A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Lorrany Inacio Cardoso dos Santos |
| *Student Number* | 2024062 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA2 Capstone Report |
| *Assessment Due Date* | 10th November 2024 |
| *Date of Submission* | 10th November 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Summary

[1. Introduction 1](#_Toc182140679)

[1.1. Objective 4](#_Toc182140680)

[1.2. Problem Definition 4](#_Toc182140681)

[2. Development 5](#_Toc182140682)

[2.1. Data Understanding and Data Preparation 5](#_Toc182140683)

[2.2. Machine Learning Algorithm 8](#_Toc182140684)

[2.2.1. Linear Regression Model 8](#_Toc182140685)

[2.2.2. Random Forest & Decision Tree Models 11](#_Toc182140686)

[2.3. Data Sources and Ethical Considerations 12](#_Toc182140687)

[Conclusion 13](#_Toc182140688)

[References 16](#_Toc182140689)

**" The multifaceted influence of various attributes such as location, property type, and amenities, on the dynamic pricing trends within the Airbnb platform "**

**Lorrany I. C. dos Santos**

# Introduction

Digital technologies have revolutionized organizational operations, prompting the need for innovative solutions and diverse applications. Despite the broad scope of technology, this research project focuses specifically on Data Analysis as the proposed solution to the identified challenge. According to the UCD website, Data Analytics involves analyzing raw data to draw conclusions about business decisions. Data Analytics is essential for predicting trends, understanding customer behavior, and optimizing budget allocation.

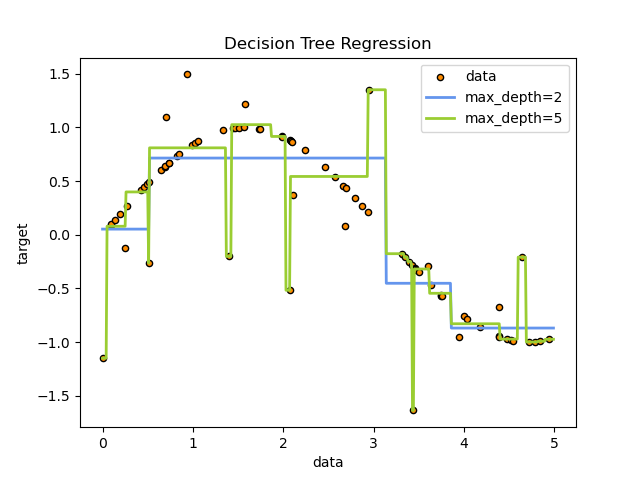
Karen and Henry explain that "data-informed" and "data-driven" are often used to discuss how data analytics helps organizations make decisions. Despite that, these two terms mean different things. Then, it is important to understand their differences and similarities (WEBBER & ZHENG,2020). It is about using decision algorithms, heuristics, and rules to make decisions. They also explain that this approach focuses on letting data guide decisions, minimizing the influence of human factors. "Data-Informed Decision Making" (DIDM) is a newer concept. It is about using data to understand situations better and provide evidence for decisions. Using data to figure out what it tells us, helps us make better decisions (WEBBER & ZHENG,2020).

Harold suggests that effective project management commences with delineating the project itself. A project, in essence, embodies a sequence of tasks and activities with a distinct goal to be achieved within predefined parameters, such as specified objectives, clear start and end dates, and, if pertinent, financial constraints. The essence of project management lies in optimizing resource utilization by fostering both horizontal and vertical flow of work within the organization. This methodology doesn't seek to dismantle the hierarchical structure but rather encourages interdepartmental collaboration to ensure seamless progress across the organization (KERZNER,2017).

According to Nick in the realm of data science, project manager holds the responsibility of overseeing the successful execution of advanced analytics and AI/Machine Learning projects. While the essence of their role aligns with traditional IT project management, it is distinguished by a focus on the intricacies of data science applications. Key tasks encompass: communicating comprehensive project roadmaps, coordinating and supervising the day-to-day activities and workflows of project teams, addressing stakeholders, defining project tasks in line with the project's vision, managing and documenting scope, and identifying and collating essential datasets required for project implementation. Additionally, some roles may necessitate technical proficiency, where the data science project manager is expected to be adept in Python, and visualization tools (HOTZ,2024).

Nick says that clients anticipate that project management will guarantee the precise fulfillment of their needs regarding the delivered data analytic solutions. This entails solutions that proficiently tackle the designated business issues or opportunities. Data analytics initiatives must harmonize with the overarching business aims and strategic objectives of the company. Clients expect project management to certify that projects are synchronized with these objectives and play a role in their attainment (HOTZ,2024).

One of the most important library that clearly addresses machine learning models is Scikit-Learn. Through the website you will find the definitions that will be seen in our project, such as a non-parametric supervised learning technique for regression and classification that is called a decision tree (DT). The objective is to learn basic decision rules derived from the data features in order to build a model that forecasts the value of a target variable. A tree can be thought of as an approximation of a piecewise constant. Using a series of if-then-else decision rules, decision trees like the Example 1 below learn from data to approximate a sine curve.



**Example 1 (Scikit-Learn, website 2024)**

Scikit-Learn proposes as well that a Random Forest is a meta estimator that employs averaging to increase predictive accuracy and manage over-fitting when fitting several decision tree classifiers on different dataset subsamples. The forest's trees employ the best split method, which is the same as giving the underlying DecisionTreeRegressor splitter="best." If bootstrap=True (default), the sub-sample size is managed by the max\_samples argument; if otherwise, each tree is constructed using the entire dataset.

