Informatica ETL Project

Overview

This Informatica project involves extracting, transforming, and loading (ETL) data from two **CSV files**: **Churn_Modelling2.csv** and **Churn_Modelling3.csv**. The objective is to perform various analyses of the data and generate specific outputs as outlined below. The project also includes extracting data from the **Oracle HR schema**, performing transformations, and loading the transformed data back into the database.

Tasks

1. Sum of Balances for Male and Female Customers:
   - Calculate the sum of balances for male and female customers.

2. Number of Active Male and Female Customers:
   - Determine the number of active male and female customers.

3. Customer Age Distribution:
   - Group customers based on age distribution into the following categories:
     - Age between 18 and 30
     - Age between 31 and 45
     - Age above 45

4. Users with Available Balance Data:
   - Identify users with available balance data (balance > 0) and rank them in ascending order.

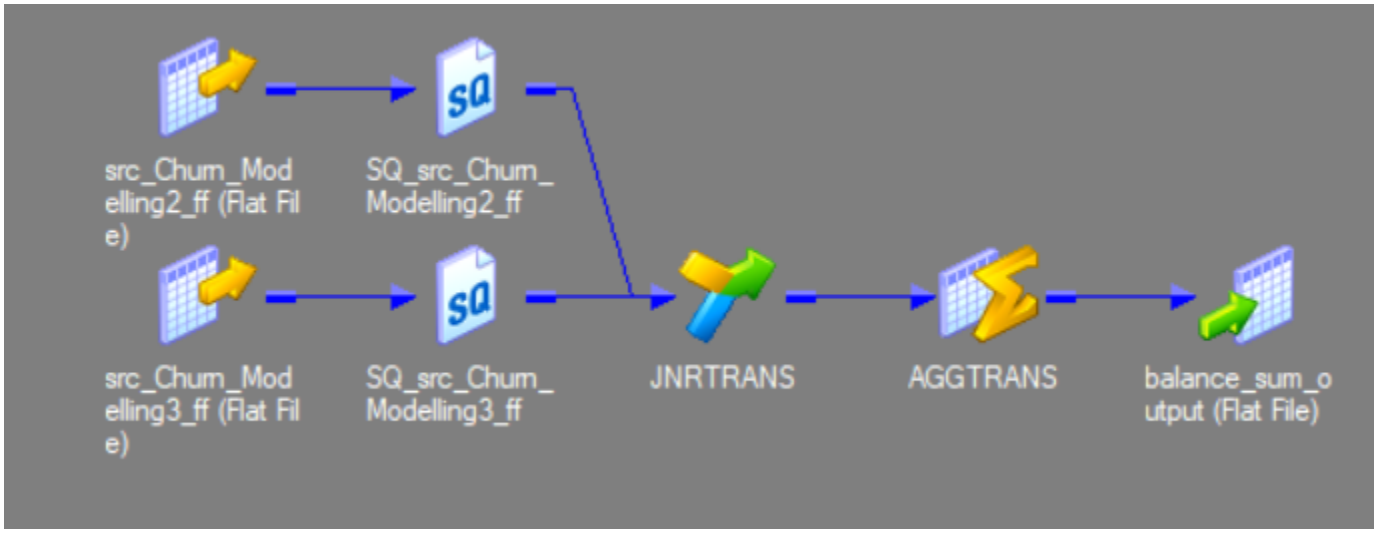
5. Top 5 Balances:
   - Retrieve all information for customers with the top 5 balances.

6. Highest and Lowest Credit Score Values:
   - Find the highest and lowest credit score values.

7. Perform the following expressions:
   - Concatenate the first name and last name.
   - Check if the commission is null, set it to zero.
   - Calculate the commission.
   - Calculate the total salary plus commission.

