**Week of Jan 24th – Feb9th**

**Blog 2: February 10th, 2019**

Last week, we successfully set up our accounts for the virtual machine and got the access to the dataset and we tried to understand different features and variables since there are no clear data description for us. We had a phone-call meeting with Mariem Ayadi on January 28th, who helped us walk through the dataset. This one-hour talking gave us a basic overview about CRM and what the dataset is supposed to look like since we were not onboarding that time. We discussed on our timeslots and arranged a weekly phone call schedule based on everyone’s availability, which is 12pm-1pm on Friday.

Before I got access to the dataset, I worked on a similar case “red hat”, a Kaggle competition project. I learned about different approaches some candidates performed and it turned out to be really helpful. Meanwhile, I did more research on pipeline by reading various papers and reports so that I can get used to the structure of the dataset immediately as soon as I got it. We also designed our potential sprints on Trello, dividing the whole tasks into several parts.

I signed the Non-Disclosure-agreement and sent it back to the company so that I would be authorized to log in the system and touch the dataset. When receiving our account information, there were many some little problems making us difficult to log into the virtual machine at first. We reflected these technical problems immediately to our clients and had them solved eventually with their kind support.

On February 8th, we were given access to the data. I started to understand the variables first while I realized there is no column names in one of the CSV file. In addition to that, all of us raised a few questions and we wrote down for next phone-call meeting. I did some basic exploratory data analysis with Python to have an overview for the dataset and researched a few features which I though important. We also submitted a two-page draft paper showing the whole framework of our project and updated the content according to professor’s feedback. In the coming few weeks, we will spend most time cleaning our data and do some pre-processing work.