

ĐẠI HỌC NÔNG LÂM TP. HỒ CHÍ MINH  
KHOA TÀI NGUYÊN MÔI TRƯỜNG - NGÀNH HỆ THỐNG THÔNG TIN  
Học phần: **CƠ SỞ ỨNG DỤNG TRÍ TUỆ NHÂN TẠO**  
**2023 - 2024**

**REPORT FORM - PRACTICE 2**

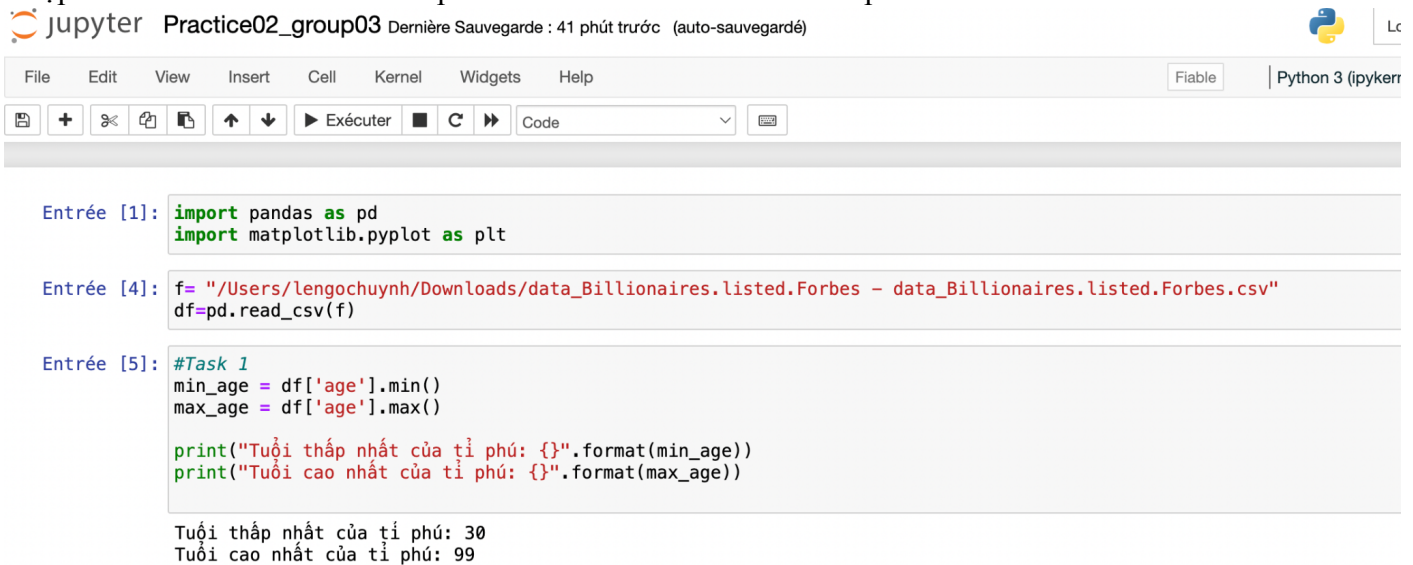
**GROUP 03**

**Ngày: 08/12/2023**

STT	Họ và tên	MSSV
1	Nguyễn Thị Lan Phương	21166153
2	Lê Ngọc Huỳnh	21166128
3	Nguyễn Minh Nhi	21166041
4	Vũ Hồng Quân	21166155
5	Nguyễn Thị Diễm Quỳnh	21166101

Task I:

– Lập trình code để tìm ra tuổi thấp nhất và tuổi cao nhất của các tỉ phú



The screenshot shows a Jupyter Notebook with the following content:

```
Entrée [1]: import pandas as pd
import matplotlib.pyplot as plt

Entrée [4]: f= "/Users/lengochuynh/Downloads/data_Billionaires.listed.Forbes - data_Billionaires.listed.Forbes.csv"
df=pd.read_csv(f)

Entrée [5]: #Task 1
min_age = df['age'].min()
max_age = df['age'].max()

print("Tuổi thấp nhất của tỉ phú: {}".format(min_age))
print("Tuổi cao nhất của tỉ phú: {}".format(max_age))

Tuổi thấp nhất của tỉ phú: 30
Tuổi cao nhất của tỉ phú: 99
```

