

Informatica Challenge

NOTE: You are given tasks for both Informatica and Tableau. Each part has its 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 an Amazon book dataset that contains information about the book's details. Your task involves cleaning the data and 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:

Cellphones_data:

```
CREATE TABLE cellphones_data ( cellphone_id NUMBER, brand VARCHAR(255), model_of_mobile VARCHAR(255), operating_system VARCHAR(255), internal_memory NUMBER, RAM NUMBER, performance_of_mobile NUMBER, main_camera NUMBER, selfie_camera NUMBER, battery_size NUMBER, screen_size NUMBER, weight NUMBER, Price NUMBER, release_date DATE, quantity NUMBER, quantity_sold NUMBER );
```

cellphones_cleaned_data

```
CREATE TABLE cellphones_cleaned_data ( cellphone_id NUMBER, brand VARCHAR(255), model_of_mobile VARCHAR(255), operating_system VARCHAR(255), internal_memory NUMBER, RAM NUMBER, performance_of_mobile NUMBER, main_camera NUMBER, selfie_camera NUMBER, battery_size NUMBER, screen_size NUMBER, weight NUMBER, Price NUMBER, release_date DATE, quantity NUMBER, quantity_sold NUMBER );
```

battery_category:

```
create table battery_category(cellphone_id number, brand varchar(255),model_of_mobile
```

varchar(255),operating_system varchar(255),battery_capacity varchar(255), Price number,battery_status varchar(255));

price_analysis:

create table price_analysis(brand varchar(255), total_amount_of_sales number,rank NUMBER);

Load Data:

- Load **cellphones_data.csv** into the **cellphones_data** table.
- File Path: ~/Desktop/Project/kickoffs-informatica-tableau-cellphone-data/inputfile/informatica/

Informatica Repository Setup:

- **Connect to Informatica Repository Manager:**
 - **Username:** Administrator
 - **Password:** Administrator
- **Create Folder:**
 - Create a folder named “**cellphone_data**”.

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:** cellphones_cleaned_data

Problem Statement: Clean the data and store the values in “**cellphones_cleaned_data**” target table.

Operations:

1. Import source table **cellphones_data** from oracle
2. Remove the rows containing null values in any of the columns.
3. Drop the duplicate rows from the data.
4. Load the cleaned data into “**cellphones_cleaned_data**” target table.

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:

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

Task 2: Price Analysis

- **Mapping Name:** m_price_analysis
- **Workflow Name:** w_price_analysis
- **Session Name:** s_price_analysis
- **Target Table:** Price_analysis

Problem Statement: Fetch the top 10 highest sum of sold price for each brand.

Operations:

1. Import table **cellphones_cleaned_data** from oracle as source.
2. Create a column “**total_amount_of_sales**” which calculates the total amount of sales. [formula: quantity_sold * price]
3. Create a column “**sum_of_total_price**” which calculates the sum of total amount of sales for each “**brand**”.
4. Finally, Rank the data based on the **sum of total price** and fetch the top
5. Fetch the **top 10** records based on the sum of total price and rank it.
6. Load this data into the **Price_analysis** target table

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_price_analysis**”
- Create Workflow “**w_price_analysis**”.
- Check Workflow Monitor and verify if the workflow ran successfully and check the data loaded successfully into the target table.

Sample output:

Task 3: Battery Category

- **Mapping Name:** m_battery_category
- **Workflow Name:** w_battery_category
- **Session Name:** s_battery_category
- **Target Table:** BATTERY_CATEGORY

Problem Statement: Analyzing the discount based on each discount category.

Operations:

1. Import table **cellphones_cleaned_data** from oracle as source.
2. Create a column named as “**batter_status**” and flag the string based on the condition below

Condition	Flag
Battery_Size>=5000	High
Battery_Size >=3500 and Battery_Size<5000	Medium
Battery_Size<3500	Low

3. Create a column named as “**battery_capacity**” which appends the string “**mAh**” at the end of the “**battery_size**” column.
[Example:]
4. Filter the data which is having the brand as “**Samsung**” and which contains the “**Galaxy**” string in the **model of the mobile** column.
5. Sort this data based on the “**battery_size**” and “**price**” in descending order for both columns.
6. Load data into the **BATTERY_CATEGORY** 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_battery_category”

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

Sample output:

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 **sampletest.ps1 file** to check the sample score
Click on the **submit** button to validate your solution.