Data Madness script:

(Blazje): John is a data science consultant. One day he got a call. It was Maastricht University. Thinking they want him to help with the cyberattack again, he shouted to the phone for the 100th time that the fact that he can programme, doesn’t mean make him a computer expert. Surprisingly enough, they called for a different reason. It seems that they finally understood what his job is about. Their problem was the following: they noticed recently that their students seems unhappy. They wanted him to look at the data and find the happiness recipe. Of course, they didn't want to (or "could" - remember how much was the cyberattack ransom) spend money on collecting a dataset based on their own students, so he had to look for a similar public dataset. After all, millennials are all the same, aren't they? He found a dataset from University of Technology in Bratislava. A questionnaire with 150 questions filled in by about 1000 students. Not great, not terrible (insert Chernobyl meme).

(Tobias): After spending some time looking at the data, John thought about whether there might exist a group that is especially happy or unhappy. He remembered his uncle Donald saying all his life: “Woman never are happy with anything.” However looking at the data it looked like men and woman were equally unhappy. No significant difference. Also when he looked at people from urban and rural areas, nothing. Number of siblings: nothing. Financial background: no difference. So John decided to use all his data science knowledge to build a model to predict happiness. Because John is a smart data scientist he used the gradient boosting algorithm xgboost. After verifying the plausibility of the model by looking at some of the produced decision trees, John put his model to the test. He could predict with up to 73% if a person was happy or not, with good precision and recall. But he knew Maastricht university won’t be happy with fancy numbers. They would want to know: why? So he analysed his model and looked at the features which were especially important to its performance and therefore to peoples happiness. Viewing the problem as continuous he could get helpful approximation of each of the features exact contribution to the model performance and looking at the data categorically he could see the importance of each feature for each happiness score. The results let him ponder for some time. He found that the most important features for predicting happiness were social aspects like not having to be lonely and having more friends. Of course John knows that the recipe to happiness is never that easy, so he also looked at the interaction between the variables. He found that for people with low energy levels not being lonely actually didn’t help too much, so that people would need extra help. But for people with a normal or high amount of energy being around friends was crucial. And suddenly he knew which message he really wanted Maastricht university to understand:

To be happy we need to stick together!