Project Proposal: Predictive Modeling for Airbnb Rental Prices in San Francisco

Problem Statement:

The project aims to predict Airbnb rental prices in San Francisco using advanced machine learning methods. The primary objective is to develop a reliable model that property owners and existing hosts could potentially use to set competitive rental prices for their listings.

Context:

With the increasing demand for Airbnb rentals in San Francisco, property owners face the challenge of determining optimal pricing strategies. The project seeks to address this by leveraging data from insideairbnb.com and applying machine learning techniques to enhance the accuracy of rental price predictions.

This project not only addresses a real-world problem faced by Airbnb property owners but also enhances my data science portfolio by showcasing the application of machine learning methods in predictive modeling for the hospitality industry.

Criteria for Success:

- The success of the project will be measured by the model's ability to make accurate predictions.
- Additionally, success involves providing actionable recommendations for property owners to optimize their rental prices based on key features.

Scope of Solution Space:

The project focuses on predictive modeling for Airbnb rental prices, emphasizing the impact of features such as property capacity, host experience, and neighborhood clusters. While the initial scope is limited to San Francisco listings, the methodology can potentially be extended to other locations.

Constraints:

The project is limited by the available data from insideairbnb.com, with a specific focus on San Francisco listings from September 2022 to September 2023.

Stakeholders:

- Property owners who want to start hosting their property on Airbnb.
- Existing hosts on Airbnb.

Data Source:

The dataset is obtained from insideairbnb.com, offering comprehensive information on major world cities' Airbnb listings. Features include location, neighborhood, pricing, reviews, and host details. The dataset spans from September 2022 to September 2023, comprising 7208 rows and 75 columns.

Proposed Solution:

The project will train and develop regression models from available properties' information.

Deliverables:

The deliverables will include well-documented code, a comprehensive document detailing the project methodology, findings, and recommendations, a slide deck for presentation purposes and a model metrics file containing the final model features, parameters, hyperparameters, and performance metrics.