**Sprint 2:**

**Start date: February 2nd, Saturday**

**End date: March 10th, Sunday**

**Time: 5 weeks**

**What items were planned to be accomplished during this sprint**

* Get the initial description of the data:
  + Basic data dictionary
  + How many categorical or how many numerical
  + Match variables in the dataset with the sales pipeline understanding
  + What is the shape of the data
* Exploratory Data Analysis using Python to give an overview of the data
  + Understanding outliers, what transformations are necessary
  + Distribution of the target variable and if any normalization is needed
* Understand the relationship between history and latest data file, and figure out how to merge them or transform them
* Construct models to predict the probability of target variable given on features we selected
* Evaluate the effectiveness of subjective probability

**Which items were accomplished?**

* The difference between history and latest data file and how could we use them
* The target variable and some other important features
* Data visualizations on these variables

**Which items were not accomplished?**

* Modeling
* The evaluation of subjective probability is still in process

**What could be improved for the next sprint?**

* Use Power BI to create visualizations and dashboard
* Get business insights from these visualizations
* Create time-series dataset