RÉPUBLIQUE FRANÇAISE

Liberté
Égalité
Fraternité

FINANCES PUBLIQUES

AI for Tax Evasion Detection in France

12.03.2025

Plan

Policy & Governance 01.

How we organize the use of AI?

Data & Models 02.

How we use the data to detect tax evasion?

Performance & results 03.

How well does it work?

1

Policy & Governance

A decade-long use of AI

Milestones

2014

Creation of a dedicated unit

2018

Strengthening inter-agency data sharing

2021

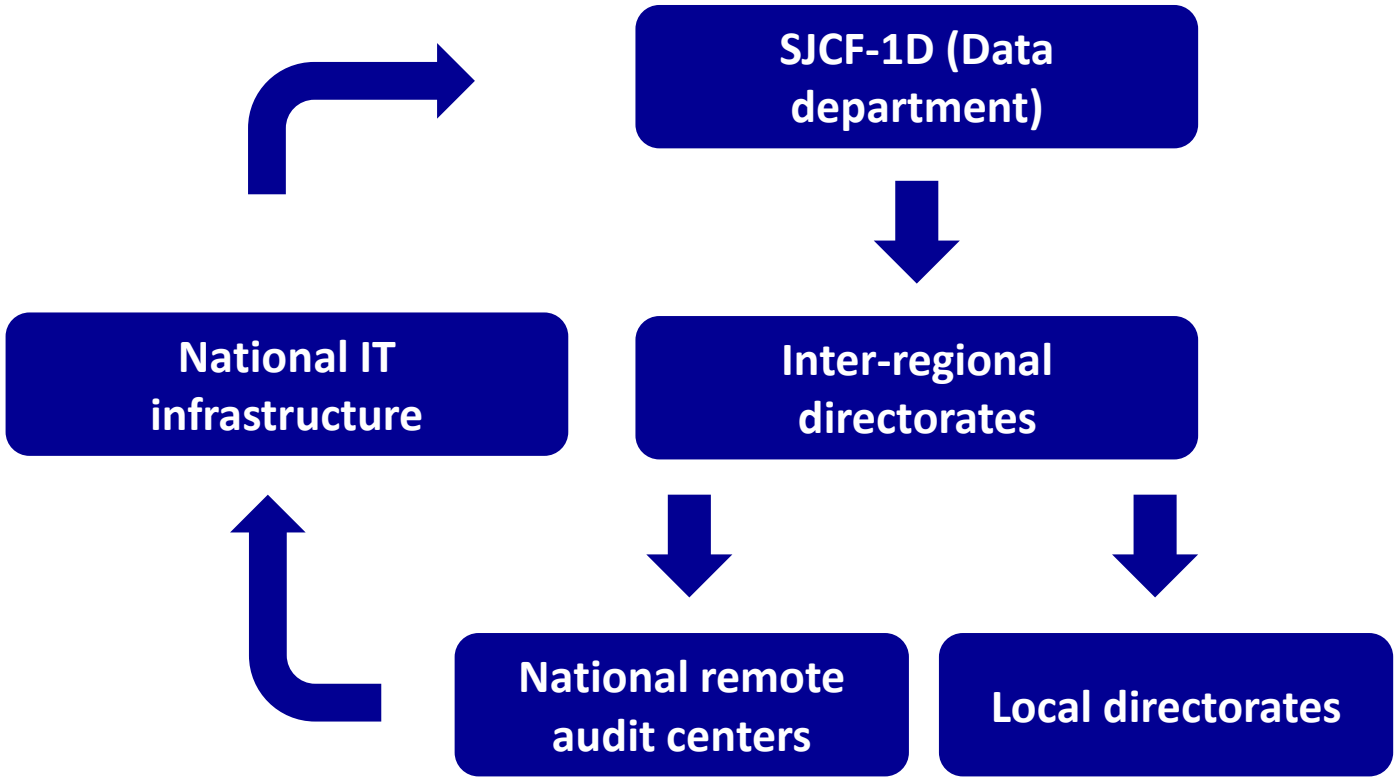
Authorisation to use social network data

2022

Deployment of the GALAXIE programme

AI is upstream in the process

Organization of the tax audit process



2

Data & Models

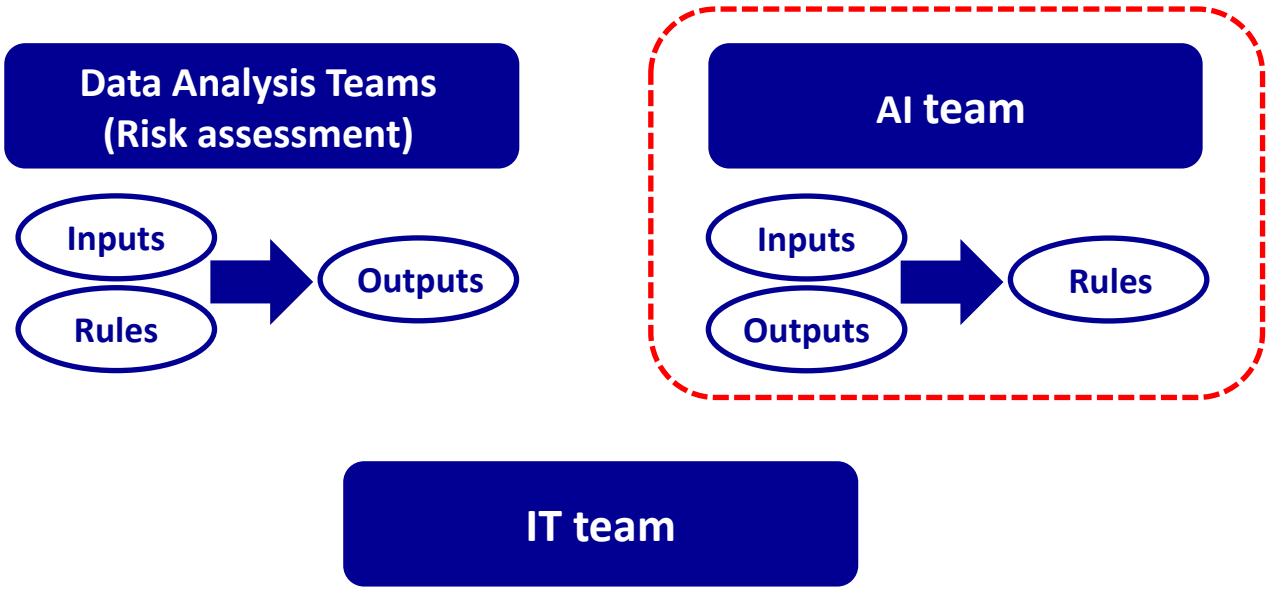
55 data flows are collected and stored

The data department stores terabytes of data every year

1	**Tax data**	Business, individual, asset, property
2	**Third party reported data**	Sharing economy platform data, staff reporting statements, fees paid, etc.
3	**Data from international automatic exchange of information**	Assets held abroad, etc.
4	**Data obtained from other government departments or ministries**	Company data, customs data, etc.
5	**Open-source data**	Register of beneficial ownership of companies
6	**Data collected from social media**	E-commerce websites

