**Members of this Assignment: Kalani Anastasia, Kourtesi Ioanna**

We are a business intelligence services provider that has been appointed by an Insurance Company in order to build a model that will classify whether or not we will suggest an insurance contract to a potential customer. The aim of our project is that, the company will increase its profit by taking the right decision.

In current assignment we are utilizing a dataset of 48,620 observations. The 80% of the data was randomly selected for the training data set – 38,896 observations and remaining 20% of the data was set aside for the testing – 9,724 observations.

We started by making some pre-processing steps, such as data cleaning. After that, we used machine learning techniques such as Logistic Regression and at last, we developed deep learning architectures such as the Multilayer Perceptron and the Recurrent Neural Network.

# Python code files – General Description

This file describes the code files reading sequence.

* 1. **prepare\_data.py**. It reads the csv files and creates the datasets that are necessary for running the modes.
  2. **ROC\_Cover4.py**. It creates ROC for Chewing Damage. Also, it reports other indices like F1-Score, etc.
  3. **Mlp.py** Code for mlp testing for Chewing Damage
  4. **Logistic.py**. Code for Logistic Regression.
  5. **Rnn.py** Code for Rnn for Chewing Damage.
  6. **graphs.py**. Examples of how to create values necessary for the “Data Features” of the documentation.
     1. **Instructions for creating charts in excel:** Initially, we have to copy and past each variable in an excel. Then, in a new column we will put the distinct values of this variables. After that, we will go to the Data tab in the ribbon and click on the Analysis group and click on the Data Analysis and we will get the graphs.
  7. **percentiles\_Cover4.py**. It creates the graph of the ordered segmentation of Average probability vs the Actual sales for Chewing Damage.
  8. **Model\_tests\_cover4.py** . Contains all the efforts we have made in modeling.

Csv we used:

1. **accidents\_data.csv and occupation\_tariff\_codes.csv** : Datasets
2. **graph test.csv**: for the creation of confusion matrices and Chewing Damage Threshold Results graph for both logistic Regression and Mlp.