AIRBNB

Optimizing Airbnb Performance in New Zealand: Pricing, Segmentation, and Guest Insights

Leveraging unbalanced panel data, k-prototypes clustering, and lexicon-driven sentiment analysis to uncover price determinants, identify distinct market segments in Auckland and Queenstown, and quantify guest perceptions—delivering actionable strategies for revenue optimisation, targeted promotions, and enhanced guest satisfaction.

August 31, 2025

PRATIK GANGULI
University Of Auckland

Introduction *(Background Information, Research Questions)*

Airbnb has established itself as a major force in New Zealand's short term accommodation sector, offering a diverse range of options to both domestic and international travellers. For property owners and managers, the ability to identify price determinants, understand distinct market segments, and capture guest sentiment is essential for maintaining a competitive advantage. This project applies advanced marketing analytics to explore these factors in depth, focusing on insights that can directly support informed decision making by hosts and managers.

Data Overview

The analysis draws on data from ***InsideAirbnb*** for New Zealand, covering the period **July 2024 to June 2025**. The dataset contains **576,621 rows** and **86 columns**, equivalent to approximately **49.6 million recorded values**, with a low and manageable **1.16 percent** rate of missing values. The data is structured as **unbalanced & irregular panel data**, meaning that each property (panel unit) is observed at multiple points in time, but the number and timing of these observations differ across properties. Some listings appear frequently across the 12-month period, while others are only recorded once or a few times, and the intervals between observations are not consistent, enabling the observation of property-level changes over time. While the dataset covers the national market, the study focuses in greater detail on **Auckland** and **Queenstown**, the two largest markets in terms of Airbnb listing volume.

Research Literature Review

The literature review summarises key academic findings that inform the analytical approach of this study. It focuses on three areas directly aligned with our research questions: the determinants of Airbnb listing prices, the segmentation of properties in major markets, and the analysis of guest review sentiment. For each theme, prior research is examined to establish the relevance of the method, identify existing gaps, and justify its application in the New Zealand context. This ensures the study is grounded in established theory while addressing market-specific insights for Airbnb hosts and managers.

1. Panel Data Regression – Price Drivers

"Over a 12-month period, what are the primary drivers of Airbnb listing prices in New Zealand, and how can host characteristics, property features, and seasonal trends be leveraged to optimise pricing strategies?"

This is impactful because pricing directly affects host revenue. *Wang and Nicolau* (2017) found that these factors significantly influence prices, and seasonal patterns add further variation. Their work supports using panel data regression to capture both cross-sectional and time-based effects in the New Zealand market.

2. Segmentation – Auckland and Queenstown

"Which are the two most populated cities in New Zealand in terms of Airbnb listings, and how can their properties be segmented based on price, review scores, amenities, and other characteristics to identify distinct market clusters?"

Gunter and Önder (2018) applied clustering techniques to Airbnb data in Vienna to segment properties into market groups defined by location, price, and review characteristics. Their results highlight how segmentation can reveal differentiated markets with distinct competitive dynamics. Using a similar methodology for Auckland and Queenstown allows for the identification of property clusters that share common attributes, supporting targeted marketing and pricing strategies.

3. Sentiment and Emotion Analysis – Guest Reviews
"What key sentiments and emotions emerge from Airbnb guest reviews across New Zealand over a 12-month period, and how do they highlight the aspects of a stay that customers value most and those they find unsatisfactory?"

This is impactful because understanding both positive and negative feedback helps hosts identify strengths to maintain and weaknesses to address. *Xie et al.* (2014) demonstrated that online consumer reviews contain valuable sentiment and emotional indicators that influence customer decision-making and business performance. Applying this approach to Airbnb guest reviews in New Zealand provides actionable insights for improving guest satisfaction.

Through a combination of statistical modelling, market segmentation, and text analytics, the project delivers actionable insights designed to help Airbnb hosts refine pricing strategies, improve property offerings, and enhance guest satisfaction.

Research Question 1 (Panel Data Regression – Price Drivers)

Descriptive Overview (Statistical Summary)

The dataset combines twelve consecutive months of Airbnb listings, covering the period from **July to June**, and was merged into a panel structure that tracks the same properties over time. This format provides richer insights than a single-month snapshot, as it allows us to see how prices and property characteristics change month by month while controlling for unobserved differences between hosts, i.e. it refers to factors that influence prices but are not directly captured in the dataset, such as hosting style, property presentation, or local reputation. The panel was both **unbalanced and irregular;** many listings appeared intermittently, and not all had the same time gaps between observations. Since the raw panel was **unbalanced** with many listings appearing intermittently, we applied a filter to retain only those properties present in at least nine out of the twelve months. This ensured a more stable base for analysis, reducing noise from short-term or event-driven listings while preserving a large, representative sample.

Two main data quality issues were identified. First, certain variables such as review scores, structural attributes like beds and bathrooms, and even prices contained missing values. Second, some records risked duplication, where the same property appeared more than once in the same month. We removed duplicate property–month pairs and treated missing values using methods that preserved accuracy while maintaining realistic market behaviour.

The distribution of **nightly prices** revealed a heavily skewed market[Figure 1.15]. The midpoint price was about **NZD 203** per night, but the average rose to **NZD 394** due to a small number of luxury listings charging more than **NZD 100,000** per night. This skew was addressed by applying a **logarithmic transformation** to prices, which produced a more balanced, bell-shaped distribution and supported more robust modelling. This also improved interpretability, as percentage-based effects (e.g., "a 5% increase per extra bedroom") are easier to communicate across a wide range of prices.

Structural characteristics showed that the typical listing accommodated four guests in two bedrooms with one bathroom. While extreme values with dozens of rooms or beds were rare, they highlighted the diversity of the platform. **Price** gains from capacity followed a pattern of diminishing returns for example, adding a second bedroom had a notable effect, whereas adding a tenth had far less proportional impact. [Figure 1.17]of price by room type confirmed systematic differences, with entire homes commanding the highest rates, followed by private rooms and then shared rooms.

Review activity was highly uneven. Although the average listing had **68** lifetime reviews, the median had only **26**, i.e. meaning that half of all listings had **26** reviews or fewer. In any given month, many listings received no reviews, while a minority saw more than **20** per month. This concentration suggests that review frequency acts as a strong signal of popularity and trust, giving certain hosts greater pricing power.

Ratings were similarly concentrated at the top of the scale. Most listings scored between 4.8 and **5.0** across categories such as accuracy, cleanliness, and communication. While this leaves little statistical variation, even small shifts for example, from **4.8** to **4.9** can meaningfully influence guest decisions and support higher rates.

Availability patterns further highlighted host strategy differences. The median property was available for only **17 nights** in the next month, **61 nights** over the next three months, and 210 nights across the year. Seasonal patterns were visible in average monthly prices, with peaks in June and lower rates in off-season months such as February. This variation suggests that some hosts use scarcity to drive higher nightly rates, while others prioritise occupancy.

These EDA findings provide the foundation for our regression analysis, highlighting the variables likely to influence price, the importance of addressing skewed distributions, and the need to account for both structural and seasonal effects in the model.

Analytical Methodology & Insights

The descriptive analysis set the stage for the regression work by showing clear patterns in how listing characteristics, quality signals, and host behaviour interact to influence Airbnb pricing. Luxury outliers(extreme values) distorted the raw price distribution, with the median nightly rate at around **NZD 203** and the mean almost double at **NZD 394** due to a small number of extreme cases priced at over **NZD 100,000** per night. Transforming prices into log terms balanced the distribution into a bell curve and allowed results to be expressed in percentage changes much easier to interpret in a business context. Property type emerged as a fundamental baseline driver, with entire homes commanding the highest rates and private/shared rooms pricing much lower. Review activity revealed that while the average listing had **68** lifetime reviews, the median had only **26**, showing that a small number of highly active hosts dominate credibility and pricing influence. The near-perfect ratings across most listings (average overall **4.83**, with communication highest at **4.89** and value lowest at 4.75) meant that even a small **0.1** unit of improvement could meaningfully shift demand. Availability patterns suggested that the median property was open for only **17** nights in the next month, **61** over the next three months, and **210** across the year, indicating hosts strategically control calendars to either maximise occupancy or create scarcity to support higher rates.

