

FINAL TEST D4E

Người ra đề: Nguyễn Lê Chí Bảo

Dataset

Cho dataset: sales.csv, bao gồm các column chính sau đây:

Order ID: Mã đơn hàng

Order Date (dd/MM/yyyy): Ngày đặt hàng

Ship Date: Ngày giao hàng

Ship Mode: Hình thức ship

Customer ID: Mã khách hàng

Product ID: Mã sản phẩm

Category: Nhóm sản phẩm

Sub- category: Dòng sản phẩm

Quantity: Số lượng sản phẩm

Sale: Doanh thu

Profit: Lợi nhuận

PHẦN 1: SQL (4 điểm)

Câu 1 (0.75 đ): Lấy thông tin về Mã đơn hàng, mã sản phẩm, mã khách hàng, số lượng sản phẩm của những dòng dữ liệu thỏa điều kiện Ship Mode là Standard Class

Câu 2 (0.75 đ): Lấy thông tin về những mã đơn hàng (orderId) của **những dòng dữ liệu** thỏa mãn điều kiện sản phẩm (Product ID) thuộc category là Office Supplies và có quantity > 3

Câu 3 (1đ) : Thống kê số lượng mã đơn hàng, số lượng các loại sản phẩm (product ID), tổng doanh thu và tổng lợi nhuận theo từng Category, sắp xếp theo thứ tự giảm dần của doanh thu

Câu 4 (1đ): Với mỗi loại Ship mode, lấy ra thông tin khách hàng (Customer ID), số lượng đơn hàng sao cho có số lượng đơn hàng của khách hàng đó theo hình thức Ship mode đang xét là nhiều nhất.

Ship_Mode	Customer_ID	TotalOrder
First Class	Dp-13240	5
First Class	JM-16195	5
First Class	SZ-20035	5
Same Day	RB-19465	3
Same Day	SU-20665	3
Same Day	JK-15640	3
Same Day	CM-12190	3
Second Class	CC-12670	6
Second Class	JD-15895	6
Standard Class	EP-13915	12

Câu 5 (1đ): Viết 1 câu query trả về 1 table với điều kiện như sau: với mỗi dòng dữ liệu, thêm 1 column có tên là **totalSaleBefore**: Tổng số doanh thu của các đơn hàng mà trước đó customer đó thực hiện (Bao gồm cả đơn hàng hiện tại). **Những đơn hàng trước đó chính là những đơn hàng có Order Date <= Ngày của đơn hàng đang xét.**

Kết quả mẫu khi **select một số column**:

Order_Line	Order_ID	Order_Date	customer_ID	Product_ID	Sales	Quantity	totalSaleBefore
1	CA-2016-152156	2016-11-08	CG-12520	FUR-BO-10001798	261.96	2	1130.62
2	CA-2016-152156	2016-11-08	CG-12520	FUR-CH-10000454	731.94	3	1130.62
3	CA-2016-138688	2016-06-12	DV-13045	OFF-LA-10000240	14.62	2	14.62
4	US-2015-108966	2015-10-11	SO-20335	FUR-TA-10000577	957.58	5	979.95
5	US-2015-108966	2015-10-11	SO-20335	OFF-ST-10000760	22.37	2	979.95
6	CA-2014-115812	2014-06-09	BH-11710	FUR-FU-10001487	48.86	7	3714.29
7	CA-2014-115812	2014-06-09	BH-11710	OFF-AR-10002833	7.28	4	3714.29
8	CA-2014-115812	2014-06-09	BH-11710	TEC-PH-10002275	907.15	6	3714.29
9	CA-2014-115812	2014-06-09	BH-11710	OFF-BI-10003910	18.5	3	3714.29
10	CA-2014-115812	2014-06-09	BH-11710	OFF-AP-10002892	114.9	5	3714.29
11	CA-2014-115812	2014-06-09	BH-11710	FUR-TA-10001539	1706.18	9	3714.29
12	CA-2014-115812	2014-06-09	BH-11710	TEC-PH-10002033	911.42	4	3714.29
13	CA-2017-114412	2017-04-15	AA-10480	OFF-PA-10002365	15.55	3	1790.51
14	CA-2016-161389	2016-12-05	IM-15070	OFF-BI-10003656	407.98	3	4930.489999999998
15	US-2015-118983	2015-11-22	HP-14815	OFF-AP-10002311	68.81	5	393.36
16	US-2015-118983	2015-11-22	HP-14815	OFF-BI-10000756	2.54	3	393.36
17	CA-2014-105893	2014-11-11	PK-19075	OFF-ST-10004186	665.88	6	665.88
18	CA-2014-167164	2014-05-13	AG-10270	OFF-ST-10000107	55.5	2	55.5

