**The Role of Guest Sentiment on Airbnb Pricing along Ireland’s Wild Atlantic Way**



This research project is submitted in partial fulfillment of the Degree of Master of Science in

Data Analytics at Technological University of the Shannon: Midlands Midwest

Submittedby

**Helmi Kaittikkattil Abraham**

**(K00302088)**

**Faculty: Business & Humanities**

**Department: Business & Financial Services**

**Supervisor: Dr. David Leonard**

**Submitted on: 31/08/2025**

**Declaration**

*I have read the Institute’s code of practice on plagiarism. I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Science in Data Analytics, is entirely my own work and has not been taken from the work of others, only to the extent that such work has been cited within the text of my work.*

*Student ID Number: K00302088*

*Name of Candidate: Helmi Kaittikkattil Abraham*

*Signature:*



*Date: 31/08/2025*

**ACKNOWLEDGEMENTS**

To my supervisor, Dr. David Leonard, who constantly encouraged me to complete this. I would like to express my heartfelt gratitude for his insightful guidance and constructive feedback during the research process. His critical observations and steady support were invaluable in shaping the direction of this thesis and ensuring its successful completion.

I am also grateful to my lecturers and classmates in the Master's in Data Analytics programme. Their stimulating teaching and enlightening discussions inspired me to develop my ideas and approaches adopted in this work.

Finally, I extend my deepest thanks to my whole family and friends, whose patience, encouragement, and unwavering belief in me provided the strength to persevere through the challenges of this journey. Their support has been the foundation of my motivation and the key to completing this work successfully.

**ABSTRACT**

This paper analyzes Airbnb price determinants along the Wild Atlantic Way (WAW) in Ireland and determines whether sentiment analysis of the guest reviews enhances predictive accuracy. The analysis was based on two research questions: (1) Does incorporating sentiment features improve Airbnb price prediction models, and (2) Which factors have the most significant impact on pricing outcomes?

The analysis used data from the Inside Airbnb project (over 30,000 listings and more than 600,000 reviews). Sentiment features obtained with transformer-based models (BERT and RoBERTa) were combined with structured listing features. After preprocessing, several machine learning models were used, such as Linear Regression, Random Forest, XGBoost, and LightGBM, and model performance was measured by R2 and RMSE.

Findings indicate that Airbnb supply is concentrated in Galway, Kerry, and Cork, with smaller counties such as Leitrim and Sligo representing niche but higher-priced segments. Reviews exhibited strong seasonality and were overwhelmingly positive, with average sentiment scores consistently high across counties. XGBoost demonstrated the best performance among the predictive models, achieving an R² of 0.71 and an RMSE of 0.35. It was also verified through the analysis of feature importance that the most salient determinants of price are property capacity, number of bedrooms, and number of bathrooms, but sentiment variables also provided an explanatory value.

The current study complements the field of tourism analytics by synthesising structured and unstructured data in a local context, which in turn expands the hedonic pricing framework and the signalling theory. Practically, the study suggests how host can improve their competitiveness by integrating traveler sentiment with the existing property features, whereas policymakers can draw on review-based measures to shape sustainable tourism strategies.

**Keywords:** Airbnb, Wild Atlantic Way, Sentiment Analysis, Price Prediction, Machine Learning, Tourism Analytics, Ireland

Table of Contents

[Chapter 1: Introduction 1](#_Toc207391618)

[1.1 Background 1](#_Toc207391619)

[1.2 Research Problem 1](#_Toc207391620)

[1.3 Aim 2](#_Toc207391621)

[1.4 Objectives 2](#_Toc207391622)

[1.5 Research Questions (RQs) 2](#_Toc207391623)

[1.6 Hypotheses 2](#_Toc207391624)

[1.7 Research Gap 3](#_Toc207391625)

[1.8 Structure of Dissertation 3](#_Toc207391626)

[Chapter 2: Literature Review 4](#_Toc207391627)

[2.1 Introduction 4](#_Toc207391628)

[2.2 Airbnb and the Sharing Economy 4](#_Toc207391629)

[2.3 Airbnb Pricing Models 6](#_Toc207391630)

[2.4 Customer Reviews as Data 8](#_Toc207391631)

[2.5 Sentiment Analysis in Hospitality Research 10](#_Toc207391632)

[2.6 Integrating Sentiment Analysis with Price Prediction 13](#_Toc207391633)

[2.7 Conceptual Framework 14](#_Toc207391634)

[2.8 Summary of Literature Review 16](#_Toc207391635)

[Chapter 3: Methodology 18](#_Toc207391636)

[3.1 Research Design 18](#_Toc207391637)

[3.2 Research Approach 19](#_Toc207391638)

[3.3 Research Strategy 19](#_Toc207391639)

[3.4 Data Collection 20](#_Toc207391640)

[3.5 Rationale for Using Secondary Data 22](#_Toc207391641)

[3.6 Data Sources 22](#_Toc207391642)

[3.7 Software & Tools 23](#_Toc207391643)

[3.8 Data Preprocessing 25](#_Toc207391644)

[3.9 Sentiment Analysis 26](#_Toc207391645)

[3.10 Feature Engineering 27](#_Toc207391646)

[3.11 Modeling Approaches 27](#_Toc207391647)

[3.12 Evaluation Metrics 28](#_Toc207391648)

[3.13 Methodological Limitations 28](#_Toc207391649)

[3.14 Summary 28](#_Toc207391650)

[Chapter 4: Results 30](#_Toc207391651)

[4.1 Introduction 30](#_Toc207391652)

[4.2 Descriptive Statistics of Listings Data 30](#_Toc207391653)

[4.3 Exploratory Analysis of Reviews 35](#_Toc207391654)

[4.4 Sentiment Analysis Results 38](#_Toc207391655)

[4.5 Price Prediction Results 39](#_Toc207391656)

[4.6 Case Comparisons and Integrated Analysis 44](#_Toc207391657)

[4.7 Summary of Findings 44](#_Toc207391658)

[Chapter 5: Discussion 46](#_Toc207391659)

[5.1 Introduction 46](#_Toc207391660)

[5.2 Interpretation of Listings Data 46](#_Toc207391661)

[5.3 Insights from Reviews and Sentiment Analysis 47](#_Toc207391662)

[5.4 Price Prediction Discussion 47](#_Toc207391663)

[5.5 Integrated Case Comparisons 48](#_Toc207391664)

[5.6 Practical Implications 49](#_Toc207391665)

[5.7 Theoretical Contributions 49](#_Toc207391666)

[5.8 Limitations 50](#_Toc207391667)

[5.9 Directions for Future Research 50](#_Toc207391668)

[5.10 Summary 51](#_Toc207391669)

[Chapter 6: Conclusion 52](#_Toc207391670)

[6.1 Overview 52](#_Toc207391671)

[6.2 Key Findings 52](#_Toc207391672)

[6.3 Contributions of the study 53](#_Toc207391673)

[6.4 Limitations 54](#_Toc207391674)

[6.5 Recommendations 54](#_Toc207391675)

[6.6 Final Conclusion 55](#_Toc207391676)

[References 56](#_Toc207391677)

**List of Tables**

[Table 2.1: Structure and Influence of Airbnb Reviews 6](#_Toc207392207)

[Table 2.2: Common Biases in Online Reviews 10](#_Toc207392208)

[Table 2.3: Summary of Sentiment Analysis Applications in Hospitality 12](#_Toc207392209)

[Table 4.1: Distribution of Airbnb Listings by County and Room Type 31](#_Toc207392210)

[Table 4.2: Average Sentiment Scores by County 39](#_Toc207392212)

**List of Figures**

[Figure 4.1: Distribution of Airbnb Prices (Wild Atlantic Way) 32](#_Toc207393312)

[Figure 4.2: Distribution of Prices by Room Type 32](#_Toc207393313)

[Figure 4.3: Correlation Heatmap of Key Variables 33](#_Toc207393314)

[Figure 4.4: Airbnb Listings Count by County (WAW) 34](#_Toc207393315)

[Figure 4.5: Average Airbnb Price by County (WAW) 35](#_Toc207393316)

[Figure 4.6: Review Volume by County 36](#_Toc207393317)

[Figure 4.7: Temporal Trends in Reviews 36](#_Toc207393318)

[Figure 4.8: Common Terms in Reviews 37](#_Toc207393319)

[Figure 4.9: R² and RMSE comparison (with vs. without sentiment, XGBoost) 41](#_Toc207393321)

[Figure 4.10: Performance comparison of ensemble models 41](#_Toc207393322)

[Figure 4.11: Feature importance ranking 42](#_Toc207393323)

[Figure 4.12: Error distribution analysis 43](#_Toc207393324)

# ****List of Abbreviations****

* **ABSA** – Aspect-Based Sentiment Analysis
* **AI** – Artificial Intelligence
* **BERT** – Bidirectional Encoder Representations from Transformers
* **CSV** – Comma-Separated Values
* **EDA** – Exploratory Data Analysis
* **GB** – Gradient Boosting
* **LightGBM** – Light Gradient Boosting Machine
* **MAE** – Mean Absolute Error
* **ML** – Machine Learning
* **NLP** – Natural Language Processing
* **RMSE** – Root Mean Squared Error
* **R²** – Coefficient of Determination
* **RoBERTa** – Robustly Optimised BERT Pretraining Approach
* **XGBoost** – Extreme Gradient Boosting
* **WAW** – Wild Atlantic Way

# Chapter 1: Introduction

## 1.1 Background

Airbnb was founded in 2008 by Brian Chesky, Joe Gebbia, and Nathan Blecharczyk, and it transformed the hospitality industry by creating an online platform that connects people seeking short-term stays with hosts offering unique spaces. According to Gao et al. (2022), Airbnb has not only disrupted hospitality but also influenced broader lifestyle and travel behaviours. Airbnb expanded to the Irish market in 2011, recognizing Ireland's compounding tourism market and popularity as a destination for international travellers. The Irish quickly rose to be one of the top destinations for co-hosts and guests of Airbnb, and hosts started developing a community to list city apartments in Dublin to rural cottages. The Wild Atlantic Way is a stunning route that follows the coast of Ireland. It starts in Donegal and goes south through Leitrim, Sligo, Mayo, Galway, Clare, Limerick, Kerry, ending up in County Cork. Tourists on this trip may find Airbnb very useful. Di Persio and Lalmi (2024) say that pricing strategies that work for the platform need to be competitive and make money. They also need to show that users value their experiences a lot.

New research (Amat-Lefort et al., 2023; Abrardi et al., 2025) has shown that service quality and how guests feel about a hotel affect how well it does its job. Reading reviews from past guests is a great way to find out how satisfied tourists were and how they felt during their stay. People who work in the hotel business are using mood analysis more and more to figure out how customers act, improve service, and make decisions. This is the computer study of what people think, feel, and say in writing (Alharbi, 2023; Erol, n.d.).

## 1.2 Research Problem

Listing prices on Airbnb are hard to guess because the type of property, its location, demand during different times of the year, and the host's methods all impact each other in intricate ways (Ghosh et al., 2023; Costa, 2025). Researchers have tried different approaches to use machine learning to predict Airbnb prices (Garcia, 2023), with many focusing on structured data such as reviews, location, and facilities. However, few incorporated the potential of text-based review data in this process.

Structured datasets could not pick up on the minor signs that reviews send about customer satisfaction, perceived worth, and service quality. Abrardi et al. (2025) say that forecast models might not be very good at showing how people really feel about the market if this information is not added.

## 1.3 Aim

To enhance the accuracy of Airbnb price prediction models for Irish listings in Wild Atlantic Way counties by integrating sentiment scores derived from customer reviews with traditional property and location features.

## Objectives

* To collect and preprocess Airbnb listings and review datasets for Ireland, ensuring high data quality and relevance for analysis.
* To perform sentiment analysis on customer reviews using advanced NLP models such as BERT and RoBERTa to quantify guest sentiments.
* To engineer sentiment-based features (e.g., average sentiment score per listing) and integrate them with traditional listing features.
* To develop and compare machine learning models for Airbnb price prediction using: Only traditional listing features, and Traditional features combined with sentiment-based features.
* To evaluate model performance using regression metrics such as RMSE, MAE, and R², and determine the contribution of sentiment to prediction accuracy.

## 1.5 Research Questions (RQs)

* **RQ1:** Does incorporating sentiment analysis of customer reviews improve the accuracy of Airbnb price prediction models for WAW counties in Ireland?
* **RQ2:** Which factors (like location, availability, sentiment, review scores) have the strongest influence on Airbnb prices?

## 1.6 Hypotheses

* **H1:** Including sentiment scores from guest reviews as features will significantly improve Airbnb price prediction accuracy compared to models without sentiment features.
* **H2:** Positive guest sentiment will be positively correlated with higher listing prices, even after controlling for location and property attributes.

## 1.7 Research Gap

Many modern Airbnb price prediction algorithms do not consider mood analysis of customer reviews when making predictions (Erol, n.d.; Alharbi, 2023). Even though there are many studies that use textual data to judge service quality, there is limited research in the field of Irish tourism that measures mood ratings based on reviews and links them to property traits to predict prices. This difference is important because guests’ and hosts' price decisions are influenced by how they feel and what feedback they get.

## 1.8 Structure of Dissertation

This dissertation will be structured as follows:

1. **Introduction** – Outlines the background, research problem, objectives, research questions, and hypotheses.
2. **Literature Review** – Critically reviews existing studies on Airbnb pricing models, sentiment analysis applications, and the role of online reviews in hospitality economics.
3. **Methodology** – Details the datasets used (listings and reviews for Ireland), data preprocessing, sentiment analysis using BERT and RoBERTa, feature engineering, and machine learning models for price prediction.
4. **Results** – Presents descriptive statistics, exploratory data analysis on the available data, sentiment analysis findings, model training outcomes, and performance metrics.
5. **Discussion** – Interprets findings, compares them with prior research, and assesses the role of sentiment in improving price prediction accuracy.
6. **Conclusion and Recommendations** – Summarizes contributions, acknowledges limitations, and suggests directions for future research, especially in integrating multi-modal data such as images and spatial analytics.

# Chapter 2: Literature Review

## 2.1 Introduction

This literature review will examine previous research on sentiment analysis, prediction models, and Airbnb pricing to provide a deeper understanding of the Irish short-term rental market and integrate their findings. By enabling people to share accommodation directly, Airbnb has transformed travel patterns worldwide. This has created both benefits and challenges for lawmakers, guests, and hosts. Many research studies have looked at what affects Airbnb prices and how reviews can change people's minds. However, not many have used advanced NLP methods like BERT and RoBERTa to find out how reviews affect prices. The goal of this project is to address these gaps that exist now and build the groundwork for a future research study that will combine sentiment-based traits with organized listing data to make price prediction models better.

The chapter is divided into several parts. Section 2.2 discusses the impact of Airbnb on travel in Ireland and globally. Section 2.3 focuses on Airbnb's pricing system, including factors that affect prices and how to predict them. Later sections explore customer feedback, tools for sentiment analysis, and prediction models that use sentiment. The aim is to show how this research contributes to broader discussions and its potential value.

## 2.2 Airbnb and the Sharing Economy

**Growth of the Sharing Economy**

Digital platforms enable people to share unused resources, which is the essence of the sharing economy. Some are often described as peer-to-peer (P2P) exchange or collaborative consumption. Key aspects include eliminating middlemen, sharing assets, and engaging based on trust. This model has evolved rapidly due to advances in technology, changing consumer preferences, and the rise of reputation systems that make people feel safer (Gao et al., 2022). Gao and his colleagues used structural topic models to look at reviews from Airbnb guests compared to hotel guests. This showed how platform-mediated services change what customers expect.

Many businesses worldwide, including those in transportation, housing, and professional services, have undergone significant changes due to the sharing economy. With Airbnb, people can rent out their extra room and make money at the same time (Di Persio & Lalmi, 2024)—a great example of this change. It was shown by Di Persio and Lalmi that regression and natural language processing can be used together to predict Airbnb cost.

The sharing economy significantly impacts the global economy. It allows some individuals and small businesses to earn money. Customers benefit from more choices, personalized experiences, and greater flexibility in pricing. However, concerns remain about regulatory compliance, the availability of goods, and the potential effects on the traditional market.

**Airbnb’s Role in Tourism**

Airbnb's efforts to bring more tourists to Ireland have helped cities like Dublin and the Wild Atlantic Way, which runs from Donegal to West Cork and is mostly countryside and along the coast. As Airbnb has become more famous in the hospitality business, it has changed how tourist infrastructure works, especially in places where there are not many traditional hotels. Airbnb has been around since 2008 and has added millions of places to stay all over the world. This has changed how people get around and how competitive places are. The platform's main benefit is that it offers one-of-a-kind living options that fit in with the neighbourhood. This makes it appealing to people from all over the world.