The AI team is a part of the data department

Alongside the analysis team for individuals and the analysis team for businesses



We use 3 main types of models

In the AI team

1 **Unsupervised models**

→ Identify homogeneous clusters, then:
- Target the most fraudulent clusters
- Target the most isolated individuals

2 **Supervised models**

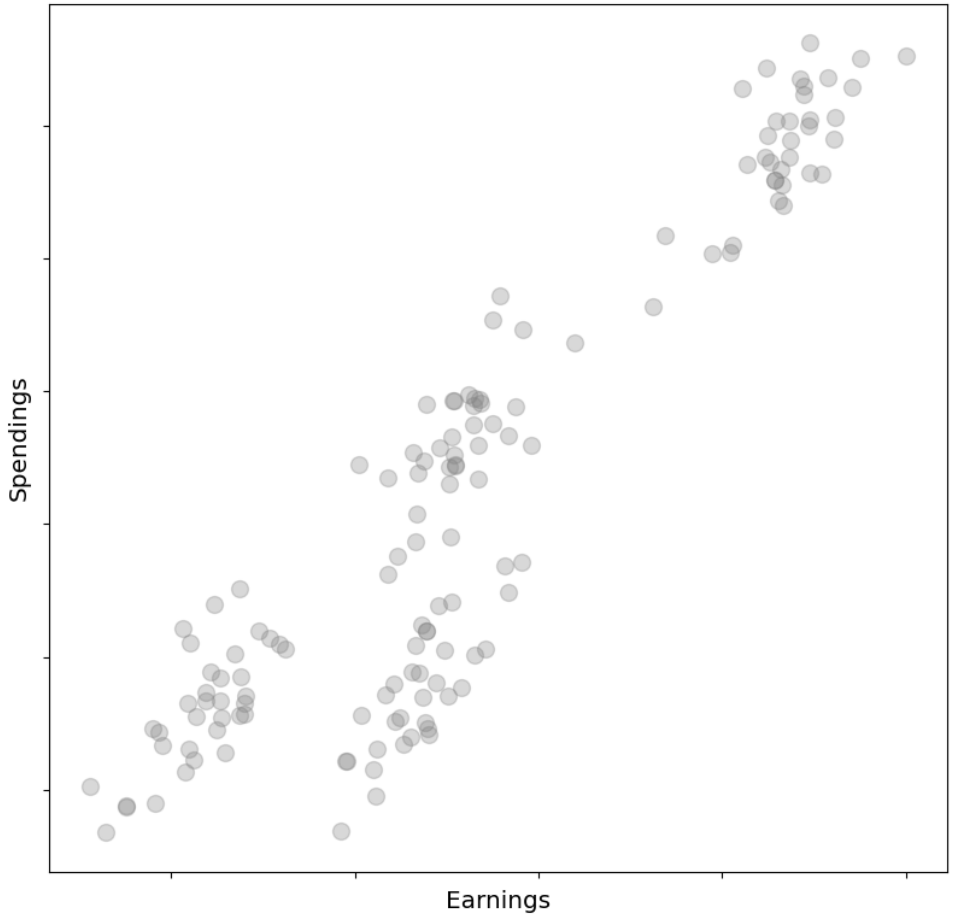
→ Learn to predict some known quantity, then compare to detect outliers

