

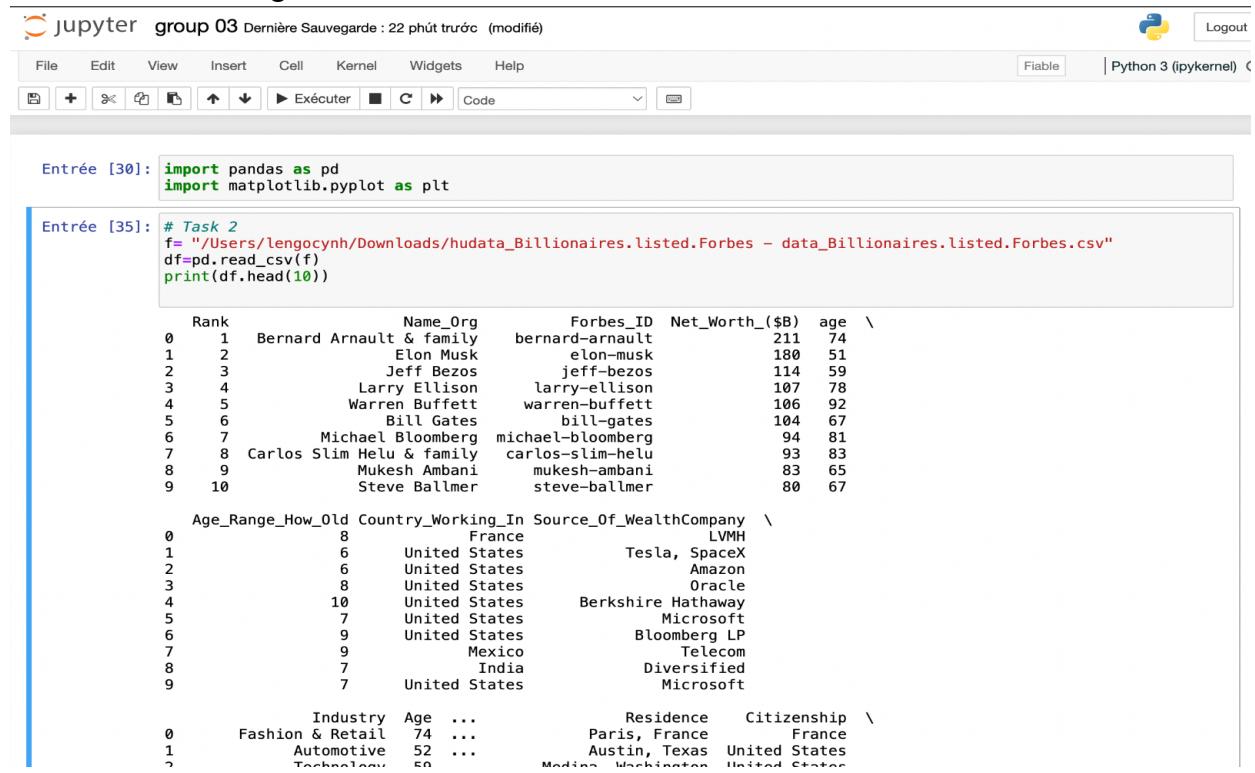
**ĐẠI HỌC NÔNG LÂM TP. HỒ CHÍ MINH**  
**KHOA TÀI NGUYÊN MÔI TRƯỜNG - NGÀNH HỆ THỐNG THÔNG TIN**  
**Học phần: Cơ sở và ứng dụng Trí tuệ nhân tạo**  
**2023 - 2024**

**REPORT FORM - PRACTICE 01**  
**GROUP 03**  
**Ngày: 01/12/2023**

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

**Task II:**

- Nhúng dữ liệu (file .csv) vào code để tiến hành phân tích dữ liệu → show 10 hàng dữ liệu đầu tiên của bảng dữ liệu



The screenshot shows a Jupyter Notebook interface with the following details:

- Header:** jupyter group 03 Dernière Sauvegarde : 22 phút trước (modifié)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help.
- Cell Buttons:** New, Cell, Run, Stop, Kernel, Help, Code.
- Output:**
  - Entrée [30]: `import pandas as pd  
import matplotlib.pyplot as plt`
  - Entrée [35]: `# Task 2  
f= "/Users/lengocynh/Downloads/hudata_Billionaires.listed.Forbes - data_Billionaires.listed.Forbes.csv"  
df=pd.read_csv(f)  
print(df.head(10))`
  - Output (Table): A table showing the top 10 rows of the Forbes Billionaires dataset. The columns include Rank, Name\_Org, Forbes\_ID, Net\_Worth\_(\\$B), age, Age\_Range\_How\_Old, Country\_Working\_In, Source\_Of\_WealthCompany, Industry, Age, Residence, and Citizenship.

