



## **SERVICE FAILURES AND PERCEIVED FAIRNESS IN MALAYSIAN HOTELS- EVIDENCE FROM LOW-RATING CHINESE-LANGUAGE REVIEWS ON CTRIP AND AN ACTIONABLE RECOVERY PLAYBOOK**

Zhang Juan<sup>1</sup>, Choo Wei Chong<sup>2\*</sup>, Yee Choy Leong<sup>3</sup>, Lin Yihuan<sup>4</sup>

<sup>1</sup> Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia

 [gs66172@student.upm.edu.my](mailto:gs66172@student.upm.edu.my)

 <https://orcid.org/0009-0008-1774-9479>

<sup>2</sup> School of Business and Economics, Universiti Putra Malaysia, Malaysia

 [wchoo@upm.edu.my](mailto:wchoo@upm.edu.my)

 <https://orcid.org/0000-0002-5797-7707>

<sup>3</sup> School of Business and Economics, Universiti Putra Malaysia, Malaysia

 [fyee@upm.edu.my](mailto:fyee@upm.edu.my)

 <https://orcid.org/0000-0002-3309-1020>

<sup>4</sup> School of Business and Economics, Universiti Putra Malaysia, Malaysia

 [gs58501@student.upm.edu.my](mailto:gs58501@student.upm.edu.my)

 <https://orcid.org/0000-0002-8391-7732>

\*Corresponding Author

### **Article Info:**

#### **Article history:**

Received date: 12.01.2026

Revised date: 02.02.2026

Accepted date: 03.03.2026

Published date: 16.03.2026

#### **To cite this document:**

Zhang, J., Choo, W. C., Yee, C. L., & Lin, Y. (2026). Service Failures and Perceived Fairness in Malaysian Hotels- Evidence from Low-Rating Chinese-Language Reviews on Ctrip and An Actionable Recovery Playbook. *Journal of Tourism Hospitality and Environment Management*, 11 (43), 350-376.

### **Abstract:**

Malaysia's tourism growth agenda heightens the need for scalable hotel service-quality governance under the public visibility of review platforms. This study analyzes low-rating Chinese-language reviews of Malaysian hotels posted on Ctrip, treating low ratings as diagnostic texts that concentrate critical incidents and fairness cues. Using review-level multi-label coding anchored in a critical incident lens and a distributive-procedural-interactive justice framework, we identify recurring failure domains, test travel-type heterogeneity with category-wise association tests under false discovery rate control, assess robustness to an alternative low-rating threshold, and examine compound failures via co-occurrence intersections. The complaint landscape is highly concentrated: check-in/out and process failures (21.6%), cleanliness and hygiene failures (21.5%), and maintenance and facilities failures (17.9%) dominate low-score narratives, whereas other domains are notably less prevalent. Incident prevalence differs systematically across travel types, with residual-based contrasts indicating distinct segment emphases within the same overall hierarchy. The incident ordering remains stable when tightening the low-rating definition from rating < 3 to rating ≤ 2. Co-occurrence analysis shows that compound failures among the top three domains are common, suggesting layered vulnerabilities spanning front-desk workflow,

housekeeping execution, and engineering response rather than isolated mishaps. Fairness-related complaints occur in a meaningful minority of low-score reviews and vary by travel type, while justice discourse is dominated by distributive and interactional concerns and procedural justice is least salient. The study contributes a monitorable recovery instrument that maps incident triggers to justice lenses, executable actions, and platform-reply SOP priorities, enabling hotels to standardize what is monitored and to audit recovery quality as a platform-facing capability.

**DOI:**10.35631/JTHER.1143022 **Keyword:**

Ctrip; Critical Incidents; Hotel Review Analytics; Low-Rating Reviews; Service Failure; Service Recovery; Perceived Justice



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [jthem@gaexcellence.com](mailto:jthem@gaexcellence.com).

## Introduction

Malaysia has recently repositioned tourism as a core engine of national growth, which makes hotel service quality a destination competitiveness issue rather than a purely firm level concern. Official tourism satellite account reporting indicates that tourism generated RM291.9 billion and contributed 15.1% to the economy in 2024, while supporting about 3.5 million jobs, which elevates the importance of scalable quality governance rather than isolated fixes (Department of Statistics Malaysia [DOSM], 2025). In parallel, national communications around Visit Malaysia 2026 articulate targets of 35.6 million tourists and RM147.1 billion in receipts, reinforcing a policy logic in which the sector's growth ambitions rely on reliable, repeatable service delivery and credible recovery capability (Tourism Malaysia, 2025). The managerial implication is straightforward. When tourism becomes a macroeconomic instrument, service failures become more than operational noise, because they accumulate into reputational frictions that platforms can amplify and that hotels must manage under public visibility.

Online reviews matter in this environment because they function as public complaint infrastructure. A large body of evidence links review valence to demand relevant outcomes, including room sales and related performance indicators, which means negative reviews are market signals rather than private dissatisfaction alone (Anderson, 2012; Ye, Law, & Gu, 2009). The same logic extends to recovery communication. Managerial responses are not merely courtesy notes, because they can influence subsequent ratings and review volume and can shape how prospective customers interpret the credibility of a hotel's service system (Xie et al., 2016). At the same time, response design can backfire when it is perceived as templated or rote, which clarifies that response governance is part of platform mediated quality rather than an optional add on (Liu, Teichert, Rossi, Li, & Hu, 2021; Lopes, Dens, De Pelsmacker, & Malthouse, 2023).

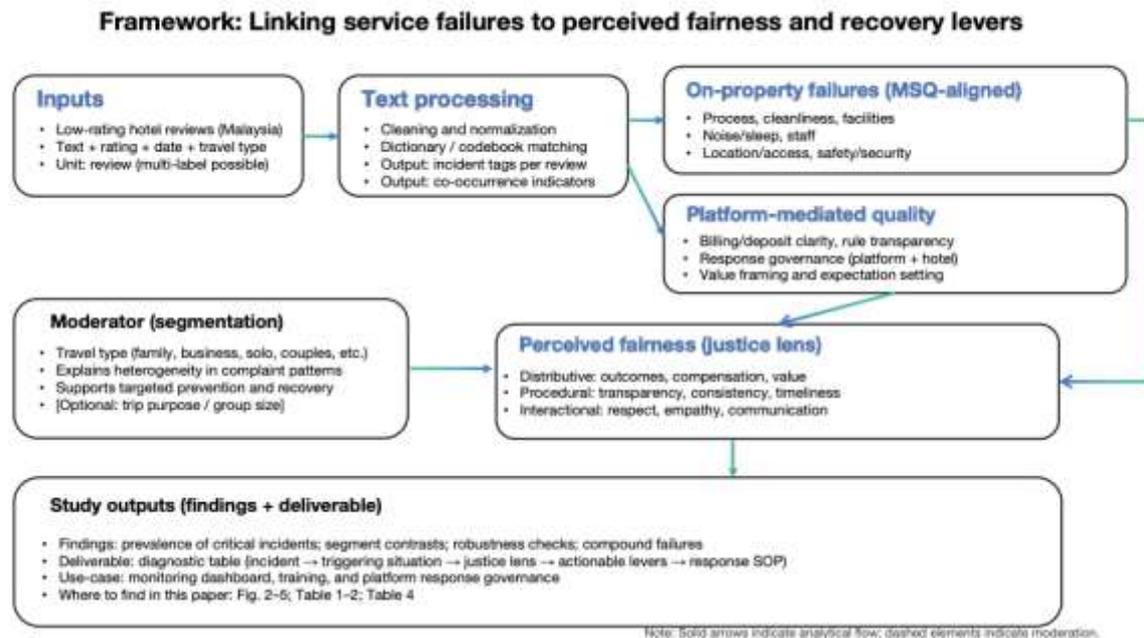
Against this backdrop, the present study examines Chinese language reviews posted on Ctrip for Malaysian hotels. This empirical site is analytically useful because Ctrip is high volume and routinized in complaint and response conventions, while language and culture shape how guests narrate breakdowns and justify low ratings. Hospitality research suggests that culturally meaningful values such as face and guanxi can influence how failures are interpreted and how recovery is judged, which implies that fairness cues in Chinese discourse can be structured differently from those commonly captured in English language corpora (Li, Qiu, & Liu, 2016). Complementing this mechanism view, work on Chinese hotel guest perceptions indicates that Chinese guests' evaluations can vary systematically across destinations even under similar brand cues, which supports the idea that language specific corpora are not simply translation variants (Sun, Tong, & Law, 2017). For Malaysian hotels, these Chinese language texts therefore represent an important governance signal because they sit at the intersection of cross border expectations, platform visibility, and recovery legitimacy.