3 **Graph models**

→ Detect fraud networks based on links between individuals and companies

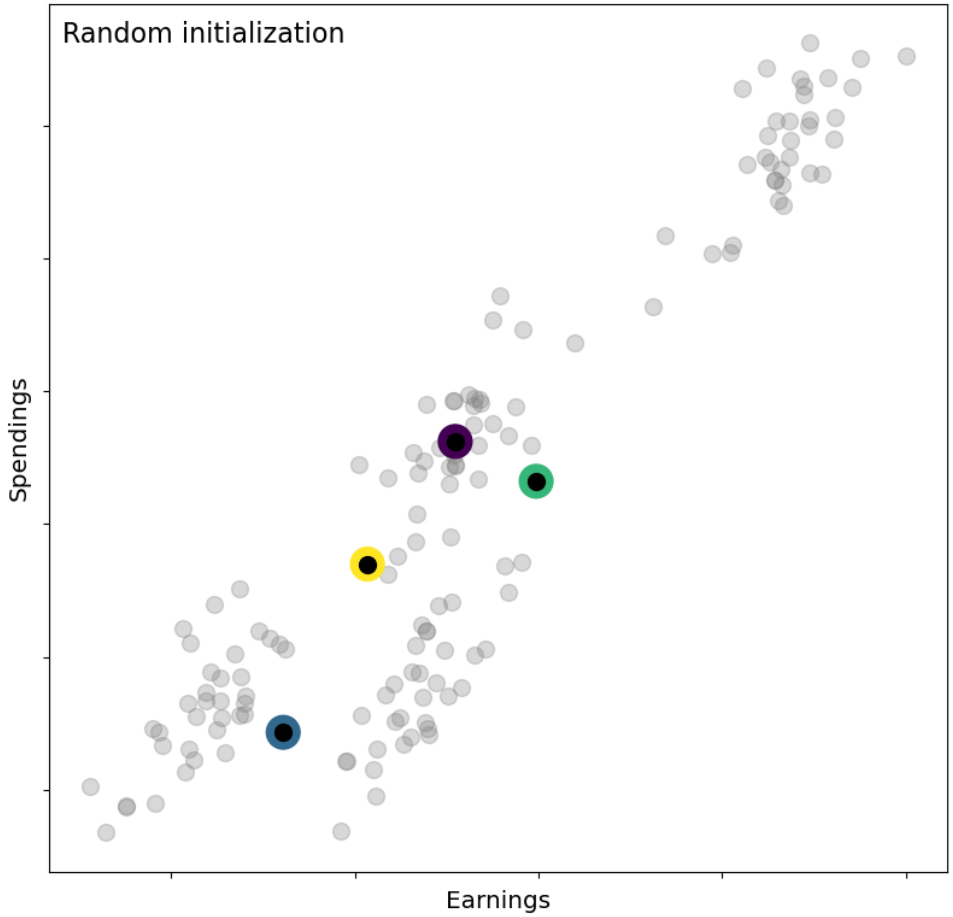
Unsupervised models

Find homogeneous clusters



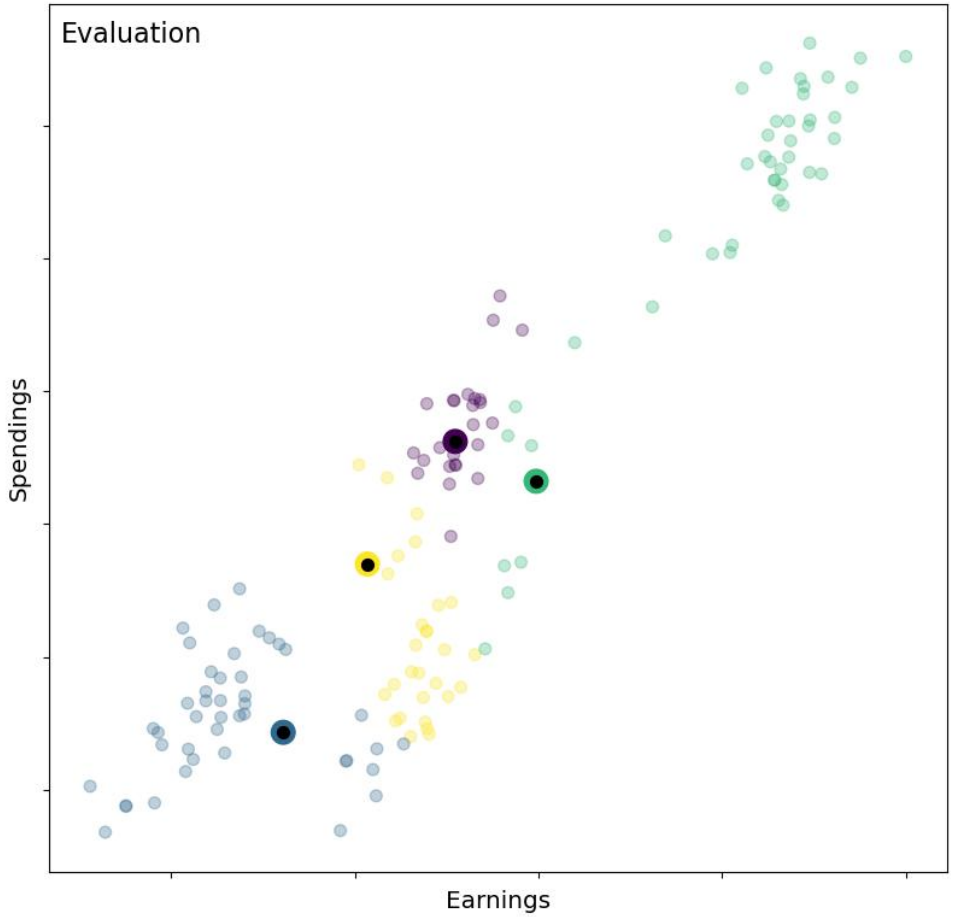
Unsupervised models

Find homogeneous clusters



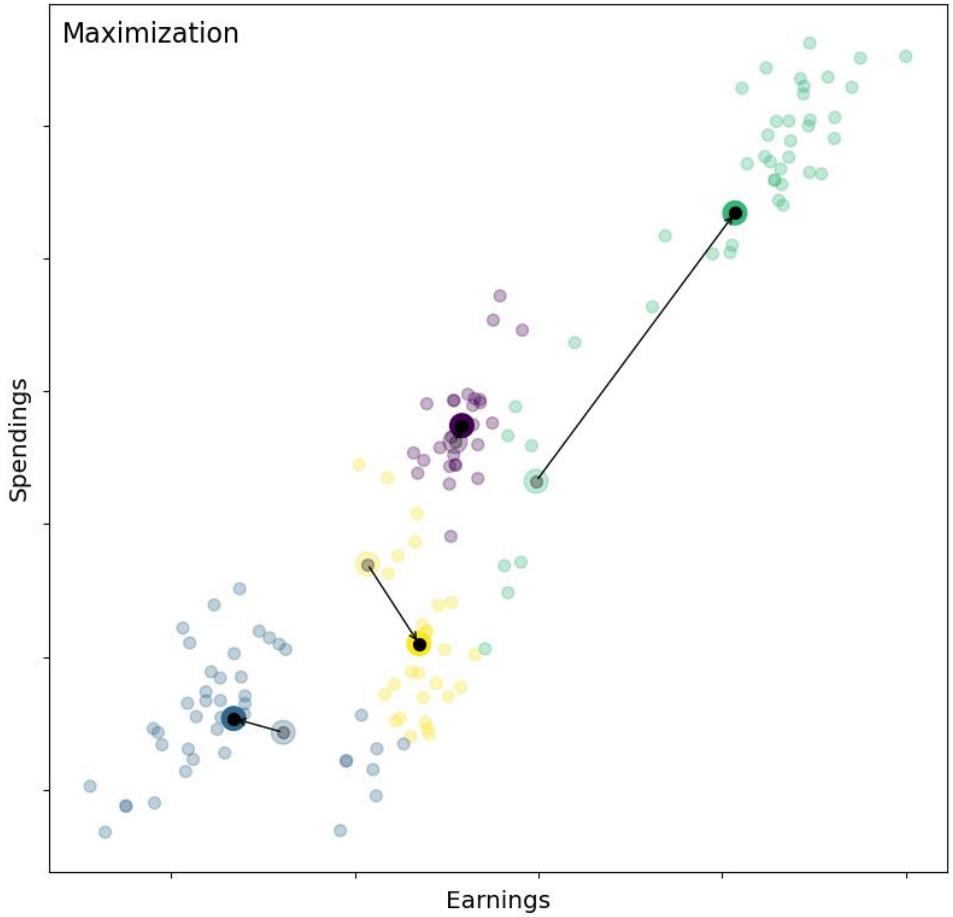
Unsupervised models

