Problem Statement – Graded Project

- Domain: Insurance
- Business Context - A XYZ life insurance company wants to predict their customers who are going to churn in next month, and want you to develop a model with proper EDA to get some recommendation. So accordingly they can design and provide offers to the customers.

- Dataset description:

Variable	Details
CustID	unique customer identifier
Mobile_num	mobile number of customer
Churn	Customer churn indicator
Age	Age of the customer
Payment_Period	Payment frequency of the customer in a year
Product	Type of Product
Cust_Tenure	Customer Tenure
EducationField	Highest education of the customer
Gender	Customer Gender
Overall_cust_satisfation_score	Overall customer satisfaction score
Cust_Designation	Designation of the customer in the current organization
CC_Satisfation_score	Satisfaction score of customer towards customer care service
Cust_MaritalStatus	Customer marital status
Cust_Income	Customer monthly income
Agent_Tenure	Tenure of the acquisition agent
Complaint	Weather customer raise a complaint
YTD_contact_cnt	Number of time company contact to the customer for Xsell the products
Due_date_day_cnt	Number of days left for due date
Existing_policy_count	Number of existing policy of the customer
Miss_due_date_cnt	Count of instance, when customer miss the due date of payment

- Create a predictive logistic model on with the below mentioned steps. Also give business insights and recommendation. So business end user can act on the recommendation.
- Learning Steps to perform:

1. Data cleaning (missing value and outlier treatment
2. EDA with proper insights
3. Split the prepared data into train and test
4. Develop model on train dataset
5. Calculate different model performance parameters on train
6. Validate the model on test dataset

Marks Distribution

- **Question 1:**

1.	Import dataset in the SAS environment and check top 10 record of import dataset (2 Mark)

- **Question 2:**

2.	Check variable type of the import dataset (2 Mark)

- **Question 3:**

3.	Checks if any variables have missing values, if yes then do treatment? (3 Mark)

- **Question 4:**

4.	Check summary and percentile distribution of all numerical variables for churners and non-churners? (5 Marks)

- **Question 5:**

5.	Check for outlier, if yes then do treatment? (3 Mark)

- **Question 6:**

6.	Check the proportion of all categorical variables and extract percentage contribution of each class in respective variables? (5 Marks)

- **Question 7:**

7.	Customer service management want you to create a macro where they will just put mobile number and they will get all the important information like Age, Education, Gender, Income and CustID (6 Marks)

- **Question 8:**

8.	Check correlation of all numerical variables before building model, because we cannot add correlated variables in model? (4 Marks)

- **Question 9:**

9.	Create train and test (70:30) dataset from the existing data set. Put seed 1234? (4 Marks)

- **Question 10:**

10.	Develop linear regression model first on the target variable to extract VIF information to check multicollinearity? (6 Marks)

- **Question 11:**

11.	Create clean logistic model on the target variables? (4 Marks)

- **Question 12:**

12.	Create a macro and take a KS approach to take a cut off on the calculated scores? (4 Marks)

- **Question 13:**

13.	Predict test dataset using created model? (2 Marks)