# CASA Data Hackathon 2024

## Description

CASA Data Hackathon is a three-day event focused on urban spatial data science, organised by UCL Centre for Advanced Spatial Analysis (CASA) and supported by the West Midlands Fire Service (WMFS), the second largest fire service in the UK.

During the event, participants will work in diverse teams to tackle real-world urban problems using the skills they have learned during their journey at CASA - exploring spatiotemporal datasets, developing data-driven solutions, and presenting their findings. By immersing students in real-world challenges, this event aims to enhance students' practical skills, foster interdisciplinary collaboration, and provide hands-on experience in tackling urban problems.

This collaborative environment will be closely supported by PhD students acting as mentors, providing guidance and expertise. Therefore, no previous hackathon experience is required, and everyone is welcome.

Evaluations will be led by a group of academic partners and industry experts from WMFS, focusing on both academic innovation through presentation and performance of quantitative models.

**Links**

- Kaggle (for task 1): https://www.kaggle.com/t/d317903b41794a42a7e79043521a1b0a
- OneDrive (for task 2 and 3): https://liveuclac-my.sharepoint.com/:f:/g/personal/ucfnaaj_ucl_ac_uk/Ema5s9V7e5VOlIZHLXGEsCIBUs62-dV4FFz1uG1053NZ-A?e=87oflV
- OneDrive presentation submission (for all tasks): https://liveuclac-my.sharepoint.com/:f:/g/personal/ucfnaaj_ucl_ac_uk/EtJxdbCzhz9KiC5alYJeEBEBSeIvDqm5gbM7z8q1VORzfw?e=PbrHFE

## Tasks

Each group is required to complete Task 0 and choose one additional task from Task 1, Task 2, or Task 3.

### Task 0: Identification of Spatial and Temporal Patterns

Analyze the spatial and temporal patterns of fire events. Identify trends, anomalies, and key insights that can help understand the distribution and frequency of fire incidents over space and time.

### Task 1: Spatiotemporal Prediction

- Predict the 'driving_seconds' – the duration of driving from the fire station to the incident site.
- Submit your prediction on Kaggle.

### Task 2: Spatial Optimization

- **Station Closure:** Due to budget constraints, WMFS has to close a certain number of fire stations. Determine which stations should be closed to minimize the impact. Demonstrate cases where n = 1, 2, 3, ...
- **Station Relocation:** WMFS can relocate a certain number of fire stations. Identify which stations should be moved and to where. Demonstrate cases where n = 1, 2, 3, ...
- **Vehicle Allocation:** WMFS plans to purchase a total of n new PRL vehicles and m new BRV vehicles for their current fire stations. Propose an optimal allocation strategy for these new vehicles.

### Task 3: Visualization

Develop an (interactive) dashboard to visualize the data. Provide insights that can assist WMFS in decision-making processes. Highlight key patterns, trends, and actionable insights through your visualizations.

## Submission

> **Important**
>
> Deadline: Friday midday (1 hour before presentation starts)

For Groups doing task 1, Please submit your final prediction and select your 2 scored private submissions by the deadline.

For all groups, please submit your presentation slides (and/or any other files that you will present, such as html files for task 3) into this OneDrive folder. You will use these materials for your presentation. No changes can be made after the deadline.
Please note, this deadline is strict. We recommend you do not submit at the last minute.

After submission, you may want to enjoy your lunch, and come back to CASA to join the final presentation at 1 pm.

## Evaluation

### Kaggle leaderboard (task 1 only)

Submissions are evaluated on mean squared logarithmic error (msle). This is a non-negative float value – a lower value means a higher accuracy. The best value is 0.

**Submission File:** For each Incident_Number in the test set, you must predict a value for the driving_seconds variable. The file should contain a header and have the following format:

```
Incident_Number,driving_seconds
382513,100
382514,200
382515,300
etc.
```

### Presentation (all groups)

Each group have 6 minutes presentation + 5 minutes Q&A. We will stop the presentation 6 minutes after the presentation.

Presentations will be evaluated by a group of academic partners and industry experts from WMFS.

We do not require every team member to present. The presenter(s) do not need to introduce background, datasets, or details of the methods, as the audiences know the background and datasets.

Presentation will be marked on the following five criteria, with 20% each:

- Insights & Relevance: Evaluate the connection and impact on social issues and practices.
- Clarity of Explanation: Assess the clarity and narrative of the presentation.

