Scores

Introduction:

This time you will create the data.

Exercise based on [Chris Albon](#) work, the credits belong to him.

Step 1. Import the necessary libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

Step 2. Create the DataFrame that should look like the one below.

```
import pandas as pd

data = {
    'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],
    'last_name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze'],
    'age': [42, 52, 36, 24, 73],
    'female': [0, 1, 1, 0, 1],
    'preTestScore': [4, 24, 31, 2, 3],
    'postTestScore': [25, 94, 57, 62, 70]
}

df = pd.DataFrame(data)

print(df)
df.head()
```

```
  first_name last_name  age  female  preTestScore  postTestScore
0      Jason    Miller   42       0             4             25
1      Molly  Jacobson   52       1            24             94
2       Tina       Ali   36       1            31             57
3       Jake    Milner   24       0             2             62
4        Amy     Cooze   73       1             3             70
```

	first_name	last_name	age	female	preTestScore	postTestScore
0	Jason	Miller	42	0	4	25
1	Molly	Jacobson	52	1	24	94
2	Tina	Ali	36	1	31	57
3	Jake	Milner	24	0	2	62
4	Amy	Cooze	73	1	3	70

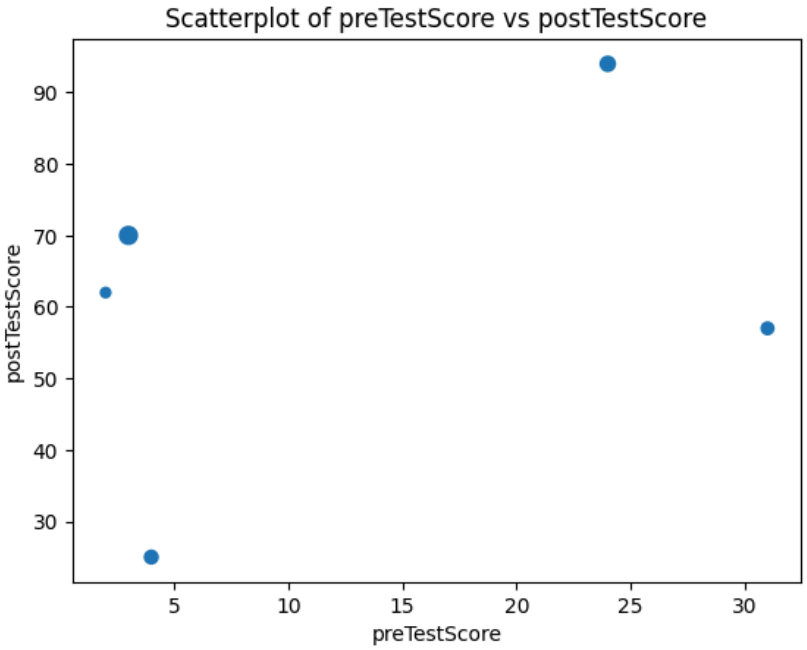
Các bước tiếp theo: Tạo mã bằng df | Xem các đồ thị được đề xuất | New interactive sheet

Step 3. Create a Scatterplot of preTestScore and postTestScore, with the size of each point determined by age

Hint: Don't forget to place the labels

```
plt.scatter(df['preTestScore'], df['postTestScore'], s=df['age'])
plt.xlabel("preTestScore")
plt.ylabel("postTestScore")
plt.title("Scatterplot of preTestScore vs postTestScore")

plt.show()
```

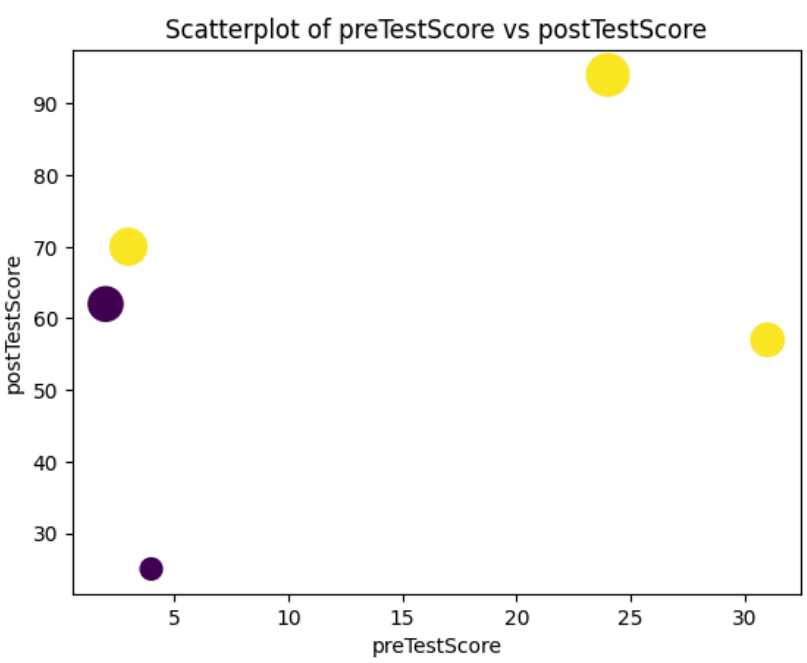


Step 4. Create a Scatterplot of preTestScore and postTestScore.

This time the size should be 4.5 times the postTestScore and the color determined by sex

```
plt.scatter(df['preTestScore'], df['postTestScore'], s=df['postTestScore'] * 4.5, c=df['female'])
plt.xlabel("preTestScore")
plt.ylabel("postTestScore")
plt.title("Scatterplot of preTestScore vs postTestScore")

plt.show()
```



BONUS: Create your own question and answer it.

Tôi có thể tính điểm preTestScore trung bình cho những người tham gia là nữ trong DataFrame như thế nào?

```
average_preTestScore_female = df[df['female'] == 1]['preTestScore'].mean()
print(average_preTestScore_female)
```

```
19.333333333333332
```