- Lọc và liệt kê top 10 lĩnh vực chuyên ngành (industry) mà có nhiều nhà tỉ phú nhất

```
total_wealth_by_industry = df.groupby('Industry')['Net_Worth_($B)'].sum()
top_10_industries = total_wealth_by_industry.sort_values(ascending=False).head(10)
print("Top 10 lĩnh vực chuyên ngành có nhiều nhà tỉ phú nhất:", top_10_industries)
```

Top 10 lĩnh vực chuyên ngành có nhiều nhà tỉ phú nhất: Industry

Technology	1075
Fashion & Retail	1001
Finance & Investments	552
Diversified	435
Food & Beverage	396
Automotive	365
Metals & Mining	273
Manufacturing	223
Media & Entertainment	184
Energy	177

Name: Net\_Worth\_(\$B), dtype: int64

- Lọc và liệt kê top 10 quốc gia (Citizenship) mà có nhiều nhà tỉ phú nhất

```
total_wealth_by_citizenship = df.groupby('Country_Working_In')['Net_Worth_($B)'].sum()

top_10_countries = total_wealth_by_citizenship.sort_values(ascending=False).head(10)

print("Top 10 quốc gia có nhiều nhà tỉ phú nhất:", top_10_countries)
```

Top 10 quốc gia có nhiều nhà tỉ phú nhất: Country\_Working\_In

United States	2316
China	544
France	460
India	276
Germany	257
Russia	235
Hong Kong	180
Mexico	129
Canada	91
Australia	90

Name: Net\_Worth\_(\$B), dtype: int64

Task II:

Thống kê số lượng tỉ phú theo bằng cấp học vị (Bachelor – Master – Doctorate)

### #Task2

```
phd_count = df['Doctorate'].shape[0]
master_count = df['Master'].shape[0]
bachelor_count = df['Bachelor'].shape[0]

print("Số lượng tĩ phú có Doctorate: {}".format(phd_count))
print("Số lượng tĩ phú có Master: {}".format(master_count))
print("Số lượng tĩ phú có Bachelor: {}".format(bachelor_count))
```

Số lượng tĩ phú có Doctorate: 221  
Số lượng tĩ phú có Master: 221  
Số lượng tĩ phú có Bachelor: 221

- Tìm kiếm quốc gia nào mà có tổng Net Worth của các nhà tĩ phú (thuộc quốc tịch) là lớn nhất

```
total_net_worth_by_citizenship = df.groupby('Country_Working_In')['Net_Worth_($B)'].sum()

max_net_worth_country = total_net_worth_by_citizenship.idxmax()
max_net_worth_value = total_net_worth_by_citizenship.max()

print("Quốc gia có tổng Net Worth lớn nhất là {} với giá trị {}".format(max_net_worth_country, max_net_worth_value))

Quốc gia có tổng Net Worth lớn nhất là United States với giá trị 2316
```

