## **[Bài tập] Phân tích bộ dữ liệu Shopee.ph**

### **Mục tiêu**

Luyện tập các bước tiến hành trực quan hóa dữ liệu

Luyện tập sử dụng thư viện Matplotlib để vẽ một số biểu đồ: biểu đồ đường, biểu đồ cột, biểu đồ tròn, biểu đồ Scatter.

### **Mô tả**

* Sự phổ biến ngày càng tăng của thời trang lấy cảm hứng từ Hàn Quốc đã tiếp quản bán lẻ điện tử ở Philippines. Các doanh nhân đang bắt kịp xu hướng ngày càng tăng này bằng cách bán các sản phẩm lấy cảm hứng từ thời trang Hàn Quốc của họ trên các nền tảng trực tuyến khác nhau: Thông tin thu thập được từ shop bao gồm:

|  |  |
| --- | --- |
| date\_collected | ngày lấy dữ liệu |
| shopid | shopid |
| name | tên shop |
| join\_month | Tháng gia nhập |
| join\_day | Ngày gia nhập |
| join\_year | Năm gia nhập |
| item\_count | số lượng sản phẩm |
| follower\_count | Số lượng theo dõi |
| response\_time | Thời gian phản hồi |
| response\_rate | Tỷ lệ phản hồi |
| shop\_location | Địa chỉ cửa hàng |
| rating\_bad | số lượt khách hàng đánh giá xấu |
| rating\_good | số lượt khách hàng đánh giá tốt |
| rating\_normal | số lượt khách hàng đánh giá trung bình |
| rating\_star | Điểm đánh giá trung bình |
| is\_shopee\_verified | cửa hàng đã được xác thực hay chưa |
| is\_official\_shop | Có cửa hàng chính thức hay không |

Yêu cầu

* + Vẽ biểu đồ so sánh số lượng shop gia nhập theo các năm.
  + Vẽ biểu đồ thể hiện mối quan hệ giữa tỉ lệ phản hồi với số lượt khách hàng đánh giá tốt.
  + Vẽ biểu đồ thể hiện mối quan hệ giữa thời gian phản hồi (đơn vị giây) với số lượt khách hàng đánh giá xấu.
  + Vẽ biểu đồ thể hiện xu hướng của số lượng shop gia nhập theo thời gian.
  + Vẽ biểu đồ thể hiện phân bố của điểm đánh giá trung bình.

### **Hướng dẫn**

* Sử dụng pandas để đọc dữ liệu
* Sử dụng kiến thức bản thân để chọn biểu đồ thích hợp.
* Sử dụng các kiến thức về DataFrame để lọc dữ liệu cần thiết.
* Sử dụng hàm Matplotlib để vẽ và tinh chỉnh biểu đồ.

Demo

import pandas as pd  
import matplotlib.pyplot as plt  
  
d = pd.read\_csv("shopeep\_koreantop\_clothing\_shop\_data.csv")  
x = [2015, 2016, 2017, 2018, 2019, 2020, 2021]  
y = [0]\*len(x)  
for idx in range(len(x)):  
 y[idx] = d[(d['join\_year'] == x[idx])].shape[0];  
  
plt.bar(x, y, width=0.5)  
plt.title('Number of New Shops Over Years', fontsize = 16)  
plt.xlabel('Year', fontsize = 14)  
plt.ylabel('Number of Shops', fontsize = 14)  
plt.show()