# Week 7 Deliverables

**Team member details:**

Group Name (give a name to your group), Name, Email, Country, College/Company, Specialization ( Data Science, NLP, Data Analyst)


- Alyssa Hubiak, alyssa.hubiak@gmail.com, United States, currently a student at Maryville University of St. Louis, Data Science
- Nasibul Hossain, Nasibulh@gmail.com, US, currently seeking a role in Data Science


**Problem description**

Develop a model to predict whether a customer is likely to buy the bank's product or not. This will help ABC Bank determine where to focus their marketing for their new term deposit product. This will need to be a binary classification model to predict whether the customer will subscribe (1) or not (0).

**Business understanding**

The end goal is to create more profit using as little resources as possible. The objective behind this project is to generate insights to help the bank achieve that goal. The first step is to predict which customers will convert and which will churn based on the data. After doing so, both kinds of customers can be further segmented. From that, the bank can focus on how to better adjust services in order to attract more customers as well as retain current customers. In terms of metrics, a low amount of false negatives with a higher amount of false positives are better because it is better to get all of the customers who will possibly convert than and waste some resources in contacting them than to not contact customers who were likely to convert and save some resources while sacrificing a lot of possible subscribers.

**Project lifecycle and deadline**

Deadline is November 30th

**Data Intake report**

See below

**Github Repo link**

https://github.com/ahubiak/Data_Glacier_DS_Project

# Data Intake Report

Name: Data Science - ABC Bank Marketing
Report date: 19 October 2022
Internship Batch: LISUM 13:30
Version: 1.0
Data intake by: Alyssa Hubiak
Data intake reviewer: Nasibul Hossain
Data storage location: https://github.com/ahubiak/Data_Glacier_DS_Project

Tabular data details:
bank-additional-full.csv

| | |
|---|---|
| Total number of observations: | 45211 |
| Total number of files: | 1 |
| Total number of features: | 20, 1 target |
| Base format of the file: | .csv |
| Size of the data: | 5699KB |

Proposed Approach:

- Mention approach of dedup validation (identification)
  - Full dataset is in csv format and can be read into python
  - Several categorical features which will need to be encoded prior to model creation
  - Model will be binary classification
  - Several boolean features which will need to be typed
- Mention your assumptions (if you assume any other thing for data quality analysis)
  - Assuming all data is correct, was gathered on phone calls so may have bias
  - There may be class imbalance which we will need to adjust for