Textual description: Did the team include a textual description of their solution? Is it understandable? Is the described process correct?

Code quality: Is the code readable and concise? Does it use the powerful features of the libraries, or does it re-implement everything from scratch?

Results: Is the final result correct? Are all the assumptions well justified? Are there textual comments/visualizations to convince you of the final result?

Textual description: - What is z-score? What do you do with your data before task 1? You write that you chose the negative class precision but use accuracy….

+ Could have been more in task 2 to understand you thought process.

Code quality: - Repetitive code in task 1, put the spearman and Pearson correlation in a method?

+Uses libraries

Results: - Task 1, visualizations? Could also have calculated some p value and based on an alpha, chance that the strength of the relationship you found happened by chance if the null hypothesis were true. Task 2. Standardize the test set with data from the training set. Dummy encode before splitting, or else you might not have the same values in the test set as in the training set! Shuffle the data before splitting maybe? Somethings looks very wrong with the plots in part B (e.g. recall at 0.9 should be very low)! This makes task C wrong also… and should be tested from 0 to 1, inclusive.

+

Review:

There could have been more textual descriptions describing the way you wanted to solve the problem, more for task 2, this would have given the reader a better insight in you thought process. When you are doing you initial data exploration and cleaning before task 1 it would be good to have a textual description here that describes what is done and why, e.g. what is the z-value that you are using to filter your data and why is this value a good value to filter on? There is also a lot of repetitive code in task 1, maybe put the spearman and pearson correlation calculation in a method would be better. Visualizing the results for task 1 could take the task to a higher-level regarding understanding as the values of pearson and spearman is not that easy to understand by them self. Some sort of null hypothesis testing could also have been done for task 1 to see if the strength of the relationship you found happened by chance if the null hypothesis were true.

For task 2 seeing that the age\_upon\_intake feature contains days, months and years is a nice find. For standardizing you test data it could be better to use the mean and standard deviation used to standardize the training set. This is because of it is hard to know and prove that the data in the test set is fully representative of the data in the training set, even though the data is shuffled before it is split (which you could have done). When dummy variable encoding the data, it would be better to do before splitting into training and testing set as you can’t be sure that all the unique variables are present in both sets. If they weren’t you couldn’t test you model with the test set as it would have had a different dimension than the training set, lucky that didn’t happen.

For the plots of the different metrics in task 2.C something is very wrong, all metric shouldn’t be like they are after threshold 0.4, e.g. recall should be very low when the threshold 0.9 as very few predictions are that certain. Sadly, these errors will affect the results for the rest of the tasks too. I think that threshold 0.0 and 1.0 should be included too so you could see what happened to precision and F1 score. Even though the results after task 2.B the code for your different methods looks to be correct.

Grades:

Textual description: 4.5

Code quality: 5

Results: 4.5