Project- Part A: Airbnb Price Prediction and Insights

1. Overview

Airbnb provides a platform for property owners to rent out their spaces to travelers. Pricing a listing effectively is critical for maximizing revenue while staying competitive in the market. For hosts, understanding what factors influence the price of their listings is essential.

This project aims to build a machine learning model to predict the price of Airbnb listings based on various features such as property type, room type, location, amenities, and host characteristics. By analyzing these factors, this project will provide actionable insights to Airbnb hosts to optimize their listing prices.

2. Problem Statement

The primary objective of this project is to develop a regression model that predicts the price of an Airbnb listing. Using features such as property type, room type, number of reviews, location, and amenities, the model will aim to estimate the price accurately.

The insights derived from this analysis will help Airbnb hosts understand the key drivers of price, enabling them to make data-driven decisions for pricing their properties. Additionally, the project will help Airbnb refine its recommendations for pricing to improve host and guest satisfaction.

3. Dataset Information

Dataset: **Airbnb_data**

The dataset information can be found in this doc **Data Information**.

4. Deliverables

1. **Data Exploration and Preprocessing (10 Marks)**:
    - Analyze the dataset for trends, missing values, and outliers.
    - Perform data cleaning, feature engineering, and transformations.
2. **Model Development (20 Marks)**:
    - Build a regression model to predict listing prices.
3. **Model Evaluation (10 Marks)**:
    - Evaluate the model's performance using appropriate metrics like RMSE, MAE, and R^2.
4. **Final Report and Presentation (10 Marks)**:
    - Final video clip of candidate summarizing the entire process of the project as the presentation. (**Note**: Video should not exceed **5** mins)

5. Success Criteria

The project will be deemed successful if:

- The regression model achieves acceptable performance metrics (e.g., RMSE and R^2) on test data.
- Insights about the factors influencing Airbnb prices are effectively communicated to stakeholders.
- Predictions for new listings can be made with reasonable accuracy.

6. Guidelines

- **Data Splitting**: Divide the data into training, validation, and testing sets to ensure robust evaluation.
- **Feature Engineering**: Extract meaningful features, such as neighborhood popularity, number of amenities, and host activity metrics.
- **Model Tuning**: Experiment with different models and tune hyperparameters to optimize performance.
- **Visualizations**: Use charts and graphs to present data trends and model results clearly.
- **Interpretability**: Ensure the model is interpretable and actionable for non-technical stakeholders.

7. Tools Required

- **Python Libraries**: pandas, numpy, scikit-learn, matplotlib, seaborn, XGBoost, etc.
- **Jupyter Notebook**: For developing and documenting code.

Part B: Customer Churn Prediction

---

1. Overview

Customer churn, or customer attrition, refers to when a customer ceases their relationship with a company or service provider. In today's highly competitive business environment, retaining customers is a critical factor for long-term success. Predicting customer churn can help organizations take proactive steps to retain customers, thus minimizing revenue loss. This project aims to build a machine learning model that can predict whether a customer will churn based on their demographic, account, and service-related data.

2. Problem Statement

The goal of this project is to develop a classification model that predicts whether a customer will churn. Using demographic data (such as gender, senior citizen status, and tenure), along with information about the services they use (such as internet service, phone service, and online security), we will attempt to build a model that helps the company identify customers who are at a high risk of churning.

By predicting customer churn, the company can proactively design retention strategies to keep these customers, thereby improving customer satisfaction and reducing financial loss.

3. Dataset Information

Dataset: **Customer_data**

The dataset information can be found in this doc **Data Information.**

4. Deliverables

- A data exploration and preprocessing notebook or report that analyzes the dataset, handles missing values, and prepares the data for modeling. **(10 Marks)**
- A machine learning model capable of predicting customer churn. **(20 Marks)**
- An evaluation of model performance using appropriate metrics (such as accuracy, precision, recall, F1 score, etc.). **(10 Marks)**
- Final video clip of candidate summarizing the entire process of the project as the presentation. (**Note**: Video should not exceed **5** mins) (**10 Marks)**

5. Success Criteria

The success of the project will be determined by the following:

- Proper interpretation of the model's output, providing actionable insights to reduce customer churn.
- Get the predictions for the new data.

6. Guidelines

- Make sure to split your data into training and testing sets to avoid overfitting.
- Tune the hyperparameters of your models to improve performance.
- Report all the steps taken in the data preprocessing, modeling, and evaluation phases.
- Provide a final model that balances accuracy with interpretability.

---

7. Tools Required

- Python (with libraries such as pandas, scikit-learn, matplotlib, seaborn, etc.)
- Jupyter Notebook or any IDE suitable for running Python code.

Submit Guidelines

- Submit your ipynb/ folder.
- Create 2 videos of maximum of 5 mins explaining the analysis for each Part and share the drivelink along with the folder.

How to ZIP a PDF file:

- Put all of the documents/ipynb you want to compress (or just one) into a new folder.

- Right click on that folder.
- Select the “Compress to ZIP file” option and then click “Compressed (Zipped) folder.”
- Anew.ZIP file will be created that contains your document(s).

Note:

- Plagiarism will result in a penalty, including possible project disqualification.
- The project will be evaluated based on the quality of analysis, depth of insights, and feasibility of recommendations.
- If the student sets their own criteria, they need to clearly mention and explain it.Marks will be given according to the specified criteria if they are acceptable.
- Remember to keep the video length less than 5 minutes with your face clearly visible.