As more people use Airbnb, it might help local economies by encouraging more shopping, dining out, and exploring local attractions. The poll also showed that hotels are having to work harder to stay ahead of the competition, and rental prices are going up in places where there is a lot of demand. Tourists like Airbnb make economies more diverse. However, it makes it harder for cities to enforce housing rules and keep people together. As Abrardi et al. (2025) say, this shows how unclear the company's social situation is. Abrardi et al. studied how buyers act when there are flaws in online ratings. They stress how important it is to handle guest feedback correctly in order to keep prices low and maintain a strong place in the market. Table 2.1 summarises the dual structure of Airbnb reviews and their economic and psychological influence on guest behaviour.

Table 2.1: Structure and Influence of Airbnb Reviews

| **Aspect** | **Description** | **Key References** |
| --- | --- | --- |
| **Structure** | Combination of numeric ratings (e.g., cleanliness, communication, overall score) and free-text comments. | Santos et al. (2022); Subroto & Christianis (2021) |
| **Influence on Decision-Making** | Reviews guide guest booking decisions, enhance trust, and reduce perceived risk. | Marti-Ochoa et al. (2025); Vassilikopoulou et al. (2024) |
| **Economic Impact** | Positive reviews increase demand, enabling hosts to charge higher prices; negative reviews can lower occupancy rates. | Erol (n.d.); Safari (2025) |
| **Psychological Effects** | Reviews act as social proof, shaping perceptions of service quality and influencing willingness to pay. | Amat-Lefort et al. (2023) |

## 2.3 Airbnb Pricing Models

**Factors Influencing Airbnb Prices**

Many studies have explored the factors influencing Airbnb prices, often using the hedonic pricing model. The most significant traits are those that are structured, with location being one of the most accurate indicators. Properties in desirable areas, such as downtown or near popular sites, typically command higher prices (Costa, 2021). Geographic factors, such as public transit routes, buildings, and town characteristics, can be useful when using geographical data.

In general, property-specific characteristics also play a major role. Airbnb listings may differ from a normal hotel in terms of amenities such as Wi-Fi, air conditioning, or parking, which directly influence pricing. Ghosh et al. (2023) identified that the number of beds, the number of people a building can accommodate, and the number of bathrooms are some other factors that affect capacity; however, their reliance on structured attributes overlooked unstructured guest feedback that may further shape perceived value. Ghosh and his colleagues used a group machine learning method that got rid of traits that were based on amenities. They were able to keep the accuracy of their predictions by using different models, such as hidden variables gathered from host and ticket trends.

Hosting factors are also important. This is clear from the higher prices, the fact that cancellations are easy, and the 'superhost' designation. Hosts with more reviews or who have been on the platform longer might be seen as more reliable, which enables them to charge higher prices (Garcia, 2023). Garcia showed that incorporating visual elements such as photos from posts into price prediction tools significantly influenced users’ perception of value, yet the study neglected textual reviews, limiting its insights into subjective guest experiences.

Prices may increase during the holidays, sports events, and other peak visiting times of the year. With dynamic pricing systems, prices change based on how much people want to buy. These steps might be manageable by the user, or a program might do it for them. More and more people are aware that COVID-19 and other events can make short-term rental markets very different in terms of prices.

**Price Prediction in the Literature**

Conventional models like linear regression and hedonic models are widespread, but often do not capture nonlinear interactions between features and therefore, the predictive power of these models in complicated market setups is reduced. These methods helped determine the factors that influence real estate prices to rise or fall. However, these models do not always take into account the complicated, nonlinear connections between factors (Costa, 2025), even though they make sense in theory and are easy to understand. AI and machine learning methods that can work with large datasets are becoming more and more popular because of this.

Recent research is increasingly using machine-learning methods (such as support vector machines, gradient-boosting algorithms, decision trees, and random forests) to more effectively capture complex interactions between features. However, in general, these methods have lower interpretability as compared to classical statistical models. According to Ghosh et al. (2023), ensemble methods may work better than traditional models because they take into account how region, host, and property factors interact with each other. Garcia (2023) showed that adding unstructured picture data to multimodal inputs made predictions more accurate and made this method more useful.

It has become more important to have unorganised data like guest reviews and listing details. Di Persio and Lalmi (2024) integrated natural language processing techniques to get mood and theme data from review articles, demonstrating predictive gains; however, their analysis was not region-specific, raising questions about applicability to markets like Ireland. Then it was incorporated into price models. Keser (n.d.) and Erol (n.d.) did the same. Both studies used scores of positive sentiment as a prediction variable and found that better selling prices and rental rates are linked to positive sentiment.

Recent years have witnessed significant progress in deep learning, improving its applicability to text-based prediction tasks. Neural networks are particularly effective as they use transformers and patterns that repeat to figure out how different parts of text and data are connected. Liu (2021) found that putting Airbnb reviews into mood categories was more accurate than using structured-only baselines. In 2024, Long et al. discovered that older methods, such as TF-IDF or a normal bag-of-words, were better at figuring out how people felt about Airbnb reviews than newer models based on transformers, such as BERT and RoBERTa.

But even with these changes, there are still limits. Some studies cannot be used in all situations because they do not look at specific regions. The Irish Airbnb market does not have a lot of study that uses these methods, and even fewer that use large datasets with organised, written, and visual features. Putting together powerful natural language processing mood analysis with well-known models for predicting prices could help close this gap. This would help the shows and leaders by giving them more exact results and useful information.

## 2.4 Customer Reviews as Data

**Nature of Airbnb Reviews**

Marti-Ochoa et al. (2025) note that Airbnb reviews are made up of two parts: organised numerical scores and free-text notes that are not structured, but relying solely on numerical scores risks overlooking nuanced guest experiences. Platforms might just rate things by giving them clear categories, like scores for cleaning, location, and general experience. Free-text reviews capture personal experiences, emotions, and nuances that numeric ratings often overlook(Santos et al., 2022).

Reviews act as a type of social evidence, and thus enhance trust and the chance of booking; yet, the effect of such reviews depends on their affective tone as well as on their objective description. They think about both intellectual and emotional things when making their choice, like how close it is to important places and how nice the host is (Subroto & Christianis, 2021). Asmat-Lefort et al. (2023) study on buyer behaviour shows that reviews with positive emotional wording raise perceived value and the chance of buying, even when the property's objective traits are the same, highlighting how text-based sentiment can capture hidden drivers of pricing.

When prices go up or down, it may be because of what people buy. As Safari (2025) says, positive reviews may give a host more control over prices by making demand more flexible. On the other hand, whether the reviews are mixed or bad, the price might need to be reduced to stay competitive. Barri et al. (2025) discovered that ratings that make people feel strongly are more likely to be recognized and have a bigger impact on the market. This is because of things that happen in our minds, like the availability bias.

**Review Bias and Reliability**

Although reviews might be helpful, their reliability is undermined by systematic biases in how they are written and published. As an example, positivity bias is when people give more positive feedback because they are afraid of giving negative feedback or because they feel pressured to do so (Abrardi et al., 2021), inflating average scores and masking service issues. Selection bias is another issue. This is when people who had really good or really bad experiences are more likely to post reviews, which makes the mood ranges uneven (Gao et al., 2022).

The tone of reviews is also influenced by culture, where collectivist cultures prefer to avoid outright criticism, whereas individualistic contexts allow openness to criticism-based feedback. One way cooperative societies may encourage positive bias is by discouraging public criticism to preserve social harmony(Santos et al., 2022). This is not the same as societies that care more about each person. Abrardi et al. (2025) did a study that showed some hosts do everything they can to keep their good scores, which lets them charge high prices. This is because hosts may charge different amounts based on these assumptions.

It is important to keep in mind the flaws that are already in the data when adding raw review data to forecast models. There are mistakes that people make all the time when they try to predict the future that could happen if this preparation is not done before sentiment-derived features are used in price modelling.

While online reviews are a valuable source of information, they are not free from limitations. Several types of bias can influence how reviews are written and interpreted, which in turn may affect the reliability of sentiment-based price prediction models. Table 2.2 summarises the most common forms of bias identified in prior research.

Table 2.2: Common Biases in Online Reviews

| **Bias Type** | **Description** | **Potential Impact on Airbnb Research** | **Key References** |
| --- | --- | --- | --- |
| **Positivity Bias** | Guests tend to leave more positive than negative reviews due to social norms and avoidance of conflict. | Inflates average scores, masking service issues. | Abrardi et al. (2025); Santos et al. (2022) |
| **Selection Bias** | Reviews are more likely from highly satisfied or dissatisfied customers, excluding neutral perspectives. | Skews sentiment distribution, affecting predictive accuracy. | Rezazadeh Kalehbasti et al. (2021) |
| **Cultural Bias** | Review tone and rating norms vary across cultures and regions. | Requires context-aware sentiment models for accuracy. | Sánchez-Franco et al. (2021) |

## 2.5 Sentiment Analysis in Hospitality Research

**Overview of Sentiment Analysis**

Sentimentanalysis refers to the computational study of opinions, emotions, and attitudes expressed in text (Sánchez-Franco et al., 2021), offering a way to quantify subjective perceptions often missed in structured data. It operates at multiple analytical levels. At the document level, it assigns an overall sentiment score to an entire review. At the sentence level, it evaluates sentiment within individual sentences, which is useful when opinions vary within the same review. Additionally, aspect-based sentiment analysis (ABSA) extracts sentiment related to specific attributes, such as "cleanliness" or "location" (Vassilikopoulou et al., 2024).

Two dominant methodological approaches exist in sentiment analysis. The first is the lexicon-based approach, which utilizes predefined sentiment dictionaries, such as SentiWordNet, to count positive and negative words. While these methods are transparent, they often fail to capture contextual nuances, sarcasm, or domain-specific language- the key challenges in hospitality reviews. The second approach is machine learning-based, which employs supervised algorithms like SVM or Random Forest, trained on annotated datasets. Machine learning methods often outperform lexicon-based methods, especially when dealing with complex, context-dependent text (Raza et al., 2022); however, they require large annotated datasets, which can be costly to obtain. Recent years have seen a paradigm shift toward deeplearning and transformer**-**basedarchitectures, enabling more sophisticated context handling.

**Applications in Hospitality and Tourism**

According to Amat-Lefort et al. (2023), mood analysis is being used more often in the hotel industry to find out how customers feel about the level of service, but most applications remain exploratory rather than embedded in predictive models. According to Santos et al. (2022), there is a strong link between positive mood ratings, a higher number of bookings, and loyal guests.

In addition to established predictors, sentiment has been shown to provide insights into demand fluctuations and vacancy rates. According to Alharbi (2023), hotels and Airbnb hosts may quickly change their prices based on what customers say in their reviews.

Many studies have found a link between price strategies and happiness. Safari (2025) found that rentals with better general mood scores could keep their prices 12–18% higher than similar listings and not lose any renters, suggesting practical pricing benefits, though findings were region-specific and may not generalise. Some researchers, including Medpalliwar et al. (2025), found that adding mood analysis to physical characteristics could make price models up to 9% more accurate.

Table 2.3: Summary of Sentiment Analysis Applications in Hospitality

| **Study** | **Method** | **Main Findings** |
| --- | --- | --- |
| **Raza et al. (2022)** | Deep learning (LSTM, BERT) applied to Airbnb review datasets. | Transformer-based models outperform traditional ML in sentiment classification. |
| **Safari (2025)** | Sentiment analysis across U.S. regions linked to pricing. | Positive sentiment correlates with higher acceptance rates and increased prices. |
| **Rezazadeh Kalehbasti et al. (2021)** | Integrated sentiment features into price prediction models. | Improved predictive accuracy when combining textual and structured data. |
| **Vassilikopoulou et al. (2024)** | Aspect-based sentiment analysis of negative reviews. | Detailed aspect analysis identifies service improvement areas beyond overall sentiment. |

**Advances in NLP for Sentiment Analysis**

According to Raza et al. (2022), BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (a highly improved BERT version) are two transformer-based models that have recently achieved state-of-the-art performance in classifying sentiment. Since these models can read text both ways, they can find the link between context and words that other models based on bag-of-words or RNNs often miss, enabling finer sentiment classification.

Adding corpora that are important to the field could make these models better in hotel settings, where words like "cosy" and "compact" may mean different things. Cross-cultural Airbnb marketplaces are made easier by versions like mBERT and XLM-RoBERTa (Marti-Ochoa et al., 2025) that support multiple languages, though their general models may underperform without domain-specific fine-tuning. These marketplaces are important for Ireland's varied tourist population.

## 2.6 Integrating Sentiment Analysis with Price Prediction

**Existing Studies**

Algorithms designed to predict prices have found that sentiment scores are valuable for making accurate predictions. Rezazadeh Kalehbasti et al. (2021) demonstrated that adding factors based on mood to regression models made Airbnb price predictions 6–8% more accurate, though the analysis was limited to US markets. Tan et al. (2024) showed that a multimodal approach that blends scores of emotions, visual features, and data about place works a lot better than a model that only uses one modality, highlighting the need for richer data integration in hospitality research.

Similar results have been found in other studies. Mahyoub et al. (2023) used machine learning techniques like Random Forest and XGBoost on ads that also had opinion ratings. When compared to models that only looked at listing traits, these methods produced higher R² values, but their study relied on small samples, limiting generalisability. Medpalliwar et al. (2025) used deep learning to look at both number and emotional factors at the same time to improve classification across places.

Sánchez-Franco et al. (2021) used a fuzzy metatopics method to sort reviews into groups based on how they made people feel and then added these groups to price models. This method kept both the certainty of numbers and the depth of meaning.

**Research Gaps**

Despite these advancements, several gaps remain in the research. Most studies focus on global datasets or major U.S. cities, with limited research having applied these methods to the Irish Airbnb market, particularly along rural and coastal tourism regions such as the Wild Atlantic Way, where pricing dynamics differ substantially from urban contexts. Additionally, many studies rely on simple polarity scores for sentiment analysis, underutilizing aspect-based sentiment, which could provide more interpretable pricing insights. Furthermore, few models integrate sentiment with advanced features such as image data (Garcia, 2023) or spatial context (Tan et al., 2024), both of which could significantly influence perceived property value. Addressing these gaps could lead to more robust, market-specific, and multimodal price prediction models, enabling hosts and platforms to optimise pricing strategies in a competitive marketplace.

## 2.7 Conceptual Framework

The overall goal of this review is to show how the main parts of the study—listing features, sentiment scores from guest reviews, and price prediction outputs—are connected in both practical and theoretical ways. These parts are part of a larger methodological framework that includes data preprocessing, feature engineering, and modelling. As an example, the Irish Airbnb market is used to show how the analytical and descriptive methods can be used to get better prediction results by combining organised and unstructured data.

At its core, the design is based on the idea that the property characteristics and opinion scores are separate, while the expected price of an Airbnb is dependent. How these factors work together is affected by how the data is prepared, how emotion is extracted, and how price forecast models that use machine learning are used.

**Inputs**

The system gets its first set of data from two main sources:

Firstly, the organised traits that come from the listings themselves make up the Airbnb profile characteristics. The property's latitude and longitude, its neighbourhood, type (apartment, house, private room, shared room), physical features (number of bedrooms, bathrooms, and capacity), availability, host characteristics (like Superhost status and response rate), amenities, and pricing history are all listed in a CSV file. Previous studies have shown that these features are important parts of hedonic price models for holiday rentals (Mahyoub et al., 2023; Medpalliwar et al., 2025; Yang, 2021).

Then, visitors' reviews with scores based on how they made people feel, based on what they wrote in their reviews. There are scores in the Excel file that show how users feel about the words they enter. Each review is turned into a number by formulas that determine its orientation (positive, neutral, or negative) and, in some cases, its strength. Studies (Santos et al., 2022; Safari, 2025) show that reviews affect how people book and show how they feel about value and quality, which may have something to do with price changes. The method uses mood scores to look at a part of how good people think the service is that is not always taken into account in price models that only look at features (Rezazadeh Kalehbasti et al., 2021; Subroto & Christianis, 2021).

**Processes**

In between gathering basic data and making predictions, there are some preparations to do with it. Cleaning the data, removing duplicates, adding missing values, and choosing which attributes to use for raw listing and reviewing the data are all parts of data preparation. To connect review mood scores to the property features they are linked to, preparation also includes joining the two databases based on the unique listing ID. Thakur et al. (2022) and Tran (2021) say that this process makes sure that the information is always the same and is great for machine learning.

Sentiment analysis is a method that evaluates reviews and rates them based on their positivity or negativity. This approach typically relies on a list of words that are already associated with specific emotional values. Tagged datasets are used to train models in machine learning, helping them categorise emotions into groups. When it comes to hospitality mood classification tasks, transformer-based natural language processing models like BERT and RoBERTa have done better than traditional methods (Raza et al., 2022; Marti-Ochoa et al., 2021).

