**Cloth Manufacturing Company Analysis**

**Introduction**

This document outlines the end-to-end process of analyzing a dataset from a cloth manufacturing company. The primary objective is to detect whether a store is located in the United States (US) or not, with the 'US' column as the dependent variable and the remaining columns as independent variables.

**Methodology**

**1. Data Preprocessing:**

**1.1 Importing Necessary Packages and Extracting Data**

The initial step involved importing essential Python packages and extracting the company data.

**1.2 Handling Missing Values and Encoding Categorical Data**

No null values were identified in the dataset. Three columns with categorical data were converted into numerical format using label encoding.

**1.3 Exploratory Data Analysis (EDA)**

EDA was conducted to understand relationships between different variables. A heatmap was used for correlation analysis, and boxplots aided in identifying outliers.

**1.4 Outlier Detection and Imputation**

Outliers were detected using boxplots, and imputation was performed using threshold values to maintain data integrity.

**1.5 Save Cleaned Data**

The cleaned data was saved separately for future use.

**2. Modeling:**

**2.1 Data Splitting and Scaling**

The 'US' column was set as the dependent variable (y), and the rest of the columns were used as independent variables (X). The data was scaled using Standard Scaler, and a train-test split of 80:20 was performed.

**2.2 Random Forest Classifier**

The data was fitted into a Random Forest classifier to predict the 'US' column for the test set.

**3. Model Evaluation:**

**3.1 Accuracy Assessment**

The model was evaluated for accuracy by predicting 'US' values for the test set. The entire process was repeated 1000 times with different random states, resulting in an overall accuracy of 98.75%.

**Conclusion**

The project successfully analyzed the cloth manufacturing company's dataset to predict whether a store is located in the United States. Through careful preprocessing, exploratory data analysis, and the implementation of a Random Forest classifier, the model achieved a high accuracy rate of 98.75%.