



**IS5126**

**Hands-on with Applied Analytics**

**Assignment 1 Technical Report**

**Group 6:**

**AVIRAL GOYAL - A0330324W**

**CHEW JUN AN EDEN - A0202214A**

**DANIEL LIM WEI ZHONG - A0154795N**

**NG YU CHENG - A0297198E**

**XIAN JIANBO - A0110375R**

**Date of Submission:**

**22nd February 2026**

## Member Contribution Summary

Team Member	Contribution
Aviral Goyal	Worked on Data Preparation and Performance Profiling
Chew Jun An Eden	Worked on Exploratory Data Analysis and Benchmarking
Daniel Lim Wei Zhong	Worked on Benchmarking
Ng Yu Cheng	Worked on Benchmarking
Xian Jianbo	Worked on System Architecture and Dashboard

Github Repository Link:

<https://github.com/lucasnyc/is5126-assignment1/tree/main/student-name-hotel-analytics>

Base Database Link (Containing 71,000 records):

<https://drive.google.com/file/d/1VpYJQXLVGk-JgXVxlbWG7LA9LVHsKeRq/view?usp=sharing>

# 1. Executive Summary

## 1.1 Business Problem & Solution Overview

Hotel managers are often "data rich but insight poor," possessing thousands of reviews but lacking the ability to objectively benchmark performance against true competitors. HospitalityTech Solutions addressed this by developing an intelligent analytics platform that transforms raw data into strategic clarity. Our solution utilizes a high-performance SQLite backend and a K-Means clustering engine to segment hotels into multi-dimensional "experience profiles". This moves beyond arbitrary star ratings, allowing managers to identify specific service gaps and optimize resources based on how they perform relative to their actual market peers.

## 1.2 Key Findings

**High-Utility Data Filtering:** We optimized the dataset from 750,000 to 71,100 reviews by retaining only the top 20th percentile of "helpful" votes. This ensures the platform's insights are grounded in high-quality, seasoned traveler feedback rather than noise. **Three-Tier Market Segmentation:** Using the Elbow Method and Silhouette Scores, we identified three distinct clusters that mirror travel industry utility: "Bang-for-Buck" (Value/Location), "Location Trap" (Convenience/Budget), and "Premium Experience" (Elite/Luxury). **Strategic Benchmarking:** The platform features a predictive benchmarking engine that assigns hotels to these clusters and calculates performance deviations. Findings show that success is defined by outperforming the cluster average in six key drivers: Service, Cleanliness, Value, Location, Sleep Quality, and Rooms. **Operational Efficiency:** Through query profiling, we ensured the system processes complex multi-dimensional clusters with minimal latency, providing non-technical users with real-time, data-driven decision support.

# 2. Data Foundation

Since the objective of this assignment is to build an analytics platform that helps hotel managers understand customer satisfaction, benchmark competitors, predict trends and optimize operational decisions, the dataset must be reliable, relevant and structured for our scalable analysis. Thus, we focus on these aspects for our data preparation: logical filtering of data aligned with the business objectives, data validation and integrity checks, performance optimization through indexing and ensuring sufficient analytical volume.

## 2.1 Data Filtering Rationale

We performed data filtering based on business use of the analytics platform. We have retained reviews only from the recent 5 years which ensures that our analysis reflects the current service standards and market conditions. Hotel operations can change over time renovations, pricing adjustments , management changes, and service improvements. Thus, older reviews could distort competitive benchmarking by reflecting outdated performances. By restricting the dataset to recent reviews, the insights will be more relevant for present day decision making. As a result of this filtering the hotel managers will be able to get the insights based on current customer experiences and not based on some historical reputation.

Helpfulness-Based Quality Filtering

To enhance the analytical reliability, reviews were filtered using the 80th percentile of helpful votes rankings. The helpful votes indicate that the other users found the review informative or valuable. Reviews above this threshold are retained by us to prioritize detailed and socially validated feedback. Practically, reviews with higher helpful votes contain more specific comments about service quality, room conditions, cleanliness, or staff behaviour. Such details can be important for identifying customer satisfaction drivers and operational improvement areas. The 80th percentile was selected as a balanced threshold as it improves the overall signal quality while preserving a sufficient number of observations. Table 1 (refer to appendices) shows the impact of different percentile thresholds.

2.2 Schema Design and Relational Structure

The dataset was structured using a relational schema to ensure clarity, scalability, and efficient querying. At the core of the design is the Review table, which contains transaction-level information such as review ID, hotel (offering) ID, author ID, review text, title, rating attributes (overall, service, cleanliness, value, location, sleep quality, rooms), review date, and helpful votes. This table acts as the central fact table. It is linked to dimension-level entities such as Hotel (Offering) and Author, where offering\_id connects reviews to specific hotels and author\_id links reviews to individual reviewers. This separation avoids redundancy, improves data integrity, and supports flexible aggregation (e.g., hotel-level benchmarking, reviewer behavior analysis, or time-based trend analysis). From a business perspective, this relational structure enables multi-level analytics required by HospitalityTech Solutions. For example, hotel managers can analyze performance at the property level, compare attributes across hotel groups, or examine patterns in customer segments. The schema supports benchmarking (hotel vs. competitors), satisfaction driver analysis (attribute ratings vs. overall score), and trend forecasting (using review dates). By organizing the data into a structured relational format, we ensure that the analytics platform remains scalable, logically consistent, and aligned with real-world business reporting requirements.

