**Week of February 18th- March 4th**

**Blog 4: March 4th**

In these weeks, we finished our initial preliminary draft including the business problems we would like to work on, the solution approach we would apply and our main objectives. We sent it out to our clients and received their feedback soon. Based on that, our group divided this project into several parts and we tried to work on the problems we want to solve step by step. There are many missing values in some columns and we started to clean the dataset, especially for the columns that we need to use immediately. Also, I merged the current dataset and historical dataset so that I can aligned what I need from both.

After being authorized with the access to the virtual system, we created a Share Drive and made sure that we could work on the project and communicate more effectively. Using Jupyterbook, we performed some basic descriptive analysis and discussed our methods with Mariem weekly. Initially, we had a lot of problems about the concept for different column name and the relations among those datasets. We sent out emails asking for clarify so that we would not work in a wrong way. By discussing with Mariem, we confirmed our targeted variable ‘status.code’ which has six levels with won(funded), won(unfunded), lost and so on. We would focus on the won or lost while considering others as outliers. Also, we asked whether we could be given a clear data dictionary for us to better understand the datasets. Meanwhile, we kept working on the exploratory data analysis.

In addition, we received feedback for the preliminary draft we submitted before and there are many helpful comments on it. Most of them are supportive with few things to eliminate and some helpful suggestions. This feedback allows us to be much clearer about what approaches we should use and how to work on the projects better. We will pay attention to their suggestions and performed some visualizations to present our result clearly.

In these few weeks, we mainly focus on sprint 2 of data explorations, initial processing and analysis. We learned about our data and did some researches on them so that we could picked up the target variable and some features we thought that were important. In the coming weeks, we will try to achieve more detailed results before building the models. Also, we would discuss more and did some researches about how to perform time series on our problem as recommended by our professor.