MINI ANALYTICS PROJECT

Exploring the Key Drivers of Customer Satisfaction in Airbnb Listings: Insights for Hosts and Airbnb

GitHub repository: <https://github.com/ntat73/Mini-analytics-project>

Submission date: Dec 16, 2024

TABLE OF CONTENTS

[INTRODUCTION: 2](#_Toc185110028)

[PROJECT BACKGROUND 2](#_Toc185110029)

[PROBLEM STATEMENT 2](#_Toc185110030)

[OBJECTIVES AND QUESTIONS 3](#_Toc185110031)

[ANALYTICS TOOLS AND APPROACH 3](#_Toc185110032)

[DATA ACCESS, ETHICS, SECURITY AND PRIVACY 4](#_Toc185110033)

[DATA SOURCE 4](#_Toc185110034)

[ETHICS, PRIVACY, AND SECURITY 6](#_Toc185110035)

[DATA RELIABILITY AND VALIDITY 6](#_Toc185110036)

[ANALYTICS TECHNIQUES, MODELS, AND APPROACHES 6](#_Toc185110037)

[TOOLS AND LIBRARIES 10](#_Toc185110038)

[MODEL ASSUMPTION AND DATA MODIFICATION 10](#_Toc185110039)

[REPORT PRESENTATION 10](#_Toc185110040)

[CONCLUSION 15](#_Toc185110041)

[REFERENCE 16](#_Toc185110042)

[APPENDIX 18](#_Toc185110043)

[APPENDIX 1: DATA PREPARATION CODE SNIPPET 18](#_Toc185110044)

[APPENDIX 2: DESCRIPTIVE ANALYTIC CODE SNIPPET 20](#_Toc185110045)

[APPENDIX 3: STATISTICAL/ COMPARATIVE ANALYTICS CODE SNIPPET 22](#_Toc185110046)

[APPENDIX 4: PREDICTIVE ANALYTICS CODE SNIPPET 23](#_Toc185110047)

[APPENDIX 5: VISUALIZATIONS CODE SNIPPET 24](#_Toc185110048)

[APPENDIX 6: AI CONSENT FORM 26](#_Toc185110049)

# INTRODUCTION:

## Project background

The rapid growth of Airbnb has significantly impacted the hotel industry financially, appealing to travelers seeking unique and personalized experiences that traditional hotels often lack with a diverse range of lodging choices (Agapitou et al., 2020). Conversely, Airbnb guests contribute to local economies by spending on dining, entertainment, and transportation, benefiting local businesses. As a result, customer expectations in the UK hospitality industry have shifted as travelers increasingly prioritize authentic, localized experiences over conventional hotel stays (Tabari, 2019). However, the rise of short-term rentals has been linked to declining hotel bookings and job losses in the hospitality sector.

London exemplifies this growth, with 117,000 active listings in 2023, contributing £1.5 billion to the local economy (Airbnb, 2024). Entire homes cater to families and groups, while private rooms attract budget-conscious solo travelers (Ding et al., 2023). Many hosts depend on Airbnb’s income for financial stability.

However, Airbnb often falls short compared to hotels' standardized services, creating gaps in guest expectations (Agapitou et al., 2020). Meanwhile, customer satisfaction is vital, driving better ratings, repeat bookings, and profitability (Svetec and He, 2023). To stand out, hosts must prioritize cleanliness, room features, and responsiveness—key factors in UK hospitality (Magnini and Zehrer, 2021). High service standards are essential for ensuring positive guest experiences and long-term success.

## Problem statement

Airbnb hosts struggle to identify factors driving higher guest ratings due to subjective reviews and a lack of standardized frameworks. Research highlights that reviews often focus on general satisfaction, offering limited actionable insights (Guttentag et al., 2018). Unlike structured hotel ratings, Airbnb’s system creates inconsistencies, complicating efforts to improve guest satisfaction (Cheng et al., 2019). Evolving expectations add to the challenge, with 67% of guests in 2022 expecting hotel-like amenities like 24/7 check-in, which many rural hosts struggle to provide (Airbnb Insights, 2022).

