**Second project Checklist -**

**Kickstarter Project Success (Crowdfunding)**

**Before starting**

1. Create Repo to contain code with branches
   1. Create gitignore and readme
2. Link github NB review tool
3. Set up conda environment
4. Create a new jupyter notebook
5. Set up note reviewNB module and explore branches
6. Shallow EDA

* Research question #1: How can the stakeholders increase her/his chances to get the funding?
* Hypothesis #1: There are some features that increase the chances of success (getting the funding).
* Dependent variable: ‘state’ ( failure:0, success:1)
* Independent variables (features):
* We want White Box models

1. Set up business case

**Starting the project**

1. Pre-processing of data
2. EDA
3. Model selection
4. Presentation
5. Production-ready code in Visual studio code

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | Required Outcome | To Do | WIP | Done |
| Shallow EDA | Document Columns.  Identify columns to be dropped  Identify columns that need transformation  Identify Target Variable |  |  |  |
| Import File Script | Create function that imports all csv files into a dataframe |  |  | AE |
| Create Business case blurb | Background and business situation explained  1st simple Hypothesis is clearly specified |  |  |  |
| Preprocessing of data | Split into train and test data sets |  |  |  |
| MVP EDA | Identify interesting attributes of Target variable via charts  Identify some potentially interesting features via charts  Show some correlations via eg corr matrix |  |  |  |
| MVP Data cleaning | Identify all necessary data cleaning and transformation actions to get dataset ready for modelling |  |  |  |
| Data cleaning function | Function is created to run all cleaning and transforms action on the selected data |  |  |  |
| Baseline model creation | Simple baseline model is created with train data to show initial results |  |  |  |
| Blurb about model selection | Discuss model selection process |  |  |  |
| Explore 3 models | Use simple application of sklearn models to identify candidate for final model |  |  |  |
| Blurb about final model selection | Write up how we chose final model |  |  |  |
| MVP Grid search on final model | Use grid search with a simple parameter space to optimise the final model |  |  |  |
| Functions for modelling | Put modelling code into stand alonefunctions |  |  |  |
| MVP Presentation | Create simple slides to present our results to non technical audience |  |  |  |
| MVP jupyter notebook | Clean up notebooks |  |  |  |
| MVP python script | Put all function together into working script |  |  |  |