## **[Bài tập] Phân tích bộ dữ liệu HousePrice \_ Đống Đa**

### **Mục tiêu**

Luyện tập các bước tiến hành trực quan hóa dữ liệu

Luyện tập sử dụng thư viện Matplotlib để vẽ biểu đồ kết hợp.

### **Mô tả**

* Một công ty mối giới bất động sản, muốn cung cấp dịch vụ tư vấn giao dịch cho khách hàng tại khu vực Đống Đa Hà Nội với loại hình nhà mặt đất. Công ty đã tiến hành thu thập thông tin của một số bất động sản ở các vị trí khác nhau: Thông tin nhân viên của công ty thu thập được gồm:
  + Area: diện tích
  + Price: giá căn nhà (tính theo đơn vị tỷ đồng)
  + postDate: Ngày chủ nhà giao bán trên mạng
  + land\_certificate: Loại giấy tờ
  + house\_direction: hướng nhà
  + balcony\_direction: Hướng ban công
  + toilet: số lượng toilet
  + bedroom: Số lượng phòng ngủ
  + Floor: Số tầng
  + Property\_Type: loại nhà đất (mặt phố hoặc trong ngõ)
* Yêu cầu vẽ biểu đồ kết hợp để:
  + Phân tích mối liên hệ giữa diện tích với giá nhà. Đồng thời, giữa số phòng ngủ với giá nhà và giữa số toilet với giá nhà.
  + So sánh giá nhà trung bình trên 1 m2 giữa các hình thức nhà (type\_of\_land). Đồn thời thể hiện tỉ lệ % bài đăng (bản ghi) giữa các hình thức nhà (type\_of\_land).
  + Vẽ biểu đồ thể hiện sự thay đổi giá nhà trung bình trên 1m2 theo số lượng phòng ngủ, theo số phòng toilet hoặc theo diện tích.

### **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

from scipy import stats  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
d = pd.read\_excel("house\_price\_đống đa.xlsx", sheet\_name='Sheet1')  
d1 = d[['area','price']]  
d2 = d1.dropna(subset=['area', 'price'])  
z = np.abs(stats.zscore(d2))  
d3 = d2[(z < 3).all(axis=1)]  
  
d4 = d[['toilet','price']]  
d5 = d4.dropna(subset=['toilet', 'price'])  
z = np.abs(stats.zscore(d5))  
d6 = d5[(z < 3).all(axis=1)]  
  
d7 = d[['bedroom','price']]  
d8 = d7.dropna(subset=['bedroom', 'price'])  
z = np.abs(stats.zscore(d8))  
d9 = d8[(z < 3).all(axis=1)]  
  
fig, ax = plt.subplots(1, 3)  
ax[0].scatter(d3['area'], d3['price'])  
ax[0].set\_xlabel('Area')  
ax[1].scatter(d6['toilet'], d6['price'])  
ax[1].set\_xlabel('Toilet')  
ax[2].scatter(d9['bedroom'], d9['price'])  
ax[2].set\_xlabel('BedRoom')  
ax[2].set\_ylabel('Price')  
plt.show()