This disconnect between expectations and capabilities results in inconsistent ratings, even when experiences are satisfactory. Key drivers—such as pricing, property type, or Superhost status—remain unclear, hindering targeted improvements (Aramendia-Muneta et al., 2023). Similarly, Airbnb faces challenges refining its platform and addressing questions on pricing, cleanliness, and property features, limiting both host and platform optimization.

## Objectives and questions

**Objectives:**

This study aims to uncover the key drivers of guest satisfaction in Airbnb listings and quantify the influence of various factors on customer experiences, specifically how aspects such as property type, cleanliness, pricing strategies, host status, and proximity to city centers impact satisfaction scores. By leveraging empirical data, the research will provide a detailed analysis of these elements to identify actionable insights.

The ultimate goal is to offer insight that helps hosts optimize their listings, ensuring they meet guest expectations effectively while also providing a guide for Airbnb to refine its platform functionalities by offering data-driven solutions to improve both host and guest experiences, as well as enhance its marketing strategies’ inclusiveness toward customers’ preferences. This dual approach is intended to support better outcomes for all stakeholders involved.

**Key questions:**

1. Which listing features (price tiers, room type, cleanliness, distance, or Superhost status) are most associated with high satisfaction scores?
2. Does the type of property (entire home, private room, shared room) influence customer satisfaction?
3. How does cleanliness impact satisfaction scores?
4. How do pricing strategies affect customer satisfaction?
5. How does Superhost status influence satisfaction scores?

## ANALYTICS TOOLS AND APPROACH

This project examines factors influencing Airbnb customer satisfaction using Python and analytics libraries, including Pandas and NumPy for data manipulation, Seaborn and Matplotlib for visualization, and Scipy and Statsmodels for statistical analysis. The overview of the project’s approaches is presented in (table 1) and will be further justified.

|  |  |  |
| --- | --- | --- |
| **Analytics approach** | **Method** | **Objectives** |
| Data Cleaning | Handle missing values, outliers, and normalize features | Ensure data accuracy and consistency for analysis. |
| Descriptive Analytics | Summarize data distributions and averages across features. | Identify patterns and trends in satisfaction scores. |
| Comparative Analytics | ANOVA, T-tests, Tukey’s test and correlation analysis. | Highlight differences and relationships between satisfaction drivers. |
| Predictive Analytics | Logistic and multiple linear regression | Understand feature contributions and forecast satisfaction outcomes. |
| Visualization | Scatter plots, box plots, bar chart, etc. | Present findings clearly and support insights visually. |

Table 1: Project’s tools and approaches overview

# Data access, ethics, security and privacy

## Data source

The datasets, *london\_weekdays,* and *london\_weekends*, were collected through automated web-scraping using Selenium WebDriver as part of a study by Gyódi and Nawaro (2021) on short-term rental impacts in European cities. Search queries targeted two-person accommodations for two-night stays in 10 major European cities, with data collected 4–6 weeks before travel. Prices included full costs, such as reservation and cleaning fees. Weekday samples formed the primary analysis dataset, with weekend samples used for robustness checks.

Searches covered various city districts while excluding listings outside administrative borders using city area shapefiles. Inner London was analyzed for consistency across cities, and listings accommodating more than six people were excluded to focus on smaller properties. The dataset, hosted on Zenodo, provides detailed insights into Airbnb listings in London, including property types, pricing, availability, host characteristics, and guest satisfaction, enabling an in-depth analysis of market dynamics and customer preferences.

**Data structure and nature:**

The dataset is structured in tabular format, with each row being a particular listing’s observation. The columns represent properties features as shown and justified below. In Figure 2, the rows in red are the key features of this analytics project, while orange represents a potential feature to be considered.

A table of numbers and a few digits

Description automatically generated with medium confidence

Figure 1: Dataset’s structure

A table with text on it

Description automatically generated

Figure 2: Columns’ explanation

## Ethics, privacy, and security

**Ethical consideration**

The dataset used in this project is publicly available and adheres to data-sharing guidelines, ensuring ethical use. Sensitive information, such as Personally Identifiable Information (PII), is anonymized per GDPR, ensuring no individual is identifiable. It is used solely for research and educational purposes, with no commercial intent, reinforcing compliance.

