Quant Competition 2023
Second Edition

Summary of the contest

⭐ Second Edition, academic contest open to **multiple Master programs** in Romania (1st and 2nd year students), in the area of Data Science / Machine Learning / Econometrics, including but not limited to:

- ➢ ASE Cibernetica – all masters
- ➢ Universitatea Bucuresti – **Master Data Science**

  Master Artificial Intelligence

  Master Probabilitati si Statistica in Finante
- ➢ Univ Babes-Bolyai Cluj – **Master Econometrie si Statistica Aplicata**
- ➢ Univ Babes-Bolyai Cluj – **Master Data Science for Industry and Society**
- ➢ Univ Babes-Bolyai Cluj – **Master Analiza Datelor Complexe**
- ➢ Univ A.I. Cuza Iasi – **Master Data Mining (Cibernetica si Statistica)**
- ➢ Universitatea de Vest Timisoara – **Master Big Data (data science, analytics and technologies)**

🎖️ Students are given a **real-world professional assignment** and will be assessed jointly by their Professor and by a jury of Deloitte & BT professionals.

🐷 The outcome of the program for students include **financial awards** for top 3 participants (below the net amounts), as well as multiple invitations for collaboration (**hiring**) in the Risk Advisory team at Deloitte and at Banca Transilvania:
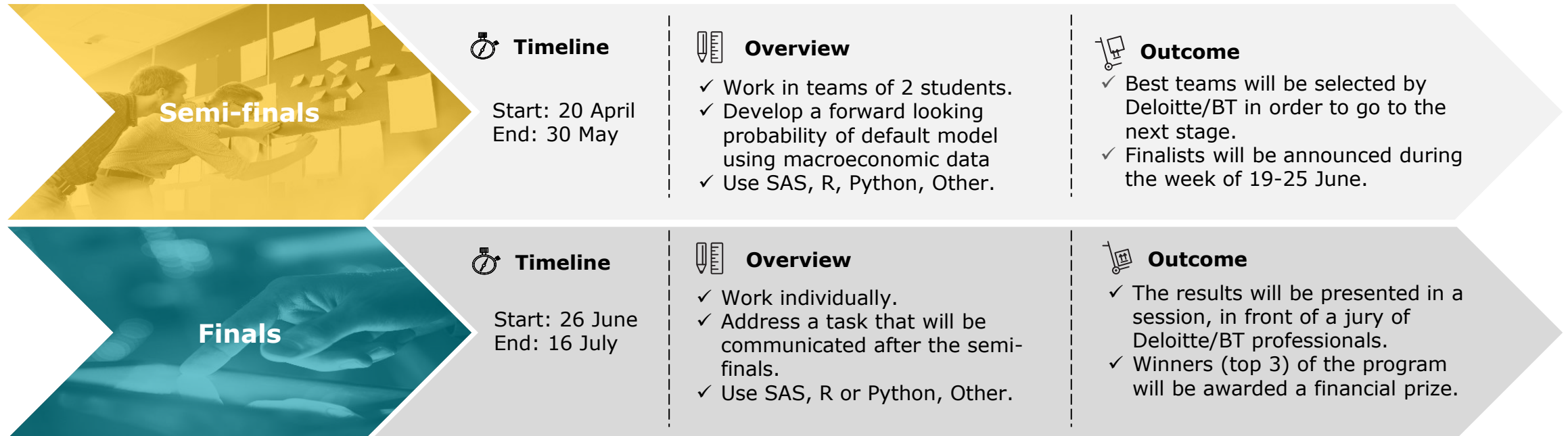
🏆 1st place: **€2000**

🏆 2nd place: **€1500**

🏆 3rd place: **€1000**

Timeline

PROBABILITY OF DEFAULT MODELLING USING MACROECONOMIC FACTORS

Semi-finals

Timeline

Start: 20 April
End: 30 May

Overview

- ✓ Work in teams of 2 students.
- ✓ Develop a forward looking probability of default model using macroeconomic data
- ✓ Use SAS, R, Python, Other.

Outcome

- ✓ Best teams will be selected by Deloitte/BT in order to go to the next stage.
- ✓ Finalists will be announced during the week of 19-25 June.

Finals

Timeline

Start: 26 June
End: 16 July

Overview

- ✓ Work individually.
- ✓ Address a task that will be communicated after the semi-finals.
- ✓ Use SAS, R or Python, Other.

Outcome

- ✓ The results will be presented in a session, in front of a jury of Deloitte/BT professionals.
- ✓ Winners (top 3) of the program will be awarded a financial prize.

Assignment - Overview and Dependent variable calculation

- Context: Time series models are often used in the estimation of forward-looking component of probability of default models used in quantification of losses according to the IFRS 9 standard. Their role is very important in capturing the effect of macroeconomic factors and other exogenous risk drivers that may affect the level of credit risk of a loan portfolio. In order to build such a model a bank usually calculates the observed PD or the so-called observed default rate (DR). The outcome of the DR is 12 months, this is equivalent to a PD for the next 12 months.

