IST 5520, Fall 2021, Chen

**Project Evaluation Form – Milestone 1: Proposal**

**Submission Due Sep 24, 11:59 PM**

**Instruction**

1. Discuss with your group members. Find a business problem that can be solved by using business analytics methods. Submit your project proposal.
2. Students should come up with their own datasets. If you have your own datasets, that is great. If you don’t have datasets at hand, you can either collect your own datasets or find some existing datasets for your project from the Internet such as the following websites:
   1. Open Government Data <https://www.data.gov/>
   2. Kaggle Datasets (especially those open competitions) <https://www.kaggle.com/datasets>
   3. Harvard Dataverse <https://dataverse.harvard.edu/dataverse/harvard>
   4. KDD Nugets <https://www.kdnuggets.com/datasets/index.html>
   5. Google Dataset Search <https://toolbox.google.com/datasetsearch>

The dataset that is appropriate for the course project must satisfy the following criteria:

1. The dataset must have potential to tell an interesting story;
2. The dataset must be in raw format. It has not been cleansed by other people;
3. There must be relatively large number of variables (>=20) that could be extracted from the dataset;
4. The dataset should have a large number of observations (>=1000);
5. The dataset and your research questions have not been fully analyzed on the Internet, especially using Python.

Each group must send their dataset to the instructor for approval. It is recommended that each group discuss their dataset with the instructor in the office hours.

1. The project proposal should include two sections:

* Introduction
* Data Source and Collection

Read the evaluation criteria carefully on the next page for the detailed requirement.

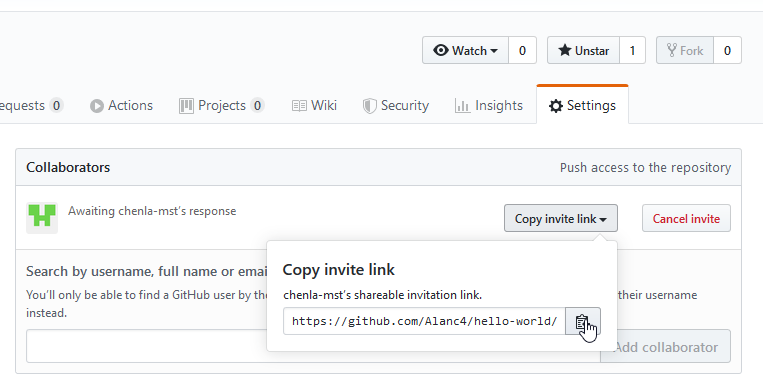
1. Use Markdown in Jupyter notebook to write your project proposal. You need to use proper Markdown syntax to format your proposal. Do not use MS Word or other format.
2. Please create your group project repository on github and use it to manage all your project documents.
3. Please submit the following two documents into Canvas:

* The project proposal written in .ipynb file;
* The Evaluation form with full project team information (see below table).

**Project Repository**

Each group needs to use github to host their project and support collaborative work. To learn how to use github, refer to <https://guides.github.com/>.

* Each group member needs to register a github ID (if you don’t have one) and work on the project repository.
* Please add the instructor (github ID: chenla-mst) as a collaborator for your group project. After you add the instructor into your group project, please copy the invitation link and send it to my email ([chenla@mst.edu](mailto:chenla@mst.edu)). See below for an example.



**Project Team Information (filled in by students)**

|  |  |  |
| --- | --- | --- |
| **Member name** | **Percent contribution** | **Activities completed by the member** |
| Sayantan Majumdar | 20 | Data collection, proposal writeup, Github repository maintenance |
| Dawit Wolday Asfaw | 20 | Proposal writeup, group coordination, proposal proofreading |
| Cong Shen | 20 | Data collection, proposal proofreading, Github repository maintenance |
| Divya Reddy Manku | 20 | Proposal writeup, group coordination, proposal proofreading |
| Akhil Reddy Annreddy | 20 | Proposal writeup, group coordination, proposal proofreading |

**Evaluation Summary – M1 (filled in by instructor)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **Target %** | **Comments** | **Evaluation** |
| * Fully explain the business problem you choose to analyze. * Introduce business scenario or setting for your project. | 30 | You need to set a more specific business scenario for your project. Without a full explanation of the project background, it’s hard to understand why those research questions make sense and are important. Specifically, why predicting EPA emission score is important.  I understand the data is from a government portal which may not contain much background information. Now it’s your project and you need to find the business scenarios/background underlying your data. You need such kind of information to tell an interesting story. | 25 |
| * Clearly specify the high-level requirements and research questions. | 20 |  | 20 |
| * Specify data source(s) [and data collection method(s) if applicable]. The dataset should satisfy all criteria as specified on page 1. * Explain how you get the data and what key information is available in the dataset. | 20 | Please also provide a data dictionary to explain the meaning of the important variables. This allows us to evaluate whether the data is appropriate to answer those research questions proposed. | 20 |
| * Format your project proposal in a professional way. * Write your project proposal by using appropriate Markdown syntax. | 20 |  | 20 |
| * Create a github repository for your project. * The repository has been shared to the instructor. * All group members have joined the repository. * Use the repository to manage all your project documents including meeting schedules, meeting minutes, and the proposal (the instructor should be invited to verify this). | 10 |  | 10 |
| **Total** | **100** | **Good job! I expect to see interesting results from your project.** | **95** |