3rd edition

The **financial challenge** of the year

NOVARTISDATATHON

online

In collaboration with

eurecat
Centre Tecnològic de Catalunya

NOVARTIS

Pharmaceutical mission

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

"Reimagine medicine to improve and extend people's lives"

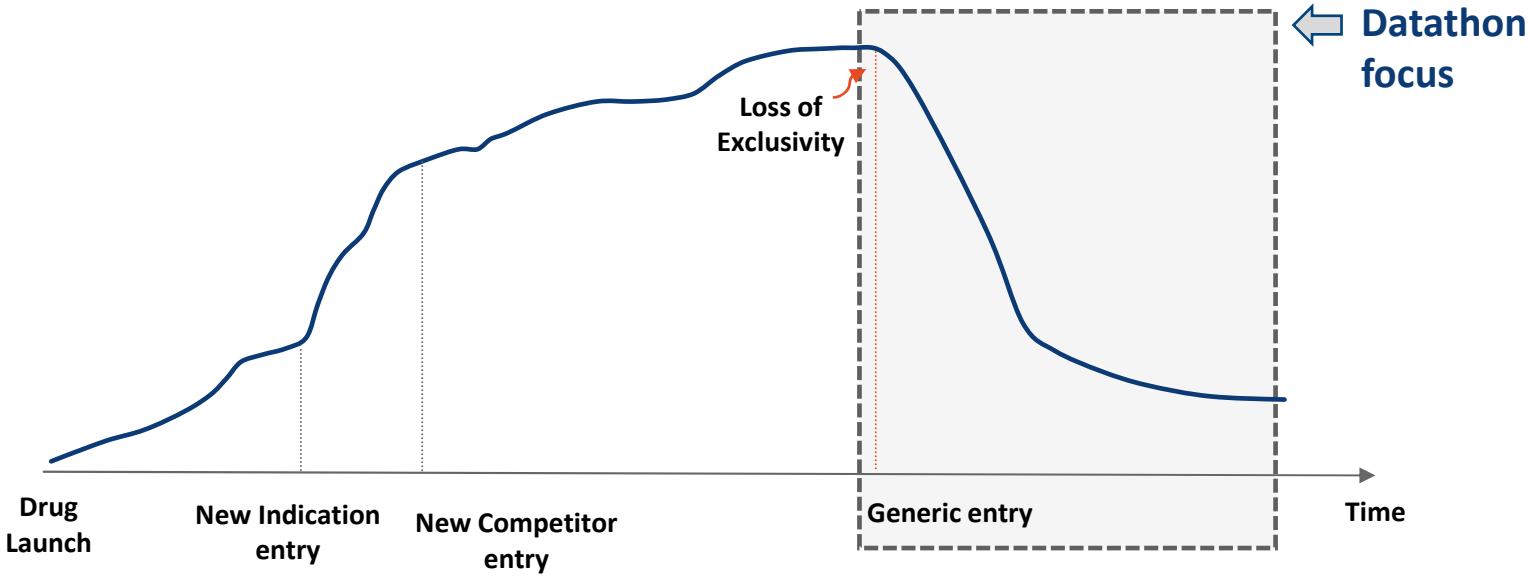
- Novartis is a medicine-focused company powered by advanced therapy platforms and data science.
- R&D of new products of our IM Division can take approximately 10 to 15 years from discovery to commercial product launch. The development process must undergo highly complex, lengthy and expensive approval processes.
- Loss of exclusivity allows generic companies to distribute the original compound.

Financial efficiency allows us to develop more and better drugs, while reaching more people

NOVARTIS

Lifecycle of a drug

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



NOVARTIS

Datathon challenge

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

1. Data Science challenge

Participants are required to provide **24 months of volume forecast** after the generic entry date together with a **95% confidence intervals** for the given prediction for all the brands in the test set.

2. Business challenge

All teams that present in front of the Jury will be asked to provide a **deep exploratory analysis** on the correlation **between features** provided and the **impact in the volume sold** after the gx entry. We encourage the participants to use visualization tools.

NOVARTIS

Datathon Criteria

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

The winner selection of the Datathon will be in 3 phases:

- **Phase 1 (Accuracy)**: There will be an objective metric to calculate the accuracy of the given predictions. This metric will be used to select the **top 10 teams** with the lowest error in the volume forecast.

- **Phase 2 (Certainty)**: Once the top 10 teams with highest forecast accuracy are selected, a second objective metric will be used to mesure the certainty of the given confidence intervals. The **5 teams** with less error in this metric will be selected to present in front of the Jury.

- **Phase 3 (Jury's criteria)**: There will be 5 teams presenting the results in front of the Jury. The members of the Jury will consist of people from both technical and business background and they can ask questions about any of the 2 challenges. Once the 5 presentations are finished, the Jury will decide which are the 3 winners of the Datathon.

NOVARTIS

Challenge: Data Provided

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

- **Target:**
  - **Volume** *: Historical (pre-gx) volume for **1078 country-brands** that went generic in the past.
    - **Train: 887 observations** for which in addition are provided 24 months of volume after the gx entry date.
    - **Test: 191 observations** for which a forecast needs to be provided for month_num = 0 (month of gx entry) to month_num = 23 after the gx entry date.
- **Features:**
  - **Therapeutic Area**
  - **Package**
  - **Panel (Channel of Distribution)**
  - **Number of gx**

*Volume can be in different units depending on the country and brand (miligrams, packs, pills, etc.)

NOVARTIS

Data Examples: Volume

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

country	brand	volume	month_num	month_name
country_1	brand_3	12911629	-4	Jul
country_1	brand_3	11470630	-3	Aug
country_1	brand_3	11876792	-2	Sep
country_1	brand_3	12056281	-1	Oct
country_1	brand_3	7695814	0	Nov
country_1	brand_3	7975224	1	Dec
country_1	brand_3	5796841	2	Jan
country_1	brand_3	4895233	3	Feb
country_1	brand_3	6053584	4	Mar

NOVARTIS

Data Examples: Features

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

brand	therapeutic_area
brand_1	Nervous_system
brand_2	Respiratory_and_Immuno_inflammatory
brand_3	Cardiovascular_Metabolic
brand_4	Cardiovascular_Metabolic
brand_5	Cardiovascular_Metabolic
brand_6	Cardiovascular_Metabolic

country	brand	presentation
country_1	brand_3	PILL
country_1	brand_4	PILL
country_1	brand_10	PILL
country_1	brand_14	PILL
country_1	brand_18	CREAM
country_1	brand_20	INJECTION

country	brand	channel	channel_rate
country_1	brand_3	B	1.18970413
country_1	brand_3	D	98.81029587
country_1	brand_4	B	0.09022942
country_1	brand_4	D	99.90977058
country_1	brand_10	B	1.01569734
country_1	brand_10	D	98.98430266

country	brand	num_generics
country_1	brand_3	3
country_1	brand_4	1
country_1	brand_10	6
country_1	brand_14	1
country_1	brand_18	1
country_1	brand_20	2

NOVARTIS

Data Insights and Hints

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

- All historical **volume** provided is at **monthly level** and since the beginning of the brand (or since the first available datapoint).
- The **gx entry date** corresponds to **month 0**. Positive months represent months after the gx entry date and negative months represent months prior to the gx entry date.
- You can train with **all** the data provided, even if you have data after the gx entry date for some test examples.
- Volume can be in different units (miligrams, packs, pills, etc.) for the different country-brands.
- Assume that categorical variables **do not** change over time.

NOVARTIS

Metric: Prediction Error

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

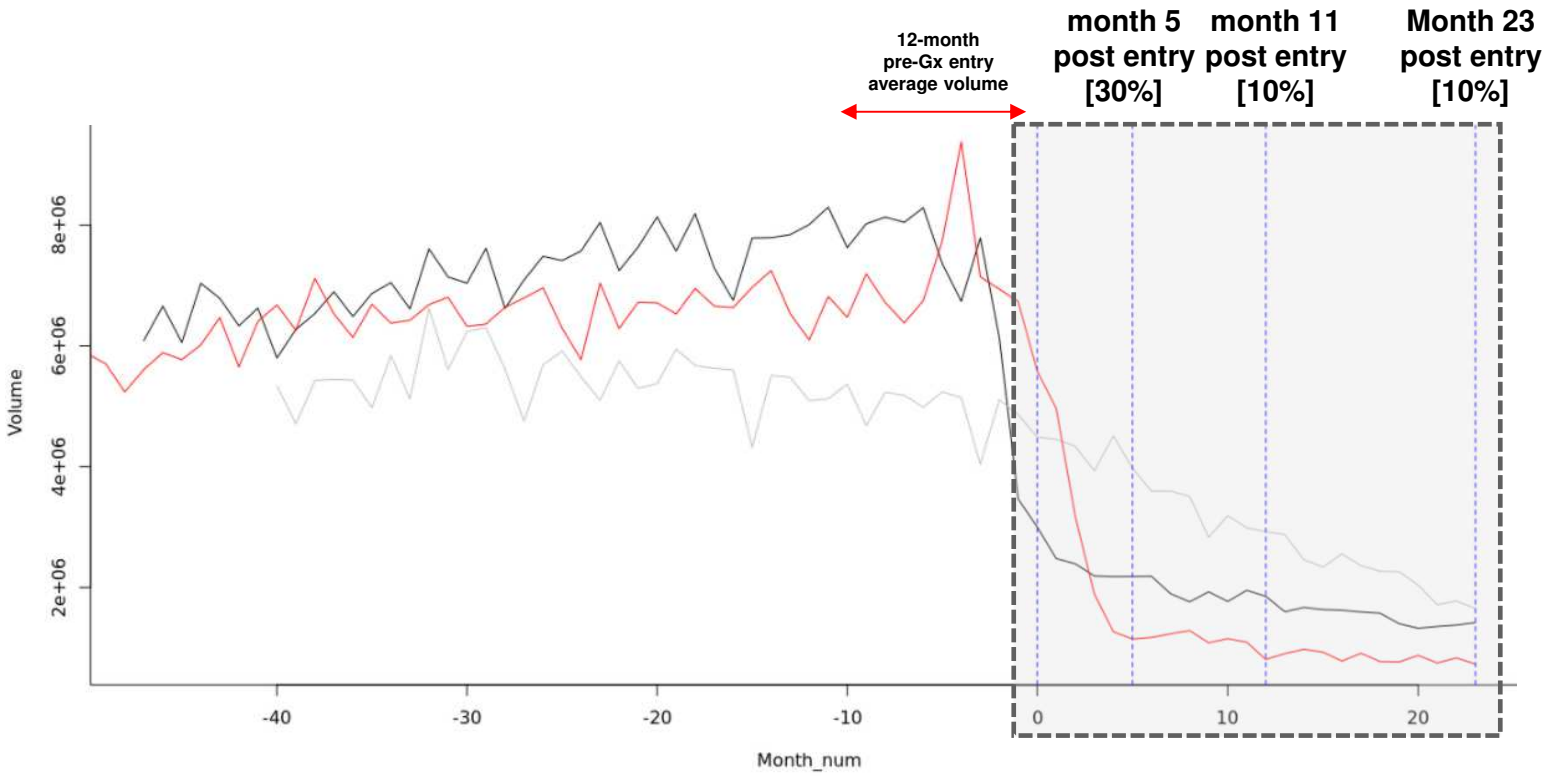
To compute the prediction error we will evaluate the difference between the predicted values vs the actual volume in four different ways weighted as follows:

1. Absolute **monthly** error of all 24 months (50%)
2. Absolute **accumulated** error of months 0 to 5 (30%)
3. Absolute **accumulated** error of months 6 to 11 (10%)
4. Absolute **accumulated** error of months 12 to 23 (10%)

All the 4 items will be normalized by the average monthly volume of the last 12 months before the generic entry in order to take into account the magnitude of the brand.

