1.Briefly describe the conceptual approach you chose! What are the trade-offs?

A. As the target column, which we need to predict is a categorical value i.e. binomial (0 or 1), so we can use any model which is suitable to predict a binomial target.

I have chosen the logistic regression model. when I saw the summary of train and test sets, I have found that the mean, mode, e.t.c values of both the sets were same even though the size of the test set is more than train set. So, if I run the model and do some predictions on train set and I can use those insights to decide the target values for test set.

I have replaced the missing values with mean (for continuous features) and mode (for categorical features) values. Also, I have chosen the probability threshold criteria from train set to decide the target values in test set.

2. What's the model performance? What is the complexity? Where are the bottlenecks?

As AIC value is suggesting us, the model performance is not up to the mark. But the problem here is we don’t know which attributes to select... because those column names are not given. if we have domain knowledge, we can remove unwanted columns. These are the complexity and bottlenecks for this data.

**This is worst case scenario model**. Even we can try for other models if we have something knowledge on the domain of the data.

3. If you had more time, what improvements would you make, and in what order of priority?

Here the problem is not with the time. It is with the data which doesn’t have any explanation of its domain.

If we have information on domain or scope related data, definitely we can apply a good model which will give more accuracy on predictions.