Please note that instructions are in grey below and there is also an attachment which is required to complete the task.

Instructions

Attached is a fake data set of 100,000 customer records with some demographic data. In the last column is information about their favourite ice cream. However, only about 10% of customers have told us their favourite ice cream, the rest have “No Answer”. For the customers that have “No Answer” can you please create a quick predictive model which tells us what type of ice cream is likely their favourite based on the information we have about customers where we know the answer.

Please present your results in a digestible format and also include a short written summary with:

•             how you went about building this model

•             why you used this type of model

•             summarise the interesting findings of your results

•             what else would you do if you had more time

Any code or other artefacts that could help explain how you went about your approach would also be helpful.

Please note that this task is not designed to take a long period of time. We are more interested in the method of building the model, and using it to kickstart a discussion about the practical implications, rather than the actual results of the model. Therefore, it should take no longer than a few hours to complete this task.