FWD Data Analytics Test

Instruction

Before Doing this assignment please do this to setup the environment

1. Download the .db file from this link
2. Open querysandbox
3. Once on the QuerySandbox website, go to the top menu and select File -> OpenSQLiteDB. Upload the downloaded .db file to the QuerySandbox.

Case Background

X Company is a well-established insurance provider offering various types of insurance policies, including health, auto, and home insurance. The company prides itself on providing excellent customer service and maintaining comprehensive records of its customers, policies, and claims.

Table Information

Table 1: Customers

customer_id	name	date_of_birth	address	phone
1	Alice Smith	1990-05-15	123 Maple Street	123-456-7890
2	Bob Johnson	1980-11-22	456 Oak Avenue	987-654-3210
3	Carol Williams	1965-08-09	789 Pine Road	555-666-7777

Columns Description

- customer_id: INT - Unique identifier for each customer
- name: VARCHAR - Name of the customer
- date_of_birth: DATE - Date of birth of the customer
- address: VARCHAR - Address of the customer
- phone: VARCHAR - Phone number of the customer

Table 2: Policies

policy_id	customer_id	policy_type	start_date	end_date	premium
1	1	Health	2022-01-01	2023-01-01	500.00
2	1	Auto	2021-05-15	2022-05-15	300.00
3	2	Home	2022-06-01	2023-06-01	1000.00
4	3	Auto	2022-07-01	2023-07-01	450.00

Columns Description

- policy_id: INT - Unique identifier for each policy
- customer_id: INT - Identifier for the customer holding the policy
- policy_type: VARCHAR - Type of the policy (e.g., health, auto, home)
- start_date: DATE - Start date of the policy

- end_date: DATE - End date of the policy
- premium: DECIMAL - Premium amount for the policy

Table 3: Claims

claim_id	policy_id	claim_date	claim_amount	status
1	1	2022-02-15	200.00	approved
2	1	2022-05-10	150.00	approved
3	2	2021-07-22	300.00	approved
4	3	2022-08-10	400.00	pending
5	4	2023-01-05	250.00	rejected

Columns Description

- claim_id: INT - Unique identifier for each claim
- policy_id: INT - Identifier for the policy the claim is made on
- claim_date: DATE - Date the claim was made
- claim_amount: DECIMAL - Amount claimed
- status: VARCHAR - Status of the claim (e.g., approved, pending, rejected)

Query Task

You are required to write a SQL query that combines data from the Customers, Policies, and Claims tables to produce the report described below. The query should meet the following criteria:

1. Filter the claims to include only those made in 2023.
2. Group the results by customer and policy type.
3. Calculate the necessary aggregates:
    - Total number of claims per customer and policy type.
    - Total claim amount per customer and policy type.
    - Average premium per customer and policy type.
4. Classify customers into age groups: under 30, 30-60, and over 60.
5. Include only customers who have made more than 2 claims in the last year.

Output Sample

Name	Policy Type	Total Claims	Total Claim Amount	Average Premium	Age Group
Michael Larsen	Auto	6	2740.06	1217.215	30-60
Michael Larsen	Health	4	1457.36	1415.92	30-60
Ivan Smith	Auto	5	2633.8	1852.47	30-60
Steven Crawford	Health	6	3191.2	1372.4416666666	30-60
Steven	Home	4	3006.55	952.99	30-60

Crawford					
Jesse Dyer	Home	3	1697.01	1071.8	Under 30
Barbara Case	Health	5	3361.43	1358.03	Over 60
Jennifer Abbott	Health	3	1054.98	907.0033333333334	30-60
Jennifer Ward	Health	3	1442.16	1202.06	30-60
Matthew Davis	Auto	3	1499.01	249.46	Over 60