PHẦN 2: POWERBI (3 điểm)

Câu 1 (1.5 điểm): Thực hiện vẽ một số biểu đồ sau:

- Vẽ biểu đồ thể hiện được tình hình doanh thu của cửa hàng theo từng ngày/tháng/ năm
- Vẽ biểu đồ thể hiện được tổng số lượng sản phẩm được bán ra theo từng category
- Vẽ biểu đồ thể hiện được tỷ trọng lợi nhuận giữa các category trên tổng số lợi nhuận của cửa hàng.

Câu 2 (1.5 điểm): Phân tích tình hình phát triển của cửa hàng qua các năm.

Câu 3 (1 điểm): Thực hiện việc tạo ra các measure để thể hiện được các nhu cầu sau:

3a) Hiện nay cửa hàng đang trên đà phát triển, tuy nhiên vì lượng dữ liệu khá lớn nên chủ cửa hàng khá khó khăn trong việc thống kê lại tổng lợi nhuận mà cửa hàng thu được từ đầu năm cho tới một ngày nào đó trong năm là bao nhiêu? Bạn hãy tạo ra 1 measure để giúp chủ cửa hàng tính ra được tổng lợi nhuận mà cửa hàng thu được từ đầu năm cho tới ngày đang xét.

Ví dụ ngày đang xét là ngày 20/7/2018=> số liệu cần xem chính là tổng lợi nhuận từ ngày 1/1/2018 cho tới ngày 20/7/2018

3b) Chủ cửa hàng hiện tại chỉ sử dụng những thống kê tổng để biết được tình hình cửa hàng từ lúc khai trương đến nay đang như thế nào, mang tiếng là cửa hàng đang có vẻ phát triển. Tuy nhiên, chủ cửa hàng muốn chứng minh điều đó chứ không phải là “có vẻ phát

triển”, vì vậy chủ cửa hàng muốn tính xem thử với mỗi ngày bất kỳ, thì lợi nhuận của ngày hôm đó so với ngày đó năm trước thì cao gấp bao nhiêu lần, ngày này tháng trước thì cao gấp bao nhiêu lần. Bạn hãy giúp cửa hàng xây dựng các measure để tính ra được lợi nhuận của ngày này tháng trước và lợi nhuận của ngày này năm trước là bao nhiêu
Ví dụ: Ngày đang xét là ngày 20/7/2018, thì ta tính lợi nhuận của ngày 20/7/2018 với 20/7/2017 và tính ngày 20/6/2018.

PHẦN 3: PYTHON (3 điểm)

Python syntax:

Câu 1 (1đ): **Số hoàn hảo** (hay còn gọi là **số hoàn chỉnh**, **số hoàn thiện** hoặc **số hoàn thành**) là một số nguyên dương mà tổng các ước nguyên dương thực sự của nó (các số nguyên dương bị nó chia hết ngoại trừ nó) bằng chính nó.

In ra số lượng các số là số hoàn hảo trong khoảng giá trị từ 1 cho tới 10000.

VD: 6, 28, 496 là các số hoàn hảo.

Câu 2 (1đ): In ra 4 giá trị như sau: tổng giá trị của các số nguyên tố, tổng giá trị của các số chẵn, tổng giá trị của các số là số chính phương, tổng giá trị của các số là bội số của 10 trong khoảng giá trị từ 1 cho tới 10000.

Python pandas (1.5đ):

Hiện thực Câu 1->3 của phần SQL bằng Pandas. Mỗi câu 0.5 điểm