**Structured Plan to Achieve Higher than 75% Accuracy on Model**

(Using external resources such as notebooks from linkedin learning. )

(feel free to cut and paste shamelessly for this part)

**Part 1: Documentation**

Identify all variables, numeric and non-numeric. Since it is a finite and not huge number of variables, document these in the form of a data dictionary, or a codebook.

Do not use the data dictionary to mark it up and inform your analysis.

Instead, keep that separate. And document plans along the way.

**Part 2: EDA**

Mimic the EDA conducted by the linkedin learning training. In fact, create 1 whole notebook dedicated solely to EDA.

This EDA should contain everything ranging from

* Numeric variable summarization
* Categorical variable recoding to numeric
* Imputing missing data
* Creating / extracting new columns from existing ones, etc.

Output a cleaned dataset.

**Part 3: Model Training**

Create at least one notebook for training your model. If necessary, create multiple notebooks, one for each model. But, as is customary, you may train and evaluate multiple models in the same notebook.

**Part 4: Model Evaluation**

Here, you may choose to place model evaluation in a separate notebook. However, you will have to rerun the model training section anyway, in order to conduct model evaluation.

Therefore, you may choose to train and evaluate up to 3 models in the second notebook. And, if the models do not suffice, overwrite them with new models, or iterations of the same notebook.

Therefore, we would expect at least 2 notebooks, that, if they become too unwieldly, can be broken out into separate notebooks, by model.