

```

#Calculate Customer Age
current_year <- as.numeric(format(Sys.Date(), "%Y"))
smartfresh$Customer_Age <- current_year - smartfresh$Year_Birth

# Calculate Total spend
smartfresh$Total_Spend <- smartfresh$Spend_Wine + smartfresh$Spend_OrganicFood +
smartfresh$Spend_Meat +
  smartfresh$Spend_WellnessProducts + smartfresh$Spend_Treats +
smartfresh$Spend_LuxuryGoods

#Label Encoding Categorical Variables
smartfresh$Education_Level <- as.numeric(factor(smartfresh$Education_Level))
smartfresh$Marital_Status <- as.numeric(factor(smartfresh$Marital_Status))

#Central Tendency & Variance
cent_tendency <- data.frame(smartfresh[, c("Customer_Age", "Annual_Income", "Total_Spend",
"Education_Level",
"Marital_Status", "Purchases_Online",
"Purchases_Store",
"Kidhome", "Teenhome")])

summary(cent_tendency)
std_dev <- apply(cent_tendency, 2, sd)
variance <- apply(cent_tendency, 2, var)
print(std_dev)
print(variance)

summary(Key_variables)

library(ggplot2)

# Histograms: Distribution of Numerical Variables
ggplot(smartfresh, aes(x = Customer_Age)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "black") +
  ggtitle("Distribution of Customer Age")

ggplot(smartfresh, aes(x = Annual_Income)) +
  geom_histogram(binwidth = 10000, fill = "lightgreen", color = "black") +
  ggtitle("Distribution of Annual Income")

ggplot(smartfresh, aes(x = Total_Spend)) +
  geom_histogram(binwidth = 100, fill = "lightcoral", color = "black") +
  ggtitle("Distribution of Total Spend")

# Boxplots: Comparing Spend by Education Level
ggplot(smartfresh, aes(x = Education_Level, y = Total_Spend, fill = Education_Level)) +
  geom_boxplot() +
  ggtitle("Total Spend by Education Level")

# Scatter Plots: Relationships between Variables
ggplot(smartfresh, aes(x = Customer_Age, y = Annual_Income)) +
  geom_point(alpha = 0.5) +
  ggtitle("Customer Age vs. Annual Income")

ggplot(smartfresh, aes(x = Annual_Income, y = Total_Spend)) +
  geom_point(alpha = 0.5) +
  ggtitle("Annual Income vs. Total Spend")

```