Bank of America Marketing Strategy

Nirja Vaishnav

Praneth Nandeesh

Introduction

Bank of America launched the first nationally licensed credit card program, which was initially called BankAmerica. After rapid and extensive adoption and growth, the program grew globally and was eventually dubbed Visa. Bank of America has been tracking the activity of its 9000 active credit card clients for the past six months. It intends to investigate its credit card customers' usage behavior and spending patterns. The company's goal is to establish cluster-based segmentation for targeted tailored marketing. This analysis will assist them in determining which clients can benefit from specific credit card features in order to increase revenue. In addition, a focus will be made to flag consumers to whom no new card or increase limit will be issued for cost optimization.

Dataset Summary

After loading the BOA dataset and running early exploratory functions like dim, summary, and str, we determined the dataset's dimensions and summarized its structure. The dataset includes factors such as the balance, the frequency with which the balance is updated, purchases, cash advance details, and the duration of credit card service, among others. The dataset contained 8950 observations and 19 variables.

Code:

```
# Load necessary libraries

library(cluster)   # For daisy and clustering plots

library(mclust)    # For model-based clustering

library(ggplot2)


#  Load the data
```

```
boa_data <- read.csv("~/Downloads/BOA.csv")


#Explore the data
dim(boa_data)
str(boa_data)
head(boa_data)
```

Data Cleaning and Reprocessing

The data was cleaned by deleting unrelated columns (X and CUST_ID) and processing missing values with functions that ignore NA. Numeric variables were standardized using the scale() function to ensure that all features had a mean of 0 and a standard deviation of 1, avoiding variables with wider ranges from dominating the clustering.

Code:

```
#  Remove the first two columns (X and CUST_ID)
boa_data <- boa_data[, -c(1, 2)]
#  Scaling numeric variables
numeric_columns <- sapply(boa_data, is.numeric)  # Identify numeric columns
boa_data_scaled <- boa_data
boa_data_scaled[, numeric_columns] <- scale(boa_data[, numeric_columns])
```