- Liệt kê các đặc trưng của dữ liệu (file dữ liệu thể hiện các thông số, nội dung gì?)

```

Entrée [36]: print(df.describe())
      Rank Net_Worth_(\$B)    age  Age_Range_How_Old    Age \
count  221.000000  221.000000  221.000000  221.000000 \
mean   110.239819  24.235294  67.963801  7.18095  68.330317 \
std    63.251600  26.239900 13.609774  1.570698 13.643994 \
min     1.000000  8.000000  30.000000  0.000000 31.000000 \
25%    56.000000 10.000000  58.000000  6.000000 58.000000 \
50%   108.000000 15.000000  68.000000  7.000000 69.000000 \
75%   165.000000 25.000000  78.000000  8.000000 78.000000 \
max   220.000000 211.000000  99.000000 10.000000 99.000000 \
                                           \
Self-Made Score Philanthropy Score Children Bachelor \
count  221.000000  221.000000  221.000000  221.000000 \
mean   6.180995  1.280543  2.936652  0.493213 \
std    1.619147  0.752590  1.588471  0.501089 \
min     1.000000  1.000000  1.000000  0.000000 \
25%    6.000000  1.000000  2.000000  0.000000 \
50%    6.000000  1.000000  2.000000  0.000000 \
75%    6.000000  1.000000  3.000000  1.000000 \
max   10.000000  5.000000 10.000000  1.000000 \
                                           \
Master Doctorate Drop Out Self_Made \
count  221.000000  221.000000  221.000000  221.000000 \
mean   0.294118  0.036199  0.072398  0.660633 \
std    0.456679  0.187209  0.259734  0.474560 \
min     0.000000  0.000000  0.000000  0.000000 \
25%    0.000000  0.000000  0.000000  0.000000 \
50%    0.000000  0.000000  0.000000  1.000000 \
75%    1.000000  0.000000  0.000000  1.000000 \
max   1.000000  1.000000  1.000000  1.000000 \

```

```

Entrée [43]: # Sắp xếp DataFrame theo một cột cụ thể (ví dụ: cột 'ColumnName' từ cao đến thấp)
sorted_df = df.sort_values(by="Net_Worth_(\$B)", ascending=False)
print(sorted_df)
      Rank Name_Org Forbes_ID Net_Worth_(\$B) age \
0      1 Bernard Arnault & family bernard-arnault 211    74 \
1      2 Elon Musk elon-musk 180    51 \
2      3 Jeff Bezos jeff-bezos 114    59 \
3      4 Larry Ellison larry-ellison 107    78 \
4      5 Warren Buffett warren-buffett 106    92 \
..    ... ... ... ... ... \
195  195 Jin Baofang jin-baofang 9    70 \
197  195 Marijke Mars marijke-mars 9    58 \
198  195 Pamela Mars pamela-mars 9    62 \
219  220 Finn Rausing finn-rausing 8    68 \
220  220 Jorn Rausing jorn-rausing 8    63 \

```

## Task III + IV:

- Sắp xếp thứ tự những tỉ phú theo thứ tự từ cao đến thấp → lưu lại thành file .csv mới

```

Entrée [43]: #task_3]
# Sắp xếp DataFrame theo một cột cụ thể (ví dụ: cột 'ColumnName' từ cao đến thấp)
sorted_df = df.sort_values(by="Net_Worth_(\$B)", ascending=False)
print(sorted_df)
      Rank Name_Org Forbes_ID Net_Worth_(\$B) age \
0      1 Bernard Arnault & family bernard-arnault 211    74 \
1      2 Elon Musk elon-musk 180    51 \
2      3 Jeff Bezos jeff-bezos 114    59 \
3      4 Larry Ellison larry-ellison 107    78 \
4      5 Warren Buffett warren-buffett 106    92 \
..    ... ... ... ... ... \
195  195 Jin Baofang jin-baofang 9    70 \
197  195 Marijke Mars marijke-mars 9    58 \
198  195 Pamela Mars pamela-mars 9    62 \
219  220 Finn Rausing finn-rausing 8    68 \
220  220 Jorn Rausing jorn-rausing 8    63 \

```

```

      Age_Range_How_Old Country_Working_In Source_Of_WealthCompany \
0                  8 France                   LVMH \
1                  6 United States           Tesla, SpaceX \
2                  6 United States            Amazon \
3                  8 United States             Oracle \
4                 10 United States        Berkshire Hathaway \
..                ... ... ... ... ... \
195                7 China          Solar panels \
197                6 United States       Candy, pet food \
198                7 United States       Candy, pet food \
219                7 Sweden            Packaging \
220                7 Sweden            Packaging \

```

```

      Industry Age ... Residence Citizenship \
0  Fashion & Retail 74 ... Paris, France  France \
1  Automotive 52 ... Austin, Texas United States \
2  Technology 59 ... Medina, Washington United States \
3  Technology 79 ... Lanai, Hawaii United States \
4 Finance & Investments 92 ... Omaha, Nebraska United States \
.. ... ... ... ... ...

