Initial Tests of models

1. Analyse the performance of various models

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **RMSE** | **MAE** | **R² Score** |
| XGBoost | 239.636461 | 138.830618 | 0.924031 |
| Random Forest | 267.938703 | 124.441007 | 0.905027 |
| Gradient Boosting | 300.454949 | 187.542087 | 0.880576 |
| Lasso Regression | 335.683611 | 192.796115 | 0.850930 |
| Ridge Regression | 337.454512 | 211.273202 | 0.849353 |
| Linear Regression | 360.589641 | 240.249588 | 0.827988 |

1. Tuning the Hyperparameters of XGBoost

|  |  |  |
| --- | --- | --- |
| **Metric** | **Default XGBoost** | **Tuned XGBoost** |
| MSE | 72095.77 | 54008.93 |
| RMSE | 268.51 | 232.40 |
| MAE | 143.07 | 137.66 |
| R² Score | 0.9046 | 0.9286 |
| CV R² Score | N/A | 0.8544 |

param\_grid = {

        'model\_\_learning\_rate': [0.01, 0.05, 0.1, 0.2],

        'model\_\_max\_depth': [3, 5, 7],

        'model\_\_min\_child\_weight': [1, 3, 5],

        'model\_\_n\_estimators': [100, 200, 300],

        'model\_\_subsample': [0.6, 0.8, 1.0],

        'model\_\_colsample\_bytree': [0.6, 0.8, 1.0],

        'model\_\_gamma': [0, 0.1, 0.2]

    }

Best Result =

model\_\_learning\_rate: 0.1

model\_\_max\_depth: 5

model\_\_min\_child\_weight: 3

model\_\_n\_estimators: 200

1. Testing Ensemble Methods

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **XGBoost** | **Voting** | **Stacking** | **Custom** |
| MSE | 54008.93 | 72087.15 | 53070.96 | 56203.43 |
| RMSE | 232.40 | 268.49 | 230.37 | 237.07 |
| MAE | 137.66 | 160.22 | 135.04 | 138.64 |
| R² Score | 0.9286 | 0.9046 | 0.9298 | 0.9256 |

Custom used (Random Forest

Gradient Boosting

XGBoost)

1. Then tested different META-Learners with no better results