Egypt University of Informatics
Faculty of Computing & Information Sciences
**C-DE211** Data Analysis Course Project (2)

# THE STATE OF HOTELS IN EGYPT

| | |
|---|---|
| **Submission Date** | 24 / 05 / 2025 |
| **Submitted By** | Jana Mohamed Abdelhafeez 23-101097<br>Mahy EmadEldin Omar 23-101079<br>Malak Mohamed Ali 23-101124 |
| **Supervised By** | Dr. Mohamed Taher ElRefaei<br>TA Shereen Elfayoumi<br>TA Nadine Elsaeed |

## 01. INTRODUCTION

This report explores what truly influences guest satisfaction when booking hotels in Egypt. The analysis is based on a data-set collected from online booking platforms, containing real reviews and ratings from people who actually stayed in these properties. This includes detailed information such as hotel names, locations, star ratings, satisfaction scores, number of reviews, room types, price ranges, amenities, and location ratings. These columns offer a well-rounded view of both guest experiences and hotel features, helping us understand which factors contribute most to a satisfying stay. By examining patterns in this data, the goal is to uncover how elements like service quality, pricing, accommodation type, and overall value influence guest perception. The findings aim to provide practical insights for improving hotel offerings and guiding travelers toward better choices.

## 02. RESEARCH QUESTION

"What are the key factors that drive guest satisfaction in hotels across different regions of Egypt?"

## 03. HYPOTHESIS

### 1. Review Count Group vs. Guest Satisfaction

Null Hypothesis ($H_0$): There is no difference in average overall ratings between different review count groups.

Alternative Hypothesis ($H_1$): There is a significant difference in average overall ratings between different review count groups.

### 2. Location vs. Guest Satisfaction

Null Hypothesis ($H_0$): There is no difference in guest satisfaction scores between hotels in tourist cities and those in non-tourist cities.

Alternative Hypothesis ($H_1$): Hotels in tourist cities receive significantly higher guest satisfaction scores than those in non-tourist cities.

### 3. Price per Night vs. Guest Satisfaction

Null Hypothesis ($H_0$): There is no correlation between the price per night and guest satisfaction scores.

Alternative Hypothesis ($H_1$): There is a significant correlation between the price per night and guest satisfaction scores (positive or negative).

## 04. POPULATION OF INTEREST

The population we're focusing on includes all guests who have stayed at hotels in Egypt and shared their experiences through reviews, ratings, or booking details. These guests represent a mix of travel types, such as families, solo travelers, business visitors, and tourists across various hotel types and locations throughout the country.

# 05. METHODOLOGY 

## 1. Data Collection

This project began with gathering a comprehensive data-set from online hotel booking and review platforms. It reflects real guest experiences across hotels in Egypt, capturing a wide range of details such as hotel names, city locations, star ratings, room types, guest reviews, and satisfaction scores. These data points offer a practical lens into traveler preferences and hotel performance throughout the country.

**Source:** https://www.kaggle.com/datasets/ziadashraf3010/egypt-hotels-dataset

## 2. Data Cleaning

**Outlier Handling:**
To ensure the integrity of our analysis, outliers—those rare values that can heavily skew results—were handled using the Inter-quartile Range (IQR) method using box plots. Instead of simply removing them, extreme values were capped within logical limits to preserve the dataset's balance.

**Missing Value Treatment:**
Entries with too many missing values were excluded, while others were addressed using imputation techniques such as filling with the mean or mode, depending on the variable. This process helped enhance the dataset's reliability without discarding valuable information.

## 3. Correlation Analysis

To explore how different factors relate to one another, we performed a correlation analysis focusing on variables like guest satisfaction, review counts, star ratings, and hotel locations. Visual tools such as regression plots and heat-maps were used to make these relationships easier to interpret. These insights offered a deeper understanding of what influences a guest's overall experience.

## 4. Hypothesis Testing

**Hypothesis:**
We hypothesized that hotels situated in Egypt's most popular tourist destinations—such as Cairo, Sharm el-Sheikh, and Hurghada—would receive higher satisfaction scores, largely due to enhanced services and better facilities.

**Testing Approach:**
To examine this, we used correlation metrics and visual analysis, comparing satisfaction levels across different cities and hotel ratings.

**Findings:**
We analyzed that guest satisfaction in Egyptian hotels is shaped more by the quality of service, value for money, and overall experience than by location or hotel type—highlighting that memorable stays are driven by how guests feel, not just where they stay or how much they pay.

## 5. Data Visualization

Throughout the analysis, visualizations played a key role in communicating findings clearly and effectively. From scatter plots and box plots to heat-m---aps and bar charts, each graphic was designed to highlight trends—such as how star ratings align with guest satisfaction or how preferences vary by city. These visuals not only made complex patterns more accessible but also provided actionable insights for decision-makers in the hospitality sector.