Conceptually, the study treats low rating reviews as diagnostic texts and relies on a critical incident lens to preserve operational specificity. Hotel service quality has long been theorized as an expectation versus perceptions judgment, which explains the enduring influence of SERVQUAL and lodging specific adaptations such as LODGSERV as organizing logics for what guests notice and how they evaluate performance (Parasuraman, Zeithaml, & Berry, 1988; Knutson, Stevens, Wullaert, Patton, & Yokoyama, 1990). At the same time, empirical hotel research has repeatedly shown that generic dimensions require contextual calibration and that attribute-based refinement can improve managerial interpretability (Akbaba, 2006). A critical incident approach is particularly useful here because it shifts attention from abstract satisfaction toward concrete episodes that guests can narrate and that researchers can categorize to diagnose system weaknesses (Bitner, Booms, & Tetreault, 1990). In a platform context, this logic becomes scalable, since incident prevalence can be monitored as a leading indicator of reliability risk rather than treated as anecdote.

Incidents, however, do not fully explain why some failures escalate into legitimacy claims and durable reputational damage. Service recovery research has repeatedly converged on perceived justice as a central mechanism because customers evaluate not only outcomes but also processes and interpersonal treatment when judging whether recovery is fair (Tax, Brown, & Chandrashekar, 1998; Smith, Bolton, & Wagner, 1999). This mechanism is especially relevant under platform visibility, where interpersonal tone is rendered as text and becomes legible to third party observers. Organizational justice theory further supports this decomposition, showing that distributive, procedural, and interactional justice capture distinct components of fairness perceptions that are consequential for evaluations and responses (Colquitt, Conlon, Wesson, Porter, & Ng, 2001). In hospitality, integrative reviews also suggest that the field increasingly expects service failure research to clarify boundary conditions and deliver implementable guidance rather than stopping at documentation of dissatisfaction (Kim & So, 2023).

These streams point to a gap that remains practically consequential. Although tourism and hospitality analytics has expanded rapidly, recent synthesis work argues that text-based evidence becomes most valuable when it is theory anchored and translated into controllable managerial levers rather than presented as descriptive clustering (Bi, Zhu, & Han, 2024). Yet incident diagnosis and fairness diagnosis are often treated as separate conversations, and platform response governance is frequently discussed as secondary to what happened on property. This study closes that gap by integrating two diagnostic layers into one monitorable

pathway. The first layer identifies which critical incident domains dominate low rating narratives, how they vary by travel type, whether the hierarchy is robust to alternative low rating thresholds, and how frequently dominant failures co-occur as compound breakdowns. The second layer traces when dissatisfaction becomes a fairness claim and how justice discourse decomposes into distributive, procedural, and interactional dimensions across segments under platform visibility.



**Figure 1: Conceptual Framework Linking Incident Diagnosis, Fairness Diagnosis, and a Monitorable Recovery Playbook**

Note: The framework distinguishes on-property quality signals (incident domains) from platform-mediated quality signals (response governance and rule transparency), and it links both to low-rating outcomes and the actionable recovery deliverable (Table 4).

Figure 1 summarizes this logic as a compact framework. On property quality is represented through incident domains that map onto modified hotel service quality dimensions, while platform mediated quality is represented through rule transparency and response governance that condition how recovery is interpreted. This framework then motivates the empirical results reported in Figures 2 to 7 and culminates in the study’s central managerial deliverable. Rather than positioning the analysis as a classification exercise, the paper delivers an actionable recovery playbook in Table 4 that maps incident triggers to justice lenses, executable actions, and platform reply SOP priorities, thereby converting recurring complaint signals into a tool that can be monitored, audited, and standardized across properties.

The remainder of the paper follows this diagnosis to translation pathway. The next section develops the literature on service failures, fairness in recovery, and platform mediated governance. The methods section details the Ctrip review corpus and the review level coding strategy for incidents and fairness. The findings report the incident hierarchy, segment contrasts, robustness checks, co-occurrence structure, and fairness prevalence and composition. The discussion then interprets these results as system vulnerabilities and governance targets, and it explains how an incident justice playbook can support tourism

growth objectives by standardizing what is monitored and how recovery quality is operationalized in a platform facing environment.

## Literature Review

### *Policy Context and Review Governance*

Tourism has recently been repositioned as a core economic engine in Malaysia, which is why hotel service quality increasingly operates as a destination competitiveness issue rather than a purely firm-level concern. Official tourism satellite account reporting indicates that tourism generated RM291.9 billion and contributed 15.1% to the economy in 2024, while supporting about 3.5 million jobs, a scale that makes quality governance and rapid learning systems more consequential than isolated fixes (DOSM, 2025).

In parallel, Visit Malaysia 2026 communications set explicit targets of 35.6 million tourists and RM147.1 billion in tourism receipts, reinforcing the value of tools that can surface recurring failures and standardize recovery routines across properties (Tourism Malaysia, 2025).

Because online reviews shape consideration and booking decisions, review platforms function as an external governance layer that amplifies failures and recoveries beyond the original dyad, and hospitality evidence links review valence to demand-relevant outcomes (Anderson, 2012; Ye, Law, & Gu, 2009).

### *Service Failures as Diagnostic Signals*

Hotel service quality has long been theorized as an expectations-versus-perceptions judgment, which is why SERVQUAL and lodging-specific adaptations such as LODGSERV remain influential as organizing logics for what guests notice and how they evaluate performance (Knutson et al., 1990; Parasuraman et al. 1988). At the same time, hospitality research consistently shows that generic dimensions require contextual calibration because what counts as quality varies by setting, purpose, and cultural expectations, which is why modified scales and attribute-based approaches continue to be used in hotel studies (Akbaba, 2006). Within this broader measurement tradition, service failures matter because they concentrate attention and become narratable episodes that guests circulate to justify low ratings, so low-rating reviews can be read as compressed diagnostic texts rather than balanced summaries of an entire stay.

A critical incident lens is especially useful here because it moves analysis away from diffuse satisfaction and toward concrete breakdowns. Classic work on the service encounter shows that customers can reliably recall favorable and unfavorable episodes and that these incidents can be categorized to identify operational vulnerabilities in hotels and related services Bitner et al. (1990). Under user-generated content conditions, the same logic becomes scalable because reviews allow incidents to aggregate into recurring clusters while preserving situational cues that managers can translate into corrective action. This orientation aligns with recent synthesis in tourism and hospitality analytics, which argues that text evidence is most valuable when it supports theory-guided interpretation and managerial translation rather than descriptive clustering alone (Bi et al., 2024).

### ***Organizational Justice in Recovery***

Service recovery research increasingly converges on the idea that guests evaluate not only whether the hotel fixed a problem, but also whether it did so fairly. The justice framework is therefore explanatory rather than decorative because it clarifies why identical outcomes can be experienced as acceptable in one case and insulting in another. In service research, perceived justice is conventionally decomposed into distributive, procedural, and interactional justice, and this triad has long been used to explain complaint evaluations and post-complaint relationship outcomes (Tax et al., 1998). Foundational recovery models similarly show that recovery evaluations integrate outcome assessments with process and interpersonal cues, which provides a direct rationale for reading fairness signals from complaint narratives (Smith et al., 1999).

In hospitality and tourism, the justice lens is further re-energized by the platform environment because fairness is assessed in public and across time rather than privately within the original encounter. A recent in-depth scoping review indicates that procedural and distributive justice remain central while cross-cultural contexts and qualitative approaches are increasingly emphasized, which is consistent with fairness judgments being socially shaped and context dependent (Golverdi, Sharifirad, & Rastegar, 2024).

At the same time, integrative review evidence suggests the field increasingly expects recovery research to specify boundary conditions and deliver implementable guidance, which keeps justice a practical mechanism when the goal is to design actionable protocols (Kim & So, 2023; Smith et al., 1999).

### ***Platform-Mediated Quality and Webcare***

Online reviews matter for hotel performance, which is why platforms operate as a governance layer over perceived service quality. Large-scale evidence indicates that review valence is associated with demand outcomes and that improvements in reputation metrics can translate into performance-relevant shifts (Anderson, 2012; Ye et al., 2009). This also means negative reviews are not merely post hoc feedback; they are market signals that guide attention and resource allocation for recovery.

This environment elevates platform-mediated quality as a distinct component of what guests evaluate. Guests assess not only what happened on property, but also how the hotel responds, whether the response is timely, and whether it matches the complaint in specificity and tone. Empirical work on managerial responses supports this shift because responses can increase subsequent ratings and review volume and shape downstream eWOM dynamics (Xie, Zhang, Zhang, Singh, & Lee, 2016). Evidence also suggests that response design matters since rote and highly templated replies can depress subsequent engagement and review valence under some conditions (Liu, Wang, Gao, & Gallivan, 2021).

Complementary linguistic analyses further show that features such as polarity, diversity, and readability in managerial responses are associated with return customers' satisfaction change, implying that interactional justice increasingly operates through written communication choices in public settings (Xu & Zhao, 2022).