Screenshots

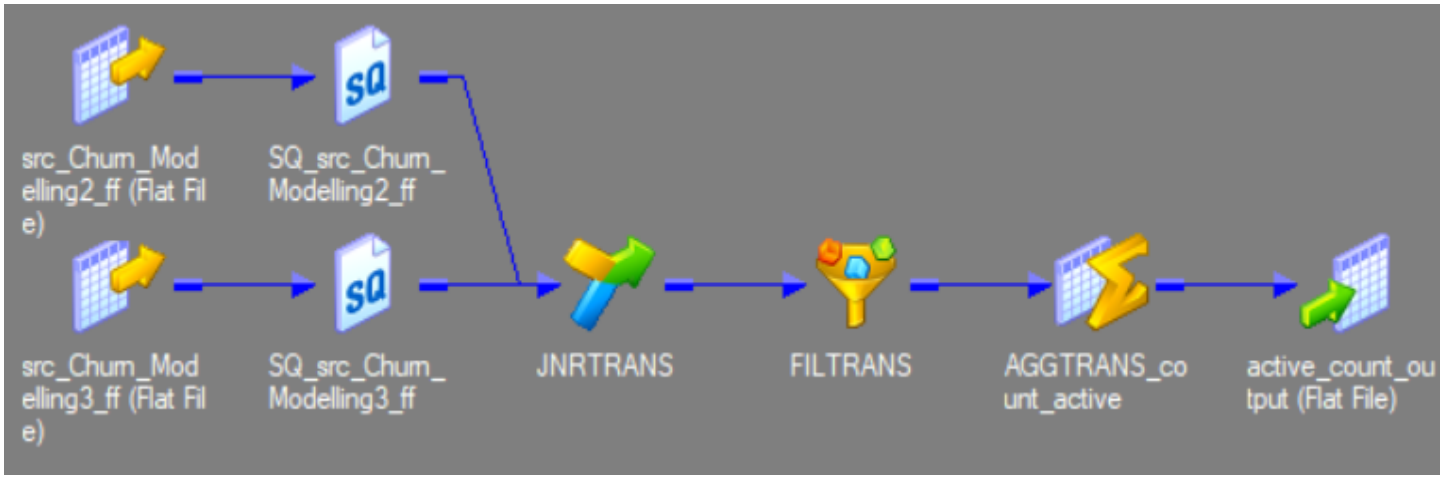
1. Sum of Balances for Male and Female Customers:



Output

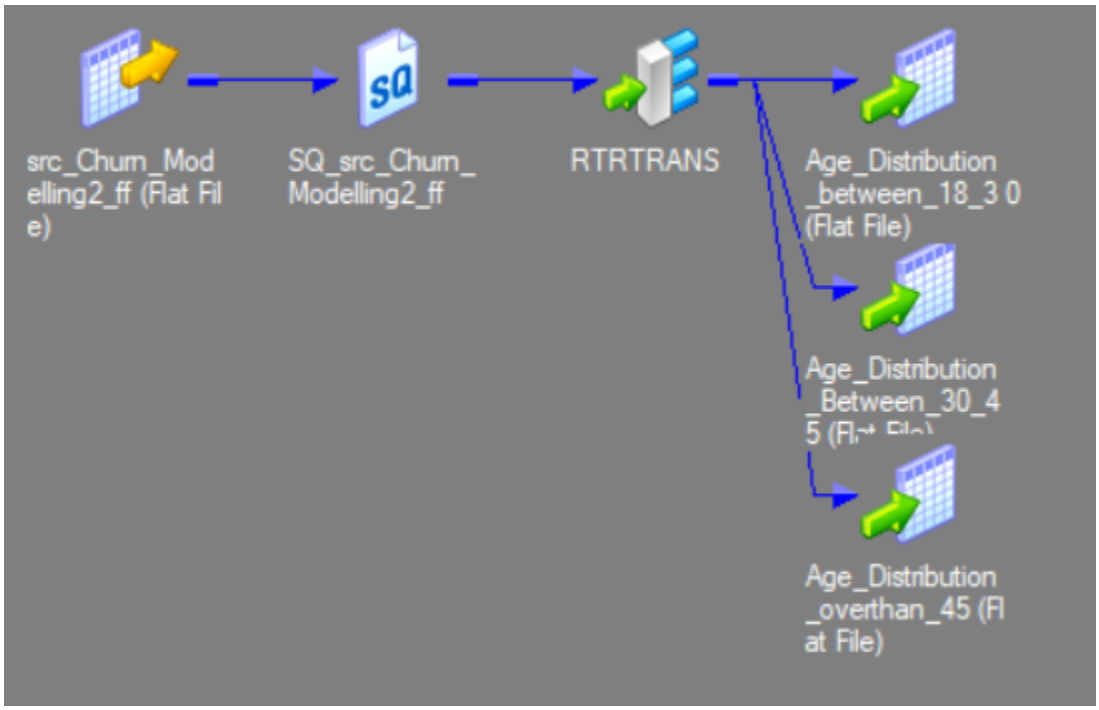
• balance_sum.csv: Sum of balances for male and female customers.

