Project Title: Data Science:: Bank Marketing (Campaign) -- Group Project
Date: February 14, 2024

Group Name: Data Department_1
- **Name**: Minseok Kim
- **Email**: mxk230041@utdallas.edu
- **Country**: United States
- **College/Company**: The University of Texas at Dallas
- **Specialization**: Data Science

Problem Statement:
A Portuguese banking institution seeks to improve its marketing strategy for selling term deposit products. The goal is to develop a predictive model that can anticipate whether a customer is likely to purchase the term deposit based on their past interactions with the bank or other financial institutions.

Business Context:
The project aims to align with the bank's objective to optimize resource allocation by targeting customers who are more likely to subscribe to the term deposit. This focus increases the efficiency of marketing channels such as telemarketing and digital marketing, thereby saving time and reducing costs.

Project Phases:
- **Phase 1**: Business Understanding - Deadline: January 24, 2024
- **Phase 2**: Data Understanding - Deadline: January 31, 2024
- **Phase 3**: Data Preparation - Deadline: February 7, 2024
- **Phase 4**: Model Building - Deadline: February 7, 2024
- **Phase 5**: Model Evaluation - Deadline: February 7, 2024
- **Phase 6**: Deployment - Deadline: February 14, 2024
- **Phase 7**: Presentation and Reporting - Deadline: February 14, 2024

Data Intake Report

Name **Data Science:: Bank Marketing (Campaign) -- Group Project** Report date: Feb 14, 2024
Internship Batch: LISUM 28
V ersion:1.0

Data intake by: Minseok Kim
Data intake reviewer: Minseok Kim
Data storage location: https://github.com/N0VA-code/Data-Glacier-Data-Based-Consulting-Project.git

Tabular data details:

Total number of observations Total number of files
Total number of features Base format of the file

Size of the data
Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

1.Deduplication validation will be performed using a combination of key feature comparisons and hashing techniques to identify and remove duplicate entries.

2.Assumptions for data quality analysis include the completeness and accuracy of the records, the consistent format of entries across all data points, and the correct encoding of categorical variables.

Data Summary:
The dataset comprises direct marketing campaign data from a Portuguese bank, with attributes ranging from client demographics to economic context indicators. Preliminary exploration suggests trends and patterns that could be predictive of term deposit subscriptions.

GitHub Repository: https://github.com/N0VA-code/Data-Glacier-Data-Based-Consulting-Project.git