Project being undertaken:

Bank Marketing (Campaign) --- Data Science

Group Name

The Data Team.

Members

Fawzi Khatib

Aurthur

Maliplema

Nigel

Bank-Additional Data Intake Report.

Report date: 03/05/2021

Internship Batch: LISP01

Version: 1.0

Data intake by: Data Team

Data intake reviewer: <intern who reviewed the report>

Data storage location: https://archive.ics.uci.edu/ml/machine-learning-databases/00222/

Tabular data details:

Total number of observations	`41188`
Total number of files	**5**
Total number of features	**21**
Base format of the file	**.csv**
Size of the data	**6.6 MB**

The 3 files consist of the **"bank-additional-full"** .csv file which contains the full data of the 21 features, the **"bank-additional"**.csv file which is a sample of 4652 observations of the bank-full file, and **"bank-additional-names"**.txt which is a text document file contains the features names and explanations.

Proposed Approach:
- "poutcome" feature: is the "outcome of the previous marketing campaign" this feature has about 86% of *Unknown* values that's why we may drop it for further working.
- We have 12718 missing-values without counting the Na values of the "poutcome" feature.

Bank Data Intake Report

Name: Bank Marketing (Campaign)
Report date: 03/05/2021
Internship Batch: LISP01
Version: 1.0
Data intake by: Data Team
Data intake reviewer: <intern who reviewed the report>
Data storage location: https://archive.ics.uci.edu/ml/machine-learning-databases/00222/

Tabular data details:

Total number of observations	45211
Total number of files	3
Total number of features	17
Base format of the file	.csv
Size of the data	5.9 MB

The 3 files consist of the **"bank-full"** .csv file which contains the full data of the 17 features, the **"bank"**.csv file which is a sample of 4120 observations of the bank-full file, and **"bank-names"**.txt which is a text document file contains the features names and explanations.

Proposed Approach:

- Bank full data contains the "balance" feature which is not in exist in the data additional dataset. Also, we can notice that there is difference in the number of observations between the two datasets, moreover the unavailability of a primary key between them makes it difficult to join them.
- "poutcome" feature: is the "outcome of the previous marketing campaign" this feature has about 81% of *Unknown* values that's why we may drop it for further working.
- We have 15165 missing-values without counting the Na values of the "poutcome" feature.