Health Insurance Lead Prediction

Your Client FinMan is a financial services company that provides various financial services like loan, investment funds, insurance etc. to its customers. FinMan wishes to cross-sell health insurance to the existing customers who may or may not hold insurance policies with the company. The company recommend health insurance to it's customers based on their profile once these customers land on the website. Customers might browse the recommended health insurance policy and consequently fill up a form to apply. When these customers fill-up the form, their Response towards the policy is considered positive and they are classified as a lead.

Once these leads are acquired, the sales advisors approach them to convert and thus the company can sell proposed health insurance to these leads in a more efficient manner.

Now the company needs your help in building a model to predict whether the person will be interested in their proposed Health plan/policy given the information about:

* Demographics (city, age, region etc.)
* Information regarding holding policies of the customer
* Recommended Policy Information

Sample Submission

This file contains the exact submission format for the predictions. Please submit CSV file only.

| Variable | Definition |  
| ID | Unique Identifier for a row |  
| Response | (Target) Probability of Customer showing interest (class 1) |

Credits:

This is dataset is released as part of a hackathon conducted by Analytics Vidhya.  
Visit <https://datahack.analyticsvidhya.com/contest/job-a-thon/#ProblemStatement> for more information.

Data Description:-

We have the following information regarding the potential-customer and the insurance at any given point in time:

* Demographics (city, age, region etc.)
* Information regarding holding policies of the customer
* Recommended Policy Information

Expected Outcome:-

* Build a model to predict whether the person will be interested in their proposed Health plan/policy given the information above.
* Grading Metric: **ROC\_AUC\_SCORE**

Problem Category:-

For the data and objective its is evident that this is a **Binary Classification Problem** in the **Tabular Data** format.

So without further ado, let's now start with some basic imports to take us through this journey of Lead prediction:-

From the column keys in problem statement we know the following information about each of the features:-

| Variable | Definition |
| --- | --- |
| ID Unique | Identifier for a row |
| City\_Code | Code for the City of the customers |
| Region\_Code | Code for the Region of the customers |
| Accomodation\_Type | Customer Owns or Rents the house |
| Reco\_Insurance\_Type | Joint or Individual type for the recommended insurance |
| Upper\_Age | Maximum age of the customer |
| Lower\_Age | Minimum age of the customer |
| Is\_Spouse | If the customers are married to each other (in case of joint insurance) |
| Health\_Indicator | Encoded values for health of the customer |
| Holding\_Policy\_Duration | Duration (in years) of holding policy (a policy that customer has already subscribed to with the company) |
| Holding\_Policy\_Type | Type of holding policy |
| Reco\_Policy\_Cat | Encoded value for recommended health insurance |
| Reco\_Policy\_Premium Annual | Premium (INR) for the recommended health insurance |
| Response (Target) | 0 : Customer did not show interest in the recommended policy, |
|  | 1 : Customer showed interest in the recommended policy |