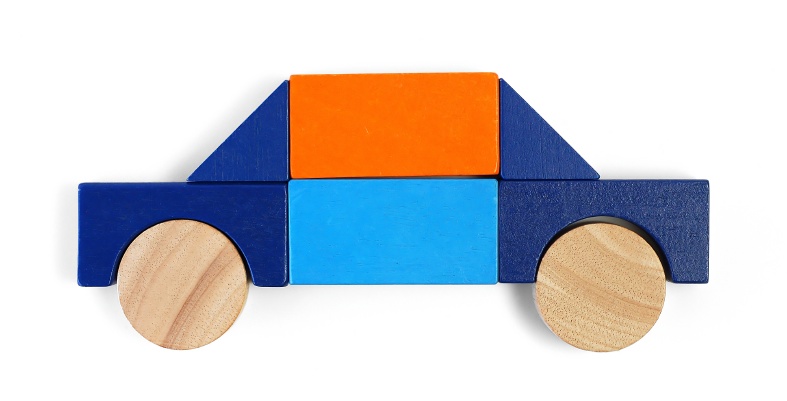
Predict a purchased policy based on transaction history



As a customer shops an insurance policy, he/she will receive a number of quotes with different coverage options before purchasing a plan. This is represented in this challenge as a series of rows that include a customer ID, information about the customer, information about the quoted policy, and the cost. Your task is to predict the purchased coverage options using a limited subset of the total interaction history. If the eventual purchase can be predicted sooner in the shopping window, the quoting process is shortened and the issuer is less likely to lose the customer's business.

Using a customer’s shopping history, can you predict what policy they will end up choosing and the price for that policy?

### Files

The training and test sets contain transaction history for customers that ended up purchasing a policy. For each customer\_ID, you are given their quote history. In the training set you have the entire quote history, the last row of which contains the coverage options they purchased. The last rows are also collected in the file train\_short.csv. In the test\_session\_history.csv set, you have only a partial history of the quotes and do not have the purchased coverage options. These are truncated to certain lengths to simulate making predictions with less history (2 sessions for each customer).

For each customer\_ID in the test set, you must predict the coverage option they end up purchasing and the price of the product.

* **train.csv:** quote history with multiple rows per customer Id. The last row shows the coverage option they ended up purchasing.
* **train\_short.csv**: this data set is a subset of train.csv. It contains only the last rows for each customer ID. The rows that show which product is purchased by the customer. This file is created for the ones who do not wish to use the entire quote history while building their models.
* **test\_session\_history.csv:** shows the partial history of the quotes (2 sessions per customer) and do not have the purchased coverage options or the price. This data is optional and you do not have to use it for your predictions.
* **test.csv:** there is one row per customer ID. You will notice that the last two columns in this data set are blank: **policy and price.** You are required to estimate these two values in the dataset for each customer ID. Policy can have values from 1, 2, 3 and 4, whereas price can have any integer values.

### What is a customer?

Each customer has many shopping points, where a shopping point is defined by a customer with certain characteristics viewing a product and its associated cost at a particular time.

* Some customer characteristics may change over time (e.g. as the customer changes or provides new information), and the cost depends on both the product and the customer characteristics.
* A customer may represent a collection of people, as policies can cover more than one person.
* A customer may purchase a product that was not viewed!

### Product Options

There are 4 options with 4 ordinal values possible: **1, 2, 3** and **4**.

The cost of a product is a function of both the product options and customer characteristics.

### Variable Descriptions

**customer\_ID** - A unique identifier for the customer  
**shopping\_pt** - Unique identifier for the shopping point of a given customer  
**record\_type** - 0=shopping point, 1=purchase point  
**day** - Day of the week (0-6, 0=Monday)  
**time** - Time of day (HH:MM)  
**state** - State where shopping point occurred  
**location** - Location ID where shopping point occurred  
**group\_size** - How many people will be covered under the policy (1, 2, 3 or 4)  
**homeowner** - Whether the customer owns a home or not (0=no, 1=yes)  
**car\_age** - Age of the customer’s car  
**car\_value** - How valuable was the customer’s car when new  
**risk\_factor** - An ordinal assessment of how risky the customer is (1, 2, 3, 4)  
**age\_oldest** - Age of the oldest person in customer's group  
**age\_youngest**- Age of the youngest person in customer’s group  
**married\_couple** - Does the customer group contain a married couple (0=no, 1=yes)  
**C\_previous** - What the customer formerly had or currently has for product option C (0=nothing, 1, 2, 3,4)  
**duration\_previous** -  how long (in years) the customer was covered by their previous issuer  
**policy** the coverage options  
**cost** - cost of the quoted coverage options