NOVARTIS

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



Absolute Monthly Error [50%]

NOVARTIS

Metric: Prediction Error

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Formula:

$$PE_j = 0.5 \cdot \left(\frac{\sum_{i=0}^{23} \left| Y_{j,i}^{act} - Y_{j,i}^{pred} \right|}{24 \cdot Avg_j} \right) + 0.3 \cdot \left(\frac{\left| \sum_{i=0}^{5} Y_{j,i}^{act} - \sum_{i=0}^{5} Y_{j,i}^{pred} \right|}{6 \cdot Avg_j} \right)$$

$$+ 0.1 \cdot \left(\frac{\left| \sum_{i=6}^{11} Y_{j,i}^{act} - \sum_{i=6}^{11} Y_{j,i}^{pred} \right|}{6 \cdot Avg_j} \right) + 0.1 \cdot \left(\frac{\left| \sum_{i=12}^{23} Y_{j,i}^{act} - \sum_{i=12}^{23} Y_{j,i}^{pred} \right|}{12 \cdot Avg_j} \right)$$

Finally the Prediction Error PE will be the average across all the prediction errors PE_j of all brands n in the test set:

$$PE = \frac{1}{n} \sum_{i=1}^{n} PE_j$$

NOVARTIS

Metric: Confidence Error

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Given the prediction intervals $\{L_j, U_j\}$ for a particular example we will measure 2 things with the following weights:

1. Whether the actual values fall inside the intervals (15%):

$$L_{j,i} \leq Y_{j,i}^{act} \leq U_{j,i}$$

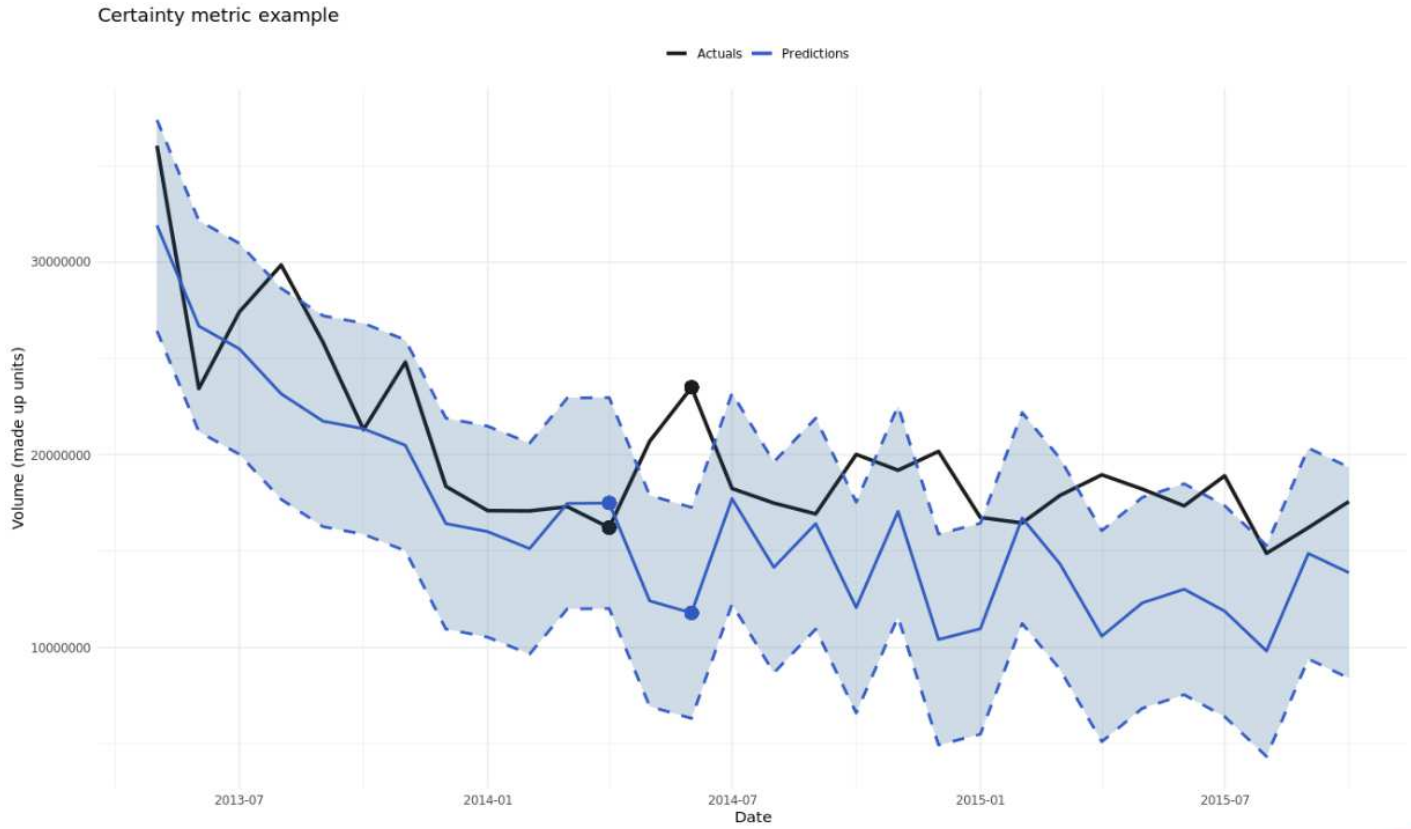
2. How accurate are the prediction intervals. In other words, we will penalize wide intervals by measuring the distance between them (85%):

$$|U_{j,i} - L_{j,i}|$$

For business reasons, the confidence Error for the first 6 months will be weighted more than the rest of the months (60% and 40% respectively). The error will be also normalized by the average monthly volume of the Brand in the 12 months prior to the generic entry.

NOVARTIS

Metric: Confidence Error



Metric: Confidence Error

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Formula:

Interval width penalization | Real point below lower interval | Real point above upper interval

$$CE_j = 0.6 \cdot \left[\frac{\sum_{i=0}^{5} \left(0.85 \cdot |U_{j,i} - L_{j,i}| + 0.15 \cdot \left[\frac{2}{0.05} \cdot (L_{j,i} - Y_{j,i}^{act})\, \mathbf{1}\{Y_{j,i}^{act} < L_{j,i}\} + \frac{2}{0.05} \cdot (Y_{j,i}^{act} - U_{j,i})\, \mathbf{1}\{Y_{j,i}^{act} > U_{j,i}\}\right]\right)}{6 \cdot Avg_j} \right]$$

$$+ 0.4 \cdot \left[\frac{\sum_{i=6}^{23} \left(0.85 \cdot |U_{j,i} - L_{j,i}| + 0.15 \cdot \left[\frac{2}{0.05} \cdot (L_{j,i} - Y_{j,i}^{act})\, \mathbf{1}\{Y_{j,i}^{act} < L_{j,i}\} + \frac{2}{0.05} \cdot (Y_{j,i}^{act} - U_{j,i})\, \mathbf{1}\{Y_{j,i}^{act} > U_{j,i}\}\right]\right)}{18 \cdot Avg_j} \right]$$

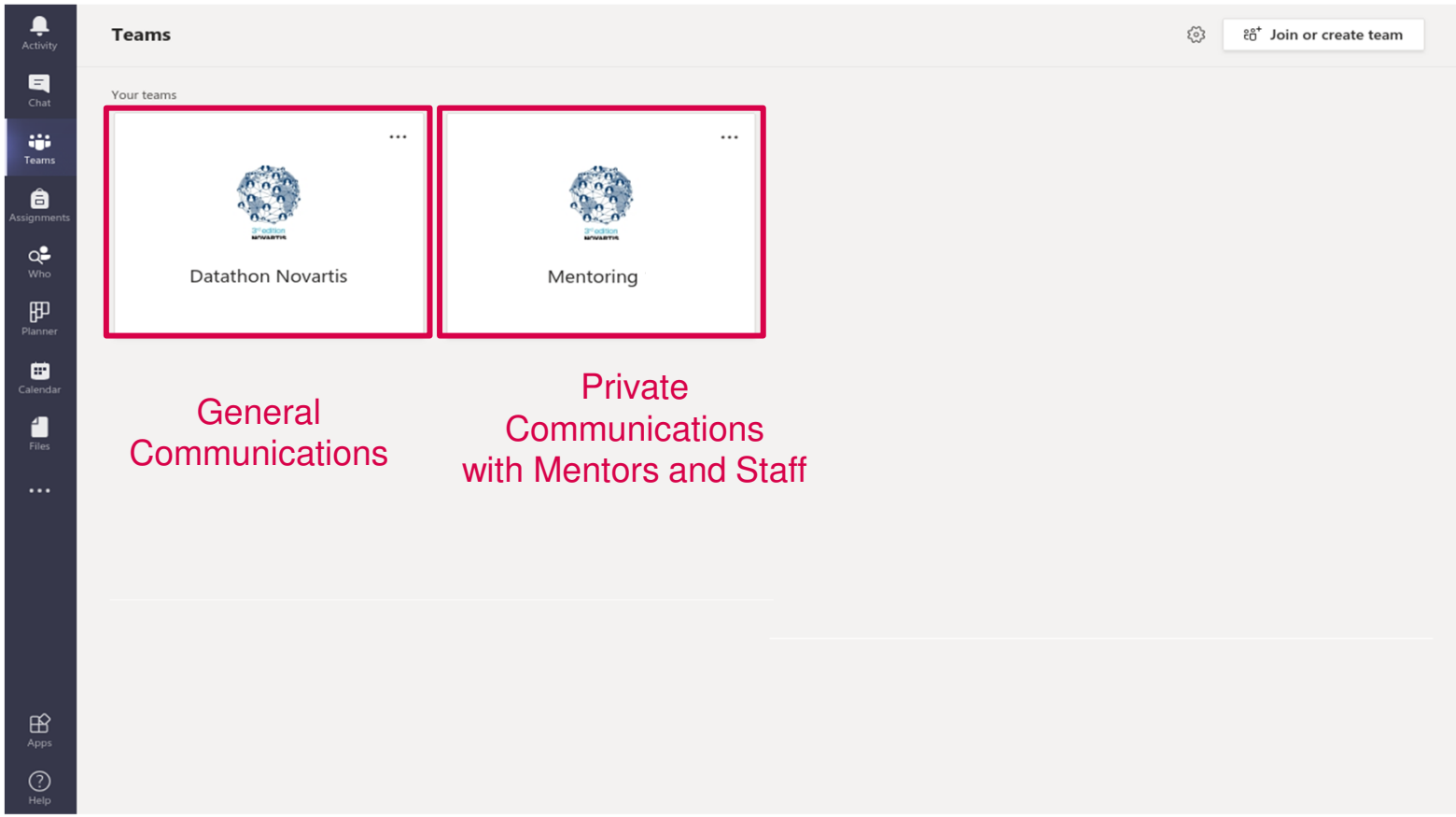
Finally the Confidence Error CE will be the average across all the confidence errors CE_j of all brands n in the test set:

