception analysis
Would_Recommend	Y = Yes, No = No , N = Neutral	Direct measure of customer satisfaction and loyalty
Competitor_Name / Price_Per_Night / Ratings	Data from similar guest houses	Used for competitive benchmarking of pricing and services

Employee Dataset Metadata

Variable	Description	Relevance
Employee_ID / Name	Unique identifier	Internal tracking, anonymized in report

Role / Department	Job position and area	Helps analyze satisfaction across work types
Tenure_Years	Duration of employment in years	Correlates with satisfaction and likelihood of resignation
Wage_Per_Month	Monthly compensation	Compared with satisfaction and exit feedback
Job_Satisfaction_Score	Rating on a scale of 1–10	Central HR metric in identifying engagement levels
Interview_Comments	Qualitative feedback from employee	Used for thematic analysis of dissatisfaction causes
Exit_Feedback / Exit_Status	Reason and type of exit (Resigned/Terminated)	Critical for analyzing attrition and organizational issues

Justification of Metadata

The above variables were selected to enable a dual-layered analysis:

1. **Business Optimization:** Guest booking variables such as room type, season, occupancy, and final price help design a **dynamic pricing model** that reflects demand patterns.
2. **HR Insight:** Employee data such as job satisfaction, tenure, and exit reasons supports a **turnover analysis**, identifying root causes behind workforce instability.

By combining financial and HR perspectives, this metadata allows a comprehensive, data-driven strategy to improve **profitability** and **operational stability** of the guest house.

Link to Data –

Booking Data – Sales_Data.xlsx

Employee Data- Employee_Data.xlsx

3.2 Descriptive Statistics

1. Guest Booking Data (May 2024 – April 2025)

Metric	Revenue (₹/day)	Occupancy (%)	Final_Price (₹/night)
Mean	₹6,012.19	49.3%	₹2,010.20
Median	₹5,800	50%	₹2,100

Mode	₹0	50%	₹2,100
Standard Deviation	₹3,590.48	26.68%	₹559.51
Range	₹0 – ₹19,500	0% – 100%	₹900 – ₹2,800

Observation:

Daily revenue and occupancy show **high variability**, reflecting seasonal patterns. While occupancy swings significantly, the **final price per night stays within a narrow range**, indicating that the guest house has **not yet fully adopted dynamic pricing**. This validates the need for a predictive model that incorporates demand and seasonality into pricing decisions.

2. Employee Satisfaction Data

Metric	Satisfaction Score (1–10)	Tenure (Years)	Wage (₹/month)
Mean	7.2	1.06	₹18,055.56
Median	8	0.6	₹15,000
Mode	8	3	₹15,000
Standard Deviation	1.92	1.11	₹12,233.26
Range	4 – 10	0.2 – 3	₹7,500 – ₹50,000

Observation:

Employees with **tenure less than 2 years** reported **lower satisfaction** and made up the majority of exits. Interestingly, salary differences were minimal across roles, suggesting that dissatisfaction may stem from **non-financial factors** such as long work hours, night shifts, and communication issues with management — all of which were evident in interview feedback.

4 Detailed Explanation of Analysis Process/Method

Data Cleaning, Structuring, and Analysis

The analysis process for this project started with the **manual collection of booking records and employee data** from the guest house, covering the period from **1st May 2024 to 30th April 2025**. These records were initially handwritten and then digitized using **Microsoft Excel**, followed by advanced analysis and modeling in **Python**.

1. Data Cleaning and Preprocessing

Raw data contained inconsistencies such as spelling errors, missing fields, duplicates, and non-standard formats. To ensure accuracy:

- **Missing values** were filled using context-aware assumptions (e.g., if total amount was missing, it was computed using per-night price × stay duration).
- **Outliers** (e.g., price entries below ₹500 or above ₹4000) were manually validated.
- **Categorical data** such as Season, Room_Type, Bargain_Status were standardized (e.g., 'Y', 'N', 'S' in discounts).
- **Derived columns** like Occupancy %, Length of Stay, and Revenue were calculated using Excel formulas.

Employee feedback, collected through interviews, was cleaned by standardizing language, removing noise, and preparing it for thematic sentiment categorization.

