

Chapter 6 - Exercise 1: Trực quan hóa dữ liệu Chipotle

Cho dữ liệu <https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv> (<https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv>).

Nhà hàng Chipotle cần phân tích dữ liệu bán được trong ngày diễn ra khuyến mãi để có thể điều chỉnh thực đơn và thực hiện các chương trình khuyến mãi phù hợp.

Dữ liệu được cung cấp trong file chipotle.tsv, hãy thực hiện các yêu cầu sau:

Yêu cầu

Câu 1:

- Đọc dữ liệu và gán vào biến chipo
- Hiển thị 10 dòng đầu của dữ liệu

Câu 2:

- Tạo dataframe df gồm 2 cột: tên món ăn và đếm các món ăn

Câu 3:

- a) Sắp xếp df theo tần suất giảm dần, và lấy 5 item đầu tiên
- b) Vẽ biểu đồ bar chart cho biết 5 món được gọi nhiều nhất, có title, xlabel, ylabel và xticks

Câu 4:

- a) Đổi kiểu dữ liệu của cột item_price sang kiểu số thực
- b) Tạo cột revenue, với $\text{revenue} = \text{quantity} * \text{item_price}$
- c) Nhóm các đơn hàng theo order_id, và tính tổng số lượng gọi và tổng giá trị của mỗi đơn hàng, in kết quả

Câu 5:

- Từ câu 4b, hãy vẽ scatterplot với x là revenue, và y là quantity, có title, xlabel, ylabel. Bạn có nhận xét gì qua biểu đồ này

Câu 6:

- Vẽ bar chart với x là các món ăn và y là tổng thành tiền. Vẽ cho 5 món có tổng thành tiền lớn nhất

```
In [1]: 1 import pandas as pd
        2 import matplotlib.pyplot as plt
```

```
In [2]: 1 # Câu 1: Đọc dữ liệu và gán vào biến chipo
        2
        3 # Hiển thị 10 dòng đầu của dữ liệu
        4
```

Nhấn vào đây để xem kết quả!

	order_id	quantity	item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	NaN	\$2.39
1	1	1	Izze	[Clementine]	\$3.39
2	1	1	Nantucket Nectar	[Apple]	\$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	\$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
5	3	1	Chicken Bowl	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	\$10.98
6	3	1	Side of Chips	NaN	\$1.69
7	4	1	Steak Burrito	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	\$11.75
8	4	1	Steak Soft Tacos	[Tomatillo Green Chili Salsa, [Pinto Beans, Ch...	\$9.25
9	5	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Pinto...	\$9.25

```
In [3]: 1 # Câu 2: tạo dataframe df gồm 2 cột: tên món ăn và đếm các món ăn
        2
```

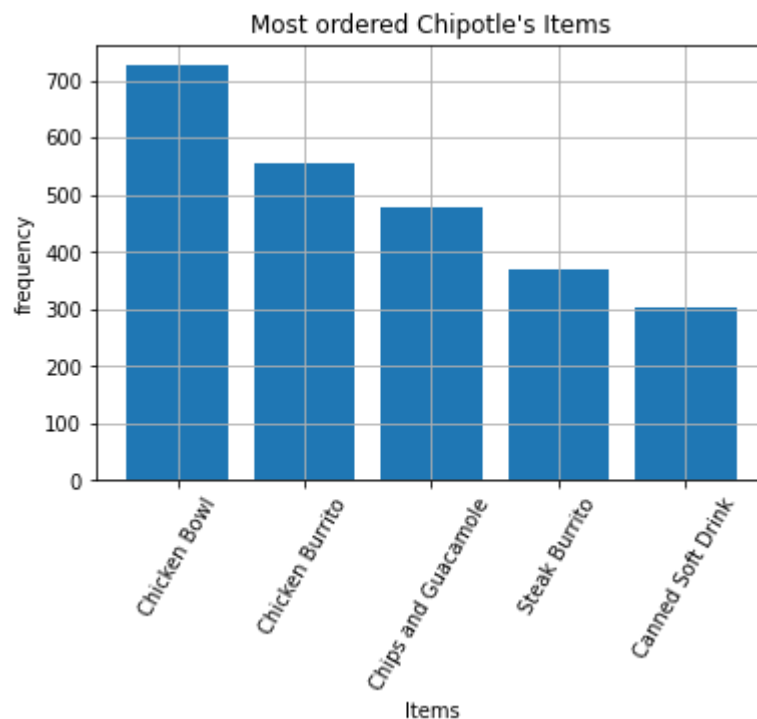
Nhấn vào đây để xem kết quả!

	item_name	cnt
0	6 Pack Soft Drink	54
1	Barbacoa Bowl	66
2	Barbacoa Burrito	91
3	Barbacoa Crispy Tacos	11
4	Barbacoa Salad Bowl	10

```
In [4]: 1 # Câu 3:
        2 # a) Sắp xếp df theo tần suất giảm dần, và lấy 5 item đầu tiên
        3
```

```
In [5]: 1 # b) Vẽ biểu đồ bar chart cho biết 5 món được gọi nhiều nhất
        2 # (có title, xlabel, ylabel và xticks)
        3
```

Nhấn vào đây để xem kết quả!



```
In [6]: 1 # Câu 4:
        2 # a) Đổi kiểu dữ liệu của cột item_price sang kiểu số thực
        3
```

```
In [7]: 1 # b) Tạo cột revenue, với revenue = quantity * item_price
        2
```

```
In [8]: 1 # c) Nhóm các đơn hàng theo order_id, và tính tổng số lượng gọi
        2 # và tổng giá trị của mỗi đơn hàng,
        3 # in kết quả
        4
```

Nhấn vào đây để xem kết quả!

	quantity	revenue
order_id		
1	4	11.56
2	2	33.96
3	2	12.67
4	2	21.00
5	2	13.70

```
In [9]: 1 # Câu 5: Từ câu 4b, hãy vẽ scatterplot với x là revenue,
        2 # và y là quantity, có title, xlabel, ylabel
        3 # Bạn có nhận xét gì qua biểu đồ này
        4
```

```
In [10]: 1 # Câu 6: vẽ bar chart với x là các món ăn,
          2 # và y là tổng thành tiền. Vẽ cho 5 món có tổng thành tiền lớn nhất
          3
```

```
In [ ]: 1
```