$$CE = \frac{1}{n} \sum_{i=1}^{n} CE_j$$

NOVARTIS

Communication Channel

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



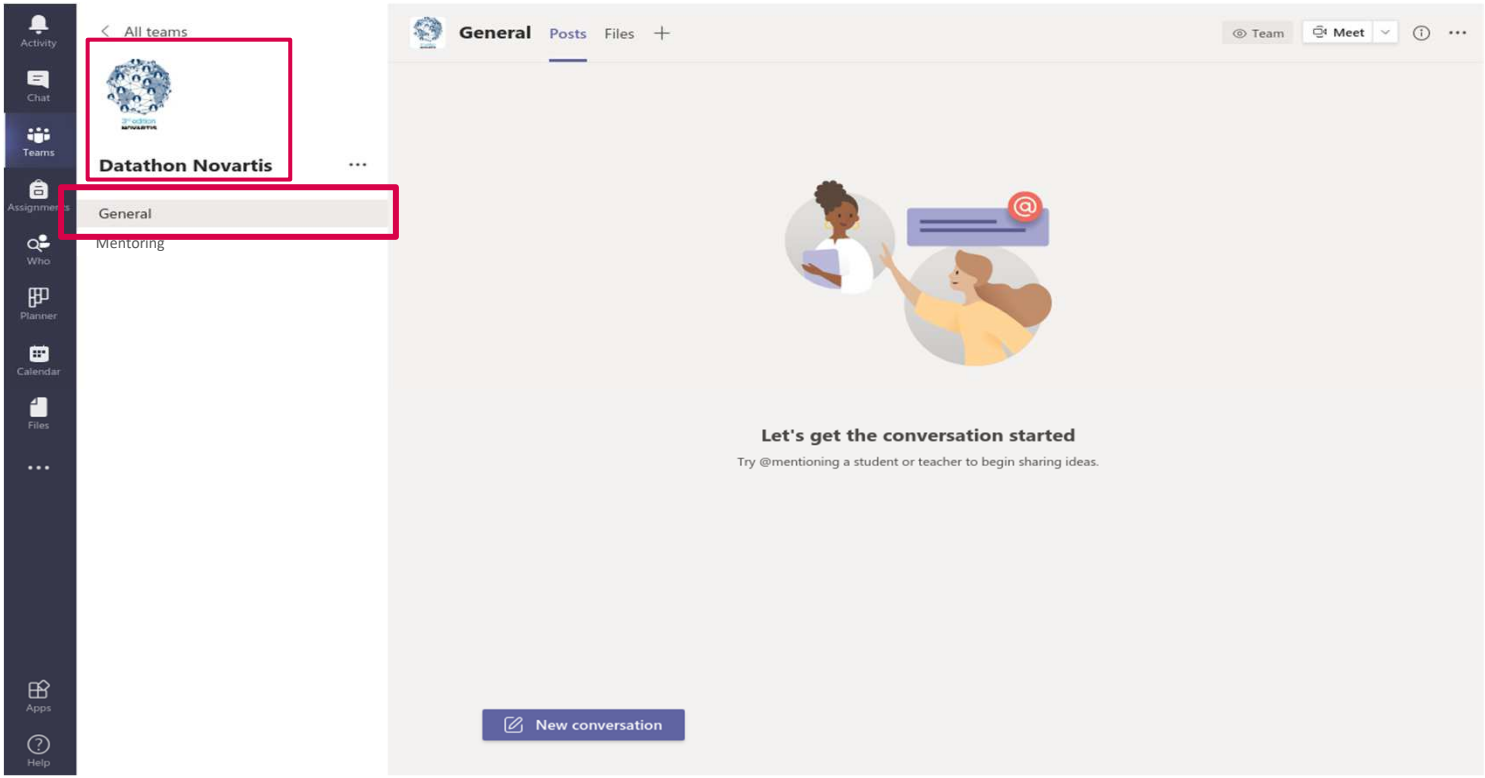
General Communications

Private Communications with Mentors and Staff

Communication Channel

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

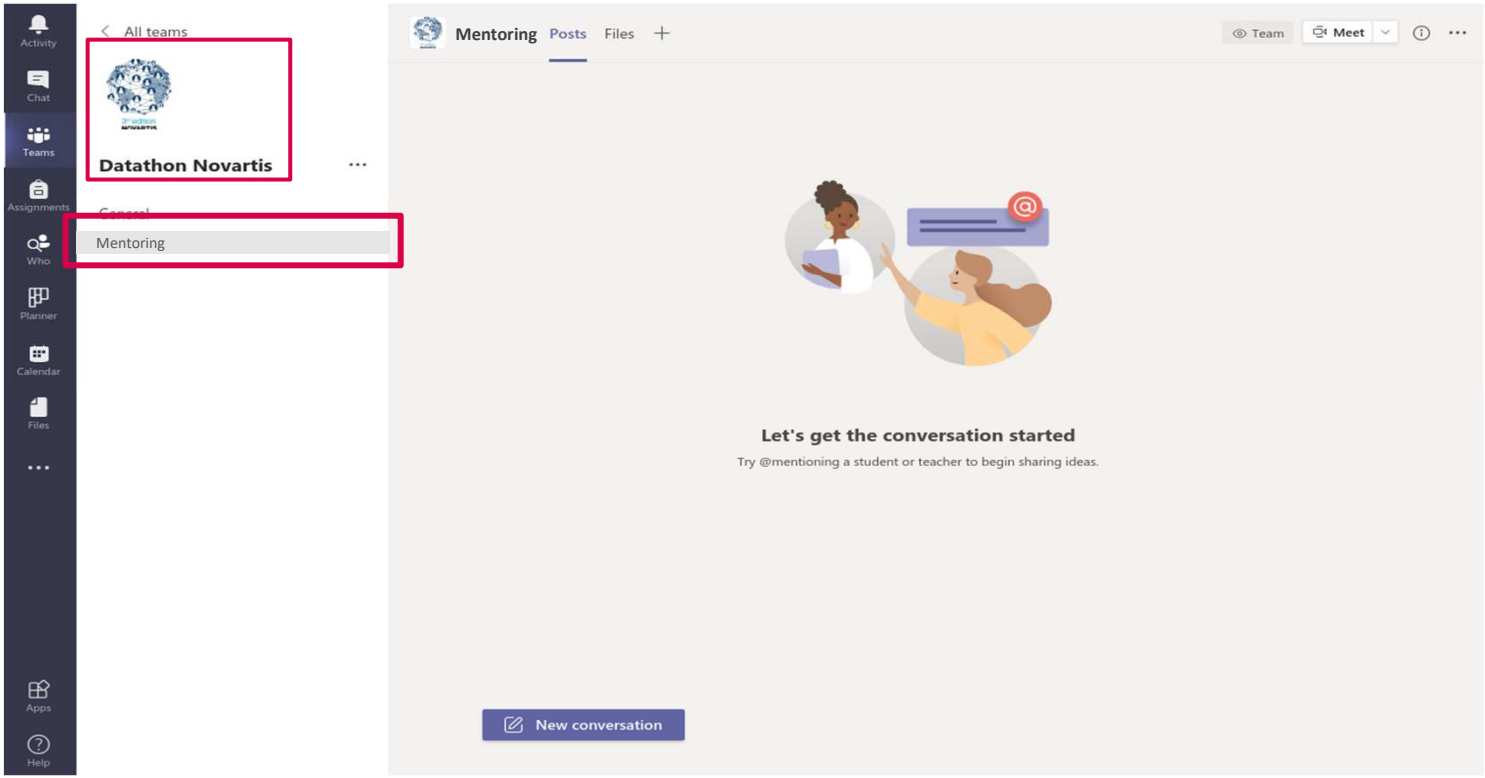
General communications



Communication Channel

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

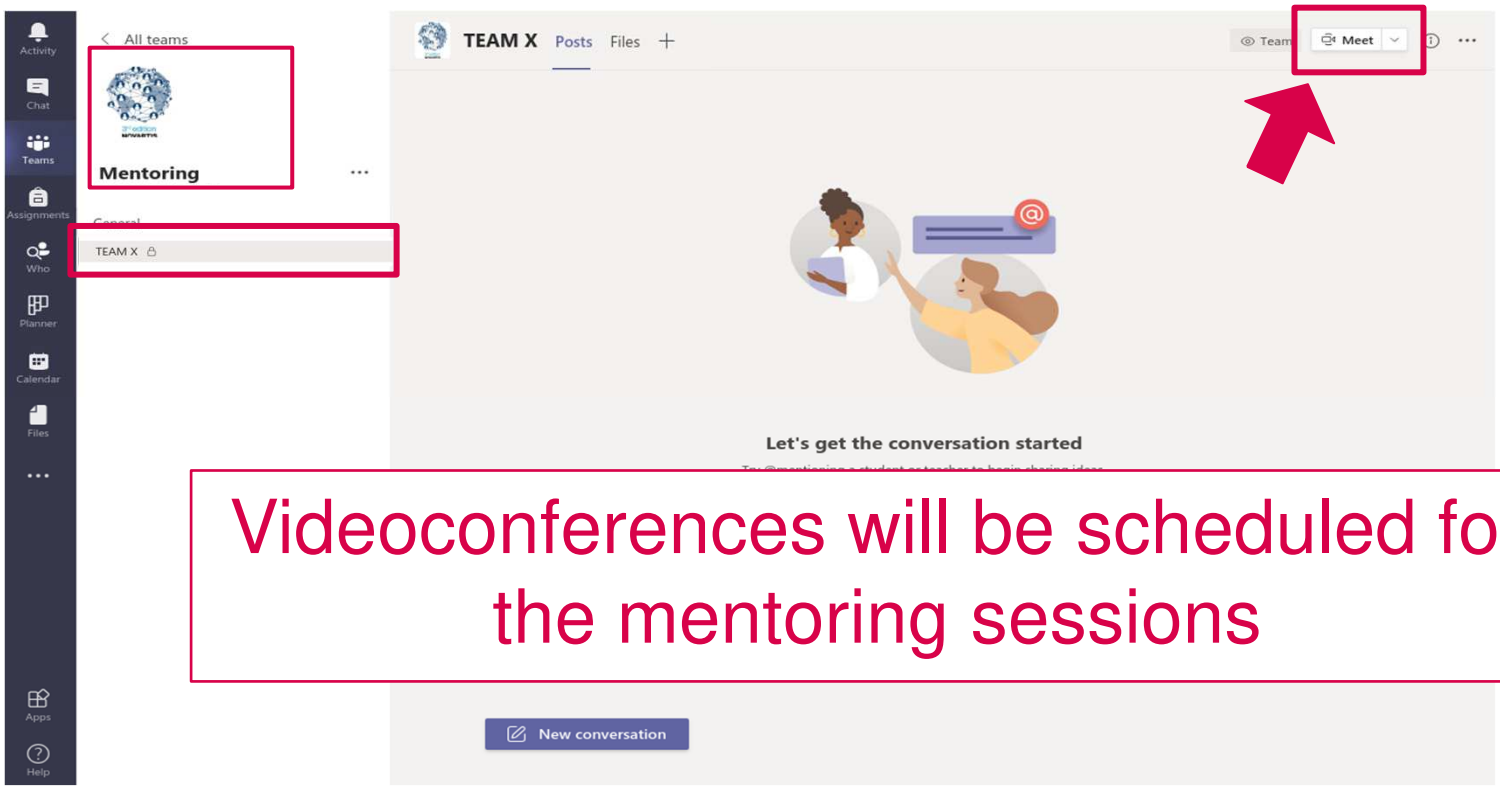
General communications



Communication Channel

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

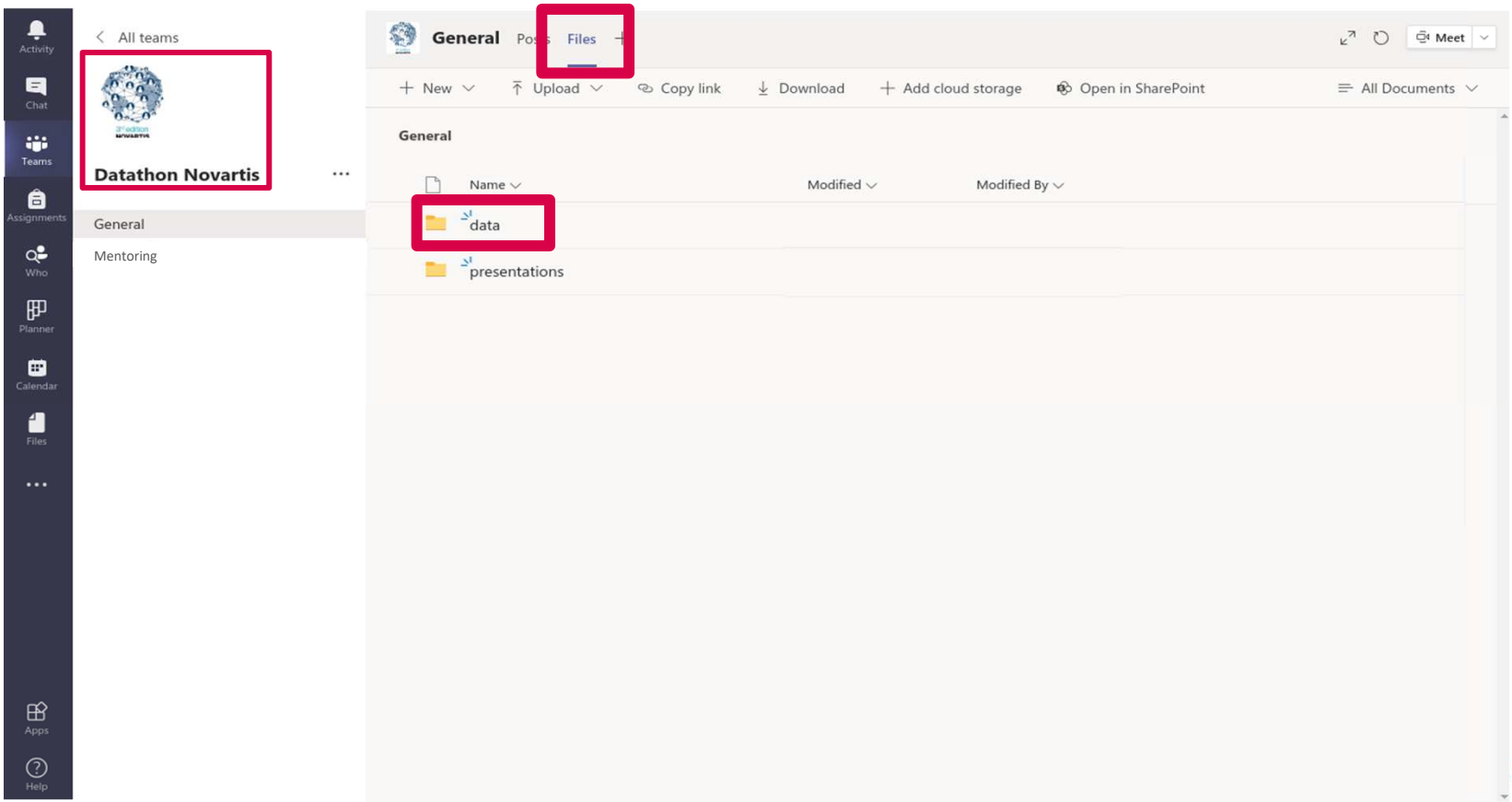
Private communications



Videoconferences will be scheduled for the mentoring sessions

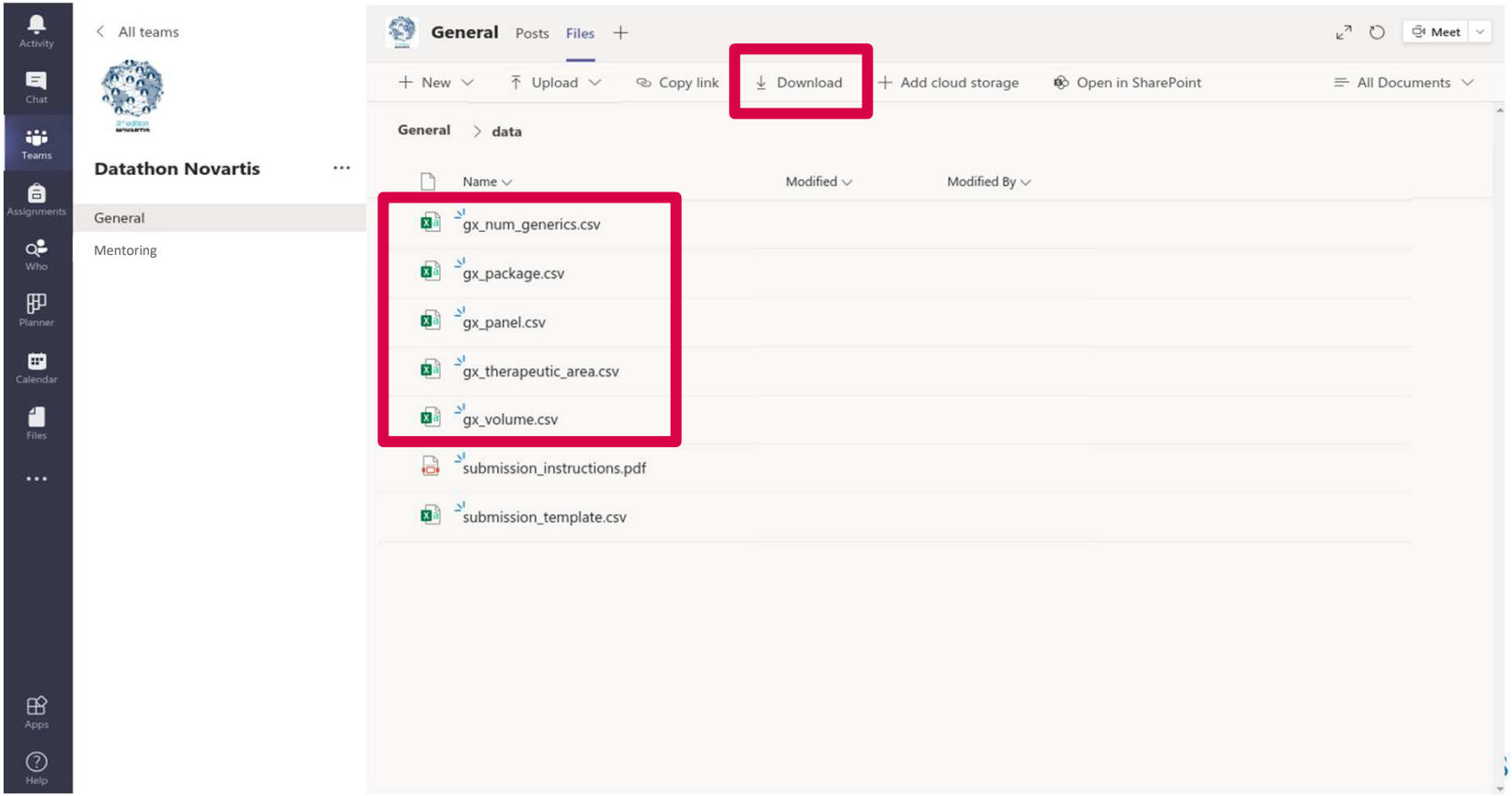
