

▼ [Thực hành] Phân tích sự biến đổi số lượng đơn hàng theo thời gian - Online Retail

Mục tiêu: Sử dụng thư viện Matplotlib để biểu đồ kết hợp trên bộ dữ liệu Online Retail. Bộ dữ liệu Online Retail mô tả tình hình kinh doanh của một doanh nghiệp bán hàng online mà bạn đã được làm quen từ các buổi trước.

Điều kiện:

- Dữ liệu đã được tiền xử lý.
- Thành thạo các câu lệnh đọc dữ liệu từ file, trích xuất dữ liệu mong muốn.
- Nắm được các khái niệm cơ bản về Data Visualization. Các bước thực hiện:

1. Chọn mục tiêu

- Biểu đồ đường thể hiện doanh thu theo từng tháng năm 2011.
- Biểu đồ cột thể hiện số lượng đơn hàng trong các tháng của năm 2011.

2. Khai báo thư viện cần dùng:

```
import pandas as pd
import matplotlib.pyplot as plt
import datetime
```

3. Load data

```
df = pd.read_csv("OnlineRetail.csv", encoding = "ISO-8859-1")
```

4. Lấy dữ liệu cần thiết

```

df['InvoiceDate'] = pd.to_datetime(df['InvoiceDate']) #chuyển InvoiceDate thành datetime object
d1 = df[['InvoiceNo', 'InvoiceDate', 'Quantity', 'UnitPrice']] #lấy các cột dữ liệu cần thiết
d1['Revenue'] = d1['Quantity'] * d1['UnitPrice'] #tính doanh thu trên mỗi row
d1 = d1.set_index(['InvoiceDate']) #chuyển mỗi giá trị InvoiceDate thành index của mỗi để search theo index
d2 = d1['2011'] #lọc những hàng mà index có chứa '2011'
d2 = d2.reset_index()
d3 = d2.groupby(by=d2['InvoiceDate'].dt.month).sum() #tính tổng doanh thu theo tháng

d4 = d1.drop_duplicates(subset = 'InvoiceNo', keep = 'first') # xóa bỏ các dòng trùng lặp của cùng một đơn hàng
d4 = d4['2011']
d4 = d4.reset_index()
d5 = d4.groupby(by=d4['InvoiceDate'].dt.month).count() #đếm tổng số đơn hàng trong tháng

```

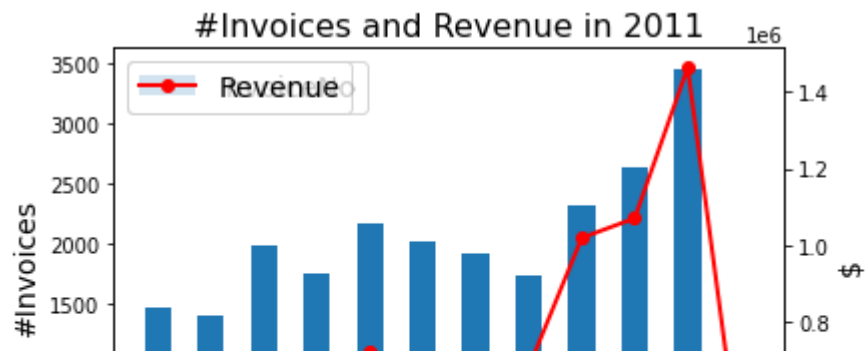
5. Vẽ biểu đồ kết hợp

```

x = d5.index.get_level_values(0)
plt.bar(x, d5['InvoiceNo'], width = 0.5, label = 'InvoiceNo')
axes1 = plt.gca()
axes2 = axes1.twinx()
axes2.plot(x, d3['Revenue'], label = 'Revenue', linewidth = 2, c = 'r', marker = 'o')
axes1.set_xlabel('Month', fontsize = 14)
axes1.set_ylabel('#Invoices', fontsize = 14)
axes2.set_ylabel('$', fontsize = 14)

axes1.legend(fontsize = 14)
axes2.legend(fontsize = 14)
plt.title('#Invoices and Revenue in 2011', fontsize = 16)
plt.show()

```



Nhận xét:

- Tháng 4 có tổng số đơn hàng nhiều hơn tháng 2. Tuy nhiên tổng doanh thu thì chỉ tương đương.
- Tháng 9, 10, 11 tổng số đơn hàng nhiều và tổng doanh thu cũng tăng nhanh chóng.

▼ Tổng kết

Qua bài thực hành này, chúng ta đã biết sử dụng thư viện Matplotlib để vẽ biểu đồ đường và cột kết hợp biểu diễn thông tin trên bộ dữ liệu Online Retail.