Business Context:

Customer value analysis is critical for a good marketing and a customer relationship management strategy. An important component of this strategy is the customer retention rate. Customer retention rate has a strong impact on the customer lifetime value, and understanding the true value of a possible customer churn will help the company in its customer relationship management. Conventional statistical methods are very successful in predicting a customer churn.

What does “churn” mean?

“Churn” is a common phenomenon that occurs in telecom Industry. By “Churn” we mean those customers, who will be leaving us in near future.

If we are able to predict in advance, the attributes of customers whom we are going to lose in near future one can take corrective action so that we can minimize this phenomenon.

Predicting the churn also helps us to approximately know the life time value of customers. If a group of customers have a 20% chance of churning this month, then we would expect them to remain customers for 5 months (1 month ÷ 20%).

If the churn were reduced to 1%, then we would expect the customers to remain for 100 months. The other application is for prioritizing customer segments. If a segment is more likely to churn, perhaps, they should not get a high value gift. May be a discount might encourage them to stay. The issue may not be clear cut, but having a churn score will definitely help in better Decision making.

The goal of this study is to apply analytical techniques to predict a customer churn and analyse the churning and non-churning customers by using data from an internet connection company. Like most companies that supply goods and services over the internet, this company mainly deals with customers remotely. This can make it difficult to determine whether a customer is satisfied with the company or not. This, in turn, makes preventing churn a particularly challenging task. One tried and tested method of retaining customers is to offer them incentives to stay. However, if little is known about the behaviour of the customers, this can be a very imprecise science, leading to an incentive that is:

- Too little, or too much: The higher the value of the customer to the company, the higher the value the incentive should be to retain them.

- Too early, or too late: The customer might be perfectly happy with the company, in which case the incentive is an unnecessary expense, or they might have already taken their business elsewhere before they receive the incentive.

Ideally the company would have a clear and early indication of customers which are likely to churn, when and why, so they can focus their resources on just targeting them with the right offer at the right time.

Background:

Internet provider client is a provider of communications, high-speed Internet and entertainment services through broadband and fiber transport networks. With over 7,000 employees, this Client serves 2.2 million access lines in 25 states with annual revenue over $2.4 Billion. This client experienced record growth in High Speed Internet connection sign-ups with a 31% annual increase in subscriptions leading the US in growth. In order to effectively support this rapid growth, there would be significant investment in capital expenditure and labor costs. The cost concerns related to adding a large permanent labor force were challenging for this client because locating and retaining a large internal support workforce was difficult to predict and prohibitively expensive. Scaling Internet help desk operations to meet the support requirements of this growth proved to be a daunting task for this Client. Initially, the Client's operation was supporting all data products for both business and residential customers. As the result of this explosive growth, this client was unable to keep pace with the increased call volume that new subscribers were generating. This challenge was reflected in high abandon rates of 20%, and with calls being answered with an average speed of answer of 7.5 minutes. The lack of customer accessibility to the support team was an impediment to sustained growth in the competitive HSI market for the client.

Objective:

Client wants to know what the factors which contribute to churn are. They would like to have a system through which they could identify such customer which could also help them decide who the one who could not be retained are and the customer who could be retained what should be the appropriate incentive for them. They have goal to bring down the churn rate to the minimum.

Data Availability:

- Case_study_data.xlsx: This workbook has two tabs (Active customers and churn customers).
- Detailed data dictionary for both data sets have been provided as follows

DATA DICTIONARY FOR ACTIVE CUSTOMERS:

