Informatica Hands-On Challenge: Super_Store Analysis

- **Introduction:** You are provided with a sample dataset from a retail store, Super_Store. This dataset contains information about orders, customers, products, and sales. Your task involves cleaning the data, analyzing sales, customer orders, customer geography, and order processing time using Informatica PowerCenter.

Data Preparation:

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

Create a table named **Super_Store** by importing the superstore_data.csv into sql developer. Provide the table name while importing data and takes the data formats as per the data in the file.

Load **superstore_data.csv** into the **Super_Store** table.

You are given data set is in the **"~\Desktop\Project\miniproject-informaticasuper_store \"**

Informatica Repository Setup:

Connect to the Informatica repository manager using the following credentials:

- **Username:** Administrator
- **Password:** Administrator

Create a folder named **Super_Store** in the repository manager.

How to Import Source Table in Source Analyzer

Following are the steps to import source table in Informatica Source Analyzer:

Step 1) Go to "Sources" option In

source analyzer

1. Click on tab "Sources" from the main menu
2. Select import from database option, after this ODBC Connection box will open.

Step 2) Create ODBC connection

- We will now create ODBC connection
    1. Click on the button next to ODBC data Source(...).
    2. On the next page, Select user DSN tab and click Add button.
    3. Select oracle wire protocol
    4. On the next page, select the general tab and enter the database details. Then click connect.
- **Data Source name** : oracle
- **Host** : localhost
- **port** : 1521
- **sid** : xe

Create Connections for Workflow Manager

To Create a Relational Connection

Step 1: In Workflow Manager

- Click on the **Connection** menu
- Select **Relational** Option

Step 2: In the pop up window

- Select **Oracle** in type
- Click on the **new** button

Step 3: In the new window of connection object definition

- Enter Connection Name (oracle)
- Enter **username - system**
- Enter **password – Admin**
- Enter **connection string - xe**

- Leave other settings as default and Select OK button **Note :** For more **credentials**, like for designer, knidly check in the **Readme** File.

Note : Please Follow the naming conventions in the problem statement

Data Cleaning:

- **Mapping Name: NEW_**Map_Cleaned_Data
- **Workflow Name: NEW_**Workflow_Cleaned_Data
- **Session Name: NEW_**Session_Cleaned_Data
- **Target Table: NEW_**Super_Store_Cleaned_Data

Operations:

- Remove duplicates from the dataset to ensure data integrity.
- Filter records where Country is 'United States' and Customer id <> ' ' to focus on domestic orders.
- Extract numeric part from Customer_ID to standardize customer identification.(EX: CH-1234, extract 1234). Extract the first letter from Region.
- Concatenate extracted part from region with Customer_ID and Customer_Name with '-' to create a unique identifier for each customer. (Ex: S-1234-Charlies, Extracted Region-Extracted_ID-Customer_name) and store it in Customer_Id_Name Column
- Drop the customer_id, Customer_name, Region Columns.
- After cleaning Load data into the **NEW_Super_Store_Cleaned_Data** target table (For columns check sample output)
- **Sample Output :** Remaining columns and additional with **'CUSTOMER_ID_NAME'** COLUMN.

CUSTOMER_ID_NAME
W-21925-Zushuss Donatelli
C-16585-Ken Black
E-21520-Tracy Blumstein

NOTE : NEW_Super_Store_Cleaned_Data table data is used for the below every tasks.

Analysis Tasks:

Task 1: Sales Summary

- **Mapping Name:** NEW_Map_Sales_Summary
- **Workflow Name: NEW_**Workflow_Sales_Summary
- **Session Name: NEW_**Session_Sales_Summary
- **Target Table: NEW_**Sales_Summary

Problem Statement: Summarize total sales and average profit for each customer. Identify customers with significant contribution to overall sales.

Operations:

- Profit percentage for each category is given below.

CATEGORY	PROFIT PERCENTAGE
FURNITURE	12%
OFFICE SUPPLIES	18%
TECHNOLOGY	25%

- Now calculate the profit from each customer and store it in PROFIT column.
- Calculate the total sales and average profit for each customer. Round profit up to 2 decimal places. Sort the data in descending order based on average profit.
- Filter customers with total sales greater than 3000 and average profit greater than 250 to focus on significant contributors.
- Drop the unnecessary columns, kindly check the sample output.
- Load data into the **Sales_Summary** target table (For columns check sample output)
- After completing of mapping, in the workflow manager.