**Privacy and security measures**

To maintain privacy and security, strict measures were implemented. Data handling adhered to privacy guidelines, with no attempts to reverse engineer or de-anonymize individuals. The analysis focused on aggregated insights, reducing privacy risks. The dataset was securely stored in a controlled environment and accessed via authorized tools like Python libraries (e.g., Pandas, Seaborn). Local systems use encryption and robust security protocols to prevent unauthorized access.

## 

## Data reliability and Validity

The dataset, sourced from Zenodo, a trusted open-access repository, ensures credibility and reliability. It represents Airbnb activity in London on weekdays, directly aligning with the research objectives. Supplementary peer-reviewed data enhances reliability by adding context and validation.

Key features such as room\_type, cleanliness\_rating, and guest\_satisfaction\_overall align with objectives, ensuring validity. Derived features like price tiers were created to maintain analytical integrity (Appendix 1). Missing data was removed, and outliers in pricing (realSum) and distance (dist) were addressed to prevent skewed results. Variables were normalized for consistency (Appendix 1). Statistical tests and cross-validation techniques ensured robust, reliable conclusions (Appendix 3), reflecting the project’s commitment to ethical, reliable, and secure data usage.Top of FormBottom of Form

# Analytics Techniques, Models, and Approaches

The project employs a combination of descriptive, statistical (comparative), and predictive analytics techniques chosen for their effectiveness in addressing the research objectives and supported by established academic literature in the fields of consumer behavior and hospitality analytics.

**Descriptive Analytics:**

Descriptive analytics lays the groundwork for exploring data distributions and trends using frequency analysis, correlations, grouping, and visualizations like bar charts and box plots. For example, price tiers (low, medium, high) were created based on room types to analyze pricing’s effect on satisfaction, while binary indicators for high satisfaction (scores ≥ 85) and cleanliness (ratings ≥ 8) aided further analysis (Appendix 2). These methods effectively summarize complex datasets, uncover key patterns, and provide preliminary insights into guest satisfaction (Law et al., 2018). Descriptive analytics is crucial for understanding customer behavior and enhancing service quality in hospitality (Mariani & Baggio, 2021).

The detailed methods will be shown in Table 2, while the Appendix 2 reveals the code snippet for descriptive analytics used in this project.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Objectives** | **Independent Variable** | **Dependent Variable** |
| **Frequency Analytics** | Identify the distribution of high satisfaction scores across categories | room\_type  price\_tier | **High satisfaction** |
| Calculate the proportion of high satisfaction across features | host\_is\_superhost |
| **Correlation** | Measure the linear relationship between continuous variables | cleanliness\_rating | **Guest satisfaction\_ overall** |
| Evaluate monotonic relationships between binary features | host\_is\_superhost |
| **Mean grouping** | Calculate average satisfaction scores grouped by features to identify trends. | room\_type  price\_tie |
| Compare mean satisfaction scores between two groups to identify significant differences. | host\_is\_superhost |
| **Visualization** | Box Plots  Scatter Plots  Bar Charts | price\_tier  room\_type  room\_type  cleanliness\_rating |

Table 2: Descriptive methods conducted for the project (Kotu & Deshpande, 2019; Mariani & Baggio, 2021).

**Statistical/ Comparative Analytics**

Statistical analytics involves inferential techniques like ANOVA and T-tests, which are crucial for identifying significant group-level differences. ANOVA evaluates variations in satisfaction scores across categories ( room type, price tier), while T-tests compare binary groups (Superhosts versus non-superhosts) (Appendix 3). These methods are justified by their robust application in identifying meaningful differences in hospitality settings, as evidenced by Magnini and Zehrer (2021), who emphasize the importance of statistical testing in uncovering actionable insights into guest satisfaction. Tukey’s test was used to identify pairwise differences following significant ANOVA results, offering further granularity in understanding satisfaction drivers.