```

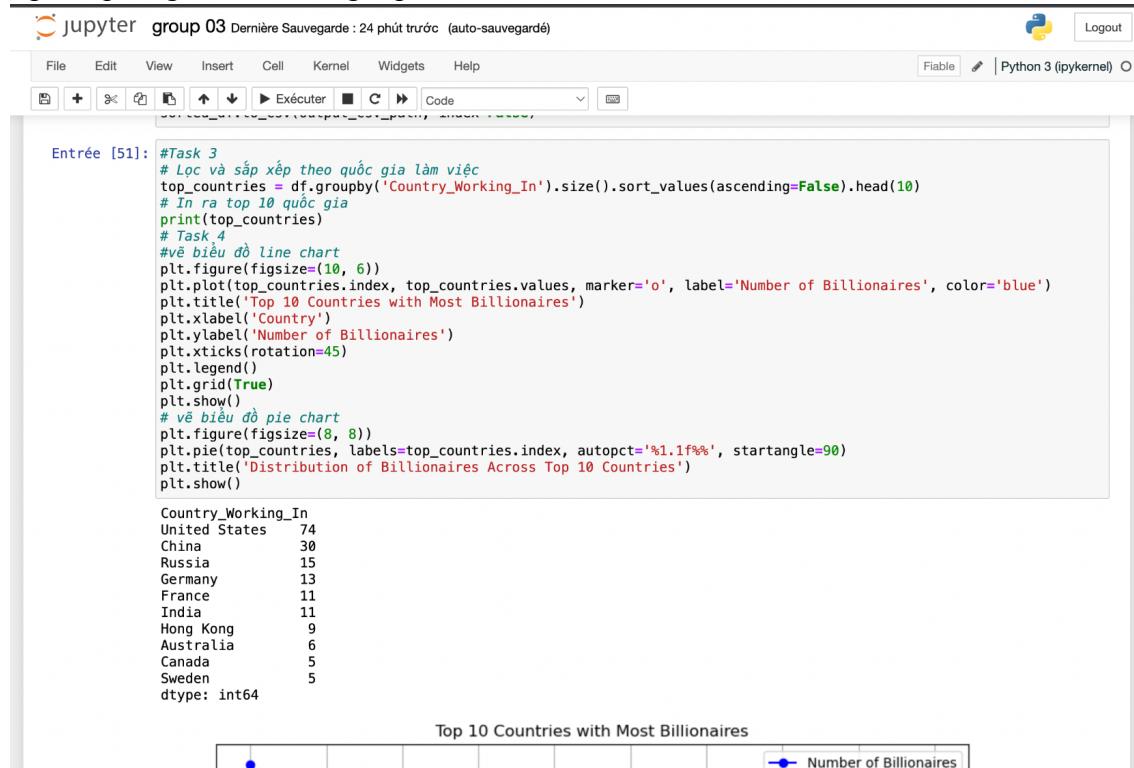
```

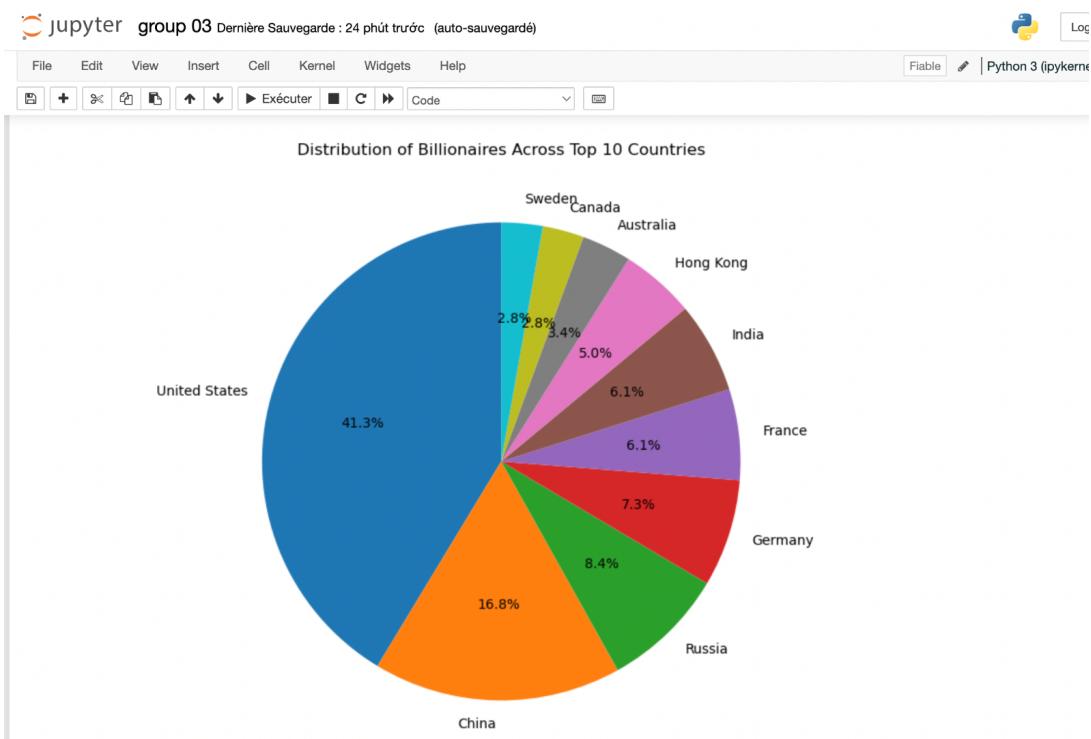
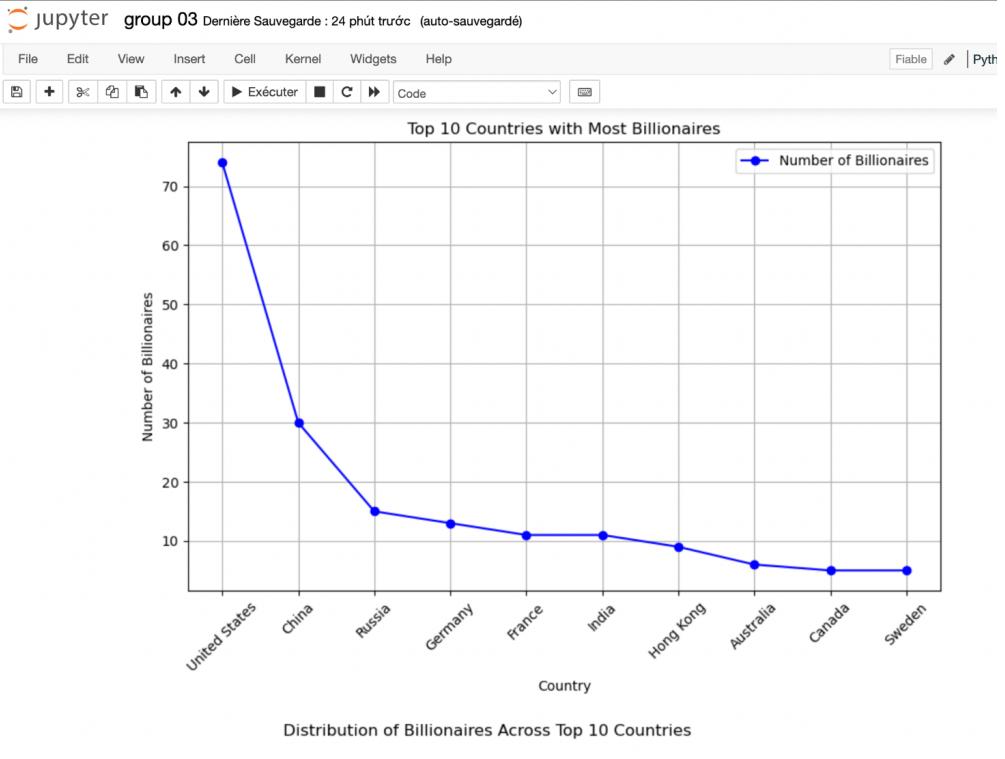
Entrée [45]: output_csv_path = '/Users/lengochuynh/Downloads/task3_1.csv'
# Lưu DataFrame đã sắp xếp vào một tệp CSV mới
sorted_df.to_csv(output_csv_path, index=False)

