Data Science Context at CSO

Growing number of data science projects underway in recent years, in areas such as

- Automated text classification - *NACE, occupation, causes of death*
- Real-time data sources - *traffic at shipping ports, road traffic*
- Earth observation data - *land use, occupancy modelling*
- Web-scraping - *prices, death statistics*
SAS to R migration : 2022-2027

Factors influencing decision include functionality, cost, strategic alignment, risks, benefits.

Challenges include huge volume of legacy code dating back many years

Beginning of migration coincided with widespread adoption of LLMs
SAS to R Transcompiler

Function
Enter SAS code, R code (with comments) is outputted

Features
• Locally hosted using R Shiny
• Wrapper prompt developed in iterative experimental approach
• Code governance via regex: screening for sensitive details
AI and ML for Official Statistics

Four-year ESSnet project with 15 participating countries

13 workpackages, of which

• 6 support the project – coordination, standards, training resources, lab, etc.

• 7 are use cases – text classification, editing, imputation, synthetic data, LLMs, etc.

Overall aim: produce a single point of entry which helps ESS staff to deploy solutions driven by AL and ML.