Find homogeneous clusters



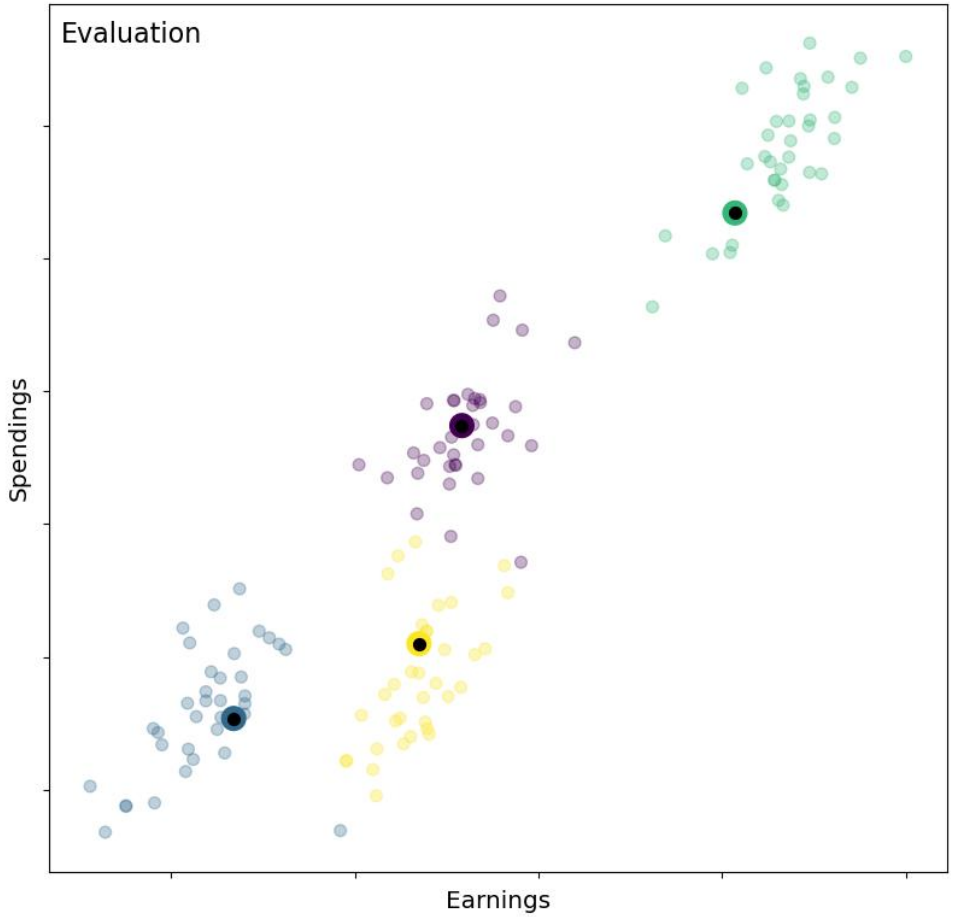
Unsupervised models

Find homogeneous clusters



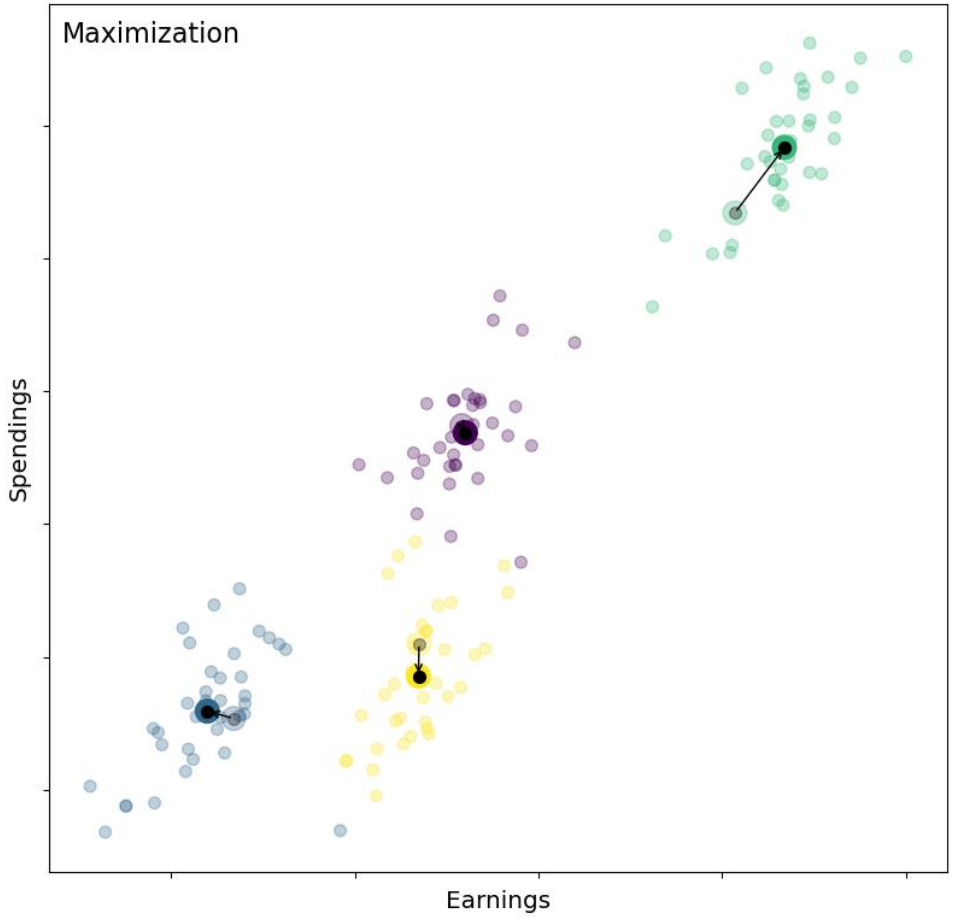
Unsupervised models

Find homogeneous clusters



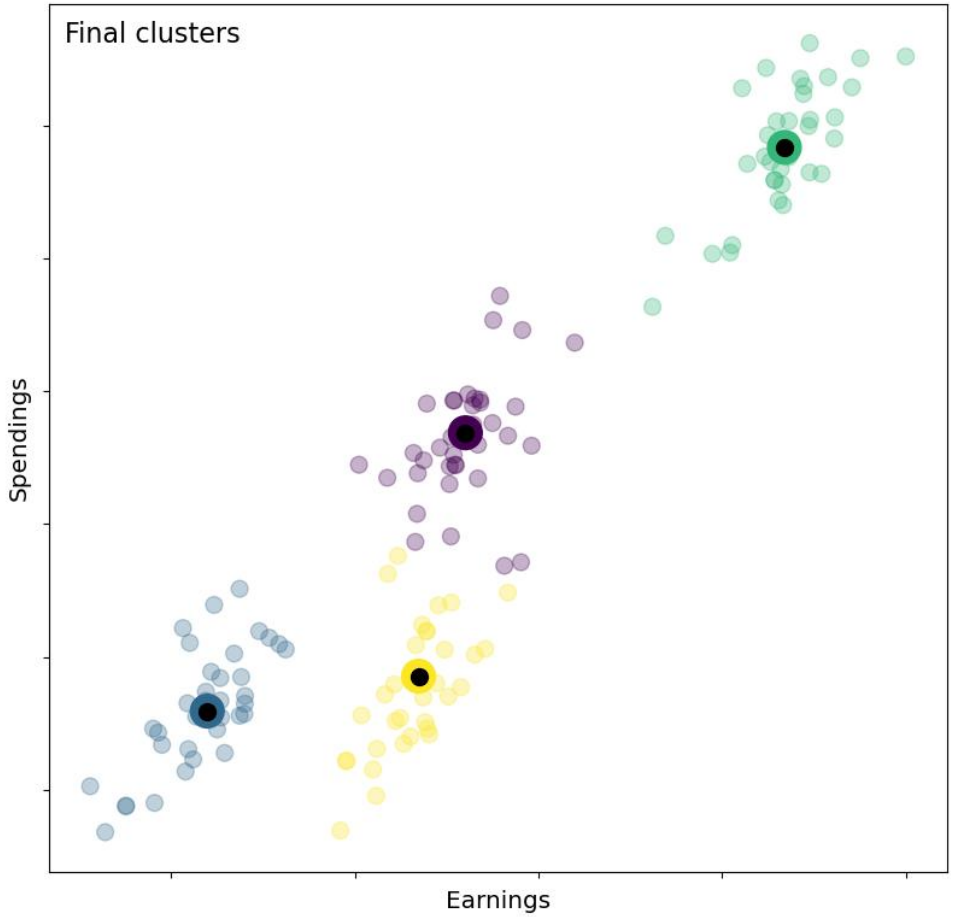
Unsupervised models

Find homogeneous clusters



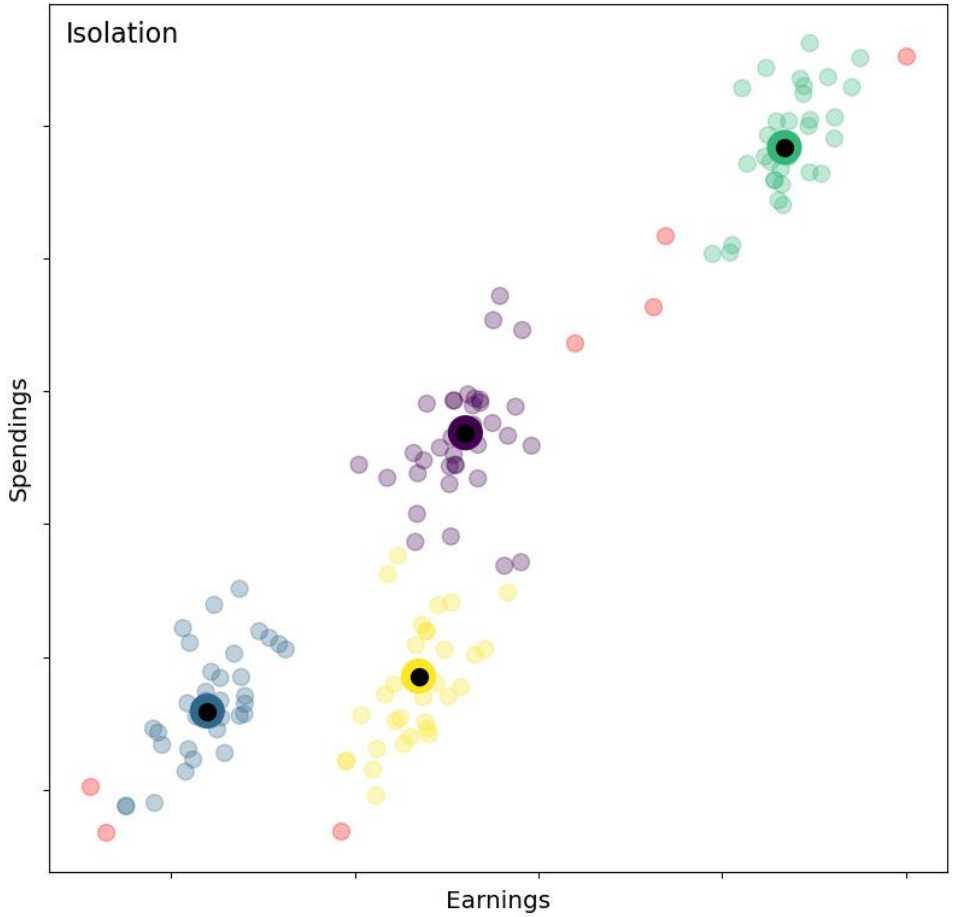