For example, the mean sentiment score per property can be used as an organised input variable in the pricing model to do the price prediction. This score is based on how good or bad each review is.

Machine learning is used to try to guess what Airbnb prices will be in the future. This is the last step in the planning process. Earlier research, like that by Mahyoub et al. (2023), Medpalliwar et al. (2025), and Tan et al. (2024), has looked into a number of methods, including decision trees, gradient boosting machines, random forests, and deep neural networks. This study uses both standard marketing parameters and mood scores to make the model more useful by including both objective property traits and subjective visitor reviews. Studies like Rezazadeh Kalehbasti et al. (2021) and Erol (n.d.) show that adding mood factors to the price system may improve the accuracy of predictions by adding another level of explanation.

**Output**

This lays out how much an Airbnb in Ireland will cost. This price estimate is based on more than just the regression result. It also takes into account the property's real aspects and the less visible parts of guests' experiences. This way, hosts can get the best deals and guests can see how much they are getting for their money.

**Justification for Including Sentiment Analysis**

It makes sense to include mood research as a separate variable based on the messaging theory of markets. As Amat-Lefort et al. (2023) say, this theory says that trustworthy signs, like customer reviews, may help close the knowledge gap between sellers and buyers. When people write reviews about Airbnb, they show that the service is great and that other people have used it. When conducting sentimental analysis for numbering the reviews, it can be used to organise both the practical and emotional parts of a stay, which could have an impact on price and demand (Safari, 2025; Santos et al., 2022).

Adding scores of how people feel about a property to its Airbnb prices shows that the prices are affected by both set property traits and market forces that depend on how people feel. One house might cost more than another, even if they look the same. Some people who stay at the first house leave lots of good reviews, but some people who stay at the second house leave mixed or bad reviews. Abrardi et al. (2025) showed that online review bias and opinion trends may have a big effect on how hosts set their prices. This link fits with their findings.

In line with recent progress in multimodal machine learning (Tan et al., 2024), this work uses mood analysis to bring together organised and unstructured data analysis.

## 2.8 Summary of Literature Review

This review looked at how mood analysis is becoming more important in the hotel business, how Airbnb prices are affected by many things, and how customer reviews may have certain traits and biases. First, it was shown how important Airbnb is to the sharing economy, especially in Ireland's tourism industry. Mahyoub et al. (2023) and Medpalliwar et al. (2025) both do studies that show the price of a home is greatly affected by its qualities, where it is located, and the state of the market. Through Section 2.4, it is clear how customer reviews can be used as data. These have both number scores and free-form notes. Review articles change more than just how much people pay. It also changes what people think and what they buy. According to Abrardi et al. (2025), scores may not reflect the real quality of service because of cultural factors, positive bias, and selection bias.

Section 2.5 explains how sentiment analysis can be used in hotel research. Then the definition, different types of analysis (document, phrase, and aspect-based), and the different ways to do it (lexicon-based vs. machine learning-based) are explained. Emotional intelligence is related to many things in the hotel and tourism industries, such as how people see service quality, how they predict demand, how many rooms are occupied, and how they set prices (Amat-Lefort et al., 2023; Alharbi, 2023; Erol, n.d.). New developments in natural language processing, especially transformer-based models, have made it much easier to accurately and contextually understand how to categorise mood in big review datasets (Raza et al., 2022; Marti-Ochoa et al., 2021).

Section 2.6 talks about how mood analysis and price forecasts might work together. Research (Rezazadeh Kalehbasti et al., 2021; Safari, 2025) has shown that adding emotion traits to machine learning models makes them more accurate. Even so, there are still big holes in the research. Irish people do not pay enough attention to Airbnb data, especially when it comes to touring the country and the coast. Also, not much study has been done on how mood scores can be used as a structural factor in price models that are based on hedonics or machine learning. Tan et al. (2024) say that mood integration is not good enough when added to more complex features like location models and picture analysis. In Section 2.7, data preparation, emotion extraction, and machine learning analysis, which combine the results into a single model, are explained. This plan linked the parts of a listing and the emotional opinions to the price that was expected. The use of mood analysis was made possible by theoretical models like signalling theory and real-world data that showed how visitors' opinions had a big impact on market prices.

It has been found that organised property data and sentiment-based information may work together to improve the accuracy of predictions about Airbnb prices. This is especially important for the Irish market because tourism has a big effect on the country's income, and tourists and the places they stay have very different experiences in different parts of the country. The goal of this study is to fix these problems so that future research and price tactics in the short-term rental market can be better.

This review highlights the importance of integrating sentiment analysis into pricing models, and the next chapter outlines the methodological approach adopted in this thesis. It will demonstrate how to apply this mental model in practice. It will cover the process of data collection, data cleaning, analysis of guest reviews, incorporating that with organised property data and the use of machine learning models to predict prices. This framework will serve as the foundation for the upcoming study.

# Chapter 3: Methodology

This section outlines the methodology adopted for the study. It focuses on the short-term rental market in Ireland (especially in the WAW counties) and the relationship between Airbnb's features, guest reviews, and price predictions. This paper describes the study's design, the strategies and methods employed, the data collection process, and the rationale behind these choices. Adopting a quantitative, data-driven approach, the study utilizes secondary data from publicly available Airbnb records. Machine learning, natural language processing (NLP), and statistical analysis are applied to develop a prediction model that combines both structured listing data and unstructured review data.

## 3.1 Research Design

The study plan lays out the methods that will be used to answer the research questions and test the theories. A secondary data analysis approach was used along with a quantitative, correlational, and predictive research method in this study.

The study aims to find numeric links between factors such as traits, ratings of how people feel about them, and expected prices, while also creating forecasting models. This creates a focus on numbers. A correlational design can be used to check the links between the independent variables (the quality of the listings and the ratings of how people feel about them) and the dependent variable (the price of an Airbnb room) without making any changes to the variables.

The central aim of this study is to predict Airbnb listing prices. Ultimately, machine learning models are built that use data from sentiment analysis to guess how much an Airbnb listing will cost. This study aims to adapt this method to the wild Atlantic way Irish market by looking at previous work that has used predictive modelling successfully in similar situations (Mahyoub et al., 2023; Rezazadeh Kalehbasti et al., 2021; Medpalliwar et al., 2025).

This method integrates different data types by combining numerical and categorical listing features with sentiment scores derived from unstructured text. As the literature review points out, there is a knowledge gap here because most of the previous studies on predicting Airbnb prices in Ireland did not use both datasets. Alternative approaches such as primary data collection (surveys/interviews) and lexicon-based sentiment analysis were considered. However, given the availability of large-scale secondary data (Inside Airbnb) and the superior contextual performance of transformer models, a quantitative, secondary-data design with advanced NLP and ML was deemed most appropriate.

## 3.2 Research Approach

This study adopts a deductiveapproach, beginning with theoretical underpinnings and proceeding to final validation. Using well-established theories and empirical data on property characteristics, consumer perceptions, and pricing (Amat-Lefort et al., 2023; Safari, 2025; Lancaster, 1966; Spence, 1973; and the hedonic pricing model), this study used the right kind of deduction.

Using a logical framework, the first step in research is to come up with well-thought-out theories about how listing features and opinion scores affect Airbnb prices. Previous research has shown that the tone of reviews can affect demand, rental rates, and prices (Rezadeh Kalehbasti et al., 2021; Santos et al., 2022). This supports the theory that this study is trying to prove. The study used statistics and machine learning methods to check if the claims were true in the Irish market.

The deductive method makes sure that the methods are consistent by going from theory to data analysis in a rational way. When a lot of data is used, this scientific method is needed to make sure that the results are real and can be tried again.

## 3.3 Research Strategy

The study strategy spells out how the plan and method will be carried out. The focus of this study is on Airbnb rentals in WAW of Ireland. It uses a case study method and stresses how important large datasets are for observational research. Case study methods are usually linked to qualitative research, but computational social scientists are slowly realising how important they are in a quantitative, data-driven setting (Yin, 2018). These are the main points that emphasise why the case study method is important here.

Hospitality and tourism in Ireland also depict a different picture across the year due to the influence of cultural events and the differing origins affirmed by the international visitors. Looking at the context of Irish tourism, especially in the Wild Atlantic Way, will give great insight into how the dynamics of Airbnb usage occur in that environment. Furthermore, Airbnb provides comprehensive data at the property level, which allows conducting an in-depth analysis of listing features and user-generated content since the information regarding the study is easily accessible. Lastly, since review analysis and price prediction in short-term rental markets are of global significance, the research findings will still be very valuable not just to the host and guests but also to the policymakers and tourism officials in Ireland.

In this method, machine learning was used by the researchers as the main tool of analysis, and a systematic set of steps was followed to achieve methodological rigour. The data was prepared by first cleaning the data, normalising, and integrating both structured features of the listings and unstructured features of the guest reviews. The foregoing preprocessing step has made the information datasets consistent, accurate, and prepared to be analyzed. Subsequently, more sophisticated natural language processing models, namely BERT and RoBERTa, were applied to the review data to perform sentiment analysis and obtain sentiment scores, so that the study is able to classify and quantify the guest emotions systematically. A set of regression and ensemble learning models, Random Forest, Gradient Boosting, and XGBoost, was then applied to both conditions with and without sentiment-related factors as explanatory factors. This comparative modelling framework was used to determine the incremental advantage of the inclusion of the guest sentiment in addition to the conventional listing features. By integrating such computational methods with theoretical underpinnings, the study was in a position to consistently relate the means of data science to the research problem that it addressed and thus come up with insights that were not just useful in describing the current dynamics but also in anticipating the future situation in the Irish Airbnb market.

The research strategy was operationalised through a structured workflow. First, relevant Airbnb datasets were collected from Inside Airbnb. Next, the data underwent preprocessing to clean, merge, and standardise variables. Sentiment analysis was then applied to guest reviews using BERT and RoBERTa models, and the resulting scores were integrated into the dataset through feature engineering. Machine learning models were subsequently trained using both structured listing attributes and sentiment-based features. Finally, predictive performance was evaluated using standard regression metrics. This systematic workflow ensures a consistent progression from raw data to interpretable model outcomes.

## 3.4 Data Collection

The secondary data used in this study comes from Inside Airbnb, a website that offers information taken from Airbnb's main website. Due to their size and organisation, Airbnb's statistics are often used in policy analysis and scholarly study (Cox, 2025). All datasets were sourced from the same monthly release of Inside Airbnb to ensure temporal consistency.

There are two main sets of data used in this study. The first is the listings dataset, which contains structured information at the property level, including listing ID, address, host details, property characteristics, features, availability, historical prices, and other relevant attributes. These organised variables form the foundation of the predictive models by providing the numerical and categorical inputs required for analysis. The second dataset is the reviews collection, which comprises both structured and unstructured information generated by guests. Along with identifiers such as reviewer ID, listing ID, and review date, it includes the textual content of guest reviews. This unstructured text serves as the essential source material for sentiment analysis, enabling the calculation of sentiment scores that can be incorporated into the modelling process alongside the structured listing data.

The following study steps were used to obtain and prepare the data to be analysed. The review and listings data were all downloaded as CSV files, and the Irish data was a simple, straightforward find within the Inside Airbnb twelve-month snapshots. All the datasets were downloaded on the same release date in order to have temporal consistency. The datasets were then read into Python with the Pandas library and cleaned, inspected, and merged. The missing values were replaced with estimates based on suitable imputation methods, including mean and mode imputation, respectively, and in other cases by deletion. Subsequent preprocessing steps were the elimination of replicated data, simplification of categorical variables by assigning them to numeric codes and removal of unnecessary rows in the data to increase data consistency. The two datasets were then joined on the listing\_id variable as the key, and thus the reviews matched with the respective listing available. The unstructured review texts also needed to be prepared before they could be used in natural language processing. This was done by eliminating undesirable entities like special symbols, lesser words and HTML code, and creating clean textual data to be analysed based on the sentiment.

There were over one million reviews, and thirty-one thousand entries of listings present in each data collection. Hugging Face Transformers was used for natural language processing, and Python's Pandas and NumPy tools were used to quickly look through them. If one intends to use Inside Airbnb data, it is important to consider ethics and stick to the guidelines for data usage and privacy protection. Since Inside Airbnb provides instructions on how to use the data, private information about renters or guests was not shared. Moreover, they do not possess any private information beyond what is already available on Airbnb accounts.

## 3.5 Rationale for Using Secondary Data

This kind of material is useful because it is important, large, and easy to find. In-depth talks or polls with Airbnb hosts and guests would have taken a lot of time and work, and they would have been skewed in a lot of ways. But Inside Airbnb puts together a huge set of data that shows how the market as a whole is doing. The objective parts of lists and the biased parts of guest reviews are both in this material.

Other researchers in this field, such as Rezazadeh Kalehbasti et al. (2021), Santos et al. (2022), and Medpalliwar et al. (2025), have used Inside Airbnb files. This facilitates direct comparison with prior studies and strengthens the contribution of this research. This comparison makes the study's addition to the field stronger by letting us directly judge how adding mood analysis might improve the accuracy of predictions compared to traditional feature-based models.

Previous research on predicting how people feel about hospitality has pointed the way for future continuous extensions of this study, such as keeping an eye on how opinion and price change over time. And this is possible because Inside Airbnb data covers a lot of time (Amat-Lefort et al., 2023; Safari, 2025).

## Data Sources

* **listings.csv (Inside Airbnb)**

The listings.csv file from Inside Airbnb serves as the primary dataset for this study, containing structured property-level information on Airbnb listings across Ireland. It includes key variables such as listing ID, host ID, property type, room type, number of bedrooms, number of bathrooms, amenities, geographic coordinates (latitude and longitude), and price. These attributes form the core source for the numerical and categorical features used in the price prediction models. The dataset also records review-related metrics, including overall rating, cleanliness, communication, and location scores. Notably, prices are stored as strings with currency symbols, which required cleaning and conversion into numeric values before further analysis. Extremely unrealistic prices (e.g., pricing for some castles) were flagged and excluded to reduce distortion

* **4\_years\_reviews.xlsx**

The 4\_years\_reviews.xlsx file is a curated dataset of guest reviews spanning a four-year period, corresponding to the listings represented in *listings.csv*. It contains variables such as review ID, listing ID, review date, reviewer ID, and the full text of each review. This dataset provides the unstructured textual data needed for sentiment analysis through natural language processing (NLP) techniques. It also enables temporal analysis of trends in guest sentiment and review volume over time. Since multiple reviews are available for each listing, aggregation strategies such as calculating mean or median sentiment scores were applied to generate representative features for modelling.

* **county\_code.xlsx**

A supplementary dataset containing the mapping of county names to county codes for Ireland was also used. This file supports geographic standardisation of listings, making it possible to filter properties located within the Wild Atlantic Way (WAW) counties and conduct spatial analysis. In addition, it ensures consistency in location-based variables when integrating multiple data sources, reducing errors in regional comparisons.

## 3.7 Software & Tools

This paper has also used a diverse mix of Python libraries, machine learning frameworks and analytic environments to aid in the preparation of the data, modelling and visualisation. The selection of tools was carried out due to their effectiveness, stability, and high use in both research and industrial spheres, which makes the approaches employed in this paper transparent, reproducible, and follow good practices.

**Core Python Libraries:** TheMajority of data preprocessing and numerical computations were done with Pandas and NumPy. Pandas was a must when it came to the reading, cleaning, transforming**,** and merging of structured data sets such as listings and reviews, and NumPy was necessary in numerical operations**,** which were particularly able to handle large computations effectively. The scikit-learn library was used as the foundation of the machine learning process and the selection of the models, which includes a nice set of tools to choose the model and train it, evaluate the model and its performance, tune it**,** and perform the process of scaling for features. All these libraries helped to process structured data efficiently and implement standardised machine learning procedures with reasonable reliability.

**Natural Language Processing (NLP):** We used the Hugging Face Transformers library that gave us access to the current state-of-the-art pre-trained models to conduct analysis in the unstructured text. To perform sentiment analysis on reviews by guests, BERT (nlptown/bert-base-multilingual-uncased-sentiment) and RoBERTa (cardiffnlp/twitter-roberta-base-sentiment) were utilised. These models, together with their tokenisers, allowed the transformation of raw textual information into structured sentiment scores, in the form of tokenisation, truncation, padding, and inference. The ability to be incorporated into GPU environments also enabled us to interactively process over one million reviews, which would otherwise be computationally expensive.

