Data dictionary

Column	Description
id_policy	Unique policy identifier
year	This dataset covers a 5 year range. This indicates the year of the contract `1-5`.
pol_no_claims_discount	This is a compulsory system in most European countries and globally. The coefficient is attached to the driver and indicates previous claims. It starts at `0.631` for new drivers (i.e. first year of insurance). Then, every year without claim, the value decreases by *approximately* `0.05` until it reaches its minimum of `0`. Without any claim, the value evolution would then be: `0.63`, `0.58`, `0.53`, `0.48`, ... , `0.03`, `0`. Every time the driver causes a claim (only certain types of claims are taken into account), the coefficient increases by *approximately* `0.2`, with a maximum of `1`. Thus, the range of `pol_no_claims_discount` extends from `0` to `1` in the dataset.
pol_coverage	There are four types of coverage: `Min`, `Med1`, `Med2` and, `Max`, in this order. `Min` policies cover only Third Party Liability claims, whereas `Max` policies covers all claims, including Damage, Theft, Windshield Breaking, Assistance, etc. The two `Med` policies represent intermediate coverage.
pol_duration	Policy duration represents how old the policy is. It is expressed in years, accounted from the beginning of the current year. There may be very old policies in the dataset for loyal customers.
pol_sit_duration	Represents how old the current policy characteristics are. This is different from `pol_duration`, because the same insurance policy could have evolved in the past (e.g. coverage, or vehicle, or drivers, ...). The `pol_sit_duration` is a metric for how up-to-date the characteristics are.
pol_pay_frequency	The price of the insurance coverage can be paid annually, bi-annually, quarterly or monthly. Be aware that **you must provide a yearly quotation** in your final submission.
pol_payd	PAYD or "Pay As You Drive" is a boolean (i.e. a string with `Yes` or `No`), which indicates whether the client has subscribed a mileage-based policy or not.
pol_usage	This indicates how the driver uses the vehicle with 4 possible values: `WorkPrivate`, `Retired`, `Professional` (which denotes a professional usage of the vehicle), and `AllTrips` which is quite similar to `Professional` (including professional tours).
drv_sex1	European rules force insurers to charge the same price for women and men. But gender can still be used in academic studies, and that's why `drv_sex1` is still available in the datasets. For our purposes, let's assume that this can be used as variable in this pricing game. In this dataset the values are `M` or `F`.
drv_age1	This is the age of the first driver in years.
drv_age_lic1	The age of the first driver's driving license in years.
drv_drv2	This boolean (`Yes`/`No`) identifies the presence of a secondary driver on the contract. There is always a first driver, for whom certain characteristics are provided, but a secondary driver is optional.
drv_sex2	The gender of the second driver if present in the data.
drv_age2	When `drv_drv2` is Yes, then the secondary driver's age is present. When no second driver is registered, this is `0`.
drv_age_lic2	The age of the second driver's driving license in years, if present.
vh_make_model	A hash representing the make and model of the vehicle. These hashes (random strings) represent text such as `Ford Focus` or `Honda Civic`.
vh_age	This variable is the vehicle's age, the difference between the year of production of the vehicle and the current year. One can consider values of `1` or `2` to correspond to new vehicles.
vh_fuel	`Diesel`, `Gasoline` or `Hybrid`. This is the fuel type of the vehicle.
vh_type	`Tourism` or `Commercial`. You'll find more `Commercial` types for `Professional` `policy_usage` than for `WorkPrivate`.
vh_speed	This is the maximum speed of the vehicle, as stated by the manufacturer.
vh_value	The vehicle's value (replacement value) is expressed in Euros.
vh_weight	The weight (in kg) of the vehicle.
population	The population of the town where the policy holder is registered multiplied by a constant scale.
town_surface_area	Approximate surface area (in hectors multiplied by a constant scale.) of the town where the policy holder is registered.
claim_amount	The amount of a claim expressed in euros for the year (a value of zero indicates that no claim was made).