Enhancing Airbnb Listings: Aligning Visual Appeal with Guest Experiences in Athens

Business Analysis with Unstructured Data

Mattia Bianchi

Karla Esmeralda Felipe Chavarria

Mateo Franco

Angel Lanto

Pearl Soderberg

Phani Teja Venigalla

Master’s in Business Analytics

Table of Contents

I. Introduction

a. Business Problem

In the highly competitive short-term rental market of Athens, Airbnb hosts often face a significant challenge which is a mismatch between how their properties are visually presented through photos and written descriptions and the actual experiences of guests, as reflected in review comments. These inconsistencies can lead to guest dissatisfaction, negative feedback, reduced trust, and ultimately a decline in booking performance. For instance, a listing may showcase bright, well-composed images and polished descriptions, yet guests may report issues such as poor cleanliness, missing amenities, or misleading representations of space or location. This disconnects between expectation and reality not only impacts guest satisfaction but can also damage a host's reputation and profitability.

To address this problem, our project analyzes two types of unstructured data from the Inside Airbnb dataset: textual data, including guest review comments, host descriptions, and "host about" sections, which offer insight into guest perceptions and concerns; and image data, such as listing and host photos, which heavily influence guest expectations and booking decisions.

The objective of this analysis is twofold: first, to identify areas of alignment or discrepancy between listing visuals and actual guest experiences; and second, to provide actionable recommendations for Airbnb hosts on how to enhance their listings in ways that improve guest satisfaction and ultimately boost profitability.

b. Overview of the Dataset

The dataset from Inside Airbnb comprises two primary files: `listings.csv.gz` and `reviews.csv.gz`, offering a comprehensive view of Airbnb activity in Athens, Greece. The `reviews.csv.gz` file contains over 572,000 entries and includes six columns detailing individual guest reviews. Each review is associated with a `listing_id`, accompanied by reviewer metadata such as `reviewer_id`, `reviewer_name`, the `date` of the review, and most importantly, the `comments` column which holds the free-text feedback from guests. This column serves as a rich source of unstructured textual data that can be analyzed for sentiment, themes, and customer satisfaction indicators.

The `listings.csv.gz` file includes detailed information for 13,984 Airbnb listings across 75 columns. It features structured and unstructured data related to the listing itself, including textual descriptions (`name`, `description`, `host_about`, `neighborhood_overview`) and image data links (`picture_url`, `host_picture_url`, `host_thumbnail_url`). Additionally, it provides geographic data (latitude, longitude), host information, review scores, pricing, availability, and amenities. These fields not only help contextualize the reviews but also allow for multimodal analysis combining text and image elements.

Together, these datasets offer a rich foundation for analyzing Airbnb listing quality, guest satisfaction, and host performance. The combination of guest comments and listing images creates an opportunity to explore how visual and textual data align in shaping guest perceptions and behaviors perfectly aligning with the assignment’s goal of generating strategic recommendations using multiple types of unstructured data.

II. Data Loading and Exploration

The data loading and exploration phase began with setting up the Python environment using a comprehensive set of libraries for data handling, visualization, natural language processing (NLP), and computer vision. These included pandas, numpy, matplotlib, seaborn, nltk, textblob, cv2, and plotly, among others. After setting the appropriate configurations and downloading required NLP corpora such as stopwords and lemmatizers, the team proceeded to load the Airbnb dataset using pandas.read_csv(). The dataset consisted of two major components: a listings dataset with 14,642 entries and 75 columns, and a reviews dataset with 776,875 records and 6 columns. Initial exploration involved assessing the structure and content of these datasets using .info(), .head(), and .describe() methods to identify key variables and detect missing values.

In the listing's dataset, features such as id, name, description, picture_url, amenities, review scores, and various host-related metrics were examined in detail. The team selected a subset of relevant columns to reduce dimensionality and focus the analysis. The price column was cleaned by stripping currency symbols and formatting the values into numeric data types. The amenities field, originally a JSON-like string, was parsed into Python lists to extract the number of available amenities for each listing, stored as a new column named amenities_count. Missing values in critical text fields such as description and neighborhood_overview were filled with blank strings to ensure compatibility with NLP preprocessing steps.

