IMPERIAL

Data Challenge 2024-2025

MSc in Statistics

09/07/2025

Program

9:30am – Introduction to this year's Data Challenge and kick off

12:00pm – Intermediate submission

2:00pm – Intermediate ranking : back in HXLY 130

4:30pm – Final submissions

5:00pm – Awards to winners

Breakout rooms:
MLC -HXLY414 and HXLY411
MSc Quiet Study Space HXLY413
MSc common room   HXLY215
Huxley Common Room HXLY549

10/07/2025

Prizes

Based on the performance and the report we will assign:

- The Winton Prize - MSc Statistics Data challenge competition - winner, £1,500

- The Winton Prize - MSc Statistics Data challenge competition - second best team £600

- The Winton Prize - MSc Statistics Data challenge competition - third best team £400

10/07/2025

Data Description

Goal – detecting credit card frauds from transactions of a European bank.

Dataset of credit card transaction in 2023:

Training set
80000 credit card transactions between 17/09 and 17/10

Test set
11557 credit card transactions between 17/10 and 21/10

Data Description

Per each transaction:

ID_TRX is the id of the transaction

ID_CARD_BEN is the id of the card

DATETIME_GMT is the time and hour of the transaction

AMOUNT is the amount of the transaction

FLAG_FRAUD is the 0,1 label you are trying to predict

10/07/2025

Data Description

There are the following additional variables the bank is creating:

Anomaly_amount: 9 numeric variables measuring the anomaly of the amount of the transaction

FLAG_BEHAVIOUR_Anomaly: 8 categorical variables measuring the anomaly of the amount of the transaction

Population_Anomaly: 8 numerical variables measuring the anomaly of the amount of the transaction with respect to the client population

Prediction Evaluation

We are going to use the F1 metric to measure the performances of predicting the fraud:

$$\text{F}_1 = 2\,\frac{precision * recall}{precision + recall}$$

where:

$$precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Submission of the predictions

You can download the data from blackboard.

You can submit your predictions using Turnit-in in a zip file containing:

1. The predictions must be :
   - in .csv format separated by ';'
   - with your prediction column name FLAG_FRAUD
   - the file name should be the name of your team

2. The report should be maximum 1 pdf page. The aim of the report is to **describe and justify your model choices**.

IMPERIAL

GOOD LUCK!