

1. Team Member's Details:

- **Group Name:** Data Innovators
 - **Name 1:** Stephanie Dawsonn-Andoh
 - **Email:** stephandoh@gmail.com
 - **Country:** Ghana
 - **Specialization:** Data Analytics
 - **Name 2:** Sreedhar Rongala
 - **Email:** rongalasreedhar@gmail.com
 - **Country:** India
 - **Specialization:** Data Analytics
 - **Name 3:** Brittney Smith
 - **Email:** brittneysintership@gmail.com
 - **Country:** USA
 - **Specialization:** Data Analytics
-

2. Problem Description:

XYZ Bank is aiming to enhance its marketing campaign by delivering personalized Christmas offers to its customers. Instead of rolling out a generic offer for everyone, they want to target specific customer segments with relevant offers. To solve this problem efficiently, XYZ Bank approached ABC Analytics Company to help with customer segmentation. The bank's requirement is to group customers into no more than 5 segments, as more segments would be inefficient for their marketing efforts.

3. Data Understanding:

Dataset Context

1. Customer Demographics and Attributes

- The dataset includes a unique identifier for each customer (ncodpers), their country of residence (pais_residencia), gender (sexo), and age. This information helps understand the geographic distribution and demographic characteristics of the customer base.

2. Customer Relationship and Status

- It captures the employment status of customers (ind_empleado), which provides insight into their financial stability. The dataset also includes an indicator for new customers (ind_nuevo), the length of their relationship with the bank (antiguedad), and the type of relationship they have with the bank (indrel). Additionally, it tracks the customer's relationship status at the beginning of the month (indrel_1mes) and the nature of that relationship (tiprel_1mes).

3. Financial Products and Services

- This section outlines various financial indicators, such as whether the customer is active (ind_actividad_cliente), their gross income (renta), and whether they hold specific products like savings accounts, current accounts, mortgages, and credit cards. Each product has a corresponding indicator that shows if the customer has that type of account or service.

4. Temporal Aspects

- The dataset records key dates, including the date the customer became the primary holder of a contract (fecha_alta) and the last date they were considered a primary customer (ult_fec_cli_1t). These temporal aspects are essential for understanding customer engagement over time.
-

4. Overview of EDA performed:

1. The bank's customer base is heterogeneous in terms of recent activity. By understanding these differences and implementing targeted strategies, the bank can improve customer retention, increase customer loyalty, and drive revenue growth.
2. The right-skewed distribution suggests that the customer base is heterogeneous in terms of product usage. A large majority of customers are relatively low-product users, while a small minority of customers are high-product users.
3. A large majority of customers are relatively low-spending customers, while a small minority of customers are high-spending customers.
4. there is a higher number of women than men in the dataset.
5. in terms of correlation:
 - a. Age vs. Seniority: Older customers tend to have higher seniority, indicating a relationship between age and tenure.
 - b. Income vs. Monetary Value: Customers with higher incomes tend to have higher monetary values, suggesting a relationship between income and spending.
 - c. Frequency vs. Monetary Value: Customers who use more products or services tend to have higher monetary values, indicating a relationship between product usage and spending.
6. observations from RFM engineering
 - a. RECENCY VS FREQUENCY
 - i. Low-Activity, Low-Product Customers: These customers have recently used the bank's products but have not used many different ones. This group may require targeted marketing efforts to encourage them to use more products or services.
 - ii. High-Activity, High-Product Customers: These customers have been using the bank's products for a longer time and have used a variety of them. This group may be valuable to the bank due to their high spending and loyalty.
 - b. FREQUENCY VS MONETARY VALUE
 - i. Low-Usage, Low-Spending Customers: These customers use few products and spend little. This group may require targeted marketing efforts to encourage them to use more products or services.

- ii. High-Usage, High-Spending Customers: These customers use many products and spend a lot. This group may be valuable to the bank due to their high spending and loyalty.
- iii. Moderate-Usage, Moderate-Spending Customers: These customers have moderate product usage and spending. This group may represent an opportunity for up-selling and cross-selling to increase their spending.

Data storage location:

https://github.com/stephandoh/Data-Glacier-Internship/tree/main/Week_10