**Task1\_EDA\_Data\_Insights**

Meeting Minutes: Date: October 1st Friday 2021,1 PM UTC

* **Translate the Entire Data Set from Spanish to English**
  + Using Google Translate to translate the entire data set from spanish to english
* **Handling of Missing Null Values**
  + Drop Columns if more than 20% NaN data exists in a Column that way we can avoid Overfitting of Model.
* **Detect & Remove Outliers**
  + Discussion on outlier analysis was held. The variance of scores in Saber pro examinations is a concern which needs to be dealt with.
* **Identify and remove Duplicate Columns**
  + Example:  #Birth date, #Fathers education ,  #Genre, #Type of document
* **Feature Selection & Reduction**
* **Data Embedding**

**Next meeting:** It will be held after 3 or 4 days