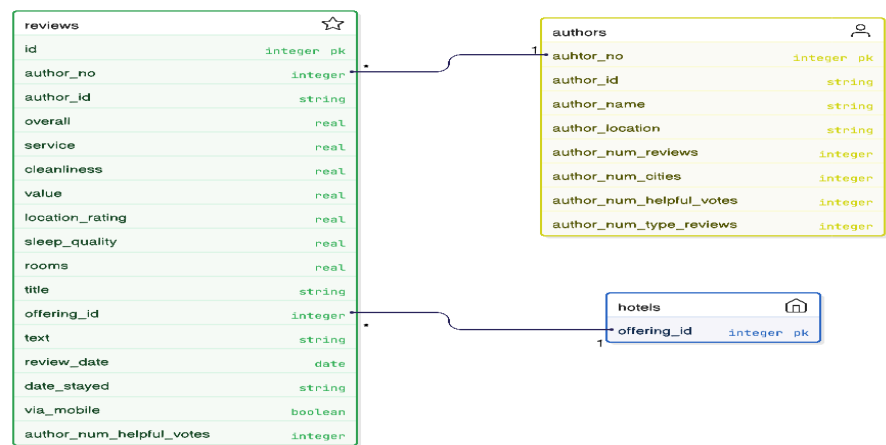


Figure 1: ER Diagram of Review Database

## 2.3 Indexing Strategy

To improve the query performance and ensure efficient analytical processing, we implemented an indexing strategy on key columns frequently used in filtering, grouping, and joining operations. Many of the business questions involve hotel-level benchmarking and competitor comparisons; indexing on `offering_id` will improve aggregation performance when calculating average ratings, review counts, and attribute-level summaries. Indexing `review_date` supports time-based trend analysis and forecasting, while indexing `author_no` will allow reviewer-level filtering and potential credibility analysis. This indexing strategy can ensure that the analytics platform can deliver insights quickly and reliably, even when handling tens of thousands of reviews. Faster query execution is important for interactive dashboards and benchmarking tools, where hotel managers expect near real-time performance comparisons. By strategically indexing high-usage columns, we reduce computational overhead, improve scalability, and enhance the overall responsiveness of the decision-support system.

## 3. Exploratory Data Analysis

### 3.1 Overview of Dataset

Exploratory analysis was conducted on a clean review dataset comprising approximately 754,000 reviews across 3,888 hotels spanning the period of December 2007 to December 2012. Each review contains an overall rating (1-5 scale), aspect-specific ratings (service, cleanliness, value, location, sleep quality, rooms), textual feedback, review date, and metadata such as mobile submission status. To ensure analytical reliability, only reviews with valid overall ratings (1–5) and non-null dates were included in rating-based analyses. Aspect ratings were allowed to be missing, reflecting platform behavior where not all reviewers rate every dimension.

### 3.2 Distribution of Ratings and General Performance Patterns

#### **Overall Rating Distribution**

The overall ratings are concentrated between 3.5 and 4.5, with a mean close to 4.0. This indicates a generally positive skew in guest satisfaction, a common phenomenon in hospitality review platforms due to self-selection bias (guests with extreme positive or negative experiences are more likely to leave reviews). Because most hotels cluster within a narrow rating band, small differences can influence competitive positioning. Managers cannot rely solely on overall rating differences; deeper aspect-level analysis is necessary to uncover meaningful differentiation.

#### **Aspect Rating Averages**

Location consistently receives the highest average score, while value perception is the lowest-rated dimension. Location appears to be a structural strength across many properties, likely reflecting geographical advantages. However, value perception emerges as the most vulnerable dimension. Even well-rated hotels risk competitive disadvantage if guests perceive price-performance imbalance. Pricing strategy and value communication therefore become critical managerial levers.

### 3.3 Correlation Analysis: Drivers of Satisfaction

Service and cleanliness are almost perfectly correlated with overall satisfaction, whereas location shows a substantially weaker relationship. Operational excellence (service and cleanliness) drives overall satisfaction most strongly. Location, while important, is less influential in shaping the final overall perception. Value perception acts as a mediating variable between operational delivery and price fairness. Investments in operational quality (staff training, housekeeping processes) are more likely to yield improvements in overall ratings than investments in location-related marketing, which is structurally fixed. For underperforming hotels, improving service quality may generate greater return than attempting to reposition based on location.

### 3.4 Mobile vs Non-Mobile Reviews

Mobile reviews had slightly lower average ratings. The difference, while not extreme, was directionally consistent. Monitoring mobile reviews can act as an early warning indicator of service breakdowns. If mobile ratings decline before overall trends shift, it may signal emerging dissatisfaction.

### 3.5 Top and Bottom Performing Hotels

The EDA identified top-performing and bottom-performing hotels based on average overall ratings and sufficient review volume thresholds. Top performers differentiate through operational strength, not just favorable location. Bottom performers often show structural operational deficiencies. Competitive benchmarking should therefore focus on comparing hotels within similar performance bands rather than across extreme tiers.

### 3.6 Implications for Segmentation and Competitive Benchmarking

The EDA revealed crucial structural insights that directly motivated clustering - Hotels differ not only in rating levels but in rating profiles. Value perception behaves differently from service and cleanliness. These findings justify **the use of multivariate clustering to identify experience-based hotel archetypes. Clustering on aggregated aspect ratings and consistency allows hotels to be grouped by underlying experience structure** rather than superficial similarity (e.g., star rating or geography).

### 3.7 Summary of Key Business Insights

The exploratory analysis yields the following actionable insights. Operational quality (service and cleanliness) is the strongest driver of overall satisfaction. Value perception is the most sensitive and differentiating dimension. Location alone does not guarantee high satisfaction. Small rating differences are strategically meaningful due to compressed rating distribution. Mobile reviews may provide early dissatisfaction signals. Collectively, these findings establish the foundation for experience-based segmentation and cluster-aware competitive benchmarking in subsequent analysis.

