Regression Model - Short Report

📌 Preprocessing Steps and Rationale

1. **Handling Missing Values:**

   ◦ Checked for null values and imputed them using mean/mode.

2. **Feature Scaling:**

   ◦ Standardized numerical features using StandardScaler.

3. **Outlier Removal:**

   ◦ Used Interquartile Range (IQR) method to remove extreme values.

4. **Dimensionality Reduction:**

   ◦ Principal Component Analysis (PCA) was applied to reduce feature space while preserving variance.

Insights from Dimensionality Reduction

• PCA helped in reducing feature redundancy and improving model efficiency.

• The first few principal components retained most of the variance, indicating that a reduced feature set could be effective.

🔍 Model Selection, Training, and Evaluation

Models Trained:

1. **Ridge Regression**

2. **Lasso Regression**

3. **ElasticNet Regression**

4. **Random Forest Regressor**

5. **Gradient Boosting Regressor**

6. **Support Vector Regressor (SVR)**

7. **XGBoost Regressor**

Best Performing Model: Lasso Regression

Model	MAE	RMSE	R² Score
Ridge Regression	581.66	775.20	0.0266
Lasso Regression	**579.82**	**773.37**	**0.0312**
ElasticNet Regression	584.96	778.60	0.0180
Random Forest Regressor	517.06	807.85	-0.0571
Gradient Boosting Regressor	566.76	794.86	-0.0234
Support Vector Regressor (SVR)	516.76	784.56	0.0029
XGBoost Regressor	556.49	793.99	-0.0212

Model Performance Summary

- **Lasso Regression** performed the best with the highest R^2 score (**0.0312**).

- **Random Forest & Gradient Boosting** had negative R^2, indicating poor generalization.

- **SVR & XGBoost** also failed to outperform linear models.

🔄 Key Findings & Suggestions for Improvement

✅ Findings:

- Linear models (Lasso, Ridge) outperformed complex models like Random Forest and XGBoost.

- Feature selection & scaling played a crucial role in performance.

🚀 Improvements:

- Implement **hyperparameter tuning** for tree-based models.

- Try **feature engineering** to extract meaningful variables.

- Experiment with **neural networks** for potential improvement.