**1. Description**: Your task is to process and analyse the data that was provided to you to answer couple of business questions. Data is not 100% clean, so feel free to clean it as per your views. Please write down what has been cleaned so that we can follow your thought process. We have prepared a logical diagram explaining overall data structure. In general, a merchant can have multiple shops, each connected to specific sales channel. End customer can make a merchandise order through that sales channel. An order can contain multiple different products (line items). Each line item can be fulfilled by different print provider and shipped by different carrier.

A diagram of a product

Description automatically generated

**2. Data**

Orders and line items data can be found in orders.csv and line\_items.csv inside data.zip provided to you. Column descriptions is share below:

**Orders**

|  |  |
| --- | --- |
| Column name | Description |
| MERCHANT\_ID | Unique ID of merchant |
| ORDER\_ID | Unique ID of order coming from merchant |
| SHOP\_ID | Unique ID of merchant's shop |
| ADDRESS\_TO\_COUNTRY | Country to where the order is shipped |
| ADDRESS\_TO\_REGION | Region to where the order is shipped |
| ORDER\_DT | Timestamp when order has been created |
| FULFILLED\_DT | Timestamp when order has been printed |
| SALES\_CHANNEL\_TYPE\_ID | ID of sales channel from which merchant generated order |
| TOTAL\_COST | Amount of money collected for product (cents) |
| TOTAL\_SHIPPING | Amount of money collected for shipping services (cents) |
| MERCHANT\_REGISTERED\_DT | Timestamp when merchant has registered in the platform |
| SUB\_IS\_ACTIVE\_FLAG | Has merchant currently subscribed to any subscription plan |
| SUB\_PLAN | What is the subscription plan that the merchant has subscribed to |
| SHIPMENT\_CARRIER | Shipping carrier chosen for shipment of the order |
| SHIPMENT\_DELIVERD\_AT | Timestamp when shipment has been delivered |

**Line items**

|  |  |
| --- | --- |
| Column name | Description |
| ORDER\_ID | Unique ID of order coming from merchant |
| PRINT\_PROVIDER\_ID | Unique ID of a print provider that is printing this specific line item |
| PRODUCT\_BRAND | Brand of blank product |
| PRODUCT\_TYPE | Product type |
| QUANTITY | Ordered quantity of specific line item |
| REPRINT\_FLAG | If line item needed to be reprinted |

**3. Analysis Questions:**

3.1 Review and clean data if necessary

3.2 What characteristics do the most successful merchants share?

3.3 What are the top two shipping carriers? Why should or shouldn’t we try to use those two for all orders?

3.4 Print Providers control the print quality, stock of items, and production time (the time from ordered to fulfilled). We want to provide a discount to the two best Print Providers and end our contracts with the worst two. Which do you choose and why?

**4. SQL Exercises:**

4.1 Write a query returning total sales, orders, and count of merchants by month;

4.2 Write a query returning merchants total sales, product count, and order count ordered by order count for merchants with more than 5 orders;

4.3 Write a query returning all ORDER\_IDs with the time the merchant has been active at the time of the order, the rank of the merchant by order count for the previous month, and the merchant's primary sales channel for the previous month;

4.4 Write a statement to create a table containing print providers with average production time, reprint percent, last order timestamp, and primary shipping carrier;

**5. Results:**

1. Analysis questions:

Feel free to do the homework in a tool or way that is most comfortable for you, but please summarise the results of analysis in Power BI covering the most important aspects and outcomes.

a. Data cleaning (3.1): please do describe the data cleaning steps you performed. This will be reviewed but you won’t be asked to present it live.

b. Other questions: In a real setting, the presentation is intended for Business stakeholders that requested the analysis. So please answer questions 3.2-3.4 keeping business stakeholders in mind.

2. Share the SQL queries for 4.1-4.4 in text/sql files.

You are welcome to make assumptions, just please write them down. Please also share work in progress files or any code you have generated.