## 4. Performance Profiling & Optimization

To evaluate the scalability and computational efficiency of the proposed analytics platform, we conducted performance profiling. The objective of the profiling is to measure execution time, query behaviour, cluster scalability, and memory consumption before and after optimization.

### 4.1 Query Profiling

We have done database-level profiling with focus on aggregation and join operations used in the competitive benchmarking dashboard. The primary aggregation query computed average rating ratings grouped by hotel (offering\_id). From Table 2 we can observe that the aggregation query became slower after indexing. This occurs because this operation requires complete table scan, and for a relatively small dataset, index traversal introduces additional overhead. This demonstrates that indexing does not always improve the performance and must be evaluated based on query type dataset size. In contrast to this, the join operations show slight improvement after indexing. This occurs because join operation relies on key-based hookups, and thus, indexing the join attribute improves retrieval efficiency. Overall, indexing improved beneficial for relational joins but not for aggregation-heavy queries in this context.

### 4.2 Code Profiling

We have conducted code profiling using python's cProfile module to measure computational cost within the analytics pipeline implemented using pandas and scikit-learn. From Table 3 we can observe that pandas aggregation is executed very efficiently due to vectorized C-level operations, confirming that data summarization is not a performance bottleneck. While K-Means Clustering required significantly more time and function calls, as most execution occurred within iterative centroid update routines. This confirms that the machine learning layer is the dominant computational cost in the system.

### 4.3 Overall Interpretation

The profiling results indicate that:

- Aggregation operations are highly optimized at both database and application levels.
- Join queries benefit from indexing
- Clustering is the most computationally intensive component
- Total execution time remain below 0.25 sec, indicating strong system responsiveness

Thus, the system demonstrates efficient performance at the current dataset scale, with scalability expected to depend primarily on clustering complexity and data volume growth.

## 5. Competitive Benchmarking Strategy

### 5.1 Methodology: Multi-Dimensional Experience Clustering

The fundamental challenge in hotel management is the "comparability paradox." Traditional benchmarking often relies on rigid, industry-standard labels like star ratings or geographic proximity. However, these labels fail to capture the nuanced expectations of different traveler

demographics. A business traveler staying in a city center hotel has vastly different priorities compared to a family on a budget. To solve this, our methodology employs **K-Means Clustering** to group hotels based on their "Experience Profiles" rather than their physical characteristics. We use six core performance dimensions (Service, Cleanliness, Value, Location, Sleep Quality and Rooms) as the feature set for our model. The justification for this approach is rooted in the reality of behavioral segmentation. Guests do not just book rooms, they book experiences. By clustering based on actual ratings, we group hotels that satisfy similar guest needs, such as those that prioritize a prime location over room luxury. Furthermore, using six features prevents a "single-metric bias," allowing for a holistic evaluation that accounts for the trade-offs guests naturally make, such as accepting lower "Value" scores in exchange for exceptional "Service" and "Rooms". Ultimately, this data-driven approach allows hotels to find their "true peers", properties that guests perceive as comparable, regardless of whether they share the same brand or star rating.

## 5.2 Selecting the Optimal Number of Clusters

Determining the optimal number of clusters was a critical step in ensuring the model's business utility. We evaluated K-Means models across a range of 2 to 10 clusters, plotting the Within-Cluster Sum of Squares (WCSS) for the Elbow Method against the Silhouette Score (Figure 2) to identify the most distinct market segments.

While initial visual inspection suggested that both 3 and 4 clusters were potentially strong fits, we computed detailed profiles for both configurations (Tables 4 & 5) to assess their distinctness.

Our review of the 4-cluster configuration revealed that this higher value introduced significant "noise"; specifically, Clusters 1 and 3 exhibited nearly identical performance levels, scoring consistently in the high 3.0 to 4.5 range across most categories. This overlap suggested redundant categorization rather than truly distinct market divisions.

Beyond statistical performance, we prioritized business utility. The 3-cluster model provides a much clearer and more interpretable segmentation that aligns with established travel industry tiers: Economy, Business, and First Class. Because guest expectations are frequently framed by these three broad levels of service and quality, a three-tier model ensures the platform's outputs are both statistically sound and intuitive for hotel managers to use in strategic planning.

## 5.3 Performance Analysis Across Hotel Groups

Our analysis identified three distinct market segments (Table 6) based on guest "experience profiles," moving beyond generic industry labels to establish mean performance baselines for each group. This multi-dimensional benchmarking allows managers to compare their properties against "true peers" rather than arbitrary standards.



## 5.4 Identification of Best Practices and Actionable Recommendations

The identification of best practices and actionable recommendations begins by passing a hotel's specific rating profile through the established K-Means model to objectively identify its appropriate peer cluster. This methodology ensures that a property is not measured against a generic industry standard, but is instead benchmarked against "true peers" that share a similar multi-dimensional experience profile across service, cleanliness, value, location, sleep quality, and room standards. Once the peer group is assigned, the hotel's individual ratings are compared directly to the average scores of that specific cluster to identify where the property sits relative to the group baseline. The resulting performance gap is calculated as a percentage deviation from the peer average, providing a clear metric for interpreting the hotel's standing. A negative percentage indicates underperformance and highlights critical operational gaps. Conversely, a positive percentage signals outperformance, identifying unique competitive strengths that can be leveraged for marketing or brand growth. This systematic comparison allows managers to move beyond broad dissatisfaction and focus their limited improvement budgets on the specific drivers that will most effectively close the gap with their competitors or facilitate migration toward a higher performance tier like the "Premium Experience" cluster.

