PartA

Extra data summarization on PowerBI (you can find others in the code)

The sum of loan amount for all applicant with management or self-employed job type (attribute)

A screenshot of a graph

Description automatically generated

The sum of loan amount for all applicant with single marriage status (attribute)

A screenshot of a graph

Description automatically generated

The sum of loan amount with good risk (attribute)

A screenshot of a graph

Description automatically generated

Part A report:

Process data in fact table and joined table. The data processing code performs several key preprocessing steps on a dataset containing numerical features. Below is a summary of the preprocessing steps along with any data transformation and quality issues encountered:

Handling missing values:

No missing values found.

Handling categorical attributes through e.g., one-hot encoding:

One hot encoding for categorial attributes like: saving account, checking account, purpose

Normalization using Min-Max Scaling:

The initial dataset contains 7 numerical columns that are unsuitable for direct use in machine learning models due to lack of normalization.

Min-Max Scaling is applied using MinMaxScaler() to scale the numerical features to a range between 0 and 1.

This scaling ensures that features with larger magnitudes (e.g., salary) do not dominate the learning process compared to features with smaller magnitudes (e.g., age or loan duration).

Data Transformation:

Before normalization, descriptive statistics are provided for the original dataset, including minimum, maximum, mean, and standard deviation values for each numerical feature.

After normalization, a similar table is generated, displaying the minimum, maximum, mean, and standard deviation values for each feature, now scaled between 0 and 1.

Final Dataset:

After preprocessing, the dataset is updated with the normalized numerical columns, resulting in a cleaner dataset ready for further analysis or modeling.

PartB

Skipped (Verified with TA)

PartC