Download data

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



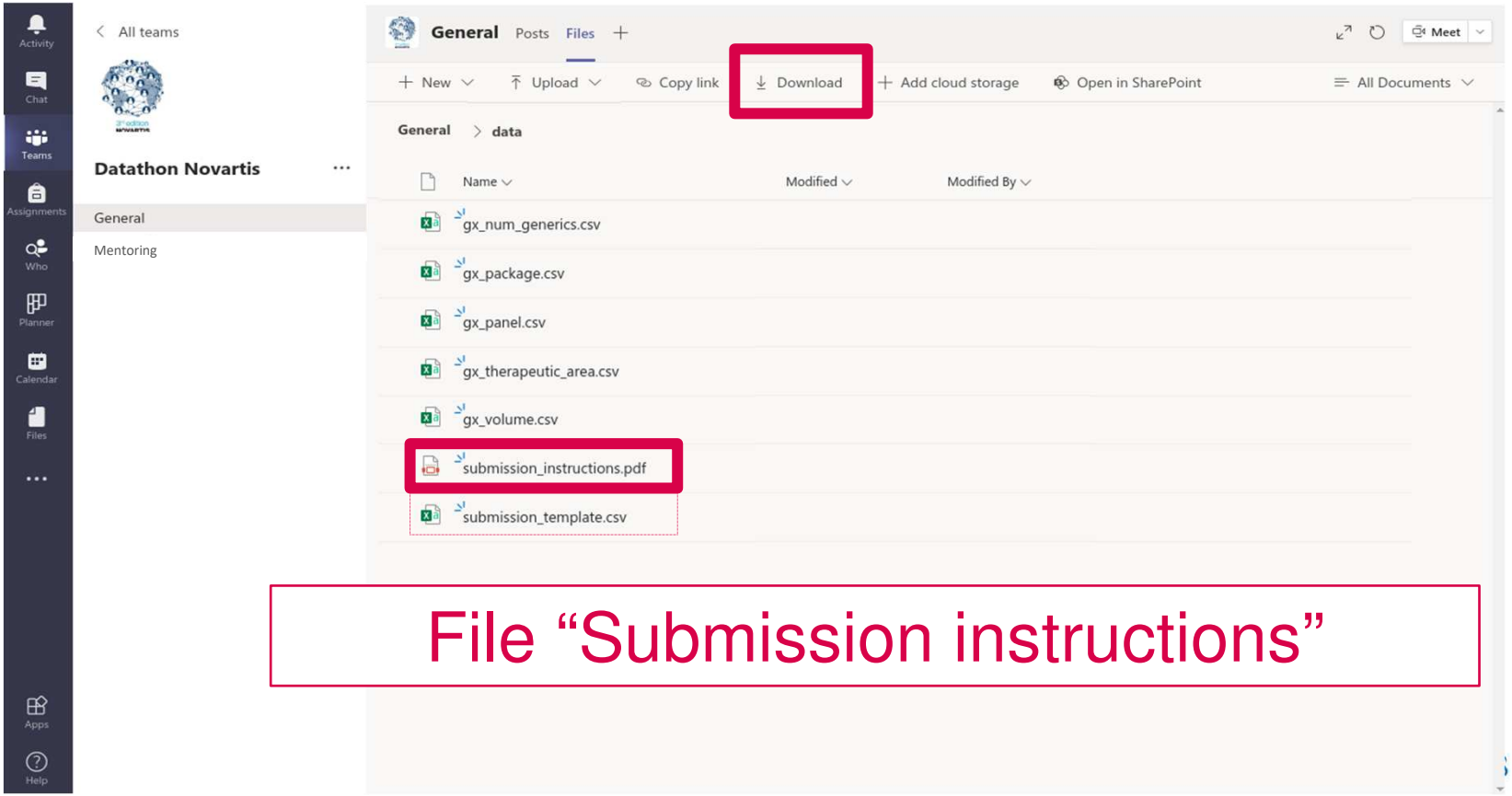
Download data

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



Download data

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



File “Submission instructions”

Submission structure

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

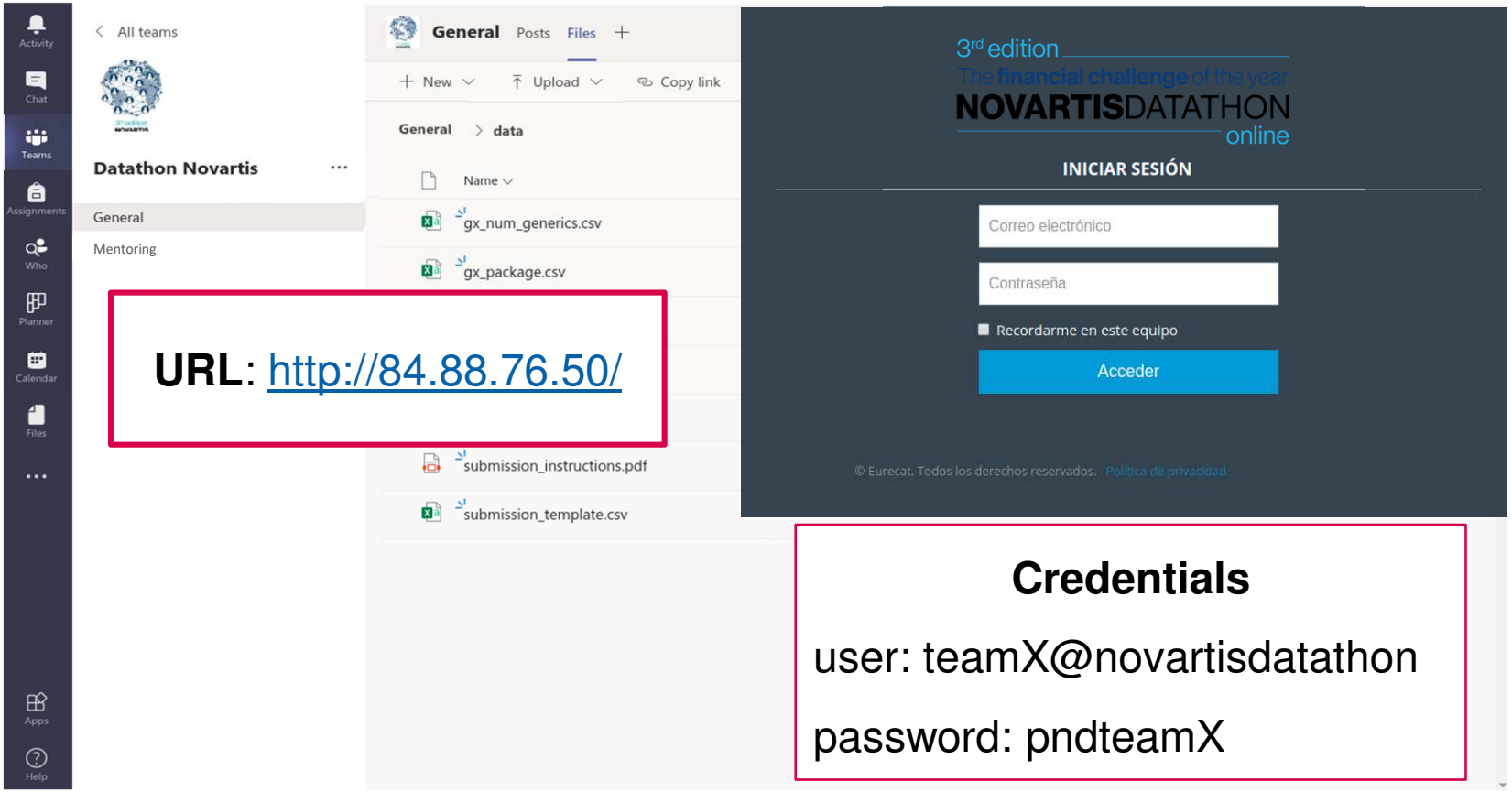
The **csv** you submit **must have**:

- a header
- same number of rows and columns as the test dataset / template
- columns in the same order as shown in the template
- comma-separated values
- pred_95_low ≤ prediction ≤ pred_95_high
- month_num should be in order from 0 to 23 per each country-brand

country	brand	month_num	pred_95_low	prediction	pred_95_high
country_1	brand_121	0			
country_1	brand_121	1			
country_1	brand_121	2			
country_1	brand_121	3			
country_1	brand_121	4			
country_1	brand_121	5			
country_1	brand_121	6			
country_1	brand_121	7			
country_1	brand_121	8			
country_1	brand_121	9			
country_1	brand_121	10			
country_1	brand_121	11			
country_1	brand_121	12			
country_1	brand_121	13			
country_1	brand_121	14			
country_1	brand_121	15			
country_1	brand_121	16			
country_1	brand_121	17			
country_1	brand_121	18			
country_1	brand_121	19			
country_1	brand_121	20			
country_1	brand_121	21			
country_1	brand_121	22			
country_1	brand_121	23			
country_1	brand_128	0			
country_1	brand_128	1			
country_1	brand_128	2			
country_1	brand_128	3			
country_1	brand_128	4			