2. Excel-Based Analysis

Excel was used to conduct **foundational analysis** of the business's operations. Key highlights include:

- **Occupancy Trends:**
  Season-wise occupancy (%) was plotted, showing that **Peak Season** had highest occupancy (~80–90%), while **Off Season** dropped below 30%.
- **Revenue Analysis:**
  Daily revenue was calculated and visualized. A clear seasonality pattern was seen — **December and May months** consistently generated the highest earnings.
- **Room Type Insights:**
  AC rooms contributed to over **70% of total revenue**, with higher prices and longer stays. **Non-AC rooms were mostly booked by group travelers** during mid or off-seasons.
- **Pricing Strategy Check:**
  Comparison of **Base Price vs Final Price** showed that only a portion of bookings had significant price negotiation, confirming that **dynamic pricing is being used inconsistently**.
- **Competitor Benchmarking:**
  Data from 8-9 nearby guest houses (collected via Google) was compiled. Most competitors offered similar rates for AC rooms but lower ratings. This indicates **scope to charge premium if service quality improves.**
- **Customer Feedback Analysis:**
  Reviews were categorized into Positive, Neutral, and Negative. "Would Recommend" data showed that **98% of guests would return**, indicating strong customer retention potential.

- **Employee Satisfaction Summary:**
  Job satisfaction data showed that **short-tenure employees (under 1 year)** had lower ratings and were more likely to exit. This highlighted **non-monetary issues** like long working hours and poor work-life balance.

3. Python-Based Predictive Modeling

To enhance the decision-making, a **multiple linear regression model** was built using Python to predict **Final Price per Night** using variables such as Room Type, Season, Base Price, and Occupancy %.

Model Results:

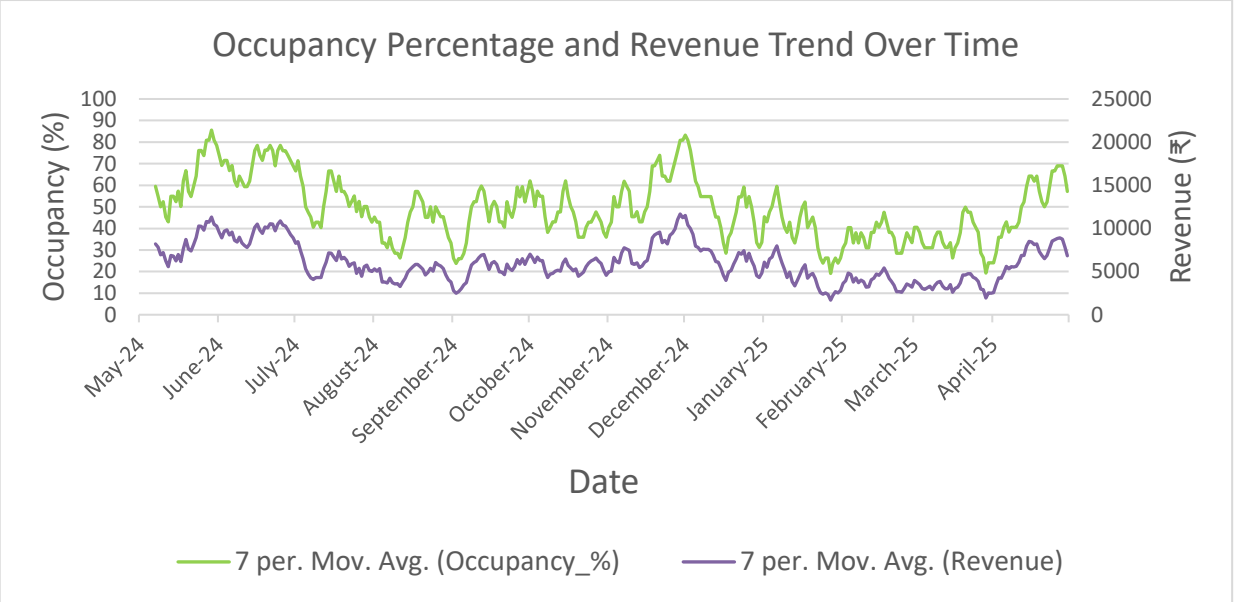
- **R^2 Score:** 0.93 (93% variance explained)
- **RMSE:** ₹150.43
- **Important Coefficients:**
  - Base_Price_Per_Night: 1.15
  - Room_Type_Non-AC: +₹85.28
  - Season_Peak: -₹81.98
  - Occupancy %: +₹0.42

This model proved that **Final Price** depends strongly on Base Price and occupancy, while seasonality and room type also play important roles.