```

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Rank	Name_Org	Forbes_ID	Net_Worth	age	Age_Rang	Country_W	Source_Of	Industry	Age	Source of	Self-Made	Philanthro
2	1	Bernard Arnault	bernard-arn	211	74	8	France	LVMH	Fashion & L	74	LVMH	6	1
3	2	Elon Musk	elon-musk	180	51	6	United States	Tesla, Spa	Automotive	52	Tesla, Spa	8	1
4	3	Jeff Bezos	jeff-bezos	114	59	6	United States	Amazon	Technology	59	Amazon, Self	8	2
5	4	Larry Ellison	larry-elliso	107	78	8	United States	Oracle	Technology	79	Oracle, Self	9	1
6	5	Warren Buffett	warren-buffet	106	92	10	United States	Berkshire Hathaway	Finance & Insurance	92	Berkshire Hathaway	8	5
7	6	Bill Gates	bill-gates	104	67	7	United States	Microsoft	Technology	67	Microsoft, Self	8	5
8	7	Michael Bloomberg	michael-bl	94	81	9	United States	Bloomberg	Media & Entertainment	81	Bloomberg	8	4
9	8	Carlos Slim	carlos-slim	93	83	9	Mexico	Telecom	Telecom	83	Telecom, Self	6	1
10	9	Mukesh Ambani	mukesh-ambani	83	65	7	India	Diversified	Diversified	66	Diversified	6	1
11	10	Steve Ballmer	steve-ballmer	80	67	7	United States	Microsoft	Technology	67	Microsoft, Self	6	2
12	11	François Pinault	francoise-pinault	80	69	7	France	L'Oréal	Fashion & Retail	70	L'Oréal	6	1
13	12	Larry Page	larry-page	79	50	5	United States	Google	Technology	50	Google, Self	8	1
14	13	Amancio Ortega	amancio-ortega	77	87	9	Spain	Zara	Fashion & Retail	87	Zara, Self	6	1
15	14	Sergey Brin	sergey-bri	76	49	5	United States	Google	Technology	50	Google, Self	9	2
16	15	Zhong Shanzhen	zhong-sha	68	68	7	China	Beverages	Food & Beverage	68	Beverages	6	1
17	16	Mark Zuckerberg	mark-zuck	64	38	4	United States	Facebook	Technology	39	Facebook, Self	8	3
18	17	Charles Koch	charles-koch	59	87	9	United States	Koch Industries	Diversified	87	Koch Industries	5	2
19	17	Julia Koch	julia-koch	59	60	6	United States	Koch Industries	Diversified	61	Koch Industries	1	2
20	19	Jim Walton	jim-walton	58	74	8	United States	Walmart	Fashion & Retail	75	Walmart	2	1
21	20	Rob Walton	rob-walton	57	78	8	United States	Walmart	Fashion & Retail	78	Walmart	4	1
22	21	Alice Walton	alice-walton	56	73	8	United States	Walmart	Fashion & Retail	73	Walmart	1	2
23	22	David Thorne	david-thorne	54	65	7	Canada	Media	Media & Entertainment	66	Media	6	1
24	23	Michael Dell	michael-dell	50	58	6	United States	Dell Technologies	Technology	58	Dell Technologies	8	2
25	24	Gautam Agarwal	agarwal-ag	47	60	6	India	Infrastructure	Diversified	61	Infrastructure	6	1
26	25	Phil Knight	phil-knight	45	85	9	United States	Nike	Fashion & Retail	85	Nike, Self	8	3
27	26	Zhang Yimou	zhang-yimou	45	39	4	China	TikTok	Technology	39	TikTok, Self	6	1
28	27	Dieter Schindler	dieter-schindler	42	83	9	Germany	Retail	Fashion & Retail	83	Retail	6	1
29	28	François Pinault	francois-pinault	40	86	9	France	Luxury goods	Fashion & Retail	87	Luxury goods	6	1
30	29	Klaus-Michael Müller	klaus-michael-mueller	39	85	9	Germany	Shipping	Logistics	86	Shipping	6	1

☐ Lọc và sắp xếp các tỉ phú theo quốc gia nơi họ làm việc (Country\_Working) → lọc ra top 10 quốc gia có số lượng tỉ phú nhiều nhất





- ❑ Lọc và sắp xếp các tỉ phú theo quốc tịch (Citizenship) → lọc ra top 10 quốc gia có số lượng tỉ phú nhiều nhất

jupyter group 03 Dernière Sauvegarde : 24 phút trước (auto-sauvegardé)

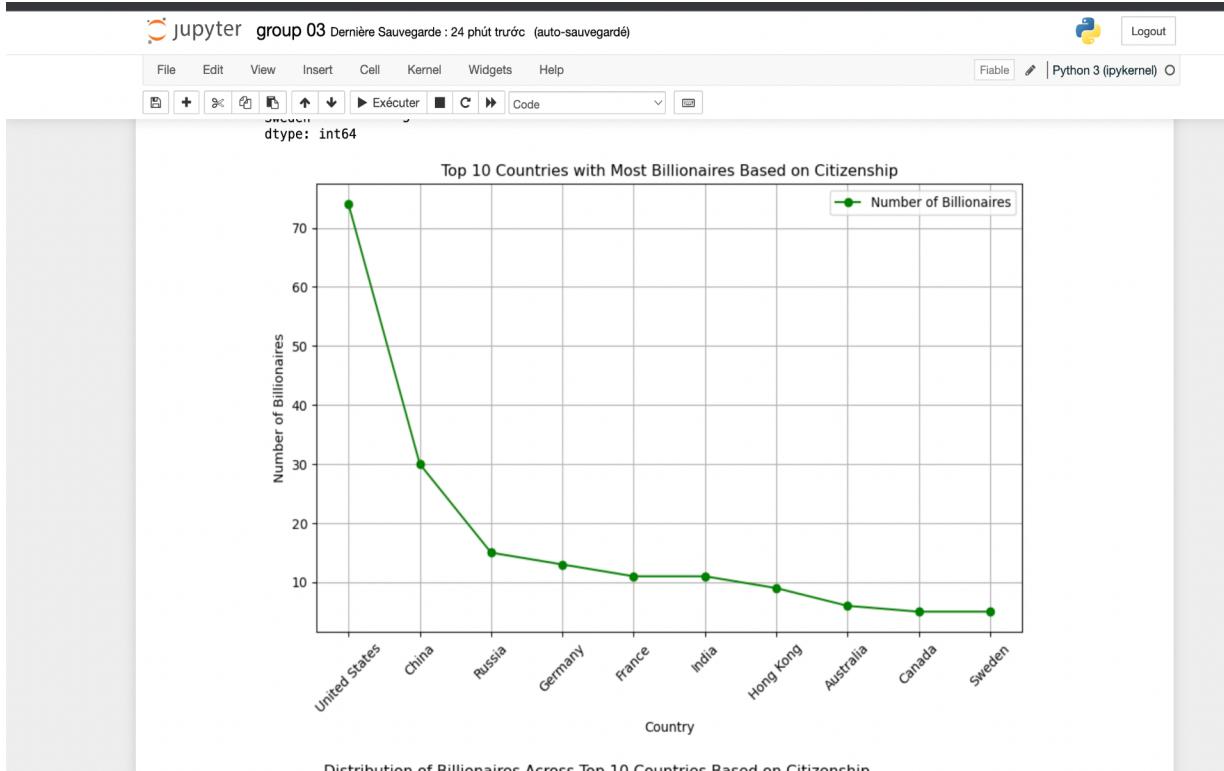
File Edit View Insert Cell Kernel Widgets Help