Building on these insights, we set out to test five key hypotheses: that larger listings command higher prices but with diminishing returns, that better review scores, especially for accuracy and communication, lead to higher prices, that review activity and availability contribute to pricing power, that some effects are amplified or constrained when combined, and that **fixed effects** (FE) models would outperform **random effects** (RE) when unobserved factors such as location quality are correlated with observed drivers[figure1.3]. In simple terms, a **fixed effects model** controls for all the time-invariant characteristics of each listing, things like its neighbourhood, property type, or view, by effectively comparing each property to itself over time. This means it focuses only on changes within a listing, removing any influence from factors that do not change month to month. By contrast, a **random effects model** assumes that these unobserved, constant factors are unrelated to the variables we measure, such as review scores or number of bedrooms. This assumption allows **RE** to use both within-listing and between-listing variation, which can make it more efficient if the assumption holds true. But also, more biased if it does not. In a market like Airbnb, where location and other fixed traits are almost certainly linked to things like price, review scores, and occupancy, the **FE** approach is generally the safer and more reliable choice.

To model this, we used a **panel regression** framework on the cleaned and filtered dataset, with prices in log form to account for skewness and enable percentage-based interpretation. This approach allowed us to track the same listings over time, comparing each property to itself and stripping out fixed factors like location or property style that do not change month to month.

Two model specifications were tested: FE, which controls for these constant characteristics, and RE, which assumes unobserved factors are unrelated to the variables we measured. [Figure 1.2]The **Hausman test** decisively favoured the FE model ($\chi^2 = 17{,}643$, df = 18, $p < 0.001$), confirming that location quality and other unmeasured traits were indeed correlated with our key predictors, making **FE** the more accurate choice.

The results[Figure 1.3] aligned closely with both theory and the descriptive patterns. Structurally, each additional bedroom or higher guest capacity significantly raised nightly prices, but with diminishing returns early upgrades provided strong gains, but the impact flattened at higher capacity levels. Review scores, particularly for accuracy and communication, had a meaningful effect: a **0.1-**point boost in these ratings was worth between **5%** and **7%** more in price, a significant uplift in a market where nearly all ratings are already "excellent." Interestingly, higher value or check-in scores were linked to lower prices, suggesting that "good value" often signals budget positioning rather than premium quality.

Activity and availability also mattered. Listings with more total reviews and higher monthly review activity commanded higher prices, reinforcing the idea that visible popularity builds trust and justifies a premium. Availability effects were smaller but still positive properties open for more nights tended to charge slightly more, reflecting hosts' ability to match supply to high-demand periods. Interaction effects showed that quality signals mattered most for smaller properties, while size and popularity reinforced each other's effect only up to a point, after which the benefit levelled off.

The model explained **42% of the variation in prices within the same property over time** (Adjusted R^2 within = 0.42), which is a strong level of explanatory power for market data of this kind. In practical terms, this means the model captures almost half of the month-to-month price changes that occur for a given listing. **Overall significance** was confirmed by the F-test ($p < 0.001$), indicating that the combined set of factors meaningfully predicts price rather than doing so by chance.

Model fit statistics also strongly favoured the **Fixed Effects (FE)** approach over **Random Effects (RE)**, with **AIC = −1,373,352 vs −738,185** and **log-likelihood = 68,695 vs 36,927**. Lower AIC and higher log-likelihood values mean the FE model explains the data better and does so more efficiently.

For managers and hosts, the takeaway is straightforward: **price is shaped by property size, perceived quality, visibility through reviews, and strategic availability management**. The returns from improvements are **not linear** meaning the first few upgrades (like adding an extra bedroom or improving cleanliness scores) bring larger gains, while later ones add progressively less. This makes it important to target investments where they will have the greatest pricing impact.

Managerial Insights

For hosts, the results point to three clear strategies for maximising revenue. First, size upgrades deliver the biggest gains early on: moving from one to two bedrooms brings a noticeable price lift, but further expansion adds less value. Second, quality matters: even small improvements in communication or accuracy scores can justify higher prices, especially in a market where most listings already score highly. Third, consistent review activity builds trust: maintaining a steady flow of guest feedback supports higher pricing power. Availability should also match business goals limiting nights can help sustain higher rates, while keeping calendars open longer can maximise occupancy.

For Airbnb as a platform, the findings highlight the importance of identifying truly exceptional hosts, as inflated review scores mean that small differences can have a big impact on guest decisions and pricing. The analysis gives confidence that these strategies are based on strong and reliable patterns in the data.

Research Question 2 (Segmentation— Auckland & Queenstown)

Descriptive Overview (Statistical Summary)

We scoped the segmentation by first identifying **where the market is strongest**. Using the *InsideAirbnb* NZ data (Jul 2024–Jun 2025), we ranked parent regions by total listings and found a clear concentration in **Auckland** and **Queenstown-Lakes**. This is visible in **[Figure 1.4]**, where Auckland sits at the top (113,268 city-month rows; **Auckland Unique Listings: 13,944**) and Queenstown follows (65,693 rows; **Queenstown Unique Listings: 7,258**). Given Auckland's role as the country's most populated urban market and Queenstown's position as the most tourism-driven destination, these two regions jointly capture the bulk of national supply and represent distinct demand contexts. That's why we anchored the segmentation on Auckland and Queenstown from the start. Across July 2024 to June 2025, we observed missing data in both markets, with **153,332 missing** fields in Auckland and **60,394** in Queenstown, indicating gaps in some listing attributes and review metrics.

Over July 2024 to June 2025, the **average monthly price** line shows **Queenstown** consistently above **Auckland** throughout the year [Figure 1.5]. Queenstown also has a clear **peak in December**, which likely reflects the summer holiday tourism surge. By contrast, Auckland's prices **crest around October**, a pattern that may relate to spring events, relocations and new roles starting, and other city activity ahead of summer. These distinct peaks suggest **different seasonal demand profiles**: Queenstown behaves like a classic tourist market, while Auckland reflects a large metropolitan market with its own timing. Complementing this, the **boxplots by region** indicate that **most listings sit in the mid-range**, but **both cities include a set of high-priced listings** that stand well above the rest [Figure 1.6]. This points to a **higher-end segment** present in each market. Each city has a different typical nightly price, with Queenstown higher than Auckland.

In summary, we focused on Auckland and Queenstown because they hold the largest share of national listings and represent New Zealand's biggest urban market and its leading tourist destination. Across the year, Queenstown is priced above Auckland and peaks in December, while Auckland peaks around October two distinct demand patterns. Box-plot evidence also reveals a small set of very high-priced listings in both cities. Through our analysis, we aim to identify and separate the small group of very high-priced listings in both cities the premium tier from the other segments.

Analytical Methodology & Insights

To prepare the data for segmentation, we first aggregated listings by **ID** to create a complete view of each property over time. The missing values identified in several attributes were retained to avoid losing potentially important information. Instead of removing them, we imputed the gaps using averages for numbers and the most common value for categories, all within each listing so individual trends weren't distorted. We also added new features to provide more context for the analysis, such as description length, which reflects how detailed a host's listing is, and amenity count, which indicates the range of facilities offered and can influence guest preferences. These steps enhanced our ability to assess listing quality and host behaviour during segmentation.

After cleaning the data, we proceeded with the segmentation stage and selected the K-prototypes model. This approach was chosen because our dataset contained a mix of both numerical and categorical variables, and **K-prototypes** is well-suited for handling this combination. To measure dissimilarity between listings, we used the **Gower distance** metric, which accommodates mixed data types by normalising numeric values and treating categorical differences appropriately. It allows the segmentation process to consider differences in numeric values alongside distinctions in categories, producing clusters that reflect the full range of listing characteristics.

To begin the segmentation for Auckland, we first selected a set of variables that capture the most important aspects of each listing. These included the type of room offered, the number of guests a property can accommodate, its nightly price, and the range of amenities provided. We also considered the minimum nights required for a booking, the length of the property description as an indicator of how much detail the host provides, and the number of reviews in the last twelve months as a measure of recent guest activity. We tested various models with different numbers of clusters and, after evaluating the results, finalised the segmentation with three clusters for the Auckland listings.

In Auckland[Figure 1.7], our clustering produced three well-defined market segments, with a silhouette score of 0.57, which suggests[Figure 1.19] that the groups are reasonably well-separated and meaningful for interpretation.

The first segment, **Family-Friendly Homes**, includes entire homes and apartments typically accommodating around four guests. They average about **NZD 327** per night, offer the highest number of amenities, and have rich descriptions, making them appealing for family stays. These listings also attract a healthy number of reviews, showing strong demand.

