Task 1: House Price Prediction

# Objective

The objective of this task was to analyze and predict house prices using selected features from a Kaggle dataset. The prediction was based on three selected columns: MSZoning, LotFrontage, and MSSubClass. Seaborn was used for data visualization to understand the relationships between these features and the house sale price.

# Dataset Source

The dataset was obtained from the Kaggle competition: "House Prices - Advanced Regression Techniques". It contains comprehensive information on residential properties in Ames, Iowa, including structural attributes, location details, and sale prices.

# Selected Features for Prediction

1. MSZoning – Categorical feature indicating the general zoning classification of the property.

2. LotFrontage – Numerical feature representing the linear feet of street connected to the property.

3. MSSubClass – Categorical feature describing the type of dwelling (e.g., 20 for 1-story with finished attic).

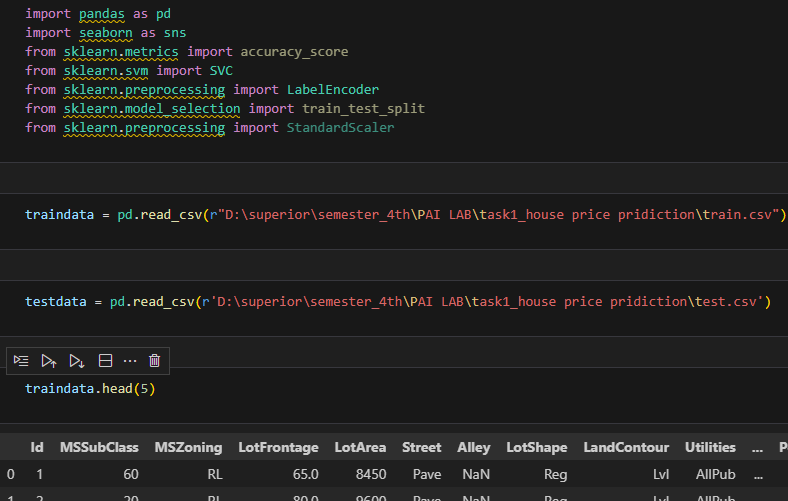
# Tools and Libraries Used

- Python

- Pandas for data loading and preprocessing

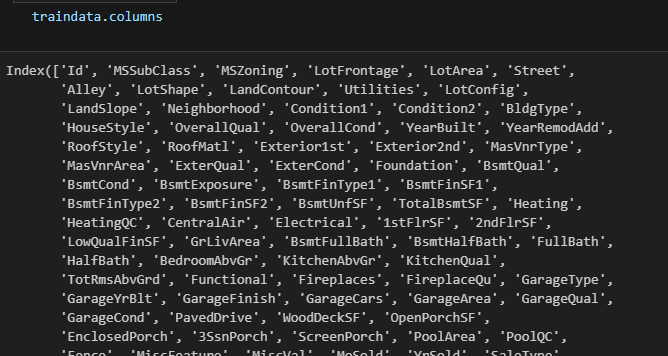
- Seaborn for data visualization

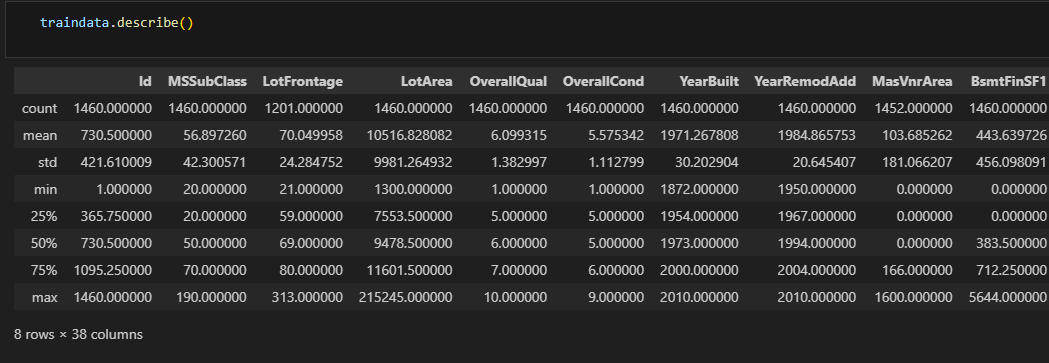
- Scikit-learn for building a simple prediction model