Fiable | Python 3 (ipykernel) C

```
Entrée [53]: #Task 3
top_citizenships = df.groupby('Citizenship').size().sort_values(ascending=False).head(10)

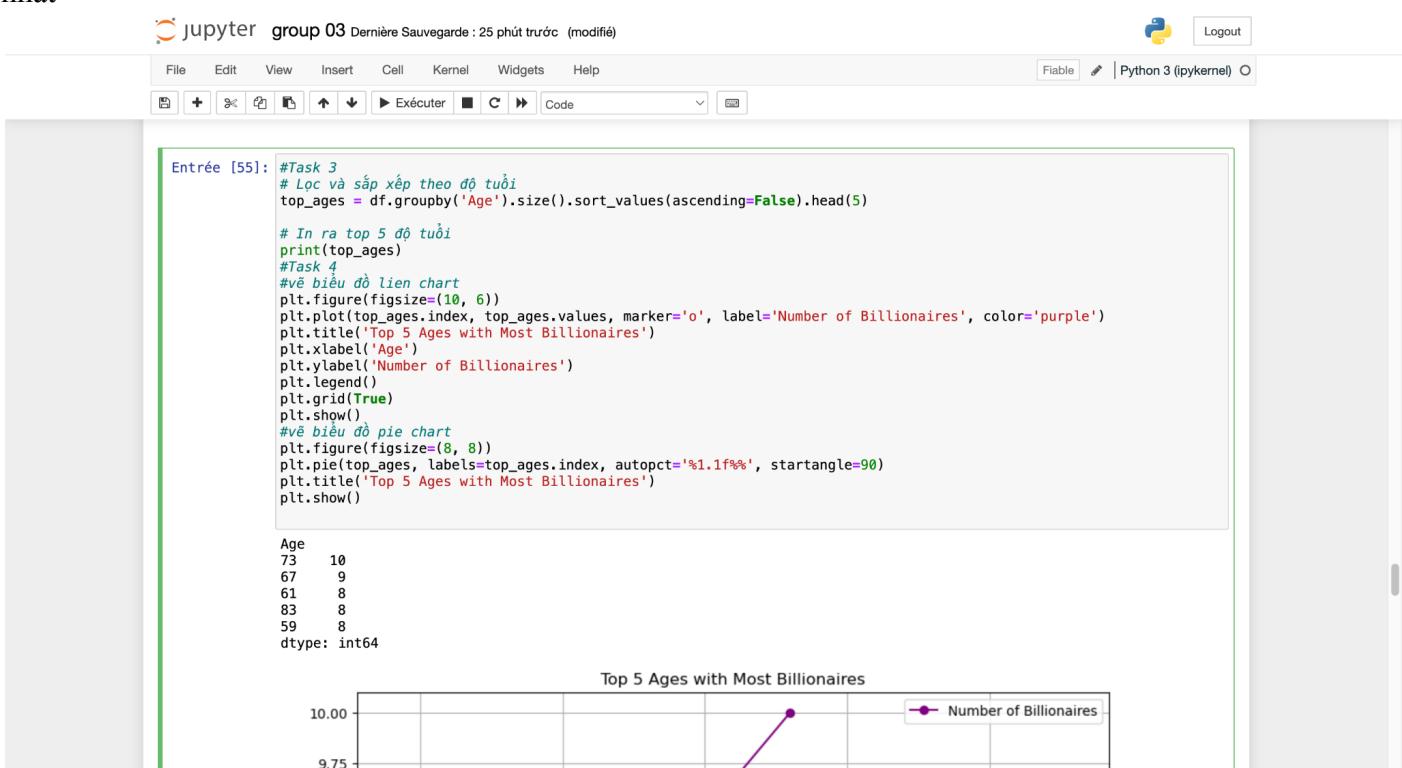
# In ra top 10 quốc tịch
print(top_citizenships)
#Task 4
#vẽ biểu đồ line chart
plt.figure(figsize=(10, 6))
plt.plot(top_citizenships.index, top_citizenships.values, marker='o', label='Number of Billionaires', color='green')
plt.title('Top 10 Countries with Most Billionaires Based on Citizenship')
plt.xlabel('Country')
plt.ylabel('Number of Billionaires')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
