# Capstone Project
# Background and Objectives

## Insurance Analytics

# Background

## 1. Background

An insurance policy is an arrangement by which a company undertakes to provide a guarantee of compensation for specified loss, damage, illness, or death in return for the payment of a specified premium. A premium is a sum of money that the customer needs to pay regularly to an insurance company for this guarantee. Just like medical insurance, there is vehicle insurance where every year customer needs to pay a premium of certain amount to insurance provider company so that in case of unfortunate accident by the vehicle, the insurance provider company will provide a compensation (called 'sum assured') to the customer.

The marketing campaign was launched for existing health insurance customers to cross sell vehicle insurance. The data for the study includes campaign response details along with customer demographics
and policy details.


TA SCIENCE
INSTITUTE

# Background

2. Primary Objectives

1. Analyze “Customer Response" and understand the factors associated with it
2. Develop Campaign Response Model
3. Implement Machine Learning Algorithms and select the best method for Response Prediction

# Background

**3. Data**

**The following datasets are available:**

1. Customer Demographics
2. Policy Details
3. **Campaign Response**

# Data : Customer Demographics

**Content**

**This dataset contains customer demographics like gender, age, driving licencse**

| id | Gender | Age | Driving_License | Region_Code |
|---|---|---|---|---|
| 1 | Male | 44 | 1 | 28 |
| 2 | Male | 76 | 1 | 3 |
| 3 | Male | 47 | 1 | 28 |
| 4 | Male | 21 | 1 | 11 |
| 5 | Female | 29 | 1 | 41 |
| 6 | Female | 24 | 1 | 33 |
| 7 | Male | 23 | 1 | 11 |
| 8 | Female | 56 | 1 | 28 |

| Columns | Description | Type | Possible values |
|---|---|---|---|
| id | Customer ID | numeric | |
| Gender | Gender | Factor | Male and Female |
| Driving_License | Whether Customer has a driving license | Factor | 0: No 1:Yes |
| Region Code | Location | numeric | |

# Data : Policy Details

**Content**

**This dataset contains information of policies availed by a customer**

| id | Previously_Insured | Vehicle_Age | Vehicle_Damage | Annual_Premium | Policy_Sales_Channel | Vintage |
|---|---|---|---|---|---|---|
| 1 | 0 | > 2 Years | Yes | 40454 | 26 | 217 |
| 2 | 0 | 1-2 Year | No | 33536 | 26 | 183 |
| 3 | 0 | > 2 Years | Yes | 38294 | 26 | 27 |
| 4 | 1 | < 1 Year | No | 28619 | 152 | 203 |
| 5 | 1 | < 1 Year | No | 27496 | 152 | 39 |
| 6 | 0 | < 1 Year | Yes | 2630 | 160 | 176 |

| Columns | Description | Type | Possible values |
|---|---|---|---|
| id | Customer ID | numeric | |
| Previously Insured | Is customer insured with the company in the past | Factor | 0: No, 1: Yes |
| Vehicle Age | Vehicle Age | Factor | <1 year, 1-2 years and > 2 years |
| Vehicle Damage | Is vehicle damaged | Factor | Yes and No |
| Annual Premium | Annual premium cost | numeric | |
| Policy_Sales_Channel | Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc. | numeric | |
| Vintage | No. of days customer has been associated with the company | numeric | |

# Data : Campaign Response

**Content**

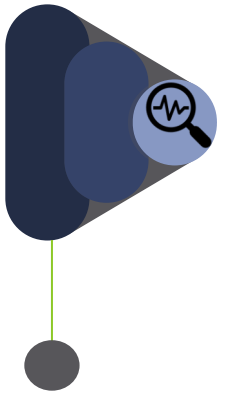
**This dataset contains** customer response

| id | Response |
|---|---|
| 1 | 1 |
| 2 | 0 |
| 3 | 1 |
| 4 | 0 |
| 5 | 0 |
| 6 | 0 |
| 7 | 0 |
| 8 | 1 |

| Columns | Description | Type | Possible values |
|---|---|---|---|
| id | Customer ID | numeric | |
| Response | Whether customer is interested in buying the policy | Factor | 0: no, 1: yes |

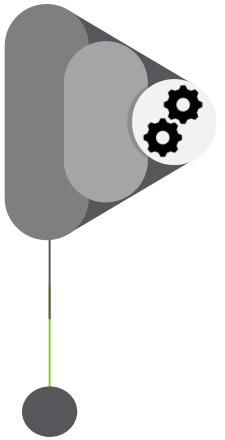
# Next steps

## Data management



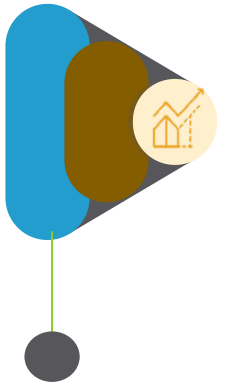
- Compile 3 datasets using Customer ID
- Data cleaning , Handling missing values and completing Basic Data checks
- Check if any variables needed to be feature coded i.e made into groups or want to be left as continuous variables

## Descriptive Statistics & Data visualization



- Summarize response rate for various subgroups in the data such as gender, age group, vintage, etc
- Explore data for response rate, are any services playing a role in response, etc.
- How can this data be presented better visually ?
- Once again post Data visualization check if any variable needs to be feature coded

## Predictive modelling



- Develop a model to predict customer response
- Using different Predictive model techniques to find Significant variables
- Ensure you follow all steps like Train and test data , checking for Multicollinearity
- Check if any other ML technique fits better

DATA SCIENCE INSTITUTE

* You can go through the Questions and Hint ppt to understand each section better / cross check if you have done all the steps mentioned