## 06. COLLECTED DATA / DATASET

**"Egypt Hotels Data-set":** Comprehensive Data-set on Egyptian Hotels: Locations, Ratings, and Amenities.

**Source:** https://www.kaggle.com/datasets/ziadashraf3010/egypt-hotels-dataset

**Rows:** 947 , **Columns:** 19

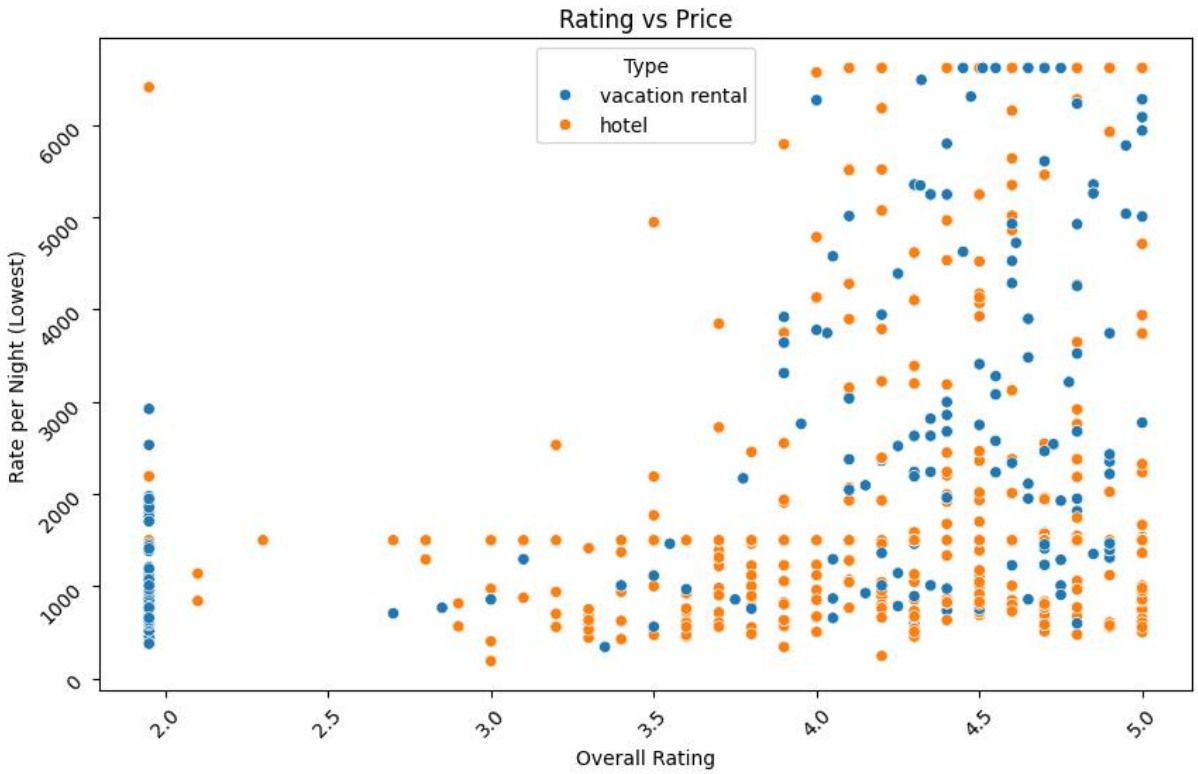
**Number of Samples used:** 947

**Columns of the data-set contain:**

- Type - hotel or vacation rental.
- Hotel Name
- Coordinates - latitude & longitude
- Check-in / Check-out Time
- Overall Rating - Star rating or customer rating.
- Reviews - Total number of customers who reviewed the place.
- Location Rating
- Price Range - A general price classification.
- Amenities - Facilities offered, like pools, gyms, etc.
- Nearby Places

# 07. DATA ANALYSIS

## A. Data Analysis




### Rating vs. Price by Property Type: "Check if higher-rated places charge more (positive slope) or not."

- The plot shows that higher prices don't guarantee higher ratings. Many highly-rated hotels charge between 500–2000 EGP, proving that good service is achievable at lower prices.
- Vacation rentals display more price variation, with some charging over 6000 EGP but not consistently receiving better ratings. This suggests price inflation without matching value.
- Overall, there's no strong correlation between price and rating. Hotels offer more predictable value, while vacation rentals may carry higher cost with uncertain quality.


