Customer Retention Case Study

Section 01: Business Objective understanding and High-level Approach

Lay down using a few slides - *refer Industry template.*

- **Overall Business objective** – put in simple words what's the scope of the analysis and what business problem is being solved *(this ensures that you've gone through the client's requirement).*
- **Understanding of the problem** in your own words *(this helps align your understanding of the business problem with client's requirement).*
- **Approach** – Very high-level view of the approach you plan to use to address the problem *(this helps the client see that you have a plan in place to attack the problem).*

Section 02: Data Health Review

Report the below results in a few slides *(ensure slide formatting and grammatical correctness).*

- Do the variables get read in Python in the right format (Integer, Float, Boolean, Date, Object)? ***List down what corrective steps were taken (if any) for the affected variables?***
- Do any variables have missing values? ***Create a list of variables for which you found missing values and what % values in these variables were affected?***
- Do any variables have outliers? ***Use suitable plots to show the outliers for all these affected variables.***
- Are there any variables that require cleaning (extra spaces, special characters, unexpected values, etc.) or replacement of values (e.g., Yes/No to 1/0)? ***List down all such variables and the kind of cleaning required.***
- Are there any duplicate records in the data? ***If yes, how many.***

1. ***Generate Extended Data Dictionary (EDD) of the provided dataset to help comment on data quality*** *– refer Industry template.*

Section 03: Exploratory Data Analysis

Perform Exploratory Data Analysis to comment on what information the dataset conveys and if it's complete/suitable to solve the given business problem. Also explore for any patterns/ insights that might guide in addressing the overall business problem. To be concrete, generate the below:

A. Observe Univariate distributions on both Object and Numeric variables.

a. Object variables – *Use suitable plots or visuals to show these distributions. Provide suitable commentary on what you observe for each variable.*

b. Numeric variables – *Use suitable plots to show these distributions. Provide suitable commentary on what you observe for each variable.*

B. Observe Bi-variate distributions

a. Scatter plots to show relationship between relevant Numeric variables, and Cross-tabulation for relevant categorical variables, etc.

Section 04: KPI/ Metric based questions – These questions have a specific ask (pin-pointed) and getting to the required outcome is quite straightforward.

1. What % of customers were active before the 6 months period. What was this % at the end of 6 months period?
2. Get the count of below customers:
   I. Customers who remained active
   II. Customers who remained inactive
   III. Active customers who became inactive
   IV. Inactive customers who became active

3. For the Four types of customers identified above, perform a comparison on the below metrics and report any significant difference
    I. Average of gross income
    II. Average of age
    III. Average of duration
    IV. Average of num_credit_cards
    V. Average of num_loans
4. For Female and Male customers, perform a similar comparison on the below metrics and report any significant difference:
    I. Average of gross income
    II. Average of age
    III. Average of duration
    IV. Average of num_credit_cards
    V. Average of num_loans
5. For customers of different segments (College, Individual,…), perform a similar comparison on the below metrics and report any significant difference:
    VI. Average of gross income
    VII. Average of age
    VIII. Average of duration
    IX. Average of num_credit_cards
    X. Average of num_loans
6. For each of the segments (College, Individual,…), find out the % of customers who:
    I. Who remained active
    II. Who remained inactive
    III. Who became inactive
    IV. Who became active

Section 05: Open-ended questions and recommendations – These are business-oriented questions which do not tell much about the kind of expected outcome, rather they require you to check if a certain phenomenon is occurring or not, or whether there is plausibility of a certain pattern. Often these questions need to be asked on your own and to answer them one needs to think through in terms of: 'what kind of output is expected', 'how to get it – which variables and by doing what' and 'whether the achieved outcome helps answer the question'. Since this can be iterative, it requires a lot of brainstorming and asking the right questions (as per the business objective).

1. In future which of the customer segments (college, individual, VIP) can be expected to remain active compared to others?
2. In future which of the customer segments (college, individual, VIP) can be expected to remain inactive compared to others?
3. In future, which of the customer segments (college, individual, VIP) can be expected to become inactive compared to others?
4. Which of Male/Female customers are more stable and do not change their activity/inactivity level much?
5. Which of Male/Female customers are more volatile and change their activity/inactivity level?
6. What impact does Income have on customer activity/ inactivity?
7. What impact does Age have on customer activity/ inactivity?
8. What impact does Duration has on customer activity/ inactivity?

End output expected is a PPT. Keep Python codes, analysis excel files, etc. as a back-up for the final presentation.