### ***Ctrip Reviews and Cultural Complaint Narratives***

Ctrip provides a relevant empirical site because it is a high-volume Chinese-language platform where complaint narratives and managerial responses are visible and routinized. Evidence using Chinese travel platforms shows that online reviews can predict hotel performance, reinforcing the managerial significance of platform data in Chinese-speaking markets (Ye et al., 2009). More recent Ctrip-based work also suggests that managerial responses are strategic artifacts rather than simple courtesies, and that response strategies can be theorized through uncertainty reduction and rapport management, which highlights the communicative governance embedded in platform routines (Zhang, La, Huang, & Xie, 2024).

Related Ctrip research further shows how service-failure topics evolve over time and how compensation-oriented interventions can be conceptualized to prevent negative review accumulation, supporting the view that platform complaints are dynamic managerial problems rather than static texts (Bi, Kong, & Gao, 2024).

Language and culture also shape how fairness is expressed. Experimental and hospitality evidence suggests Chinese customers may interpret failures through culturally meaningful values such as face and guanxi, which can alter behavioral responses to recovery (Li, Qiu, & Liu, 2016). Content analyses of Chinese hotel reviews likewise indicate that Chinese guests can differ in how they articulate and interpret service quality across destinations, implying that Chinese-language corpora capture more than translation-level variation (Sun et al. 2017).

### ***Conceptual Synthesis and Recovery Playbook Design***

Recent review work in hospitality analytics argues that the field gains most when text-mining outputs are embedded in theory and converted into managerial levers rather than remaining interpretive labels (Bi et al., 2024). In parallel, service failure and recovery syntheses emphasize a rising expectation for implementable guidance that managers can execute and evaluate (Kim & So, 2023). Against this background, integrating critical incidents with a justice lens becomes a pragmatic strategy because incidents preserve operational specificity and fairness supplies the mechanism linking incidents to recovery expectations. Together, they enable a playbook that is actionable and monitorable through recurring complaint signals and response compliance in platform-facing communication (Bitner et al., 1990; Tax et al., 1998; Xie et al., 2016; Xu & Zhao, 2022).

## **Methodology**

### ***Research Design***

This study uses a theory-informed content analysis design to translate low-rating guest narratives into an incident-based diagnosis of service failures and perceived fairness. Because online reviews capture naturally occurring complaints at scale while retaining contextual cues, they provide a useful empirical base for interpretable and auditable service diagnostics in hospitality and tourism research (Bi et al., 2024, 2024; Zheng, Huang, Wu, Sun, & Wang, 2023). At the same time, service failure and recovery scholarship emphasizes perceived justice as a core mechanism shaping post-failure evaluations, particularly in online environments where process transparency and communicative treatment become salient (Das, Jebarajarkirthy, Maseeh, Lim, & Shah, 2026; Golverdi, Sharifirad, & Rastegar, 2024).

Accordingly, the design first identifies critical incidents in low rating stays and then maps fairness cues to distributive, procedural, and interactional justice lenses, before examining segmentation, robustness, and co-occurrence.

### ***Data Source and Sample***

The empirical context is Malaysian hotels, and the data source is Ctrip, where reviews combine numeric ratings with Chinese-language narratives. We collected publicly accessible reviews and retained only Chinese-language texts to ensure linguistic comparability and to preserve interpretive validity for rule-based coding, given that complaint framing and fairness cues can be language-sensitive (Wan, 2024). The unit of analysis is the individual review, and sample characteristics, hotel coverage, observation window, and the distribution of platform-labeled travel types are summarized in Table 1. Low-rating reviews were operationalized primarily as ratings below 3 for the main analyses, and a stricter threshold of ratings of 2 or below was used for sensitivity checks.

**Table 1: Sample Profile and Travel Type Distribution**

<b>Metric</b>	<b>Value</b>
Hotels (n)	961
Low-rating reviews (n)	14120
Mean rating	2.06
Median rating	2.2
Period covered	2022-10 to 2025-12

Panel A reports dataset size and period. Panel B reports the distribution of Ctrip travel types within the low-rating sample.

<b>Travel type</b>	<b>n</b>	<b>%</b>
Family	4322	30.6
Business	2609	18.5
Friends	2367	16.8
Couples	2007	14.2
Solo	1920	13.6
Other	798	5.7
Booked for others	92	0.7
Unknown	5	0.0

Source: Ctrip (Chinese-language low-rating reviews of Malaysian hotels); authors' compilation.

## Measures And Coding Procedures

Coding was conducted at the review level and allowed multi-label assignment, since a single narrative can report multiple failures and multiple fairness claims.

### Incident Domains

Service failures were operationalized through an eight-domain incident taxonomy designed for diagnostic usefulness, covering common breakdowns such as cleanliness and hygiene, facility and maintenance issues, process friction during check-in and check-out, staff attitude and courtesy, noise and sleep disruption, safety and security, location and access, and billing or deposit disputes. Operational definitions and representative indicators are reported in Table 2. Incidents were identified using a dictionary- and rule-based matching approach anchored in these indicators, supported by synonym merging and iterative inspection to refine ambiguous terms, consistent with the interpretability emphasis in recent hospitality text analysis research (Bi et al., 2024).

**Table 2: Critical Incident Codebook and Dictionary Matching Examples (English Gloss)**

Incident domain	Definition (1 sentence)	Example indicators (English gloss)
Billing/Deposit & Value	Concerns about charges, deposits, refunds, perceived overpricing, and value-for-money issues.	unreasonable/extra charges; overcharged; fee deducted; unauthorized deduction; unreasonable charges; fee refund; refund; deposit not refunded; deposit not returned; not worth the price (poor value)
Cleanliness & Hygiene	Complaints about room cleanliness, odors, pests, and hygiene-related problems affecting comfort and perceived sanitation.	dirty; poor hygiene/cleanliness; unclean; filthy; stains; dust; hair; bedsheet; duvet cover; towel
Maintenance & Facilities Failure	Facility breakdowns and maintenance problems (e.g., air-conditioning, hot water, plumbing, Wi-Fi) that disrupt the stay.	broken; not working/unusable; malfunction; water leak; air conditioning; air conditioner (AC); AC not cooling; hot water; no hot water; water pressure
Location/Access & Environment	Problems related to location convenience, accessibility, surrounding environment, and parking/transport issues.	remote location; inconvenient; poor transport access; surroundings/nearby area; poor environment; messy/chaotic; dirty and messy; parking; hard to park; far from (distance complaint)

Noise & Sleep Disruption	Noise sources and poor sound insulation leading to sleep disruption (e.g., traffic, neighbors, construction).	noisy; noise; very noisy; poor soundproofing; cannot sleep; unable to fall asleep; late night/early morning; upstairs; next door; construction noise
Check-in/out & Process	Issues related to check-in/out procedures, waiting time, booking handling, and process-related service failures.	check-in; check-out; queue/line; long wait; waiting; procedures/paperwork; deposit; deposit; rules; policy
Safety & Security	Safety and security concerns such as theft risk, insecure locks, harassment, and perceived insecurity.	unsafe; felt afraid; poor security/safety; items stolen; theft; lost belongings; cannot lock; door lock; broken; harassment
Staff Attitude & Courtesy	Negative interactions with staff, including attitude, responsiveness, communication issues, and service courtesy.	bad attitude; poor service; impatient; indifferent/cold; perfunctory; dismissive/ignoring; unresponsive/ignored; swore/scolded; shouted; rude/aggressive

Note: The operational dictionary was constructed in Chinese. Appendix Table A2 reports the original Chinese indicators and their English glosses for transparency and replication.

### *Perceived Fairness and Justice Lens Mapping*

Perceived fairness was coded using justice lenses, namely distributive, procedural, and interactional justice, because these dimensions remain central in explaining service recovery evaluations (Das et al., 2026; Golverdi et al., 2024). Coding followed a two-stage procedure. First, an “any fairness” indicator flags whether fairness cues appear in a review. Second, flagged reviews were mapped to one or more justice lenses using rule sets anchored in outcome-related cues, process-related cues, and interpersonal treatment cues. The mapping logic is documented in the appendix materials, while the translation from fairness signals to recovery actions is consolidated in Table 4. Because platform-facing communication can shape post-failure evaluation trajectories, response guidance is treated as part of the recovery system rather than an optional add-on (Xu & Zhao, 2022).

