**Week of February 9th-February 17th**

**Blog 3: February 17th, 2019**

In these weeks, we got access to the data and began the process of initial data understanding and processing. During this, our mentor Chris Kelly got us in touch with one of his contacts to give us another view on the timing of the sales pipeline process. This gave us an opportunity to understand another perspective of how probabilities are assigned to every stage and what is the range of time periods set for each stage, giving us more understanding on the business discretion behind the pipeline process.

In our call after getting all the internal access, we made sure that we can log into the Intranet through the virtual system, get proper Share Drive access and make sure that technically we are all on the same page. We progressed with our descriptive analysis and regularly discussed our approach with Mariem Ayadi (our point of contact for weekly calls) every Friday. During the initial data exploration, we came across a lot of conceptual questions and also questions regarding some specific columns of the dataset. We got familiar with the Power BI tool for visualizations, which could be installed in our virtual machines in place of Tableau.

From the data files, we focused on some variables that would be important in our analysis from one of our csv files called APACAMER.csv. From this file, we wanted to explore the question of ideal timings for closing a deal. In addition to this, we learnt that probabilities would be distinct, and not continuous values, there are several currencies, however, we have to look at the default USD value. We tried to differentiate between the status and stage and understand the business concept behind it. We also framed questions related to coverage and product group level to understand if they are key factors or if we could find an identifiable cluster regarding these.

The other file that we worked and explored was the historical data file (Audit history), which gave us information about opportunities with fields such as opening and closing dates, amounts, modified date. This file will help in the mapping process with opportunity ID serving as the primary key.

We made some initial judgments on the variables we would like to use, what areas we would like to explore: such as coverage, product level or even region and best estimated possibilities for our target variable.