5 Results and Findings

The analysis of the guest house data using both **Excel-based visualization** and **Python modeling** revealed several key patterns and insights related to occupancy, pricing behavior, customer preferences, and employee satisfaction.

Occupancy and Revenue Trends

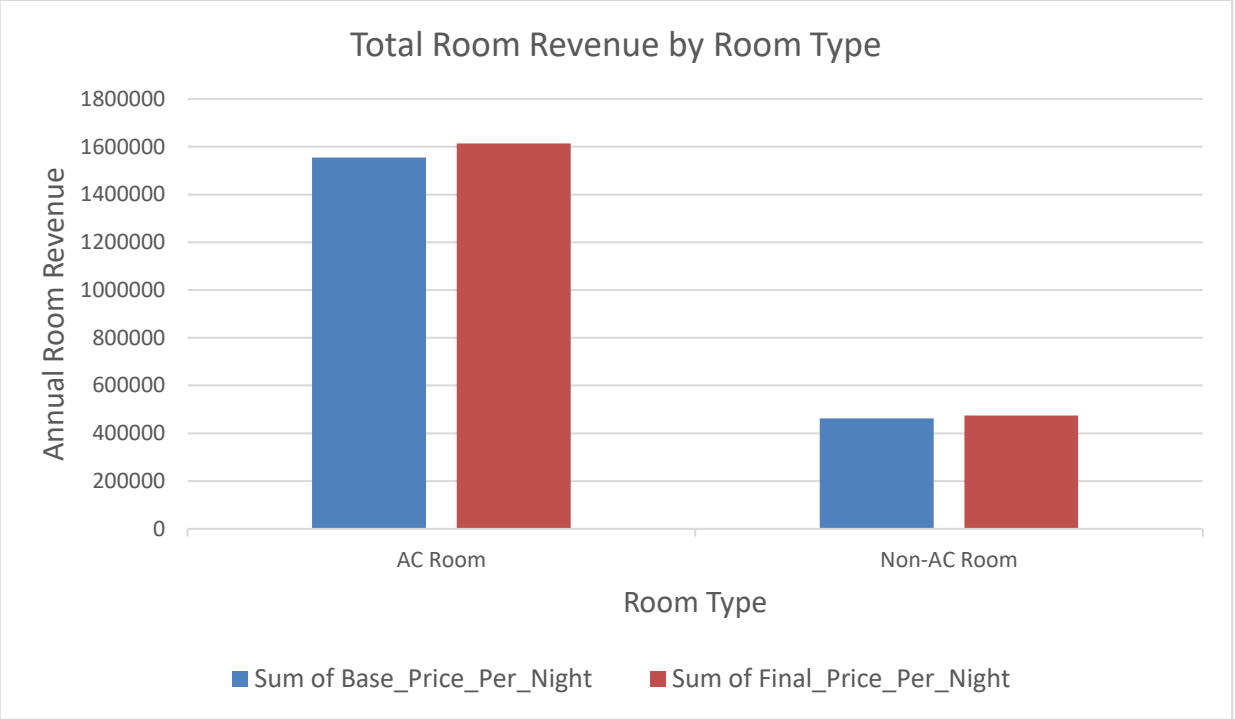


The dual-axis line chart displays the **7-day moving average** of both **occupancy percentage (green)** and **daily revenue (purple)** from **May 2024 to April 2025**, revealing a strong **positive correlation** between the two metrics. Occupancy peaked during **June, December, and April**, often crossing **70–80%**, which aligns with **peak travel seasons** such as summer vacations, year-end holidays, and spring breaks. These surges directly translate into corresponding spikes in revenue, confirming that higher occupancy drives greater earnings.

Conversely, **occupancy dropped below 30% during August and February**, indicating off-season periods when demand was lower, potentially due to extreme weather, academic schedules, or local events. The revenue line follows a similar dip, suggesting that the guest house's earnings are highly sensitive to seasonal demand.