### *Analytical Strategy*

Analyses were structured to separate operational description from substantive findings. First, incident prevalence was estimated as the proportion of low-rating reviews containing each incident domain, with results reported in Figure 2 and Table 3. Second, incident distributions were compared across platform-labeled travel types using chi-square tests of independence, while cell-level deviations were interpreted via adjusted standardized residuals and multiple testing was addressed through false discovery rate control using the Benjamini and Hochberg (1995) procedure; the segmented pattern is visualized in Figure 3. Third, co-occurrence among

dominant incident domains was examined using an UpSet visualization to represent intersection sizes in a readable manner when overlaps become complex (Lex, Gehlenborg, Strobel, Vuillemot, & Pfister, 2014), and results are reported in Figure 5. All computations and visualizations were conducted in R (R Core Team, 2024).

### *Robustness Check*

To assess threshold sensitivity, the core prevalence estimates, and key segmented contrasts were replicated under the stricter low-rating definition of ratings of 2 or below, and the comparison is summarized in Figure 4. Consistency across thresholds is interpreted as evidence that the incident hierarchy and fairness patterns reflect stable structures rather than artifacts of a single cut-off.

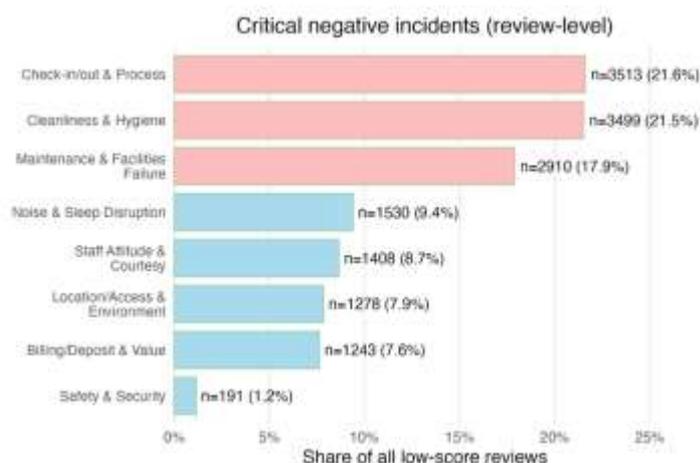
### *Ethical Considerations*

The study analyzes publicly accessible online reviews and reports results in aggregate. No personal identification information was collected. When illustrative quotations are used, the original Chinese text is retained alongside an English translation to support accessibility while allowing readers to audit interpretive fidelity.

### **Findings**

Consistent with Malaysia’s tourism growth agenda and the governance role of review platforms, the findings below report a compact set of high frequency failure signals and fairness cues that can be monitored and translated into standardized recovery actions. Incident domains are operationalized via a rule-guided Chinese dictionary summarized in Table 2, and the original Chinese indicators with English glosses are provided in Appendix Table A2 to support transparency and replication.

### *Overall Prevalence of Critical Incidents in Low Rating Reviews*



**Figure 2: Critical Negative Incidents in Low-Rating Reviews (Review-Level Prevalence)**

Note: Bars report the share of low-rating reviews (main threshold: rating < 3) that mention each incident domain. Incident coding is multi-label, so category shares are not mutually exclusive. Percent labels show prevalence within the low-rating corpus (N = 14,120).

The review level coding indicates a concentrated complaint landscape in which a small number of incident domains account for a substantial share of low rating narratives. As summarized in Table 3 and visualized in Figure.2, process related failures in check in, check out, and service procedures are the most prevalent category, appearing in  $n = 3,513$  (21.6%) of the  $N = 16,246$  low rating reviews. Cleanliness and hygiene failures follow closely at  $n = 3,499$  (21.5%), while maintenance and facilities failures rank third at  $n = 2,910$  (17.9%). This concentration is consistent with recent syntheses in tourism and hospitality analytics, which suggest that incident focused diagnosis often surfaces highly observable breakdowns that are proximal to the guest experience, particularly when studies move beyond sentiment toward operational categories (Bi et al., 2024; Song, 2024).

**Table 3: Prevalence Of Critical Incidents in Low-Rating Reviews**

Rank	Incident category	n (reviews)	% of low-rating reviews
1	Check-in/out Process	& 3513	21.6%
2	Cleanliness Hygiene	& 3499	21.5%
3	Maintenance Facilities Failure	& 2910	17.9%
4	Noise Disruption	& Sleep 1530	9.4%
5	Staff Attitude Courtesy	& 1408	8.7%
6	Location/Access Environment	& 1278	7.9%
7	Billing/Deposit Value	& 1243	7.6%
8	Safety & Security	191	1.2%

Note: Percentages are calculated over  $N = 14,120$  low-rating reviews (main threshold: rating < 3). Reviews may contain multiple incident categories (multi-label coding), therefore percentages do not sum to 100%.

In practical terms, the pattern implies a small set of monitorable failure hotspots that can be tracked as leading indicators of reliability risk, because shifts in these category rates are interpretable and directly attributable to controllable front desk, housekeeping, and maintenance routines.

Beyond the top tier, the distribution becomes more dispersed. Noise and sleep disruption appears in  $n = 1,530$  (9.4%) of low rating reviews, and staff attitude and courtesy issues appear in  $n = 1,408$  (8.7%). Location and access as well as billing, deposit, and value disputes each contribute a similar share, whereas safety and security incidents are comparatively rare at  $n = 191$  (1.2%). Because incidents are coded in a multi-label manner, these figures should be interpreted as category specific prevalence rather than mutually exclusive shares, and percentages therefore do not sum to 100%. This framing aligns with the diagnostic orientation emphasized in service failure and recovery reviews, where the managerial value of review data lies in identifying recurring breakdowns that can be targeted through recovery design (Kim & So, 2023; Sahaf & Fazili, 2024).

### Segment Contrasts by Travel Type

Although the overall incident hierarchy is stable, incident prevalence differs across travel types in ways that are statistically reliable and substantively meaningful. The category specific association tests reported in Table 3 indicate that the distribution of incident mentions varies by travel type for most categories after false discovery rate control, whereas billing, deposit, and value disputes do not show a statistically reliable travel type pattern ( $p \text{ FDR} = .131$ ). The complete set of chi square statistics and FDR adjusted p values is reported in Appendix Table A1 to keep the main text focused on substantively interpretable contrasts. Across categories, the corresponding Cramér’s V values fall in a small effect range (approximately .025 to .082), which supports an interpretation of systematic but nondeterministic segment differences. For that reason, the segment narrative prioritizes contrasts that are interpretable in operational terms rather than relying on significance alone, consistent with standard guidance for multiplicity control in multi test settings (Benjamini & Hochberg, 1995; Bi et al., 2024).

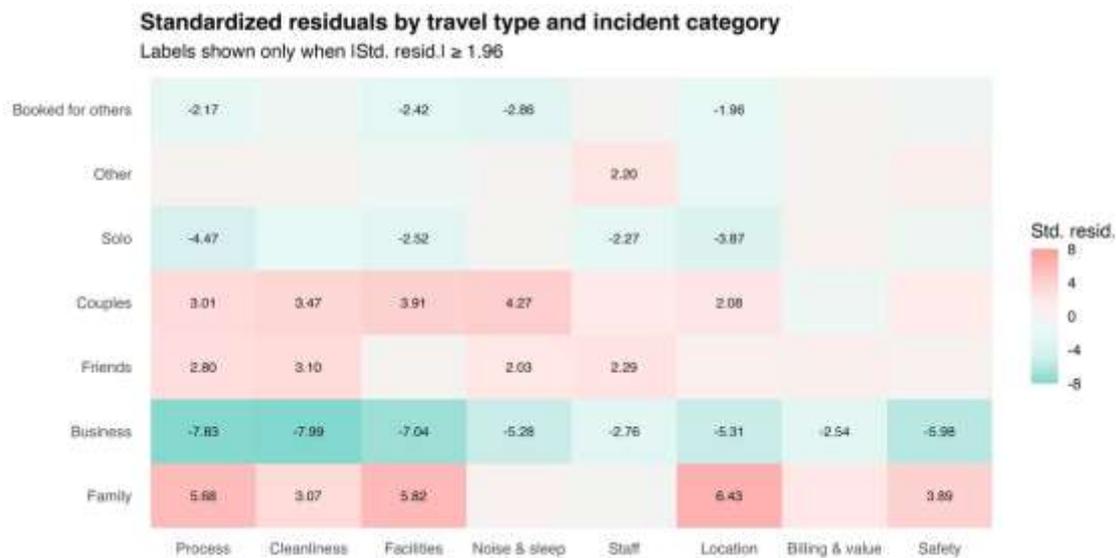


**Figure 3: Incident Prevalence by Travel Type (Within-Row Shares)**

Note: Cells report, for each travel type, the percentage of low-rating reviews that mention each incident domain. Because coding is multi-label, row totals may exceed 100%. “Booked for others” is flagged where  $n < 100$ .

