Informatica Challenge

NOTE: You are given tasks for both Informatica and Tableau. Each part has it's own input files and these two parts are independent. You can perform these parts in any order, and both should be completed.

Welcome to the Informatica challenge.

You are provided with a food dataset with various customer and order details. As a BigData Engineer, your task involves cleaning the data, analyzing the data using Informatica PowerCenter.

INSTRUCTIONS–NOTE

- **If you don't know how to create session and workflow and running it. You can refer to the below given pdf Instructions.pdf.**

Data Preparation:

Oracle SQL Setup:

- Log in to Oracle SQL Developer in **Admin** connection using the credentials:
  - **Username:** system
  - **Password:** Admin

Create Tables with the provided SQL commands:

Food: create table food(customer_id varchar(255),gender varchar(255),age varchar(255),city varchar(255),signup_date varchar(255),order_id varchar(255),order_date varchar(255),restaurant_name varchar(255),dish_name varchar(255),category varchar(255),quantity number,price number,payment_method varchar(255),order_frequency number,last_order_date varchar(255),loyalty_points number,churned varchar(255),rating number,rating_date varchar(255),delivery_status varchar(255));

Food_Cleaned_Data: create table food_cleaned_data(customer_id varchar(255),gender varchar(255),age varchar(255),city varchar(255),signup_date date,order_id varchar(255),order_date date,restaurant_name varchar(255),dish_name varchar(255),category varchar(255),quantity number,price number,payment_method varchar(255),order_frequency number,last_order_date date,loyalty_points number,churned varchar(255),rating number,rating_date date,delivery_status varchar(255));

Customer_Behaviour: create table customer_behaviour(customer_id varchar(255), gender varchar(255), age varchar(255),category varchar(255), total_spend number, total_purchase number, avg_spend number);

Citywise_Delivery: create table citywise_delivery(city varchar(255), total_orders number, total_customers number, total_price number, delivered_orders number, avg_rating number, avg_loyalty number, avg_order_value number, loyalty_score number, delivery_success_rate number);

Load Data:

- Load **food_data.csv** into the **Food** table.
- File Path: ~/Desktop/Project/kickoffs-informatica-tableau-food-data/inputfile/Informatica/

Informatica Repository Setup:

- **Connect to Informatica Repository Manager:**
  - **Username:** Administrator
  - **Password:** Administrator

- **Create Folder:**
  - Create a folder named "**food**" .

Import Source Table:

- **Source Analyzer:**
  - Go to the Sources option.
  - Click on the Sources tab in the Source Analyzer.
  - Select Import from Database. An ODBC Connection box will open.
- Create ODBC Connection:
  - Click on the button next to ODBC Data Source.
  - On the next page, select User DSN tab and click Add.
  - Choose Oracle Wire Protocol.
  - Enter the database details:
    - Data Source Name: Oracle
    - Host: localhost
    - Port: 1521
    - SID: xe
  - Click Connect

Create Connections for Workflow Manager

- Create Relational Connection:
  - In Workflow Manager, click on the Connection menu.
  - Select Relational option.
  - In the pop-up window, select Oracle as the type.
  - Click New
  - In the connection object definition window:
    - Connection Name: Oracle
    - Username: system
    - Password: Admin
    - Connection String: xe
  - Leave other settings as default and click **OK.**

Analysis Tasks:

Task 1: Cleaned Data

- **Mapping Name:** m_cleaned_data

- **Workflow Name:** w_cleaned_data
- **Session Name:** s_cleaned_data
- **Target Table:** Food_Cleaned_Data

Problem Statement: Remove all the rows containing null values, change the datatypes of some columns and store the values in **"Food_Cleaned_Data"** target table.

Operations:

1. Import table **Boxoffice** from oracle as source.
2. Remove all the null values and duplicate values.
3. Change the datatype of the column **"Signup_Date"**, **"Order_Date"**, **"Last_Order_Date"** and **"Rating_Date"** to Date/Time.
4. Load the above data into the **Food_Cleaned_Data** target table (For columns check sample output), create the target table in Oracle SQL.

NOTE:

- **Refer to the INSTRUCTION–NOTE at the top.**
- **If you know how to create session and workflow you can proceed with the given configurations below.**

Create Session and workflow:

- Create Session **"s_cleaned_data"**
- Create Workflow "**w_cleaned_data"**.
- Check Workflow Monitor and verify if the workflow ran successfully and check the data loaded successfully into the target table.

Sample Output:

customer_id, gender, age, city, signup_date, order_id, order_date, restaurant_name, dish_name, category, quantity, price, payment_method,

order_frequency, last_order_date, loyalty_points, churned, rating, rating_date, delivery_status

C6987, Other, Teenager, Islamabad, 2025-08-18 00:00:00, O10987_2, 2025-06-28 00:00:00, Pizza Hut, Pizza, Continental, 2, 1434.95, Card, 34, 2025-05-25 00:00:00, 349, Inactive, 3, 2024-12-15 00:00:00, Cancelled

C2399, Other, Senior, Peshawar, 2024-12-30 00:00:00, O6399_1, 2025-07-06 00:00:00, Burger King, Pizza, Chinese, 4, 728.85, Cash, 37, 2025-04-02 00:00:00, 268, Active, 3, 2025-08-18 00:00:00, Delivered

NOTE: Food_Cleaned_Data table data is used as source for the tasks below.