## Price Distribution by Property Type: "See which type tends to be more expensive and if prices vary widely"

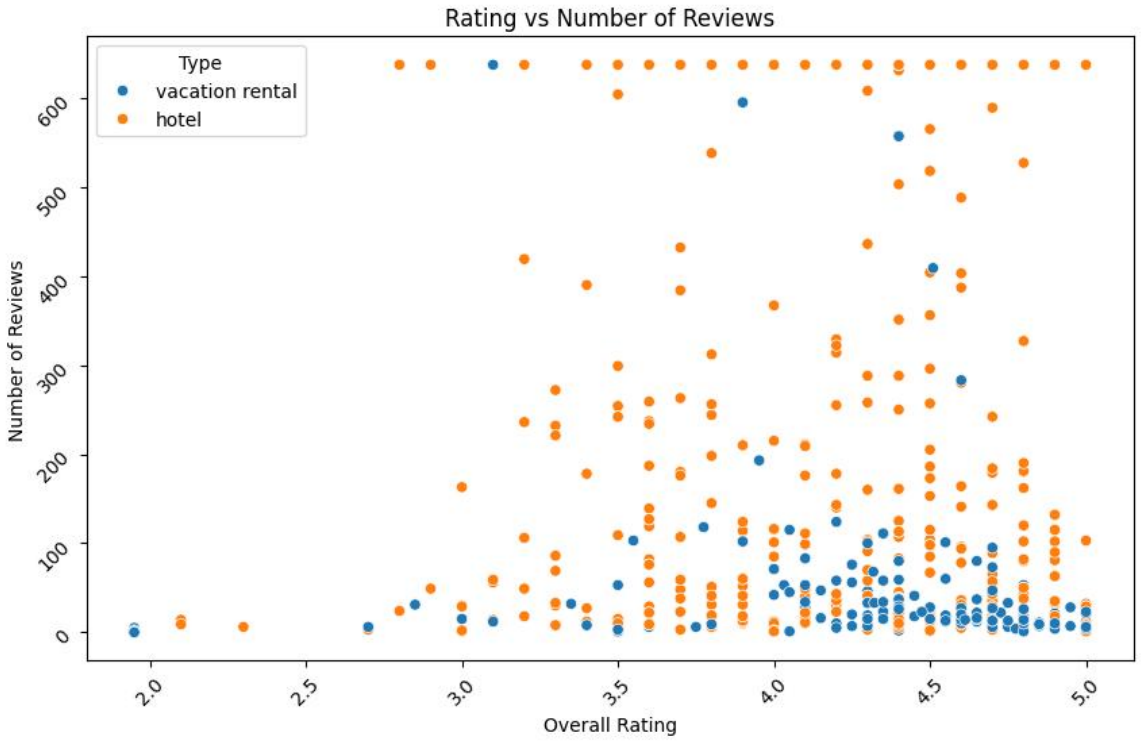
**Price Range and Distribution:** The box plot illustrate pricing differences between hotels and vacation rentals. Vacation rentals show a wider range of prices with several high-end outliers, suggesting diverse offerings in size, location, and amenities.

**Median and Variability:** Hotels generally have lower and more consistent pricing, as shown by a narrower inter-quartile range and fewer outliers. In contrast, vacation rentals display greater price variability, appealing to a broader range of traveler budgets.

**Conclusion:** Hotels tend to offer more predictable and budget-friendly prices, while vacation rentals provide flexibility and variety, though often at a higher potential cost. This makes vacation rentals attractive to travelers seeking unique stays, while hotels remain reliable for those prioritizing price consistency.

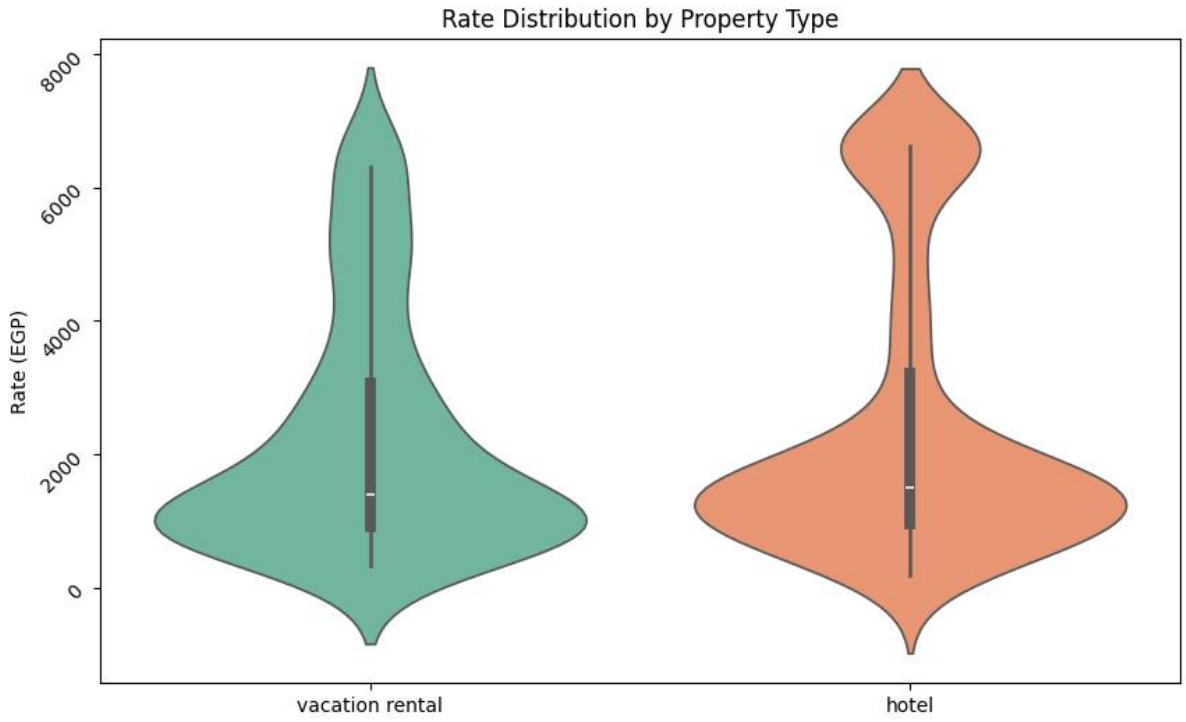

## Hotel vs. Vacation Rental Ratings: "Are hotels or vacation rentals rated higher?"

- Hotels tend to receive **higher average ratings** than vacation rentals, as shown in the bar plot. This suggests that, on average, **guests report greater satisfaction with hotels**, potentially due to more standardized services, professional management, and consistent quality.