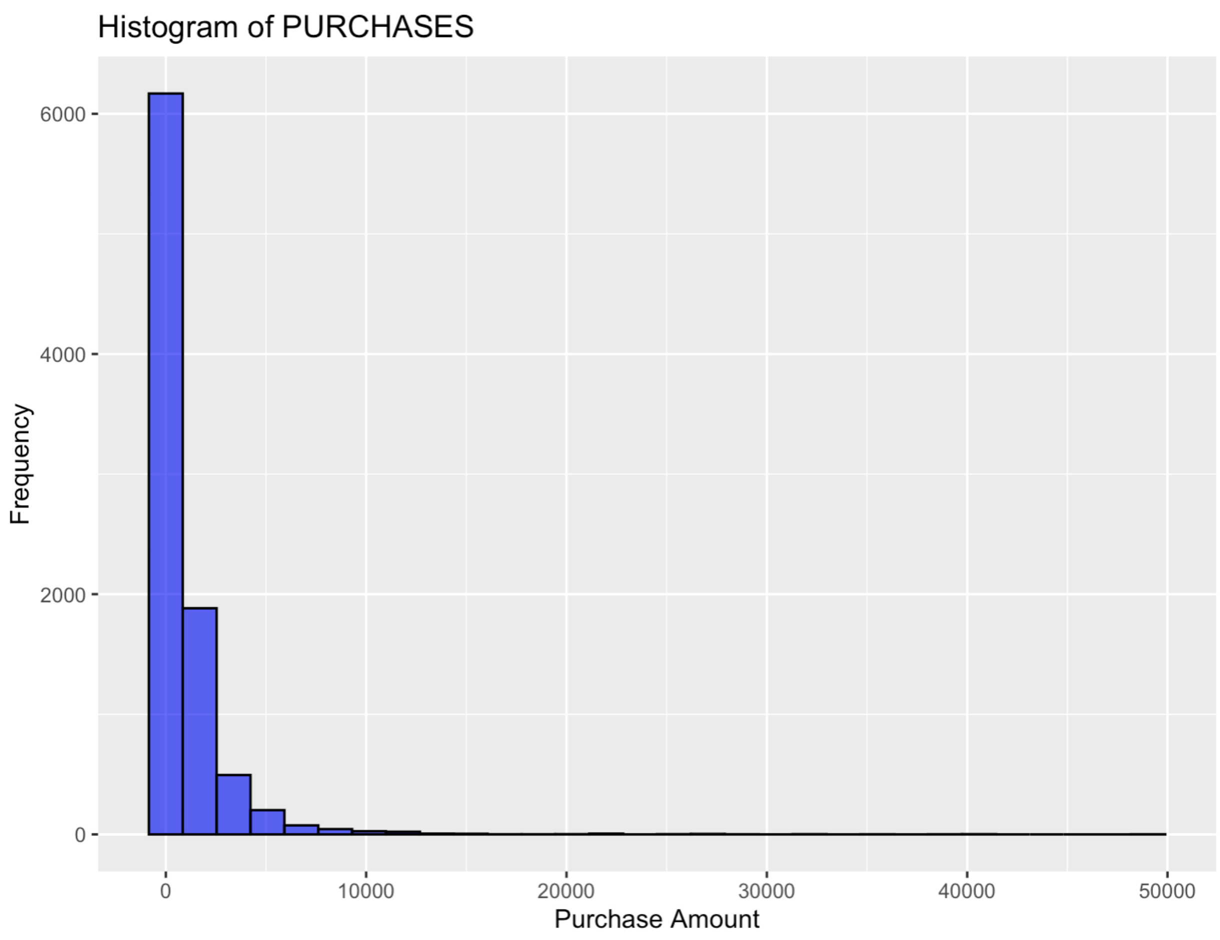
Exploratory Analysis

To better understand client spending habits, a histogram of the "PURCHASES" variable was generated.

Insights: Customers tend to make small purchases.
There are a few outliers with much greater buying values.

Code:

```
# Histogram of PURCHASES
ggplot(boa_data, aes(x = PURCHASES)) +
  geom_histogram(bins = 30, fill = "blue", color = "black", alpha = 0.7) +
  ggtitle("Histogram of PURCHASES") +
  xlab("Purchase Amount") +
  ylab("Frequency")
```



Clustering Approaches

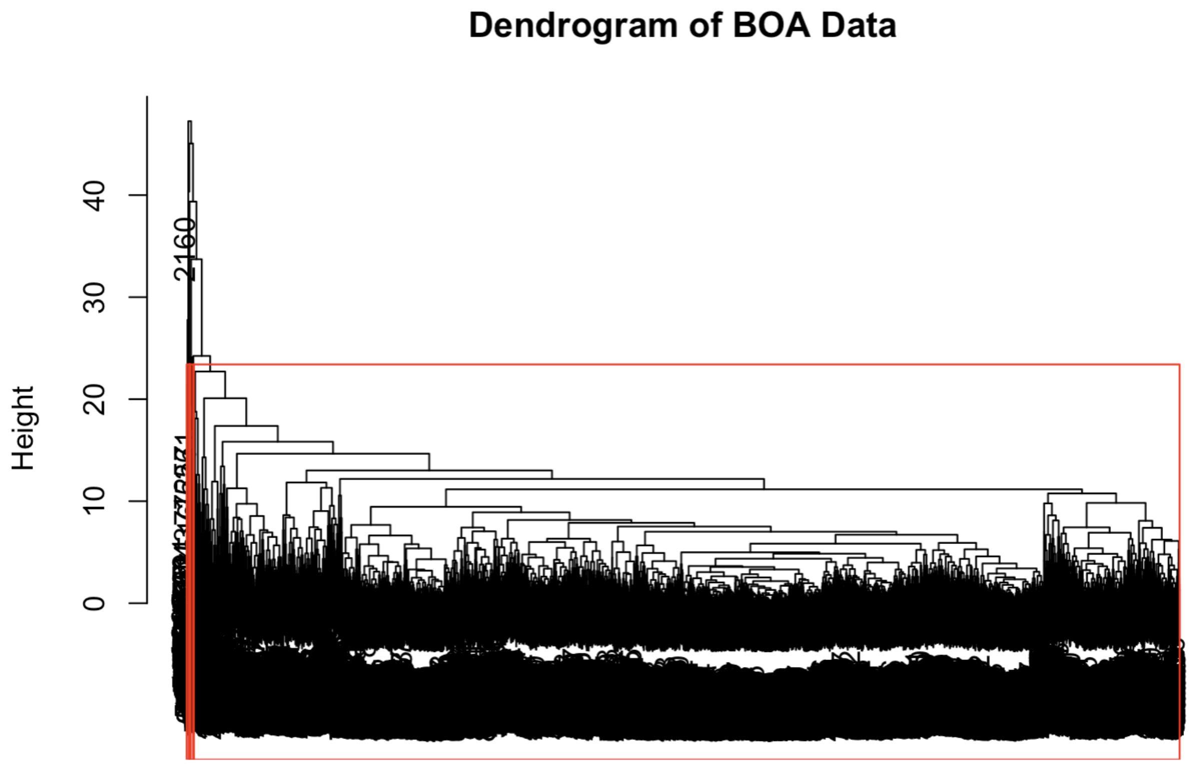
Hierarchical Clustering:

Approach:

~ Used a distance matrix computed with daisy to handle mixed data types.

