

Hack the Crash!

Hackers handbook

11-13 Oct – HackUPC

CONFIDENTIAL AND PROPRIETARY

Any use of this material without specific permission of McKinsey & Company
is strictly prohibited

Content

Problem statement

Data Dictionary

Hack Submissions

Content

Problem statement

Data Dictionary

Hack Submissions

Problem statement

Road transport in UK is very popular for various reasons, but the **rate of road-accidents and fatality in the country is high**. Pressure on roads has been on increase and the **number of vehicles is increasing by leaps and bounds**

McKinsey, as a trusted advisor for both private and public sector, was asked to **help solving the problem by predicting accident severity**. For that task, they need the best team of data scientists – that is you



Problem statement

As data scientists, you are now responsible for developing a **model predicting accident severity** based on:



Road location



Road type
and its
speed limit

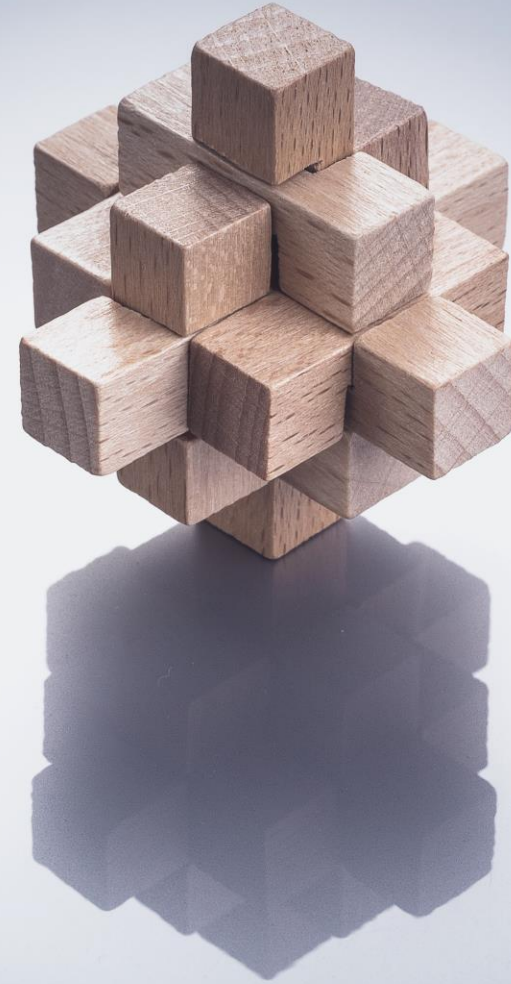


Weather
conditions



Details of
vehicles
involved in
the accident

Predicting accident severity will **help the authorities closely monitor accident prone road segments** and do a **root cause analysis** for such mishaps. This will also help them in planning the setup of emergency infrastructure such as hospitals, police stations etc.



Evaluation metrics

The solution will be evaluated in the following dimensions

F1-score

How accuracy does the model perform on private test dataset, according to F1-score?



Clarity

Is the source code and documentation structured in a concise and readable way?



Novelty

How novel is the solution in terms of feature engineering, modelling, validation, etc?



Insights & business plan

What insights are discovered during the hackathon? What are the actionable points to make the model useful in production? What business plan would you propose in order to solve this challenge in real life? How would you translate your results into an actionable plan? How would you include software and IT infrastructure to create an automated solution?



Prize

Nintendo® Switch lite per participant of the team



Content

Problem statement

Data Dictionary

Hack Submissions

Data dictionary – Accident data

Variable	Definition
Accident Index	Unique ID for accident
Number of Vehicles	Number of vehicles involved in accident
Number of Casualties	Number of casualties involved in accident
Light Conditions	Lighting conditions at the junction during accident
1st Road Class	Category of 1st road of junction
Road Type	Type of road
Junction Detail	Details of Junction
Junction Control	Traffic control facilities at the junction
2nd Road Class	Category of 2nd road of junction
Pedestrian Crossing-Human Control	Status of human-controlled pedestrian crossing
Pedestrian Crossing-Physical Facilities	Facilities for pedestrian crossing
Road Surface Conditions	Road surface conditions
Special Conditions at Site	Other conditions at the site of accident
Carriageway Hazards	Obstacles on road
Police Force	Police Force for the area (eg. Essex, South Wales, Metropolitan Police, etc)
Location Easting OSGR (Null if not known)	Location variable