Task 2: Customer Behaviour

- **Mapping Name:** m_customer_behaviour
- **Workflow Name:** w_customer_behaviour
- **Session Name:** s_customer_behaviour
- **Target Table: Customer_Behaviour**

Problem Statement: Find the Top 5 Age group of customers using the average spend and total purchase.

Operations:

1. Import table Food_Cleaned_Data from oracle as source.
2. For each Customer ID, Gender and Age, find the below columns:
    a. **TOTAL_SPEND** = Find the total price.
    b. **TOTAL_PURCHASE** = Find the count of orders.
3. Find the **AVG_SPEND** using the Total Spend and Total Purchase. **[NOTE: Make sure the output of this column must be a 2 decimal value]**

4. Arrange the data in ascending order for **"Customer ID"** and descending order for **"Total Spend"** to list all the Customers with highest spend.
5. Find the Top 5 Customer for each Age using Avg Spend.
6. Load data into the **Customer_Behaviour** target table **(For columns check sample output)**, create the target table in Oracle SQL

NOTE:

- **Refer to the INSTRUCTION–NOTE at the top.**
- **If you know how to create session and workflow you can proceed with the given configurations below.**

Create Session and workflow:

- Create Session **"s_customer_behaviour"**
- Create Workflow "**w_customer_behaviour"**.
- Check Workflow Monitor and verify if the workflow ran successfully and check the data loaded successfully into the target table.

Sample output:
customer_id, gender, age, category, total_spend, total_purchase, avg_spend
C4555, Other, Adult, Dessert, 2898.59, 2, 1449.3
C1468, Male, Adult, Italian, 2873.29, 2, 1436.65
C3602, Female, Adult, Chinese, 2867.45, 2, 1433.73

Task 3: Citywise Delivery

- **Mapping Name:** m_citywise_delivery
- **Workflow Name:** w_citywise_delivery
- **Session Name:** s_citywise_delivery
- **Target Table: Citywise_Delivery**

Problem Statement: Find the Delivery Success Rate and Loyalty score with other metrics for each city.

Operations:

1. Import table Food_Cleaned_Data from oracle as source.
2. Find the "DELIVERY_FLAG" when delivery status is 'Delivered' then set the flag as 1 otherwise 0.
3. For each City, find the below columns:
    a. TOTAL_ORDERS = Find the count of orders
    b. TOTAL_CUSTOMER = Find the count of customers
    c. TOTAL_PRICE = Find the total price.
    [NOTE: Total price must have non-decimal values]
    d. DELIVERED_ORDERS = Find the total of delivered flag
    e. AVG_RATING = Find the average of rating
    f. AVG_LOYALTY = Find the average of loyalty points
    [NOTE: Avg Rating and Avg Loyalty columns output must be a 2 decimal value]
4. Find the **"AVG_ORDER_VALUE"** using Total Price and Total Orders.
5. Find the **"LOYALTY_SCORE"** using the formula AVG_LOYALTY / (TOTAL_CUSTOMER + 0.1)
6. Find the **"DELIVERY_SUCCESS_RATE"** using the formula (DELIVERED_ORDERS / TOTAL_ORDERS) * 100
[NOTE: The output of the columns Avg Order Value, Loyalty Score and Delivery Success Rate must be a 2 decimal value]
7. Load data into the **CITYWISE_DELIVERY** target table **(For columns to check sample output)**, create the target table in Oracle SQL.

NOTE:

- **Refer to the INSTRUCTION–NOTE at the top.**
- **If you know how to create session and workflow you can proceed with the given configurations below.**

Create Session and workflow:

- Create Session **"s_citywise_delivery"**

- Create Workflow "**w_citywise_delivery"**.
- Check Workflow Monitor and verify if the workflow ran successfully and check the data loaded successfully into the target table.

Sample output:

city, total_orders, total_customers, total_price, delivered_orders, avg_rating, avg_loyalty, avg_order_value, loyalty_score, delivery_success_rate
Islamabad, 4178, 4178, 3346212, 1407, 3.02, 249.52, 800.91, .06, 33.68
Karachi, 4025, 4025, 3266012, 1257, 2.99, 253.75, 811.43, .06, 31.23
Lahore, 4276, 4276, 3429217, 1384, 2.98, 249.43, 801.97, .06, 32.37

Once you complete the challenge, make sure your output data is loaded into the respective target table in Oracle SQL

After completing the challenge, run **the sample_test.ps1 file** to check the sample score. **NOTE: Sample Score does not represent Actual Score.**

Click on the "Submit" button to validate your solution