The second segment, **Budget Private Rooms**, primarily consists of private rooms for around two guests, with an average nightly rate of **NZD 111**. These listings offer fewer amenities but remain popular, as indicated by steady review activity, appealing to cost-conscious travellers.

The third segment, **Ultra-Luxury Retreats**, is a niche category of rare, high-end entire homes with extraordinary average nightly prices of nearly **NZD 35,000**. These listings tend to have shorter descriptions, fewer reviews, and lower average lengths of stay. This could be because their target customers often high-net-worth individuals or exclusive groups, already have a clear understanding of the luxury standard expected, making lengthy descriptions unnecessary. Their stays may be shorter due to the premium pricing. This highlights a small but distinct ultra-premium tier alongside Auckland's broader family-oriented and budget-focused market.

To begin the segmentation for Queenstown, we selected variables that capture the most relevant aspects of each listing. These included the type of accommodation offered, nightly price, number of amenities provided, whether the property is instant bookable, indicators of host quality such as **Superhost** status and acceptance rate, the detail level in the property description, and the number of recent reviews over the last twelve months. This mix of features allows us to assess both the quality of the stay and the professionalism of the host. We tested several clustering solutions and settled on five clusters[Figure 1.8 & Figure 1.20], each with unique pricing, amenities, and host characteristics. The silhouette score of 0.46 indicates that the groups are well-defined and suitable for interpretation.

The **first segment**, **Reliable Comfort**, is made up of entire homes hosted by Superhosts. These properties average around **NZD 536** per night, have the highest average amenity count among all clusters, and receive a strong volume of recent reviews. This combination of high service quality, ample amenities, and proven guest satisfaction makes them ideal for families or groups looking for a well-equipped, trustworthy stay.

The **second segment**, **Standard Hosts**, also offers entire homes, but they are typically managed by non-Superhosts. They average around **NZD 472** per night and provide fewer amenities and receive fewer reviews compared to the Reliable Comfort segment. These listings still serve the mid-market traveller well but may reflect hosts who are less experienced or less responsive, resulting in slightly lower service consistency.

The **third segment**, **High-Value Superhosts**, is a premium category consisting of entire homes that are both hosted by Superhosts and instant bookable. With an average nightly price of around **NZD 665**, these properties blend high quality with booking convenience. They feature high host acceptance rates and receive excellent guest feedback, appealing to travellers who value both assurance and ease of booking.

The **fourth segment**, **Budget-Friendly Private Rooms**, offers private room accommodation at an average of **NZD 182** per night. Interestingly, these listings are still hosted by Superhosts, indicating that high service quality is maintained even in the budget segment. These properties cater to solo travellers, backpackers, or budget-conscious guests who still value reliable service.

The **fifth segment**, **Signature Stays**, represents Queenstown's ultra-premium tier. These are rare, high-end entire homes with extraordinary nightly rates averaging around **NZD 36,000**. They have fewer amenities compared to mid-range and premium clusters, shorter descriptions, and minimal review activity. This is likely because their target market high-net-worth individuals or exclusive groups already understands the luxury standard expected. Bookings may be for shorter stays due to the high price point, or for specific high-profile events. This mirrors the ultra-luxury segment identified in Auckland, showing that both markets cater to a small but distinct high-tier audience.

Managerial Insights

In Auckland, **Family-Friendly Homes** can be positioned for longer family stays through targeted seasonal packages and promotions that emphasise spacious layouts, high amenity counts, and positive reviews. **Budget Private Rooms** demonstrate steady demand and should focus on competitive pricing while maintaining strong review scores to attract budget-conscious travellers. The **Ultra-Luxury Retreats** niche, although small, represents a high-value opportunity that can be maximised through partnerships with luxury travel agencies, concierge networks, and event planners catering to high-net-worth individuals.

In Queenstown, **Reliable Comfort** and **High-Value Superhosts** should leverage their strong service quality, premium pricing, and convenience factors such as instant booking to appeal to discerning tourists. **Standard Hosts** can increase competitiveness by enhancing amenities, improving responsiveness, and building stronger review profiles. **Budget-Friendly Private Rooms**, while lower in price, can highlight their Superhost quality to attract solo travellers and budget-conscious guests seeking reliable stays. The **Signature Stays** ultra-premium tier, like Auckland's luxury segment, should focus on exclusivity, privacy, and bespoke experiences tailored for elite clientele.

In summary, Auckland shows a simpler Airbnb structure with three clear segments family homes, budget rooms, and a small ultra-luxury niche, while Queenstown has a more tiered market with five well-defined segments, ranging from affordable private rooms to instant-bookable homes and ultra-luxury exclusives. This diversity makes Queenstown a richer and more layered market, whereas Auckland remains more straightforward and primarily budget- or family-driven. Across both cities, the presence of clear premium and budget tiers highlights the need for differentiated pricing, targeted promotions, and tailored service offerings for each segment to maximise occupancy and revenue potential year-round.

Question 3 (Sentiment & Emotion Analysis)

Descriptive Overview (Statistical Summary)

To understand customer sentiments and emotions toward listings, we analysed guest reviews, as they provide valuable first-hand insights into guest experiences and perceptions. Unlike quantitative listing data, reviews capture what guests specifically liked, disliked, and the emotions they felt during their stay, revealing strengths to maintain, weaknesses to address, and the overall tone of guest experiences. Analysing the volume and timing of reviews also helps identify patterns in guest engagement and how they may align with seasonal price trends. This makes reviews not only a source of feedback for service quality but also a potential indicator of demand cycles across the year.

To begin the sentiment analysis, we first conducted exploratory data analysis (EDA) on the Airbnb guest review dataset. The raw dataset contained **33,948,195 reviews** from 2011 to 2025. The date column was initially stored as a string, so we parsed it into a proper date format to enable accurate filtering and time-based analysis. For this study, we focused only on reviews from **July 2024 to June 2025** to align with the analysis period used in our other research questions. This filtering ensured that the sentiment results were directly comparable with our pricing and segmentation findings.

During data inspection, we identified **693 missing review texts**. Since these records lacked the core content required for sentiment analysis and could not be meaningfully imputed, they were removed from the dataset. After cleaning, the remaining dataset provided a robust and comprehensive base for understanding guest perceptions over the selected 12-month period.

When examining review volumes over the July 2024 to June 2025 period, we observed high activity from July through January [Figure 1.9], with a pronounced peak in November at around **510,000 reviews**. This trend aligns with the summer holiday season in New Zealand, when both domestic and international tourism activity is at its highest. From January onward, review counts began to decline, with a sharper fall from March through June, indicating reduced guest engagement during the off-peak season. This seasonal pattern suggests that guest review activity closely mirrors broader tourism demand cycles, providing an indirect signal of Airbnb usage trends across the year.

Analytical Methodology & Insights

After completing the EDA, we focused our analysis on the sentiments expressed in reviews from July 2024 to June 2025. To achieve this, we applied **lexicon-based** text analysis, which uses predefined word lists to classify text into sentiment or emotion categories. For sentiment classification, we used the **Bing** lexicon, which categorises words as positive or negative. For emotion detection, we applied the **NRC** lexicon, which maps words to ten primary emotions such as joy, trust, and anger, alongside positive and negative sentiment categories.

We began by preparing the review text for analysis, ensuring that both the review ID and listing ID were stored as character data to preserve their full values without risk of truncation. The review comments were then broken down into individual words through tokenisation, a process that transforms free-text reviews into a structured format suitable for sentiment and emotion classification. This allowed us to analyse the frequency and distribution of words across all reviews in the dataset.

For the sentiment analysis using the Bing lexicon, we matched the tokenised review words with the lexicon's predefined positive and negative word lists. This allowed us to count the number of positive and negative words in the dataset and calculate an overall sentiment score.

The Bing sentiment analysis showed a strong positive skew in guest feedback. From [Figure1.10; Figure1.11] across all reviews from July 2024 to June 2025, we identified approximately **16.41 million positive words** and **0.98 million negative words** out of a total of **17.39 million sentiment-bearing words**, resulting in a total net sentiment score of **15.43 million**. This means **94.38%** of sentiment-bearing words were positive, while only **5.62%** were negative, giving a net positivity rate of **88.76%**.

On average, each review contained around **23.7 positive words** and only **1.41 negative words**, reinforcing the overall positive tone of guest experiences [Figure1.11]. These findings suggest that, for the period analysed, Airbnb guests in New Zealand expressed overwhelmingly favourable sentiments toward their stays, with negative feedback representing only a small fraction of the overall conversation.

