

Đề thi:

DATABASE SQL AND DATA COLLECTION FOR DATA SCIENCE

Ngày thi : 16/03/2022

***** Học viên tạo 1 thư mục là *MDS4_282_HoVaTen*, lưu tất cả bài làm vào để nộp chấm điểm *****

***** Mỗi phần tạo một tập tin jupyter để giải quyết các yêu cầu *****

***** Học viên được sử dụng tài liệu *****

***** Với mỗi câu, sử dụng Markdown để mô tả yêu cầu *****

Phần 1 : Truy vấn và Trục quan hóa dữ liệu (7 điểm)

Thời gian làm bài: 120 phút

Sử dụng cơ sở dữ liệu **MySQL** hoặc **MariaDB** để import dữ liệu từ tập tin **salesdb.sql** được cung cấp và thực hiện các yêu cầu sau :

Câu 1: Sử dụng Magic SQL hoặc Python Connector thực hiện các câu lệnh truy vấn (5 điểm)

1. Liệt kê các sản phẩm của nước Nhật theo mẫu sau, sắp tăng theo city (0.5 điểm)

City	CompanyName	ProductName	UnitPrice
Osaka	Mayumi's	Konbu	6
Osaka	Mayumi's	Tofu	23.25
Osaka	Mayumi's	Genen Shouyu	15.5
Tokyo	Tokyo Traders	Mishi Kobe Niku	97
Tokyo	Tokyo Traders	Ikura	31
Tokyo	Tokyo Traders	Longlife Tofu	10

2. Thống kê theo Số sản phẩm và Đơn giá trung bình được cung cấp theo từng quốc gia theo mẫu sau, dữ liệu được sắp tăng dần theo Số sản phẩm và giảm dần theo Đơn giá trung bình (0.5 điểm)

Country	avg_unitprice	count_product	count_city
Singapore	32.73	2	1
Spain	29.5	2	1
Denmark	10.75	2	1
Norway	20.0	3	1
Finland	18.08	3	1
Netherlands	10.67	3	1
Canada	27.31	4	2
France	76.75	5	3
Italy	27.36	5	2

3. Liệt kê các đơn đặt hàng đặt trong tháng 3/2014 theo mẫu sau, dữ liệu sắp giảm dần theo orderdate và totalAmount (0.5 điểm)

FirstName	LastName	OrderNumber	OrderDate	TotalAmount
Patricia	McKenna	543115	30/03/2014	2248.2
Yvonne	Moncada	543116	30/03/2014	2220
Jose	Pavarotti	543114	30/03/2014	1809.75
Mario	Pontes	543111	27/03/2014	15810
Elizabeth	Lincoln	543112	27/03/2014	1014
Jose	Pavarotti	543113	27/03/2014	796.5
Maria	Larsson	543110	27/03/2014	310
Roland	Mendel	543109	26/03/2014	4813.5
Maria	Larsson	543107	26/03/2014	2233

4. Thống kê sản phẩm có tổng thành tiền nhỏ nhất trong mỗi năm theo mẫu sau (1 điểm)

order_year	ProductName	total_amount
2012	Laughing Lumberjack Lager	56.0
2013	Chef Anton's Gumbo Mix	405.65
2014	Louisiana Hot Spiced Okra	17

5. Liệt kê các đơn đặt hàng có ngày đặt hàng là ngày chủ nhật của tháng 3/2014 theo mẫu sau, dữ liệu sắp giảm theo orderdate (0.5 điểm)

OrderNumber	OrderDate	customer_name	TotalAmount
543114	2014-03-30 00:00:00	Jose Pavarotti	1809.75
543115	2014-03-30 00:00:00	Patricia McKenna	2248.2
543116	2014-03-30 00:00:00	Yvonne Moncada	2220
543097	2014-03-23 00:00:00	Karin Josephs	910.4
543098	2014-03-23 00:00:00	Roland Mendel	1408
543099	2014-03-23 00:00:00	Pedro Afonso	108
543080	2014-03-16 00:00:00	Giovanni Rovelli	110
543081	2014-03-16 00:00:00	Michael Holz	482.9
543082	2014-03-16 00:00:00	Maria Anders	491.2
543083	2014-03-16 00:00:00	Thomas Hardy	4675
543064	2014-03-09 00:00:00	Renate Messner	500