- Tìm kiếm quốc gia nào mà có nhiều nhà tĩ phú trẻ nhất (< 50 tuổi)

```
young_billionaires = df[df['Age'] < 50]

young_billionaires_by_citizenship = young_billionaires['Country_Working_In'].value_counts()

max_young_billionaires_country = young_billionaires_by_citizenship.idxmax()
max_young_billionaires_count = young_billionaires_by_citizenship.max()

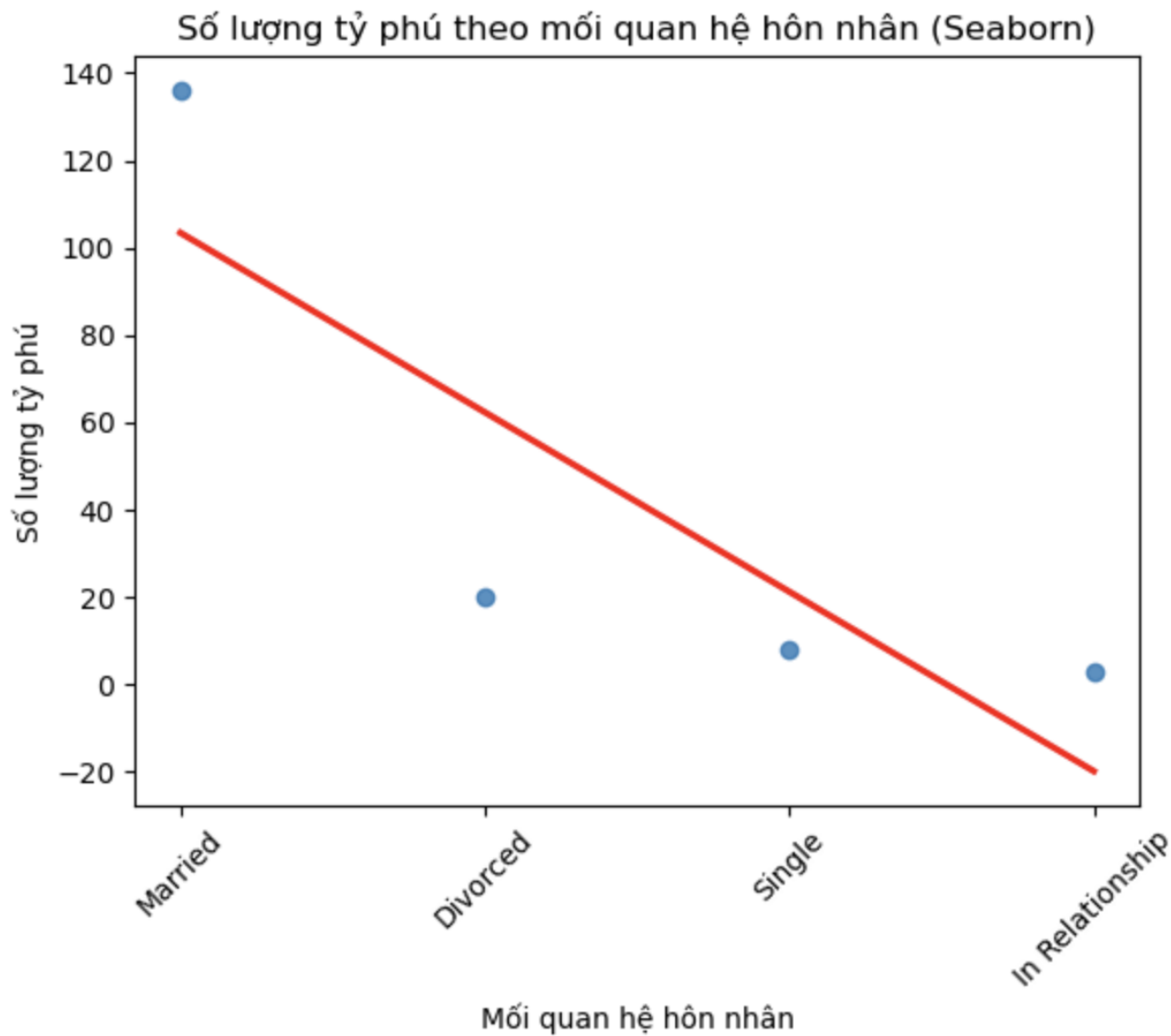
print("Quốc gia có nhiều nhà tĩ phú trẻ nhất là {} với số lượng {}".format(max_young_billionaires_country, max_young_billionaires_count))

Quốc gia có nhiều nhà tĩ phú trẻ nhất là United States với số lượng 5
```

