

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

```
data = pd.read_csv('Social_Network_Ads.csv')
data.head()
```

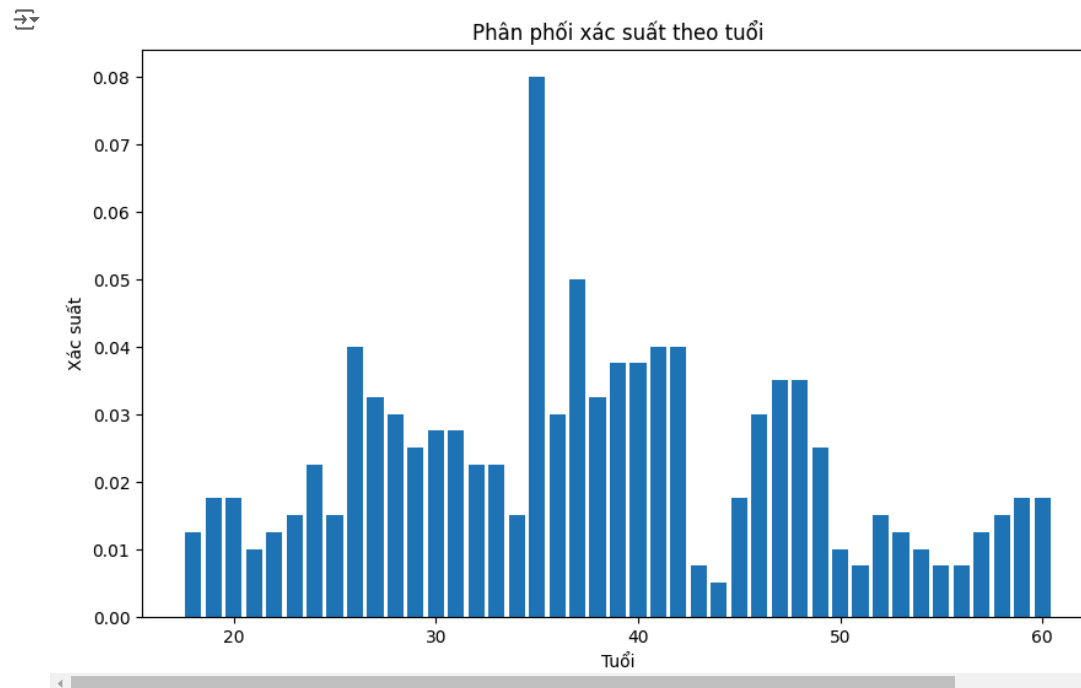
	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

Next steps: [Generate code with data](#) [View recommended plots](#) [New interactive sheet](#)

## ▼ Câu 1:

```
# Phân phối xác suất theo tuổi
pmf_age = data['Age'].value_counts(normalize=True).sort_index()

plt.figure(figsize=(10, 6))
plt.bar(pmf_age.index, pmf_age.values)
plt.xlabel('Tuổi')
plt.ylabel('Xác suất')
plt.title('Phân phối xác suất theo tuổi')
plt.show()
```



## ▼ Câu 2:

```
# Tính giá trị kỳ vọng của Age
expected_age = np.sum(pmf_age.index * pmf_age.values)

# Tính giá trị kỳ vọng của Age^2
expected_age_squared = np.sum((pmf_age.index ** 2) * pmf_age.values)

# Tính giá trị kỳ vọng của Age^3
expected_age_cubed = np.sum((pmf_age.index ** 3) * pmf_age.values)

print("Giá trị kỳ vọng của Age:", expected_age)
print("Giá trị kỳ vọng của Age^2:", expected_age_squared)
```

```
print("Giá trị kỳ vọng của Age^3:", expected_age_cubed)
```

```
↗ Giá trị kỳ vọng của Age: 37.654999999999994  
Giá trị kỳ vọng của Age^2: 1527.515  
Giá trị kỳ vọng của Age^3: 66038.255
```