~ Performed hierarchical clustering with the "complete linkage" method.

~ Cluster memberships were assigned by cutting the dendrogram at 8 clusters.



Code:

```
#  Hierarchical Clustering
boa_dist <- daisy(boa_data_scaled)  # Compute distance matrix for mixed data types
boa_hc <- hclust(boa_dist, method = "complete")  # Hierarchical clustering
plot(boa_hc, main = "Dendrogram of BOA Data", xlab = "", sub = "")
```

```
rect.hclust(boa_hc, k = 8, border = "red")  # Cut tree into 4 clusters

boa_hc_clusters <- cutree(boa_hc, k = 8)  # Cluster memberships

table(boa_hc_clusters)

hc_summary <- seg.summ(boa_data_scaled, boa_hc_clusters)

print(hc_summary)
```

Summary:

~The majority of customers (Cluster 1) have similar behavioral patterns, representing 99% of the dataset.

~Smaller clusters (Clusters 2 to 8) represent outliers or specific customer subgroups.

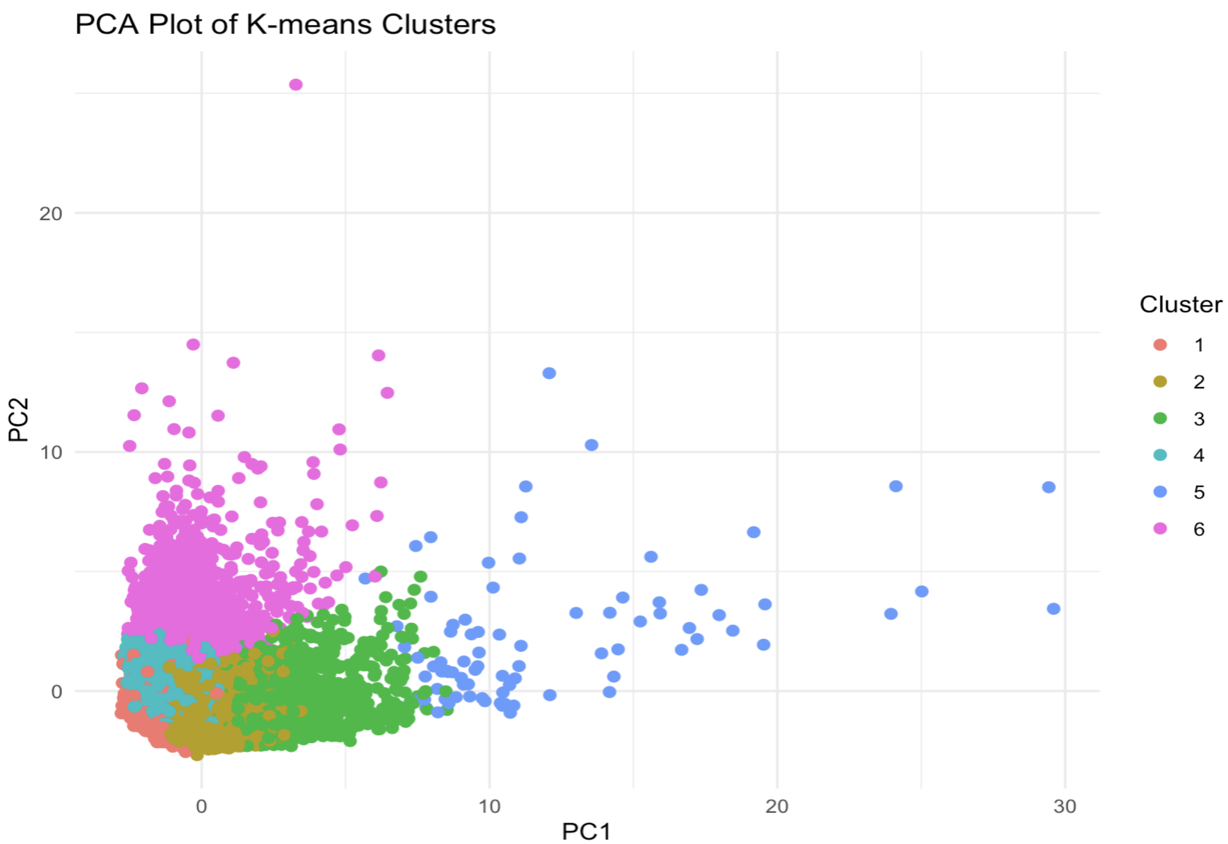
```
boa_hc_clusters
   1    2    3    4    5    6    7    8
8887   26   20    3    5    7    1    1
```