## 6. System Architecture & Dashboard

The hotel management dashboard is built around three strategic pillars. First, it delivers a statistics-based bird-eye overview that allows immediate situational awareness of the health of the hotel operation. Managers can instantly identify key strengths and weaknesses without spending a tremendous amount of time extracting and consolidating data from various sources, providing a concise and holistic performance snapshot. Second, the dashboard integrates competitive benchmarking to position the hotel within the entire market relative to its hotel peers or competitors. This allows managers to understand where the property stands, make informed decisions on commercial focuses, plan investments efficiently, and devise targeted service improvement strategies. Finally, it generates data-driven recommendations by translating guest reviews and ranking metrics into actionable items of improvement plan. These recommendations offer fresh and valuable perspectives that might otherwise be missed out, supporting more informed and proactive management decisions.

### 6.1 My Hotel Guest Experience Dashboard

The dashboard connects to a SQLite reviews database and supports analysis based on guest experience on a single hotel. A global month range time filter and optional hotel filter control output displayed, with instant validation for invalid IDs or reversed date ranges. The "Overall Rating Distribution" tab (Fig 2) shows a performance snapshot through core metrics—total reviews, average rating, and the share of high ratings—alongside a 1–5 rating distribution chart, percentage breakdown table, and time-series trends for both average rating and review volume of a hotel. The "Top and Bottom Reviews" tab (Fig 3) demonstrates top k highest- and lowest-rated reviews. Each entry includes scores in different aspects. At the bottom of the page, an expandable section is included to show full texts with reviewer profile information. The "Aspect Contributor Over Time" tab (Fig 4) plots key aspect scores (Service, Cleanliness, Value,

Location, Sleep Quality, and Rooms) over a configurable time granularity. It combines multi-line trend charts, review-volume visuals, per-aspect mini trends, and a summary statistics table for the filtered period. The Textual Feedback Analytics tab applies NLP (Fig 5) to process textual contents such as titles and content. Keywords with highest influence are extracted and separated into the positive sentiment table and the negative sentiment table. The Top Reviewer Details tab (Fig 5) examines the highest-rated reviews to profile influential contributors. It presents subset KPIs, a detailed review table with reviewer statistics, location distribution, reviewer-experience histogram, and expandable full review texts for context.

## 6.2 My Market Position Dashboard

This dashboard is an analytics page that allows hotel operators to track rating trends, benchmark against competitors in markets, monitor the evolution of key service aspects and to position his hotel. This dashboard is organised into 4 features, it provides flexible filtering by hotel and time period so users can switch between a market-wide view and a property-specific analysis. The “Overall Rating” tab (Fig 7) visualises guest satisfaction over time through a two-panel chart with a configurable time granularity. The market average is always displayed, with an optional overlay for a selected hotel to allow direct performance comparison of the selected hotel against the market average. The “Position Among Top and Bottom Performers” tab (Fig 8) benchmarks a selected hotel against the top k highest- and lowest-rated hotels within a defined date range, using ranked bar charts, an 6 aspect-dimension comparison, and a service gap-analysis table to identify competitive strengths and rooms for improvement. The “Aspect Contribution Over Time” tab (Fig 9) plots the aspect scores for the market and for a selected hotel using the same dual-line comparison approach. It is supported by review-volume trends, collapsible per-aspect mini charts, and a summary statistics table for the selected scope. Together, these views deliver a concise market perspective and a clear operational diagnostic to support data-driven decision-making. The “Hotel Cluster Classification” tab (Fig 10) assigns a hotel to our self-defined categories (1.Bang for Buck; 2.Location Trap; 3.Premium Experience) using a pre-trained K-Means model based on its average aspect ratings. After a Hotel ID is keyed in, the dashboard will display the review count, a colour-coded cluster label, and compare the hotel’s aspect scores with its peers under the same cluster, highlighting percentage gaps and over- or under-performance. A scatter plot shows the hotel’s positioning in each aspect relative to other peers in the cluster.

## 6.3 My Hotel Improvement Plan Dashboard:

This Hotel Improvement Consultant Function (Fig11) uses a SQLite review database and local open-source LLMs (e.g.TinyLlama) to generate improvement plans based on reviews in the worst performing aspect. A dropdown box enables users to choose models. Upon entering a Hotel ID, the system analyzes ratings across the 6 aspects to isolate the property’s Weakest Aspect. To ensure recommendations are provided based on actual worse reviews, the dashboard filters and displays the lowest-scoring guest reviews for that specific category. This worst review based approach provides management with direct qualitative suggestions for improvement which may be missed out and an alternative solution to tackle the worse aspect of the hotel.

## 7. Conclusion

### 7.1 Summary of Deliverables

The analytics platform optimized approximately 750,000 to 71,000 reviews to aid hotel managers in reviewing their competitive position. In contrast to traditional benchmarking based on star ratings, our platform uses a K-Means clustering to group hotels by “experience profile”. This segregates comparison based on clusters formed, ensuring a high end hotel is compared to similar profiles instead of budget hotels.

### 7.2 Key Findings

From the high correlation of Service and Cleanliness to Satisfaction, we identified these two as primary drivers across three clusters identified. Cleanliness impact on Satisfaction is also dependent on the cluster. If a hotel belonging to the “Premium Experience” cluster falls short by 10.1% in Cleanliness, it risks losing its competitive edge regardless of other drivers. On the other hand, the “Location Trap” segment’s edge is in the prime location but low averages scores for Rooms(1.79) and Cleanliness(2.02) limits the Satisfaction ceiling. For these hotels, data suggest hotel managers direct funding to housekeeping rather than room upgrades as it is deemed more effective to improve guest’s experience. From a technical view, query performance is critical to a system's reliability. We implemented targeted indexing and optimizing the data retrieval logic. Ensuring the dashboard can process multi-dimensional cluster calculations with minimal latency.

