Linear Regression XGBoost Sales Price Prediction Models

A data science project that required significant data preparation including the handling of null values to build a linear regression model and an XGBoost model for the prediction of home sale prices in Ames, Iowa.

About

This project derives from a Kaggle machine learning competition focusing on housing prices. Participants investigate a multitude of variables that may factor into the price of a home. A linear regression model and an XGBoost model leverage a sizable dataset featuring nearly eighty variables that could impact home price. The XGBoost model allowed for the identification of impactful features, which were plotted by feature importance to easily identify the key features.

Link

Dataset: https://www.kaggle.com/competitions/home-data-for-ml-course/data

Setup

This Python coding project was constructed on Jupyter Notebook. Please import the following libraries:

- Matplotlib
- Numpy
- Pandas
- Seaborn
- Sklearn
- Xgboost