**Data visualisation**: The Python-based data visualisation tools assisted data exploration and presentation, coupled with the external data visualisation tools. During EDA, Matplotlib and Seaborn were heavily utilised to generate descriptive plots, heatmaps of correlation coefficients, and distribution charts. In addition to Python, Tableau and RapidMiner could also be used, as supplementary tools to perform exploratory visualisations on location data. Such a selection of tools guaranteed the depth of statistics and the clarity of the visualization.

**Analytical environments**: The practical implementation of this study was split into several computational environments, according to the task. Jupyter Notebook became the main tool to work interactively and analyze the data as well as to document it; it helps to combine both narrative text and executable code. Google Colab was especially used in large-scale sentiment analysis, where the GPU acceleration reduced the processing time of the transformer-based models. RapidMiner was also used as an extra environment for some exploratory activities, and Tableau was incorporated as a front-end environment to offer polished visualisations that the stakeholders could interpret. Collectively, these environments provided an environment of efficient experimentation, scalability and presentable results.

By these tools and environments, it was possible to institute an end-to-end data processing, sentiment analysis, predictive modelling and visualisation pipeline. Such a systematic approach guaranteed that each of the methodology steps was facilitated with a well-proven and solid software, including initial cleaning of raw datasets and evaluation of complex machine learning models.

## 3.8 Data Preprocessing

**Cleaning Prices**: The price variable in listings.csv required preprocessing before use in modelling. Prices were originally recorded as strings containing currency symbols (such as “€” or “USD”) and formatting characters like commas for thousands separators. To ensure compatibility with numerical analysis, a cleaning procedure was implemented in Python using str.replace and astype(float) to convert the values into numeric format. Following this conversion, outliers—such as extremely high or unusually low prices far from the overall mean—were identified and flagged for further inspection to maintain data quality.

**Handling Missing Values**: For missing data, different strategies were applied depending on the variable type. Missing numerical values, such as the number of bathrooms or bedrooms, were imputed using the median values within the same property type, ensuring that extreme outliers did not overly influence the replacement. Missing categorical variables, such as property type or room type, were imputed using the mode calculated within the corresponding neighbourhood or county to preserve local patterns. Finally, rows with missing critical identifiers, such as the listing ID, were dropped entirely, as they could not be reliably linked to other datasets.

**Removing Duplicates**: Duplication of records was also checked to ensure data consistency. Duplicates were detected based on listing IDs as well as the entirety of property attributes. In areas where duplicates have been identified, only the latest or most comprehensive record was kept, and the other redundant rows were discarded. This preparation assisted in avoiding bias in the analysis, which was caused by repeated listing.

**Scaling**: Numerical features were standardised using z-score scaling where required (e.g., for regression models), while tree-based models used raw values

**Categorical** **Encoding**: Categorical variables (e.g., property type, room type) were transformed using label encoding for tree-based models, and one-hot encoding for regression models.

**Filtering for Relevant Irish Counties**: A geographic filtering was used to centre the analysis on the Wild Atlantic Way (WAW) counties of Ireland. By populating the county\_code.xlsx mapping file, only the listings in the specified WAW counties were retained. This was also satisfactory in ensuring that any international or wrongly geocoded entries that were at times present in the Inside Airbnb dataset were notified and omitted, thus improving the consistency and accuracy in the geographical domain of the research.

**Merging Datasets**: To prepare the final dataset, multiple sources were merged. First, listings.csv was combined with county\_code.xlsx to attach county codes and filter the data to include only properties along the Wild Atlantic Way (WAW). Next, the dataset was merged with 4\_years\_reviews.xlsx on the listing\_id field, linking each review to its corresponding property attributes. A left join strategy was used for this step to ensure that all listings were retained in the dataset, even if they did not have any associated reviews. These review-less cases remained useful for establishing baseline predictions in the modelling stage. The result in the form of a combined dataset included both structured listing annotations and unstructured review text and formed a new core to the analysis. At this point, data was cleaned, merged and standardised and could be used at the next stage of feature engineering, sentiment analysis, and model building.

## 3.9 Sentiment Analysis

This study evaluated lexicon-based and classical ML sentiment methods as baselines, but adopted transformer models (BERT/RoBERTa) due to their bidirectional context handling and consistently higher accuracy on review text. This choice prioritises predictive validity over transparency; model outputs are therefore complemented with feature-level interpretation at the modelling stage. Before inference, reviews were preprocessed using model-specific tokenisers, which converted raw text into token IDs. Truncation and padding were applied to ensure consistent sequence lengths across inputs. The tokenised text was then passed to the pre-trained models through the Hugging Face Transformers library, producing sentiment probabilities. These outputs were subsequently mapped to numerical sentiment scores. For BERT, the model’s native 1–5 star ratings were used directly as numeric sentiment scores. For RoBERTa, the predicted labels were mapped to numeric values: positive (label\_2) = 5, neutral (label\_1) = 3, and negative (label\_0) = 1. These scores were then aggregated at the listing level to generate an average sentiment measure for each property. Batch processing and GPU acceleration in Google Colab were used to handle the computational demands of large-scale review analysis. Although the lexicon-based TextBlob model was initially tested as a benchmark, it was excluded from the final analysis due to its limited ability to capture context compared with transformer-based approaches.

## 3.10 Feature Engineering

Feature engineering was carried out to integrate sentiment-based variables with traditional listing characteristics. First, a sentiment score per listing was generated by aggregating all associated reviews. Both mean and median sentiment values were calculated, producing two sets of features derived separately from BERT-based and RoBERTa-based models, allowing for comparative analysis of their predictive value. In addition to sentiment, other groups of features were included. Geographic variables captured spatial information such as county code, latitude, and longitude. Listing attributes reflected property-specific characteristics, including property type, room type, number of bedrooms, and number of bathrooms. Popularity indicators incorporated measures such as the total number of reviews, review scores across categories (e.g., cleanliness, communication, location), and availability metrics. Together, these features provided a balanced representation of objective property traits, spatial context, and subjective guest perceptions for use in the modelling stage.

## 3.11 Modelling Approaches

To assess the predictive potential of sentiment features together with the traditional listing attributes, a set of machine learning models was applied. The dataset was split into 80% training and 20% testing. Baseline models, Linear Regression was selected due to its simplicity and ease of interpretation and was used as the comparator against which complex models were evaluated.

Oversized models included a collection of tree-based methods of ensemble, which are more applicable in addressing non-linear relationships and heterogeneous data. Random Forest model was used to create an ensemble of decision trees and reduce overfitting to improve overall accuracy by means of bootstrapping and feature sampling. Moreover, Gradient Boosting techniques were utilised to allow complex interactions of features to be modelled by training subsequent trees to rectify the mistakes of the previous ones. Two popular implementations compared are XGBoost, which applies regularisation to regularise and provide efficient calculation, and LightGBM, optimised to have quicker training and larger datasets whilst still delivering promising predictive performance. These algorithms were selected to balance non-linear predictive power (tree ensembles) with interpretability and benchmarking (linear regression), while alternative deep models were excluded to avoid overfitting and maintain explainability.

## 3.12 Evaluation Metrics

The performance of the models was measured using three popular regression metrics. Root Mean Squared Error (RMSE) was used as a form of measure because it is more sensitive to bigger errors, and thus, ideally suited to monetary forecasts where outliers in the actual prices are critical. Mean Absolute Error (MAE) was also computed to give an easily interpretable measure of the average error of prediction stated in currency terms. Lastly, the Coefficient of Determination (R2) was employed to show what proportion of the variance in the Airbnb prices was accounted for by the model, and this gave us a simple measure of the total explanatory power. Collectively, these metrics gave a consistent perspective on measures of both accuracy and interpretability in model comparison.

## 3.13 Methodological Limitations

The present research is based on publicly accessible Inside Airbnb snapshots, which might not reflect off-platform dynamics and within-month price changes. The ratings made by guests may be positive, selective, and culturally biased, which may overstate sentiment indicators. Transformer models are computationally expensive; what is more, they can lack domain-specific fine-tuning and fail to capture the subtleties of Irish vernacular or context. Lastly, the observational design defines relationships, but not cause and effect.

## 3.14 Summary

This study was performed to figure out how much an Airbnb listing will cost in Ireland's WAW counties. It used a quantitative, data-driven method that combined organised property criteria with unstructured textual ratings. After preprocessing to correct price errors, fill in missing data, and apply location-based filters, secondary datasets from Inside Airbnb were compiled. Additional county data files were also incorporated. Here for sentimental analysis, BERT (nlptown/bert-base-multilingual-uncased-sentiment) and RoBERTa (cardiffnlp/twitter-roberta-base-sentiment), two NLP models that had already been trained, were used to get scores from guests’ comments about how they felt about their stay. Sentiment-based features were generated through feature engineering, combining them with structured property information. It was based on facts about the property, such as its location, available services, and other amenities. To predict prices, both simple methods (such as linear regression and decision trees) and more complex methods (including random forest, XGBoost, and LightGBM) were applied. Evaluation metrics, such as R², MAE, and RMSE, were used to assess the performance of these models. For data processing, Python packages like pandas, NumPy, and scikit-learn were used, while transformers were utilized for analyzing the reviews. For analyzing and visualizing the data, graphics tools like seaborn and matplotlib were used. Jupyter Notebook was primarily used for analysis and modelling, while Google Colab was also helpful for conducting sentiment analysis. In combination, these procedures directly address RQ1 by testing whether sentiment features improve predictive accuracy, and RQ2 by quantifying the relative influence of structural, locational, and sentiment variables on price.

# Chapter 4: Results

## 4.1 Introduction

This chapter presents the results of the analysis carried out on Airbnb listings and guest reviews across the nine counties of Ireland’s Wild Atlantic Way (WAW). The purpose of this stage is to provide a detailed overview of the short-term rental market in the region and to examine how structured property features and unstructured guest reviews contribute to predicting listing prices. The analysis proceeds in four stages.

First, descriptive statistics are reported for the listing data, including county-level distributions, property types, and price patterns. Second, an exploratory analysis of guest reviews is presented to highlight volume, length, and frequently occurring terms. Third, sentiment analysis is undertaken using three different approaches — TextBlob (lexicon-based), BERT (multilingual transformer), and RoBERTa (Twitter-based transformer) — to generate sentiment features that can be integrated with listing attributes. Finally, the predictive modelling stage evaluates the performance of regression and ensemble learning models with and without sentiment features to assess their effect on Airbnb price forecasting.

Throughout the chapter, results are illustrated using tables, graphs, and maps generated in Jupyter Notebook and RapidMiner. These outputs provide both statistical and visual insights into the Airbnb market along the WAW.

## 4.2 Descriptive Statistics of Listings Data

The analysis was initiated using the datasets of listings on Inside Airbnb. Of the initial 31,907 records comprising a listing of Ireland, a filter was used that retained in accordance with nine counties of the Wild Atlantic Way (Donegal, Leitrim, Sligo, Mayo, Galway, Clare, Limerick, Kerry, and Cork). This left 15,814 listings, which were used for subsequent analyses. Prices were limited to a ceiling of €1,000 at night to reduce the effect of outlying observations.

The results of the counties' comparison indicated that most of the listings were located in Galway (2,859), Kerry (2,842), and Cork (2,442). In contrast, smaller counties such as Leitrim (306), Sligo (506), and Limerick (424) accounted for a relatively minor share of listings. Regarding accommodation type, it can be seen that entire homes or apartments prevail in the WAW market and make up more than 70 per cent in most of the counties. The proportion of private rooms varies between 16 and 41 percent; shared rooms are very uncommon (<1%) in all counties. Overall average prices as shown in the WAW counties were between (144) in Limerick and (176) in Galway, Kerry, and Leitrim, with most counties falling in the range of (150-170). Surprisingly, the micro-counties (Leitrim and Sligo) use an average price equal to or higher than the larger counties. A detailed breakdown of listings by county, room type, average price, and total supply is presented in Table 4.1.

Table 4.1: Distribution of Airbnb Listings by County and Room Type

| **County** | **Entire Home (%)** | **Private Room (%)** | **Shared Room (%)** | **Average Price (€)** | **Number of Listings** |
| --- | --- | --- | --- | --- | --- |
| Clare | 75.5 | 24.0 | 0.1 | 166 | 1,503 |
| Cork | 71.7 | 28.1 | 0.0 | 158 | 2,442 |
| Donegal | 83.0 | 16.7 | 0.0 | 155 | 2,269 |
| Galway | 68.9 | 29.9 | 0.7 | 176 | 2,859 |
| Kerry | 74.8 | 24.3 | 0.1 | 176 | 2,842 |
| Leitrim | 75.2 | 24.8 | 0.0 | 176 | 306 |
| Limerick | 56.6 | 40.6 | 0.2 | 144 | 424 |
| Mayo | 75.4 | 24.4 | 0.3 | 153 | 1,441 |
| Sligo | 79.4 | 20.4 | 0.2 | 146 | 506 |

Visual inspection of the price distribution, Figure 4.1, confirmed a heavy right skew, with a small number of luxury properties priced far above the county's mean. This skewness is typical in accommodation markets and indicates the need for log-transformation of price in regression modelling.

Figure 4.1: Distribution of Airbnb Prices (Wild Atlantic Way)

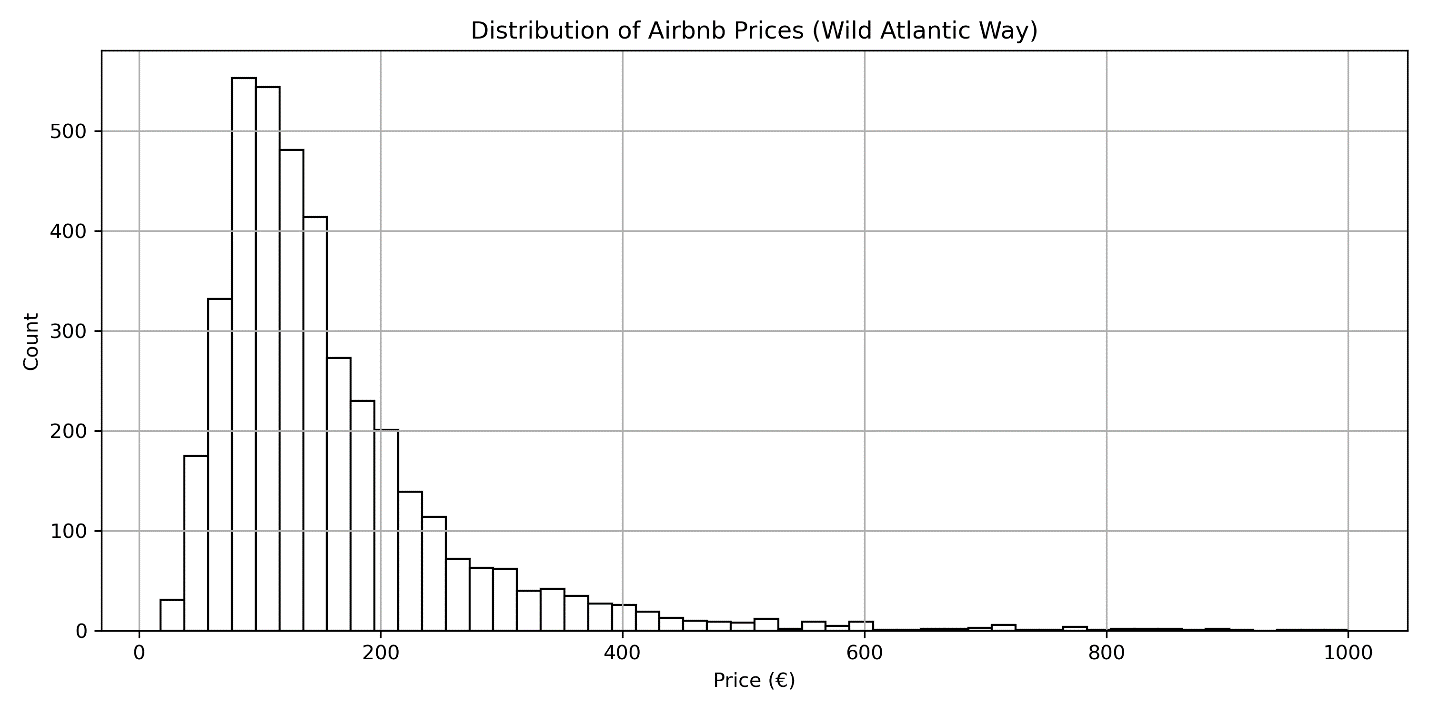
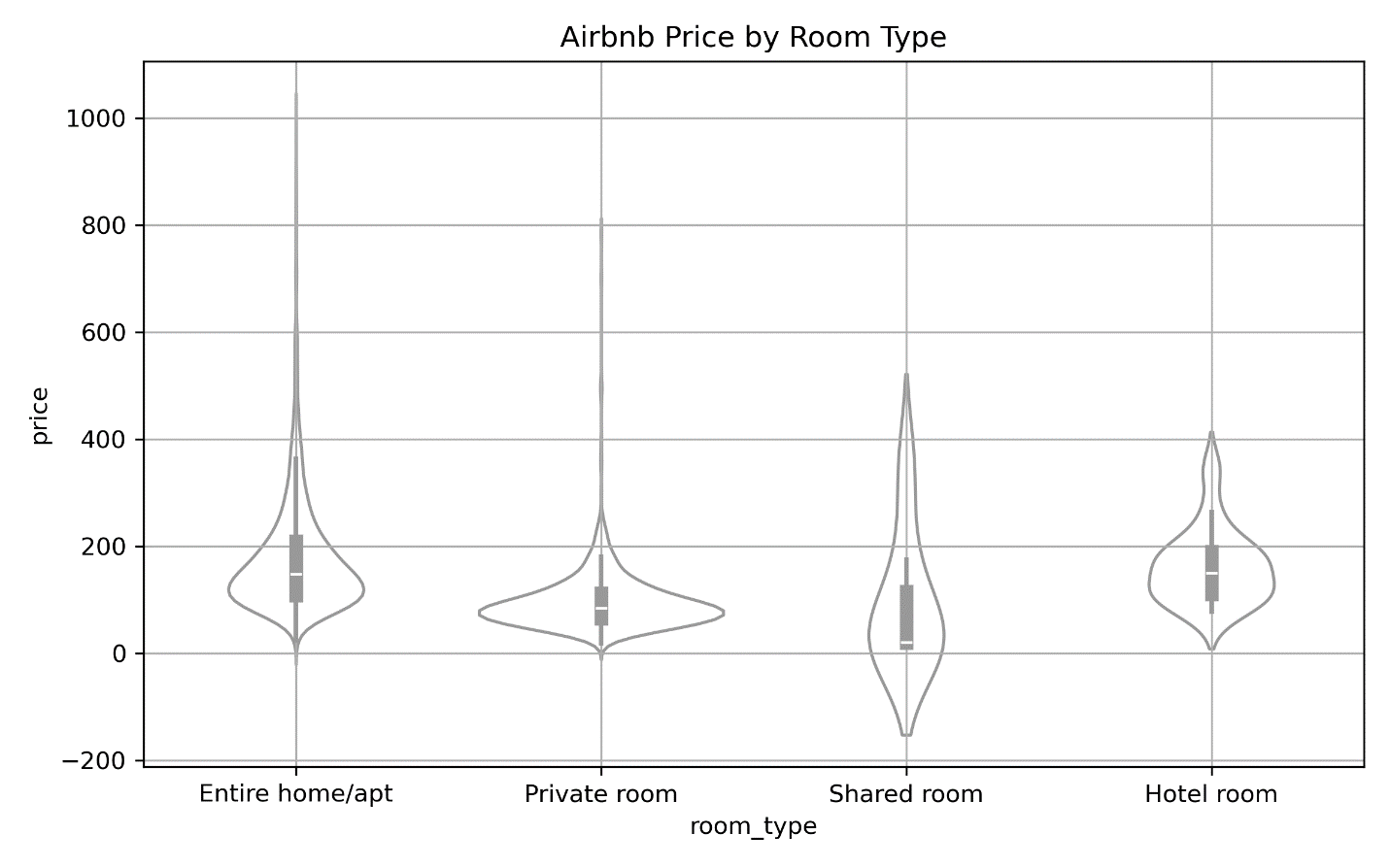