For the emotion analysis, we applied the **NRC lexicon** to the tokenised review text. This lexicon assigns each matched word to one or more of ten primary emotions joy, trust, anticipation, surprise, sadness, fear, anger, and disgust as well as overall positive and negative sentiment categories. We joined the tokenised words with the NRC dictionary, counted the occurrences of each emotion across all reviews, and calculated the percentage each emotion represented out of the total sentiment-bearing words. This allowed us to quantify the emotional tone of guest feedback at scale.

Across all reviews[Figure 1.12 & Figure 1.18], the most dominant category was **positive sentiment** (29.99%), followed by **trust** (19.17%), **joy** (18.84%), and **anticipation** (13.19%). Together, these top four categories accounted for over 80% of all emotion-related words, reflecting guest feedback that was largely positive, confident, and forward-looking. Negative emotions were far less prevalent, with **surprise** at 7.24%, **sadness** at 4.62%, **fear** at 1.49%, **anger** at 1.11%, and **disgust** at 0.82%.

When viewed alongside the Bing sentiment analysis, these results indicate that guest reviews during the analysis period were overwhelmingly favourable, with only a small share expressing strong negative emotions.

To understand what customers liked, disliked, and felt about their stays, we used word clouds to visually highlight the most common terms in guest reviews, offering an intuitive view of recurring themes. The positive word cloud [Figure 1.13] prominently featured terms such as *great*, *lovely*, *clean*, *comfortable*, *beautiful*, and *recommend*, reflecting guest appreciation for quality, cleanliness, comfort, and visual appeal. In contrast,[Figure 1.13] the negative word cloud highlighted words like *cold*, *issues*, *noise*, *problem*, and *unfortunately*, pointing to frequent concerns related to temperature, maintenance, and noise. These visualisations complement the quantitative findings by clearly showing the specific aspects of the guest experience that shaped both positive and negative perceptions.

Managerial Insights

In conclusion, Guest reviews from July 2024 to June 2025 reveal that Airbnb stays in New Zealand were overwhelmingly well-received, with positive words and emotions dominating feedback. Guests consistently valued quality, cleanliness, comfort, and overall presentation, while negative feedback was rare and mostly related to temperature issues, maintenance concerns, and noise. Seasonal peaks in review activity suggest demand is highest in the summer holiday period, offering opportunities for targeted pricing and promotions. The strong presence of trust and joy in reviews indicates that maintaining high service standards and delivering memorable experiences can further strengthen guest satisfaction and loyalty.

Limitations

Concerns and Limitations of the Applied Analysis Techniques:
In our analysis, the applied techniques provided useful insights but also had limitations that affected precision and depth.

Unbalanced and Irregular Panel Structure:
The dataset was inherently unbalanced and irregular, which is common in Airbnb market data but posed analytical challenges. Working with unequal observation periods and irregular spacing risked both data loss and reduced precision in estimates. To manage this trade-off, we chose to filter the dataset to include only listings with more than nine months of data. This approach reduced inconsistencies while retaining enough observations to capture meaningful time effects and dynamic trends, though it necessarily excluded some shorter-duration listings.

Fixing ID Formatting Issues:
To address potential identifier issues caused by RStudio's automatic conversion of large numeric IDs into **scientific notation**, we converted IDs into character format at the point of data import and, where necessary, extracted them directly from listing **URLs**.

Cluster Model Selection Without the Elbow Method:
In the k-prototypes segmentation, we could not use the traditional elbow method for determining the optimal number of clusters. The elbow method relies on within-cluster variance, which works well for purely numerical datasets. However, k-prototypes uses a combined distance metric, Euclidean for numeric variables and matching dissimilarity for categorical variables, making the scales incompatible. This can result in unstable or misleading elbow plots. Instead, we relied on silhouette scores tailored for mixed data, which provided a more reliable measure of cluster separation and interpretability.

Lexicon-Based Sentiment Analysis Limitations:
For sentiment analysis, we applied lexicon-based methods using the Bing and NRC dictionaries. While effective for large-scale text classification, lexicons have limitations compared to richer, context-aware NLP models. They do not account for multi-word expressions or context, leading to potential misclassification. For example, the phrase ***"to die for"*** conveys strong positive sentiment, but the word ***"die"*** alone is scored negatively in the lexicon. Such cases required additional handling through **stop word** removal, but this inevitably limited the depth of sentiment interpretation. More advanced context-sensitive methods, such as transformer-based models, could improve future analysis by better capturing nuance and idiomatic expressions.

Conclusion

Over the 12-month period analysis, the pricing model revealed that the most effective revenue gains come from targeted improvements rather than broad changes. Adding an extra bedroom or increasing guest capacity early on delivers a clear uplift in nightly rates, but returns diminish at larger scales, meaning investment should focus where it drives the most value. Service quality indicators such as communication and accuracy ratings were shown to have a direct impact on price, even a small **0.1 units** increase in these scores can support a higher rate in a market where most properties already score highly. Review activity also matters: a steady flow of recent reviews builds credibility and trust, which can translate into stronger pricing power. Hosts should also be mindful of how they manage their calendars restricting availability can sustain higher nightly rates during peak periods, while keeping more dates open can help maximise occupancy in off-peak months.

In **Auckland** and **Queenstown**, segmentation identified distinct market tiers, from budget-friendly private rooms to ultra-premium properties charging tens of thousands per night. This shows that different guest segments value different combinations of amenities, price points, and service levels. Aligning Airbnb property's **pricing, amenities**, and marketing message with the segment you serve will help you compete more effectively. The sentiment and emotion analysis confirmed that guests consistently value cleanliness, comfort, quality, and overall presentation, with positive emotions like **trust** and joy dominating feedback. **Negative** reviews were rare and typically focused on **temperature**, **maintenance**, or **noise**, small but fixable issues. For hosts, the opportunity is to combine these insights: know which segment you operate in, maintain the service qualities guests care about most, proactively address common complaints, and adjust pricing and availability in line with **seasonal peaks**. This integrated approach can help maximise both revenue and repeat bookings year-round.

Reference

Gunter, U., & Önder, I. (2018). Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics, 24*(3), 270–293. https://doi.org/10.1177/1354816617726983

Inside Airbnb. (n.d.). *Get the data*. Retrieved August 31, 2025, from https://insideairbnb.com/get-the-data/

RStudio Team. (2023). *RStudio: Integrated development environment for R* (Version 2023.06.1). Posit Software, PBC. https://posit.co

Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management, 62*, 120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management, 43*, 1–12. https://doi.org/10.1016/j.ijhm.2014.07.007

Appendix

Figure 1.1 Monthly Average Airbnb Prices (Jul 2024 – Jun 2025)

Hausman Test — Fixed vs Random Effects

Test	*Chi.square*	*df*	*p.value*	*Decision*
Hausman (FE vs RE)	17643	18	<1e-16	FE preferred (RE inconsistent)

Figure 1.2 Hausman Test Results: Fixed Effects vs Random Effects

Predictors	Fixed Effects (Two-Ways) Estimates	std. Error	Statistic	*p*
accommodates	0.040	0.007	5.775	**<0.001**
review scores accuracy	0.075	0.017	4.417	**<0.001**
review scores checkin	-0.021	0.006	-3.671	**<0.001**
review scores communication	0.092	0.015	6.256	**<0.001**
number of reviews	-0.000	0.000	-6.251	**<0.001**
number of reviews l30d	-0.002	0.000	-5.201	**<0.001**
bedrooms	0.011	0.003	3.310	**0.001**
availability 90 × number of reviews l30d	-0.000	0.000	-6.562	**<0.001**
accommodates × review scores accuracy	-0.005	0.001	-3.465	**0.001**
review scores accuracy × review scores communication	-0.017	0.003	-5.170	**<0.001**
accommodates × reviews per month	-0.001	0.000	-3.622	**<0.001**
number of reviews × bedrooms	-0.000	0.000	-3.003	**0.003**
Observations	466400			

