Data lessons from the *Understanding Vehicle Safety Inequalities* project

Joining the Dots, Birmingham
3 March 2026

Data lessons from the *Understanding Vehicle Safety Inequalities* project

1. Understanding Vehicle Safety Inequalities
2. The current vehicle data landscape
3. Cost-efficient matching of S19 data
4. Reviewing VRM and VIN in S19 data
5. A MAST instance for vehicle characteristics

RAC Foundation

Data lessons from the *Understanding Vehicle Safety Inequalities* project

1. **Understanding Vehicle Safety Inequalities**
2. The current vehicle data landscape
3. Cost-efficient matching of S19 data
4. Reviewing VRM and VIN in S19 data
5. A MAST instance for vehicle characteristics

Understanding Vehicle Safety Inequalities

- Sponsored by Road Safety Trust, work done by RAC Foundation and Agilysis

- bit.ly/4cQiVJD

Understanding Vehicle Safety Inequalities

- Sponsored by Road Safety Trust, work done by RAC Foundation and Agilysis

- bit.ly/4cQiVJD
- **We are making a MAST instance with additional vehicle variables**

Data lessons from the *Understanding Vehicle Safety Inequalities* project

RAC
Foundation

agilysis

The current vehicle data landscape

- S19: Make, Model (generic), years since manufacture, fuel type, engine displacement, vehicle type
- Meanwhile, there is:
  - DVLA data
  - DVLA/OEM-derived commercial datasets
  - DVSA MOT records (all non-bus/HGV vehicles after early 2000s)
- How do we link S19 to these richer datasets?

Data lessons from the *Understanding Vehicle Safety Inequalities* project

agilysis

RAC
Foundation

Cost-efficient matching of S19 data

- There is some amazing data, *but it can cost a lot*
- How to match in a way that's:
  - Reliable
  - Repeatable
  - GDPR compliant
  - Affordable
  - Efficient
- It all hinges on VRM and Vehicle Identification Number (VIN)

RANGE ROVER

22nd Jul, 2023 10:00

Windsorview Lakes 22nd July 2023 Sale

Lot 172

ⓘ Hover to zoom

2004 Range Rover believed to be Ex-HM Queen Elizabeth II

Sold £33,002

Lot details

Year: 2004
Make: Range Rover
Model: believed to be Ex-HM Queen Elizabeth II
Registration: BN04 EPU
Chassis No: SALLMAMA34A159508
Engine No: 539031830
Odometer: 109503
(a): have you read the description? could be incredible piece of provenance
(b): Believed to be owned & used by HM Queen Elizabeth II
(c): IT IS A Special order example from Jaguar Land Rover
(d): Serviced local to the Balmoral Estate with some tell tale signs

There can be few cars that are as well-known and iconic as the mighty Range Rover.

22nd Jul, 2023 10:00

Windsorview Lakes 22nd July 2023 Sale

Lot 172

ⓘ Hover to zoom

2004 Range Rover believed to be Ex-HM Queen Elizabeth II

Sold £33,002

Lot details

Year: 2004
Make: Range Rover
Model: believed to be Ex-HM Queen Elizabeth II
Registration: BN04 EPU
Chassis No: SALLMAMA34A159508
Engine No: 539031830
Odometer: 109503
(a): have you read the description? could be incredible piece of provenance
(b): Believed to be owned & used by HM Queen Elizabeth II
(c): IT IS A Special order example from Jaguar Land Rover
(d): Serviced local to the Balmoral Estate with some tell tale signs

There can be few cars that are as well-known and iconic as the mighty Range Rover.

Data lessons from the *Understanding Vehicle Safety Inequalities* project

agilysis

RAC Foundation

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

SALLMAMA34A1159508

UK is the country of origin

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

SAL|LMAMA34A1159508

UK-made Land Rover

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

SALLMAMA34A1159508

Land Rover Model and Spec

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

SALLMAMA34A1159508

The Check Digit

RAC Foundation

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

SALLMAMA3[4A1]159508

Assembly Plant and Model Year

Reviewing VRM and VIN in S19 data

- VIN: 17 characters
- *Fixed* universal vehicle identifier

SALLMAMA34A1159508

Sequential number

Reviewing VRM and VIN in S19 data

- Pre-processing:
  - Missing/invalid VINs
  - S19 cars only (vehicle type 9)
  - S19 VIN gets the same car in the MOT database
  - S19 Make matches Make as per the VIN
  - CO_2 value
  - Cherished plate change looks plausible
- Grouping by Make, Model, Year, CO2, Partial VIN
- Pre-emptive deduplication
- 218,304 to 80,256, 63% reduction