A descriptive portrait of segment heterogeneity is provided by the within segment shares in Figure.3, which shows the percentage of low rating reviews in each travel type that mention each incident category. Two features stand out. First, leisure-oriented segments consistently carry higher within segment prevalence in the dominant reliability domains. Couples show particularly high shares for cleanliness (29.1%) and process (27.6%) while also reporting substantial facilities mentions (24.3%), and friends display a similarly elevated profile for process (27.7%) and cleanliness (27.0%) alongside nontrivial facilities mentions (21.7%). Family travel also concentrates heavily in the reliability triad, with process at 27.5%, cleanliness at 26.4%, and facilities at 24.3%. Taken together, these patterns suggest that when leisure stays crystallize into low ratings, narratives tend to anchor on breakdowns that are concrete, experience proximal, and difficult to discount once expectations for basic reliability are violated, which is consistent with the diagnostic view that low score reviews foreground operationally legible failures rather than balanced accounts (Kim & So, 2023; Sahaf & Fazili, 2024).

Second, business travel shows a noticeably flatter share profile across the same reliability domains, which provides context for why the inferential evidence that follows is directionally consistent across categories. Within business travel low rating reviews, process is mentioned in 19.4% of cases and cleanliness in 19.0%, while facilities appear in 13.7%, all of which are meaningfully lower than the corresponding shares in couples, friends, or family segments (Figure.3). Similar attenuation is visible for location (7.4%) and staff (7.7%). This pattern does not imply that business travellers are uniformly satisfied. Instead, it indicates that, within this incident taxonomy, business travel low ratings are less frequently narrated through the dominant reliability domains, which is compatible with the idea that trip goals and complaint conventions shape what becomes narratively focal in short platform texts (Kim & So, 2023). The booked for others segment should be interpreted cautiously because  $n < 100$ , yet its comparatively low shares across multiple categories indicate that the reviewer role itself may condition what is reported and how confidently one attributes blame.



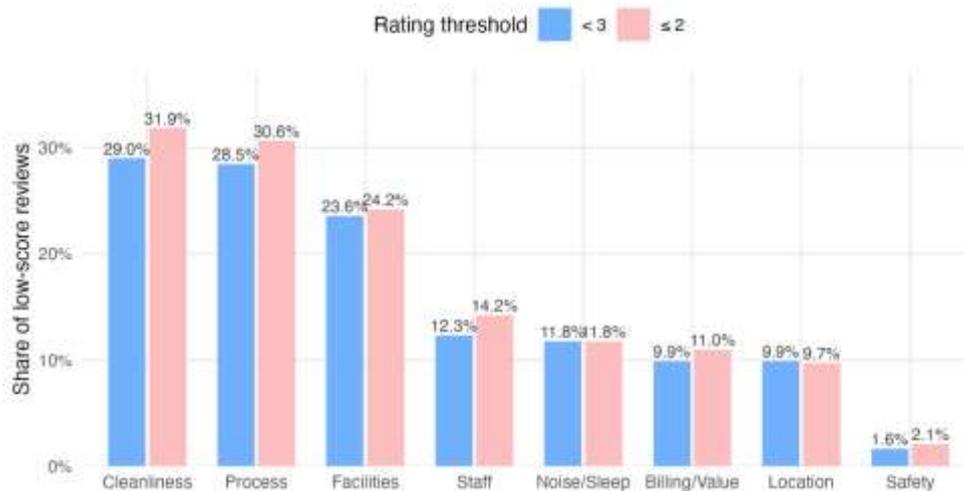
**Figure 4: Standardized Residuals by Travel Type and Incident Category**

*Note:* Heatmap displays adjusted standardized residuals from travel type  $\times$  incident category contingency comparisons. Positive values indicate over-representation, and negative values indicate under-representation relative to independence. Labels are shown only when  $|\text{Std. resid.}| \geq 1.96$ .

The descriptive contrasts in Figure.3 are then corroborated and disciplined by residual-based evidence in Figure.4, which isolates travel type by category cells that deviate meaningfully from independence and labels only those with  $|\text{Std. resid.}| \geq 1.96$ . Under this stricter evidentiary threshold, business travel is consistently underrepresented across core categories, including cleanliness (standardized residual =  $-7.99$ ), process ( $-7.83$ ), and facilities ( $-7.04$ ), which confirms that the flatter share profile is not an artefact of scaling. Conversely, family travel exhibits clear over representation for location (6.43), facilities (5.82), and process (5.68), while also surfacing elevated safety and security salience (3.89), which suggests that family oriented low rating narratives are particularly sensitive to accessibility and perceived environmental control. Couples show a pronounced positive deviation in noise and sleep disruption (4.27), aligning with their higher noise share in Figure.3 (12.6%) and supporting an interpretation in which rest quality functions as a central evaluation anchor in leisure stays. Friends register positive deviations in process (2.80) and cleanliness (3.10) and in staff related complaints (2.29), indicating that once reliability breaks down, interactional encounters may become more salient in group contexts where shared appraisal and coordination are

foregrounded (Sahaf & Fazili, 2024). These segment contrasts matter because recovery expectations are shaped by situational goals and perceived legitimacy, and justice-oriented recovery accounts emphasize that what is judged as fair often depends on context rather than on failure occurrence alone (Kim & So, 2023; Sahaf & Fazili, 2024).

### ***Robustness To Rating Threshold Choice***

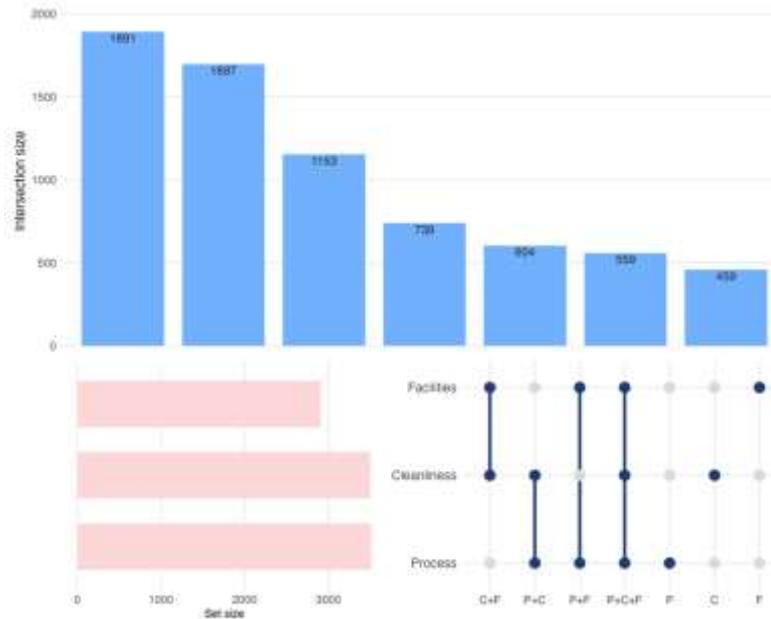


**Figure 5: Robustness Check for Low-Rating Threshold Choice (rating  $\le 3$  vs. rating  $\le 2$ )**

*Note:* Bars compare incident prevalence under the main low-rating definition (rating  $\le 3$ ) and a stricter definition (rating  $\le 2$ ). Values are percentages of reviews meeting each threshold that mention each incident domain; coding remains multi-label.

Because the operational definition of low ratings can vary across platforms and studies, the analysis evaluates whether the incident hierarchy is sensitive to the rating threshold. Figure 5 compares prevalence estimates under the main threshold of rating  $\le 3$  with a stricter definition of rating  $\le 2$ . The incident ordering remains highly stable across the two specifications, and the three dominant categories remain dominant. Although cleanliness and process increase modestly under the stricter threshold, the absolute differences remain limited, which suggests that the diagnostic priority of process, cleanliness, and facilities does not depend on the cutoff choice. This type of stability check is increasingly recommended in hospitality and tourism text analytics because it helps distinguish durable managerial signals from analytic choices (Bi et al., 2024).

