**Lab 8: Feature Engineering – Creating and Transforming Variables**

**Prelab Questions**

1. What is feature engineering, and why is it important in data analysis?
2. What are derived features? Give an example.
3. How can categorical variables be transformed for machine learning models?
4. Explain the difference between normalization and standardization.
5. Why is it necessary to handle skewed data before modeling?

**In-Lab Details**

**Objective**:

* Learn how to create new features and transform existing ones to improve data representation.

**Resources**:

* Python (Jupyter Notebook).
* Libraries: Pandas, NumPy, Scikit-learn.
* Dataset: ecommerce\_data.csv containing order amount, customer age, and product category.

**Expected Output**

1. **New Feature: Total\_Spent**: Captures spending patterns.
2. **Normalized Order\_Amount\_Scaled**: Scaled between 0 and 1 for better model compatibility.
3. **Standardized Age\_Standardized**: Mean-centered age variable with unit variance.
4. **One-Hot Encoded Product\_Category**: Converts categorical data into numerical format.

**Postlab Questions**

1. What is the purpose of feature engineering in predictive modeling?
2. Why is normalization essential when using distance-based models like KNN?
3. Compare one-hot encoding and label encoding for categorical variables.
4. How can feature selection improve model efficiency?
5. What are some common techniques for handling highly skewed features?