Life Insurance Sales Capstone

- NIMESH MARFATIA

Capstone Journey & Assessment:

Capstone Journey:

1st Mentor Connect – 21 Nov

PN1 Submission

2nd Mentor Connect – 5 Dec

PN2 Submission

3rd Mentor Connect – 19 Dec

Capstone Presentation

Final Report Submission

Assessment:

STAGES	GRADES
Project Notes – I	20
Project Notes – II	20
Capstone Presentation	20
Final report submission	40
Total Marks	100

Today's Agenda

1. Discuss the business context.
2. Understanding Data Dictionary and Data Set
3. Data cleaning and pre - processing (like outlier treatment, missing value treatment etc.)
4. How to generate insights from EDA?
5. Discuss about any finer nuances that could be used to generate insights.

Project Notes -1 Grading Parameters

Criteria – (Various Aspects of Solution & Comprehensive Understanding of the Problem)	Pts
1. Problem Understanding Defining problem statement, Need of the study/project, Understanding business/social opportunity	4
2. Data Report Understanding how data was collected in terms of time, frequency and methodology, Visual inspection of data (rows, columns, descriptive details), Understanding of attributes (variable info)	2
3. Exploratory Data Analysis Univariate, Bivariate Analysis, Missing Value, Outlier Treatment, Removal of unwanted variable, Variable Transformation, Adding new variables	10
4. Insights from EDA Data is Balanced, What to do?, Clustering or any other insights about data	4

Introduction – Life Insurance

- **Life Insurance:**

  Life insurance is a financial product that pays you or your dependents a sum of money either after a set period or upon your death as the case may be.

- **Agent:**

  An agent is a person who represents an insurance firm and sells insurance policies on its behalf.

- **Bonus:**

  Amount given to each agent for selling a policy

- Life Insurance Data (India):
- Total Life Insurance Companies: 24
- Premium Collected in FY 20: Rs. 7.31 trillion

Introduction – Life Insurance

Premium

- Life insurance premium is the amount of money you pay your life insurance company in exchange for your coverage.

Sum Assured

- Under any circumstances, such as a death, the Sum assured will be the amount that is paid by the insurance policy to the customer.

Policy Term

- Policy term refers to the period for which your insurance policy will remain active

Premium Payment Frequency

- Life insurance premium can either be a regular monthly/annual payment or a one-time payment as the case may be.

Type of Policy

- Term
- Endowment

Defining the Problem Statement

Problem Statement: Life Insurance Data

- The dataset belongs to a leading life insurance company.
- The company wants to predict the bonus for its agents so that it may design appropriate engagement activity for their high performing agents and upskill programs for low performing agents.

Basic Understanding about the Problem:

- Insurance Company doesn’t want to underpay or overpay
- Payout given for a policy is regulated by regulator (IRDA)

Why is Agent Bonus Important for Company?

1. Motivate Agent for More Sales
2. Increase Market Share of Company
3. Classify Agents into Different Bonus Categories
4. Helps in Designing New Products
5. Retain Agents
6. Increase Profits for the Company
7. Important from Regulation point of view

What Type of problem is this?

- **Regression Predictive Modelling**

- Regression is the problem of predicting a continuous quantity output.

- Regression predictive modeling is the task of approximating a mapping function (f) from input variables (X) to a continuous output (target) variable (y).

- In this case we are predicting Bonus to be paid to Agent for a policy sold by him.

- Target Variable (Y) = AgentBonus

Data Dictionary & Data Set

Variable	Description
CustID	Unique customer ID
AgentBonus	Bonus amount given to each agents in last month
Age	Age of customer
CustTenure	Tenure of customer in organization
Channel	Channel through which acquisition of customer is done
Occupation	Occupation of customer
EducationField	Field of education of customer
Gender	Gender of customer
ExistingProdType	Existing product type of customer
Designation	Designation of customer in their organization
NumberOfPolicy	Total number of existing policy of a customer
MaritalStatus	Marital status of customer
MonthlyIncome	Gross monthly income of customer
Complaint	Indicator of complaint registered in last one month by customer
ExistingPolicyTenure	Max tenure in all existing policies of customer
SumAssured	Max of sum assured in all existing policies of customer
Zone	Customer belongs to which zone in India. Like East, West, North and South
PaymentMethod	Frequency of payment selected by customer like Monthly, quarterly, half yearly and yearly
LastMonthCalls	Total calls attempted by company to a customer for cross sell
CustCareScore	Customer satisfaction score given by customer in previous service call

Problem Understanding

Defining problem statement

Need of the study/project

- Design appropriate engagement activity for their high performing agents
- Upskill programs for low performing agents.

Understanding business/social opportunity

- More Market Share
- More Profits
- Increase in awareness for Life Insurance
- More lives getting insured

Please Note:

- Cover each point, do not miss any point in the report.
- Write your observations / understandings on each point.

Data Report

- Understanding how data was collected in terms of time, Frequency and methodology
- Visual inspection of data (rows, columns, descriptive details)
- Number of Rows & Columns (Variables)
- Understanding of attributes (variable info, renaming if required)

Exploratory data analysis

Univariate analysis

- Describe the data and find patterns that exist within it.

Bivariate analysis (relationship between different variables , correlations)

- Find out if there is a relationship between two different variables

Removal of unwanted variables

- Drop the features you do not find related to the Target variable.

Missing Value treatment

- Numeric Variable – Mean / Median (if outliers are present)
- Categorical Variable – Most Common Class / Unknown Class
- If a feature has more than 15-20% missing value (15% is used as a norm, imputation of more than 15% is not recommended) and such features can be removed from the model building dataset.

Exploratory data analysis

Outlier treatment

- In this dataset there is a high probability Outlier can be actual data and required in the model.
- In practical times you can go back to business and double check on correctness of data.
- Lets take example of Sum Assured, very high chance to have an outlier here.
- Technically one can treat Outliers using Capping technique or removal of same.
- You can identify an outlier using box plot.

Variable transformation

- This is also known as Label Encoding
- In this dataset we can take example of any Categorical Variable having String Category, e.g. Edu Field
- Because you cannot perform most statistical operations on a string variable, you may want to turn the string variable into a numeric variable.

Exploratory data analysis

Addition of new variables:

- From the dataset given we can identify two important variables missing, Premium & Agent Id.
- In real world these data has to be available and you can get the same from business.
- However in case you wish to derive the same, you can try. It may or may not work.
- Premium is what a company gets and Bonus and Sum Assured is what it gives, there can be a factor by which this can be derived.
- It requires in depth knowledge and can be avoided. One needs more data to calculate Premium

Business insights from EDA

Is the data unbalanced? If so, what can be done? Please explain in the context of the business

- Checking data balance is very critical for Classification problems.
- In case of Regression, if a variable is significant then there should be adequate rows for all values of that variable.
- For e.g. If Zone is significant variable, there are only 6 records for South, or If AgentBonus is closely related to Occupation then very few rows are there for Small Business category.

Any business insights using clustering

- Clustering is generally done on Unsupervised data.
- If you wish to try you can try and build separate model for each cluster.

Any other business insights

- If you wish to add any other insights then you can add the same in this section.

Points to remember

- Understand and make clear the problem statement.
- Explain the data dictionary in detail.
- EDA should been done thoroughly.
- Do not paste only codes in your report.
- Focus should be on business report/notes.

Thank You