Interestingly, revenue exhibits **more volatility than occupancy**, especially during shoulder months like October and March, indicating that pricing decisions, group size variations, or discounting strategies could be influencing the earnings even when occupancy remains moderate. The use of **7-period moving averages** smooths daily fluctuations and reveals broader patterns, aiding in **strategic pricing and marketing planning**.

Room Type Revenue Analysis



The bar chart displays the **annual total revenue** generated from **AC and Non-AC rooms**, comparing both the **base price per night (blue)** and the **final price per night (red)**. The insights clearly reveal that **AC rooms overwhelmingly dominate revenue generation**, contributing over **₹16 lakhs annually**, while Non-AC rooms generate less than **₹5 lakhs**.

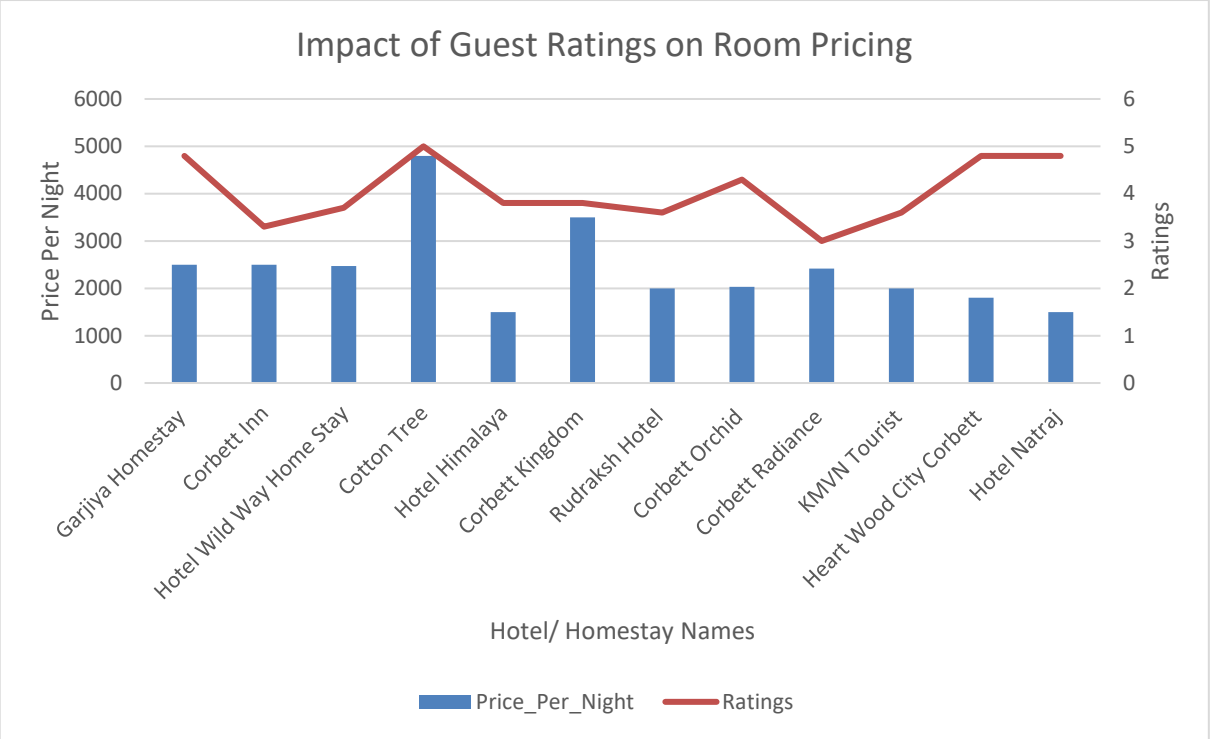
Despite the guest house having only **4 AC rooms compared to 2 Non-AC**, the revenue from AC rooms is over **three times higher**, indicating their **higher booking frequency**, **longer stays**, or **better pricing leverage**. The small difference between the base and final price bars for AC rooms suggests **limited discounts or even premium pricing**, particularly during high-demand periods.

In contrast, Non-AC rooms show almost identical base and final price totals, implying **fixed pricing with minimal bargaining or upgrades**. This could be due to these rooms being less in demand or only booked when budget constraints are high or peak season options are full.

