```sas
/*******************************************************************************
 * BIA 652-WS – Multivariate Data Analysis
 * Final Project: Understanding Customer Behavior and Churn Risk in E-Commerce
 *
 * Author: Sanjeeth Rao
 * Team: Team 3
 * Semester: Spring 2025
 *
 * Objective:
 * This program imports a cleaned customer-level dataset and applies the following
 * multivariate techniques in alignment with course methods:
 * - Descriptive Statistics & Normality Checks
 * - Exploratory Factor Analysis (EFA)
 * - Multiple Regression Analysis
 * - Logistic Regression for Churn Prediction
 * - Cluster Analysis for Customer Segmentation
 * - MANOVA for Country-Based Group Comparison
 *
 * Dataset Source:
 * UCI Online Retail Dataset (https://archive.ics.uci.edu/ml/datasets/Online+Retail)
 * Cleaned and engineered in Python, imported here for SAS-based statistical modeling.
 *
 * All procedures and logic below follow the structure and statistical techniques
 * taught and practiced in class.
 *******************************************************************************/

/* Step 1: Import cleaned Python file */
proc import datafile="/home/u64166562/sasuser.v94/retail_customer_summary (1).csv"
    out=retail_raw
    dbms=csv
    replace;
    getnames=yes;
run;

/* Step 2: Rename variables to match SAS class structure */
data retail;
    set retail_raw;

    X1 = Recency;
    X2 = PurchaseFrequency;
    X3 = TotalSpend;
    X4 = AvgOrderValue;
    X5 = AvgUnitPrice;
    X6 = TotalQuantity;
    X7 = Churn;
    X8 = ReturnedQuantity;
    X9 = ReturnRate;
    X10 = CustomerID;
    X11 = Country;
run;

/* Step 3: Descriptive stats and normality */
proc means data=retail n mean std min max;
    var X1-X6;
run;

proc univariate data=retail normal plot;
```

```sas
    var X1 X3 X4 X5 X9; /* Recency, Spend, AOV, Price, ReturnRate */
run;


/* Step 4: Factor analysis on behavioral traits */
proc factor data=retail method=principal rotate=varimax msa plots=scree nFactors=2;
    var X1 X2 X3 X4 X5 X9; /* Recency, Frequency, Spend, AOV, Unit Price, ReturnRate */
run;


/* Step 5: Regression – Predict TotalSpend */
proc reg data=retail plots(unpack);
    model X3 = X1 X2 X4 X5 X6 X9 / vif stb influence tol;
run;


/* Check residual normality */
proc reg data=retail;
    model X3 = X1 X2 X4 X5 X6 X9;
    output out=reg_out r=residuals;
run;


proc univariate data=reg_out normal plot;
    var residuals;
run;


/* Step 6a: Check class distribution for Churn */
proc freq data=retail;
    tables X7 / nocum nopercent;
run;


/* Step 6b: Logistic regression – Refined model to avoid separation */
/* Dropping TotalSpend (X3) as it is highly correlated with Churn */
proc logistic data=retail descending;
    model X7 = X1 X4 X9; /* Recency, AOV, ReturnRate */
run;


/* Step 7a: Standardize variables to mean=0, std=1 */
proc standard data=retail mean=0 std=1 out=retail_std;
    var X1 X3 X4 X5 X9; /* Recency, TotalSpend, AOV, UnitPrice, ReturnRate */
run;


/* Step 7b: Cluster analysis (K-means) */
proc fastclus data=retail_std maxclusters=4 out=clustered;
    var X1 X3 X4 X5 X9;
run;


/* Step 7c: Use MANOVA to validate cluster differences */
proc glm data=clustered;
    class cluster;
    model X1 X3 = cluster; /* Test whether Recency & Spend vary by cluster */
    manova h=cluster / mstat=exact;
run;


/* Step 7d: ANOVA Cluster Profile Validation - Test each variable across clusters */
proc glm data=clustered;
    class cluster;
    model X1 = cluster;  /* Recency */
    means cluster / tukey;
run;
```

```sas
proc glm data=clustered;
    class cluster;
    model X3 = cluster;  /* TotalSpend */
    means cluster / tukey;
run;

proc glm data=clustered;
    class cluster;
    model X4 = cluster;  /* AvgOrderValue */
    means cluster / tukey;
run;

proc glm data=clustered;
    class cluster;
    model X5 = cluster;  /* AvgUnitPrice */
    means cluster / tukey;
run;

proc glm data=clustered;
    class cluster;
    model X9 = cluster;  /* ReturnRate */
    means cluster / tukey;
run;

/* Step 7e: Export final clustered dataset to CSV */
proc export data=clustered
    outfile="/home/u64166562/sasuser.v94/retail_customer_summary (1).csv"
    dbms=csv replace;
run;

/* Step 8a: MANOVA to compare behavior across countries */
proc glm data=retail;
    class X11; /* Country */
    model X2 X3 = X11; /* Frequency and Spend */
    means X11 / hovtest=levene hovtest=bf hovtest=bartlett;
    manova h=X11 / mstat=exact;
run;

proc export data=retail
    outfile="/home/u64166562/sasuser.v94/retail_customer_summary (1).csv"
    dbms=csv replace;
run;

proc export data=clustered
```