### **For Airbnb Price Prediction Dataset:**

**Research Question:**"How accurately can we predict the price of an Airbnb listing based on its features such as location, property type, number of bedrooms, and amenities?"

### **For Tesla Stock Price Prediction Dataset:**

**Research Question:**"Can we develop a reliable model to predict the closing price of Tesla stock using historical price data and other relevant financial indicators?"

### **Prediction Objectives:**

**For Airbnb Price Prediction Dataset:**

* **Objective:** Predict the price of an Airbnb listing based on various features.
* **Features to consider:** Location, property type, number of bedrooms, amenities, etc.
* **Target Variable:** price of the Airbnb listing.

**For Tesla Stock Price Prediction Dataset:**

* **Objective:** Predict the closing price of Tesla stock using historical price data and relevant financial indicators.
* **Features to consider:** Historical price data, trading volume, market trends, financial indicators, etc.
* **Target Variable:** closing\_price of Tesla stock.

### **Dataset Overview:**

The dataset contains multiple features related to Airbnb listings, including categorical variables like property\_type, room\_type, and neighbourhood, as well as numerical variables such as number\_of\_reviews, availability\_365, and minimum\_nights.

### **Steps Followed:**

1. **Reading the Dataset**:  
   The dataset was loaded into a Pandas DataFrame for analysis and manipulation. Initial exploration revealed a diverse set of features, including categorical and numerical columns, as well as some missing values.
2. **Problem Statement Definition**:  
   The problem was defined as predicting the price of an Airbnb listing using various features related to the listing itself and its surroundings. The target variable for this prediction task was set as the log\_price of the Airbnb listings.
3. **Target Variable Identification**:  
   The target variable identified for prediction was log\_price (the natural logarithm of the price).
4. **Target Variable Visualization**:  
   The distribution of the target variable was visualized using a histogram with a Kernel Density Estimate (KDE), showing a relatively normal distribution after applying the logarithmic transformation to the price.
5. **Basic Data Exploration**:  
   The dataset was explored for basic statistics, missing values, and patterns. Several categorical variables were present, requiring conversion to numerical representations.
6. **Identifying and Rejecting Useless Columns**:  
   Features that were irrelevant or had too many missing values were identified and removed from the dataset to ensure better model performance.
7. **Visual Exploratory Data Analysis**:  
   Various visualizations, including histograms and bar charts, were used to explore the distribution of important features like property\_type, neighbourhood, and room\_type. This helped to understand how these features are distributed across the dataset.
8. **Feature Selection**:  
   Based on the distribution and correlation of the data, the most relevant features were selected for the model. These included categorical features like property\_type and room\_type, as well as numerical features like minimum\_nights and number\_of\_reviews.
9. **Outliers and Missing Values**:  
   Outliers in numerical features were handled using standard techniques, and missing values were either imputed or dropped based on their relevance to the prediction task.
10. **Correlation Analysis**:  
    Both visual (heatmaps) and statistical correlation analyses were conducted to identify the most significant features that correlate with the target variable (log\_price).
11. **Data Conversion**:  
    Categorical features were converted to numerical format using encoding techniques such as one-hot encoding and label encoding to prepare the data for machine learning algorithms.
12. **Training/Testing Sampling and Cross Validation**:  
    The dataset was split into training and testing sets, and k-fold cross-validation was applied to validate the performance of different models on unseen data.
13. **Investigating Multiple Regression Algorithms**:  
    Several regression algorithms were tested, including Linear Regression, Decision Trees, Random Forest, and Gradient Boosting. The models were evaluated based on their R² scores, and the best-performing model was selected.

### **Model Results:**

The Linear Regression model achieved an **R² score of 0.9867**, indicating a high level of accuracy in predicting the price of Airbnb listings. This model was saved using checkpoints to allow for future use.

