ICMS Modelling Camp 2025 - Challenge descriptions

Predicting house prices in Birmingham: A data-driven challenge

Facilitator: Diwei Zhou, Loughborough University
Challenge holder: Three Doctors Lets

To help understand the flow of people using A&E services

Facilitator: Xander O'Neill, Heriot-Watt University
Challenge holder: Public Health Scotland

Modelling inflation risk

Facilitator: Edilson Arruda, University of Southampton
Challenge holder: Hymans Robertson

Early warning system for assessing patient health

Facilitator: Ann Smith, University of Huddersfield
Challenge holder: These Hands Academy

Splitting train tickets: searching sparse data for the lowest prices

Facilitator: Lisa Kreusser, University of Bath
Challenge holder: Trainline

Predicting house prices in Birmingham: A data-driven challenge

Facilitator: Diwei Zhou, Loughborough University
Challenge holder: Three Doctors Lets

This project invites PhD students to explore the prediction of residential property prices in Birmingham by identifying or developing suitable predictive models and conducting a benchmarking assessment. Traditional property valuation methods often rely on simplified assumptions that may not fully capture the complex relationships between property characteristics, demographic factors, and market dynamics. This study challenges students to apply advanced analytical techniques to improve accuracy and interpretability in property price estimation.

The dataset includes detailed property attributes such as type, location, and number of bedrooms, combined with demographic information covering aspects like age, education, employment rates, and ethnicity. By integrating these diverse datasets, students will have the opportunity to explore various modelling approaches, assess their performance, and determine the most effective techniques for predicting property prices.

Participants will experiment with a range of statistical, machine learning, and AI-based methods, critically evaluating their strengths and limitations through rigorous benchmarking. The project aims to develop a robust predictive framework while providing valuable insights for real estate professionals, policymakers, and urban planners. Through this challenge, students will gain hands-on experience in real-world data analysis, model selection, and performance assessment in an applied research setting.

To help understand the flow of people using A&E services

Facilitator: Xander O'Neill, Heriot-Watt University
Challenge holder: Public Health Scotland

Public Health Scotland (PHS) is Scotland's lead national agency for improving and protecting the health and wellbeing of Scotland's people. PHS provides data and intelligence in a range of formats from publications to dashboards. Our intelligence enables our customers to inform how the system works and helps inform policy, and your work will help change what intelligence we provide. A general, but key, PHS goal is to use data and intelligence findings to support the health and social care system across Scotland and improve the experience that we, as the public, have in the system.

The health and social care system faces many challenges. One key challenge, which we will work towards here, is managing the demand of hospital beds in various regions and determine what key factors can help alleviate pressure on these hospitals. This enables stakeholders, such a health boards or the Scottish government, to plan and adapt for expected future outcomes.

To support these challenges, Public Health Scotland have been building a variety of models to simulate specific parts of the system. This would then work towards creating a Whole System model. We already have a hospital bed focused model that estimates the number of beds required in hospitals by looking at the inflows and outflows of patients in the system and have now built a national A&E model, which we are continuing to develop.

For this challenge we are looking to 'widen' our A&E model, and welcome all ideas that may be useful for us to understand the system. We provide a basic structure to follow, with steps 2 and 3 being optional based on the time available, and to be completed in any order depending on the interests of the group.

1. Create a basic model simulating the number of attendances that we get in A&E departments in Scotland.
2. How does weather influence A&E attendances and waiting times?
3. How do respiratory pathogens influence A&E attendances and waiting times?
4. Bring together everything you have learnt from steps 1-3 and present the case for what help the system needs to deal with the pressures it faces, what potential changes or improvements could be made to help it manage.

Modelling inflation risk

Facilitator: Edilson Arruda, University of Southampton
Challenge holder: Hymans Robertson

Economic scenario generators like Hymans' proprietary Economic Scenario service (ESS) offer a simulation-based tool for risk management. By simulating a joint distribution of key economic and financial variables, they allow us to assess and manage financial risk in the insurance and pensions industries. Statistical models for inflation lie at the heart of most economic scenario generators.

