Fitbod Explaratory Data Analysis

Exploratory Data Analysis (EDA) is an integral step in building robust ML models at Fitbod. Understanding the summary statistics of datasets can give us insights into model selection, feature importance and so forth. For more information see: https://en.wikipedia.org/wiki/Exploratory*data*analysis

Your task is to prepare an EDA using jupyter notebook and python 3. This is deliberately left open ended, so as to best mirror a situation you may encounter at Fitbod. If you want to use a different language please let us know. The dataset is randomly sampled for each candidate, and anonymized to protect sensitive information.

Bonus:

Build a ML model of your choosing to predict something about the data. Given the size of the data set, we don't expect it to be a very good model, just functional. Show how it preformed, and how you would go about building this model if you received a larger dataset.

Additional Thoughts:

Since the data is randomly sampled, and relatively small, the outcomes of your analysis may vary, or give uninteresting outcomes, which is fine. There may also be null values, you can choose to exclude them, or fill them, which ever method fits in with your overall strategy. This exercise is not intended to derive any meaningful outcomes from the data, but just to demonstrate your methods and strategies for approaching new datasets. Your analysis should be enough to occupy a 10-15min presentation, so no need to go overboard. Please feel free to ask us any questions, or bring up any issues you may encounter.

Deliverable:

A jupyter notebook with your results (please include plots inline, short written summaries using Mark Down cells, and code in the notebook). Please commit this to your personal git repo, instead of emailing back, if at all possible.

Presentation:

Be prepared to present this to the data team in a brief 10 - 15 min presentation after submission.

FAQ

Q: What is the unit for the weights?

A: ordinarily in our data we have a value that specifies the unit (such as lb or kg), but I didn't include it, so for your purposes you can assume one – maybe just state that in that some where in your analysis.

Q: What if weight is included in a body weight or cardio exercise?

A: Another good question, weight is generically named in this case, assume that is distance for the scope of this analysis.

Q: a user has a workoutID, and a workoutID has singlesetID(s)