The detailed methods will be shown in Table 3, while Appendix 3 reveals the code snippet for statistical analytics used in this project.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Objectives** | **Independent Variable** | **Dependent Variable** |
| **ANOVA** | Evaluate variations in satisfaction scores across categories | room\_type  price\_tier  Cleanliness\_rating | Guest satisfaction overall |
| **Tukey’s Test** | Perform pairwise comparisons to pinpoint specific group differences. |
| **T-tests** | Compare satisfaction scores between binary groups | host\_is\_superhost |

Table 3: Statistical analytics methods (Anubala, 2023)

**Predictive analytics:**

Predictive analytics uses logistic and multiple linear regression to model and understand guest satisfaction. Logistic regression predicts high satisfaction scores (high\_satisfaction = 1), incorporating predictors like price tiers, cleanliness ratings, and Superhost status (Appendix 4). This method quantifies relationships between features and outcomes, providing actionable recommendations (Agapitou et al., 2020). Multiple linear regression predicts continuous satisfaction scores (guest\_satisfaction\_overall), analyzing how features like room type, distance from the city center, and pricing impact satisfaction (Appendix 4). Together, these models equip Airbnb hosts with insights to optimize offerings, enhance guest experiences, and maintain a competitive edge in the evolving rental market.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Objectives** | **Independent Variable** | **Dependent Variable** |
| **Logistic Regression** | predict high satisfaction scores (binary outcome). | price\_tier  cleanliness\_rating  host\_is\_superhost  dist | high\_satisfaction |
| **Multiple Linear Regression** | Predict numerical satisfaction scores based on key features. | price\_tier  cleanliness\_rating  room\_type  dist | Guest satisfaction overall |

Table 4: Predictive analytics methods (Agapitou et al., 2020).

## Tools and libraries

The project relied on Python for all analyses, utilizing the following libraries:

|  |  |
| --- | --- |
| **Pandas** | For data cleaning and manipulation. |
| **NumPy** | For numerical computations. |
| **Seaborn and Matplotlib** | For creating visualizations such as box plots, scatter plots, bar charts, and heatmaps. |
| **SciPy and Statsmodels**: | For conducting statistical tests, including ANOVA and T-tests. |
| **Scikit-learn** | For building and evaluating logistic regression models. |

Table 5: Tools and Libraries

## Model Assumption and data modification

During the modeling process, several assumptions were made to ensure validity and robustness. For multiple linear regression, it was assumed that a linear relationship exists between independent variables (cleanliness\_rating, price\_tier, room\_type) and the dependent variable (guest\_satisfaction\_overall), validated through scatter plots and residual analysis. The independence of observations was presumed, and the normality of residuals was checked using Q-Q plots, with residual variance verified through residual plots. Multicollinearity was assessed using Variance Inflation Factor (VIF) values. For logistic regression, the target variable (high\_satisfaction) was assumed to be binary.

To meet these assumptions, modifications included imputing or removing missing cleanliness\_rating values and treating outliers in realSum (price) using the Interquartile Range (IQR) method (Appendix 1). Continuous variables like realSum and dist (distance) were normalized to prevent scale biases. Feature engineering added price tiers and dummy-encoded room types (Appendix 1). The dataset was split into training and testing sets for generalizability, with binary target imbalances addressed via oversampling or weighting. These steps ensured reliable and accurate analyses.

# Report presentation

**Question 1: Which listing features ( price tiers, room type, Superhost status, cleanliness, or distance to the city center) are most associated with high satisfaction scores?**

The analysis reveals that **Superhost status** and **cleanliness** are the most critical factors influencing guest satisfaction on Airbnb. Superhosts achieve the highest satisfaction rates, with 97.49% of guests expressing high satisfaction, underscoring the importance of excellent service and trust. Cleanliness similarly plays a pivotal role, with properties rated highly for cleanliness achieving an 82.57% satisfaction rate, while those with low cleanliness ratings plummet to 9.39%. Among room types, **Private Rooms** demonstrate slightly higher satisfaction (80.15%) than **Entire Homes** (74.41%) and **Shared Rooms** (73.91%), reflecting a balance between affordability and privacy.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | |  |