In the reviews dataset, the date column was converted to datetime format for temporal analysis, and entries with null or empty comments were removed to preserve analytical quality. The comments field was the central focus for text analysis, and exploratory checks were conducted to assess review lengths, language distribution, and frequency over time. A resampled time series plot of review volume per month revealed booking and review trends, including drops during pandemic years and seasonal peaks. This initial data exploration not only established a clear understanding of the dataset structure and quality but also enabled strategic preparation for advanced analysis such as sentiment scoring, topic modelling, and image-based assessments.

Note: For a deeper understanding of the analysis, we invite you to explore the accompanying Jupyter Notebook, which contains all code used for data processing, sentiment analysis, image evaluation, and multi-modal integration. The notebook is well-commented and organized by sections to reflect the structure of the report. It includes visualizations, intermediate outputs, and technical explanations that support our findings and recommendations. This resource is particularly useful for stakeholders interested in the methodological rigor, reproducibility, or further extension of this project.

III. Exploratory Analysis, Key Findings and Insights

a. Text Analysis

The text analysis component focused on extracting meaningful insights from two primary unstructured text sources: guest reviews and listing descriptions. The process began with extensive preprocessing using Natural Language Processing (NLP) techniques. This involved converting all text to lowercase, removing URLs, special characters, punctuation, and numbers, followed by tokenization and lemmatization. Stopwords were also removed using NLTK's English stopword list to retain only the most informative words. This preprocessing was applied to both the comments column in the review's dataset and the description field in the listing's dataset. The cleaned text was then saved locally to avoid reprocessing and facilitate reuse in subsequent analyses.

To gauge guest sentiment, the team performed sentiment analysis using two approaches: TextBlob and VADER (Valence Aware Dictionary and sEntiment Reasoner). Each guest comment was assigned sentiment polarity scores, which were aggregated at the listing level to derive average sentiment for each property. Visualizations revealed that the majority of reviews were positive, though there were notable variations between properties. Further, sentiment scores were compared against numerical review ratings using scatter plots and boxplots, identifying instances where high numerical scores did not align with low sentiment and vice versa highlighting potential mismatches between guest expectations and experiences.

In addition to sentiment analysis, topic modelling was conducted using Non-negative Matrix Factorization (NMF) to uncover common themes in the textual data. For guest reviews, recurring topics included keywords such as "clean," "location," "host," "apartment," and multilingual clusters (e.g., French, Greek, Spanish), indicating the dataset's linguistic diversity. Similarly, listing descriptions revealed themes around location proximity, stylishness, comfort, and amenities. The extracted topics were visualized through labelled word clouds, offering a quick glance at dominant concepts shaping guest impressions and host marketing strategies.

To gain more granular insights, the project implemented aspect-based sentiment tagging, where guest reviews were scanned for keywords related to specific aspects such as cleanliness, communication, check-in, value, amenities, and comfort. For each review, binary indicators were generated for whether an aspect was mentioned, and the average sentiment for those mentions was calculated. Results showed that reviews mentioning cleanliness and communication generally carried higher sentiment, whereas mentions of accuracy and value sometimes had more critical tones. This aspect-level breakdown enabled the team to assess which service components most influenced guest satisfaction.

Overall, the text analysis provided a robust understanding of the emotional tone, key themes, and satisfaction drivers across listings. It helped identify where presentation gaps might exist between host claims and guest experiences laying the groundwork for targeted recommendations and more accurate, trustworthy listings.

b. Image Analysis

The image analysis segment of the project aimed to evaluate the visual quality and content of listing photos and understand how these visuals aligned with guest sentiment and review scores. The process began with strategic sampling of listings based on key criteria derived from the text analysis. These included listings with high, neutral, and low sentiment scores, specific aspect mentions (e.g., cleanliness, location, amenities), and cases showing strong misalignment between sentiment and numerical review ratings. This sampling ensured a balanced representation of listings across different guest experience categories while keeping the image analysis computationally manageable.

Using the picture_url field in the listings dataset, the team downloaded photos for the selected listings, implementing delays and user-agent headers to follow ethical scraping practices. In cases where image download failed or was not feasible, mock images were programmatically generated for demonstration purposes.