### 7.3 Limitations and Future Additions

Current model relies on a static dataset of reviews and factors such as seasonality or recent hotel enhancements are not considered, to further improve the platform’s insights for hotel managers, Natural Language Processing can be employed to analyze the text of reviews. This would give us further insights on the numbers, to be more specifically categorizing low service scores causality. A general low Service Score can be pinpointed whether it is from slow check in time or negative customer experience.

## 8. Appendix

Percentile	Threshold (number of helpful votes)	Rows left after filtering
50	7	181737
55	9	156103
60	10	144946
65	12	126416
70	15	104526
75	18	87607
<b>80</b>	<b>22</b>	<b>71100</b>
85	28	53163
90	38	35219
95	61	17334

Table 1: Different Percentile Thresholds with Resulting Data Count

Operation	Before Index (s)	After Index (s)
Aggregation	0.0889	0.1977
Join	0.1663	0.1626

Table 2: Query Execution Time for Aggregation and Join Operations

Component	Execution Time (s)	Function Calls
Aggregation	0.017	2243
Join	0.218	53699

Table 3: Code Profiling Execution Time and Function Calls

Cluster	Service	Cleanliness	Value	Location Rating	Sleep Quality	Rooms
0	3.58	3.63	3.56	3.81	3.44	3.29
1	2.16	2.02	2.14	2.82	2.00	1.79

2	4.31	4.45	4.16	4.37	4.23	4.16
---	------	------	------	------	------	------

Table 4: Average rating scores for 3 clusters

Cluster	Service	Cleanliness	Value	Location Rating	Sleep Quality	Rooms
0	3.12	3.12	3.20	3.52	2.96	2.82
1	4.41	4.55	4.24	4.44	4.33	4.29
2	1.85	1.65	1.76	2.59	1.72	1.49
3	3.89	4.00	3.80	4.04	3.78	3.63

Table 5: Average rating scores for 4 clusters

Cluster	Category	Remarks
0	Bang-for-Buck (Mid-Tier Practical)	This segment represents the "Practical Traveler" who prioritizes location (3.81) over luxury. These hotels score consistently in the mid-to-high 3s, offering a reliable experience where guests accept moderate room quality (3.29) for a fair price.
1	Location Trap (Budget/Convenience)	Categorized as the "Functional/Distressed" segment, these hotels suffer from significant failures in rooms (1.79) and cleanliness (2.02). Their survival is dependent on a relatively higher location score (2.82), serving as a necessity for budget-conscious travelers near specific transit hubs.
2	Premium Experience (Luxury/Elite)	This "Experience Seeker" segment maintains exceptional consistency, with all scores exceeding 4.1. Cleanliness (4.45) and service (4.31) are the non-negotiable benchmarks for this group, where guests pay a premium for high expectations and brand loyalty.

Table 6: Identified customer segments for k=3 clusters

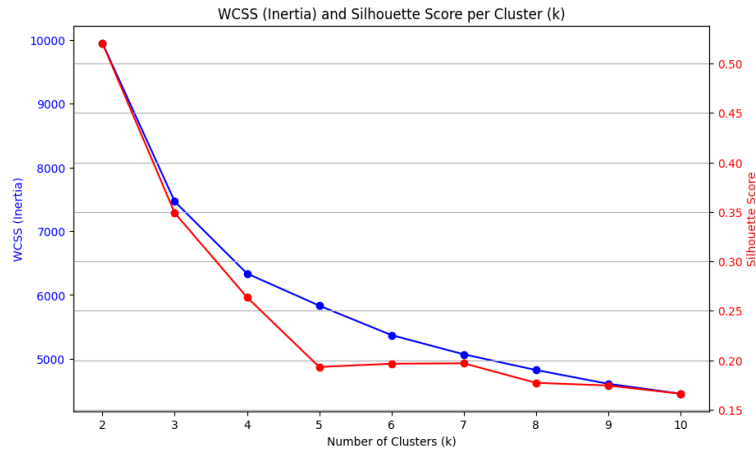


Figure 2: WCSS and Silhouette score per Cluster

## My Hotel Guest Experience Dashboard

Performance snapshot per hotel, aspect activity trends over time, and holistic text insights from guest reviews.

Filter: My Offering ID (Hotel ID)

Filter: Month-Year Date Range

Offering ID

From (Month-Year)

To (Month-Year)

93466

2010-02

2012-12

91 reviews for hotel 93466

Fig1. Hotel Filter and Date Range Filter

# Overall Rating Distribution

Total Reviews	Avg Overall Rating	High Ratings (4-5 stars)
91	3.52	56.0%

## Rating Distribution

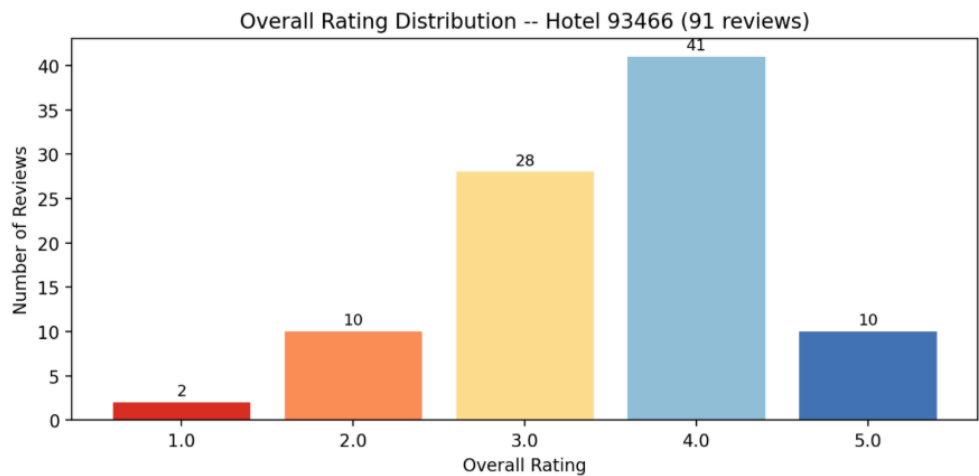


Fig 2. My Hotel Guest Experience Dashboard – Overall Rating Distribution Feature By Hotel ID