6. Thống kê số lần đặt hàng của các customer theo mẫu sau, chỉ in các customer nào có số lần đặt hàng >15, dữ liệu sắp giảm dần theo số lần đặt (0.5 điểm)

customer_name	Country	City	count_orders
Jose Pavarotti	USA	Boise	31
Roland Mendel	Austria	Graz	30
Horst Kloss	Germany	Cunewalde	28
Maria Larsson	Sweden	Bräcke	19
Patricia McKenna	Ireland	Cork	19
Carlos Hernández	Venezuela	San Cristóbal	18
Christina Berglund	Sweden	Luleå	18
Paula Wilson	USA	Albuquerque	18
Laurence Lebihan	France	Marseille	17

7. Cho biết 2 quý nào có tổng thành tiền bán cao nhất (0.5 điểm)

year	quarter	sum_totalamount
2014	1	315242.12
2013	4	193718.12

8. Liệt kê thông tin của 10 đơn đặt hàng đặt mua sản phẩm có tên là Chai theo mẫu sau, dữ liệu sắp giảm theo TotalAmount (0.5 điểm)

Id	OrderDate	OrderNumber	CustomerId	TotalAmount
444	2013-10-03 00:00:00	542821	63	10164.8
600	2014-01-22 00:00:00	542977	71	6164.9
275	2013-04-30 00:00:00	542652	44	2657.8
591	2014-01-19 00:00:00	542968	47	2584.5
784	2014-04-17 00:00:00	543161	71	2393.5
38	2012-08-20 00:00:00	542415	63	2179.2
166	2013-01-14 00:00:00	542543	41	2123.2
453	2013-10-10 00:00:00	542830	71	2048
159	2013-01-07 00:00:00	542536	62	2018.2
671	2014-03-02 00:00:00	543048	10	1930

9. Cho câu truy vấn như sau : (0.5 điểm)

```
SELECT p.ProductName, SUM(i.Quantity) AS sum_quantity
FROM products p LEFT JOIN orderitems i ON p.Id=i.ProductId LEFT JOIN orders o ON o.Id=i.OrderId
WHERE p.IsDiscontinued=0
GROUP BY p.ProductName
ORDER BY sum_quantity;
```

Và kết quả là:

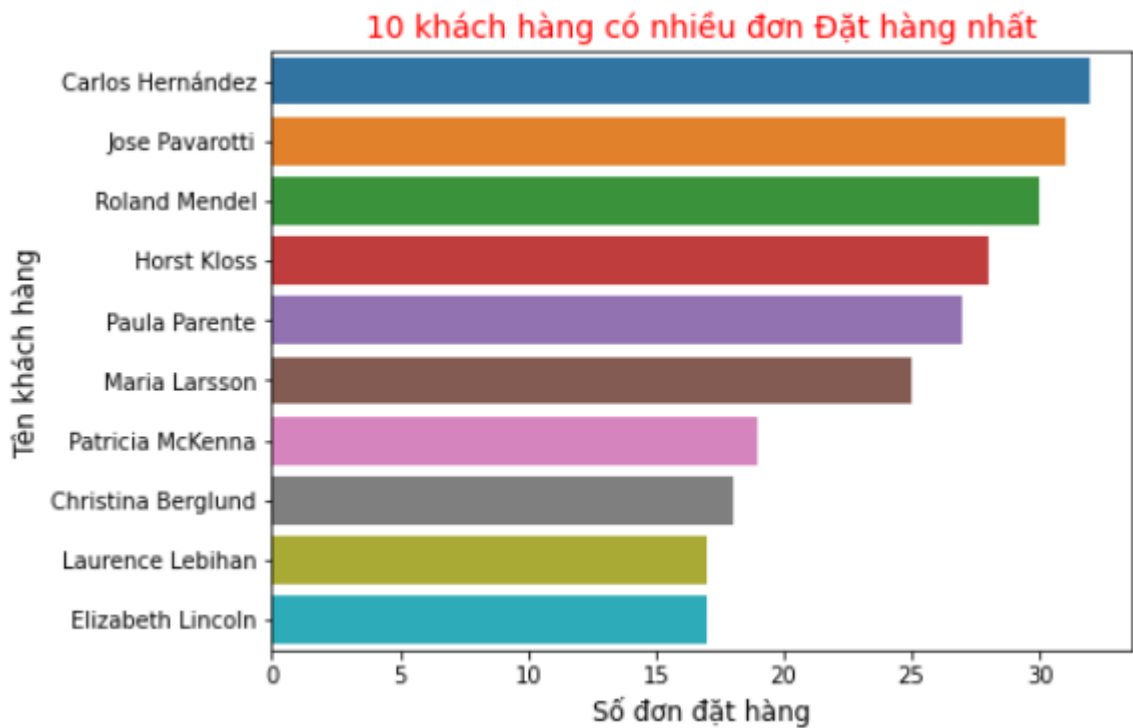
ProductName	sum_quantity
Stroopwafels	None
Genen Shouyu	122
Gravad lax	125
Chocolade	138
Laughing Lumberjack Lager	184
Valkoinen suklaa	235
Louisiana Hot Spiced Okra	239

