**Fevo Data Scientist Challenge**

**Data**

1. order\_data: This data contains 100 records of order-level data.
2. product\_sold: This data contains the product sold for each order
3. order\_with\_product\_sold: This denormalized data from the above 2 data set.

**Instructions**

Write detailed description on how you would analyze the data and/or provide query code with your results.

**Data Processing Challenge**

Using the order\_data:

1. Identify data anomalies and other potential data quality issues
2. What data transformation would you perform?
3. Identify duplicate records
4. Create a new column to sequentially rank the order within each customer by chronological order.
5. Parse and extract ‘product\_variant\_id’ and ‘price’ from the orderLineItem\_jsonb object.

**Data Analysis Challenge**

Using the data sets:

1. What additional feature would you create from the data set?
2. What are the affinities between the products?
3. Create a model to estimate the customer lifetime value
4. Create a model to predict the probability of purchase “Beam Blue Sweats”
5. What are the key drivers for repeat purchasers?
6. Forecast sales revenue for the month of July
7. How to determine the customers relationships through social cluster?
8. What additional insights can be derived from the data?

|  |  |  |
| --- | --- | --- |
| **Column** | **Data Type** | **Description** |
| id | varchar | Unique record identifier |
| order\_number | varchar | Order ID |
| order\_revision | integer | Order revision number |
| order\_created\_at\_utc | timestamp | Order created timestamp (UTC) |
| sub\_total\_amount | decimal(12,2) | Order sub total amount |
| tax\_total\_amount | decimal(12,2) | Order tax total amount |
| shipping\_total\_amount | decimal(12,2) | Order shipping total amount |
| fee\_total\_amount | decimal(12,2) | Order fee total amount |
| order\_total\_amount | decimal(12,2) | Order total amount |
| requires\_payment | boolean | Order requires payment |
| charged\_amount | integer | Charge payment amount |
| charge\_refunded | boolean | Is the charge refunded? |
| refunded\_amount | integer | Amount refunded |
| refunded\_at\_utc | timestamp | Order refunded timestamp |
| customer\_id | varchar | Customer ID |
| age | integer | Age of customer |
| gender | varchar | Gender of customer |
| income | integer | Income of customer |
| group\_id | varchar | Customer's social group |
| group\_name | varchar | Customer's social group name |
| customer\_state | varchar | Customer shipping address (State) |
| customer\_country | varchar | Customer shipping address (Country) |
| device\_type | varchar | Device the customer used to place order |
| orderlineitems\_jsonb | jsonb | Items purchased |
| price | decimal(12,2) | Price of the product |
| product\_id | varchar | Product ID (parent) |
| product\_title | varchar | Product Title (parent) |
| product\_variant\_id | varchar | Product variant ID |
| product\_variant\_title | varchar | Product variant title |