plt.show()
#Task 4
# vẽ biểu đồ pie chart
# Vẽ Pie Chart
plt.figure(figsize=(8, 8))
plt.pie(top_citizenships, labels=top_citizenships.index, autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Billionaires Across Top 10 Countries Based on Citizenship')
plt.show()

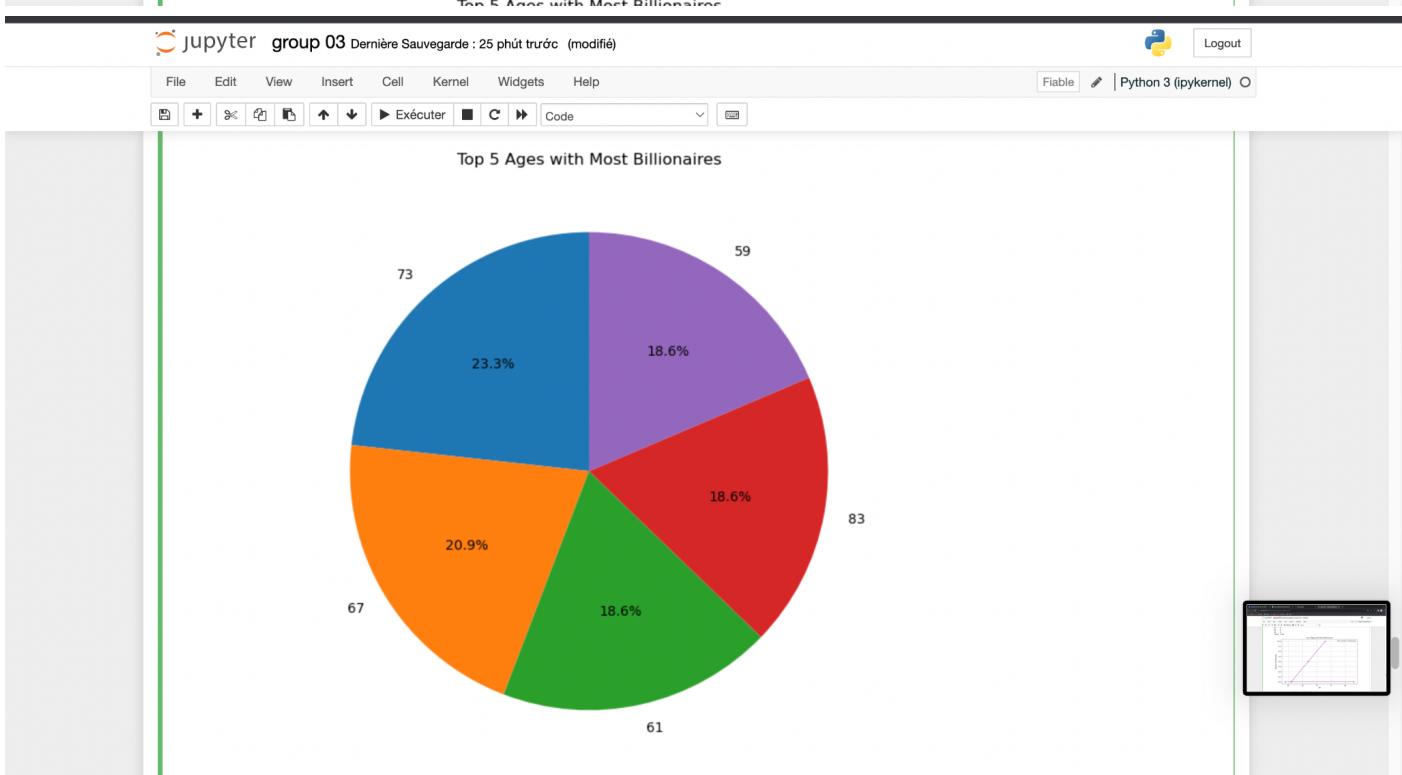
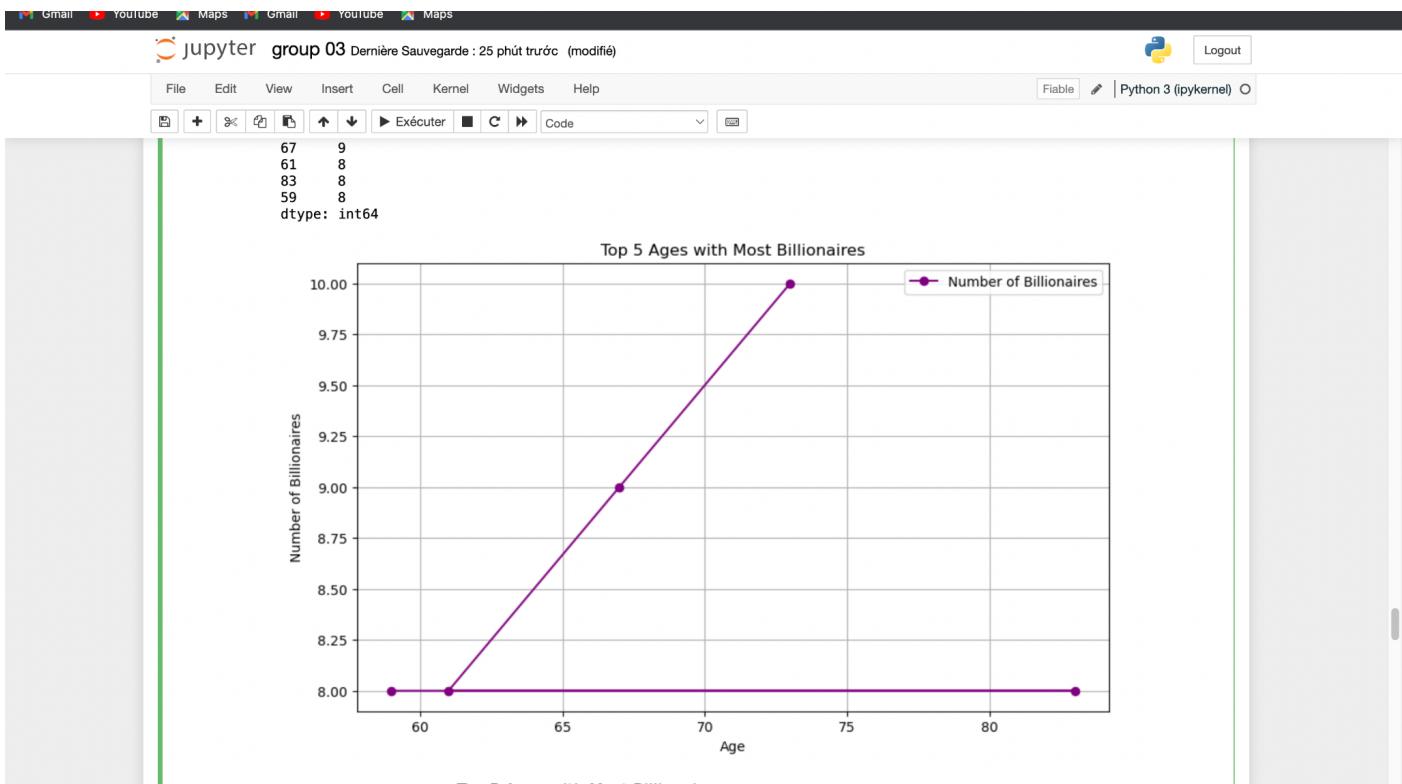
Citizenship
United States    74
China            30
Russia           15
Germany          13
France           11
India             11
Hong Kong         9
Australia         6
Canada            5
Sweden            5
dtype: int64
```