## Top and Bottom Reviews

Reviews to display per group

10

### Top Reviews (highest overall rating)

	Hotel ID	Date	Overall	Service	Cleanliness
0	93466	2011-01-05 00:00:00	5	5	5
1	93466	2011-03-27 00:00:00	5	4	5
2	93466	2011-10-14 00:00:00	5	5	5
3	93466	2012-04-25 00:00:00	5	5	5
4	93466	2012-04-27 00:00:00	5	5	4
5	93466	2012-06-03 00:00:00	5	5	4
6	93466	2012-06-24 00:00:00	5	5	5
7	93466	2012-08-07 00:00:00	5	5	5
8	93466	2012-08-20 00:00:00	5	4	5
9	93466	2012-09-01 00:00:00	5	5	5

### Bottom Reviews (lowest overall rating)

	Hotel ID	Date	Overall	Service	Cleanliness
0	93466	2012-04-09 00:00:00	1	1	3
1	93466	2012-06-18 00:00:00	1	1	1
2	93466	2010-05-31 00:00:00	2	2	4
3	93466	2010-11-21 00:00:00	2	1	2
4	93466	2011-08-30 00:00:00	2	4	2
5	93466	2011-10-27 00:00:00	2	3	3
6	93466	2011-12-12 00:00:00	2	3	3
7	93466	2011-12-28 00:00:00	2	1	2
8	93466	2012-04-26 00:00:00	2	1	4
9	93466	2012-05-25 00:00:00	2	3	3

Fig 3. My Hotel Guest Experience Dashboard – Top and Bottom Reviews by Hotel ID

## Aspect Contribution Over Time

Track how aspect scores — Service, Cleanliness, Value, Location, Sleep Quality, and Rooms — evolve over time for a specific hotel or across all hotels.

Time granularity

☒ Monthly ☐ Quarterly ☐ Yearly

### Aspect Scores Over Time

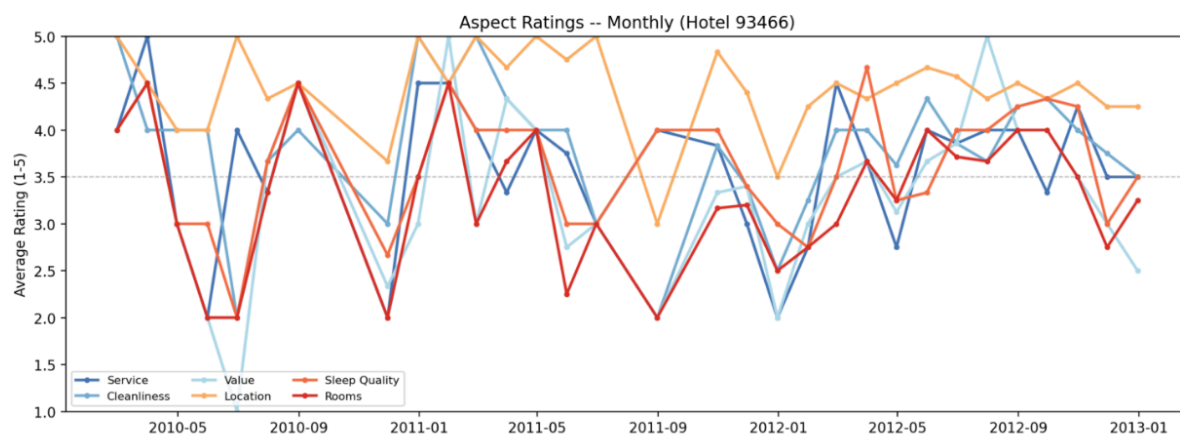




Fig 4. My Hotel Guest Experience Dashboard – Aspect Contribution Over Time by Hotel ID

Keyword Themes by Sentiment (Review Text)

Top Positive Keywords (4-5 stars, 51 reviews)

Top Negative Keywords (1-2 stars, 12 reviews)