Figure 1.3 Fixed Effects Regression Results for Airbnb Price Drivers

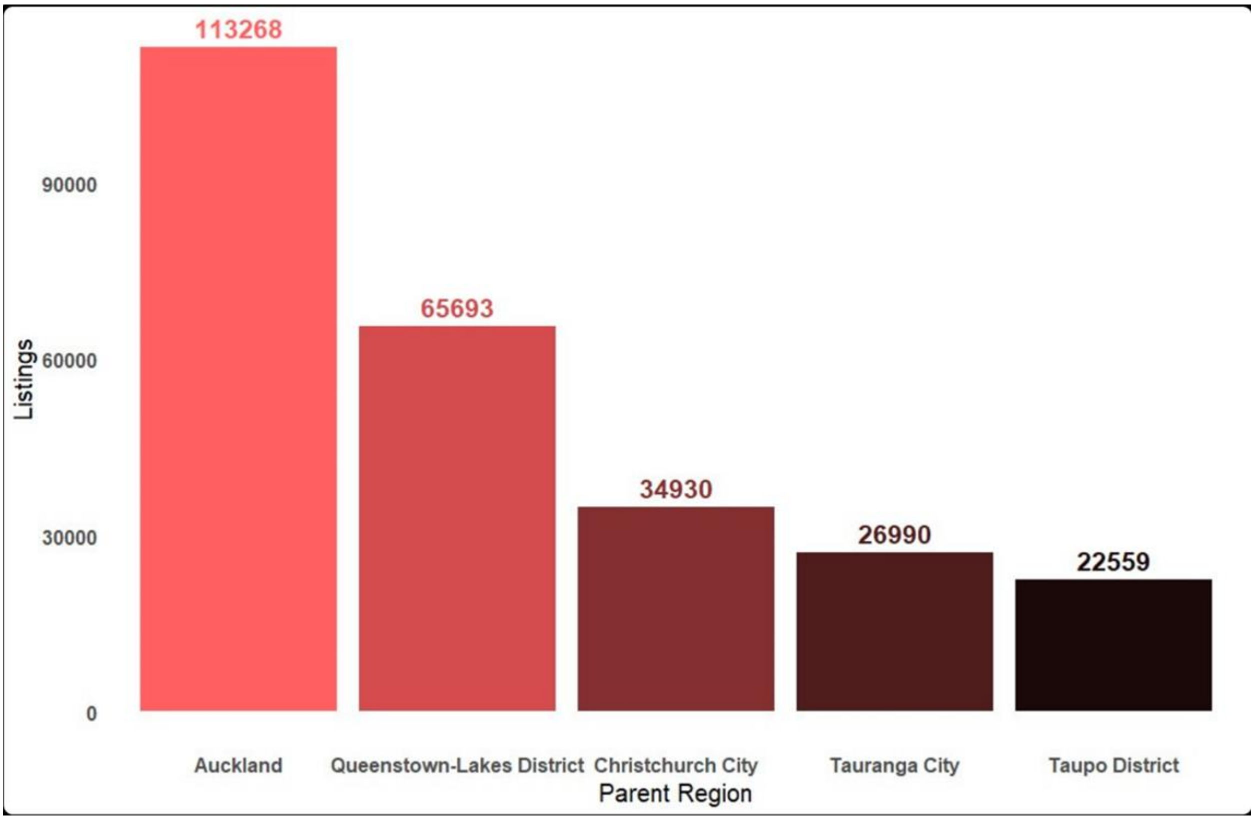

Figure 1.4 Top Five New Zealand Regions by Airbnb Listings (Jul 2024–Jun 2025)

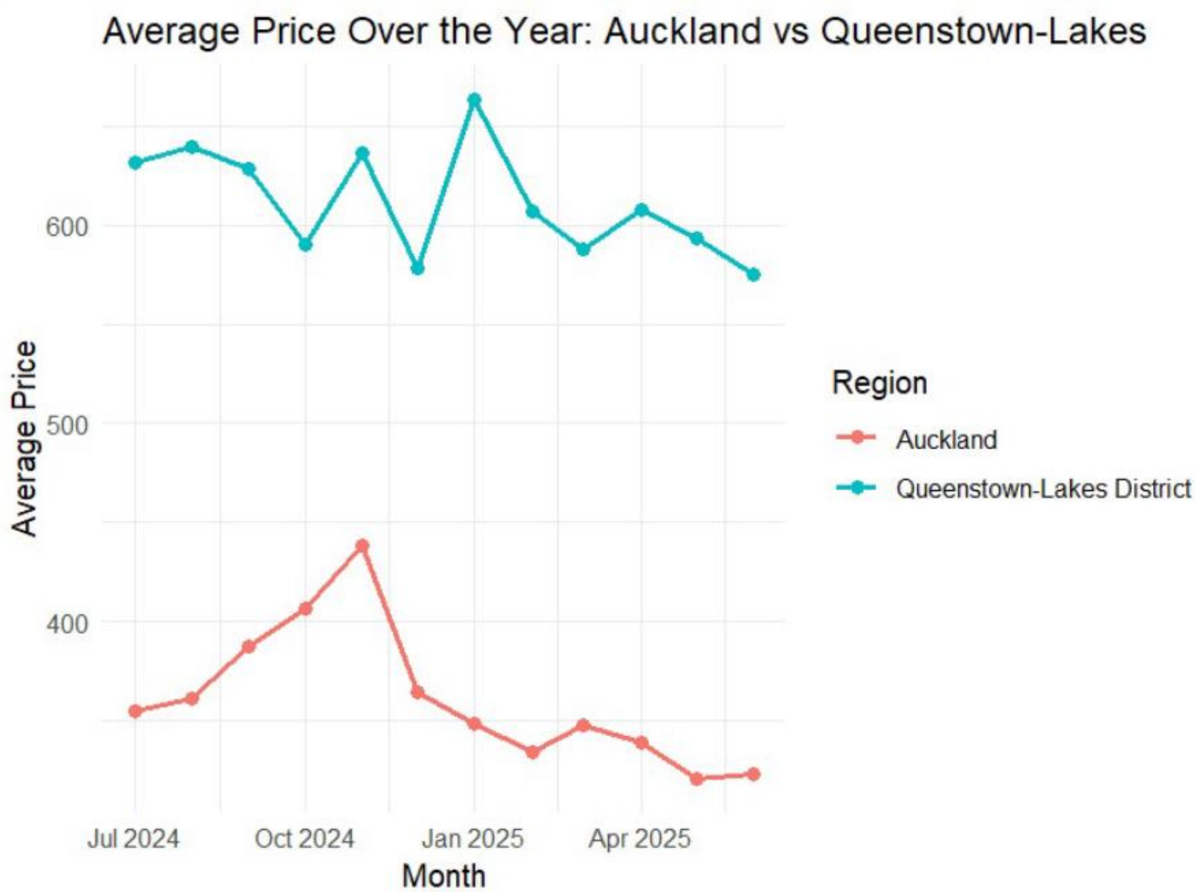

Figure 1.5 Seasonal Price Trends in Auckland and Queenstown-Lakes (Jul 2024–Jun 2025)