## Number of Reviews vs. Overall Ratings: "Expensive hotels might have fewer reviews (exclusive) or more (luxury attention)."

- The graph shows no strong correlation between the number of reviews and overall ratings. Properties with both high and low ratings exist regardless of how many reviews they've received. This suggests that more reviews don't necessarily indicate better or worse quality.
- Hotels vs. Vacation Rentals: Both property types follow similar patterns. Whether it's a hotel or a vacation rental, the number of reviews does not significantly affect the rating. This implies that the type of accommodation doesn't impact this relationship.
- Rating Concentration: Most properties tend to receive ratings between 3.5 and 5, showing generally positive guest feedback, regardless of review count.
- Outliers: Some properties with very few reviews still have excellent ratings, while others with many reviews may have average scores. This highlights that factors like service, location, or amenities play a bigger role in customer satisfaction than review quantity.
- Conclusion: The number of reviews is not a reliable indicator of quality. Both hotels and vacation rentals show similar behavior, suggesting that other factors—like guest experience, location, or value—are more important when judging a property's quality.




## Price Distribution by Type: "Shows variation and skew in prices between hotels and vacation rentals."

1. Variation in Prices:

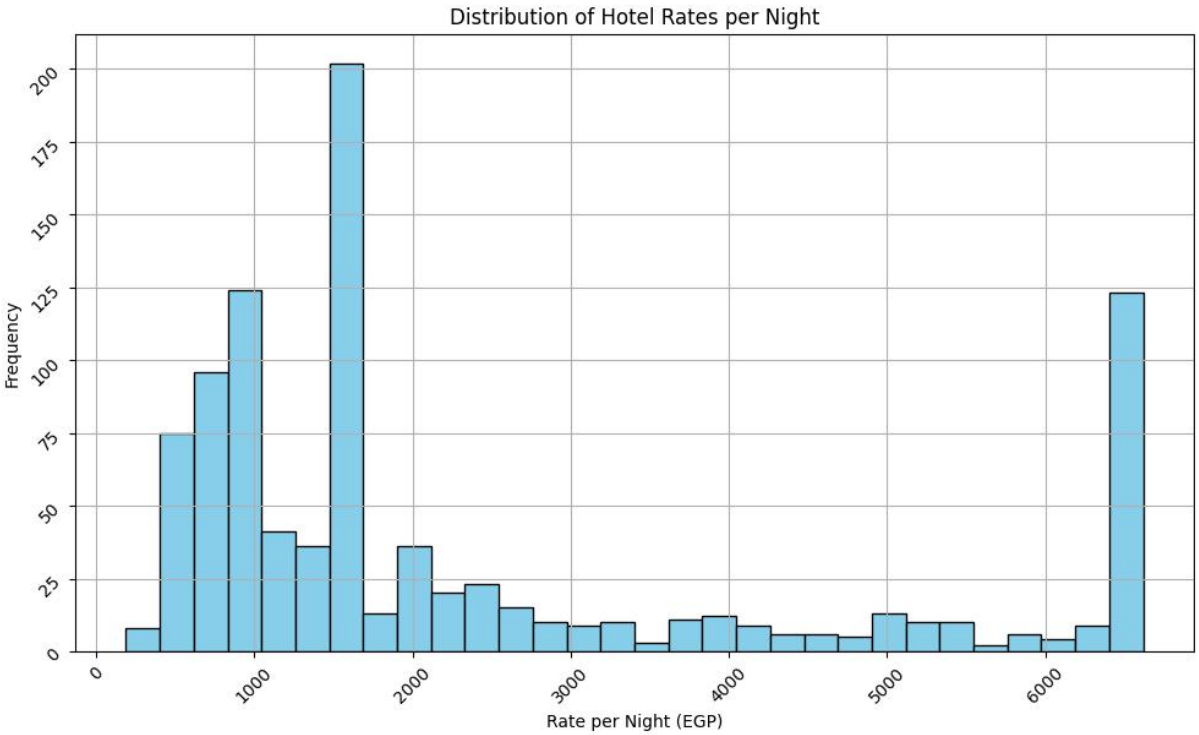
- Vacation Rentals: The plot shows a broader spread of prices, with rates ranging from very low to very high. This suggests that vacation rentals cater to a diverse market, offering both budget-friendly and luxury options.
- Hotels: The distribution of hotel rates is more concentrated, indicating less variation compared to vacation rentals. Hotels tend to have a more standardized pricing structure.

2. Skewness in Distribution:

- Vacation Rentals: The shape of the violin plot indicates a right-skewed distribution, with a noticeable number of properties offering rates on the higher end. This highlights the presence of premium vacation rentals that significantly increase the overall price range.
- Hotels: Hotel prices also exhibit some skewness, but the distribution is tighter and more centered around mid-range prices, with fewer extreme values compared to vacation rentals.

3. Conclusion:

- Vacation rentals exhibit a greater level of price diversity, making them suitable for a wider variety of travelers, from those on a tight budget to those seeking upscale accommodations.
- Hotels, on the other hand, tend to offer a more predictable pricing range, appealing to travelers looking for consistency and standardization in pricing.




## Hotel Rate Distribution: "Helps spot price clusters and popular price points."

1. Price Concentration:

- Most hotel rates fall between **500 and 2000 EGP**, indicating that the majority of accommodations in Egypt are in the **mid-range or budget-friendly** category. A distinct peak is visible around **1500 EGP**, marking it as a common and possibly competitive price point.

