

Dimension Tables:

1. Customer Table:

Table Name: customer

Description: Stores information about customers who have taken loans.

Column Names and Data Types:

hashed_social: object (Primary Key)

pseudo_phone: object

title: object

employment_length: int64b

Primary Key: hashed_social

2. Loan Table:

Table Name: loan

Description: Contains details of the loans granted.

Column Names and Data Types:

loan_id: object (Primary Key)

funded_amount: int64

funded_date: object

duration_years: int64

duration_months: int64

interest_rate_percent: float64

purpose: object

Primary Key: loan_id

3. Asset Table:

Table Name: Asset

Description: Stores information about assets associated with loans.

Column Names and Data Types:

unique_id: object (Primary Key)

property_value: int64

BUILDING_CLASS_CATEGORY: object

TAX_CLASS_AT_PRESENT: object

BUILDING_CLASS_AT_PRESENT: object

ZIP_CODE: int64

CITY: object

STATE: object

TOTAL_UNITS: int64

LAND_SQUARE_FEET: float64

GROSS_SQUARE_FEET: float64

TAX_CLASS_AT_TIME_OF_SALE: int64

Primary Key: unique_id

Fact Table:

Table Name: fact

Description: Contains payment and loan balance information.

Column Names and Data Types:

loan_id: object (Primary Key)

hashed_social: object (Foreign Key)

unique_id: object (Foreign Key)

interest_rate: float64

payments: float64

total_past_payments: int64

loan_balance: float64

snapshot_date: datetime64

Primary Key Constraint: loan_id

Foreign Key Constraint: hashed_social, unique_id (References hashed_social in customer table, loan_id in loan table, and unique_id in asset table)

Rationale:

The schema design follows the principles of data normalisation to reduce redundancy and improve data integrity.

Each table represents a distinct entity (customer, loan, asset) to organise data efficiently and facilitate data retrieval.

Foreign key constraints ensure referential integrity between the fact table and the dimension tables (customer, loan, asset).

Appropriate data types are chosen for each column to ensure accurate storage and retrieval of data.

The schema design supports the project requirements by providing a structured framework for storing and analysing loan portfolio data in GCP while adhering to data governance standards.

