ETL Project

**Introduction:**

Credit Score is a main method to predict a applicants probability of future defaults and credit borrowings. The banks will decide whether to issue a credit card to applicant by verifying the historical data.

**Extraction:**

Used 2 different datasets from the public platform Kaggle which lead to the Credit card Approval Prediction data set.

The data includes: Customer information and customer historical record.

The field of interests include:

* Client number
* Gender of customer
* Annual Income
* Marital Status
* Birthday
* Employment Period
* Status of Credit history

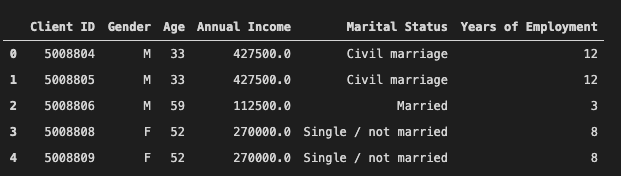
Sources for datasets used:

<https://www.kaggle.com/rikdifos/credit-card-approval-prediction?select=application_record.csv>

<https://www.kaggle.com/rikdifos/credit-card-approval-prediction?select=credit_record.csv>

**Transformation:**

* Pandas functions in Jupyter Notebook to read the CSV files.
* Reviewed the files and transformed into data frames.
* Removed operator’s columns and only record the needed columns for the focus of study.
* Used inner merge on the client id column across the data sets.
* Created queries to address the hypothesis.



**Load:**

After pulled the CSV files and loaded them to the data frame, an initial connection has also been made to the Postgres Database using PG Admin. By storing the original clean data sets into the database, an initial table schema that loaded into the postgres database has generated the first set of tables. With the queries and new tables, the relevant information will reconnect to the database.

Postgres Database:

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated