

Chapter 7

Chapter 7 - Exercise 1: Tips

Cho dữ liệu tips có sẵn trong seaborn library. Hãy vẽ những biểu đồ theo yêu cầu và nhận xét biểu đồ sau khi vẽ:

1. Vẽ violinplot cho cột total_bill.
2. Vẽ swarmplot cho cột total_bill theo sex.
3. Vẽ boxplot cho cột total_bill.
4. Tạo FacetGrid của total_bill theo 'time' và chỉ định thứ tự của các cột bằng col_order.
5. Tạo catplot chứa point plot của giá trị 'total_bill' thay đổi theo size và tách dòng theo 'time'.
6. Tạo PairGrid với scatter plot thể hiện liên quan giữa "total_bill" và "tip" .
7. Tạo Pairplot với scatter plot thể hiện liên quan giữa "total_bill" và "tip", sử dụng palette color = 'day' .

Chapter 7 - Exercise 2: Titanic

Cho dữ liệu titanic có sẵn trong seaborn library. Hãy vẽ những biểu đồ theo yêu cầu sau và cho biết nhận xét sau biểu đồ vừa vẽ:

1. Vẽ stripplot thể hiện sự phân bố của fare theo class.
2. Vẽ violinplot thể hiện sự phân bố của fare theo class.
3. Vẽ countplot đếm số lượng alive/not alive theo từng class.
4. Vẽ pointplot thể hiện khả năng sống sót 'survived' theo class.
5. Vẽ catplot dạng bar thể hiện survived của từng sex, phân loại theo class.
6. Vẽ correlation matrix (heatmap) của titanic.

Chapter 7 - Exercise 3: Visualization with Seaborn - Diamond

Nghịch lý Simpson hay hiệu ứng Yule–Simpson, là một nghịch lý trong xác suất và thống kê, trong đó một xu hướng xuất hiện trong dữ liệu sẽ bị đảo ngược khi được phân tích dưới góc nhìn khác.

Cho file dữ liệu diamonds.csv. Hãy thực hiện các yêu cầu sau, để phát hiện nghịch lý Simpson khi phân tích giá kim cương bằng các công cụ trực quan hóa dữ liệu:

1. Đọc dữ liệu diamonds.csv, đưa vào biến diamonds
2. Vẽ biểu đồ bar so sánh giá của kim cương theo color, cut và clarity. Bạn nhận xét gì về biểu đồ vừa tạo?
3. Bây giờ hãy thử phân tích thuộc tính 'carat' theo 'color' và 'clarity' qua biểu đồ catplot - bar plot.
4. Vẽ biểu đồ bar so sánh 'carat' của kim cương theo color, cut và clarity.
5. Hãy chia carat ra làm 5 khoảng giá trị, tạo cột diamonds['carat_category'] chứa khoảng giá trị tương ứng (gợi ý: dùng pd.qcut).
6. Phân tích chi tiết hơn thuộc tính 'price' theo 'clarity', 'carat_category', 'color' qua biểu đồ catplot - point plot
7. Kết luận