Once collected, the images underwent quality assessment using computer vision techniques. Key visual metrics were extracted for each image, including:

- Brightness: Average pixel intensity in grayscale.
- Contrast: Standard deviation of pixel intensity.
- Sharpness: Variance of the Laplacian, a common measure of edge clarity.
- Color Diversity: Number of unique RGB color combinations normalized to a scale.
- Saturation: Average saturation in HSV color space. Each image was assigned an overall quality score, calculated as a weighted average of these normalized metrics. Distributions of each metric were plotted to visualize the overall visual characteristics of Airbnb listings.

The team then explored correlations between image quality and guest sentiment, finding a positive trend where listings with higher sentiment often had brighter, sharper, and more colorful images. A regression line overlaid on a scatter plot of sentiment vs. image quality score suggested a moderate linear relationship. Listings in the high-sentiment group also consistently outperformed those in the low-sentiment group in terms of average visual quality metrics, as confirmed through boxplot comparisons and t-tests.

Further analysis investigated aspect-specific visual traits, comparing image characteristics like brightness across listings frequently mentioning aspects such as cleanliness, location, and comfort. For instance, listings that received positive mentions of cleanliness often featured well-lit, sharp, and high-contrast photos, visually reinforcing the positive review sentiment. Additionally, a heatmap of image metric correlations helped assess how different quality dimensions (e.g., brightness vs. sharpness) interacted across listings.

Although the project did not use deep learning-based object detection (e.g., YOLO or MobileNet) due to scope constraints, it implemented a basic object detection proxy using edge detection and contour analysis to count significant objects and estimate layout complexity. This approximation allowed the team to hypothesize room types (e.g., bedroom vs. kitchen) and assess spatial clarity, which was useful for flagging poorly framed or visually misleading listings.

Ultimately, the image analysis reinforced the findings from the text-based sentiment analysis. It confirmed that visual quality strongly contributes to guest perception, and listings that invest in professional, clean, and bright imagery tend to receive more positive feedback. The insights gathered supported the central thesis that misalignment between listing visuals and guest experiences can erode trust and satisfaction and should be a focal point for hosts aiming to improve performance.

c. Multi-Modal Analysis

The multi-modal analysis served as the integrative core of the project, combining textual insights from guest reviews and visual characteristics of listing photos to identify patterns of alignment or mismatch between presentation and guest experience. This holistic approach aimed to uncover how visual and verbal elements co-influence guest perception and booking outcomes.

The analysis began by merging sentiment data from textual reviews with quantitative image quality metrics at the listing level. Each listing, therefore, had associated data for average sentiment polarity (from TextBlob and VADER), review scores (e.g., cleanliness, communication), and image metrics like brightness, contrast, sharpness, saturation, and overall quality score. This unified dataset enabled cross-modal correlation analysis, where sentiment scores were plotted against image quality indicators to evaluate if well-presented listings were more positively reviewed. Scatter plots with regression lines suggested a positive relationship between image quality and sentiment, reinforcing the idea that guests respond favorably to visually appealing and accurately portrayed listings.

To further this, the team investigated rating–sentiment misalignment, identifying listings that had high star ratings but negative sentiment scores or vice versa. These cases were visualized on a sentiment vs. rating scatterplot, and listings with significant discrepancies were labeled as “misaligned.” This insight was critical: it suggested that numerical ratings alone may obscure deeper guest dissatisfaction, which can only be revealed through textual sentiment and visual inspection. For instance, a listing might receive five stars due to location but contain negative comments on cleanliness paired with dark, low-quality photos indicating a hidden mismatch.

The analysis also involved aspect-specific breakdowns, where sentiment related to topics like cleanliness or comfort was compared with visual metrics. Listings that were highly praised for cleanliness tended to have higher brightness and contrast scores, aligning well with textual sentiment. Conversely, listings with lower sentiment on amenities or comfort often exhibited lower saturation or sharpness in their photos suggesting a lack of visual cues for warmth, clarity, or quality. These pairings helped verify that guests use both visual and textual cues to form impressions, and that neglecting either medium may lead to lower satisfaction.