Fields/ Variables	Descriptions	Type	Var_length
Site_account_number	Account # (Unique key)	char	14
ACCOUNT_STATUS	Present Account Status	char	9
First_Communication_date	First date of communication	char	12
Upgrade_date	Date of Upgradation	char	12
SERVICE_PLAN	Plan Used	char	13
MODEL		char	7
SALES_CHANNEL	Channel through which the customer was acquired	char	15
DIRECT_INDIRECT_CHANNEL	Channel Sub Classification	char	9
SALES_SOURCE		char	51
COMPANY_SOURCE_NAME		char	13
FIRST_NAME	Account holder name	char	14
LAST_NAME		char	15
ADDRESS1	Account holder address	char	62
CITY		char	19
STATE		char	2
ZIPCODE		num	8
EMAIL_ADDRESS	Contact info	char	48
EMAIL_STATUS	Responds to emails/ feedbacks	char	19
SATELLITE		char	13
SITE_TYPE_DESC		char	16
WARRANTY_NAME		char	59
MOST_RECENT_SALES_CHANNEL		char	26
Gender	1 : Male 2 : Female	num	8
Date_of_Birth	DOB of Account holder	char	13
Age	Age of Account holder	num	8
Income	Income of Account holder (Refer Income table)	char	7
Marital_Status	0 - Single 1 - Married 2 - Divorced	char	7
Presence_of_children	0 - No children 1 - Atleast 1 child present	char	7
Computer_owner	Y - Has computer N - No computer	char	7

DATA DICTIONARY FOR CHURNED CUSTOMERS:

Fields/ Variables	Descriptions	Type	Var_length
Site_account_number	Account # (Unique key)	char	14
ACCOUNT_STATUS	Present Account Status	char	6
First_communication_date		num	8
CHURN_DATE		num	8
Upgrade_date	Date of Upgradation	char	12
SERVICE_PLAN	Plan Used	char	13
MODEL		char	7
SALES_CHANNEL	Channel through which the customer was acquired	char	15
DIRECT_INDIRECT_CHANNEL	Channel Sub Classification	char	9
SALES_SOURCE		char	51
COMPANY_SOURCE_NAME		char	13
FIRST_NAME	Account holder name	char	14
LAST_NAME		char	15
ADDRESS1	Account holder address	char	62
CITY		char	19
STATE		char	2
ZIPCODE		num	8
EMAIL_ADDRESS	Contact info	char	48
EMAIL_STATUS	Responds to emails/ feedbacks	char	19
SATELLITE		char	13
SITE_TYPE_DESC		char	16
WARRANTY_NAME		char	59
MOST_RECENT_SALES_CHANNEL		char	26
Gender	1 : Male 2 : Female	num	8
Date_of_Birth	DOB of Account holder	char	13
Age	Age of Account holder	num	8
Income	Income of Account holder (Refer Income table)	char	7
Marital_Status	0 - Single 1 - Married 2 - Divorced	char	7
Presence_of_children	0 - No children 1 - Atleast 1 child present	char	7
Computer_owner	Y - Has computer N - No computer	char	7

Income Table:

Descriptor	Description
0	Annual Income < $10K
1	Annual Income > $10K < 20K
2	Annual Income > $20K < 30K
3	Annual Income > $30K < 40K
4	Annual Income > $40K < 50K
5	Annual Income > $50K < 60K
6	Annual Income > $60K < 70K
7	Annual Income > $70K < 80K
8	Annual Income > $80K < 90K
9	Annual Income > $90K < 100K
10	Annual Income > $100K < 110K
A	Annual Income > $110K < 120K
B	Annual Income > $120K < 130K
C	Annual Income > $130K < 140K
D	Annual Income > $140K

EXPECTATIONS FROM THE TRAINEES:

1. Understand the data & perform the data preparation before perform all the analysis
2. Provide detailed insights/observations based on the analysis
3. If you build any statistical model,
    a. Understand the output from the software and explain the model fit.
    b. How would you determine what is the best model?
    c. Apply transformations to the given variables and find out the possible best model after transformations
    d. Generate the final equations
4. What are the key factors that predict customer churn? Do these factors make sense?
5. Data cleaning including missing values, outliers and multi-collinearity. Describe your predictive attrition model. How did you select variables to be included in the model?
6. Factor analysis for reduction of variables.
7. What offers should be made to which customers to encourage them to remain with company? Assume that your objective is to generate net positive cash flow, i.e., generate additional customer revenues after subtracting out the cost of the incentive.
8. Assuming these actions were implemented, how would you determine whether they had worked?
9. Provide the code with comments and generate outputs(results, plots and insights) in the format of word/pptx/html

USEFUL READINGS:

- ✓ Logistic regression (any classification technique)
- ✓ Clustering
- ✓ Nearest Neighbor Search
- ✓ Predictive Modeling
- ✓ Model Validation