Project Proposal

Team: Dragon's Crest Analysts

Introduction:
The dataset we chose is from the "House Prices: Advanced Regression Techniques" competition on Kaggle, which contains a wide variety of features related to residential homes in Ames, Iowa. This dataset includes numerical and categorical data, ranging from basic attributes like the number of bedrooms to more detailed information such as the type of roof material or the condition of the garage.

Objective:
The primary goal of this project is to build a predictive model that accurately estimates the sale prices of residential properties based on their attributes. We employ advanced regression techniques to interpret the complex relationships between various house features and their impact on sale prices.

Data Overview:
The dataset includes a training set with 1460 instances, each described by 79 features, and a sale price. Features encompass a broad spectrum of attributes, including basic structural details, quality and condition assessments, and the presence of various amenities. A separate test set is provided without the sale prices, intended for model evaluation on Kaggle.

Methodology:
Data Preprocessing: Address missing values through imputation, engineer new features to enhance the dataset, and encode categorical variables for analysis.
Exploratory Data Analysis: Conduct an in-depth analysis to identify correlations, outliers, and the overall data distribution.
Model Development: Experiment with various regression techniques such as Linear Regression, Ridge/Lasso Regression, and ensemble methods like Random Forests and Gradient Boosting. Utilize cross-validation for model evaluation, feature selection, and hyperparameter tuning to enhance performance.
Model Evaluation: Use metrics like RMSE for assessing accuracy and analyze residuals to ensure model assumptions are met.
Feature Importance: Determine the impact of different features on house prices to gain insights into the housing market

Expected Outcomes:
The successful completion of this project will result in a robust predictive model capable of estimating house prices with high accuracy. Furthermore, the feature importance analysis will provide valuable insights into the housing market, potentially informing both sellers and buyers about key factors affecting property values.