- Accuracy: Measure the correctness and precision of the model and findings.
- Technical Difficulty & Innovation: Rate the level of innovation and the complexity of methods used.
- Visualisation: Judge the quality and effectiveness of visual representations.

## Overall evaluation

For Task 1: Overall score = 40% accuracy score + 60% presentation.

For other tasks: 100% presentation.

Each group must specify which task they are presenting on the first page of the slides.

> **Important**
>
> These datasets are only used for CASA-WMFS Hackathon. Please do not use the datasets for other purposes and do not share the datasets with other people without permission from West Midlands Fire Services (WMFS) and CASA.

# Data Description (Task 1 data on Kaggle)

## Files

- **train.csv**: historical incident records, from 2015 to 2022.
- **test.csv**: historical incident records in 2023.
- **station_locations.csv**: locations and other attributes of each fire station at WMFS.

## Columns

### train.csv and test.csv:

- **Incident_number:** An unique number for each incident.
- **callsign_type:** Refers to the type of vehicle the row refers to.
- **call_time:** timestamp of the fire incident, in format of dd/mm/yyyy hh:mm, e.g. 01/01/2009 00:00
- **callsign_station:** Name of the fire station.
- **initial_incident_type:** Initial label of incident.
- **incident_classification_label:** Detailed label of incident.
- **incident_profile_label:** Another label of incident.
- **incident_classification_level1:** Type of incident, including levels of FALLS_RESPONSE, FALSE_ALARM, FIRE, OTHER, RTC, SSC.

  > **Note**
  >
  > - **FALLS_RESPONSE** is not a typo, it refers to calls to assist people who have fallen.
  > - **RTC** is Road Traffic Collision, basically just a car crash.
  > - **SSC** is Special Service Call, which is almost everything except RTCs and fires that the fire service responds to, such as floods, people trapped in lifts and hazardous materials incidents.

- **call_seconds:** The time between the call being taken and the fastest vehicle being notified.

- **reaction_sections:** The time between this vehicle being allocated for this incident and this vehicle getting ready.
- **EASTINGS:** The easting coordinate of the incident, under British National Grid projected coordinate system (EPSG:27700), in unit of meter.
- **NORTHINGS:** The northing coordinate of the incident, under British National Grid projected coordinate system (EPSG:27700), in unit of meter.

  > **Note**
  >
  > Note that the procedure of **geospatial obfuscation** has been applied to the original incident location in order to safeguard sensitive location data. That is, given an incidence with location ($x$, $y$) a random value $d$ between 500m and 1000m and a random angle $a$ between 0 and 360 is generated, this incidence is moved to the new location ($x + d * \sin a$, $x + d * \cos a$).

- **driving_seconds (only in train.csv):** Duration of this vehicle driving from the fire station to the incident site.

## station_locations.csv:

- **Station name:** Name of each station.
- **Easting:** The easting coordinate under British National Grid projected coordinate system (EPSG:27700), unit of meter.
- **Northing:** The northing coordinate under British National Grid projected coordinate system (EPSG:27700), unit of meter.
- **PRL_Count:** Number of PRL (one type of fire engines, see notes below) in this station.
- **BRV_Count:** Number of BRV (another type of fire engines) in this station.
- **Closed (Y/N):** Whether this station has been closed. N means the station is being used.
- **Opened:** Date of open, if applicable.
- **Closed:** Date of close, if applicable.

> **Note**
>
> **Details of PRL and BRV**
>
> - a PRL (Pump Rescue Ladder) is a traditional fire truck and can attend all incident types.
> - A BRV (Brigade Response Vehicle) is a 4x4 vehicle that is dispatched to 'lower' risk incidents. The aim is increase the availability of the PRLs for more serious incidents such as house fires, which the BRV would not be as equipped to tackle.
> - A PRL typically has 4-5 fire fighters, while a BRV typically has 2-4.
> - An image of one of WMFS BRVs is


# Data Description (Task 2 and 3 data on OneDrive)

Please refer to **Notes** in the above section.

## Files

- **wmfs_incidents.csv**: historical incident records, from 2010 to 2023.
- **wmfs_mobilisations.csv**: historical vehicle mobilisation records from 2010 to 2023. This dataset contains all information of the wmfs_incidents, so it is not needed to link the two datasets.
- **station_locations.csv**: locations and other attributes of each fire station at WMFS.