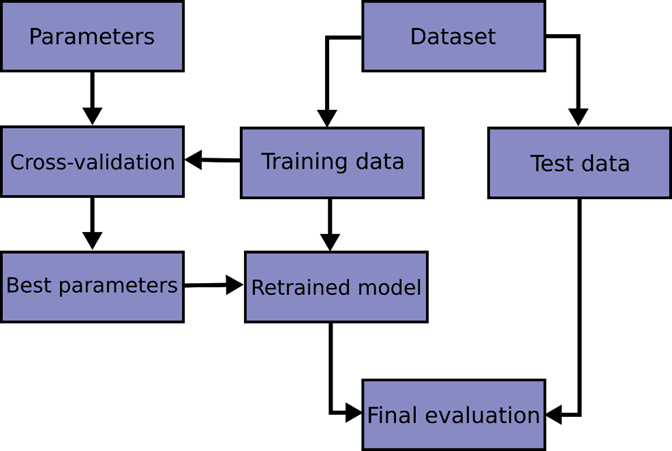
**Hyperparameter tuning**

Kizito explained hyperparameters, which are top-level parameters that control the learning process and determine the model parameters that a learning algorithm learns. They are external to the model, meaning they cannot change their values during learning or training. Machine learning engineers choose and set hyperparameter values before the model's training begins, making them external to the model. Hyperparameters are used by the learning algorithm during the learning process but are not part of the resulting model. At the end of the learning process, the trained model parameters are known as the model, and the hyperparameters used during training are not part of the model (NYUYTIYMBIY, 2020).

Hyperparameter tuning is crucial for improving the performance of a machine learning model, as it sets optimal values before the learning process and happens outside the model. Without it, errors and inaccurate results are produced due to the loss function not being minimized. Hyperparameters are like the settings of an algorithm that can be adjusted to optimize performance, similar to how we might turn knobs of an AM radio to get a clear signal. In a random forest, hyperparameters include the number of decision trees and features considered by each tree when splitting a node. Scikit-Learn provides a set of default hyperparameters for all models, but these are not guaranteed to be optimal for a problem. Tuning a model is where machine learning becomes trial-and-error engineering.

**Cross-validation: evaluating estimator performance**

Cross-validation is a methodological mistake in machine learning, where a model that repeats the labels of the samples it has just seen has a perfect score but fails to predict useful data on yet-unseen data. To avoid this, it is common practice to hold out part of the available data as a test set in a supervised machine learning experiment. The best parameters can be determined using grid search techniques, as shown in a flowchart of typical cross validation workflow in model training.



**Example 2 (Scikit-Learn, website 2024)**

# Objective

The main objective of this project is to develop and implement a plan aimed at using data tools to enhance decision-making processes. This includes improving decision-makers' abilities to make informed choices and achieve better overall outcomes. The plan encompasses various aspects of data analytics, such as understanding data, utilizing appropriate tools, and selecting optimal machine learning models. Additionally, an essential secondary objective of this project is to perform an in-depth analysis of data to uncover significant patterns, trends, and insights about competitors within the industry. This analysis will enable the organization to refine the project’s scope and objectives, ensuring that the final outcome is not only accurate but also highly relevant and aligned with the organization’s competitive landscape. By combining a focus on internal decision-making with an awareness of external trends, this project strives to deliver a balanced approach that fosters both operational efficiency and strategic advantage.

# Problem Definition

Rather of being a network of hotels, Airbnb describes itself as "a social website that connects people who have space to spare with those who are looking for a place to stay." The business operates as a middleman on a marketplace platform approach that links hosts and guests to facilitate transactions without actually owning any of the rooms (Choudary, 2013).

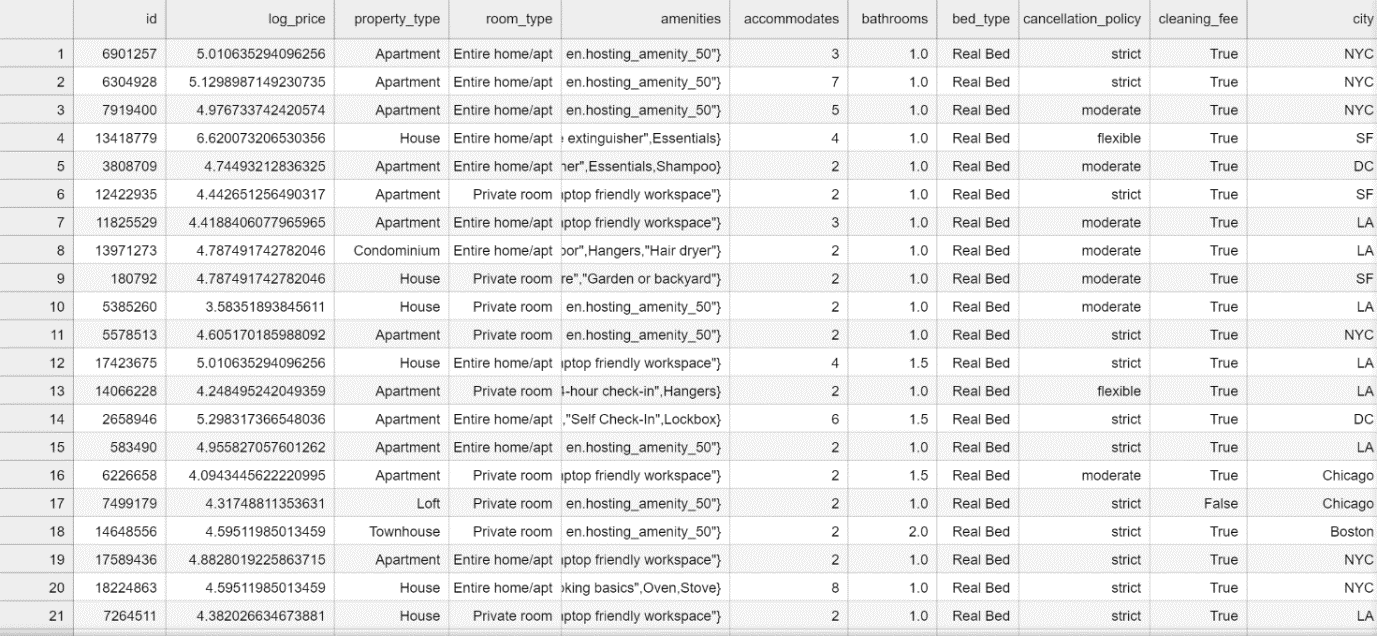
In addition to allowing users to post their flats or available rooms, Airbnb also offers insurance to the properties that are listed. In addition to offering travelers the choice of customisation for a homestay rather than a hotel, the organization offers extremely affordable solutions. The organization includes a review and rating system where guests and hosts can assess one another in an effort to gain clients' trust and the business offers a review and rating system where hosts and guests can rate one another in order to gain the trust of clients (Bashir\* and Verma )

Therefore, with the increasing demand for utilizing data and taking advantage of its opportunities, organizations are seeking clear and simple solutions and guidelines for data management. Accordingly, the project problem of this paper is *"The multifaceted influence of various attributes such as location, property type, and amenities, on the dynamic pricing trends within the Airbnb platform".*

## Development

## Data Understanding and Data Preparation

Before diving into characterization, it is essential to understand the nature and context of the data. To understand the dataset is used some codes that make it possible to analyze each piece of information in depth. In Table 1 below, we used.head(), which shows each column and what they represent.



