Advanced Customer Analytics: Segmentation, Customer Churn and Predictive Marketing Model

Team: Revanth Vemula, Harshil Chauhan, Raghava Sagar
Instructor: Chris Asakeiwicz

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®
Business Intelligence & Analytics

Introduction

In the competitive retail industry, increasing customer engagement and retention is vital for businesses to drive revenue growth and improve customer loyalty.

By leveraging advanced customer analytics, including segmentation, churn prediction, and predictive marketing models, retail companies can identify high-risk customers, understand their specific needs and preferences, and design targeted retention strategies to reduce churn.

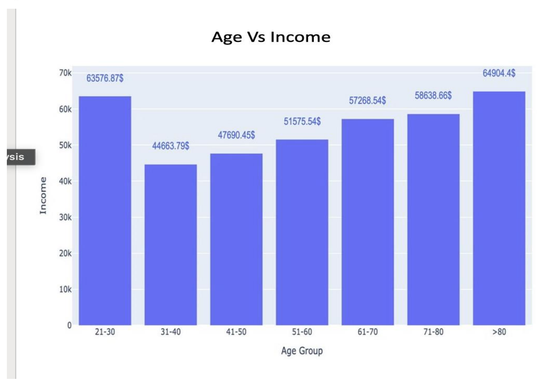
Through predictive modeling, businesses can accurately forecast customer response to future marketing campaigns and improve overall customer engagement and loyalty.

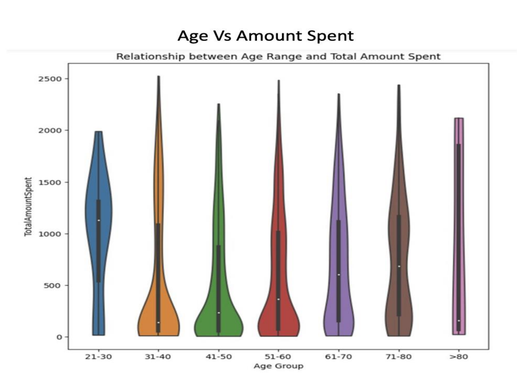
Research Question

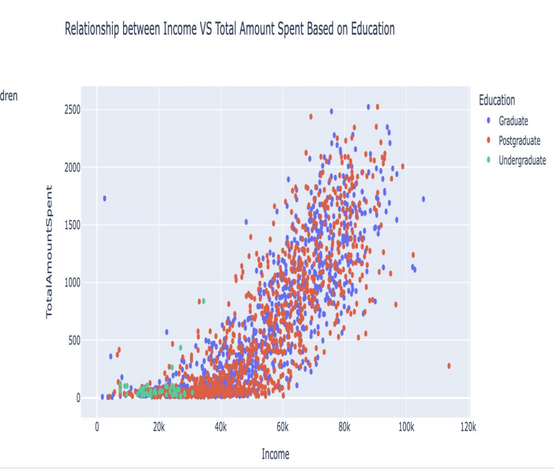
How can we segment customers based on their demographic, socioeconomic, and purchasing behavior attributes?
Which customers are at high and low risk of churning?
How to predict the response of future marketing campaigns?

Exploratory Data Analysis

Income, education level, marital status, and segmentation based on Demographics and behavior are important determinants of customer spending, with an average spending range of $0-200 per product.







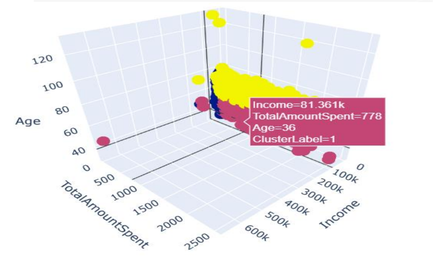
Models and Results

Clustering

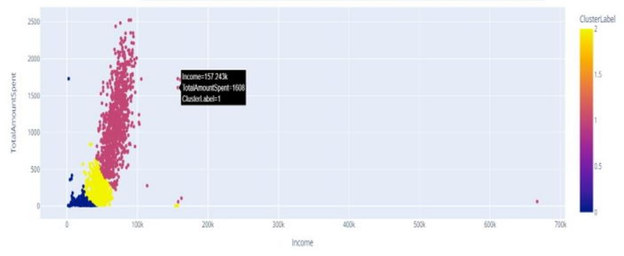
K-means clustering was used to cluster the data based on demographics and behavioral values.

Two K-means Models were built, one using two Features "Income" and "Total Amount Spent" and another using 3 features “Age”, "Income" and "Total Amount Spent”.

ClusterLabel	Age	Income	TotalAmountSpent
0	50.0	31801.0	54.0
1	45.0	67402.0	1001.0
2	66.0	62814.0	822.0



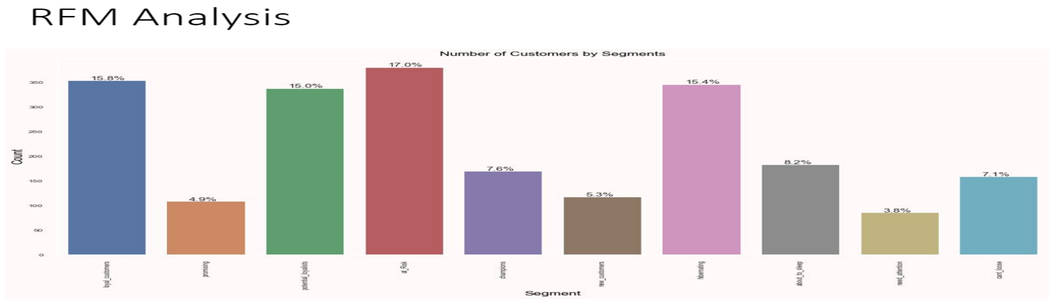
ClusterLabel	Income	TotalAmountSpent
0	25261.5	43.0
1	69084.0	1069.5
2	42641.0	145.0



Segmentation and Customer Churn

RFM analysis identifies customer segments based on Recency, Frequency, and Monetary value, with 9 distinct segments. Using logistic regression for churn prediction, we identified high-risk segments like "at_Risk" with a predicted churn rate of 60% .

By combining RFM analysis and logistic regression, we can design targeted retention strategies for high-risk segments and improve customer retention while reducing churn.



Predictive Marketing

Different machine learning models like logistic regression, SMOTE, XGBoost, and Random Forest have been applied to predict customers' response to marketing campaigns based on previous campaign response rates. Logistic regression has the highest accuracy and weighted F1 score, while random forest has the highest precision for positive class.

Regression Models	Accuracy	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)	Assessment
Logistic Regression (Unbalanced)	85%	0.50	0.16	0.24	Moderate ability to predict positive responses
Logistic Regression (SMOTE)	79.5%	0.34	0.48	0.40	Improved performance for class 1 instances
XGBoost	79%	0.33	0.45	0.38	Moderate ability to predict positive responses
Random Forest classifier	80%	0.34	0.45	0.38	Moderate ability to predict positive responses

Conclusion

By conducting exploratory data analysis and segmenting customers using K-means clustering, along with RFM analysis, businesses can identify valuable customers and prevent churn. Predictive models such as logistic regression, SMOTE, XGBoost, and Random Forest were trained to improve campaign response rates. These insights can help businesses optimize marketing and retention strategies, increase customer satisfaction, and improve profits