

SPRAWOZDANIE

Zajęcia: Eksploracja i wizualizacja danych
Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium: 2

Data: 09.03.2023

Temat: "Graficzna wizualizacja danych z użyciem bibliotek matplotlib i seaborn"

Wariant: 7

Michał Stajerski

Informatyka II stopień,
stacjonarne,
semestr 3,

Gr. 1

<https://github.com/MichalStajerski/eiwd>

1. Polecenie

Celem zajęć jest implementacja wszystkich możliwości tworzenia wykresów na podstawie tutorialu. Wariant jest określony zestawem danych. Dane do zadania zostały pobrane ze strony https://ghdx.healthdata.org/ihme_data. Wariant wybrany w zadaniu jest wariant 7:

Global Burden of Disease Study 2019 (GBD 2019) Smoking Tobacco Use Prevalence 1990-2019

2. Zadania

1 - Załadowanie potrzebnych bibliotek

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

2 - Załadowanie potrzebnych danych

```
df = pd.read_csv("IHME_GBD_2019_SMOKING_TOB_1990_2019_NUM_SMOKERS_Y2021M05D27.csv", encoding = "utf-8")  
df
```

	measure_name	location_id	location_name	sex_id	sex_name	age_group_id	age_group_name	year_id	val	upper	lower
0	Number of Smokers	1	Global	1	Male	29	15+ years	1990	8.031015e+08	8.096221e+08	7.959086e+08
1	Number of Smokers	1	Global	2	Female	29	15+ years	1990	1.891488e+08	1.930929e+08	1.855595e+08
2	Number of Smokers	1	Global	3	Both	29	15+ years	1990	9.922503e+08	1.000161e+09	9.847880e+08
3	Number of Smokers	1	Global	1	Male	29	15+ years	1991	8.138972e+08	8.200339e+08	8.069514e+08
4	Number of Smokers	1	Global	2	Female	29	15+ years	1991	1.905375e+08	1.944249e+08	1.869744e+08
...
20965	Number of Smokers	522	Sudan	2	Female	29	15+ years	2018	2.435999e+05	3.286166e+05	1.752508e+05
20966	Number of Smokers	522	Sudan	3	Both	29	15+ years	2018	2.610672e+06	2.833943e+06	2.409108e+06
20967	Number of Smokers	522	Sudan	1	Male	29	15+ years	2019	2.439150e+06	2.656579e+06	2.236450e+06
20968	Number of Smokers	522	Sudan	2	Female	29	15+ years	2019	2.500800e+05	3.345384e+05	1.816686e+05
20969	Number of Smokers	522	Sudan	3	Both	29	15+ years	2019	2.689230e+06	2.918332e+06	2.480656e+06

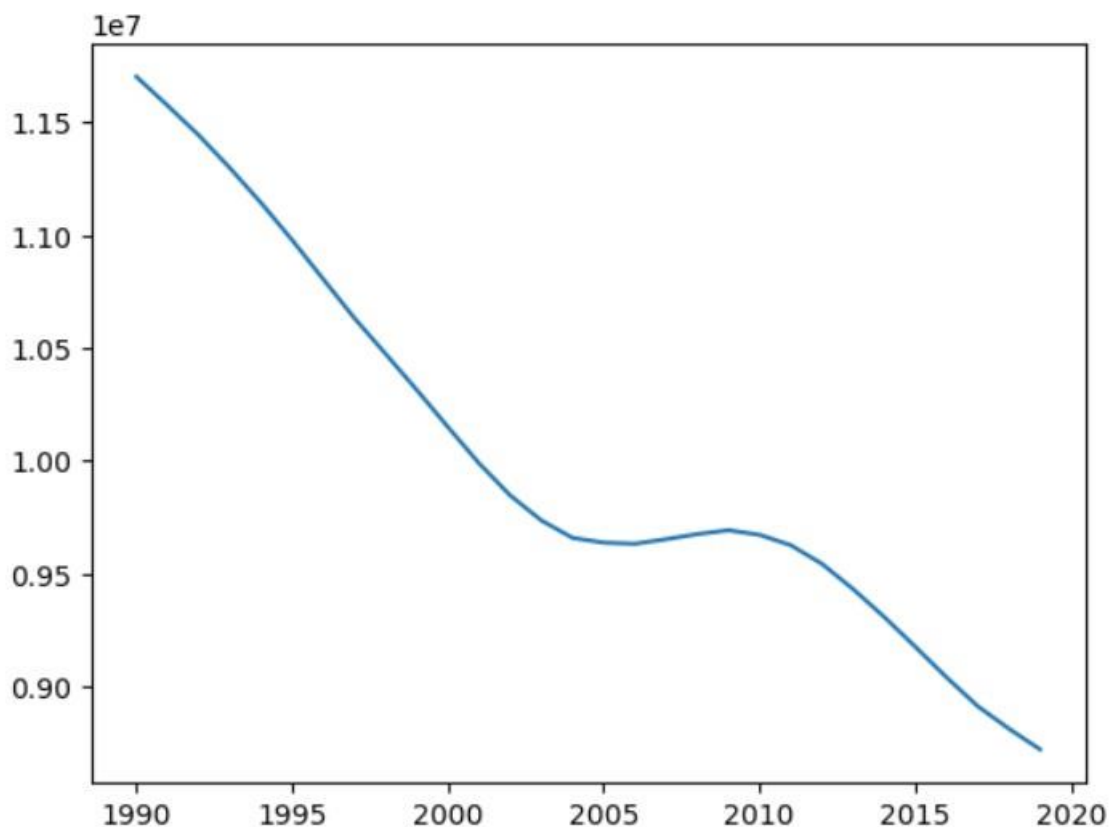
3 - użycie matplotlib

```
matplot = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]  
matplot
```