- Hãy giải thích ý nghĩa của câu truy vấn trên
- Thay thế giá trị NULL (None) của cột sum_quantity thành giá trị 0

ProductName	sum_quantity
Stroopwafels	0
Genen Shouyu	122
Gravad lax	125
Chocolade	138
Laughing Lumberjack Lager	184
Valkoinen suklaa	235
Louisiana Hot Spiced Okra	239

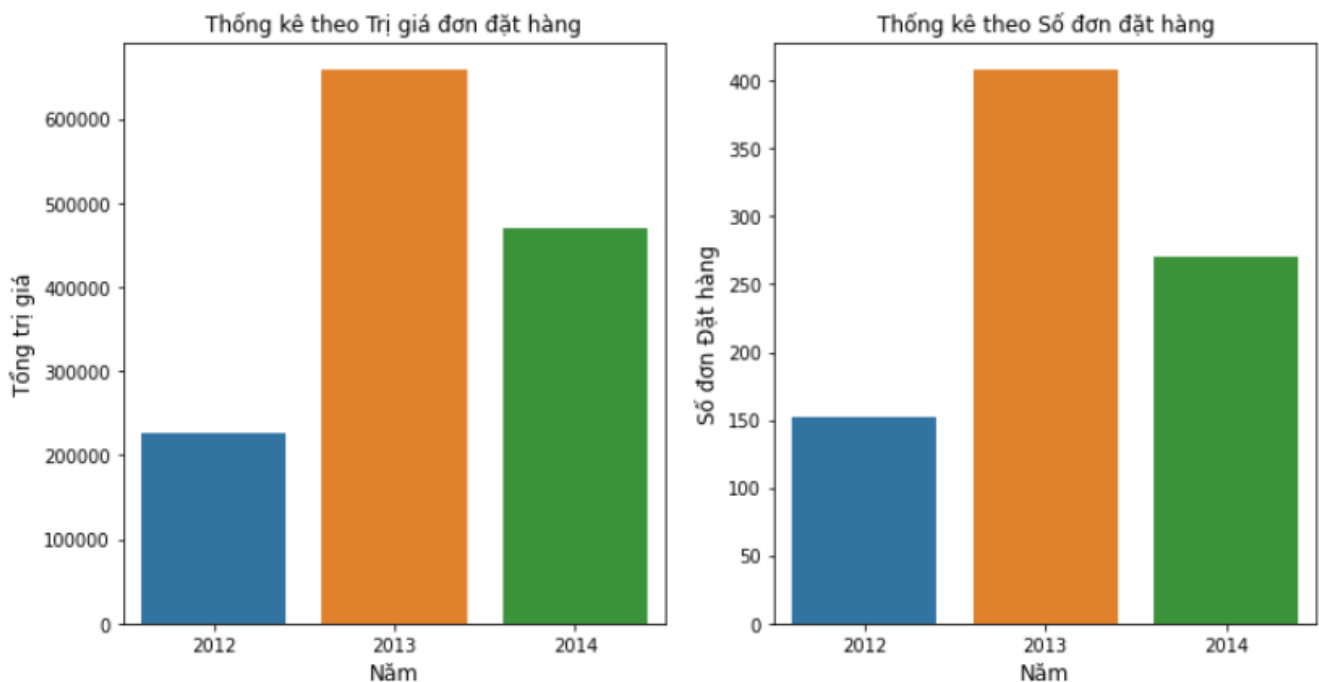
Câu 2: Trắc quan hóa dữ liệu (2 điểm)