Figure 3: Percentage of high satisfaction across feature

**Price tiers** also influence satisfaction, with higher-priced properties achieving 82.51% satisfaction compared to 71.41% for lower-priced ones, likely due to enhanced amenities and service quality. However, logistic regression indicates that price has a minimal direct effect, highlighting the need for hosts to offer value-for-money experiences. Lastly, **distance from the city center** shows a marginal positive impact, with suburban locations (10-20 miles) slightly favored due to their tranquility and balance of accessibility. While medium-tier pricing and distance play secondary roles, their impacts are context-dependent and often overshadowed by cleanliness and host quality.

|  |  |
| --- | --- |
|  |  |

Figure 4: Regression analysis

**Question 2: Does the type of property influence customer satisfaction?**

Room type significantly impacts customer satisfaction, as shown by the ANOVA test (F-statistic: 24.40, p-value: 2.88e-11) (Appendix 3) with **Entire homes/apartments** consistently achieve the highest satisfaction. Tukey’s post-hoc test confirms a statistically significant mean satisfaction difference of 2.40 between entire homes and private rooms (p < 0.05).

Satisfaction scores for all property types are generally high, with most above 80. Entire homes have the highest median satisfaction and fewer outliers, while private rooms show slightly lower median scores and more variability. Shared rooms have the lowest satisfaction and the widest spread, indicating less consistent guest experiences.

A diagram of a group of blue boxes

Description automatically generated

Figure 5: Regression analysis

Interestingly, while entire homes generally perform best, logistic regression findings reveal that room type alone is not the strongest determinant of high satisfaction as features like cleanliness and Superhost status exert a greater influence.

**Question 3: How does cleanliness impact satisfaction scores?**

Cleanliness is strongly associated with guest satisfaction, as evidenced by a correlation coefficient of 0.76 and a highly significant p-value of 0.00, indicating a strong linear relationship (Appendix 2). Properties with higher cleanliness ratings (e.g., 9 or 10) consistently achieve near-perfect satisfaction scores, while those with cleanliness ratings below five show significantly lower satisfaction. The bar plot confirms that average satisfaction scores increase steadily with higher cleanliness ratings. Properties with cleanliness ratings of 9 or 10 show the highest satisfaction, while properties rated between 2 and 5 perform poorly.

Regression analysis and the scatter plot further illustrate this positive trend, with the regression line highlighting the consistent rise in satisfaction scores as cleanliness improves. However, there is noticeable variability in satisfaction for properties with mid-range cleanliness ratings (5 to 7), suggesting other factors may influence satisfaction in this range.

|  |  |
| --- | --- |
|  |  |

Figure 5: Cleanliness Rating versus Customer Satisfaction.

**Question 4: How do pricing strategies affect customer satisfaction?**

Analysis of pricing strategies reveals significant impacts on customer satisfaction. High-price tier properties consistently achieve the highest satisfaction scores, followed by medium-tier, with low-tier properties showing the lowest satisfaction. These differences are statistically significant, as confirmed by Tukey’s HSD test. High-price tiers likely provide superior amenities, while medium tiers balance affordability and quality.

Room-type-specific pricing effects highlight nuances between price and satisfaction. Entire homes and private rooms show a weak but significant positive correlation, indicating higher prices marginally improve satisfaction, with cleanliness and amenities playing a larger role. Shared rooms show a moderate correlation, but price variability inconsistently affects experiences.

|  |  |  |
| --- | --- | --- |
| A diagram of a variety of blue boxes  Description automatically generated with medium confidence |  |  |

Figure 6: Pricing versus Customer Satisfaction.

In conclusion, pricing strategies significantly influence customer satisfaction, with high-price tiers delivering superior experiences. For entire homes and private rooms, pricing’s impact is overshadowed by other satisfaction drivers. Shared rooms display variability, indicating opportunities for targeted improvements. Hosts can use these findings to optimize pricing strategies based on property type and tier, ensuring profitability while enhancing guest experiences.

**Question 5: How does Superhost status impact satisfaction scores?**

Superhost status is strongly associated with higher guest satisfaction. The mean satisfaction score for Superhosts is 96.58, compared to 89.25 for Non-Superhosts. This is further validated by a T-test (T-statistic = 15.38, P-value = 4.61e-52), confirming the statistical significance of the higher satisfaction scores for Superhosts. A correlation coefficient of 0.22 (P-value = 4.61e-52) indicates a weak but significant positive relationship between Superhost status and guest satisfaction, suggesting that while Superhost status contributes positively, other factors also play a role.