	measure_name	location_id	location_name	sex_id	sex_name	age_group_id	age_group_name	year_id	val	upper	lower
4322	Number of Smokers	51	Poland	3	Both	29	15+ years	1990	1.170448e+07	1.201222e+07	1.138875e+07
4325	Number of Smokers	51	Poland	3	Both	29	15+ years	1991	1.157647e+07	1.188570e+07	1.125896e+07
4328	Number of Smokers	51	Poland	3	Both	29	15+ years	1992	1.144506e+07	1.175827e+07	1.112616e+07
4331	Number of Smokers	51	Poland	3	Both	29	15+ years	1993	1.129936e+07	1.162928e+07	1.097590e+07
4334	Number of Smokers	51	Poland	3	Both	29	15+ years	1994	1.114488e+07	1.147066e+07	1.080890e+07
4337	Number of Smokers	51	Poland	3	Both	29	15+ years	1995	1.097960e+07	1.132458e+07	1.062839e+07
4340	Number of Smokers	51	Poland	3	Both	29	15+ years	1996	1.080558e+07	1.115800e+07	1.045887e+07
4343	Number of Smokers	51	Poland	3	Both	29	15+ years	1997	1.063264e+07	1.098329e+07	1.027121e+07
4346	Number of Smokers	51	Poland	3	Both	29	15+ years	1998	1.047442e+07	1.083031e+07	1.012083e+07
4349	Number of Smokers	51	Poland	3	Both	29	15+ years	1999	1.031418e+07	1.068428e+07	9.960286e+06
4352	Number of Smokers	51	Poland	3	Both	29	15+ years	2000	1.014975e+07	1.051794e+07	9.794037e+06
4355	Number of Smokers	51	Poland	3	Both	29	15+ years	2001	9.988362e+06	1.032469e+07	9.655642e+06
4358	Number of Smokers	51	Poland	3	Both	29	15+ years	2002	9.845932e+06	1.014307e+07	9.532336e+06
4361	Number of Smokers	51	Poland	3	Both	29	15+ years	2003	9.735058e+06	1.002336e+07	9.443244e+06
4364	Number of Smokers	51	Poland	3	Both	29	15+ years	2004	9.657597e+06	9.922015e+06	9.384474e+06
4367	Number of Smokers	51	Poland	3	Both	29	15+ years	2005	9.636582e+06	9.878019e+06	9.380627e+06
4370	Number of Smokers	51	Poland	3	Both	29	15+ years	2006	9.621582e+06	9.875022e+06	9.384114e+06