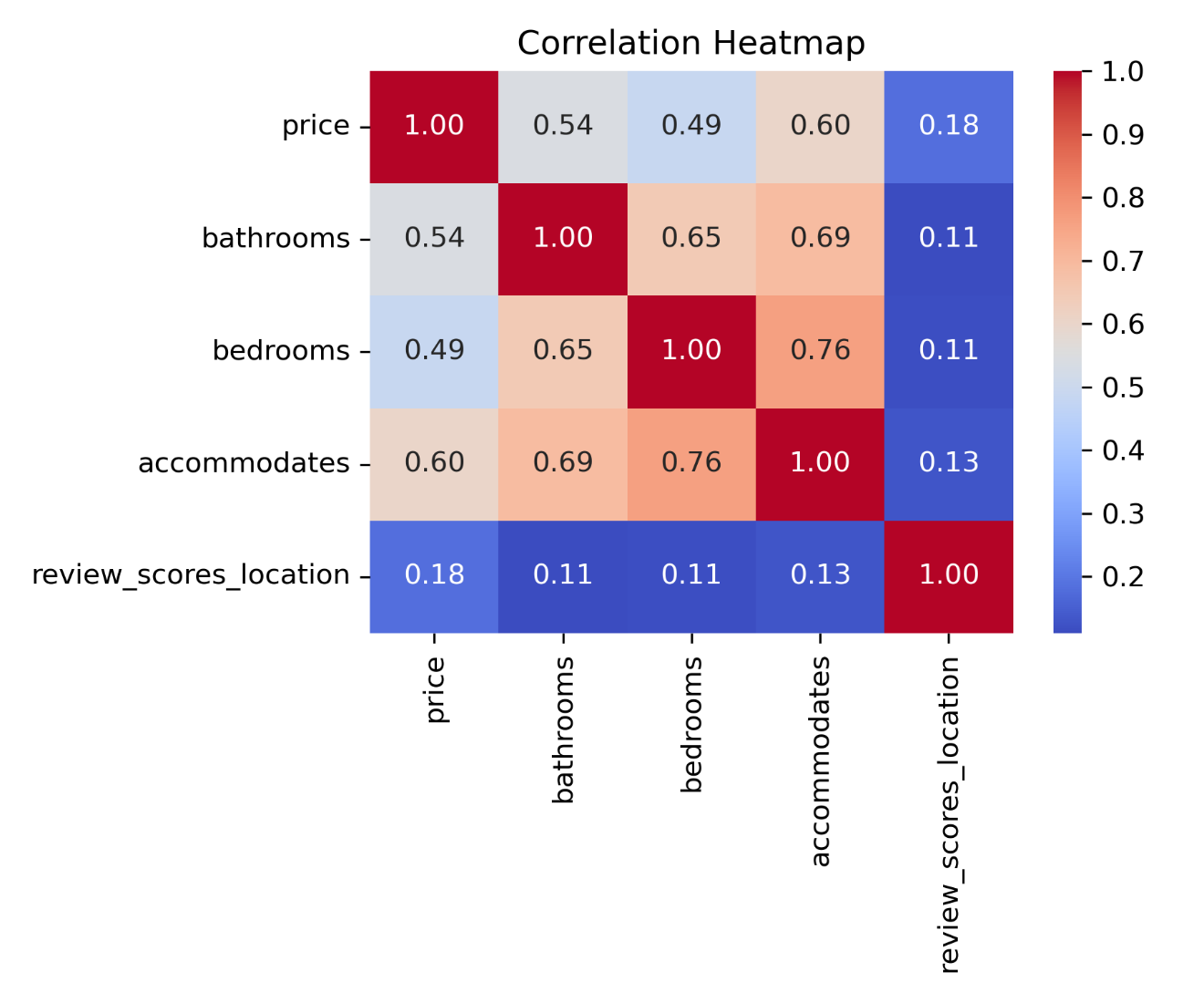
Figure 4.2: Distribution of Prices by Room Type



The violin plot (Figure 4.2) shows the distribution of nightly rates of various types of rooms in the WAW dataset. The most populous category is entire homes and apartments, with the most significant prices reaching above the whole-home outlay of over 800. The low end is more in shared rooms (although nearly all rooms cost under 100), with some cheaper private rooms as well. Hotel rooms are also less common in the data set, but follow a similar pattern of distribution as entire homes, at a slightly lower median price. The elongated tails in all levels indicate the existence of the outliers, which justifies the pricing capping of the values at 1000 Euros per night.

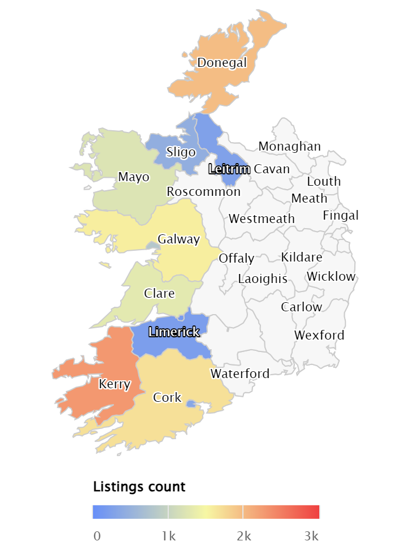
Further exploratory analysis of correlations showed that price is moderately associated with structural attributes such as the number of bedrooms and bathrooms, while review scores and availability metrics showed weaker correlations.

Figure 4.3: Correlation Heatmap of Key Variables



The heat map (Figure 4.3) demonstrates the levels of correlation concerning price and a few structural and review-related features. Price correlates most strongly with ‘accommodates’ (0.60), then with bathrooms (0.54) and bedrooms (0.49). The two variables are also closely related to each other (bedrooms-accommodates 0.76). In comparison, review\_scores\_location correlates with price only 0.18.

Figure 4.4: Airbnb Listings Count by County (WAW)

****

The map (Figure 4.4) represents the allocation of the Airbnb rentals in the nine counties of the Wild Atlantic Way (WAW). The geographical concentration is the highest in the cities of Cork, Kerry, and Galway, each of which has more than 2,000 properties. Leitrim and Sligo have, by comparison, significantly lower densities of available listings.

The aim of the choropleth map (Figure 4.5) is to show the average nightly price of Airbnb listings in the counties within the WAW. Results indicate that Cork, Kerry, and Galway had the most expensive hotels, averaging over 150 euros per night. The averages in Clare, Limerick, and Mayo lie in the mid-range; they are rather average in price, and Leitrim, Sligo, and Donegal are located at the bottom of the chart (less than 100).

Figure 4.5: Average Airbnb Price by County (WAW)

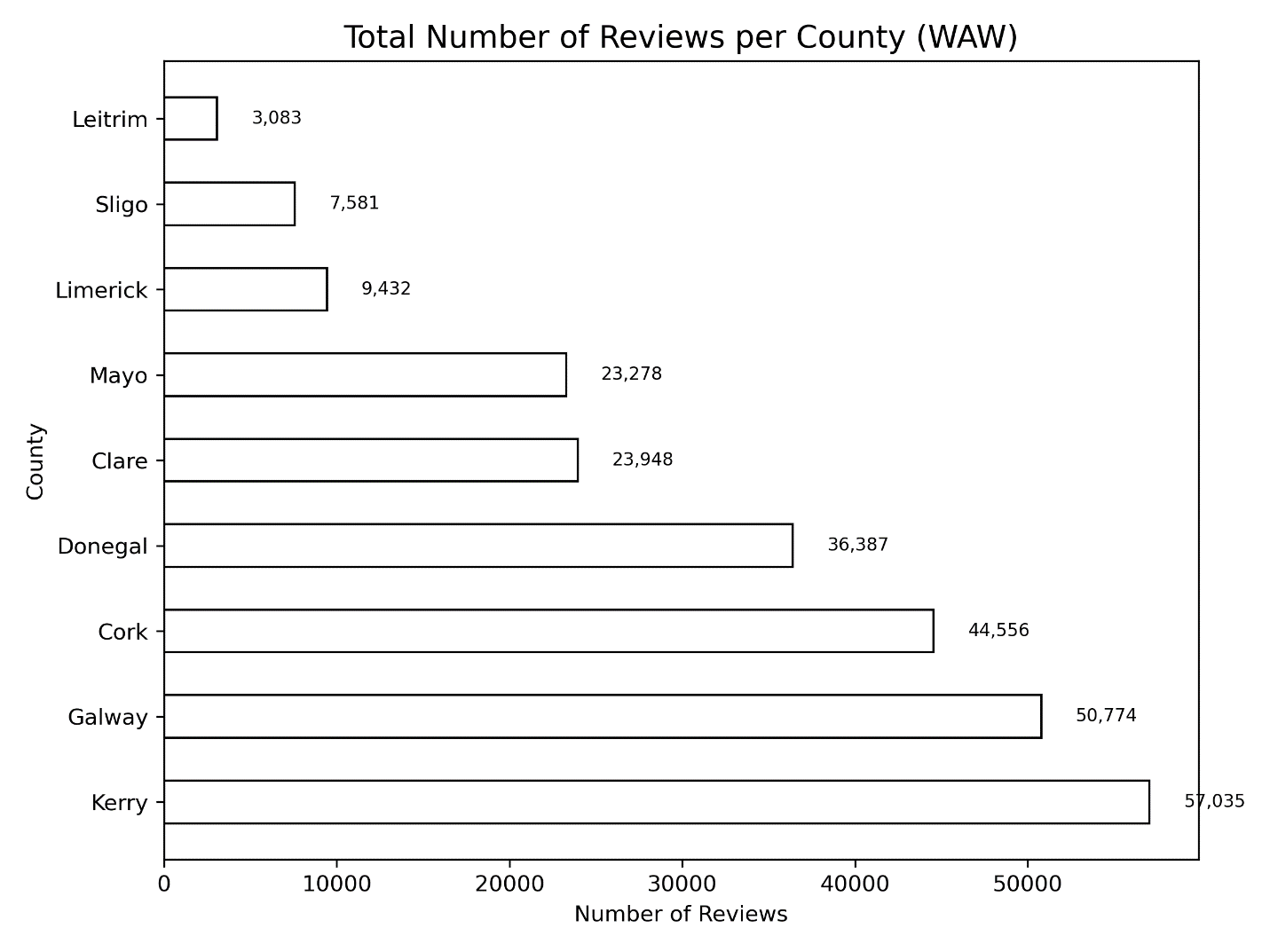
****

## 4.3 Exploratory Analysis of Reviews

The second strand of the analysis focuses on the guest reviews dataset, which contains more than 600,000 individual reviews associated with the Wild Atlantic Way (WAW) listings. Reviews represent unstructured text data and therefore provide a complementary perspective to the structured property-level information analysed in Section 4.2. Exploring this dataset is a critical step in understanding the volume, content, and tone of guest experiences before applying advanced sentiment analysis techniques.

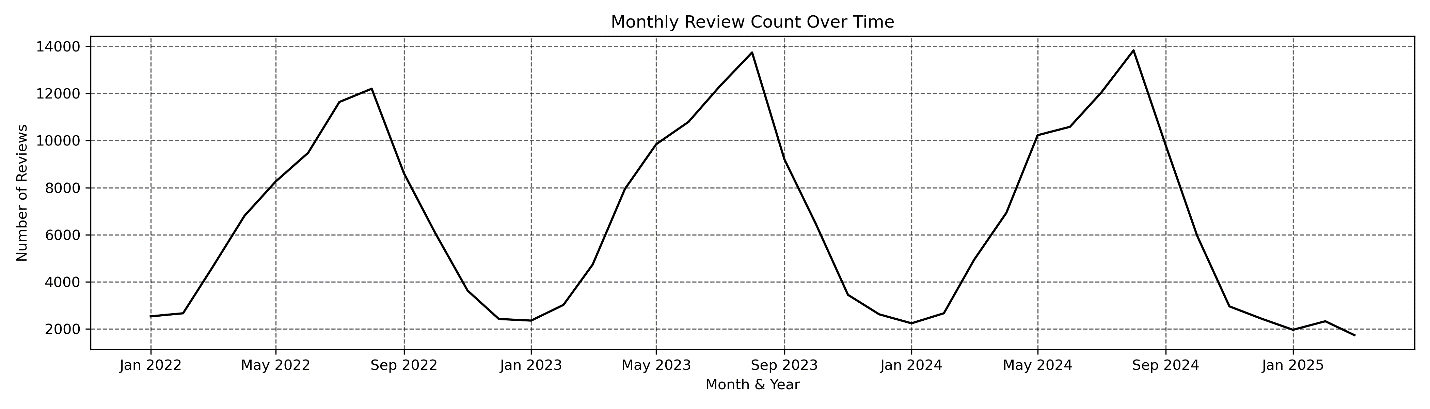
The horizontal bar graph (Figure 4.6) shows how Airbnb guest reviews are distributed in the nine counties of the Wild Atlantic Way (WAW). One can easily see the imbalance in the volumes of reviews between the southern and the northern counties. The data demonstrate that Kerry, Galway, and Cork are the dominant scores of reviews. Together, these three counties boast over fifty per cent of all the reviews.