A graph of a number of scores

Description automatically generated with medium confidence

Figure 7: Satisfaction Scores by Superhost Status.

The box plot demonstrates that Superhosts achieve consistently higher median satisfaction scores with lower variability, indicating more predictable guest experiences. In contrast, Non-Superhosts exhibit wider variability, with more outliers at lower satisfaction levels, highlighting less consistent performance. Hence, superhost status emerges as a critical driver of satisfaction, reinforcing the importance of high service standards and reliability in enhancing guest experiences.

# Conclusion

This study explored factors influencing guest satisfaction in Airbnb listings, focusing on property type, cleanliness, pricing, Superhost status, and proximity to city centers. Its goal was to provide actionable insights for Airbnb hosts and refine user experiences through data-driven improvements.

Key findings showed Superhost status and cleanliness as the strongest predictors of satisfaction. Superhosts achieved higher scores due to reliability and superior service, while cleanliness ratings of 9 or 10 correlated with exceptional guest experiences. Pricing strategies also played a role, with high-price tiers offering better amenities and ratings, medium tiers balancing affordability and quality, and low tiers underperforming due to limited features. ANOVA highlighted significantly higher satisfaction for entire homes, though regression suggested private rooms could achieve high satisfaction if other features like cleanliness and Superhost status were optimized.

In terms of self-evaluation, the analysis demonstrated notable strengths, such as identifying cleanliness and Superhost status as key drivers of satisfaction. The use of statistical tests provided robust insights into group-level differences, while predictive models quantified the importance of specific features. However, limitations arose in addressing outliers and interpreting weaker correlations, such as the influence of price tiers, highlighting areas for methodological refinement. Therefore, even though despite time constraints, the analysis succeeded in delivering actionable insights but still indicated the need for further validation of statistical assumptions and deeper exploration of feature interdependencies.

Hence, future projects would benefit from allocating additional time to exploratory analysis and validating assumptions more rigorously. Enhancing expertise in advanced modeling techniques and seeking collaborative feedback would ensure greater accuracy and depth in future analytical projects.

# Reference

Agapitou, C., Liana, A., Folinas, D., & Konstantoglou, A. (2020). Airbnb Is Customers’ Choice: Empirical Findings from a Survey. *Sustainability*, *12*(15), 6136. https://doi.org/10.3390/su12156136

Airbnb. (2024, September 3). *EY report finds Airbnb has little to no significant impact on UK housing*. Airbnb Newsroom. https://news.airbnb.com/en-uk/ey-report-finds-airbnb-has-little-to-no-significant-impact-on-uk-housing/

Anubala, R. (2023). *The Future of Hospitality: Predictive Analytics in Hotel Management.* ResearchGate. https://10.13140/RG.2.2.27132.76161.

Aramendia-Muneta, M. E. (2024). Why do guests stay at Airbnb versus hotels? An empirical analysis of necessary and sufficient conditions. *Unavarra*. https://www.academia.edu/104969623/Why\_do\_guests\_stay\_at\_Airbnb\_versus\_hotels\_An\_empirical\_analysis\_of\_necessary\_and\_sufficient\_conditions

Cheng, M., & Jin, X. (2018). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, *76*, 58–70. https://doi.org/10.1016/j.ijhm.2018.04.004

Ding, K., Choo, W. C., Ng, K. Y., Ng, S. I., & Song, P. (2021). Exploring Sources of Satisfaction and Dissatisfaction in Airbnb Accommodation Using Unsupervised and Supervised Topic Modeling. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.659481

Ding, K., Niu, Y., & Choo, W. C. (2023). The evolution of Airbnb research: A systematic literature review using structural topic modeling. *Heliyon*, *9*(6), e17090. https://doi.org/10.1016/j.heliyon.2023.e17090

Guttentag, D. (2019). Progress on Airbnb: a literature review. *Journal of Hospitality and Tourism Technology*, *10*(4), 814–844. https://doi.org/10.1108/jhtt-08-2018-0075

