**Exercise for Data Scientist candidates**

**Background:**

Part of ACC’s mandate is to reduce the severity and incidence of injuries through injury prevention initiatives. Initiatives are rolled out in many different ways, but an example is our Ride Forever programme which helps prevent motorcycle accidents by offering subsidised training. While these initiatives are ultimately managed by non-analytical injury prevention experts, ACC’s comprehensive datasets are invaluable for informing their design.

The objective of this exercise is to use ACC’s data to help inform an initiative designed to prevent injuries occurring in the gym. Specifically, the initiative will comprise an app which provides information on safe gym practices, and we have budget to deploy this app to 1,000 individuals. For the purposes of the exercise, assume that the dataset comprises everyone who might be eligible to receive the app, and we’re capable of providing the app to any one of them. You can also assume that the app is yet to be designed; and we have no prior knowledge which might be useful for this purpose (e.g., what kinds of people experience these injuries, how should we prevent them etc).

The specific brief from the programme manager (i.e., the person ultimately responsible for designing and deploying the app) is to provide insights on who to target. These insights could be specific (i.e., which individuals to deploy the app to) and/or general (e.g., what insights can you provide which might be useful for developing the app and tailoring it to the “right” audience). You may not have time to do both.

**Specific tasks:**

1. Write up your key findings, insights and/or recommendations for the programme manager. Assume that they are not analytically-inclined. This section does not need to be long.
2. Describe your methodology step-by-step, and if relevant, include some appropriate model diagnostics/outputs. There are no right answers, so please justify why you did what you did. Assume your audience is another data scientist.
3. Please provide your code.

**Some notes:**

* The exercise could be approached in a number of ways. There are no right answers, and there are probably more analyses that could be done than you’ll have time to do. The important thing is to justify why you did what you did.
* You can assume that the dataset is “correct” (i.e., you don’t need to check for coding errors), but outliers and missing values may or may not be present.
* Please be prepared to discuss your exercise during the interview.
* A dataset and data dictionary have been provided.