Tools Used in This Project

1. Programming Language

- **Python** – The primary programming language used for data analysis, preprocessing, and model development.

2. Development Environment

- **Jupyter Notebook / Google Colab** – For interactive coding, visualization, and experimentation.

3. Libraries and Frameworks

- **Pandas** – For data manipulation, cleaning, and transformation.
- **NumPy** – For numerical computations and handling arrays.
- **Matplotlib** – For data visualization and plotting graphs.
- **Seaborn** – For advanced and visually appealing statistical plots.
- **Scikit-learn (sklearn)** – For implementing machine learning models, preprocessing, and evaluation metrics.
- **Joblib / Pickle** – For saving and loading machine learning models or encoders.

4. Data Handling & Encoding Tools

- **LabelEncoder / OneHotEncoder** – For encoding categorical variables.
- **train_test_split** – For splitting data into training and testing sets.

5. Machine Learning Models Used

- **Linear Regression / Logistic Regression** – For predictive modeling.
- **Random Forest / XGBoost / Gradient Boosting** – For ensemble learning and performance optimization.

6. Evaluation Metrics

- **Accuracy / RMSE / R^2 / Precision / Recall / F1-score** – For evaluating model performance.
- **Silhouette Score** – For clustering model validation.

7. Version Control and Documentation

- **Git & GitHub** – For version control, collaboration, and repository management.
- **Markdown** – For documentation and explanation of project workflow.

8. Deployment (Optional)

- **Streamlit / Flask** – For building simple web apps (if applicable).

This setup ensures an efficient, reproducible, and scalable machine learning workflow — from data preprocessing to model evaluation and documentation.