**Final Analysis Project Brief**

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

**Business problems:**

A company aims to enhance its approach to customer engagement and retention to maximize revenue and loyalty. To achieve this, the organization intends to utilize data analysis to segment customers, forecast their probability of leaving, and formulate customized retention tactics to cater to the unique needs and preferences of high-risk customers.

**Approach:**

To enhance customer segmentation and optimize marketing efforts, clustering techniques will be utilized to group customers based on relevant attributes, and previous campaign response rates will be analyzed. Additionally, machine learning algorithms will be employed to predict future campaign responses. To improve customer retention, a predictive model will be developed to identify high-risk customers by analyzing relevant attributes, and machine learning algorithms will be employed to classify customers as either high or low churn risk. Targeted retention strategies will be designed based on the results of the model and the unique requirements of high-risk customers.

**Results Of Our Work:**

* **Income** is one of the key indicator used to determine the amount a customer would spend.​
* Customers with graduate education and above tends to spend 12 times higher than other customers.
* Average spending range for each product is about $0-200.
* RFM analysis done using existing customer data based on Recency, Frequency, and Monetary values. Each factor is scored from 1-5, and a weighted average is calculated to derive a final RFM score by ID. This score is used to segment customers.
* The "at\_Risk" segment has a high risk of churn, with a predicted churn rate of 47%.​
* The "hibernating" segment has a medium risk of churn, with a predicted churn rate of 24%.​
* The "need\_attention" segment also has a medium risk of churn, with a predicted churn rate of 27%.
* The logistic regression model has the highest accuracy (0.85) and weighted F1 score (0.81) among all the models.
* If the focus is on identifying positive responses (i.e., customers who are likely to accept the offer), the random forest model has the highest precision for the positive class (0.34) among all the models.

**Business Impact:**

The advanced customer analytics techniques used in this project have the potential to significantly impact the business outcomes of the retail company. By segmenting customers based on their demographics, income, and spending habits using K-means clustering, the company can better understand its customer base and target its marketing efforts more effectively. Additionally, RFM analysis allows the company to identify its most valuable customers and those at risk of churning, enabling the company to design targeted retention strategies to keep high-risk customers engaged and loyal. Furthermore, the predictive models developed using logistic regression, SMOTE, XGBoost, and Random Forest techniques can help the company predict future campaign responses and optimize its marketing efforts, leading to increased customer satisfaction and improved bottom line.

Overall, the business impact of this project is significant, as it can help the retail company improve customer engagement, increase revenue, and build long-term customer loyalty.