2. Number of Active Male and Female Customers:



Output • active_count.csv: Number of active male and female

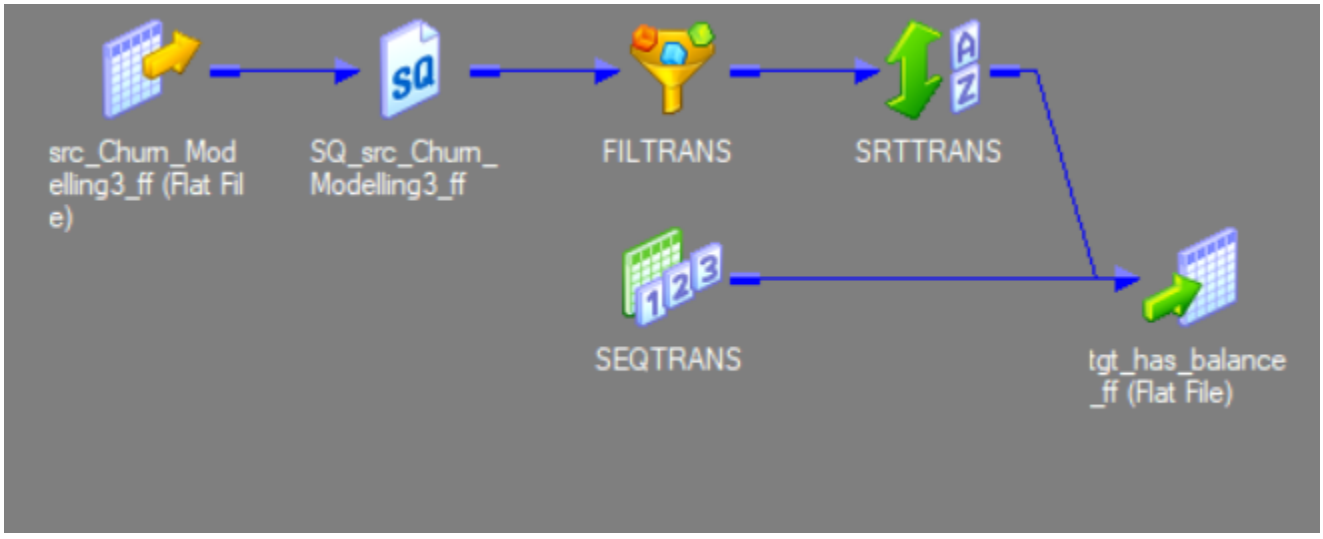
3. Customer Age Distribution:



Output

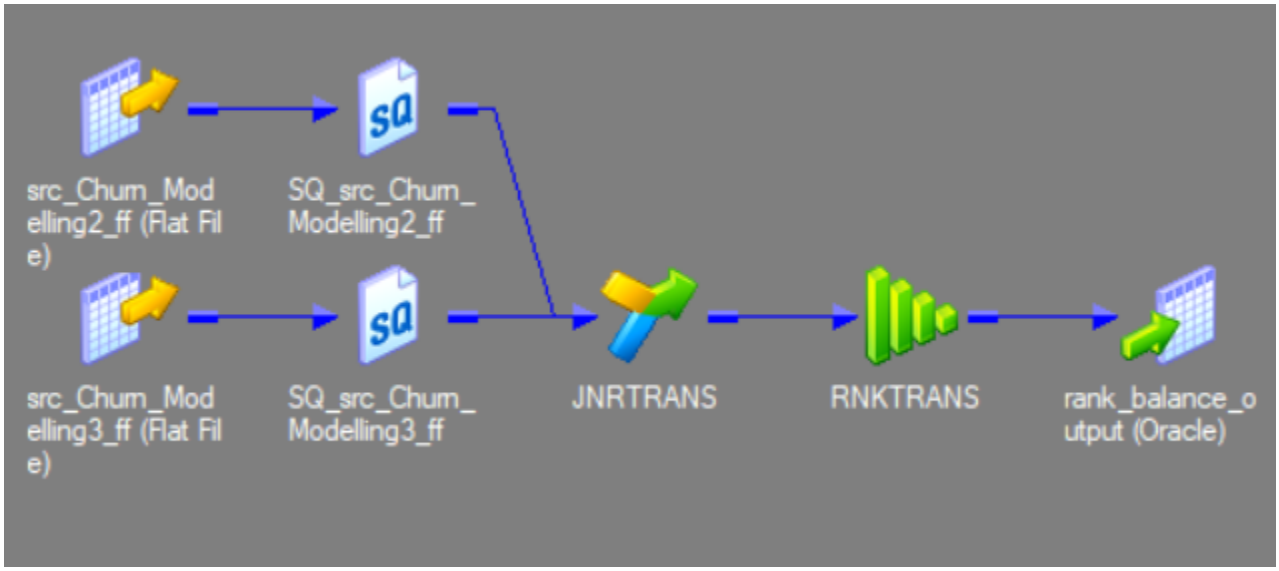
- Age Distribution Tables.

4. Users with Available Balance Data:



Output • tgt_has_balance_ff.csv: Users with available balance data.

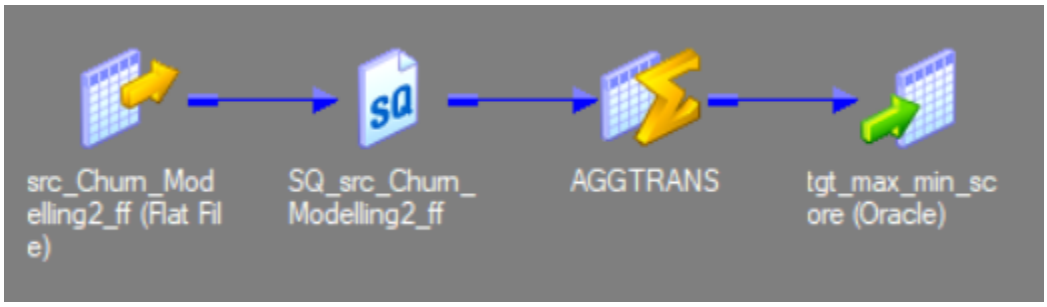
5. Top 10 Balances:



Output

• rank_balance.csv: Information for customers with the top 10 balances.

6. Highest and Lowest Credit Score Values:



Output

• tgt_max_min_score.csv: Highest and lowest credit score values.

7. Perform the following expressions:

Output

• tgt_employees_db.csv: Total SAL + Cal Commission.

Usage:

1. Ensure that Informatica PowerCenter is installed and configured.
2. Import the CSV files Churn_Modelling2.csv and Churn_Modelling3.csv, and import data from the Oracle HR schema into Informatica PowerCenter.
3. Create mappings, transformations, and workflows to perform the specified tasks.
4. Execute the workflows to generate the specified output files.

Note:

- Ensure the input CSV files are correctly mapped within Informatica PowerCenter.
- Validate the transformations and mappings to ensure accurate data processing.
- Monitor the workflow execution for any errors or issues.
- Adjust any parameters or configurations as needed to optimize performance and accuracy.