**Project Description**

Machine learning is very popular in the insurance market, which has long been renowned for being data-driven.

Working on behalf of On the Road car insurance, you'll investigate their customer data to identify the feature that produces the most accurate logistic regression model for them to predict whether a claim will be made against a policy!

**Project Instructions**

* Identify the single feature of the data that is the best predictor of whether a customer will put in a claim (the "outcome" column), excluding the "id" column.
* Store as a DataFrame called best\_feature\_df, containing columns named "best\_feature" and "best\_accuracy" with the name of the feature with the highest accuracy, and the respective accuracy score.

Insurance companies invest a lot of [**time and money**](https://www.datacamp.com/workspace/external-link?url=https%3A%2F%2Fwww.accenture.com%2F_acnmedia%2Fpdf-84%2Faccenture-machine-leaning-insurance.pdf) into optimizing their pricing and accurately estimating the likelihood that customers will make a claim. In many countries insurance it is a legal requirement to have car insurance in order to drive a vehicle on public roads, so the market is very large!

Knowing all of this, On the Road car insurance have requested your services in building a model to predict whether a customer will make a claim on their insurance during the policy period. As they have very little expertise and infrastructure for deploying and monitoring machine learning models, they've asked you to identify the single feature that results in the best performing model, as measured by accuracy, so they can start with a simple model in production.

They have supplied you with their customer data as a csv file called car\_insurance.csv, along with a table detailing the column names and descriptions below.

**The dataset**

| **Column** | **Description** |
| --- | --- |
| id | Unique client identifier |
| age | Client's age:   * 0: 16-15 * 1: 26-39 * 2: 40-64 * 3: 65+ |
| gender | Client's gender:   * 0: Female * 1: Male |
| driving\_experience | Years the client has been driving:   * 0: 0-9 * 1: 10-19 * 2: 20-29 * 3: 30+ |
| education | Client's level of education:   * 0: No education * 1: High school * 2: University |
| income | Client's income level:   * 0: Poverty * 1: Working class * 2: Middle class * 3: Upper class |
| credit\_score | Client's credit score (between zero and one) |
| vehicle\_ownership | Client's vehicle ownership status:   * 0: Does not own their vehilce (paying off finance) * 1: Owns their vehicle |
| vehcile\_year | Year of vehicle registration:   * 0: Before 2015 * 1: 2015 or later |
| married | Client's marital status:   * 0: Not married * 1: Married |
| children | Client's number of children |
| postal\_code | Client's postal code |
| annual\_mileage | Number of miles driven by the client each year |
| vehicle\_type | Type of car:   * 0: Sedan * 1: Sports car |
| speeding\_violations | Total number of speeding violations received by the client |
| duis | Number of times the client has been caught driving under the influence of alcohol |
| past\_accidents | Total number of previous accidents the client has been involved in |
| outcome | Whether the client made a claim on their car insurance (response variable):   * 0: No claim * 1: Made a claim |