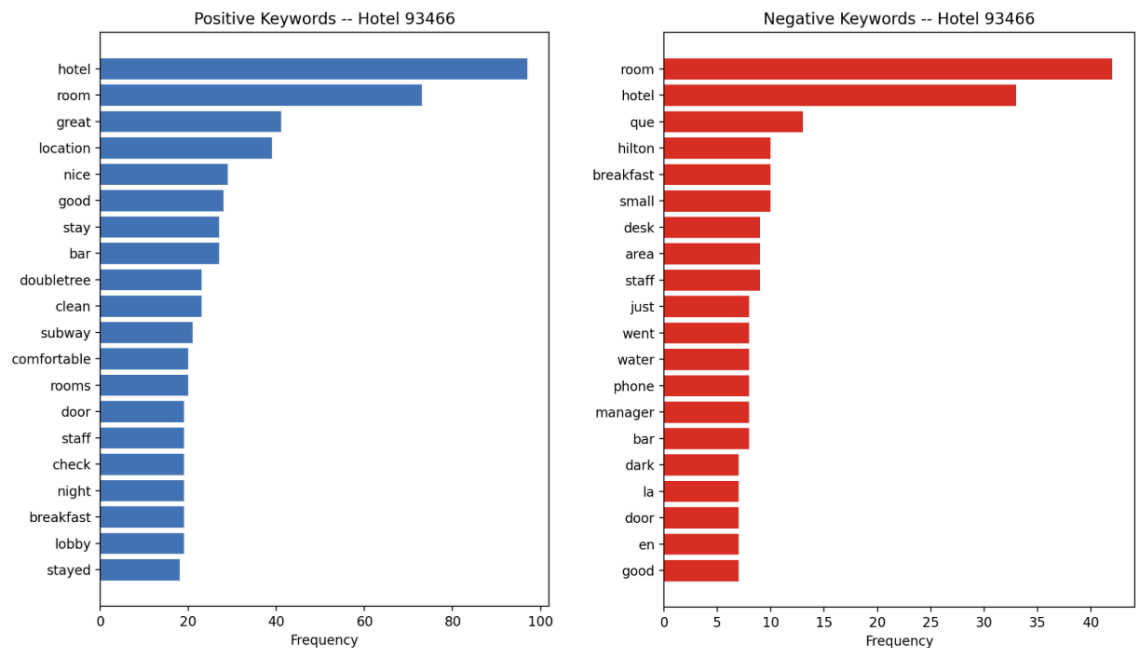
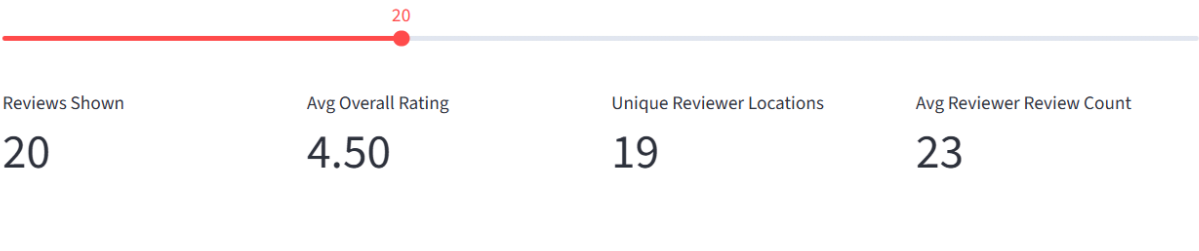


Fig 5. My Hotel Guest Experience Dashboard – Textual Feedback Analytics

## Top Reviewer Details

Detailed view of the highest-rated reviews, enriched with reviewer location and contribution statistics.

Number of top reviews to show



## Review Details

	Hotel ID	Date	Overall	Title	Reviewer Name	Reviewer Location
0	93466	2011-01-05 00:00:00	5	"Buen precio por un excelente hotel"	HumbertoMexico	Toluca, Mexico
1	93466	2011-03-27 00:00:00	5	"Great location and Fabulous value"	bowerbirdschest	Melbourne
2	93466	2011-10-14 00:00:00	5	"Excellent customer service!!! Quality hotel"	NickWelford	Middlesbrough, United Kingdom
3	93466	2012-04-25 00:00:00	5	"Great Location and Deal"	scmcb	Fairfax, VA
4	93466	2012-04-27 00:00:00	5	"Great location, awesome staff"	James W	Johns Creek, Georgia

Fig 6. My Hotel Guest Experience Dashboard – Details of Top reviewers

## Overall Rating Trend

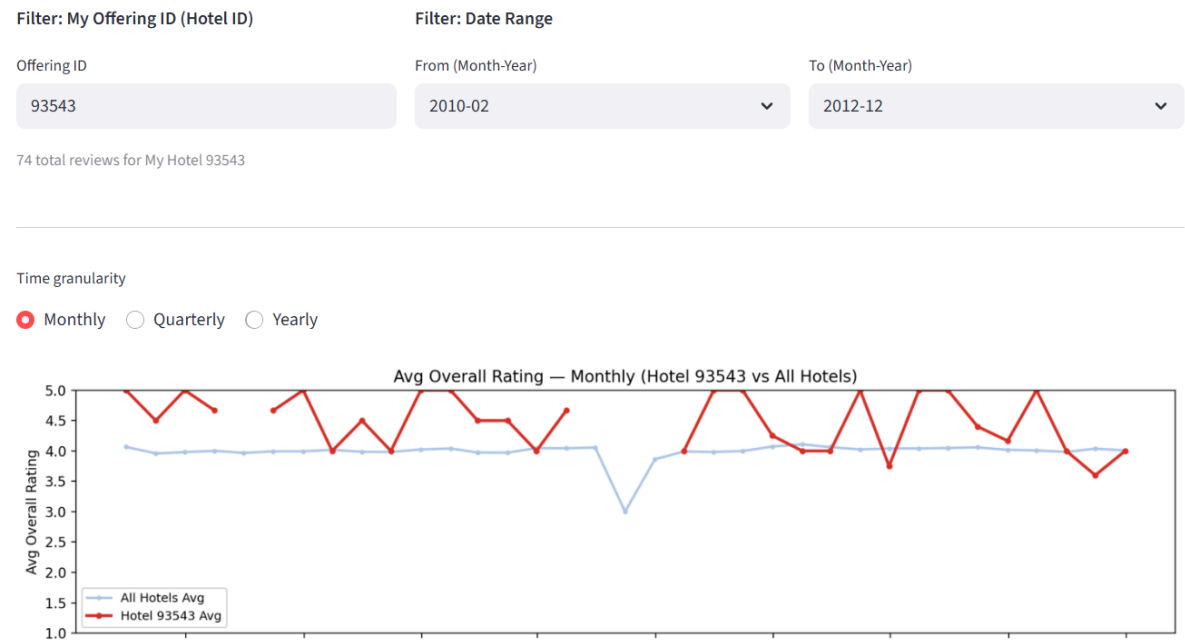


Fig 7. My Market Position Dashboard – My Hotel Overall Rating Trend

## Top & Bottom Performing Hotels

Filter: My Offering ID (Hotel ID)