2. High-End Outliers:

- While the bulk of rates are concentrated at lower levels, there are **noticeable outliers exceeding 6000 EGP**. These likely represent **luxury hotels** targeting high-income travelers or offering exclusive services.

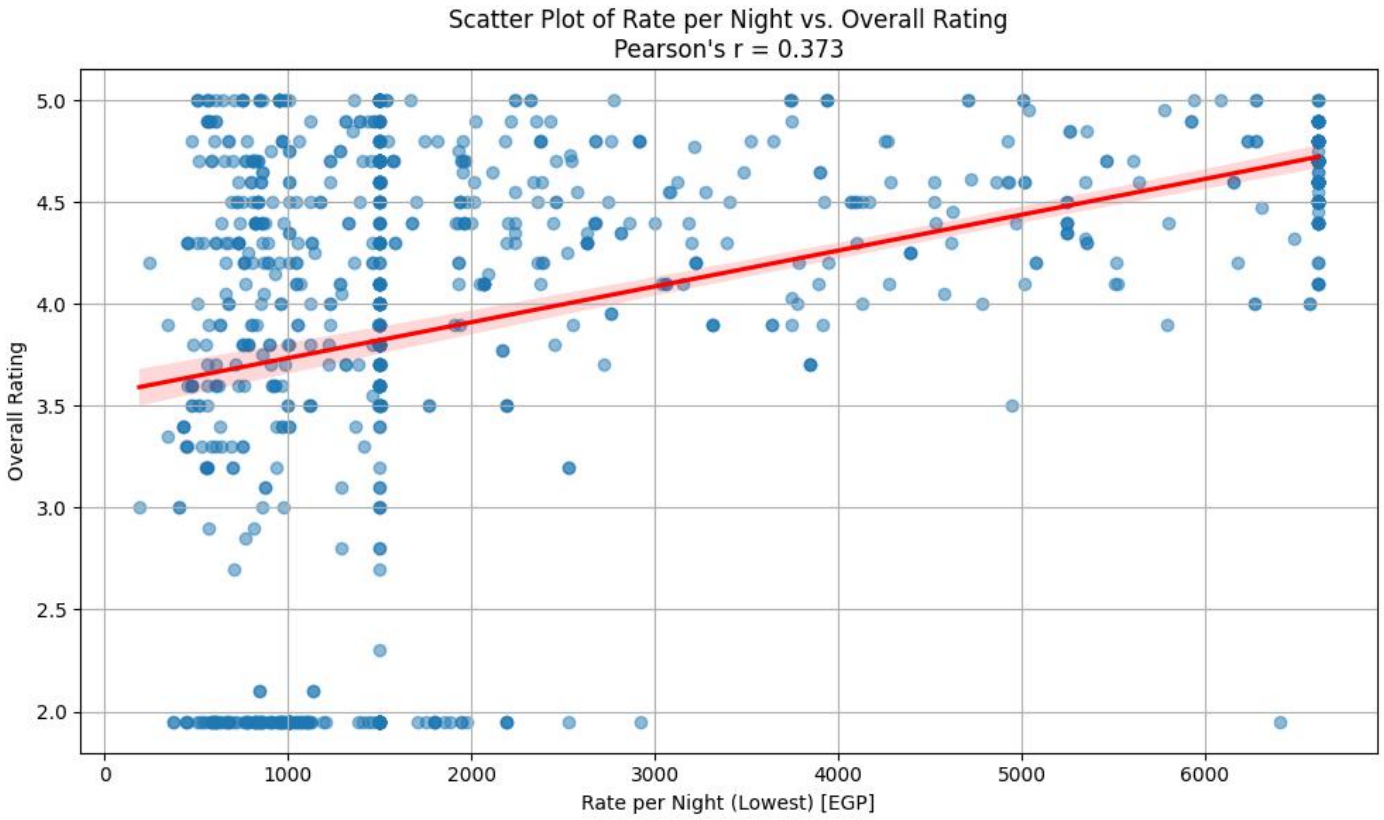
3. Distribution Shape:

- The distribution is **right-skewed**, showing that **high prices are less common** and most hotels cluster around lower price points. This suggests that affordability is a key focus for the majority of properties.

4. Conclusion:

- This histogram distribution indicates that Egypt's hotel market is primarily concentrated in the mid-range pricing segment, with distinct peaks at specific price points. This reflects high competition in the affordable segment, while luxury offerings remain limited. These findings can inform strategic pricing adjustments and targeted marketing efforts aligned with prevailing traveler budget preferences.

