Overview

In this challenge, you will act as an insurance company, where you build a pricing model and compete against other players (other insurance companies) for profit. In other words, the player that maximises competitive profit is the winner.

The market in this challenge will be a `cheapest-wins market`. That means every insurance company offers every customer an annual premium price, and the customer will always pick the company that offers the cheapest price to them (e.g., using a price comparison website).

In order to create your pricing model, you are given historical insurance data: 60K real historical car insurance policies for 4 consecutive years. Each policy concerns 1 vehicle, its drivers and an accident history over 4 years. This data has been provided by a large car insurance provider in a European country and is a uniform sample for their entire portfolio.

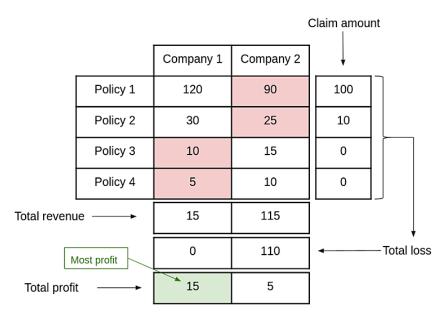You are asked to produce a model to price contracts for incoming policies for the 5th year.

The company that makes the most profit in this market, wins the challenge.

🎭 Cheapest-wins market

As a player or team you represent one insurance company. You will have to provide a premium quote for every policy (or customer) you encounter. So does every other company. But the customer will pick the cheapest price offered to them. This is illustrated bellow:

	Company 1	Company 2
Policy 1	120	90 ← Cheapest offer for Policy 1
Policy 2	30	25
Policy 3	10	15
Policy 4	5	10
Total revenue →	15	115

Now that companies 1 and 2 each have a set amount of revenue, they have to pay out the cost of the claims accociated with policies 1-4. So it will look like:

Claim amount

	Company 1	Company 2	
Policy 1	120	90	100
Policy 2	30	25	10
Policy 3	10	15	0
Policy 4	5	10	0
Total revenue →	15	115	
	0	110	← Total loss
Total profit →	15	5	

Most profit

So, once the claims are taken into account in the cheapest-wins market, we can see that Company 1 wins as it has the most competitive profit.

🏆 Leaderboards

As an insurance company in this market, you will be responsible for a portfolio of policies that pay you annual premiums. In exchange, you cover their risk. If they make a claim, you will have to pay.

Therefore, to make money, you must:

Estimate the expected loss for your portfolio, so your contracts are profitable

Come up with a pricing strategy that allows you to compete with others, so that you win contracts

To reach these goals we provide you with two leaderboards.

ROOT MEAN SQUARED ERROR (RMSE) LEADERBOARD

This leaderboard always displays your _best_ RMSE submission.

It uses your `predict_expected_claim` function and measures how well you can estimate the risk of policy IDs. It functions by measuring the root mean squared error (RMSE) of your premiums compared to the cost of the claim.

The optimal way to minimise RMSE is done by a model that predicts the expected claim for each contract most accurately.

This allows you to compare the quality of your loss prediction model with your competitors however, there is no explicit reward for performing well on this leaderboard.

This leaderboard is refreshed instantaneously upon submissions, and will always use the same data, disjoint from the other leaderboard.

Please note: your RMSE score is computed in 4 stages. When you submit a model, your model makes predictions for:

Year 1 with access to data from year 1.

Year 2 with access to data from years 1 - 2.

Year 3 with access to data from years 1 - 3.

Year 4 with access to data from years 1 - 4.

Predictions from steps 1 - 4 are then used in the standard RMSE formula to compute your final RMSE score.

In this way you are to use past data to inform the present. For example predictions for year 3 can be informed by what happened in years 1 - 2.

COMPETITIVE PROFIT LEADERBOARD (UPDATED EVERY SATURDAY AT 10PM CET)

By default this leaderboard uses your most recent successful submission, but you can choose another submission as well thought this form that is also displayed at the top of the leaderboards.

It leaderboard uses your `predict_premium` function. It measures your average competitive profit in a market of size 10 when playing against other players.

Each week uses a new set of data to ensure that you don't price the same policy many times, like a real market.

This measures `competitive profit`. That means your profit is averaged over many markets that you play in.

To make sure that results are stable, we keep putting you in markets until your leaderboard rank no longer changes from market to market.

Note: This leaderboard is updated every _**Saturday at 10PM CET**_ with your most recent submission.

If you do consistently well on this leaderboard, you will likely do well in the final evaluation.

⚖️ Evaluation metric

The final evaluation metric for this challenge is `competitive profit`: how much money does your company make in a _realistic market_. However we also provide a leaderboard using root mean squared error.

ROOT MEAN SQUARED ERROR (RMSE)

Given a set of list of claims y and a set of expected claims $\hat{y}$ predicted by your model, the RMSE is computed as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

COMPETITIVE PROFIT (THE FINAL METRIC)

The evaluation process is as follows:

Compute average profit rank. First the average competitive profit that your model makes in a market of size 10 (i.e. with 9 other random players) is computed. This gives your model a profit rank.

Compute realistic `competitive profit`**.** In a _realistic market_ models that don't perform well don't exist (i.e. go bankrupt). So to compute the realistic `competitive profit`, we place your model in a market of size 10 with 9 other models picked from the **top 10%** of the the ranking obtained in step 1.

