INFORMATICA POWER CENTER – ETL PROJECT

1. Project Overview

This project demonstrates an end-to-end ETL implementation using **Informatica PowerCenter**. The goal is to process customer churn data from CSV files, apply business rules, and generate analytical outputs that can be used for reporting and decision-making.

2. Source Data

The ETL process uses the following CSV files as source systems:

File Name	Description
Churn_Modelling2.csv	Customer churn data – part 1
Churn_Modelling3.csv	Customer churn data – part 2

3. ETL Architecture

Source → **Transformation** → **Target**
- Source Definition: Flat Files (CSV)
- ETL Tool: Informatica PowerCenter
- Target: Output files and tables for analytics

4. Business Requirements and ETL Logic

Requirement 1: Calculate the total balance for male and female customers.

ETL Approach: Use Expression and Aggregator transformations. Group by Gender and calculate SUM(Balance). Output to a flat file.

Requirement 2: Count the number of active male and female customers.

ETL Approach: Filter Active members, then aggregate by Gender to count customers. Output to a flat file.

Requirement 3: Create age group distributions.

ETL Approach: Use Expression transformation to create age buckets: • Age between 18–30 • Age between 30–45 • Age greater than 45 Then aggregate counts per group and load into output tables.

Requirement 4: Rank customers with available balance (Balance > 0).

ETL Approach: Filter Balance > 0, then use Rank transformation to rank customers in ascending order of balance.

Requirement 5: Retrieve customer details for top 5 balances.

ETL Approach: Use Rank transformation (Top = 5, descending by Balance) and load full customer records to output.

5. Key Informatica Transformations Used

• Source Qualifier – Reads data from CSV files

• Expression – Data cleansing and derived columns

• Filter – Apply business conditions

• Aggregator – Summarization and grouping

• Rank – Ranking customers by balance

• Target Definition – Store results in output files/tables

6. Project Outputs

The ETL process generates the following outputs: - Total balance by gender (file) - Active customer count by gender (file) - Age group distribution (tables) - Ranked customers with positive balance (file) - Top 5 customers by balance (file)

7. Interview & Portfolio Value

This project demonstrates strong ETL fundamentals, real-world business logic, and Informatica PowerCenter best practices, making it an excellent portfolio and interview project.