### **Conclusion:**

The project successfully built a predictive model for Airbnb prices with a high degree of accuracy. The analysis revealed that key features such as room\_type, property\_type, and the number of bedrooms played a significant role in determining the price of listings. Further improvement could be made by incorporating more advanced feature engineering or testing additional machine learning algorithms.

This summary covers all essential aspects of the Airbnb price prediction project and highlights the key steps and outcomes.

### **Summary of Tesla Stock Price Prediction Model**

**Objective**:  
The goal of this project was to develop a reliable model to predict the closing price of Tesla stock using historical price data and relevant financial indicators. The project focused on preparing and analyzing the dataset, followed by training machine learning models for predictive analysis.

### **Dataset Overview:**

The dataset contained historical stock price data for Tesla, including features such as Date, Open, High, Low, Close, Adj Close, and Volume. The target variable for this analysis was the Close price, representing the price of Tesla stock at market close on each day.

### **Steps Followed:**

1. **Reading the Dataset**:  
   The Tesla stock dataset was read into a Pandas DataFrame for analysis. The data covered several years of daily stock prices and included various price indicators.
2. **Problem Statement Definition**:  
   The problem was defined as predicting Tesla’s closing stock price using the historical stock data provided. This is a typical time-series forecasting problem in the field of finance.
3. **Target Variable Identification**:  
   The target variable identified for prediction was the Close price of Tesla stock.
4. **Target Variable Visualization**:  
   The distribution of the Close price was visualized using line plots and histograms. Time-series trends and seasonality were observed over the dataset’s period, with notable price fluctuations during certain years.
5. **Basic Data Exploration**:  
   The dataset was explored for summary statistics, missing values, and patterns in stock price behavior. Relationships between the features and the target variable were identified, with some features showing stronger correlations with the closing price.
6. **Identifying and Rejecting Useless Columns**:  
   Columns that did not contribute meaningful information to the prediction (such as the Adj Close column, which was nearly identical to the Close column) were identified and rejected from the analysis.
7. **Visual Exploratory Data Analysis**:  
   Line charts and bar charts were used to visualize key features such as Volume, Open, High, Low, and their respective relationships to the closing price. This helped understand the stock’s volatility and trends over time.
8. **Feature Selection**:  
   Based on data exploration and correlation analysis, relevant features such as Open, High, Low, and Volume were selected for the machine learning model to predict the closing price.
9. **Outliers and Missing Values**:  
   Outliers in stock prices were identified using statistical techniques, and missing values were handled through appropriate imputation methods to ensure a clean dataset for model training.
10. **Correlation Analysis**:  
    Correlation analysis was performed to determine the relationship between features and the closing price. Strong correlations were found with features such as Open, High, and Low, making them critical for the prediction model.
11. **Data Conversion**:  
    The Date feature was converted to datetime format and used for trend analysis but was not included as a direct input feature in the model. Numerical features were scaled to ensure that the model could converge efficiently.
12. **Training/Testing Sampling and Cross Validation**:  
    The data was split into training and testing sets, and cross-validation was applied to validate the performance of different machine learning models on unseen data.
13. **Investigating Multiple Regression Algorithms**:  
    Several regression algorithms were investigated, including Linear Regression, Decision Trees, Random Forest, and Gradient Boosting. These models were trained and tested, and their performance was evaluated using R² scores.

### **Model Results:**

The **Linear Regression model** was selected as the best-performing model, achieving an **R² score of 0.9867**, indicating excellent predictive accuracy for Tesla’s closing price. This model was saved using checkpointing techniques, allowing for future predictions without retraining.

### **Conclusion:**

The project successfully developed a robust predictive model for Tesla's stock price. The analysis showed that key financial indicators, such as Open, High, Low, and Volume, were significant predictors of the closing price. The Linear Regression model achieved high accuracy, and the entire process from data exploration to model deployment was successfully completed.

Further improvements could include the incorporation of advanced time-series modeling techniques or the inclusion of external economic indicators to refine predictions during periods of extreme market volatility.