**Table 1 (“.head()” - Jupyter Notebook, 2024)**

Understanding the business context helps ensure that the data analytics project is aligned with the organization's overarching objectives. This alignment ensures that the insights derived from the analysis are relevant and actionable for the business. This phase is understanding the goals and constraints and defining the common problems the AI model needs to solve. Doug emphasizes the importance of understanding data before diving into analysis. He explains the significance of data understanding as a foundational step in the data science process, emphasizes that before diving into analysis or modeling, it is crucial to thoroughly understand the data being used. Rose discusses various aspects of data understanding that will be talked about in this project (ROSE, 2020).

Using data analysis project management methodologies, it is possible to achieve good development throughout a data lifecycle. Initially, it is necessary to select a dataset to serve as the basis for the entire project. To understand the dataset, you need some codes to deeply understand each variable. The most important ones will only be discussed below:

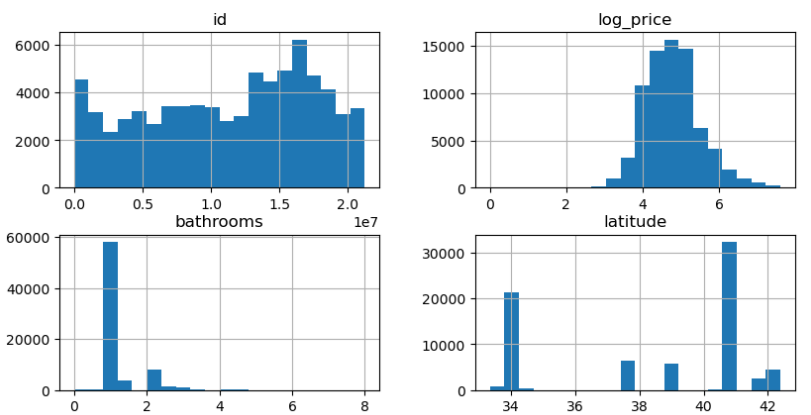
• First code: “.head()”, this function shows the first n rows and all the columns. It is useful for quickly checking of data in it, as shown in table 1 below;

• Second code: “.shape”, this function shows the total of rows and columns;

• Third: “.dtypes”, with this code you can identify if columns are numerical or categorical.

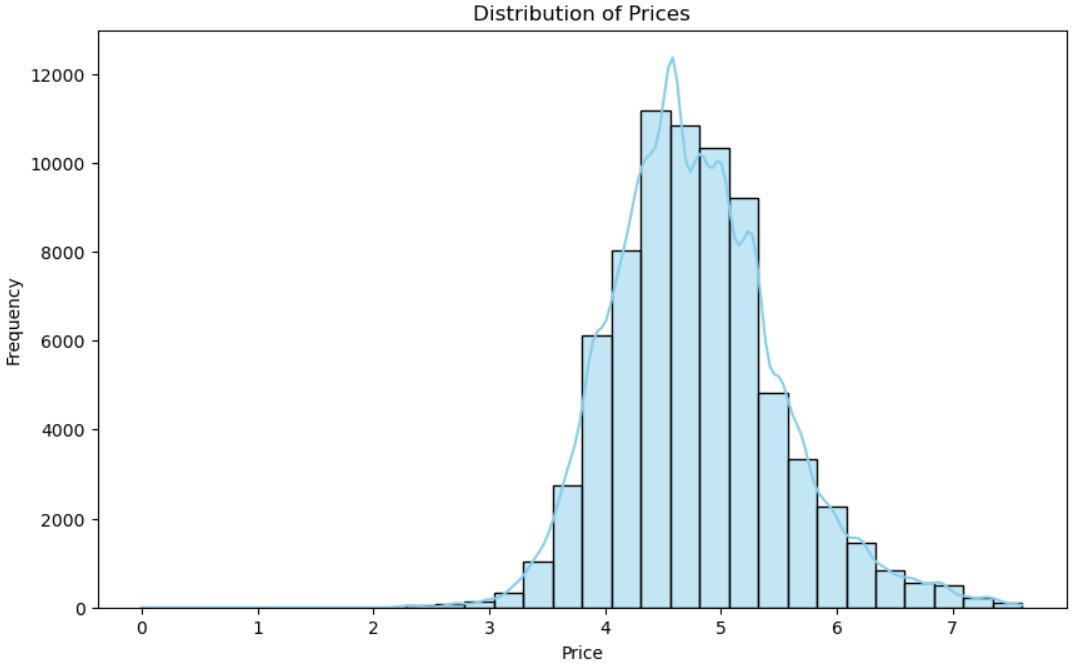
Intensify the analysis and making understanding the visual part, there are some functions that can be used to facilitate this action. Conducting Exploratory Data Analysis (EDA) to gain insights into the structure, distribution, and relationships within the data. Generally used to investigate what data can reveal beyond the formal modeling or hypothesis testing task, and it provides a better knowledge of the data set variables and their interactions.

Using the graphical library Matplotlib, it is possible to visualize the data relating to the price column through a histogram in Figure 2 and distribution of variables in Figure 1.