K-means Clustering

Process:

~ Applied to the scaled dataset with six clusters (k=6), using 25 random initializations for robustness.

~PCA was used to visualize the cluster separations in two-dimensional space.



The PCA plot of K-means clusters illustrates consumer segmentation based on financial behavior. Each point represents a client, and the colors indicate cluster memberships. Clusters 1 and 2, located in the bottom-left corner, reflect low-spenders or inactive consumers. Clusters 3 and 4, which stretch to the center and right, represent moderate-to-high spenders with increased transaction activity. Clusters 5 and 6, shown in the upper right corner, represent affluent or high-value consumers with significant credit utilization. This segmentation identifies separate client groups, allowing for targeted marketing methods such as loyalty programs for lower clusters and premium services for higher-value clusters.

Cluster Distribution:

~ Cluster 5 is notable for having the smallest population (83 customers), with extreme purchase behaviors.

~ Clusters 1 and 4 represent over 50% of the dataset, indicating general customer behavior.

```
> table(boa_kmeans$cluster)

   1    2    3    4    5    6 
1310 2227 1323 3037   83  970 
```

Key Findings:

~ Cluster 5: High purchase frequency and credit limit.

~ Cluster 6: Reliant on cash advances, with negative purchasing behavior.

Code:

```
#  K-means Clustering
set.seed(1234)  # For reproducibility
boa_kmeans <- kmeans(boa_data_scaled, centers = 6, nstart = 25)  # K-means with multiple starts
table(boa_kmeans$cluster)
kmeans_summary <- seg.summ(boa_data_scaled, boa_kmeans$cluster)
print(kmeans_summary)

#  Visualize K-means Clusters Using PCA
library(ggplot2)
pca <- prcomp(boa_data_scaled[, numeric_columns])
pca_data <- data.frame(pca$x[, 1:2], Cluster = as.factor(boa_kmeans$cluster))
ggplot(pca_data, aes(PC1, PC2, color = Cluster)) +
  geom_point(size = 2) +
  ggtitle("PCA Plot of K-means Clusters") +
  theme_minimal()
```

Model-Based Clustering (Mclust)

Process:
Gaussian finite mixture modeling was applied to the dataset, evaluating models with up to ten components. The best-fit model (EEV) was selected, which assumes ellipsoidal clusters with equal volume and shape, and identified six distinct clusters.

```
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EEV (ellipsoidal, equal volume and shape) model with 6 components:

 log-likelihood    n  df       BIC       ICL
       -19581.6 8950 940 -47716.65 -49123.95

Clustering table:
   1    2    3    4    5    6
2144 2119 1065 2342  436  844
```