This analysis strongly supports **revenue-maximizing strategies** such as:

- Further premium pricing of AC rooms during peak seasons.
- Repositioning or improving Non-AC offerings to boost demand.
- Introducing variable pricing for Non-AC rooms based on occupancy.

Competitor Pricing vs Ratings Analysis



The chart illustrates the relationship between **room prices (blue bars)** and **guest ratings (red line)** for 13 nearby hotels and homestays. A clear, though **non-linear**, pattern emerges where **higher guest ratings tend to correlate with premium pricing**, but **brand identity and property type** also play a key role.

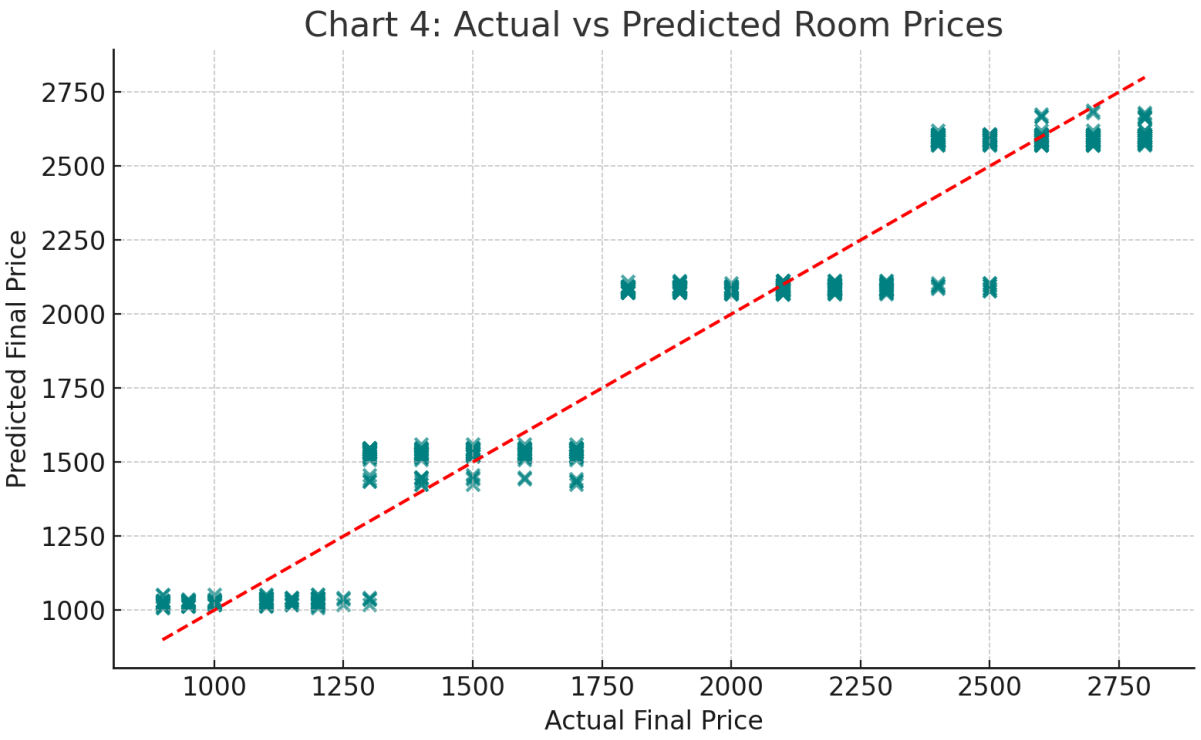
For example, **Hotel Himalaya**, with a **perfect 5-star rating**, charges the **highest price (~₹4,800)** per night, indicating that guests are willing to pay a premium for quality and satisfaction. Similarly, **Cotton Tree**, also rated highly, maintains above-average pricing. On the other hand, **Hotel Natraj** has the **lowest rating and the lowest price (~₹1,500)**, showing the direct impact of poor guest perception on pricing capability.

However, there are outliers like **Gairiya Homestay** and **Corbett Inn**, which charge mid-range prices (~₹2,500) despite having above-average ratings, suggesting a **value-for-money positioning** strategy. Conversely, properties like **Corbett Kingdom** and **Rudraksh Hotel** charge relatively high prices with only moderate ratings, indicating possible overpricing or location-based demand.