**Figure 1 (Jupyter Notebook, 2024)**

Distribution of Prices allow us to understand the behavior of prices and, with this, adjust the model, understand possible more common price ranges or even make market segmentations based on the distribution of values. For example, if the graph shows a peak on the left and the distribution decreases as the price increases, this indicates that most accommodations have a lower price (lower log\_price values). If there are significant bars on the right, this indicates that there are also accommodations with higher prices, but in smaller numbers.

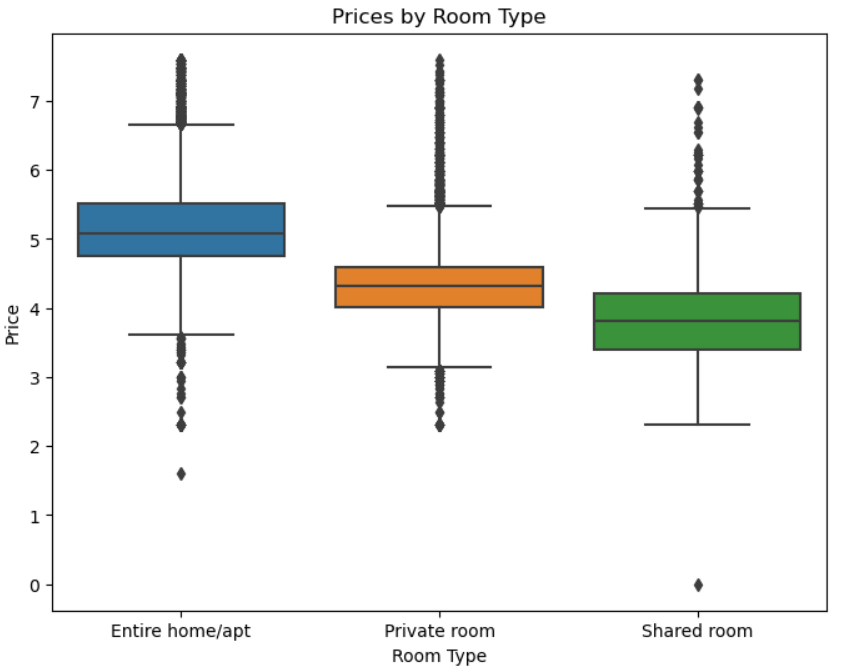


**Figure 2 (Jupyter Notebook, 2024)**

Through Box Plot shows:

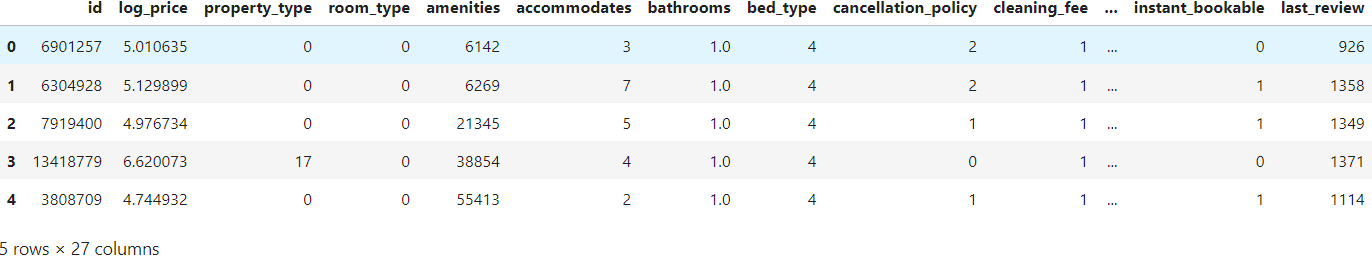
* Entire home/apt: more expensive, with greater price variability.
* Private room: moderately expensive, less variability.
* Shared room: the cheapest option, with some outliers, but generally lower prices.

This type of chart helps identify how prices behave depending on the type of accommodation, making decision-making easier for potential guests or for those analyzing the pricing of different types of accommodation on the market.



**Figure 3 (Jupyter Notebook, 2024)**

Another fundamental point in all data analysis is Data preparation. Moment when most of the project is dedicated. Knowing this, a good data project manager needs to use means that shorten this data cleaning time, and this can be done through using precise codes. Table 2 below, obtained using the same code (.head), is the result of a complete data cleaning, where columns were removed, null values were removed, words were exchanged for numbers, and several other necessary actions, which only during the entire process will you identify and adapt to the project.



**Table 2 (Jupyter Notebook, 2024)**

## Machine Learning Algorithm

First of all, it is necessary to introduce what Machine Learning is. According to Zhi-Hua, Machine learning is a method that enhances how systems work by learning from past experiences through computer-based techniques. In computer systems, these past experiences are in the form of data. The primary goal of machine learning is to create algorithms that learn from this data to build models. By giving the learning algorithm access to this data, we create a model capable of making predictions based on new observations (ZHOU, 2016).

## Linear Regression Model

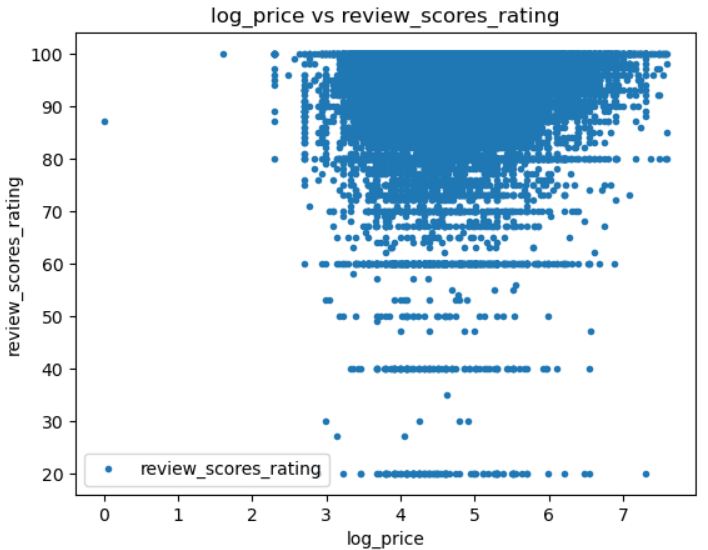
For this project, we began by selecting Linear Regression as the first model, given its simplicity and interpretability. Linear Regression is one of the foundational algorithms for predictive modeling and is particularly useful in understanding relationships between variables. According to the Scikit-Learn library, Linear Regression constructs a model based on a linear relationship, represented by coefficients w=(w1​,…,wp​) where each wi corresponds to the weight of an individual feature in predicting the target variable. The goal is to minimize the sum of squared differences between the actual target values in the dataset and those predicted by the model. This approach, known as Ordinary Least Squares (OLS), helps the algorithm fit a straight line (or a linear hyperplane in higher dimensions) through the data points to capture the underlying trend.