Unsupervised models

Find homogeneous clusters



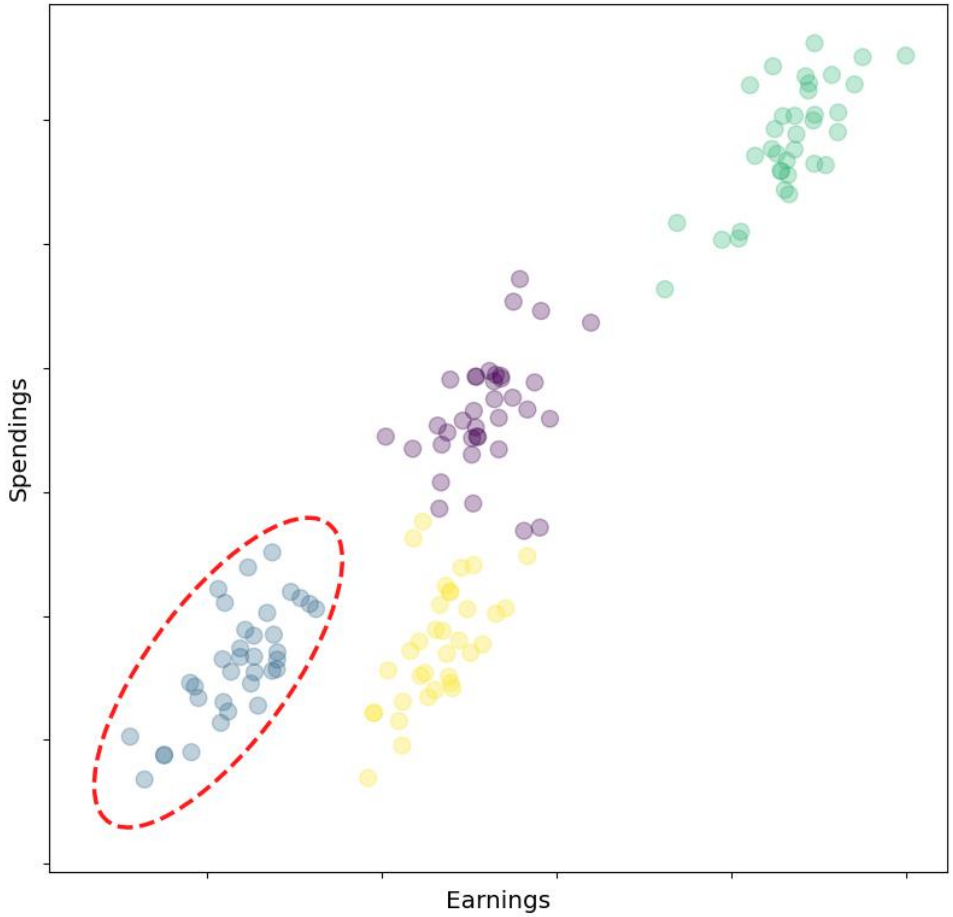
Unsupervised models

Find homogeneous clusters



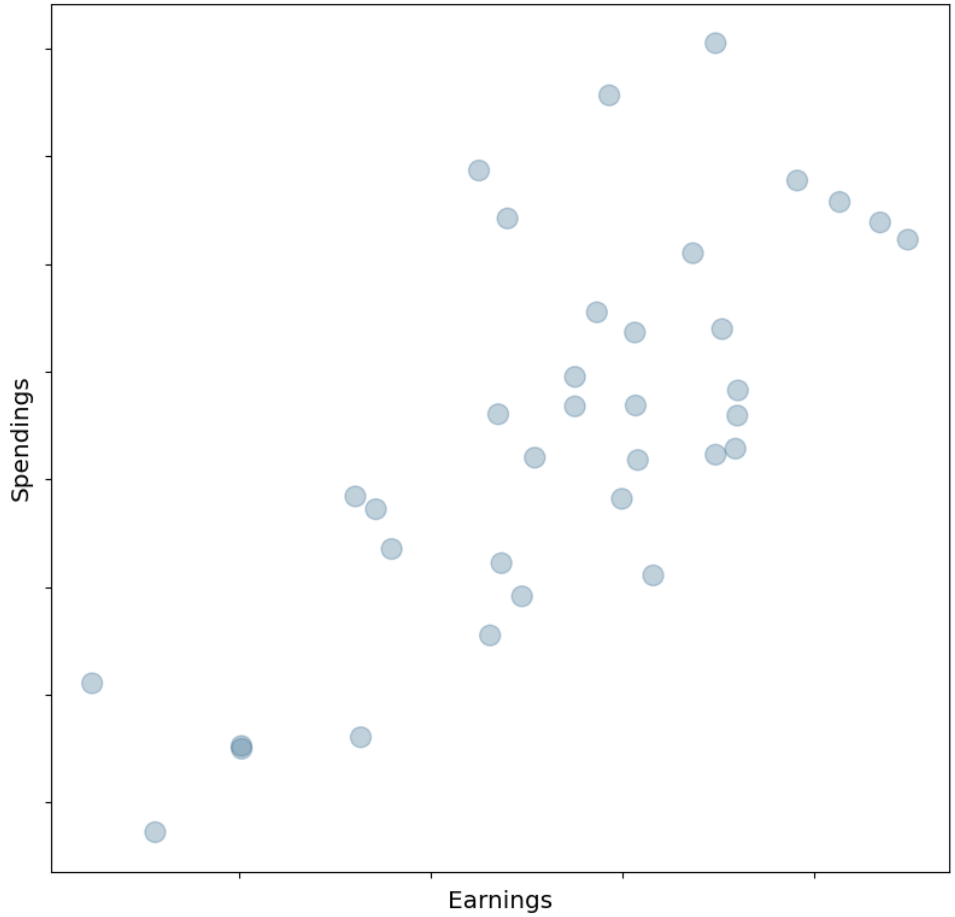
Unsupervised models

Find homogeneous clusters



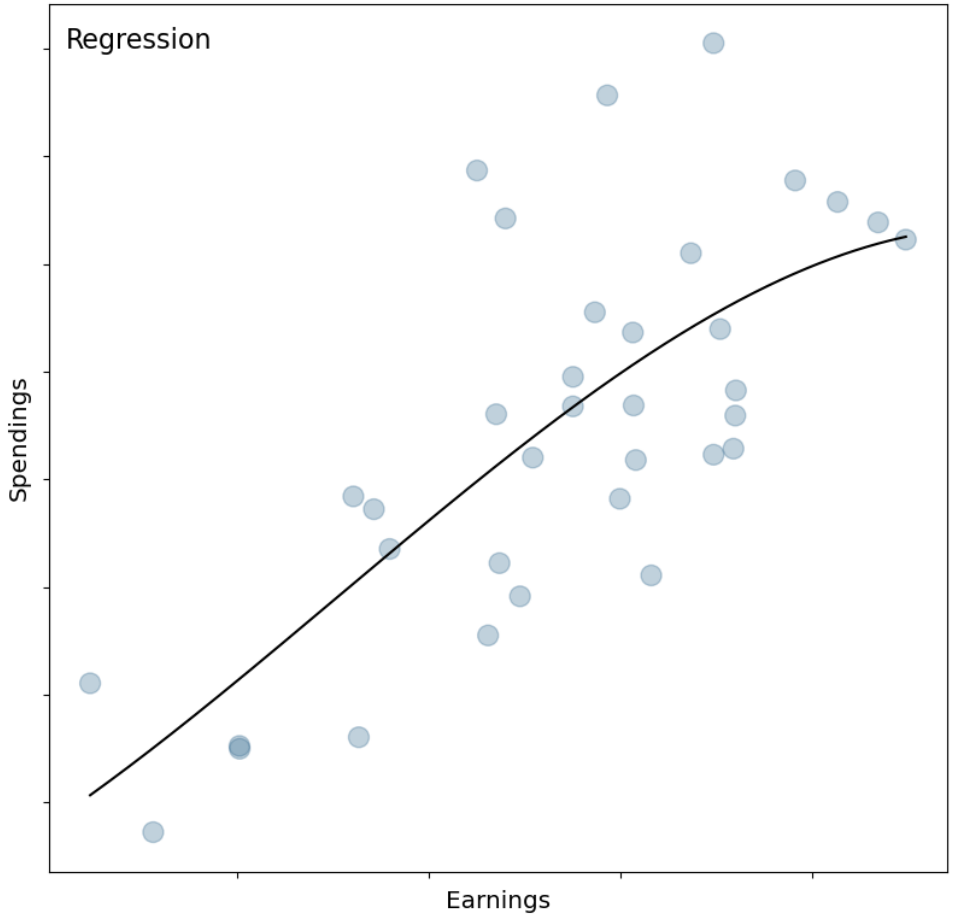
Supervised models