1. Vẽ biểu đồ bar như sau, gồm 10 customer có nhiều đơn đặt hàng nhất (1 điểm)



2. Vẽ 2 biểu đồ bar tương ứng với các năm 2012, 2013 như sau : (1 điểm)

Thống kê theo năm



Phần 2 : Thu thập dữ liệu (3 điểm)

Thời hạn nộp bài: 12g00 – 19/03/2022









Yêu cầu: Sử dụng các thư viện đã học để thu thập dữ liệu theo yêu cầu sau

1. Truy xuất trang <https://babycuatoi.vn/xep-hinh-thong-minh>

Đồ chơi xếp hình lắp ráp

Sắp xếp ▾

Babycuatoi.vn chuyên cung cấp các loại đồ chơi xếp hình thông minh cho bé chất lượng cao, sản xuất theo tiêu chuẩn châu Âu. Các sản phẩm đồ chơi lắp ghép xếp hình giúp trẻ phát triển trí thông minh sáng tạo. Bố mẹ đến ngay showroom Babycuatoi để mua đồ chơi xếp hình lắp ráp cho bé tại Hà Nội, TpHCM,... với giá rẻ, chất lượng tốt nhất nha ...[Xem thêm](#)

 <p>KM : click để xem KM Bộ bàn ghế xếp hình LEGO đa năng kèm 105 chi tiết xếp hình cỡ lớn UG401-1</p> <p>580.000đ -20%</p> <p>★★★★★ (1 đánh giá)</p>	 <p>KM : click để xem KM Bộ bàn ghế xếp hình LEGO đa năng kèm 105 chi tiết xếp hình cỡ lớn UG401-2</p> <p>580.000đ -20%</p> <p>☆☆☆☆☆</p>	 <p>KM : click để xem KM Bộ bàn chơi LEGO đa năng 2 ghế kèm bộ xếp hình 100 chi tiết cỡ lớn 188T-8/188T-9</p> <p>1.100.000đ -20%</p> <p>★★★★★ (10 đánh giá)</p>	 <p>KM : click để xem KM Đồ chơi xếp hình ngôi nhà trái cây 336 chi tiết 5218</p> <p>199.000đ</p> <p>☆☆☆☆☆</p>
 <p>KM : click để xem KM Đồ chơi xếp hình cho bé gái siêu thị 490 chi tiết 6571</p> <p>399.000đ</p> <p>★★★★★ (1 đánh giá)</p>	 <p>KM : click để xem KM Đồ chơi xếp hình khủng long máy Mecha Gorgila 560 chi tiết 1510</p> <p>379.000đ</p> <p>★★★★★ (2 đánh giá)</p>	 <p>KM : click để xem KM Đồ chơi xếp hình xe ô tô Lamborghini cho bé 341 chi tiết 5103</p> <p>199.000đ</p> <p>☆☆☆☆☆</p>	 <p>KM : click để xem KM Đồ chơi xếp hình thế giới khủng long 352 chi tiết SY1505</p> <p>199.000đ -20%</p> <p>★★★★★ (1 đánh giá)</p>

- Thu thập các thông tin liên quan đến **tất cả** các sản phẩm trong trang đồ chơi xếp hình (**kể cả các sản phẩm ở các trang sau**): Tên sản phẩm, Giá bán, Tỷ lệ giảm giá, Thông tin mô tả sản phẩm, link Hình sản phẩm.
- Xử lý chuẩn hóa dữ liệu thu thập được và lưu vào tập tin **Products.csv** gồm các thông tin: Tên sản phẩm, Giá bán, Tỷ lệ giảm giá, Thông tin mô tả, Tên tập tin hình sản phẩm (trong đó giá bán và tỷ lệ giảm giá phải có kiểu số).
- Xử lý lưu trữ các hình ảnh thu thập được vào thư mục **Images**

--- Chúc các bạn làm bài tốt ---