### Co-Occurrence And Compound Failure Mechanisms In The Top Three Categories



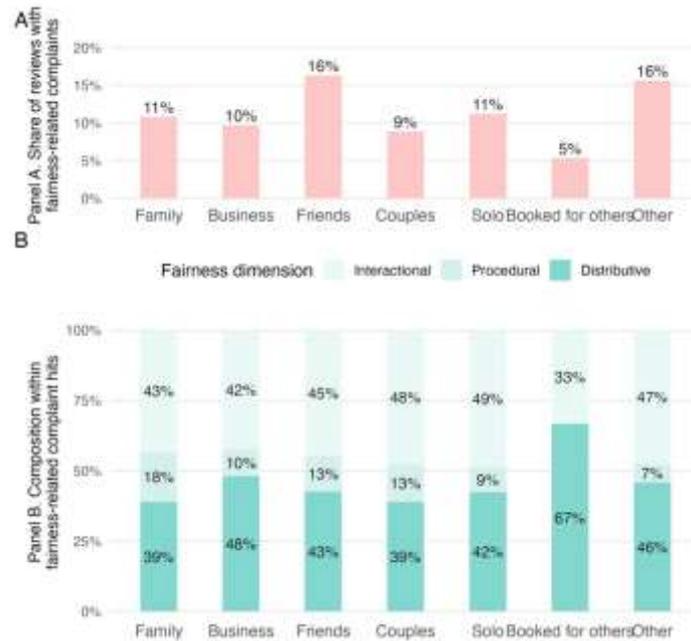
**Figure 6: Co-Occurrence Structure of the Top Three Incident Domains (UpSet Intersections)**

*Note:* UpSet visualization summarizes intersections among the top three incident domains (Process, Cleanliness, Facilities). Intersection sizes are computed within the subset of reviews that mention at least one of these three domains ( $n = 7,102$ ), so counts represent compound-failure configurations among dominant incidents rather than whole-corpus frequencies (Lex, Gehlenborg, Strobel, Vuillemot, & Pfister, 2014).

Prevalence rankings alone can understate the operational reality that guests often report multiple failures within the same stay. For that reason, Figure.6 examines the co-occurrence structure of the top three incident domains using an UpSet intersection visualization, which is designed to represent set intersections and their sizes without the ambiguity of multi way Venn diagrams (Lex et al., 2014; Strobel, Vuillemot, & Pfister, 2014). Importantly, the reported intersection sizes are computed within the subset of reviews that mention at least one of the three dominant incident domains, which totals  $n = 7,102$  reviews, or 43.7% of the low rating corpus.

The results show that single domain failures remain common, yet compound failures are far from exceptional. Within this top three subset, single domain mentions account for 66.8% of cases, whereas multi domain overlaps account for 33.2%, indicating that roughly one in three dominant failure narratives involves at least two simultaneous breakdowns. The largest two-way overlap is between cleanliness and facilities ( $n = 739$ ; 10.4% of the subset), followed by process with cleanliness ( $n = 604$ ; 8.5%) and process with facilities ( $n = 559$ ; 7.9%). A nontrivial intersection includes all three domains ( $n = 459$ ; 6.5%). Substantively, this clustering suggests that a meaningful subset of low rating narratives reflects layered breakdowns spanning front desk workflow, housekeeping execution, and engineering response. Such patterns align with service failure and recovery research arguing that recovery effectiveness depends on diagnosing whether a failure is isolated or compounded, since remedies that are adequate for a single breakdown are often judged insufficient when multiple failures accumulate (Kim & So, 2023; Sahaf & Fazili, 2024).

**Fairness Related Complaints and Their Justice Composition**



**Figure 7: Fairness-Related Complaints by Travel Type (Panel A Prevalence; Panel B Composition by Justice Dimension)**

Note: Panel A reports the share of reviews within each travel type that contain fairness-related complaint signals. Panel B reports the composition of justice dimensions within the subset of fairness hits for each travel type, decomposed into distributive, procedural, and interactional justice; shares sum to 100% within each travel type's fairness subset.

Where incident coding captures what went wrong, fairness coding captures how guests evaluate legitimacy once a failure is experienced, especially when the episode involves perceived loss, disrespect, or inadequate repair. Because travel type metadata is available for a broader set of reviews than the low rating subset, Figure.7 reports fairness prevalence and composition over the travel type labelled review corpus (N = 28,882), while the diagnosis above is estimated on low rating reviews (N = 16,246). As shown in Figure.7A, fairness-related complaints occur in a meaningful minority of reviews and vary by travel type. Across the travel type labelled corpus, fairness coded narratives appear in n = 3,248 reviews (11.2%). Segment level prevalence is highest among friends' travel (16.3%) and the residual other category (15.6%), whereas booked for others shows the lowest prevalence (5.4%). This variation reinforces the view that perceived unfairness is not simply a direct function of failure occurrence. Rather, it reflects how guests interpret responsibility, repair adequacy, and interpersonal treatment in context, which is consistent with justice-based accounts of recovery evaluations (Kim & So, 2023; Sahaf & Fazili, 2024).

The composition results in Figure.7B show a stable structure across segments, in which distributive and interactional justice dominate fairness discourse, whereas procedural justice constitutes the smallest share. Aggregated across all fairness hits, interactional justice accounts for 45.1% (n = 1,670), distributive justice accounts for 41.8% (n = 1,549), and procedural justice accounts for 13.1% (n = 486), with the total number of justice hits (n = 3,705) exceeding the number of fairness coded reviews because a single review can contain multiple justice dimensions. This pattern directly motivates Table 4, because Table 4 operationalizes the justice lenses into a recovery mapping that links triggering situations to actionable levers and platform

reply to guidance, thereby translating narrative signals into recovery actions that can be standardized as response protocols. Read as an implementation aid, Table 4 supports a simple operational flow in which managers first identify the dominant trigger and then apply the corresponding justice prioritized action set and response elements, which enables more consistent recovery governance on review platforms.

**Table 4: Fairness-Based Recovery Playbook (Justice Lens to Actionable Levers)**

Justice lens	Common signals (Top terms)	Mapped failure domain(s)	Recovery levers (ops)	Reply SOP (platform)
Distributive	Value for money / Not worth it; Refund / Return money; Deposit / Down payment; Extra charges / Hidden fees; Wrong billing / Overcharge; Compensation / Make-up; Price mismatch / Difference; Overpriced / Too expensive	Billing/Value	1) Price transparency: disclose total price + all fees before payment; 2) Standardize deposit/refund rules with clear timelines; 3) Billing audit + dispute channel (same-day verification); 4) Offer proportional compensation (voucher/partial refund) for...	Acknowledge issue → clarify charges/rules → offer remedy (refund/waiver/upgrade) → timeline → follow-up.
Interactional	Rude attitude / Bad manner; Indifferent / Perfunctory; Unresponsive / Ignore; Front desk attitude; Disrespect / Not polite; Service not professional; No explanation / Poor communication; Staff blamed guest	Staff	1) Interaction training: empathy script + apology-first; 2) Communication standard: explain 原因-方案-时限; 3) Service recovery etiquette: respectful language, no blame-shifting; 4) Monitor with spot checks + post-complaint follow-up message.	Acknowledge feelings → apologize with empathy → explain briefly (no blaming) → offer remedy → invite direct contact & follow-up.
Procedural	Housekeeping process issues; Booking /	Process	1) Reduce check-in friction: staffing by peak	Acknowledge process failure → explain what happened → commit to

System issue; Delay / No follow-up; Handling complaint poorly; Slow check-in / Queue; No solution / Pass the buck; Unclear rules / Not transparent; Refund process complicated	hours + queue triage; 2) SOP for complaint handling: acknowledgement ≤30 min, resolution target ≤24 h; 3) Rule clarity: publish house rules (noise, parking, deposits) in booking + at front desk; 4) Escalation...	fix (SLA/steps) → timeline → confirm resolution.
---	---	--

Note: Common signals are synonym-merged phrases extracted from low-rating reviews. Recovery levers summarize operational actions; reply SOP provides platform response guidance.

Moreover, because the platform setting makes managerial responses visible and consequential, the prominence of interactional justice is especially salient. Prior research suggests that the linguistic and discourse features of online managerial responses shape how recovery efforts are interpreted and evaluated over time, which implies that response governance is part of the recovery system rather than a peripheral add on (Xu & Zhao, 2022; Wan, 2024).

## Discussion

### *From Incident Counts to System Vulnerabilities*

The incident hierarchy shown in Figure.2 and Table 3 is most informative when it is read as a diagnostic map of repeated breakdown points in the service delivery system that are severe enough to trigger low ratings. Process related failures, cleanliness and hygiene issues, and facilities and maintenance problems dominate the corpus, which suggests that low ratings in Chinese language Ctrip reviews of Malaysian hotels are often anchored in breakdowns that are concrete, experience proximal, and difficult for guests to reinterpret once expectations of basic reliability are violated. This diagnostic reading aligns with recent syntheses in tourism and hospitality analytics, which emphasize that review mining becomes managerially meaningful when categories remain operationally interpretable and traceable to controllable processes rather than being reduced to sentiment polarity alone (Bi et al., 2024).

At the same time, the hierarchy should not be interpreted as a universal ranking of what matters most to all guests. Instead, it indicates where failures are most consistently narrated as decisive in low score texts, which is precisely why they are suitable as monitoring targets. Process friction likely reflects coordination strain across front desk routines, queue handling, and cross role handoffs, while cleanliness complaints point to variability in housekeeping execution and inspection discipline. Facilities and maintenance failures then become visible signals of neglected upkeep because they constrain comfort in ways that guests experience as non-negotiable. Read this way, the results move beyond description and toward vulnerability diagnosis, and this framing prepares the segment interpretation that follows.