### Task III:

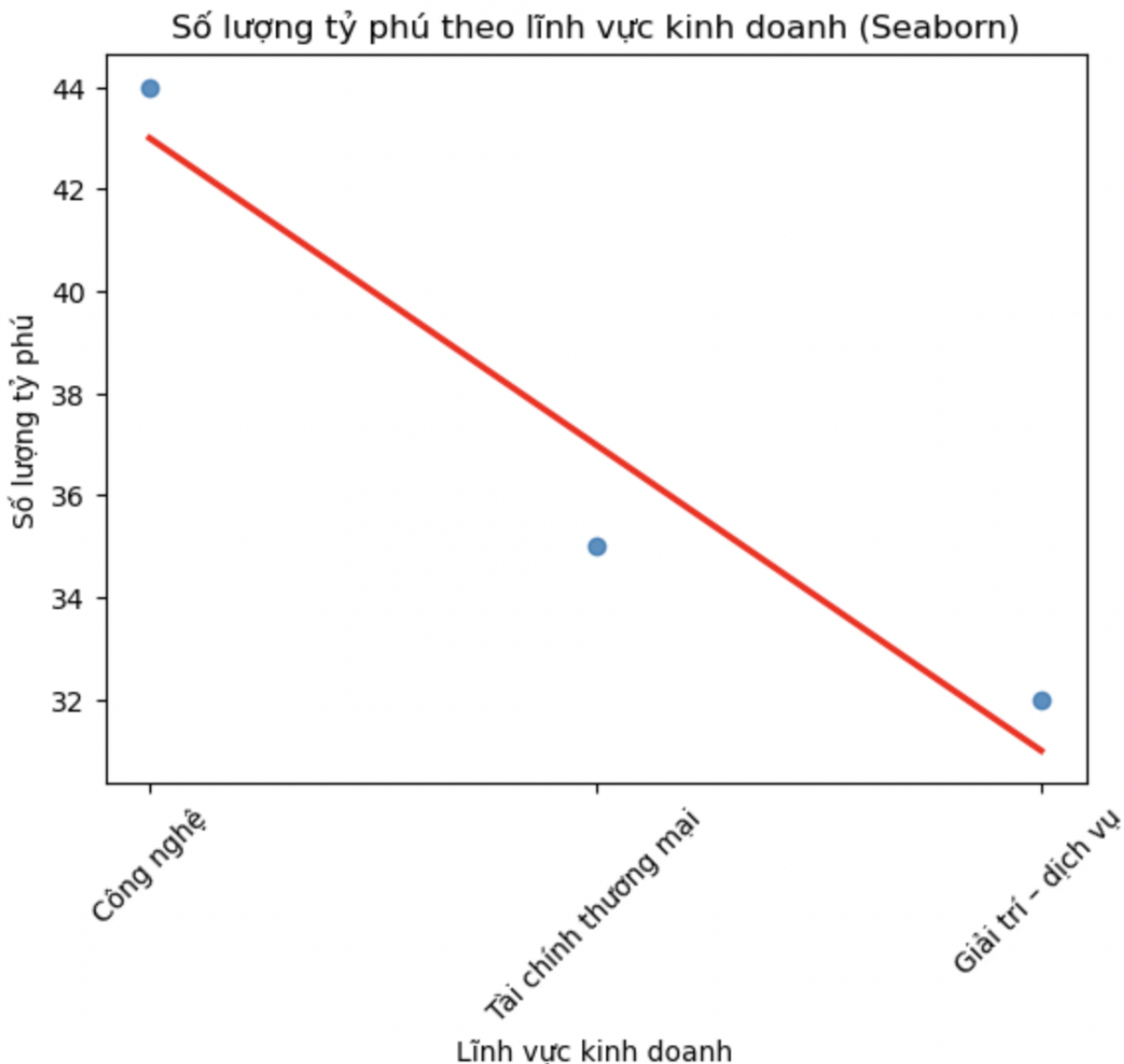
- Thống kê số lượng tĩ phú theo mỗi quan hệ hôn nhân (single, married/in relationship, divorced)

```
#Task 3.1
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
f = "/Users/lengochuynh/Downloads/data_Billionaires.listed.Forbes - data_Billionaires.listed.Forbes.csv"
df = pd.read_csv(f)
# Lọc dữ liệu theo các tình trạng hôn nhân đã chọn
selected_marital_statuses = ['Single', 'Married', 'In Relationship', 'Divorced']
filtered_df = df[df['Marital Status'].isin(selected_marital_statuses)]
# Thống kê số lượng tĩ phú theo mỗi quan hệ hôn nhân đã chọn
marital_status_counts = filtered_df['Marital Status'].value_counts()
# Vẽ đồ thị minh họa theo mô hình regression bằng Seaborn
sns.regplot(x=np.arange(len(marital_status_counts)), y=marital_status_counts, ci=None, line_kws={"color": "red"})
# Đặt nhãn và tiêu đề
plt.xticks(range(len(marital_status_counts)), marital_status_counts.index, rotation=45)
plt.xlabel('Mỗi quan hệ hôn nhân')
plt.ylabel('Số lượng tĩ phú')
plt.title('Số lượng tĩ phú theo mỗi quan hệ hôn nhân (Seaborn)')
# Hiện thị đồ thị
plt.show()
```