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



URL: http://84.88.76.50/

Credentials

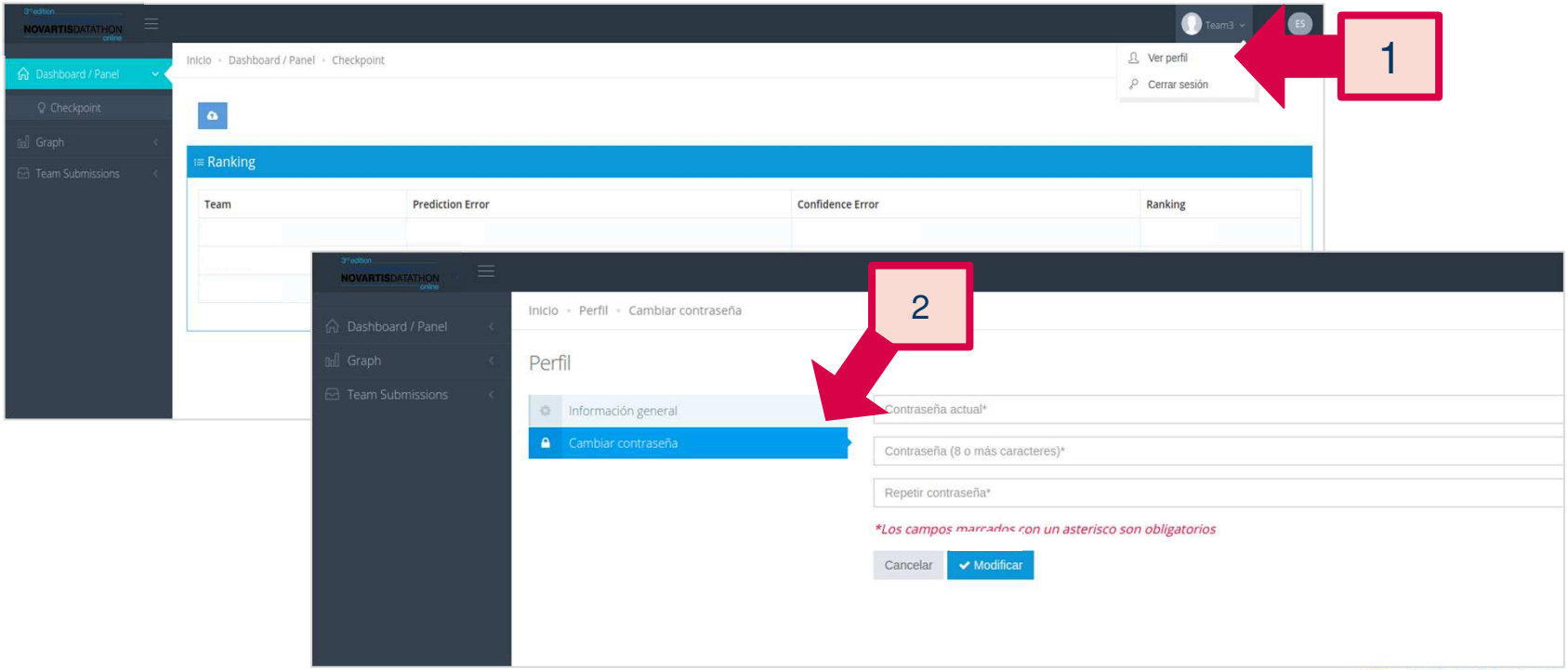
user: teamX@novartisdatathon

password: pndteamX

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

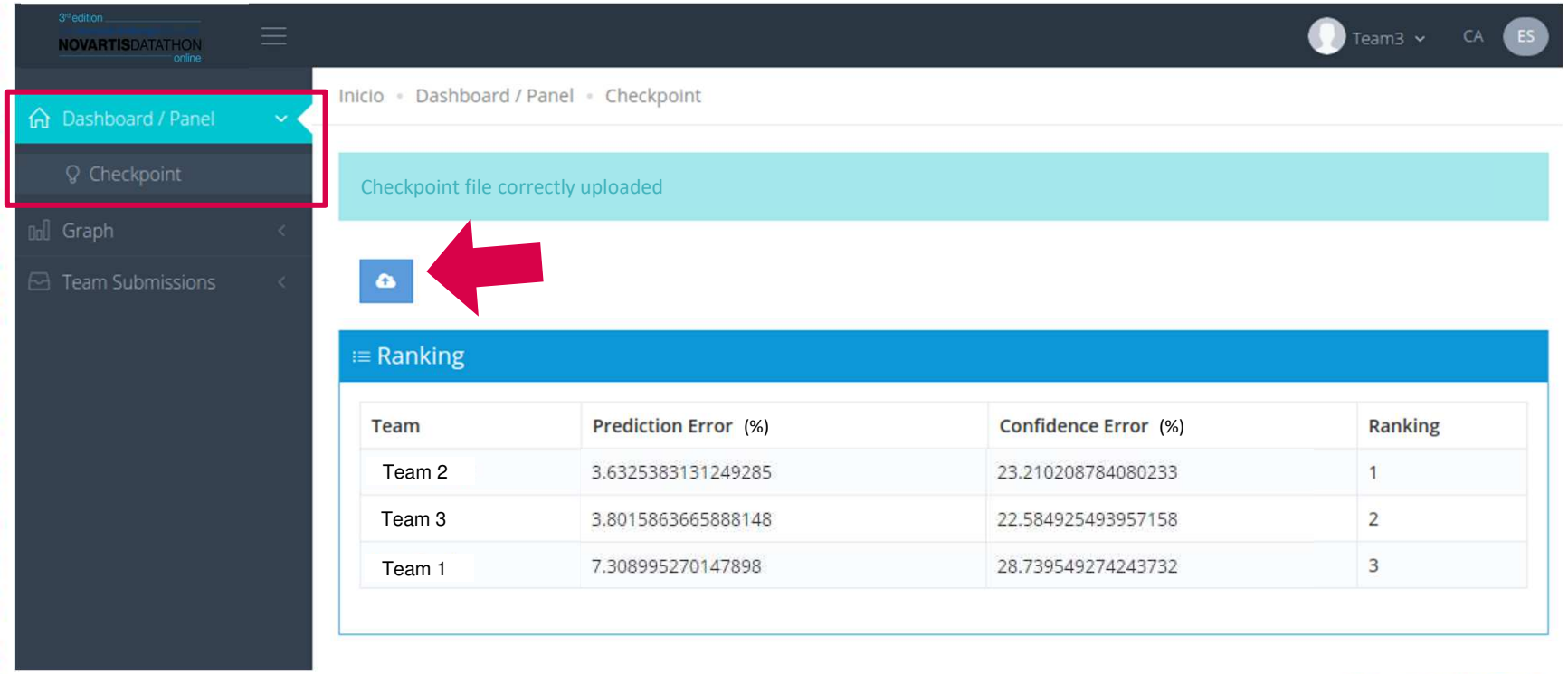
Please **change** the password



How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Submission

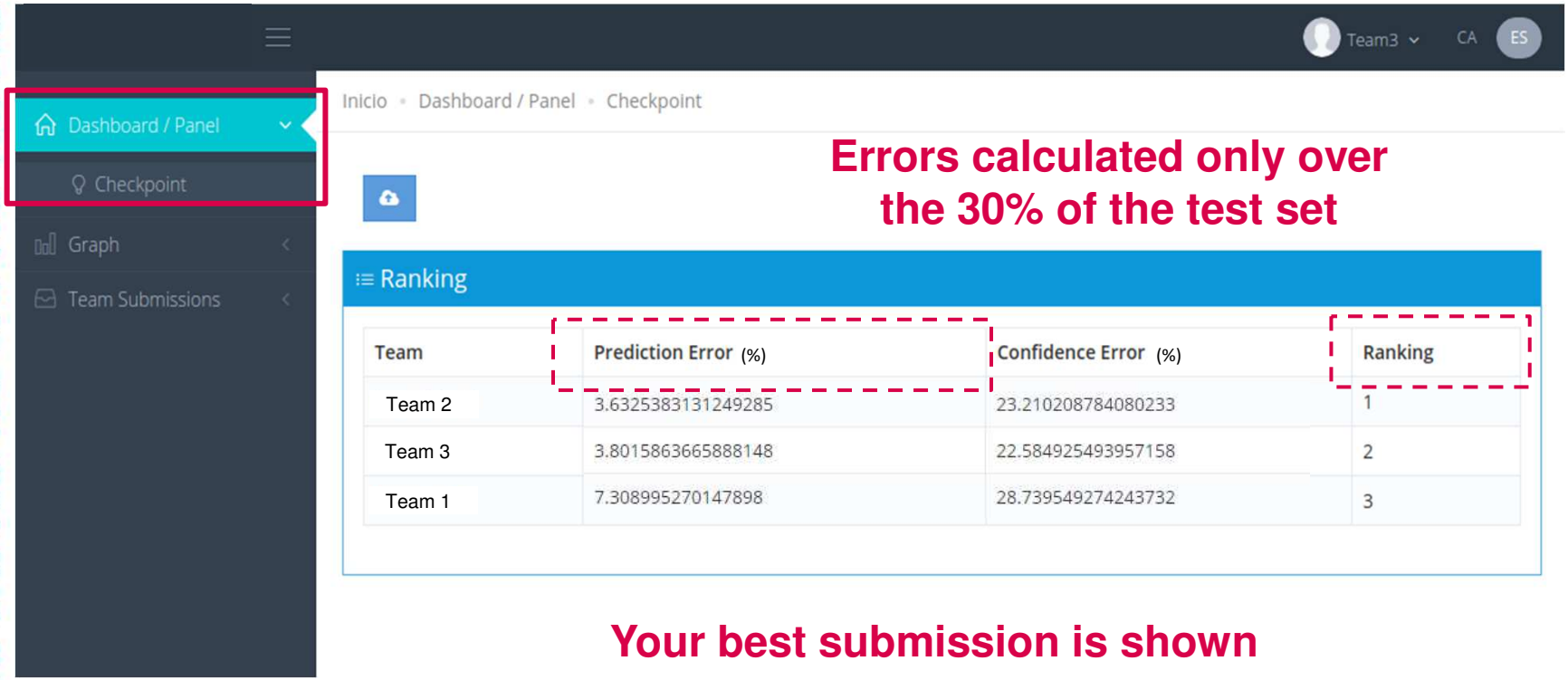


Team	Prediction Error (%)	Confidence Error (%)	Ranking
Team 2	3.6325383131249285	23.210208784080233	1
Team 3	3.8015863665888148	22.584925493957158	2
Team 1	7.308995270147898	28.739549274243732	3

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Ranking checkpoint



Team	Prediction Error (%)	Confidence Error (%)	Ranking
Team 2	3.6325383131249285	23.210208784080233	1
Team 3	3.8015863665888148	22.584925493957158	2
Team 1	7.308995270147898	28.739549274243732	3