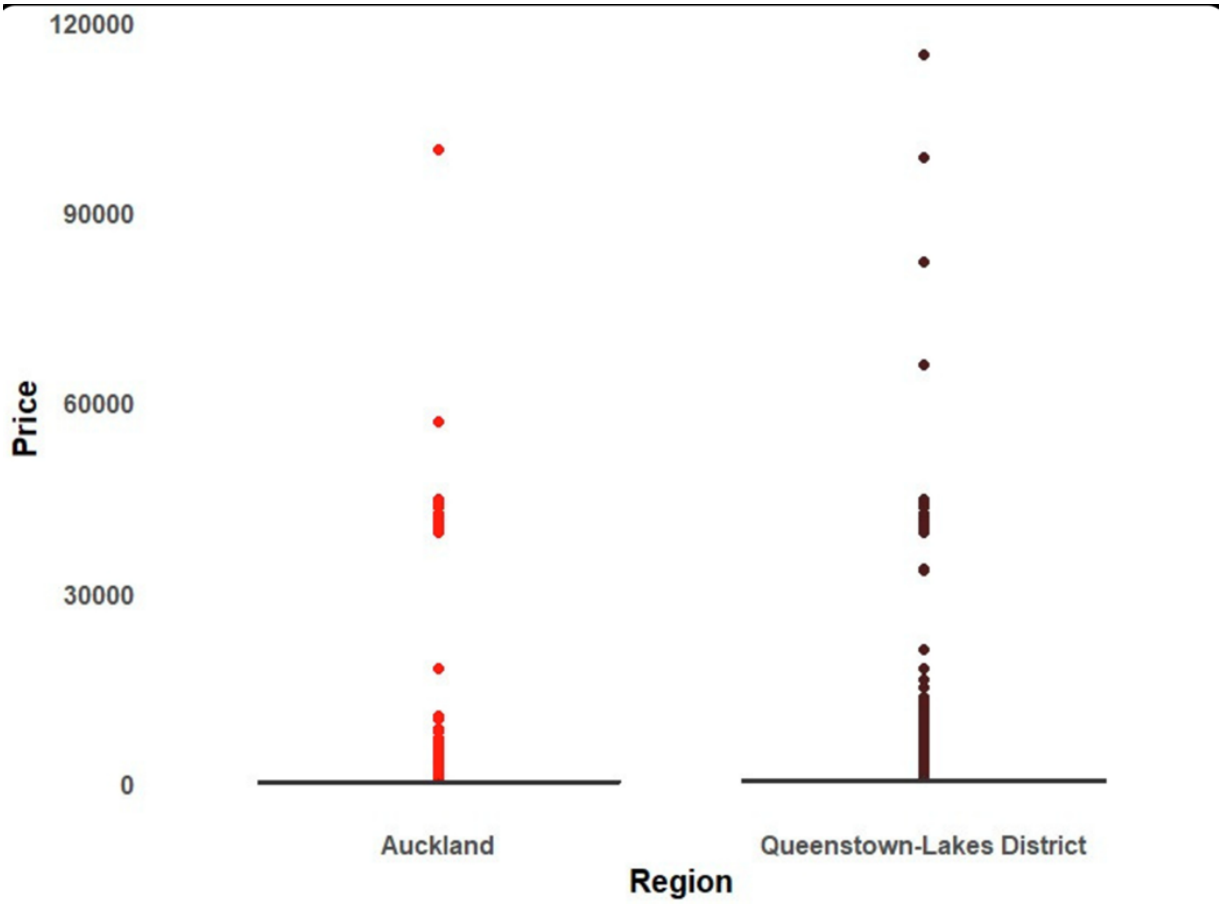

Figure 1.6 Nightly Price Distribution — Auckland vs Queenstown-Lakes

CLUSTER	SIZE	ROOM TYPE	ACCOMODATES	PRICE	AMENITIES	DESC_LENGTH	NO. OF REVIEWS	MIN_NIGHTS
Family- Friendly Homes	9935	Entire Home	~4	~327	~37	400	~10	~3
Budget Private Rooms	3974	Private Room	2	~111	~26	~349	~6.11	~2
Ultra- Luxury Retreats	35	Entire Home	5	~35000	~24	~191	~1	~1

Figure 1.7 Auckland Airbnb Market Segments — Key Characteristics

CLUSTER	SIZE	ROOM TYPE	PRICE	SUPERHOST	INSTANT BOOKABLE	NO. OF REVIEWS	AMENITIES _COUNT	DESCRIPTIO N_LENGTH	HOST_ACCEP TANCE_RATE
Reliable Comfort	2057	Entire Home	~536	Yes	No	~18	44	~385	~94
Standard Hosts	1970	Entire Home	472	No	No	~5	~35	~386	83
High Value Superheats	2390	Entire Home	~665	Yes	Yes	~15	42	~390	~98
Budget-Friendly Private Rooms	824	Private Room	~182	Yes	No	~16	29	~388	~88
Signature Stay	17	Entire Home	~36,3 50	No	Yes	~1	26	~166	~99

Figure 1.8 Queenstown Airbnb Market Segments Key Characteristics

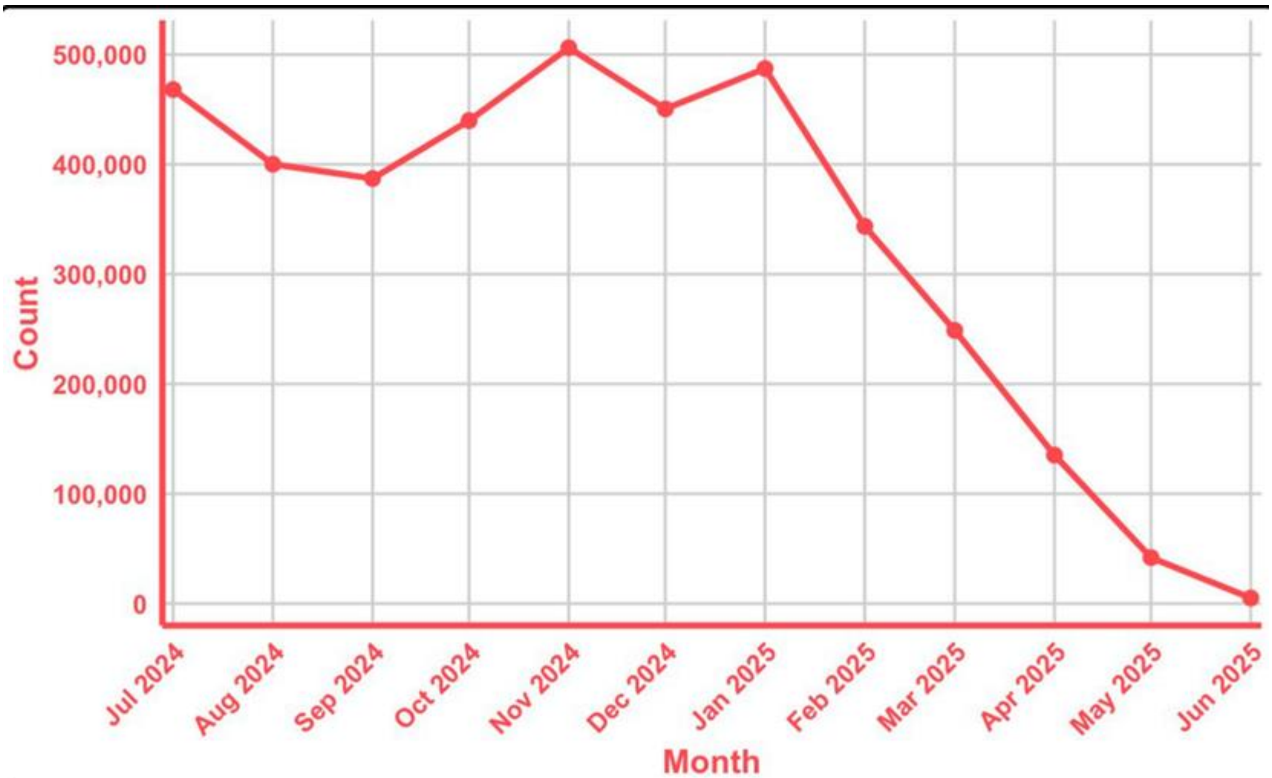

Figure 1.9 Monthly Airbnb Guest Review Counts in New Zealand

Total Positive Words	Total Negative Words	Total Words	Percentage Positive
16411762	977465	17389227	94.4

Figure 1.10 Summary of sentiment word counts.

Percentage Negative	Net Positive Word(%)	Average Positive Words	Average Negative Words
5.62	88.8	23.7	1.41

Figure 1.11 Summary of sentiment percentages and averages.