- Thống kê tỉ trọng 3 lĩnh vực kinh doanh liên quan (công nghệ, tài chính thương mại, giải trí – dịch vụ) đối với tổng số ngành kinh doanh của các tỉ phú.

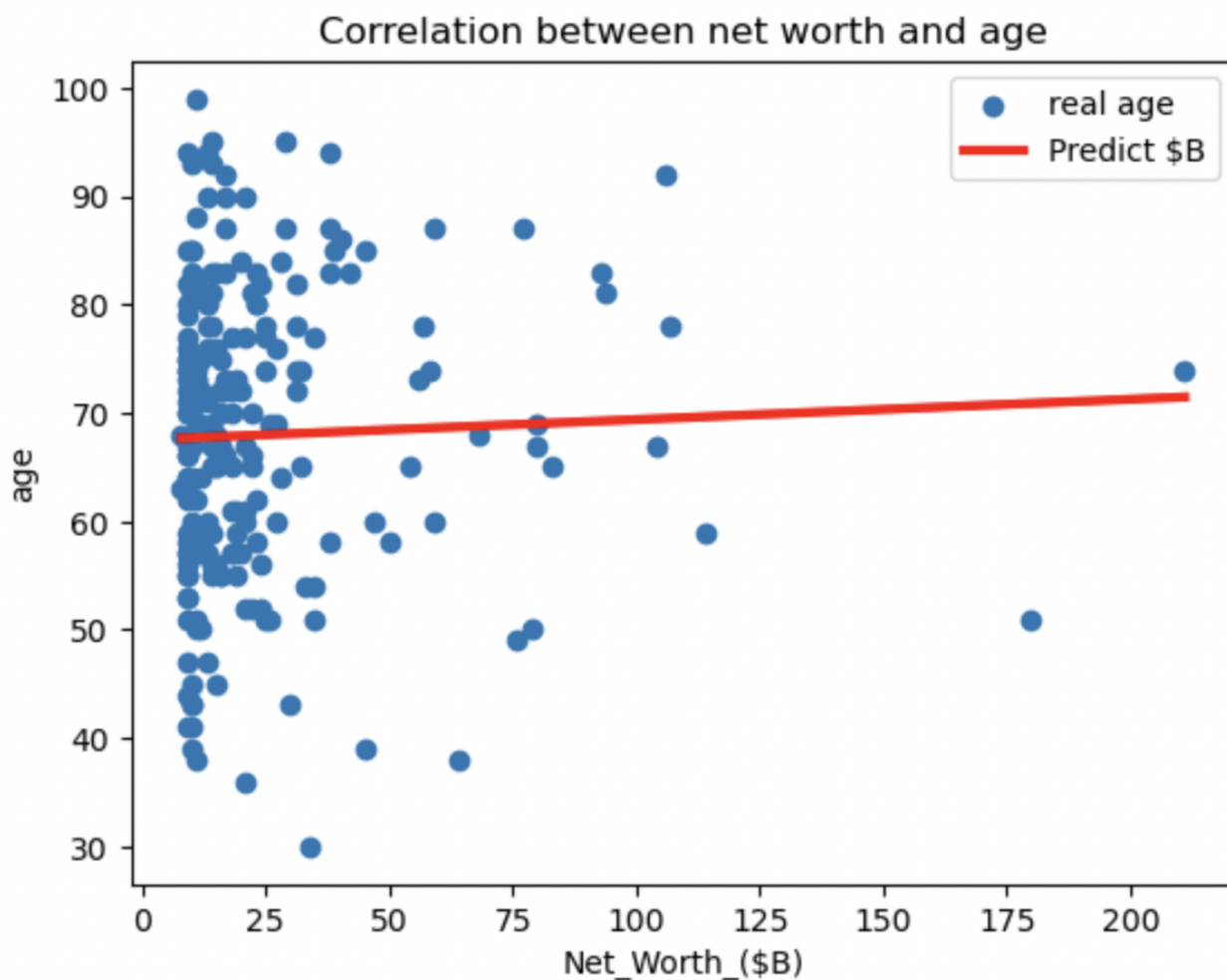
```
#Task 3.2
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
f= "/Users/lengochuynh/Downloads/data_Billionaires.listed.Forbes - data_Billionaires.listed.Forbes.csv"
# Gộp giá trị của các lĩnh vực kinh doanh
df['Industry'] = df['Industry'].replace({'Technology': 'Công nghệ',
                                         'Energy': 'Công nghệ',
                                         'Finance & Investments': 'Tài chính thương mại',
                                         'Logistics': 'Tài chính thương mại',
                                         'Shipping': 'Tài chính thương mại',
                                         'Fashion & Retail': 'Giải trí – dịch vụ',
                                         'Service': 'Giải trí – dịch vụ'})
selected_Industry = [ 'Công nghệ', 'Tài chính thương mại', 'Giải trí – dịch vụ']
# Lọc dữ liệu theo lĩnh vực kinh doanh
filtered_df=df[df['Industry'].isin(selected_Industry)]