agilysis

RAC
Foundation

Data lessons from the *Understanding Vehicle Safety Inequalities* project

1. Understanding Vehicle Safety Inequalities

2. The current vehicle data landscape

3. Cost-efficient matching of S19 data

4. Reviewing VRM and VIN in S19 data

5. A MAST instance for vehicle characteristics

agilysis

RAC Foundation

A MAST Instance for Vehicle Characteristics

Offline Data Warehouse

Collisions

Vehicles

Casualties
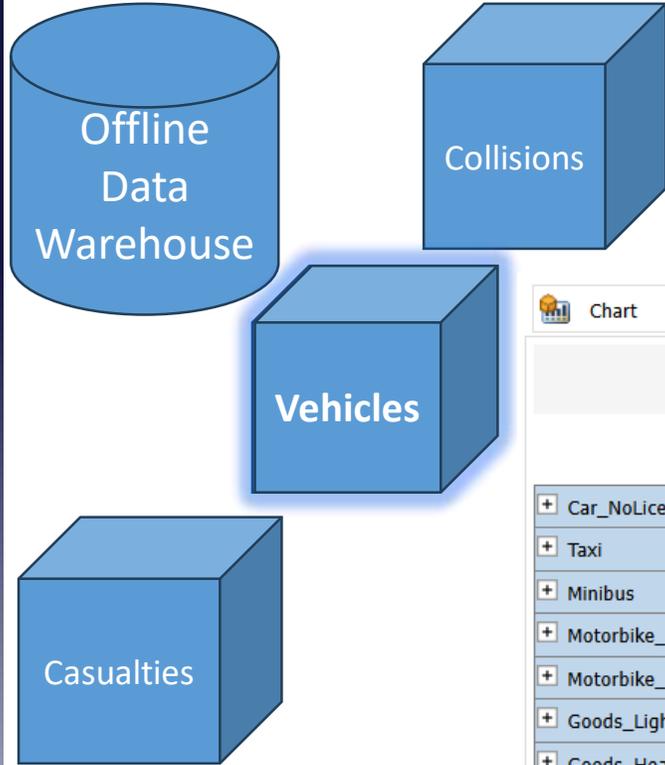
| Chart | Grid | Split | | Reset Zoom |

Vehicles by Vehicle Age by Vehicle Type

Vehicles	0	1	2-3	4-5	6-9	10-19	20-29	30-39	40+	NK	Total
Car_NoLicensedTaxi	6,016	10,680	21,608	28,668	60,426	94,920	6,508	276	116	33,231	262,449
Taxi	149	275	560	1,164	2,161	884	6	1		823	6,023
Minibus	2	10	26	32	68	86	3	1		413	641
Motorbike_UpTo125cc	2,069	3,239	4,735	3,084	3,892	2,399	300	55	57	2,024	21,854
Motorbike_Over125cc	684	998	1,654	1,444	2,093	3,410	1,728	190	164	2,134	14,499
Goods_Light	818	2,326	4,175	3,729	4,853	4,593	280	5	2	3,275	24,056
Goods_Heavy	323	596	1,144	1,162	1,559	469	35	8	1	3,282	8,579
Bus	85	172	454	829	1,846	1,956	62	1	3	780	6,188
Cycle										32,002	32,002
Horse										135	135
Tractor	38	46	102	76	89	111	34	13	8	239	756
Other	40	54	124	138	140	134	36	10	11	2,127	2,814

Edit Slicer Axis Item

Z-A A-Z

- 2011
- 2012
- 2013
- 2014
- 2015
- 2016
- 2017
- 2018
- 2019
- 2020
- 2021
- ☑ 2022
- ☑ 2023
- 2024
- Unknown

OK Cancel

| Rows (Category) | Columns (Series) | Filter (Slicer) |

| Add Dimension | Crash Year : Multiple |

RAC Foundation

A MAST Instance for **Vehicle Characteristics**

Make, Model, Version, Body style, No of doors, Max power kW, Max power hp, Fuel type, No of seats, published kerb weight, overall width (mm), overall height (mm), overall length (mm), Acceleration 0-60mph, Insurance group, Power to Weight ratio, Power steering, Autonomous driving, Night vision, GSR compliant, Accident data recorder, Emergency call, Low tire pressure indicator, Seat belt warning, Speed alert, Traffic sign recognition, Daytime running lights, Distance & object warning technology, Headlight control, Automatic emergency braking, Heads up display, Media control, Touch control, ... airbags, Knee airbags, Head ... number of airbags, ... belts, Anti collision system, Pop up bonnet, ... assistance, Overtaking sensor, Adaptive ... cornering lights/curb illumination, Stability control, Electronic traction control, Lane departure warning, Cruise control, Adaptive cruise control, 360 degree camera, Park distance system, Front Display screen size (cm), number of disc brakes, MOT failure history, MOT compliance at date of crash, odometer estimate at date of crash, odometer comparator value, MOT advisory history, country of manufacture

A MAST Instance for Vehicle Characteristics

Basic vehicle physics continuous data presented in bands

- **Dimensions**
  - overall width
  - overall height
  - overall length
- **Acceleration**
- **Power**
  - Max power kW
  - Max power hp
- **Weight**
  - Published kerb weight
  - Mass in running order
- **Power to Weight ratio**

+ 0		10,224
+ 1		18,396
	2	16,812
− 2-3	3	17,770
	Total	**34,582**
	4	20,110
− 4-5	5	20,216
	Total	**40,326**
	6	20,829
	7	20,376
− 6-9	8	18,840
	9	17,082
	Total	**77,127**

A MAST Instance for Vehicle Characteristics

Vehicle history and economics

- **True or false?**
    - Valid MOT at date of collision?
    - MOT passed *after* date of collision?
    - Compliant with EU General Safety Regulation 2?
- **Vehicle safety** by MOT outcome (as % of tests taken)
    - Advisories
    - Dangerous fail reasons
- **Mileage**
    - Estimated odometer at time of crash
    - Typical odometer for make, model, year
- **Value**
    - P11D
    - Insurance estimate of value at date of collision
- **Insurance Group**

A MAST Instance for Vehicle Characteristics

And much, much more ... categories with examples

- **Assistive technology**
  - Lane assist
- **Braking and handling**
  - Anti-lock braking system
- **Crash protection**
  - Airbags (by position)
- **Messages, data and information**
  - eCall
- **Visibility**
  - Distance and object sensing technology
- **Driver distraction**
  - Text to speech

agilysis

RAC
Foundation

Questions?

Dr Ivo Wengraf FRGS FCIHT
ivo.wengraf@racfoundation.org
www.linkedin.com/in/ivowengraf

Bruce Walton BA AMRGGB
bruce.walton@agilysis.co.uk
www.linkedin.com/in/waltonbruce

agilysis

RAC
Foundation