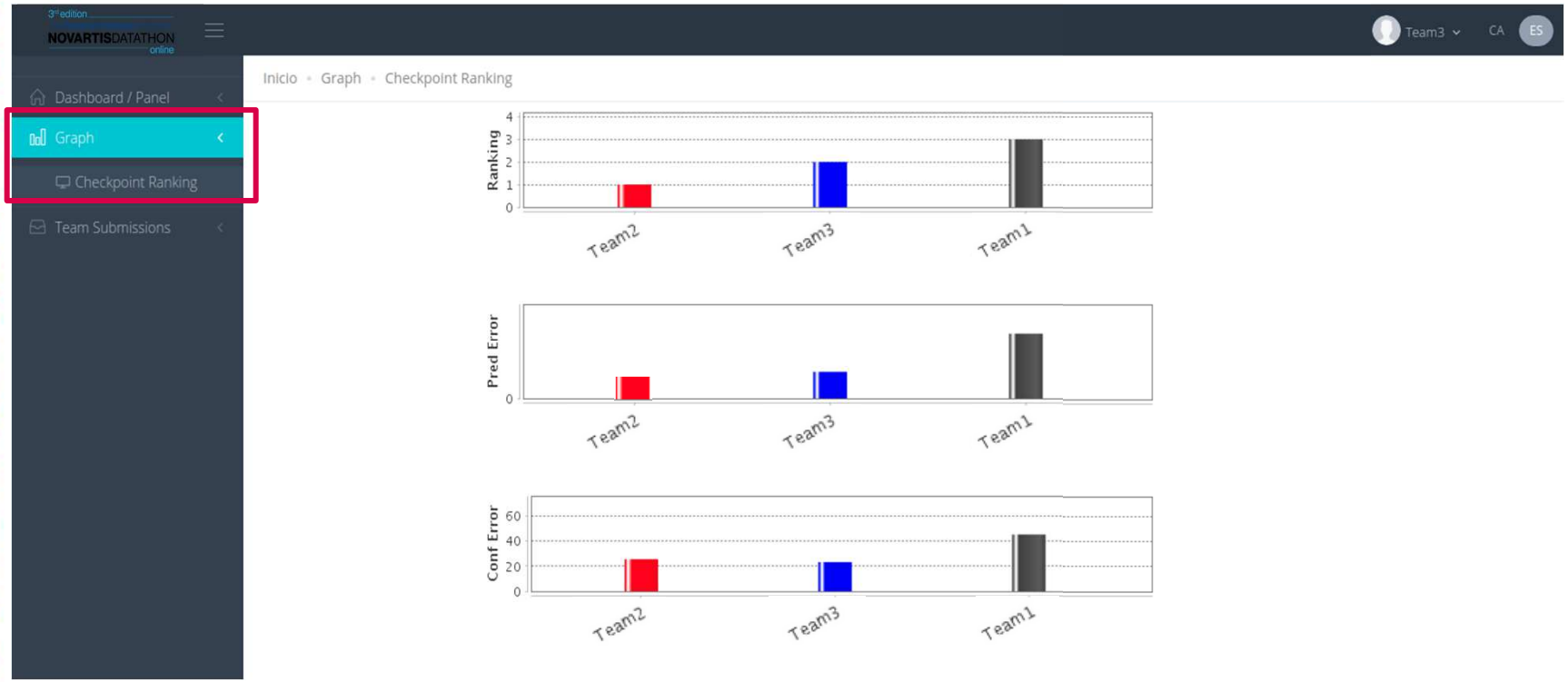
Errors calculated only over the 30% of the test set

Your best submission is shown

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

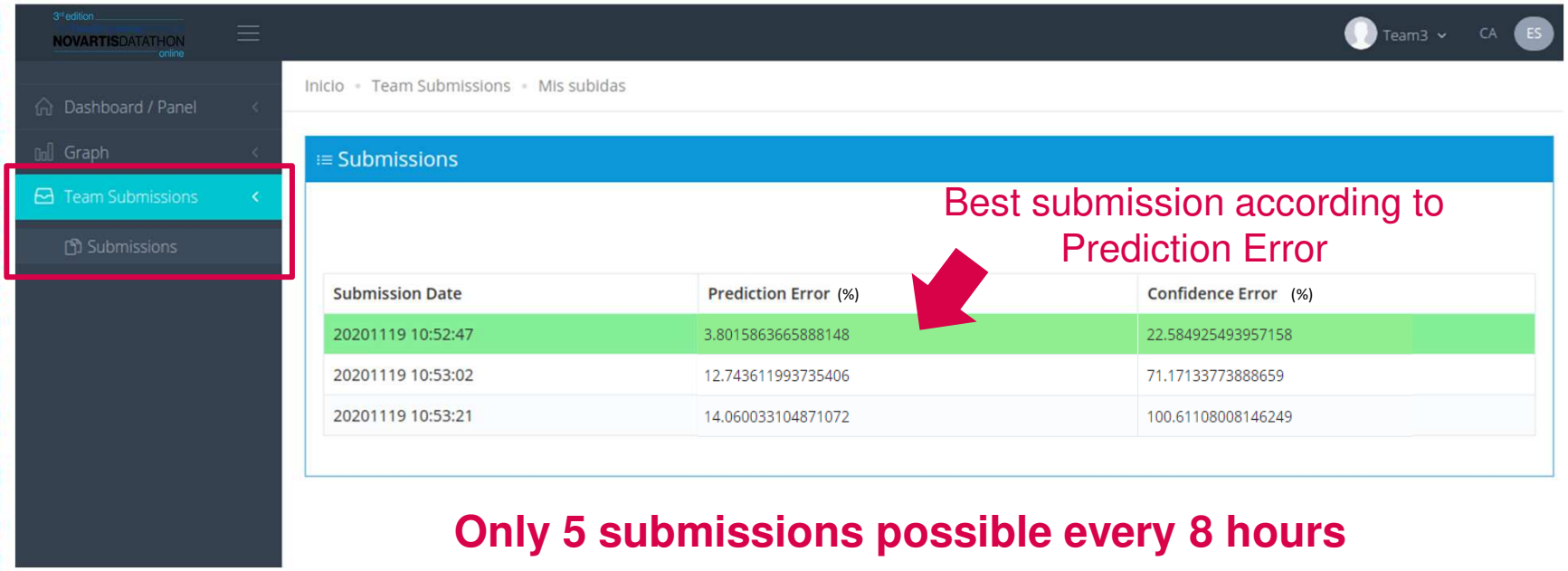
Ranking checkpoint



How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

History of submissions



Submission Date	Prediction Error (%)	Confidence Error (%)
20201119 10:52:47	3.8015863665888148	22.584925493957158
20201119 10:53:02	12.743611993735406	71.17133773888659
20201119 10:53:21	14.060033104871072	100.61108008146249

Only 5 submissions possible every 8 hours

4am-12pm | 12pm-8pm | 8pm-4am *

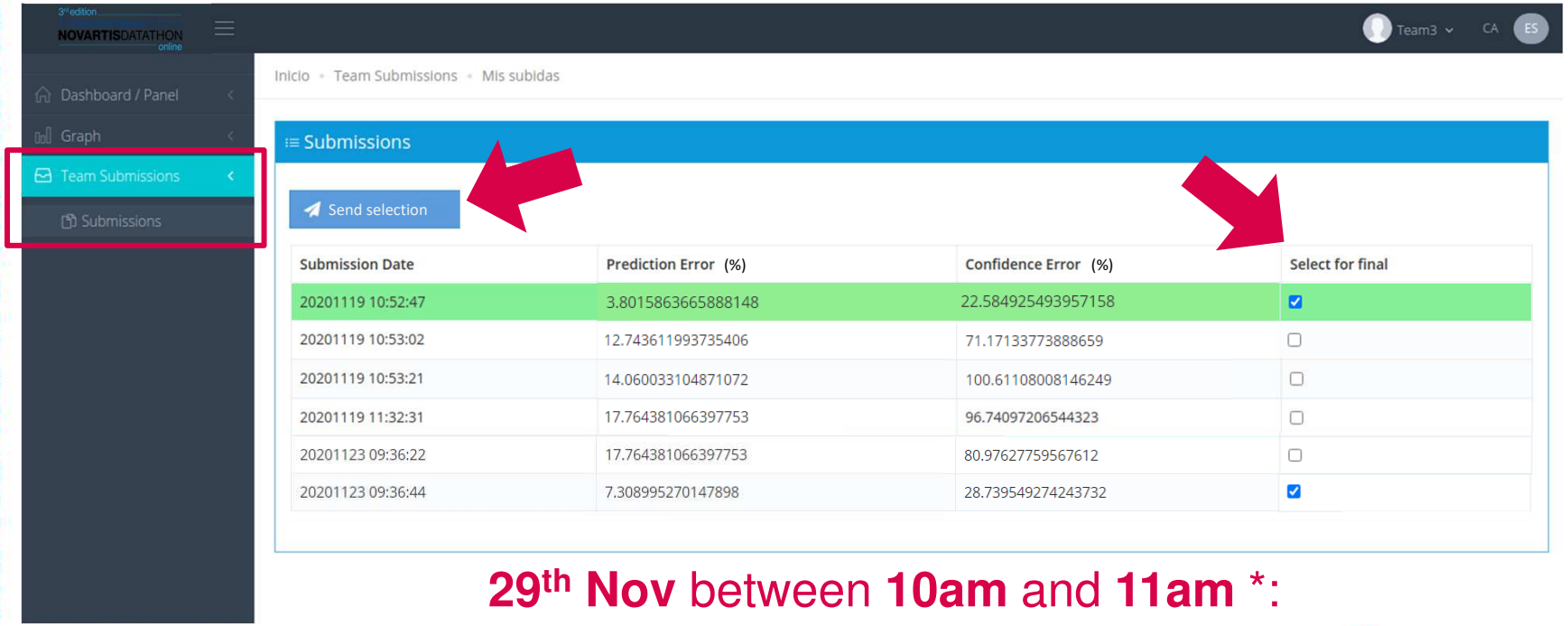
*Central European Time - Barcelona, UTC +1h

NOVARTIS

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Final submission (last hour)



29th Nov between **10am** and **11am** *:

select maximum **two** submissions

*Central European Time - Barcelona, UTC +1h

NOVARTIS

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

FINAL results calculated over the 100% of the test set
once the datathon is over
(29th Nov 11am*)

