

# Visualize\_RealEstate\_Đề\_bài

August 31, 2021

## 1 Visualization

Trong bài tập này, bạn sẽ được rèn luyện kỹ năng trực quan hóa. Trực quan hóa ngoài việc hỗ trợ trong khâu tiền xử lý dữ liệu, còn giúp bạn phân tích và rút ra được các thông tin có ý nghĩa từ dữ liệu

**Tập dữ liệu** được sử dụng được lấy từ trang batdongsan.com.vn

```
[ ]: #Import các thư viện cần thiết

import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
import warnings; warnings.filterwarnings(action='once')
import random
import math
```

```
[ ]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

```
[ ]: cd /content/drive/MyDrive/Preprocessing_practice/2.Assignment/Batdongsan.com.vn
```

/content/drive/MyDrive/Preprocessing\_practice/2.Assignment/Batdongsan.com.vn

```
[ ]: #Load dữ liệu từ file
newdf = pd.read_csv("batdongsan.com.vn.csv", dtype = {"month": "string" , "id" :
    ↪ "string"})
print("DONE")
```

DONE

```
[ ]: newdf.columns
```

```
[ ]: Index(['id', 'month', 'project', 'investor', 'square', 'bedrooms', 'bathrooms',
          'direction', 'balcony', 'district', 'ward', 'price'],
          dtype='object')
```

```
[ ]: # Xem 5 hàng dữ liệu đầu tiên, các thông tin ban đầu: df.info(), df.describe(),
      ↪ ...
      # Code ở đây:
```

Đưa ra các câu lệnh để thực hiện các yêu cầu sau:

1. Minh họa boxplot của các trường số như price, bedrooms, square, ...

- Gợi ý: sns.boxplot(data\_field)

```
[ ]: #Code for 'price' field
```

```
[ ]: #Code for 'square' field
```

```
[ ]: #Code for 'bedrooms' field
```

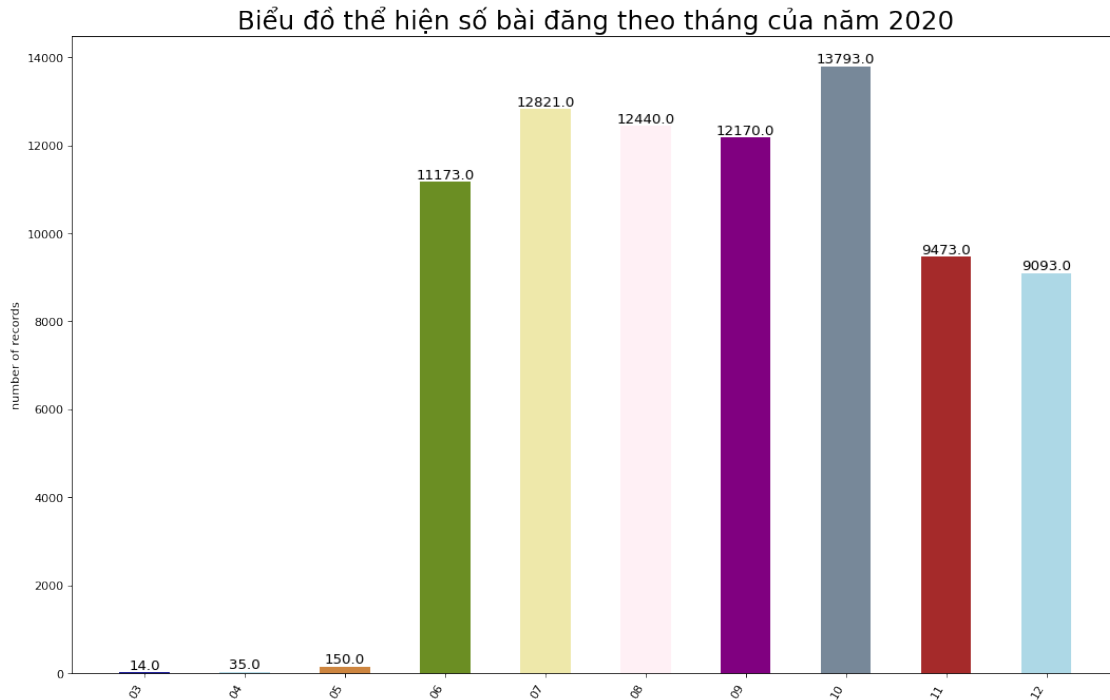
```
[ ]: #Code for 'bathrooms' field
```

2. Cho đoạn code minh họa Biểu đồ thể hiện số bài đăng theo tháng của năm 2020. Hãy thực hiện các yêu cầu cho các câu hỏi dưới đây:

```
[ ]: df1 = newdf.groupby('month').size().reset_index(name='number of records')
n = df1['month'].unique().__len__()+1
all_colors = list(plt.cm.colors.cnames.keys())
random.seed(100)
c = random.choices(all_colors, k=n)

plt.figure(figsize=(16,10), dpi= 80)
plt.bar(df1['month'], df1['number of records'], color=c, width=.5)
for i, val in enumerate(df1['number of records'].values):
    plt.text(i, val, float(val), horizontalalignment='center',
    ↪verticalalignment='bottom', fontdict={'fontweight':500, 'size':12})

plt.gca().set_xticklabels(df1['month'], rotation=60, horizontalalignment=
    ↪'right')
plt.title("Biểu đồ thể hiện số bài đăng theo tháng của năm 2020", fontsize=22)
plt.ylabel('number of records')
plt.show()
```



- Biểu đồ thể hiện giá nhà đất trung bình theo tháng (VND/m<sup>2</sup>)