### ▼ Bài 3:

```
# Tính phương sai của Age  
variance_age_1 = expected_age_squared - (expected_age ** 2)  
  
print("Phương sai của Age:", variance_age_1)  
  
variance_age_2 = data['Age'].var()  
  
print("Phương sai của Age:", variance_age_2)
```

```
↗ Phương sai của Age: 109.61597500000062  
Phương sai của Age: 109.89070175438596
```

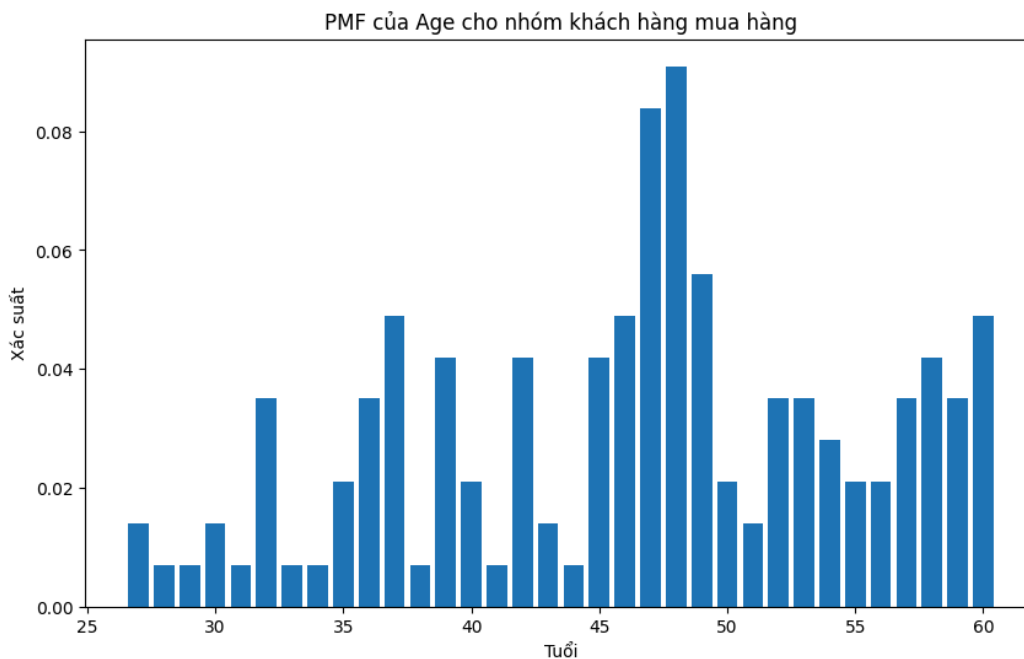
### ▼ Bài 4:

```
# Phân nhóm khách hàng mua hàng và không mua hàng  
purchased_data = data[data['Purchased'] == 1]  
not_purchased_data = data[data['Purchased'] == 0]
```

```
# Tính PMF của Age cho nhóm khách hàng mua hàng  
pmf_age_purchased = purchased_data['Age'].value_counts(normalize=True).sort_index()
```

```
plt.figure(figsize=(10, 6))  
plt.bar(pmf_age_purchased.index, pmf_age_purchased.values)  
plt.xlabel('Tuổi')  
plt.ylabel('Xác suất')  
plt.title('PMF của Age cho nhóm khách hàng mua hàng')  
plt.show()
```