The team also examined neighbourhood effects and room-type patterns across the combined dataset. By grouping listings by room type (e.g., Entire home/apt, Private room), they evaluated whether certain types of accommodations showed stronger alignment between image quality and sentiment. Results indicated that “Entire home” listings generally had better visual metrics and higher sentiment alignment, while “Shared room” listings had more inconsistencies.

Another important aspect of the multi-modal analysis was the creation of cluster profiles. Using K-Means clustering (or similar logic), listings were grouped based on combined text and image features. This allowed the identification of distinct segments, such as:

- High Sentiment, High Visual Quality (ideal listings)
- High Sentiment, Low Visual Quality (good experiences but poor presentation)

- Low Sentiment, High Visual Quality (visually deceptive)
- Low Sentiment, Low Visual Quality (poor overall)

These clusters were visualized and described to inform strategic recommendations. For instance, listings in the third category were flagged as misleading and targeted for corrective action (e.g., more realistic photography or revised descriptions).

Overall, the multi-modal analysis revealed powerful insights that could not be achieved using text or image data in isolation. It highlighted the importance of consistency between what is promised visually and described textually versus what is delivered in reality, and provided a data-driven framework for identifying, segmenting, and improving listings based on their multi-dimensional guest experience profile.

IV. Business Recommendations and Conclusion

We examined the alignment of Airbnb listing presentations presented primarily images first and second presented as text descriptions, with the lived experience of guests as represented in review comments. Using sophisticated natural language processing (sentiment and emotion recognition), topic modelling, and image processing techniques, we identified significant misalignments between what hosts presented and what guests experienced. Even though reviews were on average positive there were significant misalignments with respect to cleanliness, accuracy, and amenities most often because of expectations created by previous listing images. These misalignments can lead to guest dissatisfaction, diminished trust, and ultimately poor or no bookings.

Our multimodal analysis indicated listings with highly audacious visuals and a poor guest sentiment often associated with overly dressed up or disingenuous pictures and or a description too garish to be believable. Alternatively, listing with decidedly less presentation but substantially consistent positive reviews suggest missed opportunities to capitalize on effective visual storytelling. These findings lead to the conclusion that aligning the visual and textual representation of a listing with the lived guest experience is an important lever to draw on to improve performance in a competitive market like Athens.

This leads us to our recommendations for action for Airbnb hosts:

1. **Prioritize Visual Authenticity over Aesthetic Virtuosity:**

Problem Addressed: Visual misalignment between heavily edited listing photos and actual property conditions creates expectation gaps, leading to guest disappointment and negative reviews despite high numerical ratings.

Implement a "Day in the Life" photography approach by capturing spaces at different times: morning shots of living areas with natural light (8-10am), afternoon photos of the kitchen during typical usage hours, and evening images with realistic lighting conditions. Adopt a "70-30 Rule" where 70% of photos show everyday appearance while 30% highlight optimal features. Focus on cleanliness signifiers by including close-ups of well-maintained surfaces, bedding with natural wrinkles, and storage spaces shown at practical capacity. Maintain freshness with quarterly photo updates, particularly for seasonal features and outdoor spaces, while documenting property improvements. This balanced approach creates an honest visual narrative that sets accurate expectations while still showcasing your property's genuine appeal.

2. **Locate Descriptions with Guest Sentiment Trends from Reviews:**

Hosts can incorporate guest-validated strengths (e.g., “quiet neighborhood”, “very responsive host”) into your listing description by first analyzing positive themes from previous review comments, and then at the same time, addressing documented concern comments (e.g., “no elevator”, or “thin walls”).

3. **Position Review sentiment similar to a continuous feedback loop:**

Hosts don't always need to analyze guest feedback – simply use guest feedback inclusion sentiment or aspect-based directed for guest feedback. This can identify patterns, and lead to action. If “comfort” had consistently low sentiment – the host may need to invest in new mattresses. Reviewing guests' willingness to trust the host's attempts to ensure happiness or comfort can set a tone for superior guest relations for subsequent visitors.

4. Improve listings before guests with visual and textual balance:

Use data (i.e. your guests) to inform which images – pick images which enhance themes in prior review comments (e.g. cozy living-room, or great view). These images don't need to be long descriptions or verbiage – the image conveys meaning and emotion, and you only need just enough text to draw the right audience to books the listing.