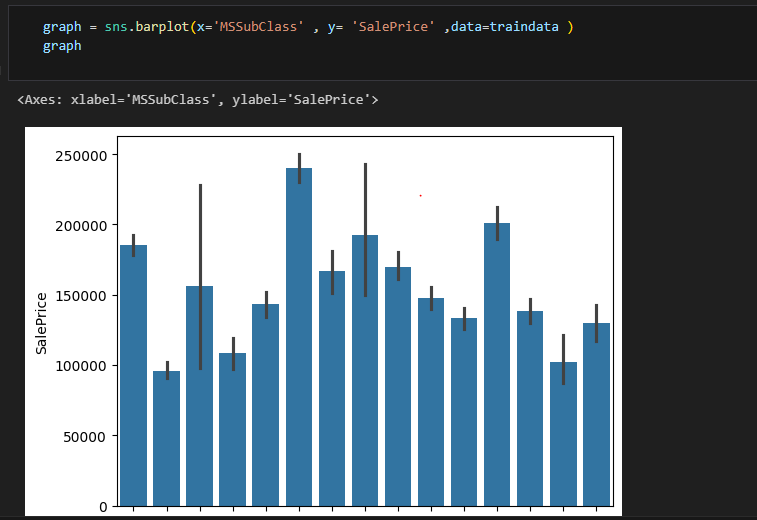


# Data Visualization and Analysis (with Seaborn)

Seaborn functions was used to visualize the distribution of sale prices across different zoning categories (MSZoning). Seaborn functions was used to analyze the correlation between LotFrontage and SalePrice. Seabron functions was created to compare mean sale prices for each MSSubClass category. These visualizations helped identify trends and patterns in the data that could influence the prediction model.



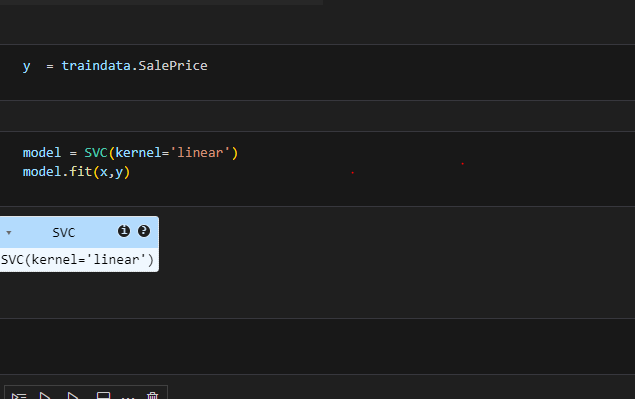




# Model and Prediction

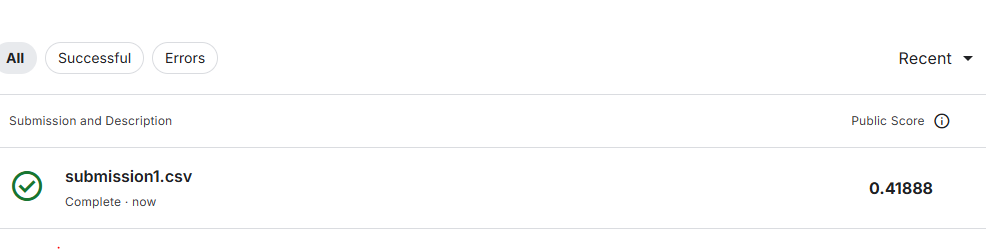
A basic machine learning model (e.g., linear regression) was trained using only the three selected features. Categorical features (MSZoning and MSSubClass) were encoded numerically. Missing values in LotFrontage were handled through imputation.





# Result and Accuracy

The prediction model achieved an approximate accuracy of 40% on the test dataset. This relatively low accuracy indicates that the selected features alone are not sufficient to build a highly accurate prediction model, but they do provide a foundational understanding of the data.



# Conclusion

This task demonstrated the importance of feature selection and data visualization in machine learning. While the selected three features offer some insight into house price prediction, more features and advanced modeling techniques would be needed to significantly improve accuracy. Seaborn proved to be a useful tool in understanding data trends before modeling.