To implement Linear Regression in Python, it is crucial to first import the necessary libraries that support both the model itself and the processes of splitting the data, evaluating model performance, and visualizing relationships. For this project, we imported the following packages:

* from sklearn.linear\_model import LinearRegression: This imports the Linear Regression model from Scikit-Learn’s linear model module;
* from sklearn.model\_selection import train\_test\_split, cross\_val\_score: The train\_test\_split function allows us to divide the dataset into training and testing subsets, which helps in evaluating how well the model generalizes to new data. cross\_val\_score enables cross-validation, an essential technique for assessing the model’s stability and reducing overfitting;
* from sklearn import metrics: This module provides various evaluation metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R² score, which are crucial for quantifying the model's performance.

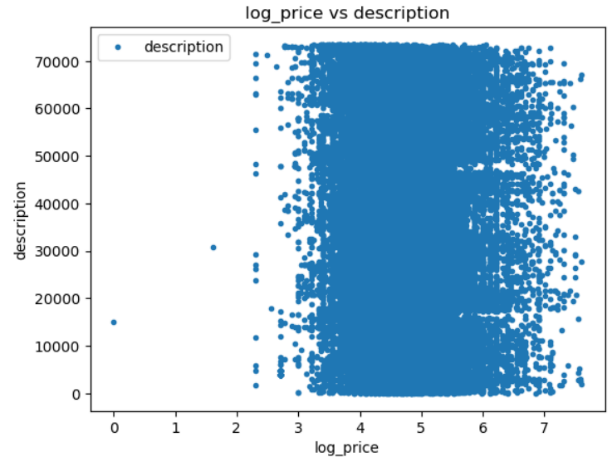
Once the libraries were in place, we used Linear Regression to model the relationship between our target variable (e.g., price) and multiple predictors. The model computes a weighted sum of the input features to generate predictions for the target. This makes Linear Regression a straightforward yet powerful tool for determining how each feature contributes to the target variable. Additionally, the model's coefficients provide insights into the ATR of each feature, which can be beneficial for feature selection and understanding the dataset.

To further explore the data, we created scatter plots to visualize relationships between key variables. For example, Figure 4 displays the relationship between log\_price and review\_scores\_rating, two important variables in the dataset. This scatter plot helps us understand whether there is a trend between the price of an Airbnb listing and its average review score. Interestingly, the scatter plot suggests that review scores do not change significantly as the price rises or falls. This indicates that while price may vary widely among listings, it does not have a strong correlation with review scores. Such a finding could imply that higher prices don’t necessarily guarantee better quality or customer satisfaction, which can be an important insight for hosts on Airbnb.



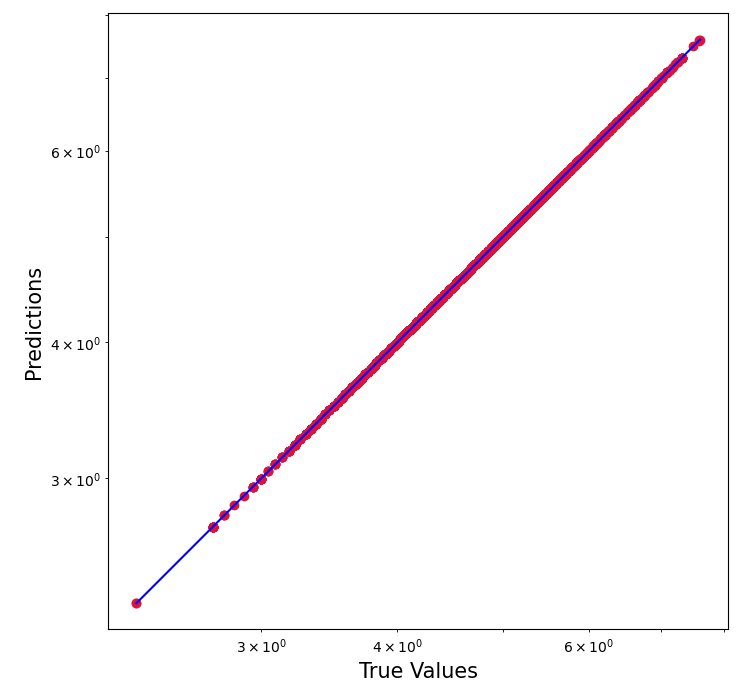
**Figure 4 (Jupyter Notebook, 2024)**

Instead, criteria including service quality, description accuracy, and cleanliness may have a greater influence on ratings than pricing. This investigation may lead to the notion that price is not a significant predictor of guest happiness, which is interesting for hosts who want to maximize their rating without raising rates.

****

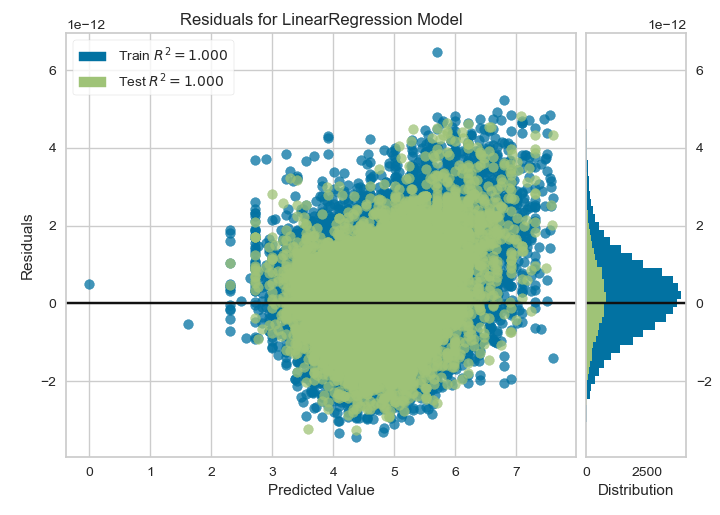
**Figure 5 (Jupyter Notebook, 2024)**

Fox and Weisberg (2018) refer to a "predictor effect display" One excellent method to understand what your regression model is saying is to examine the types of predictions it makes. The simplest approach to do so is to select a predictor in the model and compute predicted values overall values of that predictor while maintaining everything else constant as you can see below.