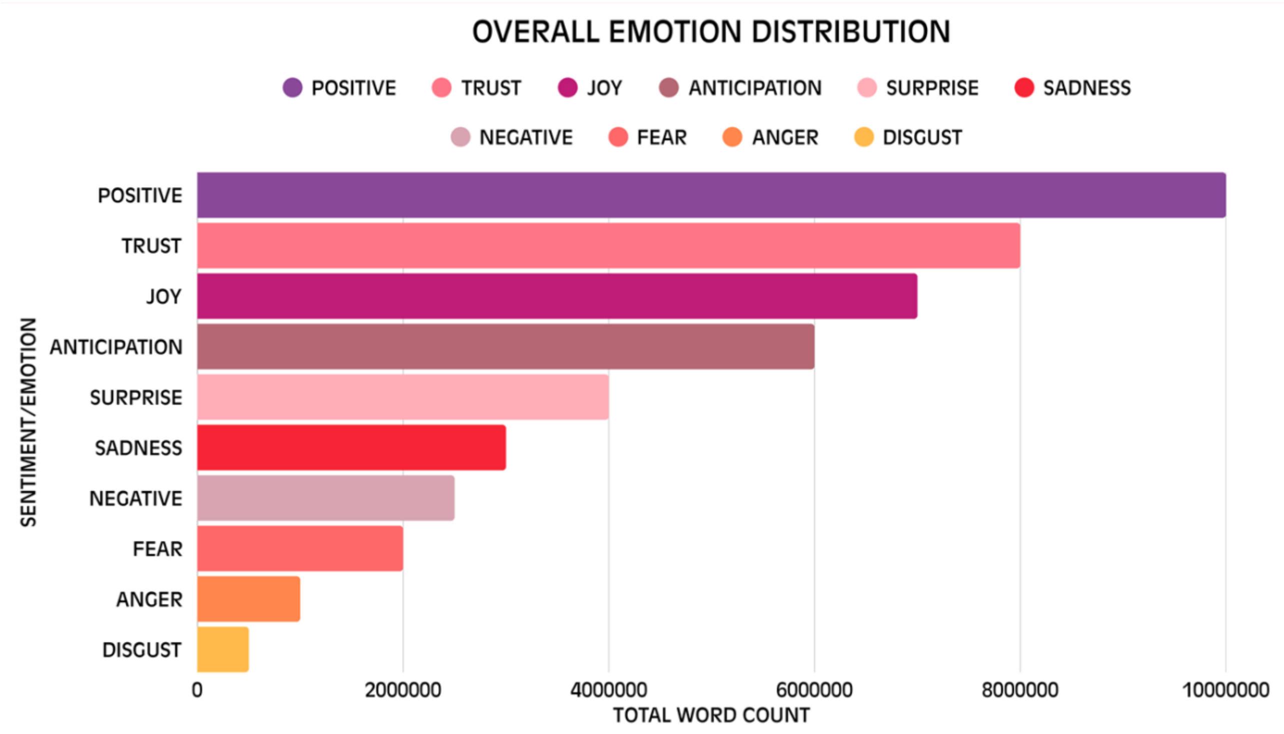

Figure 1.12 Overall distribution of emotions in Airbnb guest reviews.

Figure 1.13 Word cloud of the most frequent positive terms in Airbnb guest reviews.

Figure 1.14 Word cloud of the most frequent negative terms in Airbnb guest reviews.

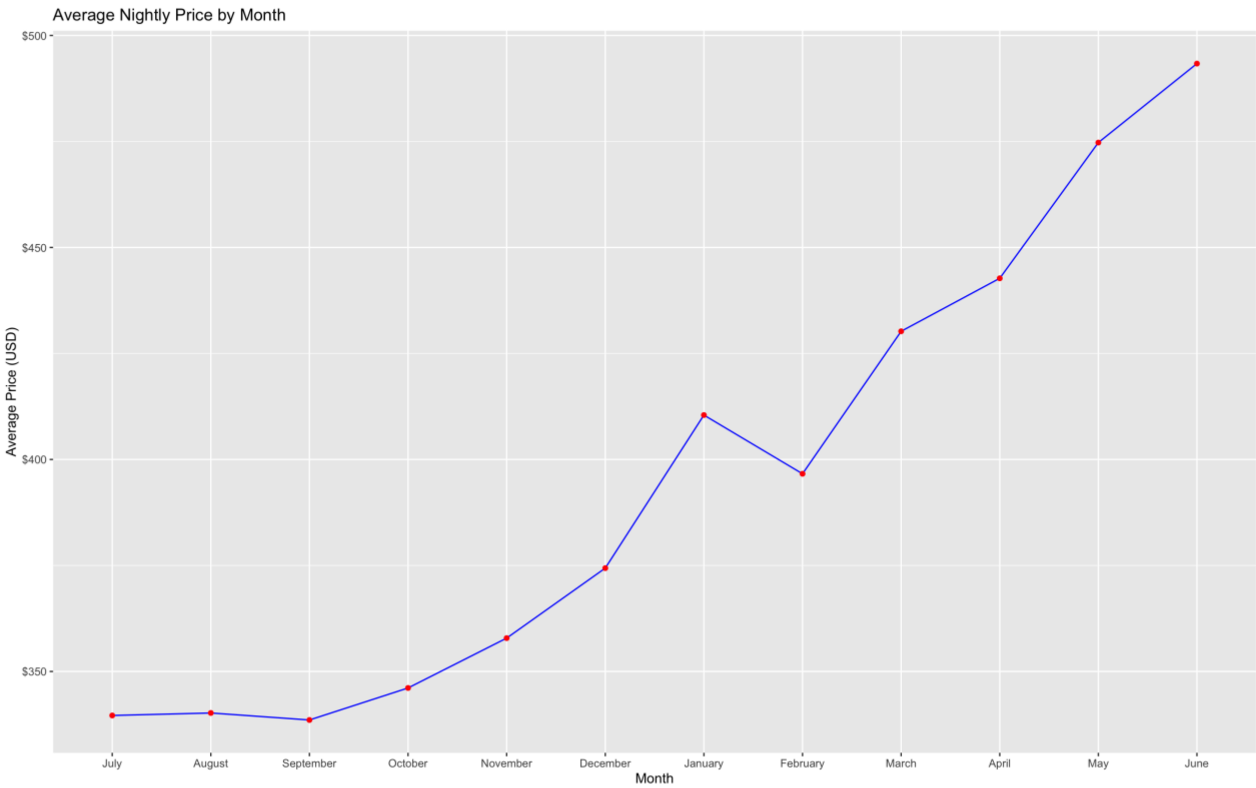

Figure 1.15 Average nightly price trends by month.