Figure 4.6: Review Volume by County



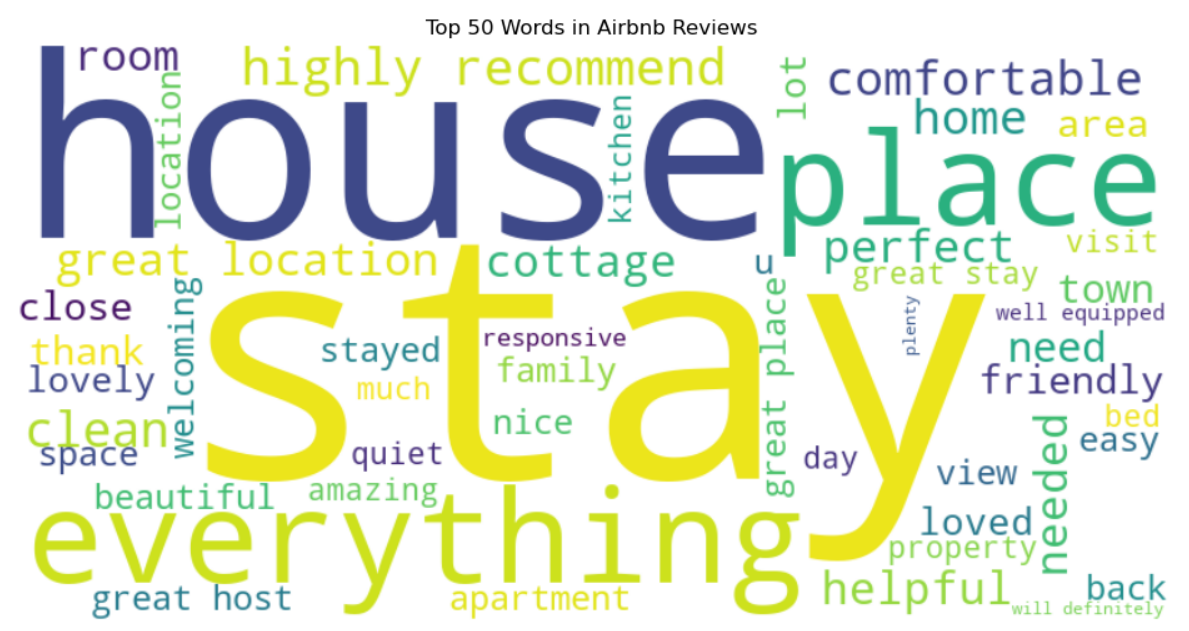
The county of Donegal also indicates a significant number of reviews (36,387). The number is moderate in Clare (23,948) and Mayo (23,278). The counties with the fewest numbers are Limerick (9,432), Sligo (7,581) and Leitrim (3,083).

Figure 4.7: Temporal Trends in Reviews



The review dataset covers the timeframe of four years, which allows the review of how guest feedback has been evolving. The monthly distribution of the reviews in all the WAW counties is presented in Figure 4.7. A distinct seasonality can be found with high values during the summer months (June-August). Also, a significant spike-down trend can be observed in 2022,

Figure 4.8: Common Terms in Reviews

****

Because the numerical analysis of the review counts is rather limited, a word frequency analysis has been conducted to provide more insight into what the guests constitute the highest (or the lowest) numbers review frequently in their feedback. Here, stop-words, punctuations and other relevant parsings have been discarded, and only significant words were used during analysis. These 50 most frequent terms are denoted in a word cloud (Figure 4.8).

Obviously, through word cloud, one can note that words like stay, house, place, everything, and clean are prevalent in the reviews. The most prominent of the terms are “friendly, recommend, perfect, cottage, and helpful, which support the importance of host behaviour and property amenities, augmenting the perception of the guest. Negative descriptors like cold or noise tend to occur less than the positive acts. Sentiment analysis outcomes and lexical patterns both showed recurring emphasis on host, location, and cleanliness.

## 4.4 Sentiment Analysis Results

The review dataset on the guest employers was sentimental to numerically measure the experiences of the travellers hidden along the Wild Atlantic Way. Three models were evaluated: lexicon-based (TextBlob) and transformer-based deep learning (BERT and RoBERTa). Since TextBlob produced less reliable results and lacked contextual sensitivity, its outputs were not included in the detailed reporting, with the focus instead placed on BERT and RoBERTa. As a multi-model strategy, this aspect brought resilience because some parts could be replaced by a more elaborate contextual embedding.

**Positive vs Negative Classification**

In addition to analysing the sentiment pattern, to interpret the rating, review scores were binarised as positive (3.0 and above) and negative (less than 3.0) depending on the outputs of BERT and RoBERTa models. The categories of reviews as positive were predominant in both models. Particularly, BERT assigned positive labels to 3,982 reviews, with just three reviews labelled as negative, whereas RoBERTa labelled all but one of the reviews (n = 3,985) as positive. The highly biased positive assessments are reflected in the almost complete lack of negative labels in Airbnb tourist reviews along the Wild Atlantic Way.

**Average Sentiment Scores by County**

Differences in guest experiences across different regions were investigated by the computation of the average sentiment scores per county in each of the models (BERT and RoBERTa). The scores of sentiment with the county are aggregated and show interesting geographic variation in the Wild Atlantic Way guest experiences. By indicating the mean values of the ratings made by four visitors on a scale of 1-5, being very high (ranging between 4.5 and 5), Table 4.2 shows that the two models recorded quite high ratings consistently throughout all counties and had a mean ranging between 4.52 and 4.86 using a five-point scale rating. The results show that there is little difference among the counties. The most positive reviews were made in Donegal (BERT = 4.79, RoBERTa = 4.83) and Leitrim (BERT = 4.77, RoBERTa = 4.86). In comparison, Galway scored a bit lower (BERT = 4.69, RoBERTa = 4.53).

Table 4.2: Average Sentiment Scores by County

| **County** | **Average** BERT**score** | **Average** RoBERTa **score** |
| --- | --- | --- |
| Clare | 4.75 | 4.70 |
| Cork | 4.75 | 4.68 |
| Donegal | 4.79 | 4.83 |
| Galway | 4.69 | 4.53 |
| Kerry | 4.77 | 4.70 |
| Leitrim | 4.77 | 4.86 |
| Limerick | 4.73 | 4.71 |
| Mayo | 4.78 | 4.75 |
| Sligo | 4.76 | 4.79 |

## 4.5 Price Prediction Results

To investigate the determinants of Airbnb prices in the Wild Atlantic Way counties, a few machine learning models were trained with the attributes of the listing advertisements and sentiment-based variables. Before running the model, prices at night were limited to a cap of 1,000 euros as outlier prices could interfere with the regression estimates. The goal was to evaluate how well structured listing features, combined with sentiment features extracted from guest reviews, could explain and predict pricing patterns.

**Model Development**

To benchmark performance, regression models were implemented: Linear Regression, Random Forest, Gradient Boosting, XGBoost, and LightGBM. The dataset consisted of structured listing attributes (e.g., county, room type, property size, number of bedrooms and bathrooms, accommodates), which were preprocessed through one-hot encoding for categorical variables and standardisation for numerical variables. The data was split into training (80%) and testing (20%) sets to enable robust evaluation. Two modelling setups were considered in this study. The first involved baseline models, which relied exclusively on structured listing features such as county, property type, and room configuration. The second involved extended models, which incorporated additional sentiment features in the form of average review sentiment scores derived from BERT and RoBERTa outputs.

**Evaluation Metrics**

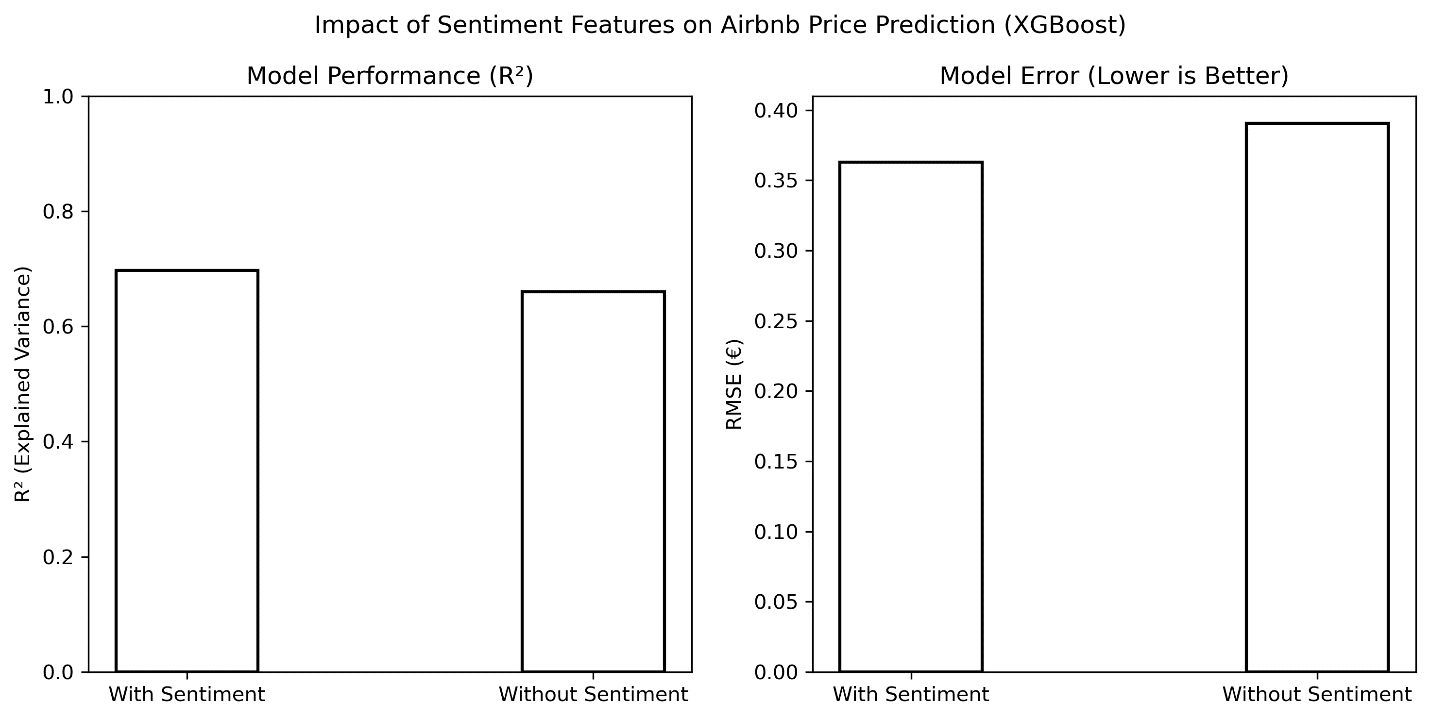
To assess the quality of Model accuracy, two metrics of standard regression were used. One is the Coefficient of Determination (R2), which is a metric to measure how much of the very elements of nightly price can be attributed to the predictors. The larger the values, the more explanatory power they have. The other is Root Mean Squared Error (RMSE), which measures the amount of average error between the predicted and the actual price, but the large deviations are penalised. The lower the results, the more accurate the predictions. The two measures are complementary in the sense that R2 provides an indication of goodness-of-fit, whilst RMSE gives the monetary value of the predictive accuracy.

**Impact of Sentiment Features**

The important experiment involved comparing the performance of the model with and without sentiment features. Figure 4.9 demonstrates the performance of the XGBoost model, which has produced the best performance out of the algorithms that were tested.

As seen, the added sentiment features helped the performance: The increases in the R2 in the baseline were successful (0.71, 0.66) with models with sentiment explained more of the price variation. RMSE was reduced (RMSE -0.35 to RMSE -0.39, no sentiment), and this indicates lower error in the prediction when the guest emotion was taken into consideration.

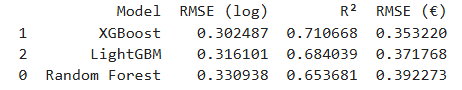
Figure 4.9: R² and RMSE comparison (with vs. without sentiment, XGBoost)



**Model Performance Summary**

The three ensemble learning models used were XGBoost, LightGBM, and Random Forest; the difference between their performance was determined to evaluate whether they were suitable in predicting Airbnb prices along the Wild Atlantic Way (Figure 4.10). The risk assessment was done on the basis of two performance KPIs: the Root Mean Squared Error (RMSE) on a log-transformed and raw euro scales, and the Coefficient of Determination (R2)

Figure 4.10: Performance comparison of ensemble models

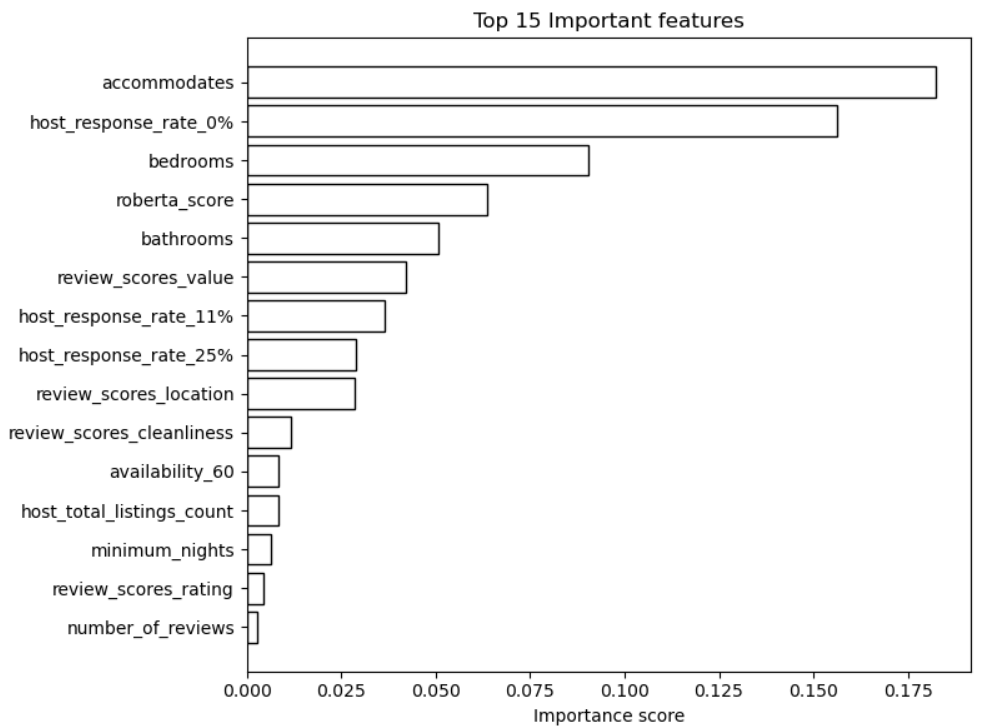


The results indicate that XGBoost delivered the best overall prediction and can achieve the lowest RMSE (0.353) and the highest R2(0.711). This shows that the model has been able to explain only about 71 per cent of the variation in Airbnb prices, which means that the predictive power of the model is very high. LightGBM worked slightly worse, with an R2 =0.684 and RMSE = 0.372, but it is still highly effective on tabular data. On the other hand, the Random Forest model provided the worst predictive performance, with an R2 = 0.654 and a more substantial RMSE = 0.392, which indicates the inability to reflect the nonlinear interactions of features as well as the boosting types of algorithms.

**Feature Importance Analysis**

To understand more about the factors that influence Airbnb pricing along the Wild Atlantic Way, the extracted values were obtained from the XGBoost model, which had portrayed the highest predictive accuracy. The importance ranking is important in that it also shows the relative importance of each variable in the elimination of prediction error, thus giving an idea of which factors are more influential in terms of the component of listing prices.