Learn to predict some known quantity, then compare to detect outliers



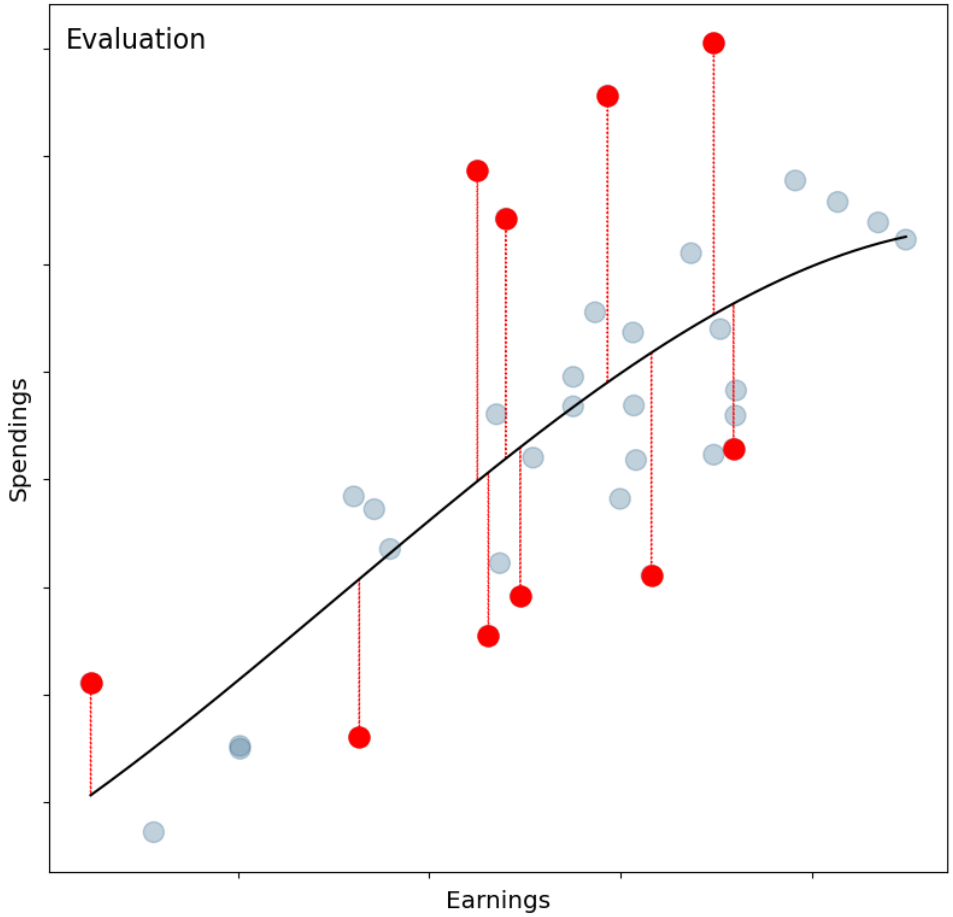
Supervised models

Learn to predict some known quantity, then compare to detect outliers



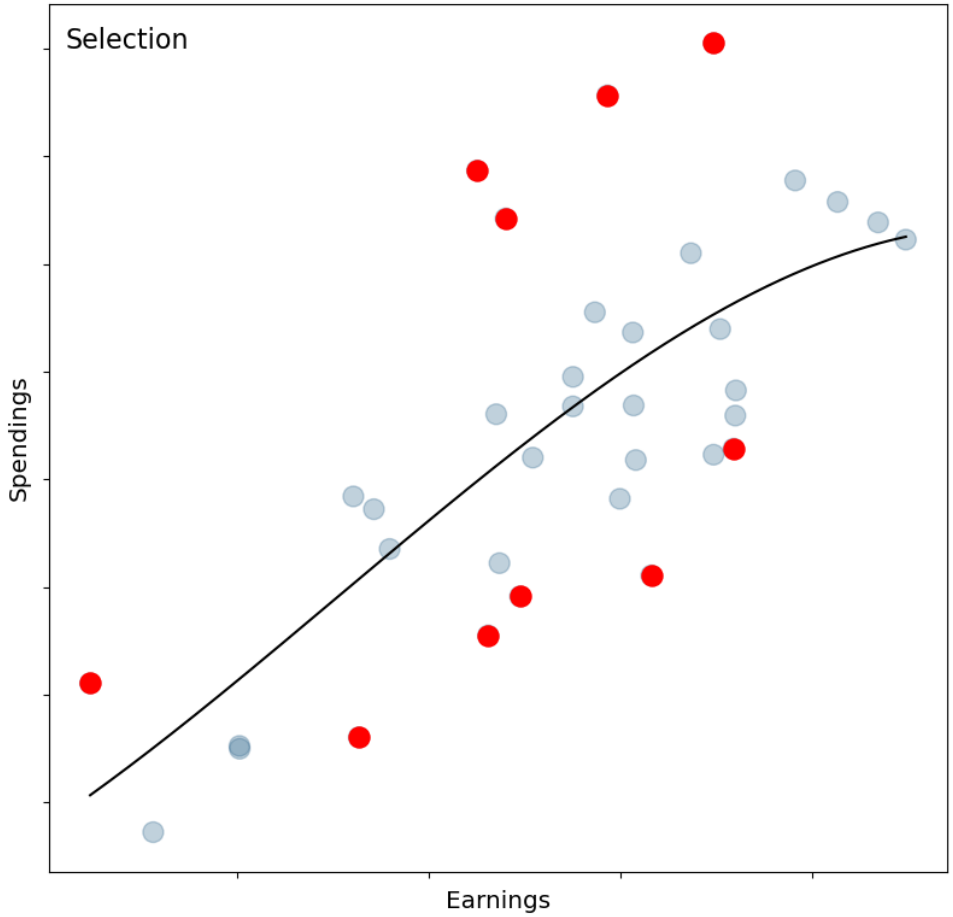
Supervised models

Learn to predict some known quantity, then compare to detect outliers