****

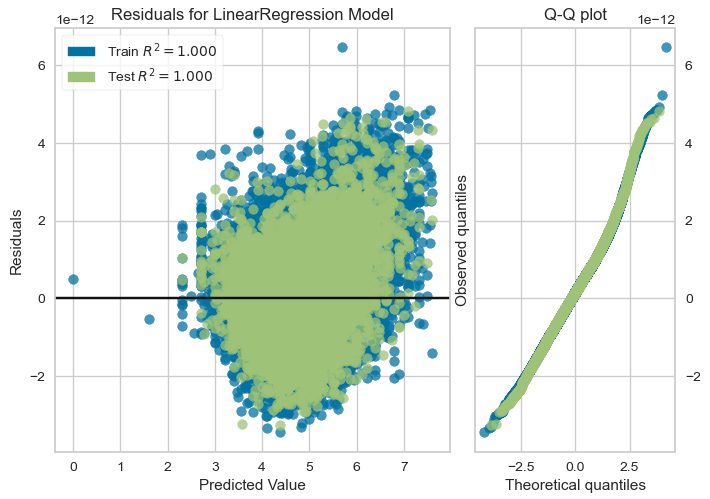
**Figure 6 (Jupyter Notebook, 2024)**

Residuals, in the context of regression models, are the difference between the observed value of the target variable (y) and the predicted value (ŷ), i.e. the error of the prediction. The residuals plot shows the difference between residuals on the vertical axis and the dependent variable on the horizontal axis, allowing you to detect regions within the target that may be susceptible to more or less error (SCIKIT-YB, 2024).

****

**Figure 7 (Jupyter Notebook, 2024)**

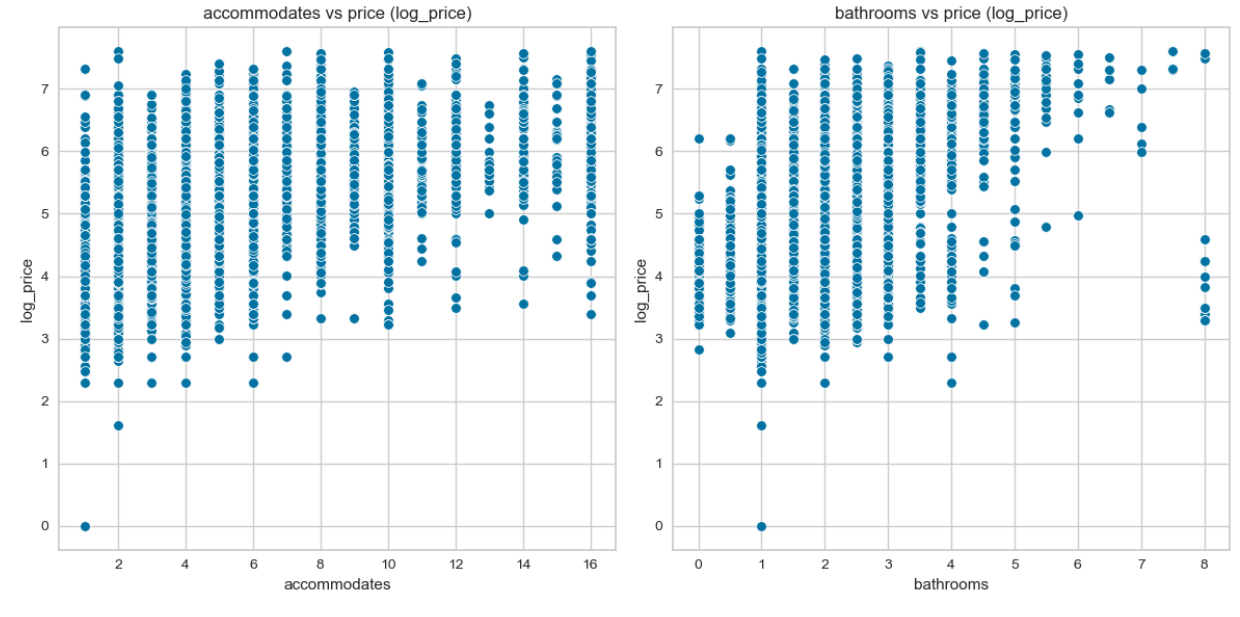
The residuals plot is commonly used to assess the variance of the regressor's error. If the points are randomly distributed along the horizontal axis, a linear regression model is typically acceptable for the data; otherwise, a non-linear model is better. In the preceding example, we see a very random, uniform distribution of residuals versus the target in two dimensions. This appears to show that our linear model is doing effectively. The histogram also shows that our error is regularly distributed at 0, indicating a well-fitted model (SCIKIT-YB, 2024).

****

**Figure 8 (Jupyter Notebook, 2024)**

## Random Forest & Decision Tree Models

Emily explains that Decision Trees, and particularly Random Forests, are powerful tools for modeling complex datasets, especially when there are intricate interactions between attributes. Random Forests are an ensemble learning method that builds multiple decision trees and combines their results to improve accuracy and reduce overfitting, making them well-suited for handling diverse and complex data. One of the main advantages of Random Forests is their ability to identify non-linear correlations and interactions between features, such as the interplay between location and amenities in predicting property prices. This allows them to capture relationships that simpler models might miss. Additionally, Random Forests offer high predictive accuracy with minimal need for extensive parameter tuning, making them relatively user-friendly and effective for real-world applications (NEWMAN, 2024).