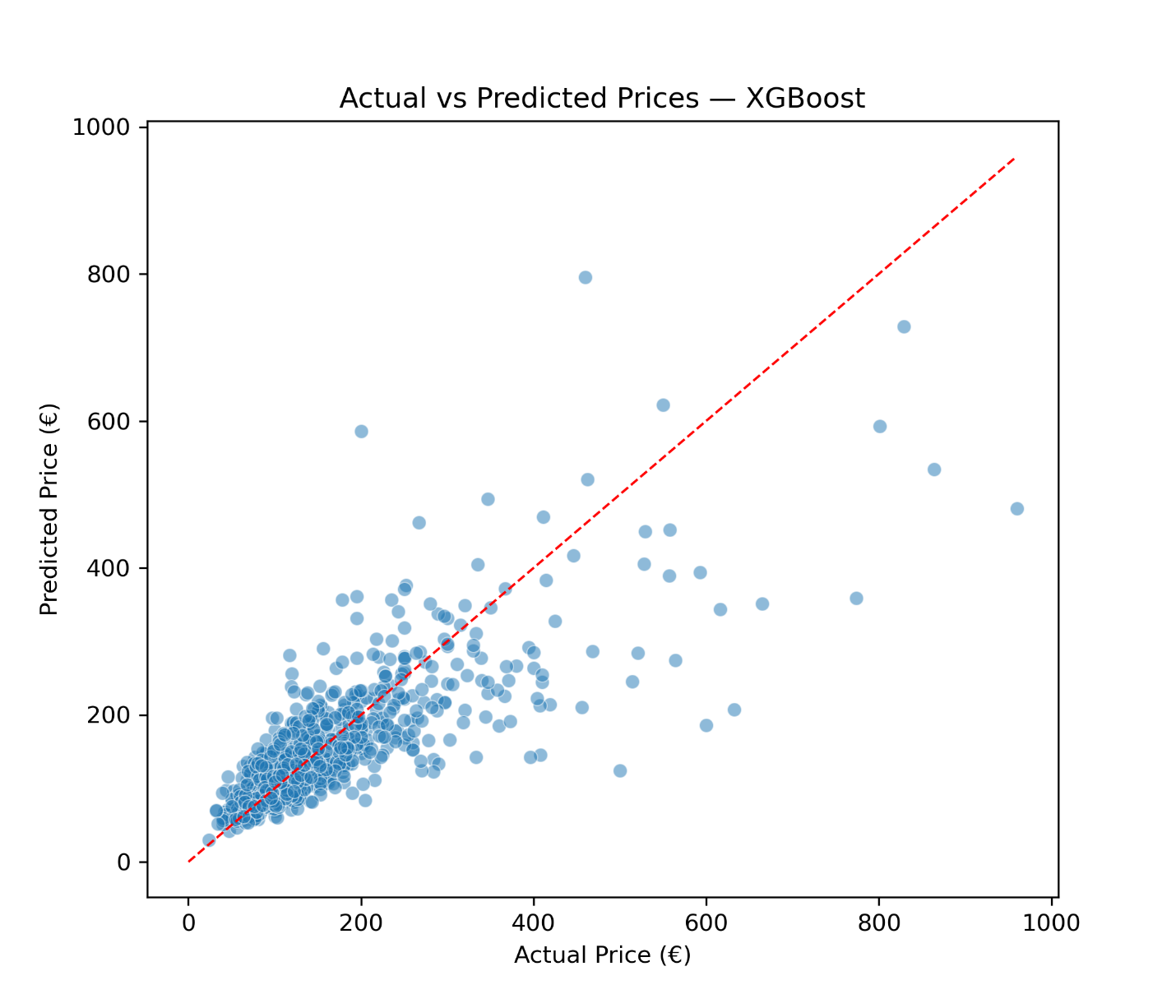
Figure 4.11: Feature importance ranking



These results reflect the vitality of the grouped permutation importance (Figure 4.11) as the most influential features in predicting Airbnb prices along the Wild Atlantic Way. The most popular driver is accommodates, which reflects the number of guests that the property can host. The rate of host response categories was also extremely influential, with the high downbeat of the value. Structural features, including the bedroom and bathroom, were also important. However, the RoBERTa-derived sentiment score is within the top five. Other, review-related features, including value, location, and cleanliness review scores, also provided non-insignificant contributions, but their relative importance was lower than that of structural and capacity-related features.

**Residual Analysis**

Figure 4.12: Error distribution analysis



In order to evaluate the performance of the models beyond R2 and RMSE, residual analysis was done. Figure 4.12 shows a scatter plot of actual and predicted prices in terms of the XGBoost model. The red diagonal line shows the ideal forecast in which there would be congruence between actual and predicted. The findings show that the model works reasonably well in most listings within the low to mid range prices (below Euro 300), with the predictions being very close to the diagonal line. However, the increase in prices causes the widening of the spread of points, which indicates the higher variance and systematically underestimates the high-value properties.

## 4.6 Case Comparisons and Integrated Analysis

Although the subsequent parts presented descriptive, sentiment and modelling results individually, it is important to discuss how these findings have common grounds. The integrated analysis shows the interconnections between the review sentiment, listing prices, and review trends along the Wild Atlantic Way.

**Sentiment score and Price.**

Average sentiment scores were uniformly positive, but county-level price patterns were not homogeneous. Kerry and Galway had both higher listing volumes and higher average nightly prices. Leitrim and Sligo had fewer listings, and smaller changes in average nightly rate and insignificant sentiment change. Sentiment scores were uniformly high across counties (BERT and RoBERTa), limiting their differentiation of county-level price variation.

**Review Trend and Price Patterns.**

Temporal analysis of reviews had seasonality, with the highest review volumes recorded in summer months. During summer peaks, higher review volumes coincided with higher average prices in Kerry and Galway. Leitrim and Sligo showed relatively low seasonal variations in the activity of the review and the price. The most salient price determinant was the structural features like the number of bedrooms and accommodation capacity, with sentiment factors and the provision of the number of reviews playing supplemental but not the foremost explanatory variables.

## 4.7 Summary of Findings

In this chapter, a detailed investigation of Airbnb activity along the Wild Atlantic Way in Ireland has been carried out, with both pre-structured listing features and unstructured guest feedback as sources of information. Such results explored how the geographic location, property characteristics and guest reviews affect the structure of the short-term rental marketplace along the WAW. On a descriptive point of view, the WAW counties have definite differences in supply and prices. The main markets that appeared to be dominant were Galway, Kerry and Cork, which comprised almost half of all the listings and reviews done. In comparison, Leitrim and Sligo experienced relatively low supply levels but achieved average prices mostly on par with or higher than bigger counties. Price distributions were strongly right-skewed across counties due to the presence of a small number of luxury listings, which justified log-transformation in modelling. The discussion of the reviews of more than 600,000 guests shows that the southern counties have high volumes of reviews, which have strong seasonal characteristics, which are the highest in the summer months. The sudden drop before 2022 was the result of the COVID-19 pandemic, causing losses in travel demand, but it was seen to improve after 2022. Reading through the feedback of the guests also showed that positive words like clean, friendly, and location were predominant and negative terms like poor were few, as there was a tendency towards positivity in online reviews.

Sentiment analysis carried a weight to these observations. Alternatively, both BERT and RoBERTa transformer models offered extremely high scores on sentiment of between 4.5 and 4.8 points on a scale of five points. Although Donegal and Leitrim ranked slightly higher in their levels of sentiment, Galway happened to be just below. The sentiment landscape was generally very positive, thereby constraining its discriminating power on a regional basis but providing value at the listing level.

The predictive modelling output showed the usefulness of incorporating review-based sentiment in price forecasting. The strongest performance was recorded among the tested algorithms with XGBoost (R 2 = 0.71, RMSE = 0.353), which outpaced the other two algorithms, Random Forest and LightGBM. Notably, the sentiment features were added to the model and increased the accuracy by a small percentage to lower the error rates and add explanatory power. The significance of the predictors was established through the property capacity (accommodates), structural attributes (bedrooms, bathrooms) and host responsiveness, which were the most significant predictors. However, the sentiment scores based on RoBERTa fell among the top five features.

# Chapter 5: Discussion

## 5.1 Introduction

This chapter analyses the evidence in Chapter 4 in the theoretical and empirical context that was developed in Chapters 1 and 2. It combines the results of descriptive statistics, sentiment analysis, and predictive models to answer two research questions: (1) Do sentiment-based features increase the predictive accuracy of Airbnb price, and (2) which determinants have the most significant effect on price movement along the Wild Atlantic Way (WAW). The talk is based on hedonic pricing theory, which describes the prices as a compound of structural characteristics (size, amenities, location), and in signalling theory, which introduces guest reviews as reputation signals in informationally asymmetric markets (Lancaster 1966; Rosen 1974; Spence 1973). These models explain the relevant contributions of the structured listing characteristics and unstructured review content to the development of the Airbnb WAW market.

## 5.2 Interpretation of Listings Data

**Geographic concentration:** Airbnb listings are highly uneven across WAW counties. The dominance falls on Galway, Kerry and Cork, which can be attributed to their deep-rooted tourism infrastructure and global recognition (Fáilte Ireland, 2023). Smaller counties such as Leitrim and Sligo account for marginal supply, consistent with their limited tourist flows.

**Accommodation preferences:** Over 70% of the listings in most of the counties surveyed are of whole homes or apartments; between 16 and 41% are of individual rooms, and the number of shared rooms is virtually nonexistent. This distribution indicates a strong area bias towards self-contained accommodation, which supports the previous research, which highlights the importance of independence and flexibility in the rural Airbnb markets (Guttentag, 2015).

**Pricing patterns:** The average nightly prices range between €144 in Limerick and €176 (Galway, Kerry, Leitrim). It is interesting to note that smaller counties like Leitrim and Sligo, though with a small supply, are showing similar prices to the larger hubs, suggesting scarcity effects and niche positioning may sustain higher rates in smaller markets. The skewness in the prices is significantly to the right due to the luxury listings, hence the reason to use a log-transformation in the development of the models.

**Correlations:** After pairwise analysis, it shows price correlates strongly with property size variables (accommodates, bedrooms, bathrooms) but weakly with review scores, which supports hedonic pricing theory, which postulates structural characteristics to dominate price formation (Rosen, 1974).

## 5.3 Insights from Reviews and Sentiment Analysis

**Volume and seasonality:** The review counts concentrate more on Galway, Kerry, and Cork. Highest peak in summer months, with smaller December peaks, while volumes fell sharply before 20202 due to COVID-19 restrictions and recovered post-2022. This trend shows how seasonality and external shocks affect the guest feedback.

**Content and positivity bias:** The word-frequency analysis highlights terms, including clean, friendly, location and comfortable, while negative words such as cold and noise are rare. This trend is symptomatic of the positivity bias that is common in hospitality-related reviews, which is explained by self-selection effects and reputation issues (Zervas et al., 2017).

**Sentiment scores:** Transformer models like BERT and RoBERTa produce high averages (4.5–4.8 on a five-point scale). Compared regional analysis shows that Donegal and Leitrim scores marginally better than Galway, and it is possible that lower-density destinations are more capable of creating more personalized experiences, and high-traffic hubs face more guest expectations. The net homogeneity of positive sentiment suppresses cross-county discrimination, but sentiment analysis is more informative at the listing level, and underlying differences are used to form competitive positioning.

**Methodological implications:** The binary classification is ineffective when more than 85% of reviews are positive, but continuous sentiment scores offer incremental value. The future methodological progress could come in the form of aspect-based sentiment analysis (ABSA) or topic modelling that separates such dimensions as cleanliness, location, and host responsiveness, which can afford more sophisticated explanatory power.

## 5.4 Price Prediction Discussion

**Model accuracy:**  XGBoost and LightGBM are gradient boosting models, which have proved to be better than Random Forest and linear regression. XGBoost achieved an R2 of 0.71 and an RMSE of about 0.35, which indicates a strong predictive ability of tabular property data.

**Effect of sentiment features:** The addition of sentiment variables yielded modest but significant gains: R² rose from 0.66 to 0.71, and RMSE decreased slightly. Such findings show that while textual reviews encode responses to perceptions of value and quality, it is not entirely represented in the responses to the structural attributes. An example is that two similar properties will be different in price due to differences in host responsiveness or cleanliness, as indicated by the sentiment scores.

**Feature importance**: The most salient predictors (i.e., accommodates, bedrooms, bathrooms) and host responsiveness came out as capacity-related variables. It is important to note that a sentiment score estimated using the RoBERTa model was in the top five of predictors, which shows that intangible guest perceptions do have a measurable impact. The results are consistent with the hypothesis that the price reflects both physical facilities and reputational cues (Li et al., 2019; Zhang et al., 2022).

**Residual patterns:** The residual analysis showed that there was an underestimation of the luxury listing and an overestimation of very cheap listings, which reflects Airbnb's long-tail distribution, where extreme cases have different dynamics. Ensemble models are excellent at capturing broad price trends, but are limited in making predictions for rare, unusual properties.

Thus, the dominant factors of price are structural features, but sentiment adds another explanatory dimension, as hypothesised by hedonic and signalling theories.

## 5.5 Integrated Case Comparisons

**Sentiment–price relationships:** There is no linear relationship between the sentiment at the county level and prices. It is interesting to note that Kerry and Galway share high prices and high sentiment, while Donegal and Leitrim share high sentiment but lower prices. These results show that the price is not the only determinant of guest satisfaction; it also relates to the scenic quality, authenticity and nature of interpersonal interactions.

**Seasonality and pricing:** Empirical studies show that the time peaks in consumer reviews in the three counties of Galway, Kerry, and Cork are pretty much proportional to the respective rise in pricing, hence the presence of demand-based price cycles in the counties. However, the smaller, less populated counties exhibit relatively flatter trends and thus support the existence of weaker demand pressures in these places.

**Interplay of features:** Pricing is supported by structural attributes, and sentiment is used to discriminate listings in competitive markets. As an example, two similar properties in Galway can have their price difference influenced by a sentiment-related feature like cleanliness or host behaviour.

## 5.6 Practical Implications

**For hosts:** Investments in property dimensions, specifically bedrooms, bathrooms, and overall capacity, yield direct price increases. However, the power of guest experience cannot be ignored; preservation of cleanliness, responsiveness, and correct descriptions of the property will lead to the development of competitive advantage, especially in markets that are already saturated, where sentiment is a distinguishing factor.

**For tourism planners:** Market concentration in the counties of Galway, Kerry, and Cork signals an imperative for targeted redistribution strategies. Smaller jurisdictions, including Leitrim and Donegal, which record higher sentiment indices, have a substantial level of reputation capital that may attract more people to visit the area, provided there is long-term investment in marketing, building infrastructure and improvement of accessibility.

**For policymakers:** TheFindings show that Airbnb pricing is highly responsive to supply-side characteristics. Regulation should balance growth opportunities with sustainable development, ensuring affordability and fair competition. Since reviews provide measurable predictive power, policy should safeguard trustworthy review systems and protect consumers against fraudulent practices.

## 5.7 Theoretical Contributions

The research makes a contribution to the existing hedonic pricing literature by showing that the prices of Airbnb along the WAW are not only dependent on some structural factors (e.g., size, capacity, location) but also on some softer reputational aspects that are based on the reviews of the guests. Although structural features are still dominant, sentiment provides further explanatory power, hence the extension of hedonic models to the use of user-generated data.

The paper also contributes to the study of online reviews using natural language processing models called BERT and RoBERTa, which are no longer limited to methods based on lexicon. The application of transformer models provides a scalable model of the guest perceptions, which supports the viability and performance of deep-learning models in the tourism analytics sphere. Methodologically, the study illustrates the advantages of gradient boosting methods for price prediction in heterogeneous tabular datasets and demonstrates the value of dual-model setups (baseline vs sentiment-enhanced) for testing the incremental contribution of unstructured data.

Finally, by focusing on the concept of the WAW as a regional tourist route, this study adds contextual nuances to the literature that exists to date, explaining how uneven spatial distribution, seasonal variations, and urban-rural polarities influence Airbnb market dynamics.

## 5.8 Limitations

Despite its useful observation on the Airbnb pricing behaviours along and within the Wild Atlantic Way, the study has a few limitations, which need to be mentioned to place the findings into perspective.

First, the dataset (Inside Airbnb) is secondary and can include either incomplete or duplicated listings. Second, the study is cross-sectional, thus restricting the ability to capture dynamic changes over time, like seasonality or one-time events like COVID-19. Third, ensemble models were faced with severe price cases due to the long-tail distribution of Airbnb. This is typical in predictive modelling in the tourism field and inhibits generalizability. Fourth, sentiment analysis was affected by the positivity bias and class imbalance, which made it less discriminative. Also, they can fail to comprehend sarcasm, humour, cultural insertion/emphasis, or idioms in guest reviews. Lastly, simplifications were forced by methodological choices (e.g., price cap, listing-level sentiment aggregation). These conditions point out the careful interpretation. Even collectively, these shortcomings do not devalue the results, but point to venues where care should be used in making an interpretation of the results.

## 5.9 Directions for Future Research

Future studies can take up the exploration of the risks and constraints of this research and use it to support research into the role of Airbnb markets and short-term rentals along the Wild Atlantic Way and in other areas of Ireland. It should use longitudinal modelling to better describe the heterogeneity in time and exogenous shocks, such as pandemics. The use of aspect-based sentiment analysis (ABSA) or topic modelling would also help provide a more specific disaggregation of review content based on such dimensions as cleanliness, comfort, and host behaviour. Along with nightly rates, additional performance measures such as occupancy, revenue, and booking numbers may provide a more complete measure of host performance.

Hybrid modelling techniques combining gradient boosting algorithms with deep-learning architectures have the potential to enhance predictive accuracy, particularly where there are outliers. The external validity of the derived findings could be tested by comparative analyses performed in several regions, like the North Coast 500 in Scotland and the Algarve in Portugal, and can show how the market structure and the expectations of visitors can be different in different destinations.