## Columns

### wmfs_incidents.csv:

- **call_time:** timestamp of the fire incident, in format of dd/mm/yyyy hh:mm, e.g. 01/01/2009 00:00
- **incident_classification_label:** Detailed label of incident.
- **incident_profile_label:** Another label of incident.
- **incident_classification_level1:** Type of incident, including levels of FALLS_RESPONSE, FALSE_ALARM, FIRE, OTHER, RTC, SSC.
- **prl_count:** number of PRL engines sent for this incident.
- **brv_count:** number of BRV engines sent for this incident.
- **EASTINGS:** The easting coordinate of the incident, under British National Grid projected coordinate system (EPSG:27700), in unit of meter.

- **NORTHINGS:** The northing coordinate of the incident, under British National Grid projected coordinate system (EPSG:27700), in unit of meter. [see notes on **geospatial obfuscation** in the above section]
- **call_seconds:** the time between the call being taken and the fastest vehicle being notified.
- **reaction_seconds:** duration of reaction, i.e. the time between a resource being allocated for this incident and the resource getting ready.
- **driving_seconds:** duration of driving from the fire station to the incident site (2023 data is removed).

## wmfs_mobilisations.csv:

- **Incident_number:** If two rows have the same incident_number then that means that there were multiple vehicles going to the same incident. This column is an unique number for each incident and should not be used to link to wmfs_incidents.
- **callsign_type:** refers to the type of vehicle the row refers to.
- **call_time:** the time when this call was made.
- **callsign_station:** Name of the fire station.
- **initial_incident_type:** Initial label of incident.
- **incident_classification_label:** Detailed label of incident.
- **incident_profile_label:** Another label of incident.
- **incident_classification_level1:** Type of incident, including levels of FALLS_RESPONSE, FALSE_ALARM, FIRE, OTHER, RTC, SSC.
- **call_seconds:** refers to the time between the call being taken and that specific vehicle being notified (rather than the fastest).
- **reaction_sections:** the time between this vehicle being allocated for this incident and this vehicle getting ready.
- **driving_seconds:** duration of this vehicle driving from the fire station to the incident site.
- **on_scene_seconds:** duration of this vehicle on scene.
- **EASTINGS:** the easting coordinate of the incident, under British National Grid projected coordinate system (EPSG:27700), in unit of meter.
- **NORTHINGS:** the northing coordinate of the incident, under British National Grid projected coordinate system (EPSG:27700), in unit of meter.

## station_locations.csv:

- **Station name:** Name of each station.
- **Easting:** The easting coordinate under British National Grid projected coordinate system (EPSG:27700), unit of meter.
- **Northing:** The northing coordinate under British National Grid projected coordinate system (EPSG:27700), unit of meter.
- **PRL_Count:** Number of PRL (one type of fire engines) in this station.
- **BRV_Count:** Number of BRV (another type of fire engines) in this station.
- **Closed (Y/N):** Whether this station has been closed. N means the station is being used.
- **Opened:** Date of open, if applicable.
- **Closed:** Date of close, if applicable.

# Open Data

You may want to combine with other open data. Here are some examples:

- Road network data: Ordnance Survey (OS) MasterMap Highways - Roads, available on Digimap (https://digimap.edina.ac.uk/, login with your UCL account)
- Traffic speed data: Pilot Speed Trail Products -> average speed and speed limits, available on Digimap
- POI data: Ordnance Survey (OS) Boundary and Location Data -> Points of Interest, available on Digimap
- Dwelling data: Available on CDRC (https://data.cdrc.ac.uk/dataset/dwelling-ages-and-prices)
- Census data: https://www.nomisweb.co.uk/sources/census

# Timeline

## Day 1: Wednesday 5 June

10:00: Kick-off and introductions at CASA (hybrid).

After introductions: Hackathon starts! CASA rooms 106/107 will be open for group work all day.

15:00-16:00: Hybrid optional drop-in sessions. Two PhD mentors will be available in-person at CASA 106/107 and online via teams meeting.

## Day 2: Thursday 6 June

CASA will be occupied until 14:30 by other events. We encourage you to find a suitable workspace on campus. After 14:30, CASA rooms 106/107 will be open for you.

15:00-16:00: Hybrid optional drop-in sessions. Two PhD mentors will be available in-person at CASA 106/107 and online via teams meeting.

## Day 3: Friday 7 June

Morning: CASA 106/107 will be open for group work.

13:00-16:00: Final presentations and award ceremony. We will wrap up by 16:00, and head to a pub for a celebratory drink!