5. Quality assurance caveats for high-impact listings:

For hosts listing active premium price point or neighborhoods, it may not be possible to afford the loss in guest sentiment to proscribed management standards that inform and communicate guest experiences – hosts may need to audit guest experience sentiment and the presentation in their listing. Sound quality assurance practices discern the properties true potential and detangle conformation of actual outcome with listing image and copy. Comprehensive quality assurance of this nature, to maintain the possibility for 5-star ratings and incidental repeat bookings, will prove invaluable to both host and host communities.

V. Ethical Considerations

This project explores mismatches between visual presentations of Airbnb listings such as photos, descriptions, and guest reviews in Athens. The goal is to provide thoughtful insights to hosts, improve listing quality, satisfaction, and increase booking potential. While the analysis seeks to benefit customer service and overall Airbnb experience, it also raises several ethical considerations that could present risks for both business and customers.

Some of these risks are:

- Data Privacy: Although the dataset does not contain sensitive information such as addresses, financial details, or medical records, the name of each user is exposed and URLS linked to specific images and personal reviews. This creates a potential risk of re-identification, especially if it is combined with external data sources. One effective way to address this risk is to anonymize usernames and remove identifiers such as reviewer_name, listing_url, and puicture_url from any published results or public-facing views. Another option to minimize this risk is to use only small and unfocused samples of images in visual analysis. Lastly, in a real-world case scenario, especially when models are intended to analyze or predict customer behavior, it is essential to obtain consent from hosts if the images are to be used in model training, reports, or dashboards.
- Algorithmic Bias: Tools like TextBlod or VADER are typically trained on standardized English. As a result, when analyzing reviews written by non-native Englis speakers, these tools misinterpret meaning, sarcasm or cultural expressions-leading to distorted sentiment assessments (Mehrai et al., 2021). To reduce this type of issues, it is essential to validate models with more diverse samples and to consider linguistical or cultural adjustments during model development and evaluation (Barocas, Hardt & Narayanan, 2019)
- Transparency and Accountability: Methodological transparency is a cornerstone of ethical data science. Clear documentation of data preprocessing, model selection, and analytical decisions is essential for ensuring auditability and facilitating peer review (Gebru et al., 2018). When analyzing this type of data, it is crucial not only to share insights, but also to provide explanations for each phase of the process. This includes describing the rationale behind each decision and function used to analyze customer behavior, this with the aim of ensuring the customers their information is being used ethically and responsibly.

Although the code may be detailed, well structured, and clearly annotated, additional justification is needed for key decisions. For example, when removing reviews without detailed comments, the rationale should be explicitly documented. Similarly, when cleaning the data for instance, directly replacing currency symbols. It is extremely important to include explanations of these steps to maintain transparency about the analysis.

In conclusion, while data analysis is used to support business decisions, it should include privacy preserving techniques, fairness audits and transparent reporting. Beyond technical accuracy, ethical data analysis should promote equity is promote and build trust among affected users but also provide a mindful analysis that helps to improve and enhance decision making in a fair and inclusive manner.

VI. References

Airbnb. (n.d.). *Community Standards and Terms of Service*. Retrieved from https://www.airbnb.com/help/article/2865/community-standards

Association for Computing Machinery (ACM). (2018). *ACM Code of Ethics and Professional Conduct*. Retrieved from https://www.acm.org/code-of-ethics

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. Retrieved from https://fairmlbook.org/

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management, 55*, 62–73. https://doi.org/10.1016/j.tourman.2016.01.013

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Inside Airbnb. (n.d.). *Inside Airbnb Athens Dataset*. Retrieved from http://insideairbnb.com/athens

Ma, X., Hancock, J. T., Mingjie, K. L., & Naaman, M. (2017). Self-disclosure and perceived trustworthiness of Airbnb host profiles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 2397–2409). ACM. https://doi.org/10.1145/2998181.2998269

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR), 54*(6), 1–35. https://doi.org/10.1145/3457607

UK Government. (2018). *Data Ethics Framework*. Retrieved from https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How much is an image worth? Airbnb property demand estimation leveraging large scale image analytics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2964874