# **Project Meeting Minutes - II**

Date of meeting(s): **24-02-2023**

Group Number: **11**

Group members present (Name, ID):

* **Vaibhav Patel (0772934)**
* **Avikumar Patel (0790966)**
* **Yashkumar Patel (0797825)**
* **Smit Rana (0792056)**

**Specific Activities that were completed/worked on:**

* We scheduled data cleansing for this week. Cleaning our data is a vital step in preparing it for analysis. It ensures that the dataset is precise, impartial, and acceptable for modelling.

Table

Description automatically generated

* **Num\_Bank\_Accounts, Num\_Credit\_Card, Interest\_Rate, Num\_Credit\_Inquiries,and Total\_EMI\_per\_month** columns contain meaningless values. (As we can see from the maximum values of these columns in the describe method). We shall establish a general threshold for these numbers while preparing data. Then we'll look for outliers and deal with them if necessary.
* Also, columns like **Num\_of\_Delayed\_Payment, Age, Annual\_Income, Num\_of\_Loan, Changed\_Credit\_Limit, Occupation** and many other columns contain string and superfluous symbols as prefixes and suffixes. Therefore, we must fix them during data cleansing. **Credit\_History\_Age** contains values in as *xx years and xx months*. Therefore, we shall convert these to strictly quantitative values that indicate the number of months by standardizing them. We convert Month values to numeric ones.

Graphical user interface, table

Description automatically generated with medium confidence

* We had number columns, but their present values contain unwelcome symbols and strings, turning the column into an object.
* **Type\_of\_loan** and **Credit\_Score** requires one hot encoding because it carries categorical data.

**Specific Output from work:**

* In order to clean the data, we did various tasks such as eliminating duplicate records, dealing with missing data, finding and removing outliers, correcting data skewness, and removing unnecessary variables after evaluating dirty and inappropriate data.
* In addition, we developed Assessment 2 with a Gantt Chart, Addressing Research Questions, an Ethical Assessment using the Data Science Ethics Checklist, and EDAs.

**On Target:**

* Indicate the current status of your project:

**✅ green:** everything on track for completion by due date

yellow: a small number of tasks are off track and completion by due date is at risk.

red: many tasks are off track and the project will not be completed by due date.

**Challenges/Disagreements:**

* The most difficult aspect of this week was evaluating and cleaning each data column. Another was responsible for completing and setting appropriate thresholds for columns with overlimit values.

**Planned Activities for coming week:**

* As soon as we have clean data and have completed the requisite transformation and selection, we will perform feature engineering. Then, we begin data modelling and produce visualizations and dashboard concurrently.