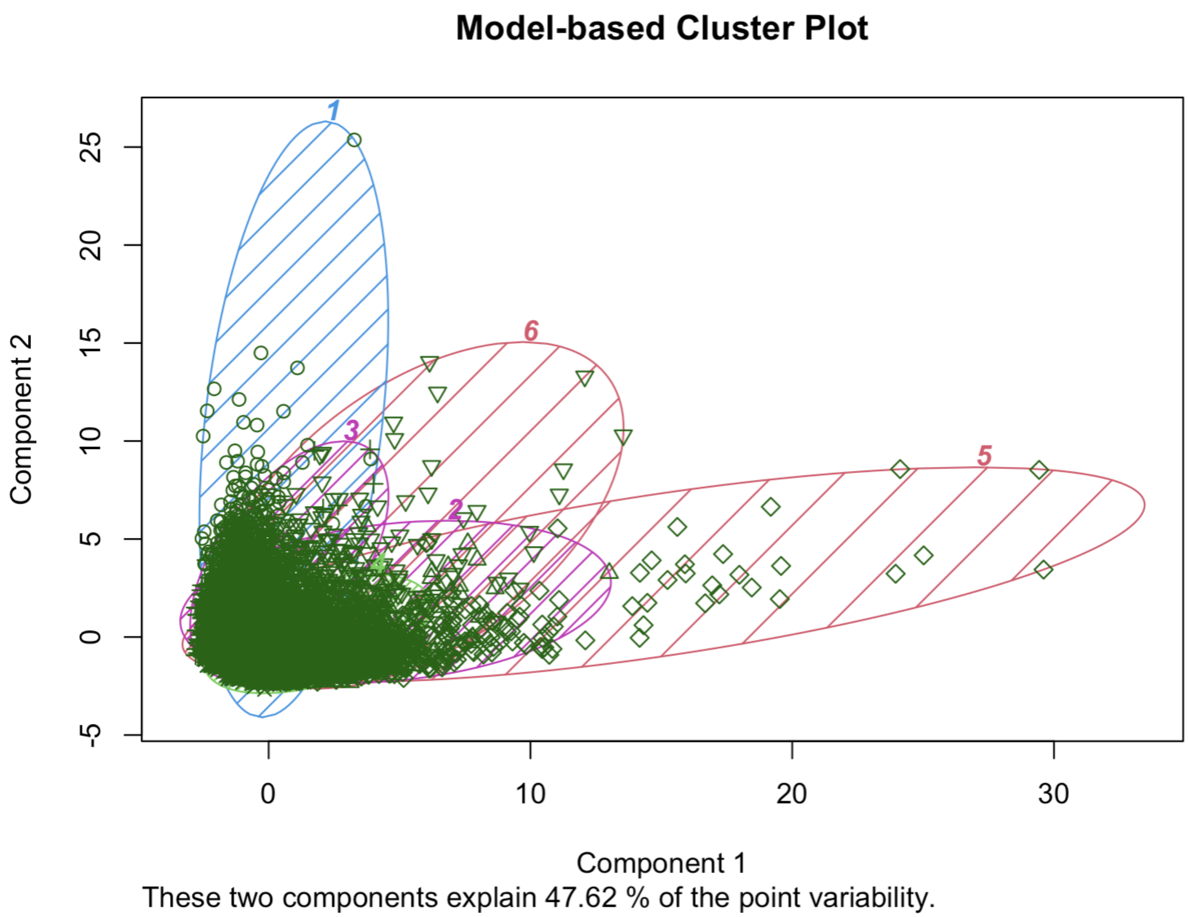
Cluster Insights:

Cluster 4: High credit limit, frequent purchases, and consistent payment behavior. Likely represents high-value, financially active customers.

Cluster 2: Low spending and minimal credit utilization, reflecting conservative financial behavior.

Cluster 3: Moderate spending but less frequent engagement, overlapping with Cluster 4 in purchasing behavior.

Visualization:
The model-based cluster plot reveals significant overlaps, particularly among mid-spending customers (e.g., Clusters 3 and 4), indicating shared financial patterns or behavior variability within this group.



Code:

```
#  Model-based Clustering (Mclust)
boa_mclust <- Mclust(boa_data_scaled, G = 1:10)  # Fit with 1 to 10 clusters
summary(boa_mclust)
clusplot(boa_data_scaled, boa_mclust$class, color = TRUE, shade = TRUE,
        labels = 4, lines = 0, main = "Model-based Cluster Plot")
mclust_summary <- seg.summ(boa_data_scaled, boa_mclust$class)
print(mclust_summary)
```

Strategy Based on Clustering Analysis

Based on the segmentation outputs from the hierarchical clustering (hclust), K-means, and model-based clustering (Mclust) methods, the following customer profiles and strategies are proposed.

Customer Profiles

Cluster 1: Low Spend, Low Risk

Customers with low spending activity, minimal cash advances, and consistent payments.

Strategy:

Retain these customers with loyalty rewards for maintaining low-risk financial behavior and incentivize them to explore new services like higher credit limits or installment plans.

Cluster 2: High Cash Advances, Financial Strain

Customers with frequent cash advances but limited purchase activity, indicating financial stress.

Strategy:

Offer financial planning services, debt management tools, or lower-interest credit products to ease financial burden and reduce risk.

Cluster 3: Moderate Purchases, Low Full Payments

Customers with average purchase behavior but low rates of full payments, potentially posing medium risk.