- The assignment of this contest will challenge you to make a model that is suitable to quantify the forward-looking component, you will start with the calculation of the target/dependent variable and potentially end with producing the output of such a model, the forecasts.

- Database: information at client level. It will be provided prior to the start date, via a Deloitte Shared Server. This will be the source of calculation for the dependent risk driver.

- Macroeconomic info: The independent risk drivers should be of macroeconomic nature, it is public information. The participants are expected to gather these risk drivers from the internet (ex: Eurostat, INSSE, BNR, World Bank etc.)

- Main assignment ("**the semi-finals**"):

  ➢ The model can be developed in SAS/R/Python/Eviews/SPSS/Excel/etc. The exercise will take place in teams of 2 students and has two parts:

  1) Calculate the dependent variable to be modeled from the Database: observed default rate with outcome of 12 months.

     - At a given reporting date the observed default rate is calculated as: number of defaulted clients in the next 12 months divided by the total number of performing clients at reporting date. For example, if there are 150 performing (non-default) clients in December 2021 (e.g. a reporting date) out of which 30 have at least one default event during Jan-Dec2022, then the observed default rate at December 2021 is 30/150=20%.

Assignment – Time series model and Forecasts

2) Develop a time-series statistical model for the estimation of the default rate using macroeconomic risk drivers. The dependent variable is the observed default rate from point 1). You can identify and download drivers from INSSE, BNR, Eurostat, etc.

- Potential models: linear regression, autoregressive (AR), autoregressive moving average (ARMA), autoregressive with distributed lags (ARDL), models to address seasonality endogenously (e.g. SARIMAX), random forest, neural networks, RNN, LSTM, etc.

  o Assess model performance through specific validation tests: p-values, R-squared, relevant business sign of macro drivers (e.g. negative GDP sign), best linear unbiased estimator BLUE tests (no collinearity, stationarity of drivers, homoscedasticity, no autocorrelation, normality of errors, linear functional form). If BLUE tests fail, we expect an explanation for the failure, it is not necessary that all BLUE tests are passed. For non-linear models, apply BLUE tests where appropriate and discuss, and also apply relevant techniques for model explainability.

  o Bonus points: (1) generate a second "alternative" model using a different modeling technique (e.g. if the main model is a linear regression, an alternative model can be AR), (2) model with more than 2 relevant macroeconomic drivers, (3) obtaining at least a 70% R-squared or equivalent with all validation tests accounted for on linear regression (4) usage of transformed risk drivers (e.g. LAGs, differences, logs, etc. and determining their optimal combination in the model); (5) conduct scenario analysis to study the effect of shocks on predicted default rates, determine an appropriate size of the shock and provide economic interpretations;

  o Prepare model documentation (minimum 5 slides in PowerPoint).

  NOTE: You do not need to apply all modelling techniques or use all types of drivers. You are free to choose one or more from the requirements above. Apply the relevant statistical tests, show your work and ensure it is reproducible.

➢ Additional assignment ("**the finals**"): will be communicated to qualified teams. The work in the finals will be done individually by each student.

Evaluation

The analyses will be evaluated by a jury consisting of Deloitte and Banca Transilvania professionals.

Different criteria will be evaluated:

Model

- Correctness of the observed default rate calculation
- Relevant value for R square, > 40%.
- Validation tests outcome
- Correctness of the forecasts calculated with the model

➢ Objective measure

25%

Quant bonus points

- 5%: Alternative model
- 10%: More than 2 independent risk drivers
- 10%: Getting very good R2 and validation outcome on linear regression
- 10%: Usage of transformed independent risk drivers
- 10%: Effect of shocks

➢ Semi-objective measure

45%

Quality of deliverables

1. Presentation: comprehensive presentation of steps and conclusions (PPT and analytical results);
2. Technical implementation: clean, functional and efficient codes that a third party can rerun to obtain the same results.

➢ Subjective measure

30%

Registration

- For registrations, please send to ppetroiu@deloittece.com one e-mail per team with the table below filled for the 2 team members:

	Team member 1	Team member 2
First name		
Last name		
Email address		
University		
Master		

- Registrations will be open during the interval 1 – 19 April.
- The database and model assignment will be made available through a dedicated server

Submission of assignment results

- Team and individual assignment materials will be delivered to the jury through the SharePoint.
- For the first stage (the semi-finals), each team of 2 students will receive a Sharepoint of the team, where all results will need to be uploaded before the deadline. Each team will have access only to its allocated sharepoint.
- All results need to be submitted in English and will need to include:
  - Programming codes used to generate the results and that a member of the Jury could rerun to check consistency.
  - Analytical files (i.e. excel files) presenting each step of the modeling project.
  - Documentation (i.e. PPT of minimum 5 slides) summarizing model results.