# Thống kê số lượng tỷ phú theo lĩnh vực kinh doanh
Industry_counts=filtered_df['Industry'].value_counts()
# Vẽ đồ thị minh họa theo mô hình regression bằng Seaborn
sns.regplot(x=np.arange(len(Industry_counts)), y=Industry_counts, ci=None, line_kws={"color":"red"})
# Đặt nhãn và tiêu đề
plt.xticks(range(len(Industry_counts)), Industry_counts.index, rotation=45)
plt.xlabel('Lĩnh vực kinh doanh')
plt.ylabel('Số lượng tỷ phú')
plt.title('Số lượng tỷ phú theo lĩnh vực kinh doanh (Seaborn)')
# Hiển thị đồ thị
plt.show()
```





#### Task IV:

```
#Task 4
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
f= "/Users/lengochuynh/Downloads/data_Billionaires.listed.Forbes - data_Billionaires.listed.Forbes.csv"
df=pd.read_csv(f)
X= df[['Net_Worth_($B)']]
Y= df[['age']]
model= LinearRegression()
model.fit(X,Y)
df['Predicted_age']= model.predict(X)
plt.scatter(df['Net_Worth_($B)'],df['age'],label='real age')
plt.plot(df['Net_Worth_($B)'],df['Predicted_age'],label = 'Predict $B',color='red', linewidth=3)
plt.xlabel('Net_Worth_($B)')
plt.ylabel('age')
plt.legend()
plt.title('Correlation between net worth and age')
plt.show()
```



Sự tương quan không theo mô hình tuyến tính