### ***Segment Heterogeneity as Expectation Work***

The travel type contrasts in Figure.3 and Figure.4 indicate that low ratings are not narrated through a single complaint template, even when the overall hierarchy remains stable. Instead, complaint narratives appear to be filtered through trip goals and contextual expectations, which resonates with recovery research that treats perceived severity, responsibility attributions, and recovery needs as contingent rather than universal (Kim & So, 2023; Sahaf & Fazili, 2024). Methodologically, the residual-based interpretation disciplines the discussion because it anchors inference in deviations that exceed an interpretable threshold, which is essential in large corpora where many comparisons can become statistically significant. In parallel, false discovery rate control offers a principled response to multiplicity in category wise testing (Benjamini & Hochberg, 1995).

Substantively, the segment contrasts imply that recovery design should be context sensitive. Standardization remains valuable, yet its effectiveness depends on acknowledging segment priorities, because a generic response can solve the wrong pain point for the segment most likely to complain in a given situation. In this sense, the travel type patterns are not merely descriptive, since they indicate where reliability monitoring and response emphasis should be adjusted without abandoning a common recovery architecture.

### ***Robustness as a Credibility Test***

The threshold sensitivity checks in Figure.5 strengthens the credibility of the diagnostic claims by showing that the incident ordering is not fragile to a reasonable change in what qualifies as a low rating. When the definition tightens from ratings below three to ratings of two or below, the leading incident remain leading and the prevalence shifts are modest in absolute terms. In tourism and hospitality text analytics, robustness checks are increasingly treated as a quality standard because they help distinguish structural signals from artefacts introduced by operational cutoffs or analytic choices (Bi et al., 2024).

This stability matters beyond method because it supports a practical claim about monitoring. If the hierarchy persists under a stricter threshold, then tracking these incident domains becomes a defensible way to audit reliability risk over time in a manner that is interpretable to managers and comparable across reporting cycles.

### ***Why Co-occurrence Changes the Recovery Problem***

The co-occurrence structure in Figure.6 is not a decorative complement to prevalence ranking because it changes what recovery needs to accomplish. Using an UpSet representation enables transparent reporting of intersection sizes when multiple overlaps exist and avoids the interpretive ambiguity of multi way Venn reasoning (Lex et al., 2014). Importantly, the intersections are computed within the subset of reviews that mention at least one of the top three domains, which means the results describe the internal configuration of dominant failures rather than overall corpus frequencies.

Substantively, the overlaps suggest that a meaningful subset of low rating stays involves layered breakdowns spanning front desk workflow, housekeeping execution, and engineering response. This configuration plausibly amplifies perceived severity because it signals systemic unreliability rather than a single mishap. Classic recovery theory helps clarify why these

matters, since post failure satisfaction integrates outcome assessments with process and interpersonal cues, and partial remedies are often judged insufficient when multiple losses accumulate (Smith, Bolton, & Wagner, 1999). Consequently, the co-occurrence evidence supports a bundled intervention mindset, because improvements in one subsystem may not shift evaluations unless adjacent subsystems are addressed in parallel.

### ***Fairness as the Interpretive Bridge in Platform Mediated Recovery***

The fairness patterns in Figure.7A and Figure.7B clarify why some narratives escalate from dissatisfaction into legitimate claims. Justice theory remains a strong interpretive backbone because it specifies how people evaluate outcomes, procedures, and interpersonal treatment when judging whether an organization responded fairly (Tax et al., 1998; Colquitt et al., 2001). In the present corpus, distributive and interactional justice dominate fairness discourse, while procedural justice occupies a smaller share. This composition is consistent with the idea that platform complaint texts privilege narratable losses and interpersonal cues, which are often easier to articulate than procedural illegitimacy in compressed reviews.

What makes this particularly consequential is the platform environment, because responses are public, persistent, and part of visible recovery governance. Evidence from hotel contexts indicates that linguistic characteristics of managerial responses can shape satisfaction trajectories, which makes interactional justice a practical lever rather than a purely interpretive category (Xu & Zhao, 2022). Recent work also documents systematic language differences between generative responses and manager written responses in online complaint settings, which reinforces the need to specify response standards that preserve interpersonal nuance rather than relying on ad hoc replies (Wan, 2024). Framed this way, fairness becomes the bridge between incident diagnosis and the managerial deliverable in Table 4, because Table 4 links triggers justice prioritized action sets and platform facing response elements.

### ***Contributions as a Monitorable Tool, not a Classification Exercise***

The central contribution of the study is not that it classifies complaints, but that it delivers a monitorable diagnostic instrument that can be embedded into routine quality governance. Analytically, the design integrates incident prevalence with segment heterogeneity, threshold robustness, and compound failure structure, which responds to calls for transparency and operational interpretability in tourism and hospitality text analytics (Bi et al., 2024). This instrument is monitorable because each component can be audited over time in a way that produces indicators managers can track rather than labels they merely interpret. Incident prevalence functions as a category rate within low rating reviews, co-occurrence functions as a compound failure proportion among dominant domains, and fairness functions as the share and composition of justice claims that signal when recovery is evaluated through legitimacy rather than performance.

At the translational level, the incident justice mapping consolidated in Table 4 functions as an actionable playbook that specifies triggering situations, the dominant justice lens, executable actions, and response priorities. This mapping enables hotels to move from post hoc interpretation to repeatable recovery practice, and it also supports governance on review platforms by making response quality auditable rather than purely discretionary.

### ***Limitations and Boundary Conditions***

Interpretation should remain bound by the corpus and platform context. The evidence is drawn from Chinese language reviews on a single platform, and it therefore reflects discursive conventions, rating practices, and response visibility specific to that environment. In addition, because the incident analysis focuses on low ratings, it should be read as a diagnostic account of severe dissatisfaction rather than a full representation of average service quality across the population of stays.

For auditability, Appendix Table A1 documents the full set of travel type association tests, while Appendix Table A2 provides the underlying Chinese dictionary indicators and English glosses that support the incident taxonomy. These materials make the coding logic more transparent and portable, which positions multilingual and cross platform extensions as a natural next step while keeping the present claims appropriately bounded.

### ***Transition to the Conclusion***

Overall, the discussion traces a coherent pathway from description to governance. Low ratings concentrate around a small set of high visibility reliability failures, segments differ in how they narrate those failures, compound breakdowns are common enough to warrant bundled interventions, and fairness discourse crystallizes when perceived losses and interpersonal treatment appear misaligned under public platform visibility. The conclusion can therefore distill three actionable takeaways, first by specifying which incident domains are most suitable as monitoring signals of reliability risk, second by clarifying when segment context and compound failures require differentiated recovery emphasis, and third by explaining how an incident justice playbook operationalizes recovery as a standardized yet fairness sensitive capability (Sahaf & Fazili, 2024; Xu & Zhao, 2022).

---

**Acknowledgements:** The author would like to thank the academic staff and librarians at Universiti Putra Malaysia for their support in accessing bibliographic databases and reference materials used in this study. Sincere appreciation is also extended to colleagues and peers who provided constructive comments on earlier drafts of the manuscript. Any remaining errors or omissions are the author's sole responsibility.

**Funding Statement:** No Funding

**Conflict of Interest Statement:** The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the Journal of Tourism, Hospitality and Environment Management (JTHER).

**Ethics Statement:** This study did not involve any human participants, animals, or sensitive data requiring ethical approval. The authors confirm that the research was conducted in accordance with accepted academic integrity and ethical publishing standards.

---

---

**Author Contribution Statement:** All authors contributed significantly to the development of this manuscript. Zhang Juan was responsible for the conceptualization, methodology, and overall supervision of the study, and also handled data collection, data analysis, and interpretation of the results. Choo Wei Chong and Yee Choy Leong provided guidance and critically reviewed the manuscript. Lin Yihuan contributed to the language editing and polishing of the paper. All authors read and approved the final version of the manuscript prior to submission.