When modelling inflation rates, we distinguish between implied and realised inflation. Implied inflation is a forward-looking metric which can be thought of as expected inflation. This can be observed in the market and is often presented as the market-implied forward inflation curve. Data on market implied inflation curves is available from the Bank of England. Realised inflation is a backward-looking measure of prices and is often presented as a price index. The Office of National Statistics maintain a range of price indices including the Retail and Consumer Price Indices.

The research question for this project is: how well do market implied forward inflation curves predict the future level of a price index?

The analysis will focus on fitting and assessing a range of candidate models for inflation rates, as measured by a price index, at various time horizons. For example, how does a simple auto-regressive process compare to more complex Stochastic Volatility Diffusion with jumps? How does our analysis change if we need to model a joint distribution over several price indices? Because these models are designed for risk management, it will be crucial to understand how well they model the upper and lower tails of the distribution of future inflation rates.

Early warning system for assessing patient health

Facilitator: Ann Smith, University of Huddersfield
Challenge holder: These Hands Academy

Early warning score (EWS) systems are important tools in helping to identify patients at risk of deterioration and in escalating them to get appropriate treatment as promptly as possible. However, the current use of different scoring systems across care settings is detrimental to patient safety. Staff moving between care settings end up speaking at cross purposes, warning signs are missed, and patient care can ultimately be compromised.

The National Early Warning Score (NEWS) represents an opportunity to maximise the benefits that EWS systems can bring, by ensuring that staff across the NHS and beyond operate using the same language to provide patients with the right care at the right time.

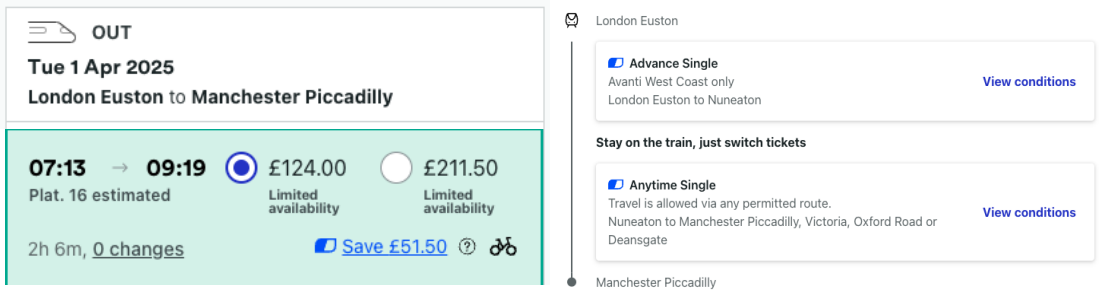
Care providers use routine health checks, measuring vitals such as blood pressure, oxygen levels, heart rate, and temperature, to monitor patient well-being. The National Early Warning Score (NEWS2) is a widely used system that helps assess patient risk based on these measurements. However, interpreting these scores alongside other clinical concerns remains a challenge.

Our client Tara Marshall of These Hands Academy has supplied a coded real-world health data set. Your task is to use the data, alongside appropriate research in the domain, to determine how different factors contribute to early detection of deterioration in care settings and to validate the effectiveness of the NEWS system.

Splitting train tickets: searching sparse data for the lowest prices

Facilitator: Lisa Kreusser, University of Bath
Challenge holder: Trainline

Every year Trainline saves its customers £344 million on train tickets through Split Ticketing. This involves splitting tickets into smaller units that together are cheaper than the total price for a journey.



Finding these Split ticket opportunities is not easy. Exhaustive search is not an option. Instead, we must use graph theory, machine learning and other techniques to find the best combinations.

In this challenge you will be given data including train network data, pricing data and user searches. You are also welcome to use external public datasets. You will build a recommender that can identify the best places to split a journey in order to maximise ticket savings. Whether through historic data or simulation, you will need to show that your approach can effectively find Split ticket opportunities and is computationally feasible.

Participants in this challenge acknowledge that Trainline may take ideas forward for further consideration.