❑ Lọc và sắp xếp các tỉ phú theo độ tuổi (age) → top 5 độ tuổi có số lượng tỉ phú nhiều nhất





❑ Lọc và sắp xếp các tỉ phú theo nhóm ngành nghề của họ → top 10 ngành nghề có số lượng tỉ phú nhiều nhất

jupyter group 03 Dernière Sauvegarde : 6 phút trước (auto-sauvegardé)

File Edit View Insert Cell Kernel Widgets Help

Fiable Python 3 (ipykernel) O

```
Entrée [56]: # Task 3
top_industries = df.groupby('Industry').size().sort_values(ascending=False).head(10)

# In ra top 10 ngành nghề
print(top_industries)
# Task 4
# vẽ biểu đồ line chart
plt.figure(figsize=(10, 6))
plt.plot(top_industries.index, top_industries.values, marker='o', label='Number of Billionaires', color='purple')
plt.title('Top 10 Industries with most Billionaires')
plt.xlabel('Age')
plt.ylabel('Number of Billionaires')
plt.legend()
plt.grid(True)
plt.show()
# vẽ biểu đồ pie chart
plt.figure(figsize=(8, 8))
plt.pie(top_industries, labels=top_industries.index, autopct='%1.1f%%', startangle=90)
plt.title('Top 10 Industries with most Billionaires')
plt.show()
```

Industry	Number of Billionaires
Technology	32
Fashion & Retail	30
Finance & Investments	28
Food & Beverage	20
Diversified	16
Manufacturing	15
Metals & Mining	14
Energy	12
Healthcare	12
Automotive	11

dtype: int64

Top 10 Industries with most Billionaires

Industry	Number of Billionaires
Technology	32
Fashion & Retail	30
Finance & Investments	28
Food & Beverage	20
Diversified	16
Manufacturing	15
Metals & Mining	14
Energy	12
Healthcare	12
Automotive	11