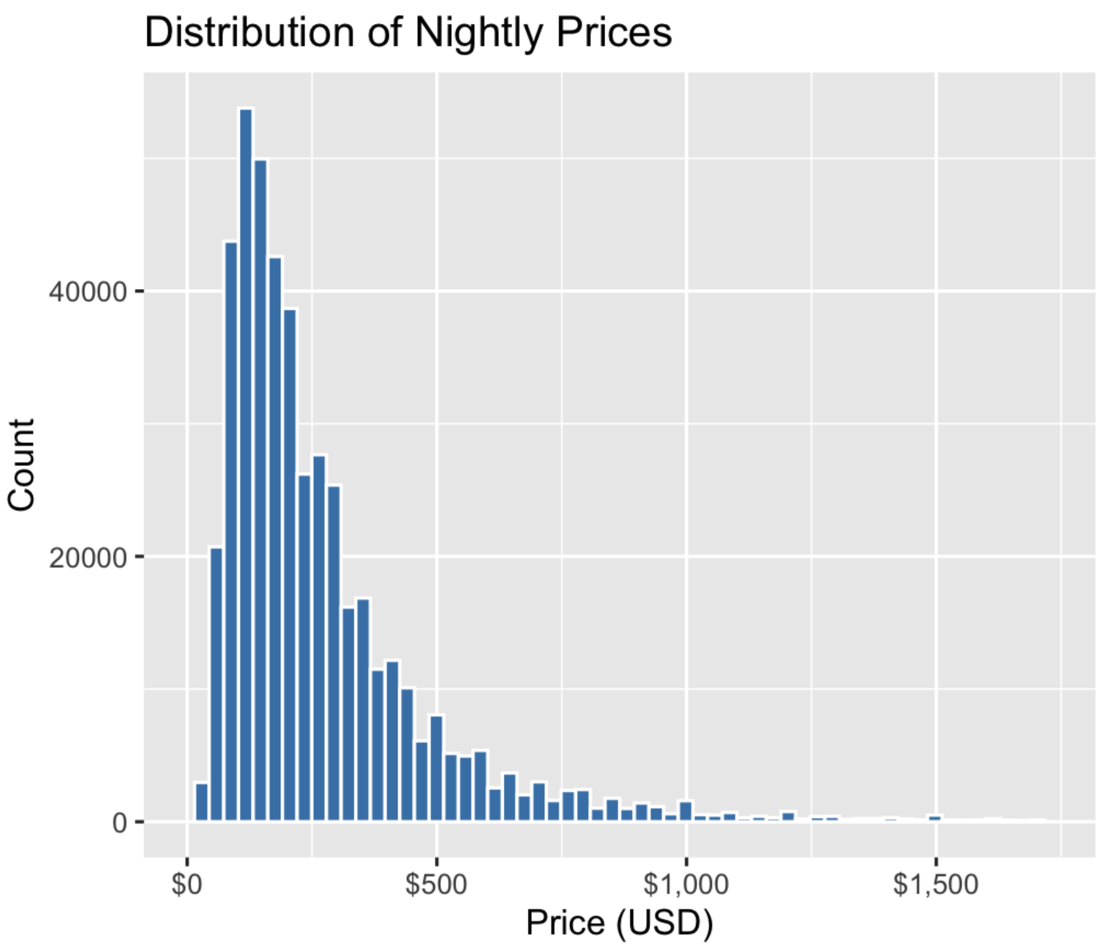

Figure 1.16 Distribution of Airbnb Nightly Prices

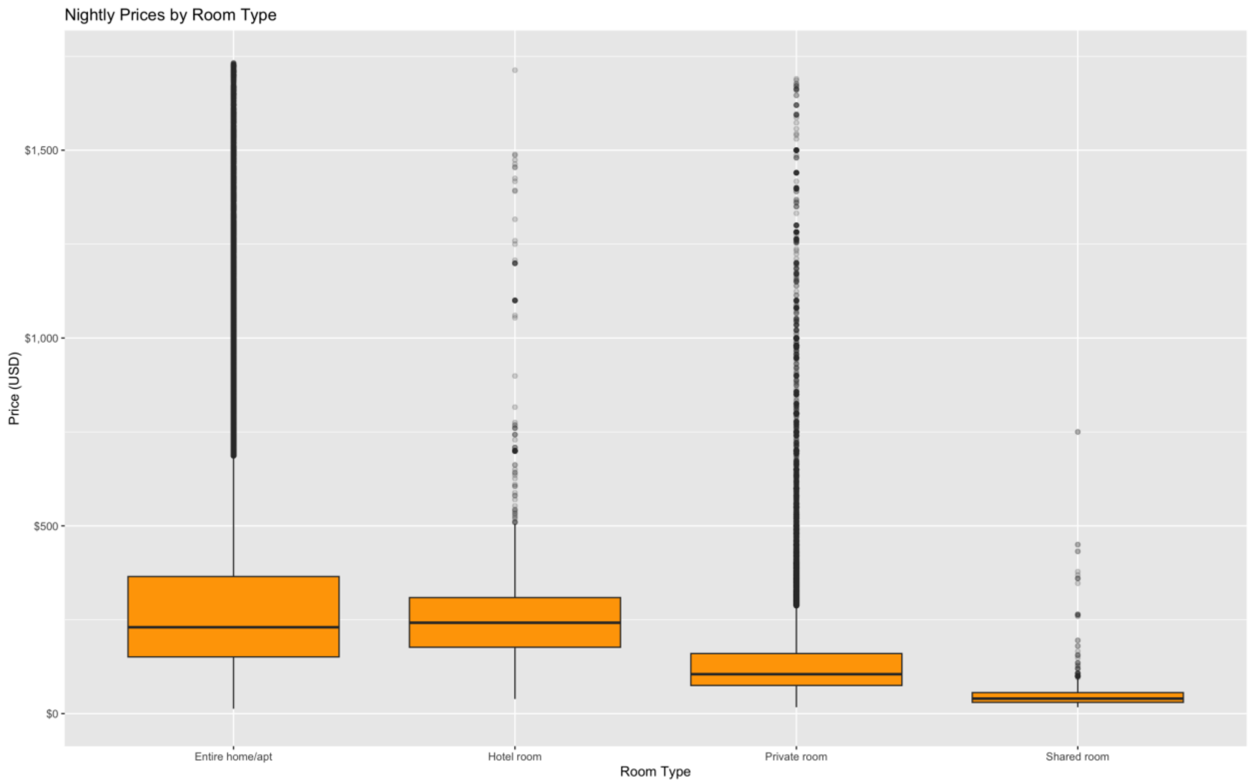

Figure 1.17 Nightly Prices by Room Type

	sentiment	count	perc
	<chr>	<int>	<dbl>
1	anticipation	4472649	13.2
2	joy	6388379	18.8
3	positive	10166747	30.0
4	trust	6498791	19.2
5	surprise	2454949	7.24
6	sadness	1567064	4.62
7	fear	505754	1.49
8	negative	1192592	3.52
9	disgust	278853	0.82
10	anger	377980	1.11

Figure 1.18 Sentiment and Emotion Distribution in Guest Reviews

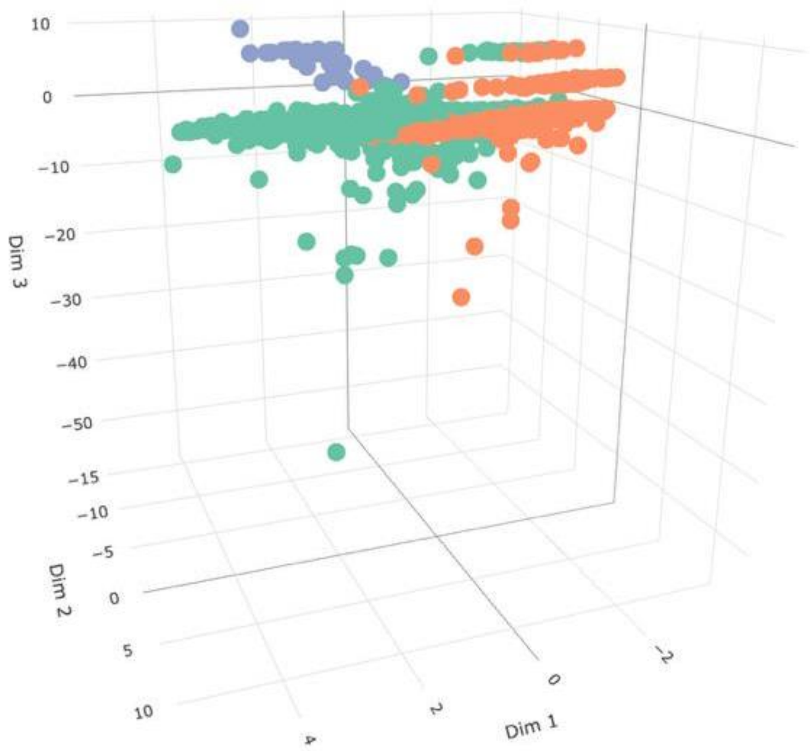

Figure 1.19 3D Cluster Visualisation of Auckland Listings

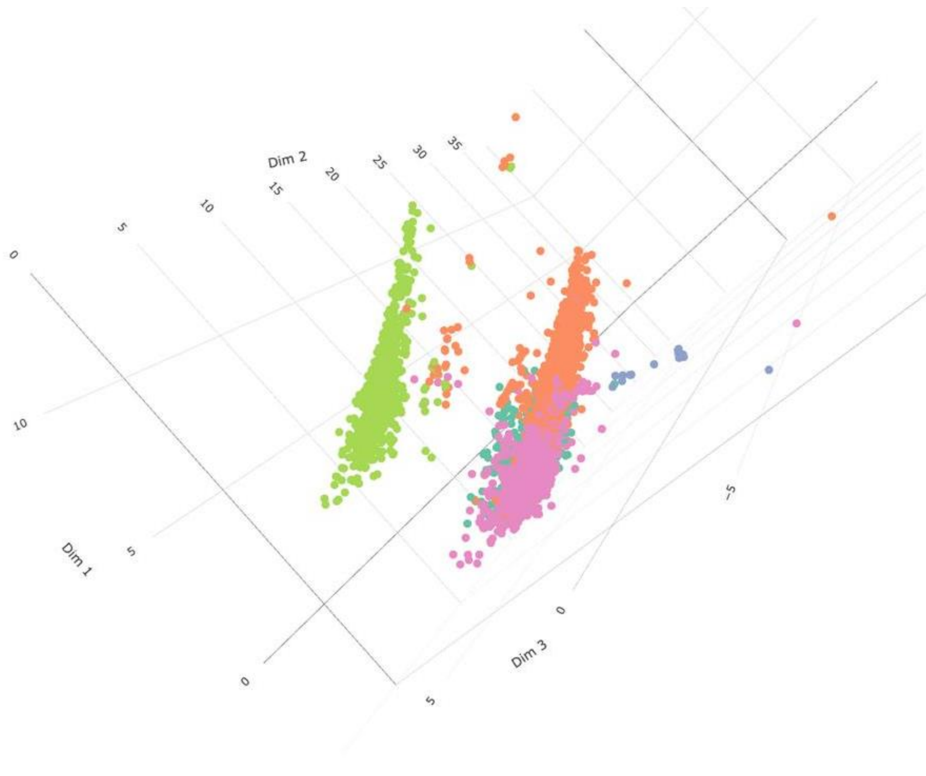

Figure 1.20 3D Cluster Visualisation of Queenstown-Lakes Listings