## 5.10 Summary

The discussion reveals that the dynamics in Airbnb along the WAW are influenced by both the physical characteristics of the property and the experience of the guest. Structural features like area and facilities continue to be the most significant price determinants, but sentiment adds more explanatory value, particularly at the listing level. It is emphasised that reviews show overwhelmingly positive perceptions of the guests, but it should be noted that the reviews lack differentiation between counties due to their homogeneity. Predictive modelling validates the high performance of gradient-boosting models; sentiment variables modestly improve predictive accuracy.

The results have theoretical implications in the extension of hedonic and signalling models, and have practical implications to hosts, tourism planners and policymakers. Limitations such as dataset constraints and positivity biases suggest that generalisation should be done with caution, and further studies are required on temporal modelling, further sentiment analysis, comparative contexts, and so on. All in all, the paper has shown the superiority of using both structured and unstructured data in order to learn and forecast Airbnb market behaviour.

# Chapter 6: Conclusion

## Overview

This dissertation examined the determinants of the Airbnb pricing in the Wild Atlantic Way (WAW) in Ireland, especially the combination of structured listing features and unstructured guest reviews. The study was conducted through the two research questions:

1. Does incorporating sentiment analysis of customer reviews improve the accuracy of Airbnb price prediction models for WAW counties in Ireland?
2. Which factors (such as location, availability, sentiment, and review scores) have the strongest influence on Airbnb prices?

The study used descriptive statistics, sentiment analysis, both lexicon-based and transformer-based, and predictive modelling based on a variety of machine-learning algorithms to answer the research questions. These data mergers of structured and unstructured data helped to undertake a thorough analysis of the Airbnb market in a local tourism context.

## Key Findings

There are a few notable patterns that were found in the WAW Airbnb market during the research.

**Spatial concentration:** Atotal of 31,907 Airbnb listings were located in Ireland, of which 15,814 were in WAW counties. Listings are concentrated in Galway, Kerry, and Cork, while smaller counties such as Leitrim and Sligo remain underrepresented. Entire homes dominate the supply, with private rooms playing a minor but present role.

**Pricing structure:** Price distributions are right-skewed across counties due to a small number of luxury listings, making a log transformation necessary for modelling. The average night price differed by a range of between 144 in Limerick and 176 in Galway, Kerry, and Leitrim.

**Review behaviour:** More than 600,000 guest reviews revealed strong seasonal tendencies with the highest activity through the summer seasons and a sharp drop during the COVID-19 restriction time. The analysis of words supported the idea that most reviews were positive, as such notions as cleanliness, host behaviour, and location became prominent, while Negative descriptors were very rare.

**Sentiment analysis**: Both BERT and RoBERTa showed high sentiment scores in all the counties of 4.5 to 4.8 out of 5, with Donegal and Leitrim slightly higher than Galway. Such results suggest that more individualised experiences of the guest may be possible in the rural or less populated areas, and in urban centres, there is a higher pressure on the services and increased expectations.

**Predictive modelling**: Structural attributes (accommodates, bedrooms, bathrooms) are predominant; still, the inclusion of sentiment covariates produced very slight increases in predictive accuracy. Of the analysed algorithms, gradient-boosting methods, in particular, XGBoost, demonstrated the best performance with a resulting R2 of about 0.71 and RMSE of about 0.35. Sentiment scores of RoBERTa were among the best-performing explanatory variables. Limited residual diagnostics showed that the luxury listing was underpredicted systematically.

Taken together, the findings show that Airbnb pricing in the WAW is shaped primarily by property attributes, but guest experiences provide complementary signals that refine competitiveness, particularly at the listing level.

## 6.3 Contributions of the study

**Theoretical contributions**: This work extends hedonic pricing models by incorporating structural features with the sentiment of the guest as reputational indicators. The integration highlights the applicability of the signalling theory in the market of peer-to-peer accommodation, where reviews serve as a sign of credibility in the market of information asymmetry.

**Methodological contributions**: This research proves the effectiveness of gradient-boosting algorithms, namely, XGBoost and LightGBM, to handle heterogeneous tabular datasets. Moreover, the comparison between the baseline and sentiment-enhanced models develops a systematic framework regarding the evaluation of the incremental value of unstructured data. Through the application of transformer-based sentiment analysis, the research contributes substantially to the literature on natural language processing in tourism studies, outperforming traditional lexicon- or star-rating-driven methods.

**Empirical Contributions:** This paper provides systematic reviews of Airbnb activity along the Wild Atlantic Way, thus providing a continuation of the body of literature on this subject beyond metropolitan markets to include a rural-coastal tourism corridor. Through an analysis of more than 15,000 listings and 600,000 reviews, the study records spatial imbalances in supply, explains how structural properties and guest sentiment jointly affect pricing, and finds temporal disruptions attributable to seasonality and the COVID-19 pandemic. The developed insights provide new empirical studies on the nature of rural tourism, thus supplementing previous studies on urban settings.

**Practical Contributions:** For hosts, the results affirm the importance of property capacity and simultaneously reveal the incremental revenue potential that comes with maintaining high-quality guest experiences. For regional planners, it points out potential to advance smaller counties with high reputational performance but low visibility. For policymakers, it is important to ensure that transparent, trustworthy review systems are in place to ensure that market credibility is upheld.

## 6.4 Limitations

The study is subject to several limitations. The reliance on secondary data that is provided by Inside Airbnb poses a question of data completeness and representativeness since inactive or duplicated listings can be biases of the results. The cross-sectional analysis style is a barrier to the holistic representation of time dynamics, including seasonal dynamics or a long-term demand response. The predictive models could not properly predict the extreme price situations because of the long-tailed nature of Airbnb markets, where luxury properties present a very different behaviour than the market at large. The imbalance of classes and positive bias affected sentiment analysis because over 85 per cent of reviews were positive, which reduced the discriminatory strength. Lastly, some methodological decisions were made, namely, capping the prices by the value of 1,000 and aggregating sentiment by the listing level, which simplified the modelling process but at the same time restricted the depth of interpretation.

## 6.5 Recommendations

**For hosts**: The most effective price driver is investment in property capacity, which includes the addition of bedrooms or bathrooms. At the same time, the quality of service, including cleanliness, responsiveness, and correct descriptions, is part of the reputational capital of listings, which, in turn, makes it possible to differentiate in the markets with high competition, including Galway and Cork.

**For tourism planners:** The level of market concentration in Galway, Kerry, and Cork requires strategic interventions that would lead to the equalisation of the demand along the West Atlantic Way (WAW). Other counties with attractive visitor feeling that are limited in supply like Leitrim, Sligo and Donegal can be developed as alternative destinations by intensive investment in infrastructure, improved accessibility and aimed marketing strategies.

**For policymakers:** The law should create a balance between raising economic opportunity and sustainable development of a community. Consumer protection and transparency in reviewing mechanisms cannot be looked upon to maintain trust in peer-to-peer accommodation systems. Since the guest sentiment is proven to influence the price, it is urgent to introduce the measure of protection against fake or cheated reviews.

**For future research:** Scholars should adopt longitudinal methods of analysis, to reflect seasonality and exogenous shocks. More subtle determinants of satisfaction may be explained by aspect-based sentiment analysis or topic modelling. Besides the pricing aspect, a wider evaluation of host performance could be provided by adding other outcome variables (e.g., occupancy rates, the number of bookings, or revenue income per occupied night). To assess the external validity of the present results, comparative studies with other tourism destinations (like the North Coast 500 in Scotland or the Algarve in Portugal) might be used. Also, existing constraints on detecting outliers may be addressed by the hybrid modelling frameworks that combine gradient-boosted trees with deep-learning architectures. Future studies can also combine external information sources, e.g., the occupancy of hotels, the number of flights arriving, or socio-economic variables, and can analyze host-level strategies and platform algorithms to clarify the behavioural determinants of pricing.

## 6.6 Final Conclusion

To sum up, this study has, firstly, shown that Airbnb pricing along the Wild Atlantic Way is based on structural features of properties and, at the same time, is supplemented by the experiences of the guests in the form of reviews. Combining structured and non-structured data provides a more detailed representation of the short-term rentals market, proving that both variables share the power to define price results. The insights have theoretical, methodological, and practical implications and offer research and policy directions in the future to facilitate balanced and sustainable growth of tourism in Ireland.

# References

Abrardi, L., Raguseo, E., & Rondi, L. (2025). Sellers’ behavior and online rating bias: A sentiment analysis on airbnb reviews. *Tourism Economics*, 13548166251315855. <https://doi.org/10.1177/13548166251315855>

Alharbi, Z. H. (2023). A sustainable price prediction model for airbnb listings using machine learning and sentiment analysis. *Sustainability*, *15*(17), 13159.

Amat-Lefort, N., Barravecchia, F., & Mastrogiacomo, L. (2023). Quality 4.0: Big data analytics to explore service quality attributes and their relation to user sentiment in Airbnb reviews. *International Journal of Quality & Reliability Management*, *40*(4), 990–1008.

Costa, J. R. G. A. (2025). *Predicting Airbnb prices using machine learning algorithms* [PhD Thesis, Instituto Superior de Economia e Gestão]. <https://repositorio.ulisboa.pt/handle/10400.5/100356>

Di Persio, L., & Lalmi, E. (2024). Maximizing Profitability and Occupancy: An Optimal Pricing Strategy for Airbnb Hosts Using Regression Techniques and Natural Language Processing. *Journal of Risk and Financial Management*, *17*(9), 414.

Erol, b. (n.d.). Investigating the effect of sentiment scores on rental house price prediction of airbnb listings [PhD Thesis, tilburg university]. Retrieved August 11, 2025, from <http://arno.uvt.nl/show.cgi?fid=160855>

Gao, B., Zhu, M., Liu, S., & Jiang, M. (2022). Different voices between Airbnb and hotel customers: An integrated analysis of online reviews using structural topic model. *Journal of Hospitality and Tourism Management*, *51*, 119–131.

Garcia, S. G. (2023). *Evaluating the Impact of Image Features on Airbnb Price Predictions: A Machine Learning Approach to Hedonic Pricing* [Master’s Thesis]. <https://openaccess.nhh.no/nhh-xmlui/handle/11250/3090237>

Ghosh, I., Jana, R. K., & Abedin, M. Z. (2023). An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. *International Journal of Contemporary Hospitality Management*, *35*(10), 3592–3611.

Keser, M. (n.d.). Predicting airbnb listing prices in istanbul using machine learning and sentiment analysis [PhD Thesis, tilburg university]. Retrieved August 11, 2025, from <http://arno.uvt.nl/show.cgi?fid=161110>

Kirkos, E. (2022). Airbnb listings’ performance: Determinants and predictive models. *European Journal of Tourism Research*, *30*, 3012–3012.

Krishna, M. (2024). *Price Prediction and Recommendation of Airbnb Property Listings Using Machine Learning*. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5214669>

Li, X. S. (2021). Short or long review?-Text analytics and machine learning approaches to online reputation. *International Journal of Business and Management Research*, *9*(1), 28–40.

Liu, P. (2021). *Airbnb price prediction with sentiment classification*. <https://scholarworks.sjsu.edu/etd_projects/979/>

Long, C. K., Solanki, V. K., Anh, N. V., Bach, L. H., & Son, C. N. (2024). Machine Learning-Based Prediction Models for Sentiment Analysis on Online Customer Reviews: A Case Study on Airbnb. *Annals of Computer Science and Information Systems*, *42*, 103–116.

Madhi, H. A. B., & Alhammad, M. M. (2021). What drives Airbnb customers’ satisfaction in Amsterdam? A sentiment analysis. *International Journal of Advanced Computer Science and Applications*,*12*(6).<https://www.researchgate.net/profile/MunaAlhammad/publication/353092032_What_Drives_Airbnb_Customers'_Satisfaction_in_Amsterdam_A_Sentiment_Analysis/links/67a31e14645ef274a46d5a49/What-Drives-Airbnb-Customers-Satisfaction-in-Amsterdam-A-Sentiment-Analysis.pdf>

Mahyoub, M., Al Ataby, A., Upadhyay, Y., & Mustafina, J. (2023). AIRBNB price prediction using machine learning. *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, 166–171. <https://ieeexplore.ieee.org/abstract/document/10099909/>

Marti-Ochoa, J., Ferrer-Rosell, B., & Martin-Fuentes, E. (2025). Analysis of Sentiments in Airbnb Experiences. In L. Nixon, A. Tuomi, & P. O’Connor (Eds.), *Information and Communication Technologies in Tourism 2025* (pp. 155–164). Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-83705-0_13>

Medpalliwar, A., Choube, D., Kute, G., Kaushik, K., & Meshra, P. (2025). Airbnb rental price prediction using machine learning techniques. *AIP Conference Proceedings*, *3233*(1), 020039. <https://pubs.aip.org/aip/acp/article-abstract/3233/1/020039/3331634>

Raza, M. R., Hussain, W., & Varol, A. (2022). Performance analysis of deep approaches on airbnb sentiment reviews. *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*, 1–5. <https://ieeexplore.ieee.org/abstract/document/9800816/>

Rezazadeh Kalehbasti, P., Nikolenko, L., & Rezaei, H. (2021). Airbnb Price Prediction Using Machine Learning and Sentiment Analysis. In A. Holzinger, P. Kieseberg, A. M. Tjoa, & E. Weippl (Eds.), *Machine Learning and Knowledge Extraction* (Vol. 12844, pp. 173–184). Springer International Publishing. <https://doi.org/10.1007/978-3-030-84060-0_11>

Safari, A. (2025). *Sentiment Analysis of Airbnb Reviews: Exploring Their Impact on Acceptance Rates and Pricing Across Multiple U.S. Regions* (No. arXiv:2504.14053). arXiv. <https://doi.org/10.48550/arXiv.2504.14053>

Sánchez-Franco, M. J., Troyano, J. A., & Alonso-Dos-Santos, M. (2021). Fuzzy metatopics predicting prices of Airbnb accommodations. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 1879–1891. <https://doi.org/10.3233/JIFS-189193>

Santos, A. I. G. P., Perinotto, A. R. C., Soares, J. R. R., Mondo, T. S., & Cembranel, P. (2022). Expressing the experience: An analysis of Airbnb customer sentiments. *Tourism and Hospitality*, *3*(3), 685–705.

Subroto, A., & Christianis, M. (2021). Rating prediction of peer-to-peer accommodation through attributes and topics from customer review. *Journal of Big Data*, *8*(1), 9. <https://doi.org/10.1186/s40537-020-00395-6>

Tan, H., Su, T., Wu, X., Cheng, P., & Zheng, T. (2024). A Sustainable Rental Price Prediction Model Based on Multimodal Input and Deep Learning—Evidence from Airbnb. *Sustainability*, *16*(15), 6384.

Thakur, N., Jain, R., Mahajan, A., & Islam, S. M. (2022). Deep neural network based data analysis and price prediction framework for Rio de Janeiro Airbnb. *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)*, 1–7. <https://ieeexplore.ieee.org/abstract/document/9824383/>

Tran, R. (2021). *Airbnb Price Prediction in the Age of Social Distancing*. University of California, Los Angeles. <https://search.proquest.com/openview/34db932d42c7aee0491b60ba015f8c0e/1?pq-origsite=gscholar&cbl=18750&diss=y>

Vassilikopoulou, A., Kamenidou, I., & Priporas, C.-V. (2024). Negative Airbnb reviews: An aspect-based sentiment analysis approach. *EuroMed Journal of Business*, *19*(2), 191–207.

Yang, S. (2021). Learning-based airbnb price prediction model. *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)*, 283–288. <https://ieeexplore.ieee.org/abstract/document/9406836/>

**Bibliography**

Fáilte Ireland. (2023). *Tourism performance in Ireland: Regional trends.* Dublin: Fáilte Ireland.

Fitzgerald, R., Chen, X., & Johnson, M. (2020). Predictive modeling for short-term rental markets: Pricing and guest satisfaction. *International Journal of Hospitality Management, 88*, 102524.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.

Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism, 18*(12), 1192–1217.

Hutto, C. J., & Gilbert, E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text.* Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14).

Li, X., Zhang, Z., & Wang, X. (2019). Customer reviews on Airbnb: What influences satisfaction? *Journal of Hospitality & Tourism Research, 43*(6), 879–900.

Zervas, G., Proserpio, D., & Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Journal of Marketing Research, 54*(5), 687–705.

Zhang, Y., Zhao, H., & Chen, L. (2022). Text mining of guest reviews in short-term rental markets. *Tourism Management Perspectives, 44*, 101020.

**Data source**

<https://insideairbnb.com/get-the-data/>