Overall, this graph highlights that **guest ratings significantly influence room pricing**, but **factors like branding, amenities, and target market segment** also play a decisive role. This

insight can guide dynamic pricing strategies and quality improvement efforts for competitive advantage.

Actual vs Predicted Room Prices



The chart visualizes the relationship between actual room prices (x-axis) and the prices predicted by the model (y-axis). Each dot represents a room listing, while the red dashed line indicates the ideal case where predicted and actual prices are equal.

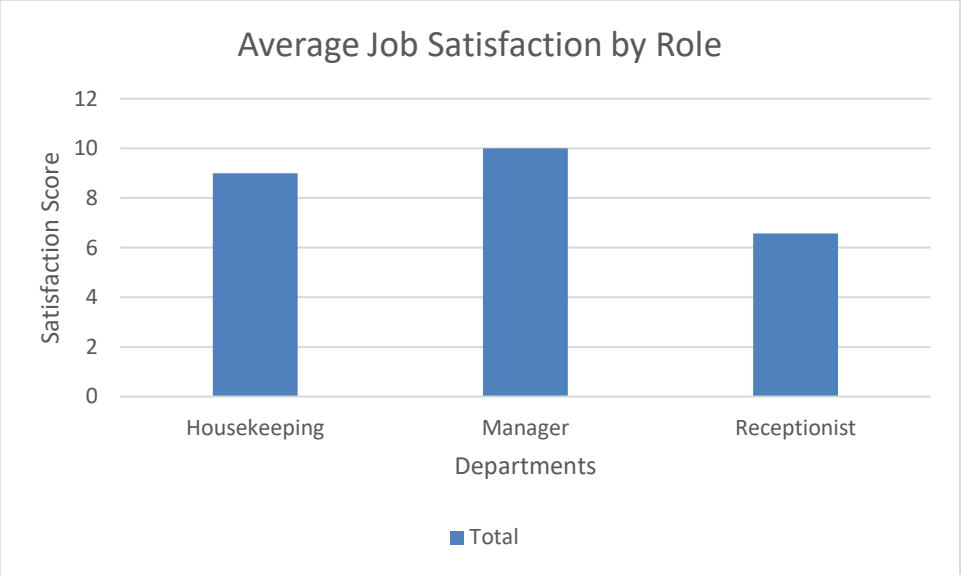
A clear trend is observed where predicted prices increase with actual prices, suggesting that the model has learned the basic pricing pattern. However, the model struggles with accuracy in the mid to high price range. For example, many actual prices between ₹1600–₹2100 are underpredicted, with predicted values often falling below the red line. Similarly, for higher-end rooms above ₹2300, the model's predictions appear clustered and less responsive to actual price variations, hinting at saturation or lack of training data in that range.

In contrast, for lower-priced rooms (below ₹1500), predictions are much closer to actual values, showing better model reliability in this segment.

Overall, the graph highlights that while the model captures the general pricing trend, it underestimates mid-range prices and shows limited sensitivity at the upper end. These patterns

indicate the need for model improvement, especially by including more pricing drivers like ratings, amenities, or seasonal demand.

Average Job Satisfaction by Role



The chart compares average job satisfaction scores across three key roles in the hospitality workforce: Housekeeping, Manager, and Receptionist. A clear hierarchy emerges in satisfaction levels, which may reflect differences in work environment, responsibilities, and perceived value.

Managers report the highest satisfaction score of **10**, indicating strong alignment with their role, possibly due to better compensation, autonomy, and respect within the organization. Housekeeping staff follow with a respectable score of **9**, suggesting a generally positive work experience, likely influenced by teamwork and clear role expectations.

However, Receptionists report the lowest average satisfaction at **7**, which may highlight underlying challenges such as customer-facing stress, shift-related fatigue, or lack of upward mobility. This relative dissatisfaction signals an opportunity for targeted improvement in support systems or recognition mechanisms for this role.

Overall, the chart highlights that while managerial roles enjoy the highest satisfaction, frontline roles—especially reception—require attention to boost morale and retention. Addressing these gaps can lead to a more balanced and motivated workforce across all departments.