Gyódi, K., & Nawaro, Ł. (2021a). Determinants of Airbnb prices in European cities: A spatial econometrics approach (Supplementary Material) [Dataset]. In *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.4446043

Gyódi, K., & Nawaro, Ł. (2021b). Determinants of Airbnb prices in European cities: A spatial econometrics approach. *Tourism Management*, *86*, 104319. https://doi.org/10.1016/j.tourman.2021.104319

Kotu, V., & Deshpande, B. (2014). Association Analysis. In *Elsevier eBooks* (pp. 195–216). https://doi.org/10.1016/b978-0-12-801460-8.00006-9

Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, *26*(5), 727–750. https://doi.org/10.1108/ijchm-08-2013-0367

Magnini, V. P., & Zehrer, A. (2020). Subconscious influences on perceived cleanliness in hospitality settings. *International Journal of Hospitality Management*, *94*, 102761. https://doi.org/10.1016/j.ijhm.2020.102761

Mariani, M., & Baggio, R. (2021). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, *34*(1), 231–278. https://doi.org/10.1108/ijchm-03-2021-0301

Tabari, S. (2024). Impact of Airbnb on Customers’ Behavior In the UK Hotel Industry. *Cardiff*. <https://www.academia.edu/88620971/Impact_of_Airbnb_on_Customers_Behavior_In_the_UK_Hotel_Industry>

OpenAI. (2024). ChatGPT [Language model]. OpenAI. https://openai.com/chatgpt

# Appendix

### Appendix 1: Data Preparation code snippet

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

### Appendix 2: Descriptive Analytic code snippet

Frequency analytics:

A screenshot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

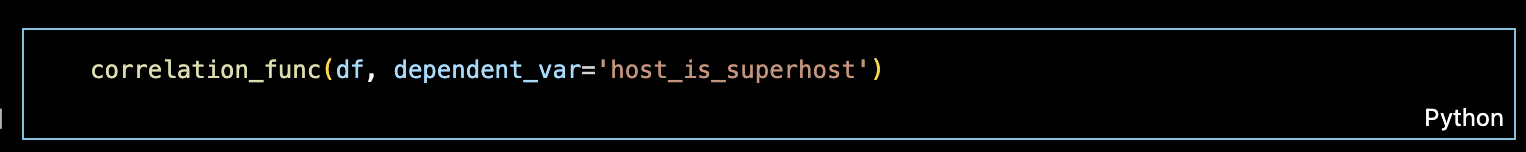
Correlation analytics:

A screen shot of a computer program

Description automatically generated

A black screen with text

Description automatically generated



Mean Grouping:

A computer screen shot of a black background

Description automatically generatedA screen shot of a computer

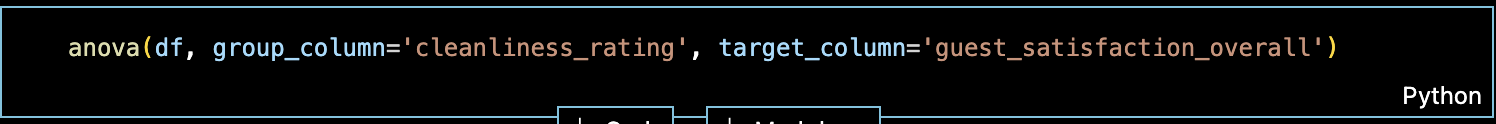
Description automatically generated

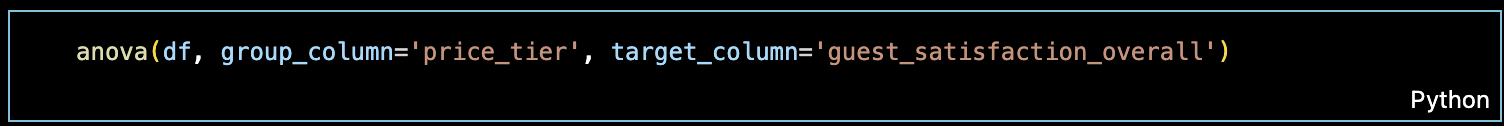
### Appendix 3: Statistical/ Comparative Analytics code snippet