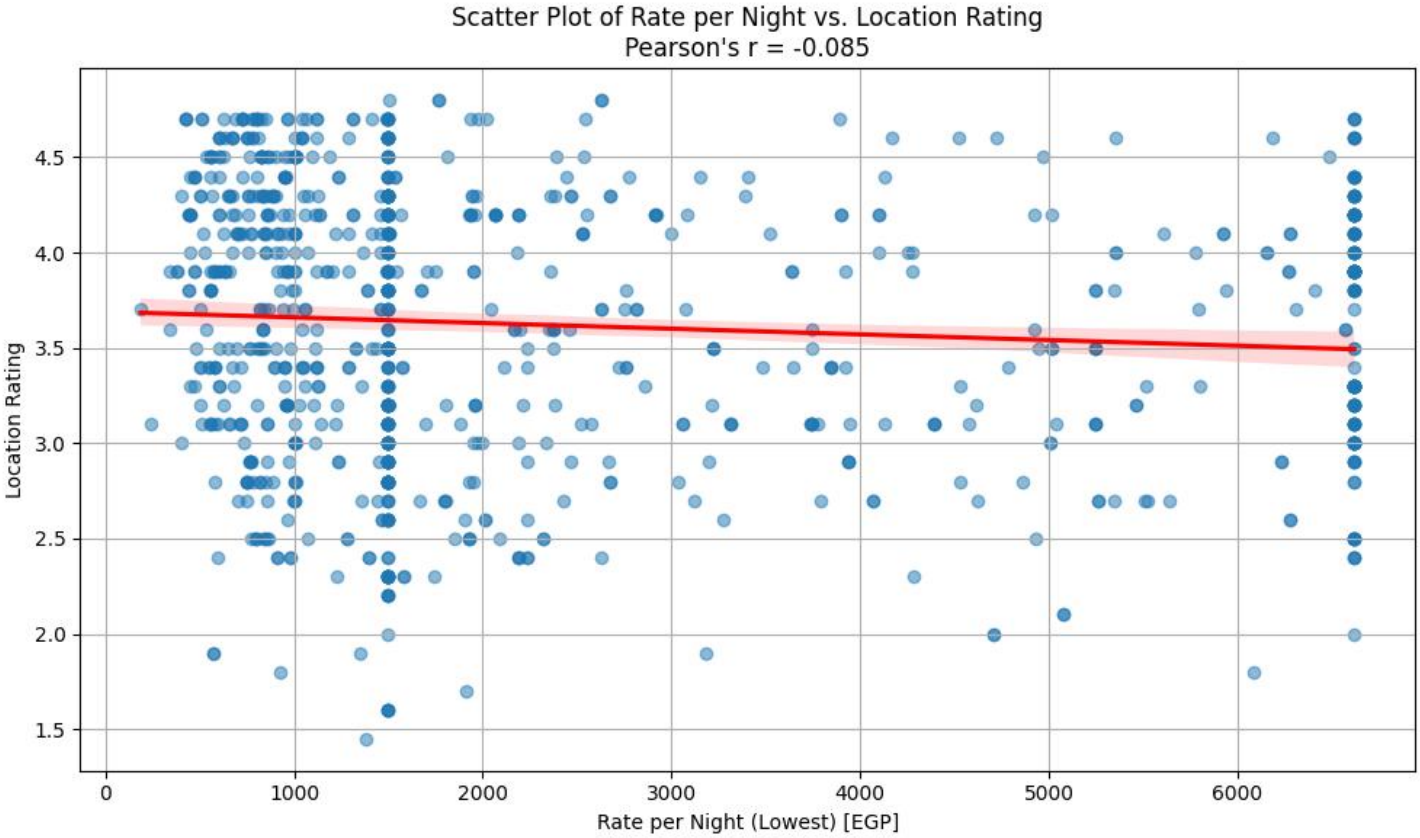
# B. Correlation Analysis:




## Rate per Night vs. Overall Rating:

Pearson's r = 0.373

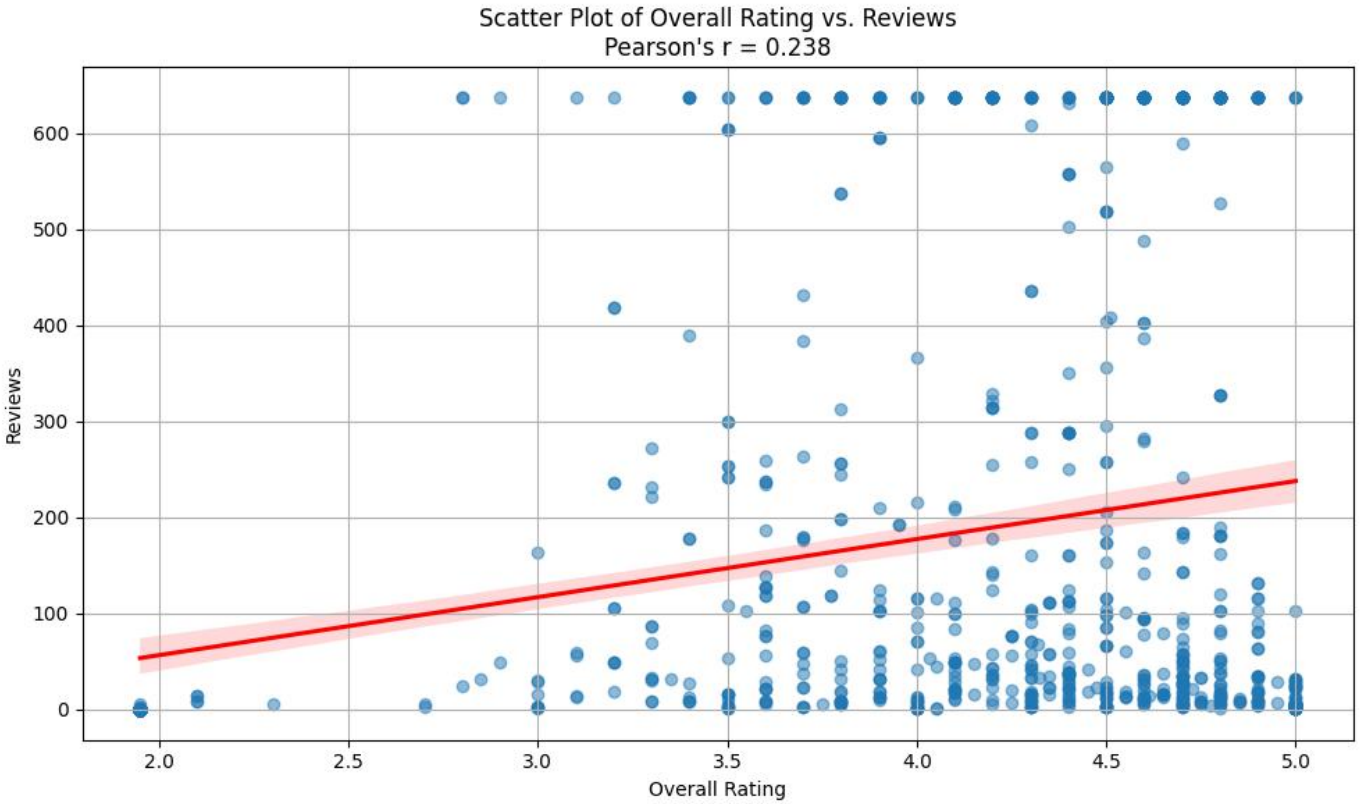
This moderate positive correlation indicates that as the nightly rate increases, the overall rating tends to improve slightly. This suggests that guests may associate higher prices with better service, amenities, or overall experience. However, the relationship is not extremely strong, meaning other factors beyond price also influence ratings.




