# Investigating the loss of Insurance Renewal

**Project Description:**

The objective of this data analysis in “Investigating loss of Insurance Renewal” is to clearly bring out the losses that are incurred by insurers based on various issues like policy cancellation, theft of vehicles accidental cases etc.

**Project Scope:**

The scope of project covers all the common phases of data analysis, which includes data collection, cleaning, implementation and visualization. To proceed with investigation, the following datasets are to be gathered: vehicle thefts, policy cancellation and subsequent claim cases. The analysis can help the insurer to decide, the area to focus on and to show where they are gaining or losing customers based on policy renewals. With this data, the insurer can ascertain why loss is caused and why the insurance is not renewed.

**To define the collection of datasets for the mentioned strategies include:**

1. **Theft cases**:

Dataset should contain the following details: Make and model of the vehicle, Year of manufacture, state of registration and count of thefts cases.

1. **Policy Cancellation cases**:

Dataset should contain details like insurance coverage type, premium paid, vehicle’s age, NCB availed, age of the insured etc.

1. **Subsequent claim cases**:

Dataset should contain the details like age of insured, gender, number of children, annual premium amount, total number of claims availed.

**Implement the most common data analysis approach like:**

1. **Data Collection**: Raw data to be collected through various websites for the defined strategies like Theft cases, Policy Cancellation and Subsequent claim cases.
2. **Data Processing & Cleaning**: To check for invalid data and to correct the same. Convert as an excel sheet or create as data table to proceed further steps.
3. **Model and Algorithm Implementation**: Adopt supervised learning technique using K means (with multiple clustering) and use decision tree technique to predict the losses and also implement linear regression representation for all columns in datasets.
4. **Data Visualization and reporting the analysis results**: Task involves abstract the data sources, collect the data, then filter those data and represent the data to end users graphically as charts and graphs. The report shall include pie charts, scatter plots and bar charts such that the data can give a trend for visual check.