Sample Output:

CUSTOMER_ID_ NAME	TOTAL_SALES	AVG_PROFIT
S-14635-Grant Thornton	8167	1015.03
C-20290-Sean Braxton	5580	697.5
W-16540-Kelly Collister	3355	419.1

Task 2: Customer Order Analysis

- **Mapping Name:** NEW_Map_Order_Analysis
- **Workflow Name:** NEW_Workflow_Order_Analysis
- **Session Name:** NEW_Session_Order_Analysis

- **Target Table:** NEW_Order_Analysis

Problem Statement: Analyze customer orders to determine the most frequent buyers and their order patterns.

Operations:

- Filter records for customers in ship_mode is ' Standard Class ' and Segment in ' Consumer ' to analyze Consumer behavior.
- Calculate the count of orders for each subcategory to determine their order frequency.
- Categorize based on the number of orders. if Orders are less than 10 then 'Need to improve sales', orders are between 10-20 then 'Sales are going on', orders are greater than 20 then 'Good Profits'.
- Sort the results by order count in descending order to identify the most frequent buyers and get only bottom 10 records.
- Generate a unique number for each row into column ID
- Drop the unnecessary columns, kindly check the sample output.
- Load data into the NEW_**Order_Analysis** target table (For columns check sample output)

Sample output:

SNO	CUSTOMER_ID_NAME	ORDERS_COUNT	ORDERS_CATEGORY
1	Copiers	6	Need to improve sales
2	Machines	12	Sales are going on

3	**Bookcases**	**20**	**Sales are going on**

Task 3: Customer Geography Analysis

- **Mapping Name:** NEW_Map_Geography_Analysis
- **Workflow Name:** NEW_Workflow_Geography_Analysis
- **Session Name:** NEW_Session_Geography_Analysis
- **Target Table:** NEW_East_Customer_Base, NEW_South_Customer_Base, NEW_West_Customer_Base

Problem Statement: Analyze customer distribution across different regions to identify potential market segments.

Operations:

- Determine the region based on the customer_id_name column. Store it in region column.
- Calculate the total sales for each customer and store in total_sales column.
- Filter the data into respective tables based on the region.
- Drop the unnecessary columns, kindly check the sample output.
- Load data into the NEW_**East_Customer_Base, NEW_CENTRAL_Customer_Base, NEW_West_Customer_Base** target table (For columns check sample output)

Sample output: NEW_**EAST_CUSTOMER_BASE**

CUSTOMER	TOTAL_SALES
E-10060-Adam Bellavance	**4439**

E-10075-Adam Hart	**842**
E-10135-Adrian Shami	**54**

Sample output: NEW_**WEST_CUSTOMER_BASE**

CUSTOMER	**TOTAL_SALES**
W-10165-Alan Barnes	**199**
W-10180-Alan Dominguez	**4189**
W-10210-Alan Hwang	**80**

Sample output: NEW_**CENTRAL_CUSTOMER_BASE**

CUSTOMER	**TOTAL_SALES**
C-10180-Alan Dominguez	**601**
C-10360-Alice McCarthy	**57**
C-10375-Allen Armold	**178**

Task 4: Order Processing Time Analysis

- **Mapping Name:** NEW_Map_Order_Processing
- **Workflow Name:** NEW_Workflow_Order_Processing
- **Session Name:** NEW_Session_Order_Processing
- **Target Table:** NEW_Order_Processing **,** NEW_Year_Wise_Sales

Problem Statement: Evaluate order processing efficiency by analyzing the time taken between order placement and shipment,

Operations:

- Filter data where SHIP_MODE <> ' ' to avoid null values.
- Extract the year out of the order date and store it in new column.
- Calculate the total sales of each subcategory in each year and pick top 8 based on their sales performance. Load this into NEW_Year_Wise_Sales table.
- Count the number of orders falling with in each Ship mode for each customer.
- Filter the data and pick each record for a customer with highest orders in ship mode.
- Drop the unnecessary columns, kindly check the sample output.
- Load data into the NEW_**Order_Data,** (For columns check sample output)

Sample Output:

NEW_**Year_Wise_Sales**

YEAR	SUBCATEGORY	TOTAL_SALES
2015	Accessories	5775
2016	Accessories	2805
2017	Accessories	5774

NEW_**Order_Data**

CUSTOMER	SHIP_MODE	ORDERS_COUNT
C-10180-Alan Dominguez	Second Class	1
C-10360-Alice McCarthy	Standard Class	2
C-10375-Allen Armold	First Class	2