Woolsocks assignment BI Data Analyst

SQL Task

1. Write a query to structure the data from the files *woolsocks.backend_events_acc.events_rewards_giftcards_orders.csv* and *woolsocks.bi_giftcards_acc.giftcard_configs.csv.*

   The first file includes events of our gift card giveaway flow. A user can give away a gift card to another person (that is not on the Woolsocks platform).

   Each gift card (orderID) can have up to 3 statuses:
   a. waiting to be claimed: the gift card is sent to a possible new user
   b. claimed: the gift card receiver registered its email at Woolsocks
   c. received: the gift card receiver fully completed the signup and received the gift card

   Conditions:
   - GiftCardTemplateID cannot be blank
   - Order status is 'WAITING_TO_BE_CLAIMED' &'ORDER_CLAIMED' & 'DELIVERED'

   Output:

   A table that contains one row per OrderID. The table should contain at least OrderID, BuyerUserID, GiftCardTemplateID, name, GiftcardAmount, Currency, CommisionPercentage and the Order statuses including the date of update Order status.

2. Write a query that returns all user_ids that were the first user_id to connect a unique IBAN. This query will help you identify users that connect a bank account already connected by another user_id.

   - To determine if a bank account is unique you can use the IBAN (this is hashed value).
   - Registration_date is the registration date of the user_id
   - Connect_datetime is the date that the bank (IBAN) was connected.

   File: *woolsocks.bankaccounts.csv*

3. Write another query to determine which country and which registration period (month or weeks) Woolsocks encountered the most fraud by users that connected a bank account that was already connected by another user.

   File: *woolsocks.bankaccounts.csv*

* Provide query including notes.

Task Engineering/Visualization

- Store bank transaction data and build a visualization showing market trends (this is an open-ended task).

  Subtasks:

- Store the file in a place where you can retrieve it via REST call (or do a manual dump into a database (like Google Bigquery)
- Extract data from the shared CSV file and transform the data
- Load the transformed data into a database (for example Google Bigquery)
- Create visualizations including market insights (for example market share, market penetration and segmentation on demographics). You can do this by using any visualization tool (PowerBi, Tableau, Google Data Studio etc.)
- You can create your own user table including country, gender, age group etc.
- Provide documentation on the design and implementation of the data pipeline

Dataset

The dataset is a fake dataset of supermarket purchases in the Netherlands for the period of September through December. The following merchants are included:

- Albert Heijn
- Aldi
- Dirk
- Hoogvliet
- Jan Linders
- Jumbo
- Lidl
- Plus
- Vomar
- Picnic