Filter: Date Range

Offering ID

From (Month-Year)

To (Month-Year)

93466

2010-02



2012-12



91 total reviews for Hotel 93466

Number of hotels to show

10

### Top Performers

### Bottom Performers

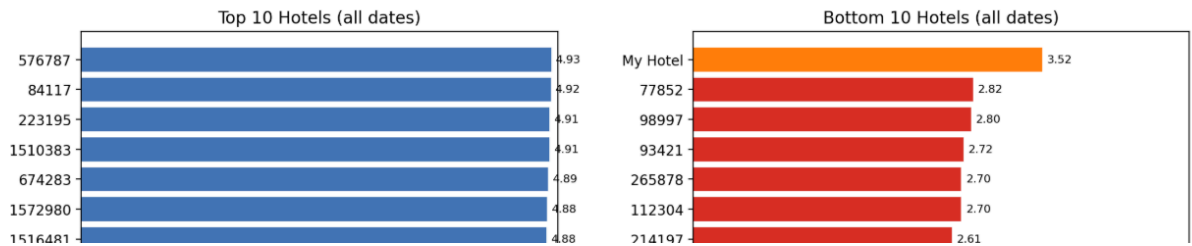


Fig 8. My Market Position Dashboard – My hotel Position by Average Overall Ranking

## Aspect Scores Over Time

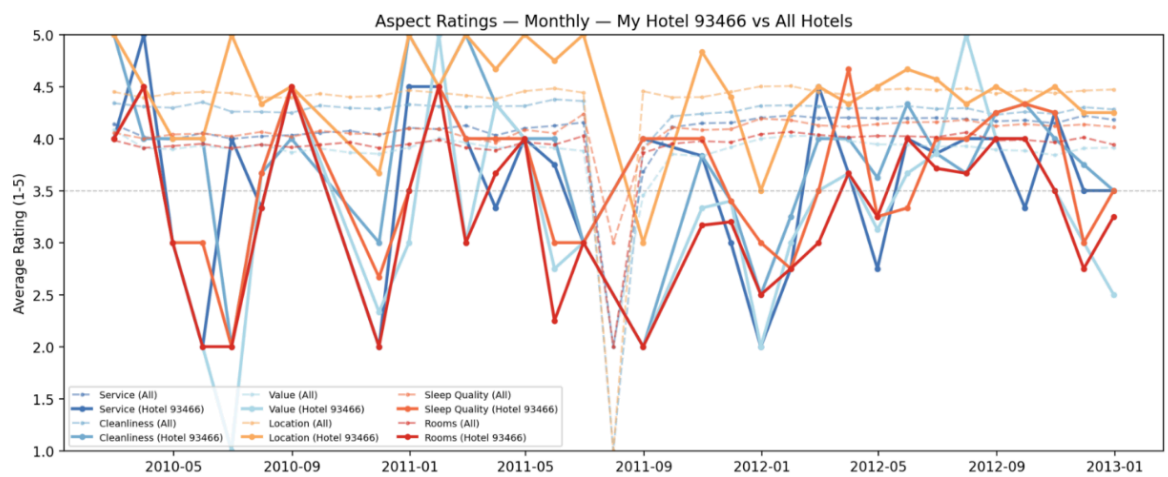


Fig 9. My Market Position Dashboard – Aspect Scores over Time Vs All Hotel Statistics

# Hotel Cluster Classification

Classify a hotel into one of three market segments based on its averaged aspect ratings. Segments are derived from a K-Means model trained on all hotels with  $\geq 50$  reviews.

Filter: My Offering ID (Hotel ID)

Offering ID

624442

16 total reviews for Hotel 624442

Cluster Segments

• Bang for Buck — Strong value-for-money; guests appreciate affordability relative to quality.

• Location Trap — Prime location drives ratings, but service and comfort lag behind.

• Premium Experience — Top-tier across all aspects; guests pay for and receive excellence.

Hotel 624442 → Premium Experience

Top-tier across all aspects; guests pay for and receive excellence.

Fig 10. My Market Position Dashboard – Hotel Cluster Classification

Hotel Info Based Bivariate Analysis

Review Info Based Bivariate Analysis

Time Series Analysis

Multivariate Analysis

Correlation Heatmap

Intelligent Analytics Platform

My Hotel Guest Experience Dashboard

My Hotel Market Position Dashboard

My Hotel Improvement Plan Dashboard

My Hotel Investment Return Prediction ...

View less

Model Configuration

Model

TinyLlama/TinyLlama-1.1B-Cha...

Max new tokens

256

Number of weak reviews to analyse

5

AI-Generated Improvement Plan

Generate Improvement Plan

Improvement Plan for Hotel 93466 Rooms

Room Quality Analysis: The hotel has made efforts to improve room quality, but it falls short when compared to similar establishments. A number of customers have reported dark and musty smelling rooms and walls covered in yellow stains along with tattered mattresses covered in torn sheets. The decor is also lackluster and does not match the overall ambience of the property. Some guests expressed dissatisfaction with the lack of amenities such as tea and coffee making facilities, making it difficult for them to conduct work meetings or study. The staff, while helpful, can be unhelpful and unresponsive.

Quick Wins (within 1 month): a) Increase the frequency of housekeeping services. b) Add additional amenities such as air conditioning, Wi-Fi, and telephones in all rooms. c) Upgrade the cleanliness standards of public areas and common spaces.

Medium-Term Improvements (1-6 months): a) Conduct thorough reviews of room amenities and offer additional perks such as free bottled water, complimentary newspapers, and complimentary parking vouchers. b) Im

Download Plan (txt)

Download Reviews Used (txt)

> View prompt sent to model

Fig 11. My Hotel Improvement Plan Dashboard – AI-Generated Improvement Plan