**CaData Science case study**

Wed 13th April 2022

### Background

Vehicle insurance fraud involves conspiring to make false or exaggerated claims involving property damage or personal injuries following an accident. Some common examples include staged accidents where fraudsters deliberately “arrange” for accidents to occur; the use of phantom passengers where people who were not even at the scene of the accident claim to have suffered grievous injury or make false personal injury claims where personal injuries are grossly exaggerated.

### Data

The data set consist of 1000 auto incidents and auto insurance claims from Ohio, Illinois, and Indiana in the United States from 01 January 2015 to 01 March 2015. The data set has a total of 39 variables relating to the insured, their policy, vehicle, and the accident.

The target is to detect if a claim is fraudulent (‘fraud\_reported’).

### Task

As per the background above, conduct your analysis using an analytical tool (one of R or Python is preferred). Prepare a set of presentation slides outlining your findings and recommendations. We don’t expect you to spend more than three hours on the exercise.

Please send through your code and presentation slides to Raj by 5pm on Monday 18th April. Reproducible notebooks are ideal for the code. Your code should cover the end-to-end process you followed, should be concise, and should execute without errors.

Please prepare for a twenty-minute presentation with ten minutes of questions about the case study and your presentation. Your presentation should be focused at non-technical executives in the Fraud and Intelligence team.

This will be followed by thirty minutes for us to chat further about the role and your experience and answer any questions you may have.