Two important notes:

- The profit rank in step 1 is not used in the leaderboard. Only the ranking in step 2 is used in the leaderboard.

- Rankings from both steps are generated as a result of many many runs of different random markets. On average you can expect your model to have competed against every other model present at least once.

🚗 Market rules

There are two rules that your submissions to the profit leaderboard, and final submissions must follow:

Non-negative training profit. Your models must be profitable on the training data. That is, the sum of your premiums must not be less than the sum of the claims.

Participation rule. Your model must participate (i.e. win at least 1 policy) in *5% or more* of the markets it is placed in.

📊 Weekly Market Feedback

In a real insurance market, every time you participate, you will get some feedback. In this game, each week thousands of markets are run! and you will get feedback about your performance in those markets.
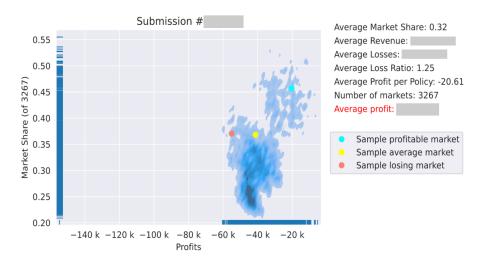
You get two types of feedback:

A plot and some KPIs

Summary statstics about policies you have won

Below you can see examples of both of these. For more details on how they are computed please see here.

EXAMPLE FEEDBACK PLOT (SEE HERE FOR DETAILS)

ONE EXAMPLE OF THE SIX FEEDBACK TABLES (SEE HERE FOR DETAILS)

Driver Summaries

Policies won	drv_sex1	drv_age1	drv_age_lic1	drv_drv2	drv_sex2	drv_age2	drv_age_lic2
often: 40.1 - 100.0%	M (55%)	55.75	33.17	No (73%)	F (55%)	49.57	27.95
sometimes: 13.4 - 40.1%	M (61%)	54.60	32.47	No (68%)	F (65%)	46.93	25.00
rarely: 0.0 - 13.4%	M (62%)	54.80	32.76	No (63%)	F (64%)	44.96	23.28
in sample profitable market	M (59%)	55.45	33.08	No (70%)	F (62%)	49.80	27.31
in sample average market	M (61%)	53.37	31.19	No (68%)	F (62%)	48.53	26.09
in sample losing market	M (59%)	53.22	30.96	No (68%)	F (61%)	48.47	26.13

💾 Dataset

You can download the dataset from the resources tab.

The dataset contains a total of 100K real historical car insurance policies over 5 years in the recent past.

This has been provided by a large car insurance provider in a European country and is a uniform sample from their entire portfolio.

You can find the data dictionary under the resources tab.

The majority of the data concerns third-party liability but there are also other types of car insurance (e.g. theft) present.

For this challenge, the data is split in the following way:

TRAINING DATA

This is 60K policies with 4 years of history (~240K rows). It can be downloaded from the resources tab.

RMSE LEADERBOARD

This contains 5K policies with 4 years of history (~20K rows).

9 WEEKLY PROFIT LEADERBOARDS

This contains 15K policies with 4 years of history (~60K rows) split evenly between 9 disjoint, weekly profit leaderboards. In addition, the data is split such that each week has more recent data. For example:

- Week 1 contains mostly year 1 data
- Week 9 contains mostly year 4 data

TEST DATA

The final test dataset, where the final evaluation takes place, includes 100K policies for the 5th year (100K rows). To simulate a real insurance company, your training data will contain the history for some of these policies, while others will be entirely new to you.
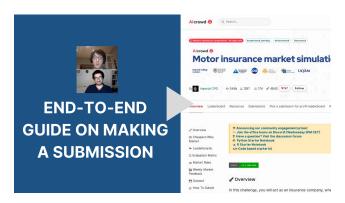
✉️ How to submit

There are two methods of submission available in this challenge:

Submission through Google Colaboratory notebooks.

Submission from a `.zip` file.

SUBMISSION VIA COLABORATORY NOTEBOOKS

Python. Visit the starter notebook here.

R. Visit the starter notebook here. Here is an end-to-end walkthrough for a simple R model.

The notebooks are self-contained and submissions are made through the notebooks themselves.

SUBMISSION VIA A ZIP FILE

Visit the starter-kit available here and follow the instructions.

Once you have prepared your `.zip` file, you can then submit it using the link in the top right corner of the page here.

Note: baseline models are automatically excluded from the profit leaderboards.

🏆 **Prizes**

	Amount	Prize sponsor
1st	$6000 USD	PartnerRe
2nd	$3500 USD	SINGAPORE ACTUARIAL SOCIETY
3rd	$1500 USD	actuartech
4th	$1000 USD	Institut des ACTUAIRES

ADDITIONAL COMMUNITY ENGAGEMENT PRIZES

In addition to the 4 cash prizes for performance in the challenge. We also have community engagement prizes which are:

- **A shiny mug**
- **A great T-shirt**

For more information on how to win these please see there.

📅 Timeline

Launch: 18 December 2020 - 5:00pm UTC

Deadline: 07 March 2021 - 11:59pm UTC

Team formation deadline: 31 January 2021 - 11:59pm UTC

🌐 Research sponsors

This research is supported in part, by the following institutions.

📞 Contact

- Ali Farzanehfar
- Florimond Houssiau