## Rate per Night vs. Location Rating:

Pearson's r = -0.085

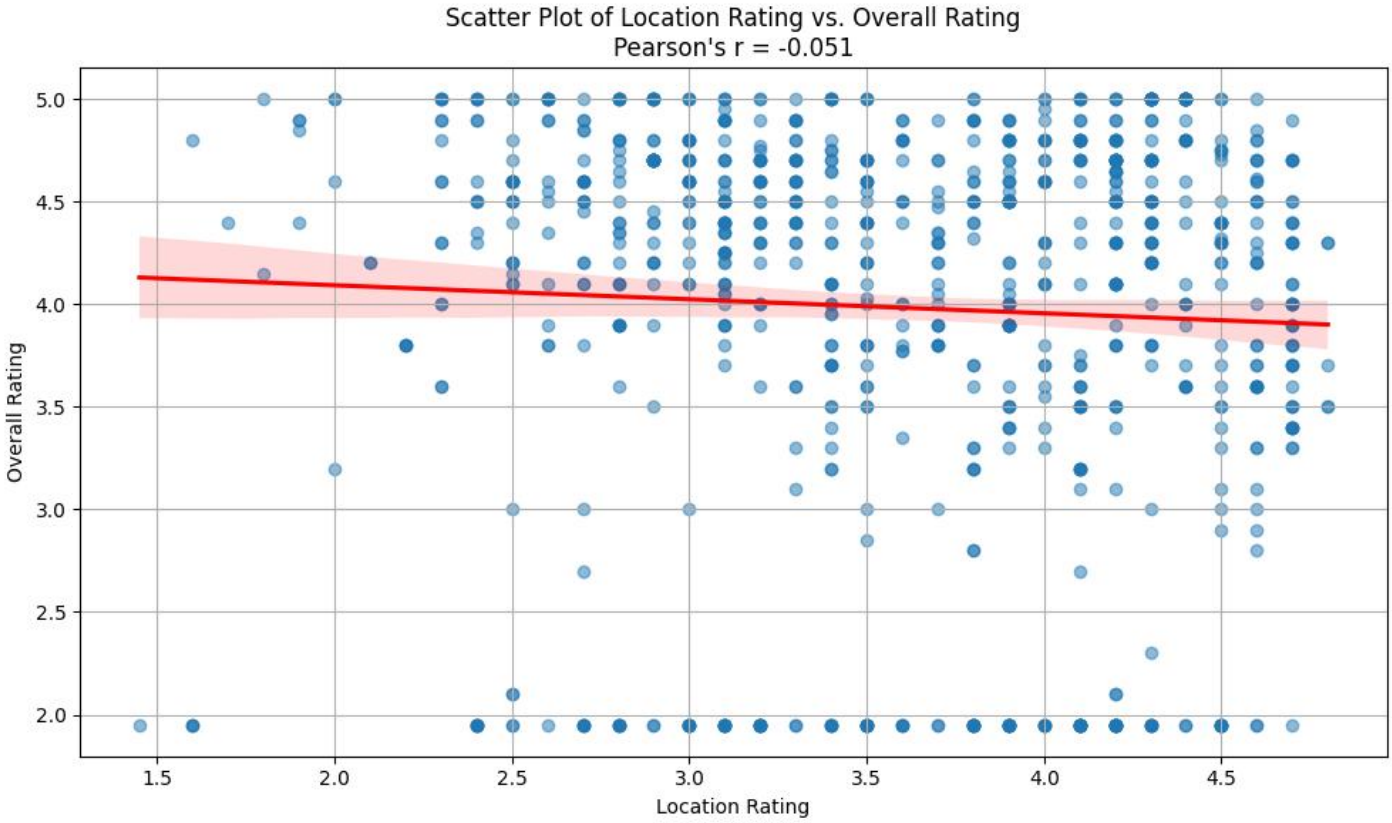
The weak negative correlation between location rating and nightly rate suggests that a hotel's perceived location quality has almost no linear influence on its pricing, indicating that other factors likely play a far more significant role in determining costs. This lack of meaningful correlation implies that travelers should not assume higher-rated locations command premium prices, nor should hotels prioritize location improvements solely for pricing leverage without considering broader competitive differentiators.