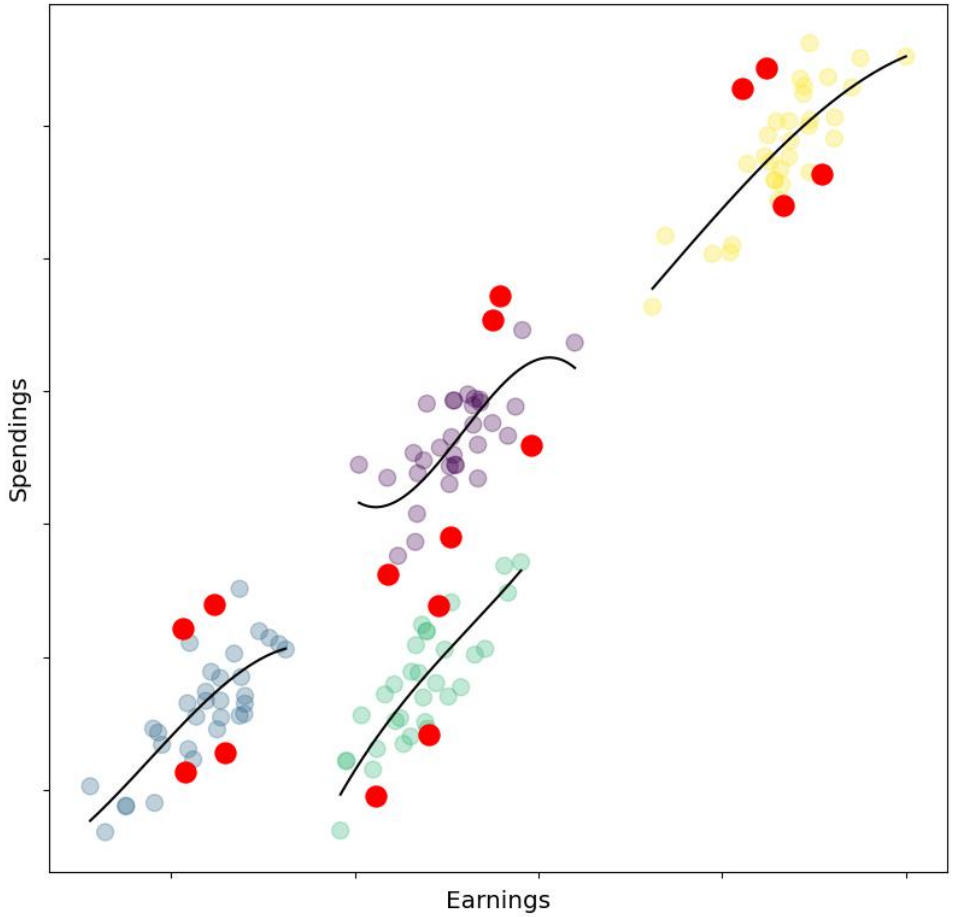
Supervised models

Learn to predict some known quantity, then compare to detect outliers



Supervised models

Learn to predict some known quantity, then compare to detect outliers

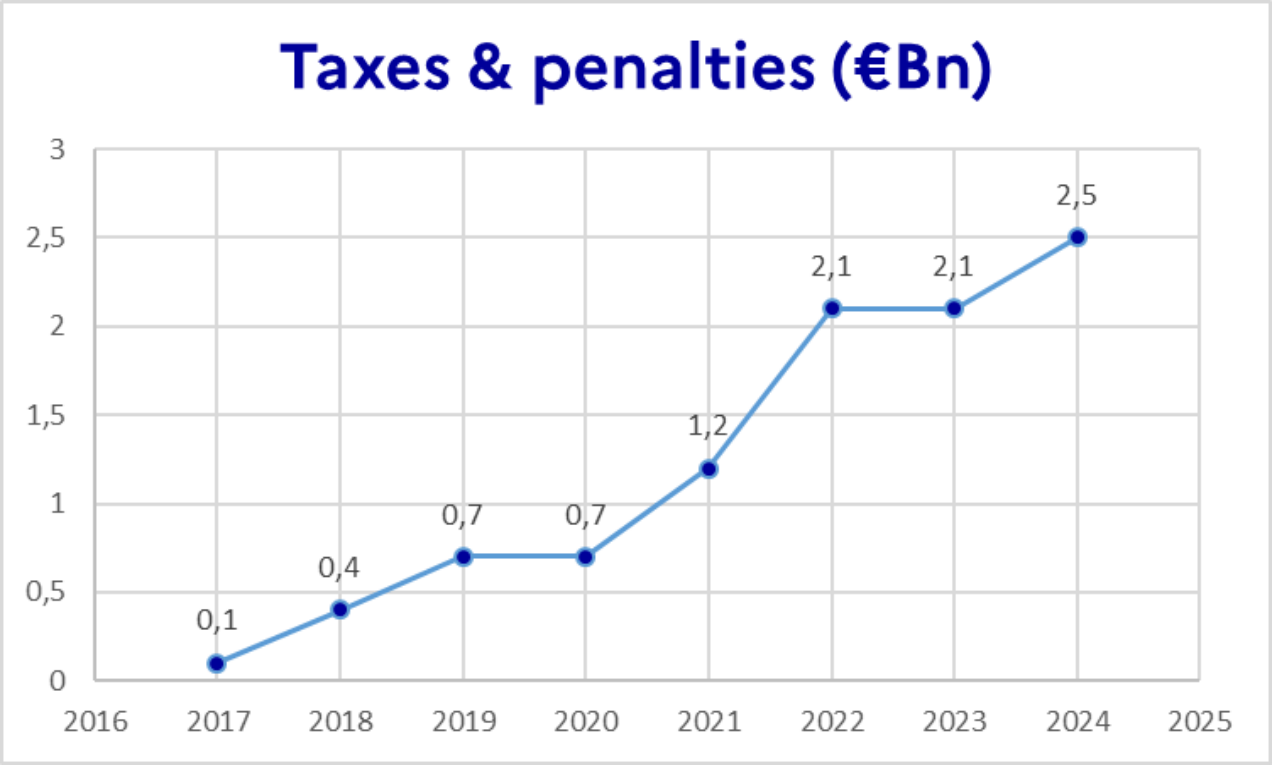


3

Performance & results

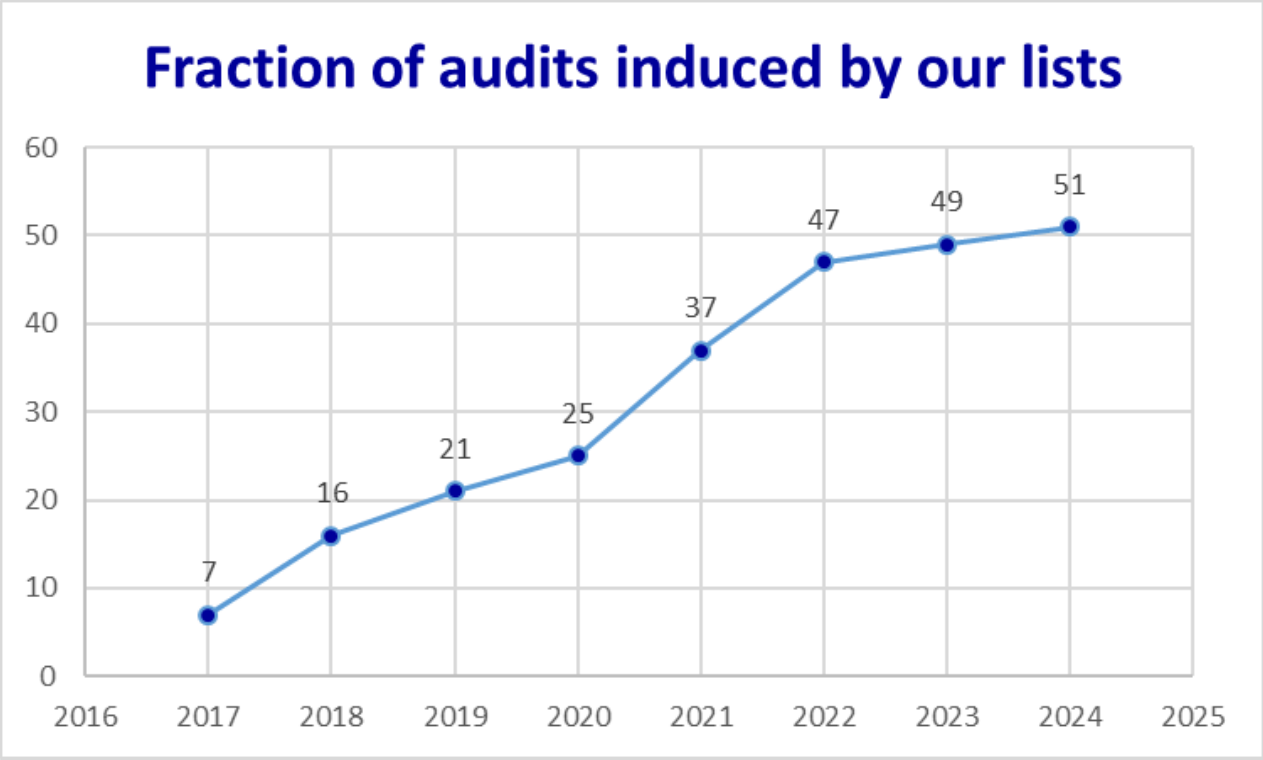
More than 2.5 €Bn detected in 2024

Including 1.6 €Bn from businesses



Our lists are used in more than 50% of tax audits

Other tax audits include research & events



Data analysis setup as of today

270

Topics covered for businesses

60

Topics covered for individuals

287,000

Cases sent through our lists

131,000

Audits completed