****

**Figure 9 (Train data, Jupyter Notebook, 2024)**

## Data Sources and Ethical Considerations

In conducting this capstone project, General Data Protection Regulation (GDPR) will address several key ethical considerations. Firstly, data privacy measures must be implemented to safeguard sensitive information, including obtaining appropriate permissions for data access, anonymizing personal data, and securely storing data to prevent unauthorized access.

The dataset in the study will be collected for free on the Kaggle website. Kaggle is a place where people who work with data and machine learning can compete and connect online. On Kaggle, you can find data, create and test models, and collaborate with others to solve problems (MUO. 2023-04-17).

Through the Kaggle website, an excellent platform for data research, from which the dataset under study was taken. Paramvir the owner provides a comprehensive view of a wide range of Airbnb accommodations across the globe, spanning from lively urban apartments to peaceful rural getaways. Equipped with detailed data on property attributes, pricing, guest reviews, and host profiles, researchers and enthusiasts can delve into trends and preferences within shared accommodations, enhancing their grasp of contemporary hospitality. Whether examining market shifts, evaluating the economic ramifications of tourism, or investigating worldwide travel tendencies, this dataset is an invaluable tool for exploration and analysis (PARAMVIR,2024).

Firstly, Arihant says that data privacy measures must be implemented to safeguard sensitive information, including obtaining appropriate permissions for data access, anonymizing personal data, and securely storing data to prevent unauthorized access. Additionally, it is crucial to reflect on the potential societal impacts of the project, including addressing biases in data collection or analysis and mitigating any adverse consequences, such as discrimination or unfair treatment, because if there is no regulation of those important aspects, people can have very bad implications with online entertainment. Following that Arihant proposes some keys to how to do it:

* Embracing an ethical framework – Establishing a robust ethical framework guiding the entire data science process is imperative. This framework should encompass clearly defined principles, values, and best practices emphasizing fairness, transparency, and accountability across the data lifecycle.
* Detecting and mitigating bias – Proactively identifying and mitigating biases within algorithms is essential. Techniques like fairness-aware machine learning and routine model audits to ensure equitable outcomes among diverse demographic groups can effectively handle challenges associated with algorithmic biases.
* Implementing ethical training initiatives – Equipping team members with a thorough understanding of the established ethical framework cultivates a culture of ethical decision-making throughout the data science process.
* Regularly updating ethical guidelines – Given the evolving nature of technology, it is crucial to periodically update ethical guidelines;
* Advocating for external audits – External audits, verifying adherence to ethical guidelines and fostering transparency in data science endeavors (PATNI, 2023).

## Conclusion

The report provides a summary of how the project came about, highlighting important goals that were achieved during the modeling stage. There are some difficulties, especially during the data cleaning phase, where most of the time is spent on project management. The modeling results will be interpreted, considering the important things learned by analyzing the data and showing any major trends or patterns found.

To understand the results obtained, you need to understand the following terms:

The R² metric, also known as the coefficient of determination assesses how well the predictions align with the actual values, indicating the degree of "goodness of fit." It ranges from zero to one, where zero signifies no fit and one denotes a perfect fit.

The Mean Squared Error quantifies the average squared difference between predicted and observed values (residuals). Similar to mean absolute error (MAE), it offers an overall estimation of the error's magnitude. Taking the square root of MSE restores the units to the original scale of the output variable, enhancing interpretability.

Mean Absolute Error computes the average of the absolute differences between predictions and actual values. It's a linear score, treating all differences equally in the average calculation. MAE provides insight into the extent of prediction errors without considering their direction (e.g., over- or underpredictions).

Running with 100 numbers of estimators I found the following results:

|  |
| --- |
| Linear Regression  Mean Absolute Error: 7.855667999719299e-13  Mean Squared Error: 1.038369681942538e-24  Root Mean Squared Error: 1.0190042600217813e-12 |
| Decision Tree - RMSE: 0.5207786466455262  Decision Tree - R²: 0.3990228931116542 |
| Random Forest - RMSE: 0.36329150337712496  Random Forest - R²: 0.7075429648968661 |

**Table 3**

A result of "CV Mean: 1.0" and "STD: 0.0" in a linear regression model indicates that the model is providing perfect accuracy on the cross-validation data. The chosen machine learning model, Linear Regression, proves how it can provide results with good accuracy. Because "CV Mean: 1.0" means that the average of performance metrics (like R², for example) calculated across multiple cross-validation folds is 1.0. A score of 1.0 indicates that the model can predict the target values ​​in the validation data perfectly. Zero standard deviation (STD) suggests that there is no variation in the performance score between the different cross-validation folds. This implies that the model is consistently achieving the maximum score across all cross-validation folds.

In the first part of the project, a single model was created, Linear Regression. However, now with the final project, I decided to include two more models so that a comparison of results could be made.

Results and Interpretation:

* Linear Regression: Although the error is extremely low, it is likely that the model was overfitted or tested on the same training data, resulting in misleading accuracy. This model may not generalize well to new data;
* Decision Tree: The decision tree model does not adequately capture the complexity of the data, reflected by the relatively high RMSE and low R². Individual decision trees tend to be limited in their ability to generalize;
* Random Forest: The best model among the three, with a lower RMSE and higher R², indicating that it is able to explain most of the variability in the data and performs better than the other models.

These steps provide an initial analysis of the dataset with decision tree and random forest models. This helps to understand the influence of each attribute on the price.

