Schema Design Summary

Data Source

Original dataset: bank.csv loaded into a DuckDB table named bank_raw.

Schema Structure

Fact Table: bank_customer_interactions

- Primary Key: customer_id (generated using ROW_NUMBER()).

- Contains: Customer interaction details (e.g. contact, age, balance, etc.).

- Foreign Keys:

  - job_id -> dim_job

  - marital_id -> dim_marital

  - education_id -> dim_education

  - finance_id -> dim_finance

Dimension Tables

dim_job: job_id (PK), job

dim_marital: marital_id (PK), marital

dim_education: education_id (PK), education

dim_finance: finance_id (PK), default, housing, loan

Key Design Decisions

- Used ROW_NUMBER() to generate surrogate primary keys.

- Fact and dimension tables are linked using these surrogate keys.

- Finance-related fields grouped into one dimension for simplicity.

Assumptions

1. Row Order Alignment:

   All tables rely on ROW_NUMBER() matching the order of bank_raw, assuming no rows are dropped or

reordered.

2. 1-to-1 Mapping:

   Each row in bank_raw maps directly to one row in each dimension, which may not be valid if the source

data has duplicates or inconsistencies.

3. Simplification for Analysis:

   Dimension values are stored separately to enable easier slicing and aggregation during analysis.