```
[ ]: #Code ở đây
```

- Biểu đồ mật độ giá nhà của tháng 4 và tháng 12 (VND/m<sup>2</sup>)

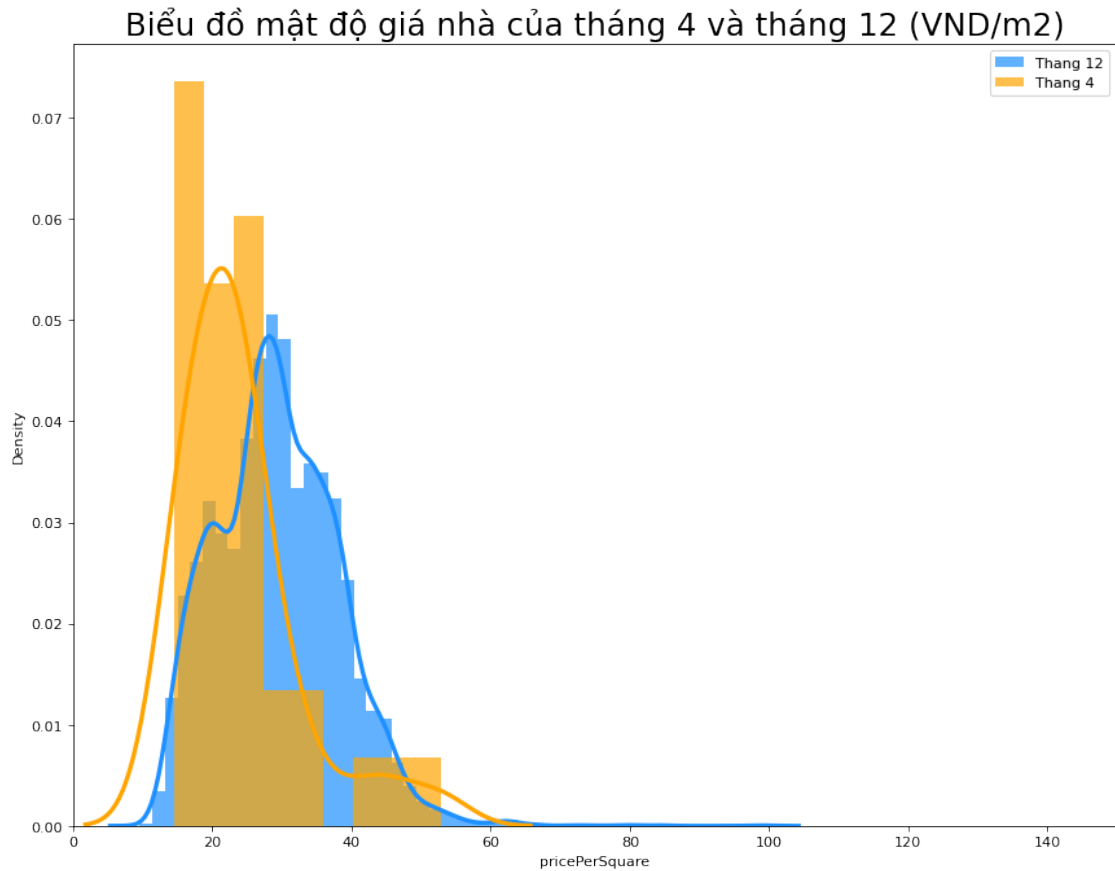
```
[ ]: #Gọi ý: sns.displot()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
```

```
warnings.warn(msg, FutureWarning)
```



- Số lượng bất động sản phân theo số lượng bedrooms

[ ]: `#Code ở đây`

- Giá bất động sản trung bình theo số lượng bedrooms

[ ]: `# Code ở đây`

- Giá nhà đất trung bình theo quận (VND/m<sup>2</sup>)

[ ]: `# Code ở đây`

- Giá nhà đất trung bình theo các phường quận Hoàn Kiếm (VND/m<sup>2</sup>)

[ ]: `# Code ở đây`

- Giá nhà đất trung bình theo nhà đầu tư (VND/m<sup>2</sup>)

[ ]: `# Code ở đây`

- Giá nhà đất trung bình theo dự án (VND/m<sup>2</sup>)

```
[ ]: # Code ở đây
```

- Giá bất động sản phân bố theo top 10 nhà đầu tư triệu /m2

```
[ ]: # Code ở đây
```

- Giá bất động sản phân bố theo dự án

```
[ ]: # Code ở đây
```

- Giá nhà đất trung bình theo hướng nhà (VND/m2)

```
[ ]: # Code ở đây
```

```
[ ]: # Code ở đây
```

### 3. Tương quan giữa các trường thuộc tính

- [https://vi.wikipedia.org/wiki/H%E1%BB%87\\_s%E1%BB%91\\_t%C6%B0%C6%A1ng\\_quan](https://vi.wikipedia.org/wiki/H%E1%BB%87_s%E1%BB%91_t%C6%B0%C6%A1ng_quan)
- biểu đồ thể hiện hệ số tương quan giữa các cặp biến số thực bằng biểu đồ heatmap

```
[ ]: # Gợi ý: sns.heatmap(cmap='RdYlGn', ...)
```

- biểu đồ thể hiện mối quan hệ giữa các biến số thực với biểu đồ phân tán

```
[ ]: #Gợi ý: sns.pairplot(kind='scatter', ...)
```

### 4. Mật độ giá nhà trung bình theo quận

```
[ ]: #Gợi ý: sns.displot()
```

End :))