ANOVA:

A screen shot of a computer program

Description automatically generated

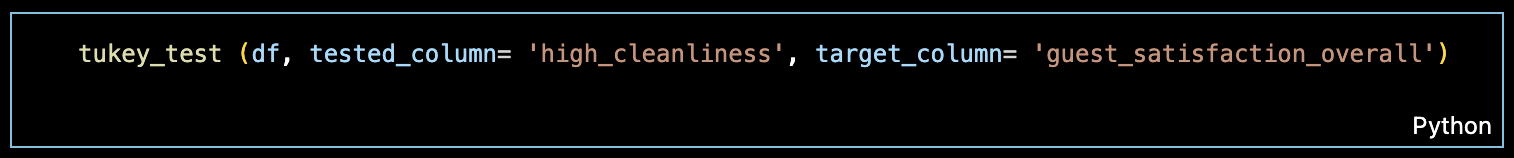
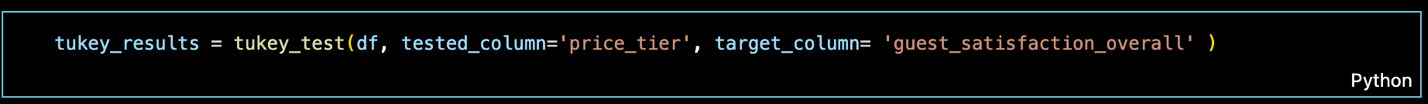




TUKEY’S TEST:

A screen shot of a computer code

Description automatically generated



T-TEST:

A screen shot of a computer program

Description automatically generated

### APPENDIX 4: PREDICTIVE ANALYTICS CODE SNIPPET

Logistic regression

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

Multiple Linear Regression:

A computer screen shot of text

Description automatically generated

### Appendix 5: Visualizations code snippet

A screen shot of a computer program

Description automatically generated

A black background with white text

Description automatically generated

A screen shot of a computer code

Description automatically generated

A computer screen shot of code

Description automatically generated

A screen shot of a computer code

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

### Appendix 6: AI Consent form

A close-up of a logo

Description automatically generated

|  |  |
| --- | --- |
| Assignment Cover Sheet | |
| Candidate Number | 025954 |
| Module Code | BEMM457\_A\_1\_202425 |
| Module Name | Topics in Business Analytics |
| Assignment Title | Coursework 2: Mini business analytics project |

*Within the Business School we support the responsible and ethical use of GenAI tools, and we seek to develop your ability to use these tools to help you study and learn. An important part of this process is being transparent about how you have used GenAI tools during the preparation of your assignments.*

*The below declaration is intended to guide transparency in the use of GenAI tools, and to assist you in ensuring appropriate referencing of those tools within your work.*

*The following GenAI tools have been used in the production of this work:* Chat GPT, Grammarly.

*I have used GenAI tools for brainstorming ideas.*

​​ *I have used GenAI tools to assist with research or gathering information.*

​​​ *I have used GenAI tools to help me understand key theories and concepts.*

​​ *I have used GenAI tools to identify trends and themes as part of my data analysis.*

​​​ *I have used GenAI tools to suggest a plan or structure of my assessment.*

​​ *I have used AI tools to give me feedback on a draft.*

​​ *I have used GenAI tool to generate images, figures or diagrams.*

​​​ *I have used AI tools to proofread and correct grammar or spelling errors.*

​​ *I have used AI tools to generate citations or references.*

​​ *Other [please specify] …………………………………………………………………………………………………………………*

​​ *I declare that I have referenced use of GenAI tools and outputs within my assessment in line with the* [*University referencing guidelines*](https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Flibguides.exeter.ac.uk%2Freferencing&data=05%7C02%7Cesm.buildingone%40exeter.ac.uk%7Cf648da2e0e514f17ff0808dcef4ae06d%7C912a5d77fb984eeeaf321334d8f04a53%7C0%7C0%7C638648352064065071%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=XtT64F9dPrvOQT2VctmEQFhuY7otvEvyQ8PLT5lzbus%3D&reserved=0)*.*