**Case Study**

We sell automobile insurance policies where customers pay us an annual premium to insure their vehicle for a year. Assume that each customer only has one vehicle, and one vehicle only has one driver (i.e., the customer). Also assume that if they get into an accident, they will always open a claim with us.

## Objective

The task is to create a model that predicts whether a customer will get into an accident and submit a claim.

## Input

In the attached dataset, you will see the following fields:

* *Claim ID: customer unique ID*
* *Gender: gender of customer*
* *Age: age of the customer at the time of policy submission*
* *Driving license: valid driver's license at the time of policy submission (1 = has a driving license, 0 = no driving license)*
* *Region code: region code of the customer’s address*
* *Previously insured: customer is previously insured with Northbridge (1 = yes, previously insured with Northbridge, 0 = no, not previously insured with Northbridge)*
* *Vehicle age: age of the customer’s vehicle at the time of policy submission*
* *Previous\_Vehicle\_Damage: customer’s vehicle has previous damage (1 = has previous damage, 0 = no previous damage)*
* *Annual premium: annual premium that the customer pays for their automobile policy*
* *Policy\_Sales\_Channel: encoded sales channel that customer use to purchase their policy*
* *Response: whether customer submitted a claim (1 = submitted claim, 0 = did not submit claim)*

Assume your audience will be a peer who will be bouncing ideas with you to improve this model. It is expected that your code will be readable and sharable within the team during model development, but not necessarily production ready. During the upcoming interview, you may be asked about this assignment and your considerations.

*Note: the data provided is based on a sample of dummy data, so it is not expected that the candidate will produce an accurate result but more to* understand the candidate’s thinking and considerations when creating a model.

## Submission

1. Summarize the model’s performance and prepare a report explaining the steps taken to process the data, the model approach taken, the rationale behind model selection, and any limitations of the analysis.
2. Submit your script (such as a Jupyter Notebook, Python script, R script) with appropriate documentation. If you had to make any assumptions, do state it.
3. Also, prepare a one-pager summary that is intended for the presentation to the business stakeholders and include key findings, recommendations, or any model insights.