

2022 Travelers University Modeling Competition



1) Business problem

You work for Peace of Mind Insurance Company in their personal automotive insurance department as a modeler. For personal auto, anyone that asks for an estimated price for a policy from your company (also known as a *quote*) will receive one. However, of those quoted, only a fraction will choose your company as their insurer (versus other companies they also received quotes from). Your team has been asked to build what's known as a *conversion model*, with the goal of understanding the population of policies Peace of Mind Insurance Company is most likely to *write* (a.k.a. *issue* or *convert*). In other words, what types of customers is your company writing over its competitors?

Your goals in this competition are as follows:

- Identify quoted policies that your company will convert (a.k.a. issue)
- Understand key characteristics of policies your company tends to write, as well as those they tend not to write (e.g. understand quoted policies with both high and low conversion rates)
- Provide a recommendation on how this information could be leveraged at Peace of Mind Insurance

While there are no constraints on the type of model you build, the colleagues that have asked your team to get this information are not modelers and have no statistical background. It'll be important to remember this when explaining your findings and recommendations.

Keywords

Policy – account associated with a given customer.

Quote – Price given to a potential customer for an insurance policy.

Issue/write/convert – If your company 'writes/issues/converts' a policy, it means the customer chose your company as their insurer.

Conversion rates – the percent of quoted policies your company issues

2) Data Description

The data provided to you and your team consists of variables describing customers that asked for quotes. There are two datasets – one at the policy level one at the driver level. In other words, each row of the policy dataset corresponds to a single policy, which could have multiple drivers associated with it. The driver level dataset has one row per driver. You can assume that drivers found in the driver dataset will belong to only one policy in the policy dataset. You can also assume that all members on a policy live at the same address.

Your company was only able to convert a fraction of the policies found in this sample. The policy dataset also has a training and test split variable called *split*. Note that the conversion indicator (the response variable) is missing for policies in the test split. Your task is to build a model on the training data and apply your model to predict the conversion indicator for each policy in test data.

Policies.csv

<i>policy_id</i>	Unique customer identifier
<i>quote_dt</i>	Date the quote was submitted
<i>quoted_amt</i>	Quote amount (US dollars)
<i>prior_carrier_grp</i>	Prior carrier group
<i>cov_package_type</i>	Level of coverage needed
<i>discount</i>	Whether or not a discount was applied to the quote amount
<i>number_drivers</i>	Number of drivers
<i>credit_score</i>	Credit score of primary policy holder

<i>num_loaned_veh</i>	Number of vehicles on policy that have a loan associated with them
<i>num_owned_veh</i>	Number of owned vehicles on the policy
<i>num_leased_veh</i>	Number of leased vehicles on the policy
<i>total_number_veh</i>	Total number vehicles on the policy
<i>primary_parking</i>	Where car(s) are primarily parked
<i>CAT_zone</i>	Catastrophe risk zone
<i>home_policy_ind</i>	Whether or not customer has existing home insurance policy with Peace of Mind
<i>zip</i>	US zip code of policy holder
<i>state_id</i>	State of policy holder
<i>county_name</i>	County of policy holder
<i>agent_cd</i>	Unique agent code (8 digits)
<i>split</i>	Train/Test split
<i>convert_ind:</i>	Conversion indicator (0=no, 1=yes). This is the response variable

Drivers.csv:

<i>policy_id</i>	Unique customer identifier
<i>gender</i>	Gender of driver
<i>age</i>	Age of driver
<i>high_education_ind</i>	Higher education indicator
<i>safety_rating:</i>	Safety rating index of driver
<i>living_status:</i>	Driver's living status (levels = 'own', 'rent', 'neither')

Vehicles.csv:

<i>policy_id</i>	Unique customer identifier
<i>car_no</i>	Unique car identifier (per policy)
<i>ownership_type</i>	Whether the car is loaned, owned or leased
<i>color</i>	Vehicle color
<i>age</i>	Vehicle age
<i>make_model</i>	Make and model of the vehicle

3) Modeling

- Each group can make at most 3 submissions per day:
- Work together within group on data analysis, but not between groups
- You can build the model with any software
- Work with your group members and ask for clarification from discussion section
- You are only permitted to use the data provided – third party data sources are not to be used
- Recommended team size is 2-5 members. Maximum team size is 5 and minimum is 1 (individual contributor is allowed.)

4) Benchmark Model

Score from a simple LightGBM model will be the competition benchmark.

5) Model Evaluation

Models will be evaluated by AUC.

The teams scoring better than the benchmark will move on to the second stage, the virtual live presentations. Each qualifying team will give a 5-7 minute presentation on the above questions followed by a 3 minute Q&A session. The winning campus team will join other winning teams for a virtual job shadow day at the Travelers Hartford campus and make final presentations to a panel who will determine the ultimate winner!

6) Timeline

Project Kickoff	Friday Oct 21
Final model submission & Group presentations	Friday Nov 4
Campus Winners announced	Week of Nov 7
Winners' Job Shadow Day at Travelers	TBD

7) Presentation Instructions

- What methods did you consider (you don't have to try all these methods; just ones that you think would work for this problem)?
- What method did you choose in the end, and why?
- How did you do your variable selection?
- What other variables not in the data set do you think might be useful?

8) Contacts

Wenye Qiu: WQIU@travelers.com

Elliot Yang: EYANG3@travelers.com

Haining Zhang: HZHANG2@travelers.com

Disclaimer

Peace of Mind Insurance Company and the data is a fictitious example used for the purpose of this competition only.