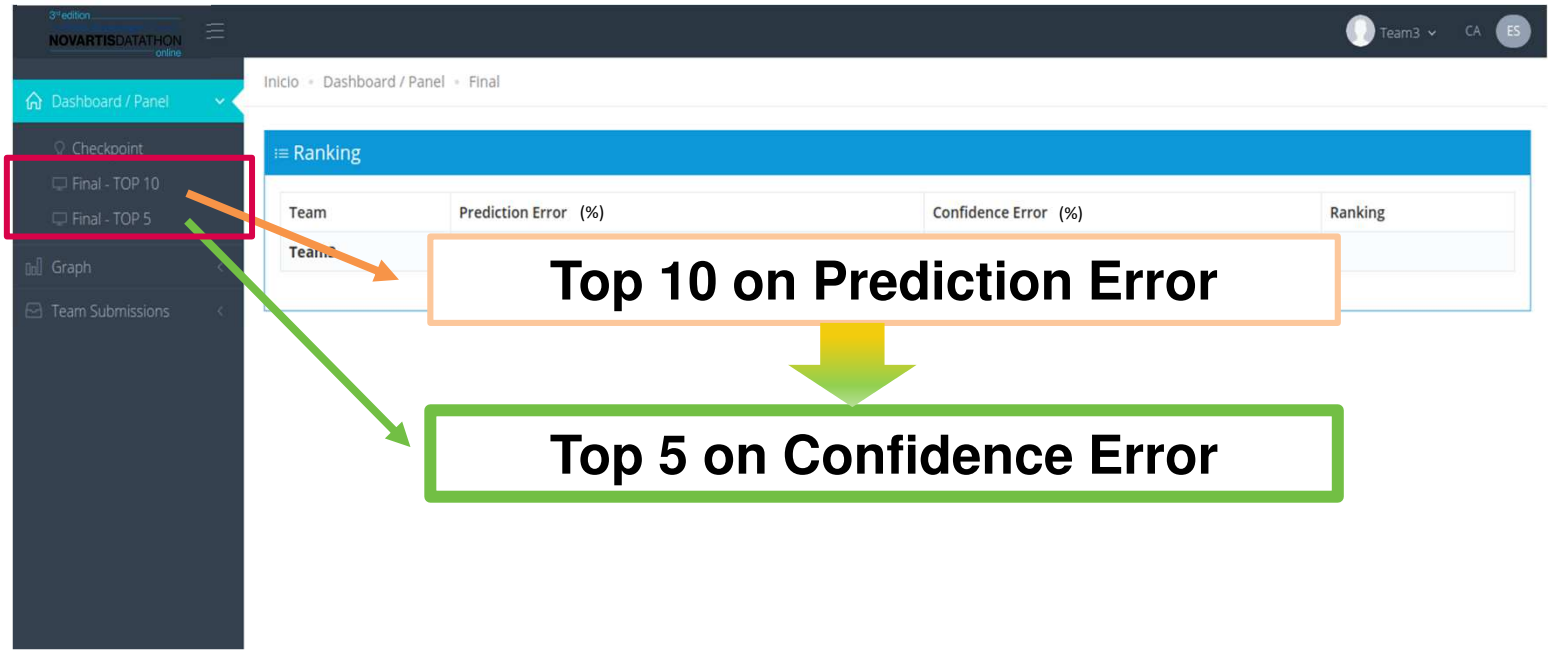
*Central European Time - Barcelona, UTC +1h

NOVARTIS

How to submit results

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Final results: Deadline **11am*** on Sunday

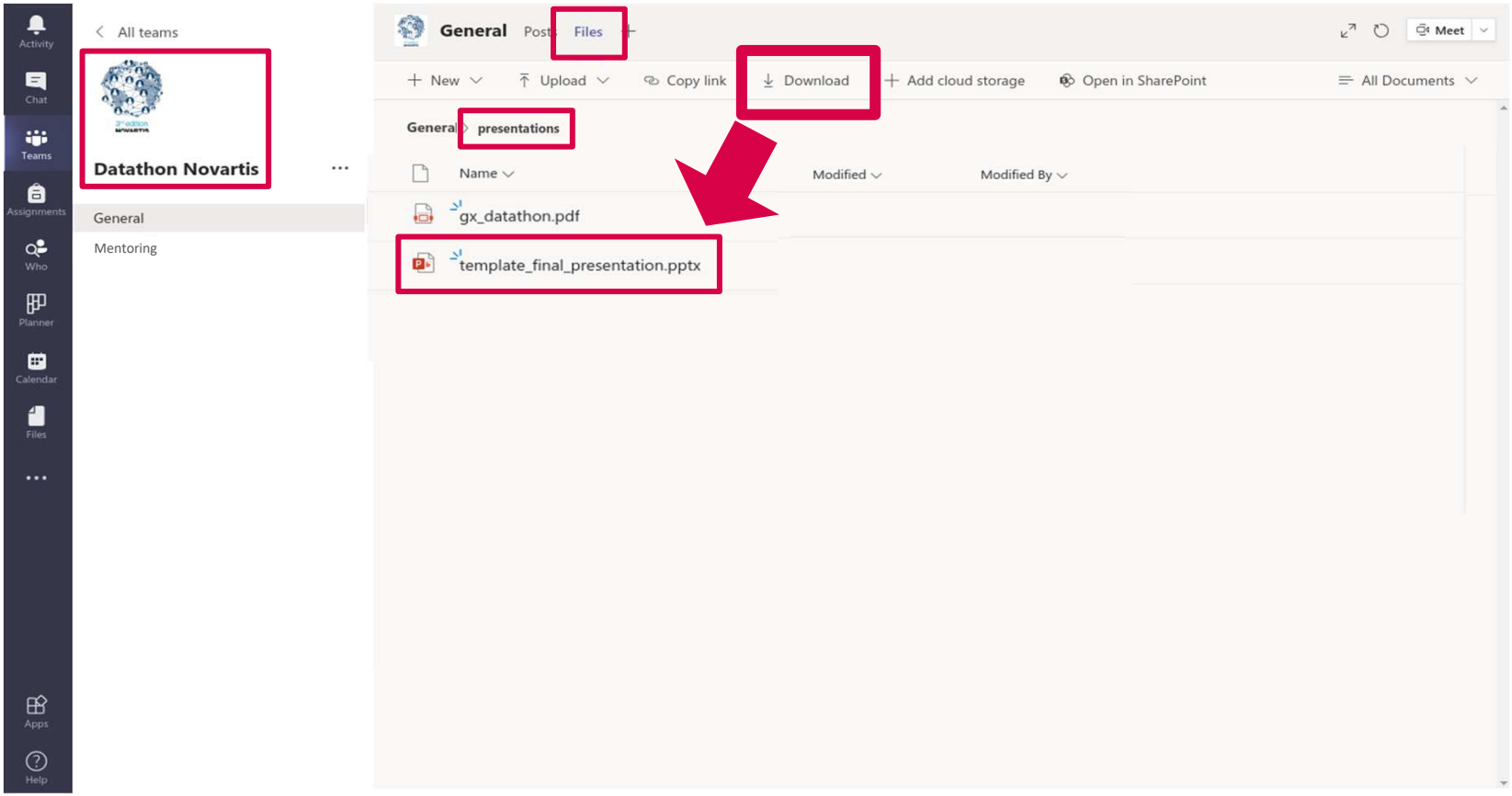


*Central European Time - Barcelona, UTC +1h

NOVARTIS

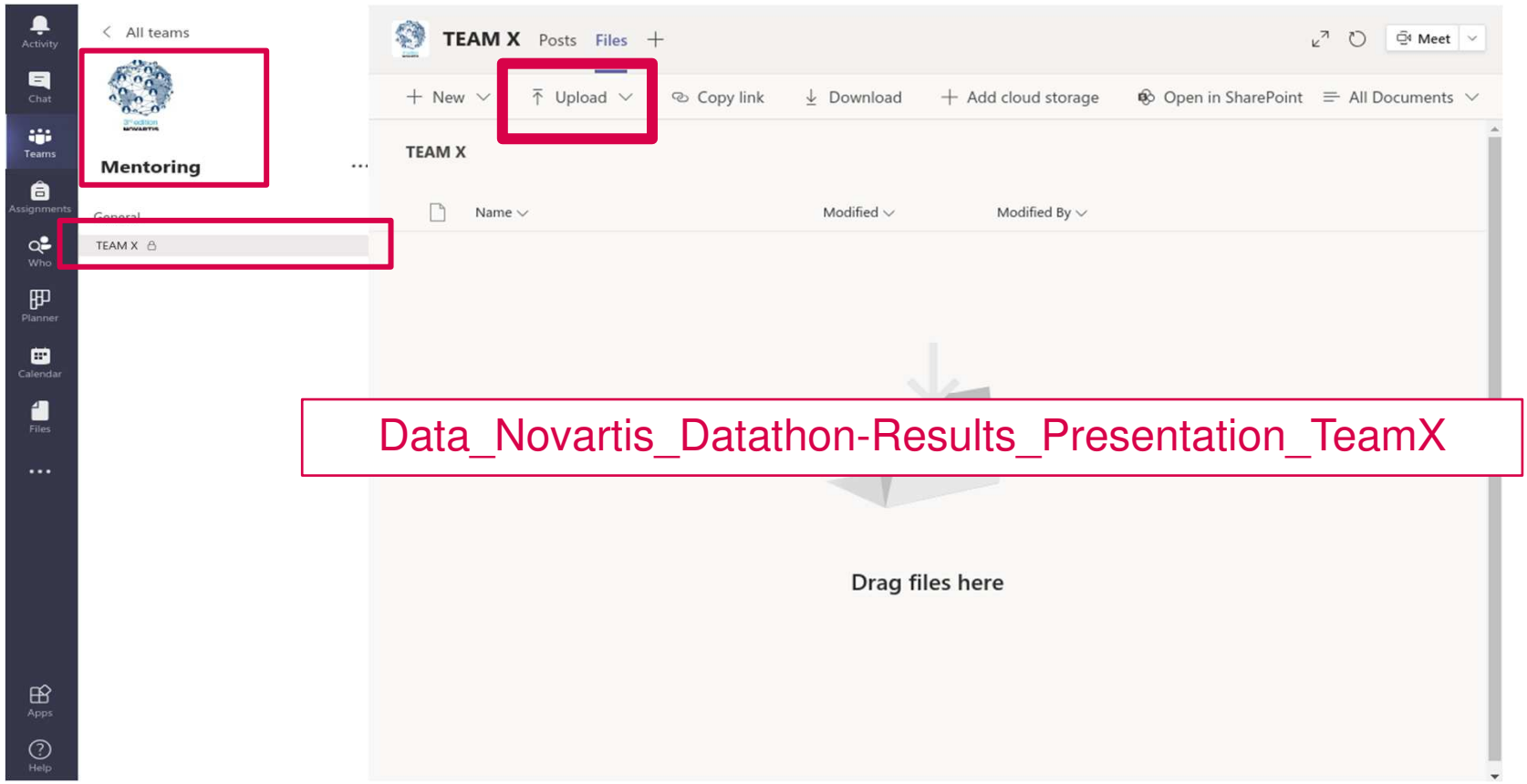
Submit presentation TOP 5

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



Submit presentation TOP 5

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online



Data_Novartis_Datathon-Results_Presentation_TeamX

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

AGENDA

THU 26 November

16:00h – 17:00h Kick-off

17:00h – 18:00h Welcome and Intro

FRI 27 November

09:00h – 18:00h Attendance of questions & Mentoring

SAT 28 November

09:00h – 18:00h Attendance of questions & Mentoring

SUN 29 November

09:00h Welcome and Jury introduction

Attendance of questions

11:00h Ranking 5th Finalists

13:00h – 14:30h Presentations

14:30h – 15:00h Jury deliberates

15:00h Announcement of the Winners

Case work from Thursday 26th 18:00h onwards

*Central European Time - Barcelona, UTC +1h

NOVARTIS

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

GOOD LUCK

NOVARTIS

3rd edition
The financial challenge of the year
NOVARTISDATATHON
online

Organizer

NOVARTIS

In collaboration with

eurecat
Centre Tecnològic de Catalunya