**Data Preprocessing**

After collecting responses via survey all the data store in excel online (one drive) through the process of Microsoft power automation we are connecting it to google drive from there we are fetching that raw data on google cloud without filtering dataset and store it as one Main table after that with the use of Big Query we are breaking the main table into 7 different table based on similar kind information. From google cloud All the 7 table then saved as csv file to google drive for data preprocessing.

For Data preprocessing we are fetching that raw data table from Google Drive to Google Colab.

Data preprocessing is an important step in data analysis that involves transforming raw data into a more usable format. The following are some common steps involved in data preprocessing:

**Data Cleaning:** The following steps can be taken for data cleaning:

* Handling missing values
* Outlier detection and treatment
* Shortening long values
* Separating semi-colon separated values
* Cleaning text field data
* Encoding categorical variables

**Data Integration:** Our data is partitioned into 7 tables, and we must need to integrate relevant data from multiple tables to conduct bivariate analysis for observing variable correlations and machine learning for satisfaction score prediction.

**Data Transformation:** In order to analyze data accurately, it must be transformed into a format that is suitable for analysis. This involves techniques such as normalization, standardization, and scaling. Normalization ensures that all data values fall within a specific range, while standardization scales data to have a mean of zero and a standard deviation of one.

**Data Reduction:** As our dataset is not large enough then data reduction may not be necessary or may be less important. However, if the data contains many features or variables, then data reduction techniques can still be useful for simplifying analysis and improving efficiency. Some data reduction techniques that can be used in this case include:

* Feature selection
* Feature engineering
* Clustering

Overall, data reduction techniques can be useful even for smaller datasets, particularly if the data contains many features or variables. However, the specific technique used will depend on the nature of the data and the goals of the analysis.

**Data Encoding:** In order to analyze our data, we need to transform categorical variables into numerical format using encoding techniques like one-hot encoding and label encoding.

Overall, data preprocessing is a critical step in data analysis that can help improve the accuracy and reliability of the results obtained.