**Project lifecycle**

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Section / Tasks | Deadline | Completion mark |
| **1.** | **Week 7** | **Dec 19, 2022** |  |
|  | **Report form: github repo link (1) PDF document** |  |  |
| 1.1. | Problem description |  |  |
| 1.2. | Business understanding |  |  |
| 1.3. | Project lifecycle along with deadline |  |  |
| 1.4. | Data Intake Report |  |  |
|  |  |  |  |
| **2.** | **Week 8** | **Dec 26, 2022** |  |
|  | **Report form: github repo link (1) PDF document** |  |  |
|  | Problem description |  |  |
|  | Data understanding |  |  |
|  | Data types |  |  |
|  | Data problems (missing values, outliers, skeweness, etc.) |  |  |
|  | What approaches you are trying to apply on your dataset to overcome problems and why? |  |  |
|  |  |  |  |
| **3.** | **Week 9** | **Jan 02, 2023** |  |
|  | **Report form: github repo link (1) PDF document, (2) IPYNB notebook, (3) peers review comments (in what form?)** |  |  |
| 3.1. | Problem description (?) |  |  |
| 3.2. | Data cleansing and transformations done on the data | *Dec 28, 2022* |  |
| 3.3. | Each member should code and review peers work. (Review comment should be present in the github repo) |  |  |
|  | ***NOTES:***  *(1) Each team member should work on different data cleansing approach*  *(2) Try at least 2 techniques to clean the data for NA values: (mean/median/mode/Model based approach to handle NA value/WOE)*  *(3) Try different techniques to identify and handle outliers as well*  *(4) You are allowed to merge the code of each individual and work together to get good result*  *(5) Code of each team member should be placed at provided URL (single repository for whole team)* |  |  |
|  |  |  |  |
| **4.** | **Week 10** |  |  |
|  | **Report form: github repo link (1) PDF document, (2) IPYNB notebook with EDA** |  |  |
| 4.1. | Problem description (?) |  |  |
| 4.2. | EDA performed on the data |  |  |
| 4.3. | Final Recommendations |  |  |
|  |  |  |  |
| **5.** | **Week 11** | **Jan 16, 2023** |  |
|  | **Report form: github repo link (1) PDF document** |  |  |
|  | Problem description (?) |  |  |
|  | EDA presentation for business users |  |  |
|  | Last slide of EDA should be dedicated to technical user which should contain recommended models for this dataset |  |  |
|  |  |  |  |
| **6.** | **Week 12** | **Jan 23, 2023** |  |
|  | **Report form: github repo link (1) (?)** |  |  |
| 6.1. | Select your base model and then explore 1 model of each family (Linear models, Ensemble model, Boosting model, other models if you have time (like stacking)) |  |  |
|  | ***NOTES:***  *(1) Selected model should fit in your business requirement. For example: if your business does not want black box model then select only those models which can be used to explain the prediction*  *(2) You are allowed to merge the code of each individual and work together to get good result*  *(3) As this is group assignment hence upload the code of each team member and other deliverables in the single repo and share the URL of that repo* |  |  |
|  |  |  |  |
| **7.** | **Week 13** | **Jan 30, 2023** |  |
|  | **Report form: github repo link (1) Report, (2) Power point presentation** |  |  |
| 7.1. | As it was group assignment hence go far a call with your team and discuss the solution of each member and select that solution which is best and is per the requirement |  |  |
|  | ***NOTE:***  *(1) You are allowed to merge the code of each individual and work together to get good result* |  |  |

Note: All PDF reports should contain:

* Team member's details : Group Name (give a name to your group)
* Name
* Email
* Country
* College/Company
* Specialization: Data Science