---

## References

- Akbaba, A. (2006). Measuring service quality in the hotel industry: A study in a business hotel in Turkey. *International Journal of Hospitality Management*, 25(2), 170–192. <https://doi.org/10.1016/j.ijhm.2005.08.006>
- Anderson, C. K. (2012). The impact of social media on lodging performance. *Cornell Hospitality Report*, 12(15), 1–14.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bi, D., Kong, J., & Gao, Y. (2024). Modeling the effect of service failures and hotel response on customer satisfaction: Evidence from online reviews. *Journal of Hospitality and Tourism Management*, 58, 269–285. <https://doi.org/10.1016/j.jhtm.2024.01.006>
- Bi, J. W., Zhu, X. E., & Han, T. Y. (2024). Text analysis in tourism and hospitality: A comprehensive review. *Journal of Travel Research*, 63(8), 1847–1869. <https://doi.org/10.1177/00472875241247318>
- Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. *Journal of Marketing*, 54(1), 71–84.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86(3), 425–445. <https://doi.org/10.1037/0021-9010.86.3.425>
- Das, G., Jebarajakirthy, C., Maseeh, H., Lim, W. M., & Shah, Z. (2026). Online service failure and recovery: A bibliometric analysis and integrative review. *Journal of Business Research*, 180, 115752. <https://doi.org/10.1016/j.jbusres.2025.115752>
- Department of Statistics Malaysia. (2025). Tourism satellite account 2024 [Press release]. Retrieved from Department of Statistics Malaysia website.
- Golverdi, M., Sharifirad, M. S., & Rastegar, R. (2024). Mapping organizational justice in tourism, hospitality and events: A scoping review. *Journal of Hospitality and Tourism Management*, 60, 22–32. <https://doi.org/10.1016/j.jhtm.2024.06.004>
- Kim, H., & So, K. K. F. (2023). The evolution of service failure and recovery research in hospitality and tourism: An integrative review and future research directions. *International Journal of Hospitality Management*, 111, 103457. <https://doi.org/10.1016/j.ijhm.2023.103457>
- Knutson, B., Stevens, P., Wullaert, C., Patton, M., & Yokoyama, F. (1990). LODGSERV: A service quality index for the lodging industry. *Hospitality Research Journal*, 14(2), 277–284.
- Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- Li, M., Qiu, S. C., & Liu, Z. (2016). The Chinese way of response to hospitality service failure: The effects of face and guanxi. *International Journal of Hospitality Management*, 57, 18–29.
- Liu, S., Teichert, T., Rossi, M., Li, H., & Hu, F. (2021). To be similar or to be different? The effect of hotel managers' rote response on subsequent reviews. *Tourism Management*, 86, 104324. <https://doi.org/10.1016/j.tourman.2021.104324>
- Lopes, A. I., Dens, N., De Pelsmacker, P., & Malthouse, E. C. (2023). Managerial response strategies to eWOM: A framework and research agenda for webcare. *Tourism Management*, 98, 104739. <https://doi.org/10.1016/j.tourman.2023.104739>

- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.
- Sahaf, M. A., & Fazili, A. I. (2024). How do we measure and assess service failure and recovery? A hybrid review. *International Journal of Consumer Studies*, 48(1), e12974. <https://doi.org/10.1111/ijcs.12974>
- Smith, A. K., Bolton, R. N., & Wagner, J. (1999). A model of customer satisfaction with service encounters involving failure and recovery. *Journal of Marketing Research*, 36(3), 356–372.
- Sun, S., Tong, K. T. L., & Law, R. (2017). Chinese hotel guest perception of international chain hotels under the same hotel brand in different travel destinations. *Journal of Vacation Marketing*, 23(2), 172–188.
- Tax, S. S., Brown, S. W., & Chandrashekar, M. (1998). Customer evaluations of service complaint experiences: Implications for relationship marketing. *Journal of Marketing*, 62(2), 60–76.
- Tourism Malaysia. (2025). Launch of Visit Malaysia 2026 campaign [Media release]. Retrieved from Tourism Malaysia website.
- Wan, M. (2024). Language differences in online complaint responses between generative artificial intelligence and hotel managers. *Informatics*, 11(3), 66. <https://doi.org/10.3390/informatics11030066>
- Xie, K. L., Zhang, Z., Zhang, Z., Singh, A., & Lee, S. K. (2016). Effects of managerial response on consumer eWOM and hotel performance: Evidence from TripAdvisor. *International Journal of Contemporary Hospitality Management*, 28(9), 2013–2034. <https://doi.org/10.1108/IJCHM-06-2015-0290>
- Xu, F., & Zhao, Y. (2022). Polarity, diversity, and readability in online managerial responses: How do they affect repeat customers' satisfaction change? *International Journal of Hospitality Management*, 102, 103146. <https://doi.org/10.1016/j.ijhm.2022.103146>
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180–182. <https://doi.org/10.1016/j.ijhm.2008.06.011>
- Zhang, Y., La, X., Huang, Y., & Xie, K. L. (2024). Managerial response strategies to online reviews on Ctrip: Effects of uncertainty reduction and rapport management. *Tourism Management*, 103, 104911. <https://doi.org/10.1016/j.tourman.2024.104911>
- Zheng, T., Huang, X., Wu, Z., Sun, K., & Wang, B. (2023). Exploring hotel online review research: A scientometric analysis and agenda for future research. *Tourism Management Perspectives*, 47, 101101. <https://doi.org/10.1016/j.tmp.2023.101101>

**Appendix**

**Appendix Table A1: Category-Wise Chi-Square Tests And FDR-Adjusted p Values By Travel Type**

Incident category	$\chi^2$	df	FDR-adjusted p	Cramer's V	n_total	p_fdr
Process	108.71	6	3.8E-21	0.082	16246	3.04E-20
Facilities	89.37	6	4.1E-17	0.074	16246	1.64E-16
Cleanliness	80.84	6	2.39E-15	0.071	16246	6.38E-15
Location	75.82	6	2.6E-14	0.068	16246	5.21E-14
Noise/Sleep	50.14	6	4.4E-09	0.056	16246	7.04E-09
Safety	44.14	6	6.95E-08	0.052	16246	9.27E-08
Staff	22.38	6	0.001033	0.037	16246	0.00118
Billing/Value	9.85	6	0.131156	0.025	16246	0.131156
Facilities	108.71	6	4.1E-17	0.082	16246	1.64E-16

Note: This appendix table reports category-wise Pearson chi-square tests assessing whether the prevalence of each incident category differs across travel types. Because incident coding is multi-label at the review level, each test is conducted on a binary indicator of whether a review mentions the focal category, and results should be interpreted as differences in category-specific prevalence rather than mutually exclusive shares. Reported statistics include  $\chi^2$ , degrees of freedom (df), FDR-adjusted p values using the Benjamini and Hochberg procedure, and Cramér's V as an effect-size measure. The sample size is constant across tests (N = 16,246 low-rating reviews). Scientific notation is used for very small p values.

**Appendix Table A2: Chinese Dictionary Indicators And English Glosses (Ctrip Reviews)**

Incident domain	Example (Chinese)	indicators	English gloss
Billing/Deposit & Value	乱收费; 多收费; 扣费; 扣款; 收费不合理; 退费; 退款; 押金不退; 不退按金; 不值		unreasonable/extra charges; overcharged; fee deducted; unauthorized deduction; unreasonable charges; fee refund; refund; deposit not refunded; deposit not returned; not worth the price (poor value)
Cleanliness & Hygiene	脏; 卫生差; 不干净; 污; 污渍; 灰尘; 头发; 床单; 被套; 毛巾		dirty; poor hygiene/cleanliness; unclean; filthy; stains; dust; hair; bedsheet; duvet cover; towel
Maintenance & Facilities Failure	坏; 不能用; 故障; 漏水; 空调; 冷气; 不制冷; 热水; 没热水; 水压		broken; not working/unusable; malfunction; water leak; air conditioning; air conditioner (AC); AC not cooling; hot water; no hot water; water pressure
Location/Access & Environment	位置偏; 不方便; 交通不便; 周边; 环境差; 很乱; 脏乱; 停车; 停车难; 离		remote location; inconvenient; poor transport access; surroundings/nearby area; poor environment; messy/chaotic; dirty and

Noise & Sleep Disruption	吵; 噪音; 很吵; 隔音差; 睡不着; 无法入睡; 凌晨; 楼上; 隔壁; 施工	messy; parking; hard to park; far from (distance complaint) noisy; noise; very noisy; poor soundproofing; cannot sleep; unable to fall asleep; late night/early morning; upstairs; next door; construction noise
Check-in/out & Process	入住; 退房; 排队; 等很久; 等待; 手续; 押金; 按金; 规定; 政策	check-in; check-out; queue/line; long wait; waiting; procedures/paperwork; deposit; deposit; rules; policy
Safety & Security	不安全; 害怕; 治安差; 被偷; 偷窃; 丢东西; 锁不住; 门锁; 坏; 骚扰	unsafe; felt afraid; poor security/safety; items stolen; theft; lost belongings; cannot lock; door lock; broken; harassment
Staff Attitude & Courtesy	态度差; 服务差; 不耐烦; 冷漠; 敷衍; 爱答不理; 不理人; 骂; 吼; 凶	bad attitude; poor service; impatient; indifferent/cold; perfunctory; dismissive/ignoring; unresponsive/ignored; swore/scolded; shouted; rude/aggressive

*Note:* The table lists representative Chinese indicator terms/phrases used to flag each incident domain in the review-level dictionary, alongside English glosses for interpretation. Indicators were applied in a rule-guided, multi-label manner and support the incident prevalence and segment comparisons reported in the main text.