



A MEMBER OF VOLKSWAGEN GROUP

Practical exercise

Introduction

Please implement a solution Using Python (Jupyter preferably) or R (Rmarkdown), analyzing the road accident safety data found at <https://data.gov.uk/dataset/road-accidents-safety-data>.

- This analysis needs to include a model that predicts if a police officer is likely to attend an accident or not.
- Please focus initially only on the accidents from the 2017 dataset for your analysis
- Deliver a presentation (30 mins), as if you were presenting to a customer, that summarizes your initial findings. The audience will include very technical and non-technical people
- This is not an exercise that develops a final solution – more of first iteration
- We would like you to describe how you might design and implement for a long term approach a data infrastructure for processing all additional data (large amount including all available years) given the desired use case above. Please describe what technology and approaches you would use/recommend for managing (Storing, automate, update) the data and make it available to the consumer in a reasonable way.

Submission

Your implementation, code, and results will be reviewed. Please send back all the necessary material (PDF, HTML, Notebook, readme...) to allow the interviewers to understand your approach.

We wish you much success!

Data Lab, Customer Technologies Team