

Project Overview

Objective:

Develop a predictive model to classify customer segments based on various retail-related features. This project involves using Python for data preprocessing and exploratory data analysis (EDA), applying machine learning algorithms for model building, and utilizing Tableau for visualization and insights.

Tasks to Perform

1. Data Preprocessing

- **Load the Dataset:** Import the dataset into a Python environment.
- **Clean the Data:** Handle missing values, correct data types, and remove or handle outliers.
- **Encode Categorical Variables:** Convert categorical variables into a numerical format using techniques such as one-hot encoding or label encoding.
- **Normalize/Scale Features:** Normalize or scale numerical features to ensure they contribute equally to the model.

2. Exploratory Data Analysis (EDA)

- **Descriptive Statistics:** Calculate summary statistics for numerical and categorical features.
- **Data Visualization:** Use libraries like Matplotlib and Seaborn to visualize the distribution of features, relationships between features, and any patterns or trends in the data.
- **Correlation Analysis:** Identify correlations between features and the target variable to understand potential predictors of customer segments.
- **Outlier Detection:** Detect and visualize outliers to understand their impact on the analysis.

3. Feature Engineering

- **Create New Features:** Derive new features from existing ones that might help in classifying customer segments.
- **Select Important Features:** Use techniques such as correlation analysis, feature importance from tree-based models, or statistical tests to select the most relevant features for the model.

4. Model Building and Evaluation

- **Train-Test Split:** Split the dataset into training and testing sets to evaluate model performance.
- **Baseline Model:** Start with simple classification models (e.g., Logistic Regression) to establish a baseline performance.

- **Advanced Models:** Train more complex models (e.g., Decision Trees, Random Forest, Gradient Boosting) to improve performance.
- **Model Evaluation:** Evaluate models using metrics such as accuracy, precision, recall, F1-score, and confusion matrix to compare their performance.
- **Hyperparameter Tuning:** Optimize model parameters using techniques such as grid search or random search to improve model performance.

5. Visualization and Insights

- **Tableau Dashboard:** Create interactive dashboards and visualizations in Tableau to present key insights and findings from the data analysis and modeling.
- **Feature Impact:** Visualize the impact of different features on customer segments to provide actionable insights for retail decision-makers.
- **Summary of Findings:** Summarize the key findings from the analysis and modeling, highlighting the most significant predictors of customer segments.

6. Reporting

- **Documentation:** Document the entire process, including data preprocessing steps, EDA findings, feature engineering, model building, and evaluation results.
- **Presentation:** Prepare a presentation or report summarizing the methodology, key insights, and recommendations based on the analysis.