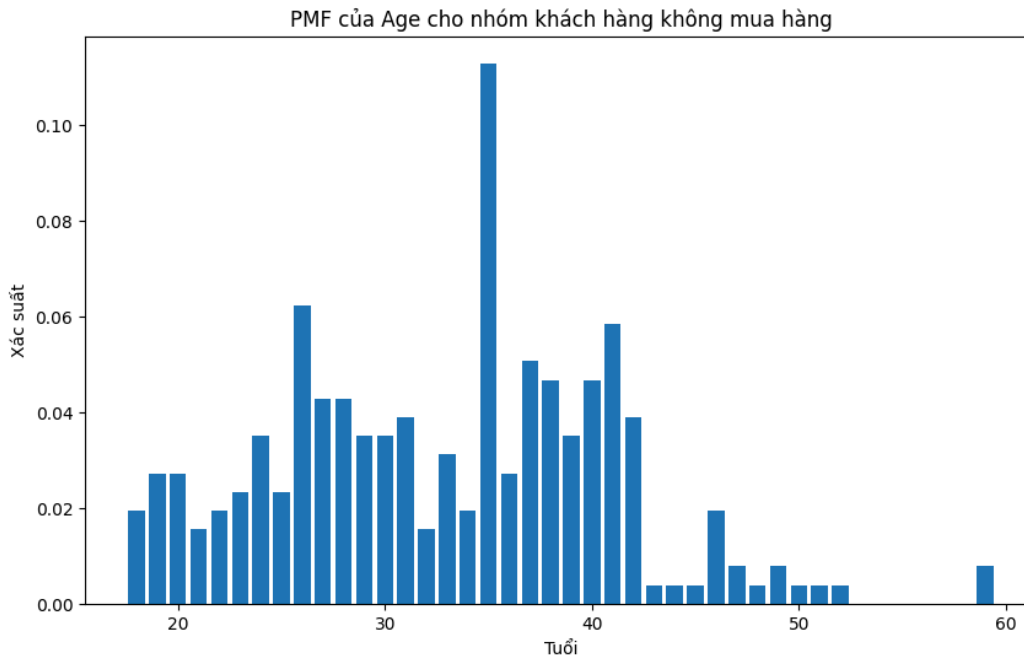
```
↗
```



```
# Tính giá trị kỳ vọng, phương sai cho nhóm mua hàng
expected_age_purchased = np.sum(pmf_age_purchased.index * pmf_age_purchased.values)
expected_age_squared_purchased = np.sum((pmf_age_purchased.index ** 2) * pmf_age_purchased.values)
expected_age_cubed_purchased = np.sum((pmf_age_purchased.index ** 3) * pmf_age_purchased.values)
variance_age_purchased = expected_age_squared_purchased - (expected_age_purchased ** 2)

# Tính PMF của Age cho nhóm khách hàng không mua hàng
pmf_age_not_purchased = not_purchased_data['Age'].value_counts(normalize=True).sort_index()

plt.figure(figsize=(10, 6))
plt.bar(pmf_age_not_purchased.index, pmf_age_not_purchased.values)
plt.xlabel('Tuổi')
plt.ylabel('Xác suất')
plt.title('PMF của Age cho nhóm khách hàng không mua hàng')
plt.show()
```



```
# Tính giá trị kỳ vọng, phương sai cho nhóm không mua hàng
expected_age_not_purchased = np.sum(pmf_age_not_purchased.index * pmf_age_not_purchased.values)
expected_age_squared_not_purchased = np.sum((pmf_age_not_purchased.index ** 2) * pmf_age_not_purchased.values)
expected_age_cubed_not_purchased = np.sum((pmf_age_not_purchased.index ** 3) * pmf_age_not_purchased.values)
variance_age_not_purchased = expected_age_squared_not_purchased - (expected_age_not_purchased ** 2)

print("Giá trị kỳ vọng của Age cho nhóm không mua hàng:", expected_age_not_purchased)
print("Giá trị kỳ vọng của Age^2 cho nhóm không mua hàng:", expected_age_squared_not_purchased)
print("Giá trị kỳ vọng của Age^3 cho nhóm không mua hàng:", expected_age_cubed_not_purchased)
print("Phương sai của Age cho nhóm không mua hàng:", variance_age_not_purchased)
```



```
Giá trị kỳ vọng của Age cho nhóm không mua hàng: 32.793774319066145
Giá trị kỳ vọng của Age^2 cho nhóm không mua hàng: 1138.95719844358
Giá trị kỳ vọng của Age^3 cho nhóm không mua hàng: 41625.914396887165
Phương sai của Age cho nhóm không mua hàng: 63.525564353737764
```