**Case Study**  
 **Data Engineer**

We have five datasets. Let us have a short overview of what each dataset contains.

|  |  |
| --- | --- |
| Dataset Name | Description |
| customer\_details | Contains customer information. |
| order\_details | Contains order information, which is ordered by customers and quantity. |
| product\_details | Contains product information with other details. |
| carier\_details | Contains carier details with product id. |
| carrier\_price\_details | Carier price details contains price information for each carier. |
| currency | Currency contains currency |
| parameter | Contains discount rate for each category. |

We have customer’s order data which has come from different data sources in a excel format. Order details containing customer who has placed purchase order for product with quantity.

We need to calculate selling price for each order.

Tasks that needs to be done:

We need to perform some manipulations on the given dataset, to produce an Output Report, which should contain the information based on the requirements.

Below we have the requirements for our Output Report:

|  |  |  |
| --- | --- | --- |
| Column Name | Description | \*\*\* |
| User ID | User id |  |
| User name | User Name |  |
| Order Id | Order Id |  |
| Product Id | Product Id |  |
| Product Name | Product Name |  |
| Product Category | Product Category |  |
| quantity | Quantity of the order. |  |
| Actual Price | Calculate Actual price in EUR. | It there is multiple price found take the valid latest price. order date should be less than the valid to of the carier price. |
| Discounted Price | Calculate Discounted price in EUR(Mention discount percentage) | Apply all discount percentage to the actual price for the entire category. |
| Currency | Currency information |  |
| savings | Percentage of savings for each order. |  |

**Tools & Technologies that can be used** – ETL tool such as Airflow, R, and Python.

Note:- If the price information is not found for a order, average of all the latest(always) carier price can be used to calculate missing order price information.