## Reviews vs. Overall Rating:

Pearson's r = 0.238

The correlation between Overall Rating and Reviews is weak (Pearson's r = 0.24), indicating no meaningful relationship between these variables. While there's a slight tendency for higher-rated businesses to have more reviews, the effect is statistically insignificant in practical terms. This suggests that improving ratings alone would likely have little to no impact on review volume, as other unmeasured factors dominate this relationship.




## Location Rating vs. Overall Rating:

Pearson's r = -0.051

The scatter plot shows a very weak negative correlation between Location Rating and Overall Rating, as indicated by Pearson's r of -0.051. This suggests that there is almost no linear relationship between the two variables, meaning changes in Location Rating do not predictably affect Overall Rating. The slight negative trend is negligible, implying that other factors likely play a more significant role in determining Overall Rating. Therefore, Location Rating alone is not a meaningful predictor of Overall Rating in this data-set.

# 08. HYPOTHESIS TESTING STEPS

## Review Count Group vs. Guest Satisfaction

### Step 1: Define null and alternative hypothesis:

Null Hypothesis ($H_0$): There is no difference in average overall ratings between different review count groups.

Alternative Hypothesis ($H_1$): There is a significant difference in average overall ratings between different review count groups.

### Step 2: Choose the appropriate test:

One-way ANOVA (comparing means across multiple review count groups).

### Step 3: Calculate the p-value:

F-statistic: 7.650567

P-value: 0.0005108351

### Step 4: Determine the statistical significance:

Since the p-value (0.0005108351) is **less than** 0.05, we **reject the null hypothesis**.

Conclusion: There is a significant difference in guest satisfaction between different review count groups.

# Location Group vs. Guest Satisfaction

## Step 1: Define null and alternative hypothesis:

Null Hypothesis ($H_0$): There is no difference in guest satisfaction scores between hotels in tourist cities and those in non-tourist cities.

Alternative Hypothesis ($H_1$): Hotels in tourist cities receive significantly higher guest satisfaction scores than those in non-tourist cities.

## Step 2: Choose the appropriate test:

One-way ANOVA (used here because you grouped by latitude ranges).

## Step 3: Calculate the test statistic

F-statistic: 1.18422693

P-value: 0.31613398

## Step 4: Determine the statistical significance:

Since the p-value (0.31613398) is **greater than** 0.05, we **fail to reject the null hypothesis**.

Conclusion: There is no significant difference in guest satisfaction across different location groups.

# Price per Night vs. Guest Satisfaction

## Step 1: Define null and alternative hypothesis:

Null Hypothesis ($H_0$): There is no correlation between the price per night and guest satisfaction scores.

Alternative Hypothesis ($H_1$): There is a significant correlation between the price per night and guest satisfaction scores (positive or negative).

## Step 2: Choose the appropriate test:

Pearson correlation (used to measure the strength and direction of a linear relationship between two continuous variables).

## Step 3: Calculate the p-value:

Correlation coefficient (r): 0.37299877176153806

P-value: ≈ 0.000000000000

## Step 4: Determine the statistical significance:

Since the p-value is less than 0.05, we **reject the null hypothesis**.

Conclusion: There is a significant correlation between price per night and guest satisfaction scores.

# 09. CONCLUSION

This analysis revealed key insights into guest satisfaction across Egyptian hotels, emphasizing that factors like service quality and value for money outweigh simple metrics like price or location. Statistical testing supported a moderate correlation between pricing and satisfaction but showed no significant advantage for hotels in tourist cities over non-tourist locations. These findings suggest that qualities such as amenities and service consistency maybe the primary drivers of guest experiences.

Strategic opportunities exist for enhancing offerings in underperforming regions and aligning prices with perceived value in vacation rentals. Addressing contextual gaps and accounting for confounding variables can further refine future analyses, enabling the hospitality sector to better meet evolving traveler expectations.

# 10. ANY POTENTIAL ISSUES

## 1. Confounding Variables:

Unmeasured factors such as global economic conditions, marketing efforts, or regional political stability could influence guest satisfaction and obscure the true impact of the analyzed variables.

## 2. Subjective Bias in Ratings:

Guest satisfaction scores are inherently subjective and may not consistently represent objective service quality, leading to potential variability in results.

## 3. Contextual Gaps:

External disruptions such as pandemics or political unrest may have influenced both guest satisfaction and data collection processes, potentially limiting the completeness and accuracy of the analysis.

## 4. Sampling Bias:

The data-set may overrepresent popular tourist destinations or highly reviewed properties, underestimating smaller or less visible hotels, which could skew findings.