Strategy:

Encourage full payments through cashback offers, lower interest rates on balances paid in full, and educational campaigns about reducing interest charges.

Cluster 4: High Purchases, High Risk

High-spending customers with significant one-off purchases but minimal full payments, indicating a higher risk of default.

Strategy:

Closely monitor credit activity, set tailored credit limits, and proactively engage these customers with repayment assistance or rewards for timely payments.

Cluster 5: Elite Customers

Customers with the highest purchase activity (both one-off and installment) and a consistent pattern of full payments, presenting the lowest risk.

Strategy:

Prioritize these customers with premium offers, exclusive benefits, and personalized recommendations for high-value services like investments or wealth management.

Cluster 6: Average Behavior

Customers with balanced purchase activity and low cash advances, representing low engagement.

Strategy:

Increase engagement through personalized campaigns highlighting product benefits, promotional offers, and upgrade incentives.

Implementation Plan

Customer Retention: Strengthen relationships with low-risk and elite customers through loyalty programs and premium services.

Risk Mitigation: Proactively manage high-risk segments with financial education and repayment support tools.

Engagement Increase: Use targeted marketing to activate low-engagement segments, fostering greater utilization of BOA's products.

This segmentation and strategy ensure BOA optimizes revenue, mitigates risk, and enhances customer satisfaction, forming a robust foundation for sustained growth and operational efficiency.

APPENDIX

```
# Load necessary libraries
library(cluster)   # For daisy and clustering plots
library(mclust)    # For model-based clustering
library(ggplot2)

#  Load the data
boa_data <- read.csv("~/Downloads/BOA.csv")

#Explore the data
dim(boa_data)
str(boa_data)
head(boa_data)

#  Remove the first two columns (X and CUST_ID)
boa_data <- boa_data[, -c(1, 2)]

# Histogram of PURCHASES
ggplot(boa_data, aes(x = PURCHASES)) +
  geom_histogram(bins = 30, fill = "blue", color = "black", alpha = 0.7) +
  ggtitle("Histogram of PURCHASES") +
  xlab("Purchase Amount") +
  ylab("Frequency")

#  Scaling numeric variables
numeric_columns <- sapply(boa_data, is.numeric)  # Identify numeric
columns
boa_data_scaled <- boa_data
boa_data_scaled[, numeric_columns] <- scale(boa_data[, numeric_columns])

#  cluster summarization function
boa.summ <- function(data, groups) {
  aggregate(data, list(Group = groups), function(x) mean(as.numeric(x),
na.rm = TRUE))
}

#  Hierarchical Clustering
boa_dist <- daisy(boa_data_scaled)  # Compute distance matrix for mixed
data types
boa_hc <- hclust(boa_dist, method = "complete")  # Hierarchical clustering
plot(boa_hc, main = "Dendrogram of BOA Data", xlab = "", sub = "")
rect.hclust(boa_hc, k = 8, border = "red")  # Cut tree into 4 clusters
boa_hc_clusters <- cutree(boa_hc, k = 8)  # Cluster memberships
table(boa_hc_clusters)
hc_summary <- boa.summ(boa_data_scaled, boa_hc_clusters)
print(hc_summary)

#  K-means Clustering
```

```
set.seed(1234)  # For reproducibility
boa_kmeans <- kmeans(boa_data_scaled, centers = 6, nstart = 25)  # K-means
with multiple starts
table(boa_kmeans$cluster)
kmeans_summary <- boa.summ(boa_data_scaled, boa_kmeans$cluster)
print(kmeans_summary)

#  Visualize K-means Clusters Using PCA
library(ggplot2)
pca <- prcomp(boa_data_scaled[, numeric_columns])
pca_data <- data.frame(pca$x[, 1:2], Cluster =
as.factor(boa_kmeans$cluster))
ggplot(pca_data, aes(PC1, PC2, color = Cluster)) +
  geom_point(size = 2) +
  ggtitle("PCA Plot of K-means Clusters") +
  theme_minimal()

#  Model-based Clustering (Mclust)
boa_mclust <- Mclust(boa_data_scaled, G = 1:10)  # Fit with 1 to 10
clusters
summary(boa_mclust)
clusplot(boa_data_scaled, boa_mclust$class, color = TRUE, shade = TRUE,
         labels = 4, lines = 0, main = "Model-based Cluster Plot")
mclust_summary <- boa.summ(boa_data_scaled, boa_mclust$class)
print(mclust_summary)
```