|  |
| --- |
| **Business Case** |
| State the business problem (do not copy paste from the question. Instead write based on your understanding). Explain from business perspective why they are interested in this problem |

Business Problem:

The Insurance company is facing losses due to fraud claims in the fields of Vehicle Insurance, Property damage insurance and Life Insurance. The benefits they offer are being misused to claim insurance money and its difficult to track the behaviour of the frauds being committed.

Why??

* To reduce financial loss to the company due to fraud claims.
* To make sure the benefits they offer are not miss utilised.
* To identify traits of a fraud being committed.

|  |
| --- |
| **ML problem statement** |
| Describe how the business problem is converted into a Machine Learning problem and how it will help the business |

The solution to the business problem can be forged with the help of Machine Learning.

* The company has the data regarding the Customer, Policy, Claims, Demography of the respective claims and the vehicle details.
* They are interested to identify traits and patterns of fraud committed to claim insurance money in the fields of Vehicle damage, Property Damage and Health Insurance.

|  |
| --- |
| **Data Exploration** |
| Describe what data you have and give a snapshot of rows and columns and describe some important parameters.  If you have merged some datafiles, briefly talk about it but do not exaggerate on merging procedure.  Explain your data and exploratory efforts with visualization (follow the instructions on visualization) |

* Generalized ML models can be built on the data provided by the company to identify certain behaviours and instances where the customer is attempting a fraud claim, hence able to avoid them.

There are 5 files each consisting different data regarding the problem statement

1. Claim: This table consists of details regarding the insurance claims which include the Amount of claims, the type of incidents and related details.

Important Parameters: Incident Address, Type of Incident, Severity of Incident, Claim Amounts, Incident State.

|  |  |  |  |
| --- | --- | --- | --- |
| Claim | 28836 rows | 19 columns | Continuous and Categorical data |

1. Demography: This table consists of details regarding the customer like education, gender, age and others.

Important Parameters: Insured Hobbies, Insured Education, Insured Occupation.

|  |  |  |  |
| --- | --- | --- | --- |
| Demography | 28836 rows | 10 columns | Continuous and Categorical data |

1. Policy: This table consists of details regarding the Insurance Policy which includes the Policy number, Policy Annual Premium and other details.

Important Parameters: Umbrella Limit, Insured Relationship, Policy Combined Single Limit.

|  |  |  |  |
| --- | --- | --- | --- |
| Policy | 28836 rows | 10 columns | Continuous and Categorical data |

1. Vehicle: This table consists of details regarding the vehicle, the model, chassis number, year of make and others. Later the rows are split into respective columns.

Important Parameters: Vehicle Model, Vehicle YOM.

|  |  |  |  |
| --- | --- | --- | --- |
| Vehicle | 115344 rows | 3columns | Categorical data |
| Vehicle | 28836 rows | 4columns | Categorical data |

1. Train File: This table consists of two columns. The Customer ID and Reported Fraud column.

Merging:

I have merged the 5 tables together on Customer ID forming a single table.