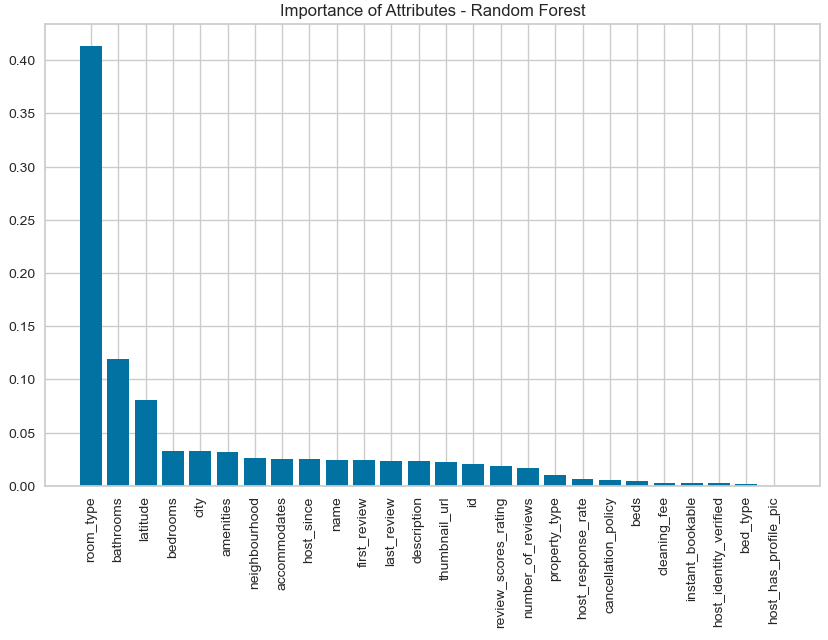
For this dataset, Random Forest is the best choice among the three models tested, as it combines low prediction error and substantial explanation of the data variability. This suggests that it can be trusted to predict prices (or other target variable) with good accuracy, while other models (especially Linear Regression) may not generalize well to new data.

During the analysis development, I had some challenges running some models. And so I had to reduce the number of estimates, for example, I reduced to 10 estimates, which gave me slightly different results.

|  |
| --- |
| Decision Tree - RMSE: 0.5207786466455262  Decision Tree - R²: 0.3990228931116542 |
| Random Forest - RMSE: 0.38142967254946786  Random Forest - R²: 0.6776107577164981 |

**Table 4**

Figure 10 demonstrates the use of a forest of trees to assess the importance of features in an artificial classification task. The blue bars represent the feature importances of the forest, with error bars indicating their variability across different trees.

****

**Figure 10 (Jupyter Notebook, 2024)**

In conclusion, this project investigated the elements that influence dynamic pricing on Airbnb, with a particular emphasis on location, home type, and amenities. The study used machine learning models—Linear Regression, Decision Tree, and Random Forest—and discovered that while Linear Regression demonstrated near-perfect accuracy, this was most likely due to overfitting, revealing limitations in its predictive power for fresh data. The Decision Tree model gave insights but struggled to capture data complexity, as shown by higher RMSE and lower R² values. The Random Forest model was found to be the best effective predictor of Airbnb prices, with low prediction error and high R².

This analysis illustrates the importance of model selection and cross-validation in developing robust predictive models for dynamic pricing. Despite challenges in data management and model tuning, this project has provided valuable insights into pricing trends, setting the groundwork for further refinement and application of these models within the Airbnb platform.

# References

Arihant Patni, “Ethical Considerations And Challenges In Data Science”. Linkedin. <https://www.linkedin.com/pulse/ethical-considerations-challenges-data-science-arihant-patni-sbnlc/>

[*"A Beginner's Guide to Kaggle for Data Science"*](https://www.makeuseof.com/beginners-guide-to-kaggle/). MUO. 2023-04-17*. Retrieved 2023-06-10*.

Kerzner, H. (2017). PROJECT MANAGEMENT A Systems Approach to Planning, Scheduling, and Controlling. New Jersey, Tenth Edition, page 29.

Karen L. Webber, Henry Y. Zheng (2020). Big Data on Campus: Data Analytics and Decision Making in Higher Education, pages 6-7.

Nick Hotz. “The Project Manager in Data Science”. Data Science Process Alliance. <https://www.datascience-pm.com/data-science-project-manager/>

Provost, F. (2013) and Fawcett, T.(2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. United States of America, O’Reilly Media, page 41.

PARAMVIR, “AirBNB\_Data”. Kaggle Website. <https://www.kaggle.com/datasets/paramvir705/airbnb-data>

Unidentified Author. “What is Data Analytics?”. UCD Professional Academy. <https://www.ucd.ie/professionalacademy/resources/what-is-data-analytics/>

Unidentified Author, “Model Evaluation”. Scikit Learn. <https://scikit-learn.org/stable/modules/model_evaluation.html#mean-squared-error>

Zhou, Zhi-Hua. (2016). Machine Learning, pages 2-3

Github, <https://github.com/lolocaardoso024/CA-2-Capstone-Report/blob/ebf237d72bacca69a1cf07e78ab4f206b3146069/CA%202%20Capstone-Report.docx>

Verma Rajesh, “AIRBNB DISRUPTIVE BUSINESS MODEL INNOVATION: ASSESSING THE IMPACT ON HOTEL INDUSTRY”. Research Gate. <https://www.researchgate.net/profile/Makhmoor-Bashir/publication/306193997_Airbnb_disruptive_business_model_innovation_Assessing_the_impact_on_hotel_industry/links/5b06460ca6fdcc8c2522ac1e/Airbnb-disruptive-business-model-innovation-Assessing-the-impact-on-hotel-industry.pdf>

Newman Emily, 2024, “8 Key Advantages of Decision Tree”, Automate Support Orp Yony X, <https://corp.yonyx.com/automate-support/advantage-of-decision-tree/>

Unidentified Author, “Exploratory Data Analysis”. IBM Website. <https://www.ibm.com/topics/exploratory-data-analysis>

Nyuytiymbiy Kizito, “Parameters and Hyperparameters in Machine Learning and Deep Learning”, Towards Data Science Website. <https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac#:~:text=Hyperparameters%20are%20parameters%20whose%20values,parameters%20that%20result%20from%20it>.

Long Jacob, 2024, “Visualizing regression model predictions”, Cran R Project Website, <https://cran.r-project.org/web/packages/jtools/vignettes/effect_plot.html>

Unidentified Author, “Residuals Plot”. Scikit Learn. <https://www.scikit-yb.org/en/latest/api/regressor/residuals.html>