Data dictionary – Accident data (Cont'd)

Variable

Definition

Location Northing OSGR (Null if not known)

Location variable

Longitude (Null if not known)

Location variable

Latitude (Null if not known)

Location variable

Local Authority (District)

Local authority for the district

Local Authority (Highway Authority – ONS code)

Local authority for the highway

Urban or Rural Area

Area of accident

Speed limit

Vehicle speed limit

LSOA of Accident Location

Location variable, Lower Layer Super Output Areas of accidents

1st Road Number

Code for the 1st road of junction

2nd Road Number

Code for the 2nd road of junction

Date (YYYY-MM-DD)

Date of accident

Time (HH:MM)

Time of Accident

Weather Conditions

Weather conditions during accident

Target

(Target) Severity of accident

Data dictionary – Vehicle data

Variable	Definition
Accident Index	Unique ID for vehicle
Vehicle Manoeuvre	Vehicle manoeuvre during accident
Skidding and Overturning	Skidding or overturning of vehicle during accident
Hit Object in Carriageway	Vehicle hit any object in the road
Hit Object off Carriageway	Vehicle hit any object off the road
Journey Purpose of Driver	Journey Purpose of Driver
Sex of Driver	Sex of driver
Age of Driver	Age of driver
Vehicle IMD Decile	The English Index of Multiple Deprivation (IMD) for the vehicle
Driver IMD Decile	The English Index of Multiple Deprivation (IMD) for the driver
Driver Home Area Type	Type of driver's native area
Vehicle Location-Restricted Lane	Location of vehicle on restricted lanes
Junction Location	Location of vehicle at junction
Vehicle Leaving Carriageway	Vehicle Leaving Carriageway

Data dictionary – Vehicle data (Cont'd)

Variable	Definition
Vehicle Type	Type of vehicle
Towing and Articulation	Type of tow/articulation
1st Point of Impact	Point of impact for vehicle
Was Vehicle Left Hand Drive	Binary variable (Yes/No)
Engine Capacity	Vehicle's engine capacity in cubic centimetres
Vehicle Propulsion Code	Fuel type of vehicle
Age of Vehicle (manufacture)	Age of vehicle in years
Vehicle Reference	Reference number for vehicles

Content

Problem statement

Data Dictionary

Hack Submissions

Submit the hack

Registration

1. Set up an account with the information you will receive on your email
2. Sign-up on the Box folder
3. Access all information there. You will find the following files:
 - Training:
 - » accidents.csv
 - » vehicles.csv
 - Test: test.csv
 - Booklet with problem statement – this document

Access Problem Statement/ dataset and build solution

You can submit your solutions multiple times via the file sharing platform

Name your final submission as Final Version and your team members

Assessment

We will run a script that will assess your F1 scores

Your business case will also be assessed on both format and content

Submit the hack (Cont'd)

Type of Files required in a regular submission

A valid submission contains predicted values for the test set. It is a csv file

Participants can make multiple submissions during the contest

What is a Final Submission and how to set it?

Final submission must contain the files mentioned below:

- Final submission file in csv format that contains predicted values of test set
- Zip file containing (Upload under code file section)
 - » Code file to generate the same results as submission.csv file
 - » A Presentation/Document explaining the approach
 - › A brief on the approach, which you used to solve the problem.
 - › What data-preprocessing / feature engineering ideas really worked? How did you discover them?
 - › What does your final model look like? How did you reach it?
 - › What are your takeaways from the challenge?
 - › What actions would you propose in order to solve the problem you are facing?
 - › What would be a tentative business plan?
 - › How would you include the use of software and IT infrastructure in your solution?
 - › What would be the economic impact of your solution?
 - › What other applications do you think these results/data could serve to?
 - › How would you translate these insights into technology improvements?