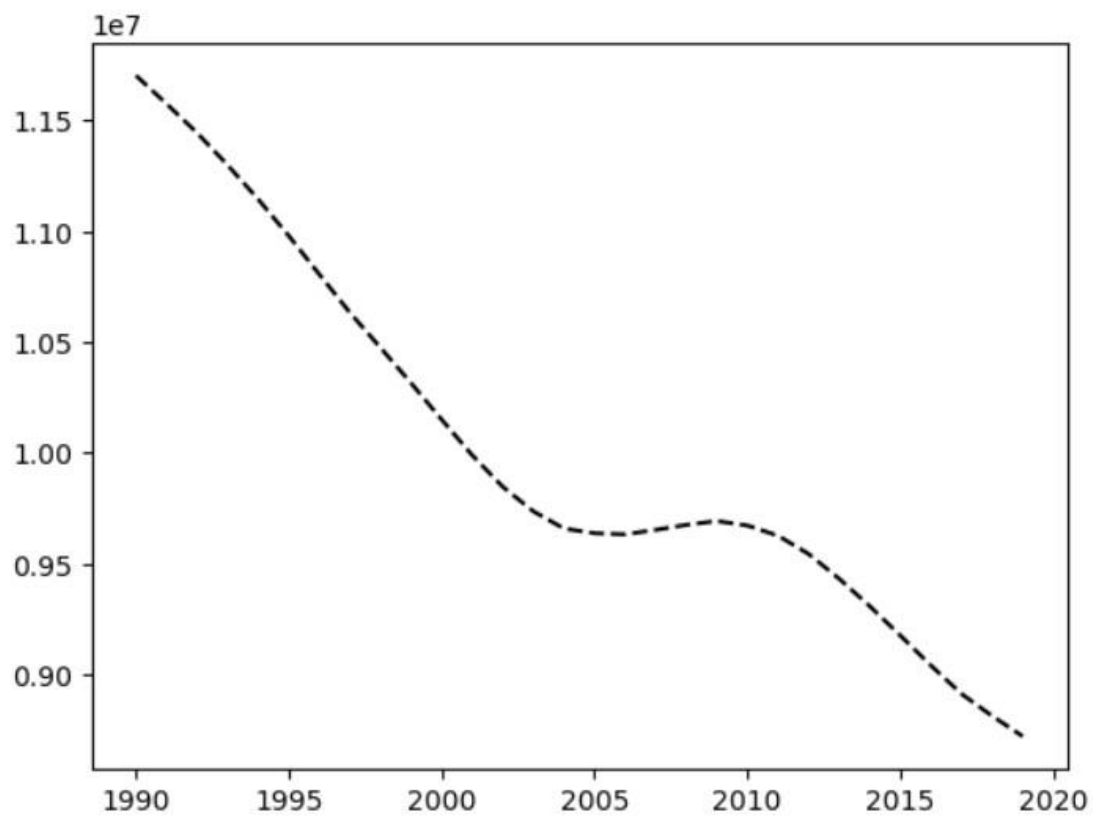
```
: plt.plot(matplot["year_id"], matplot["val"])
```

```
: [matplotlib.lines.Line2D at 0x261892ac340]
```



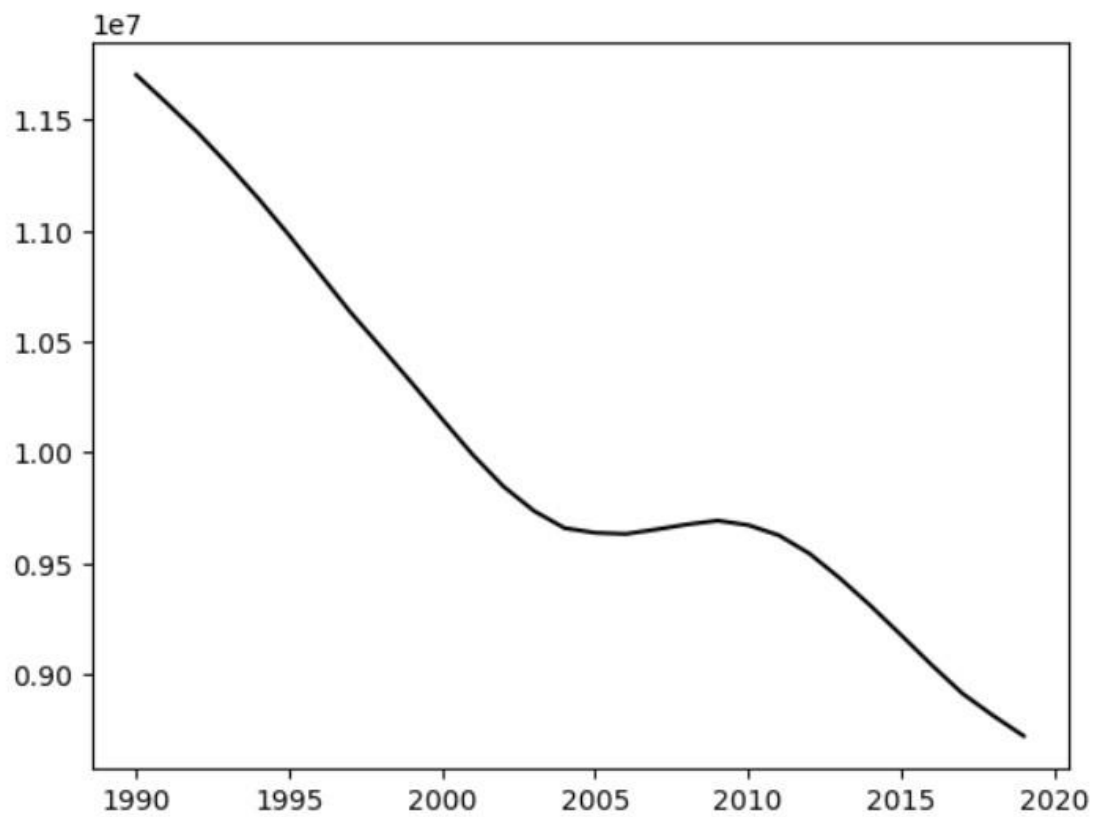
4 - przedstawienie różnych stylów wykresów

```
plt.plot(matplot["year_id"], matplot["val"], 'k--')  
[<matplotlib.lines.Line2D at 0x2618938d600>]
```



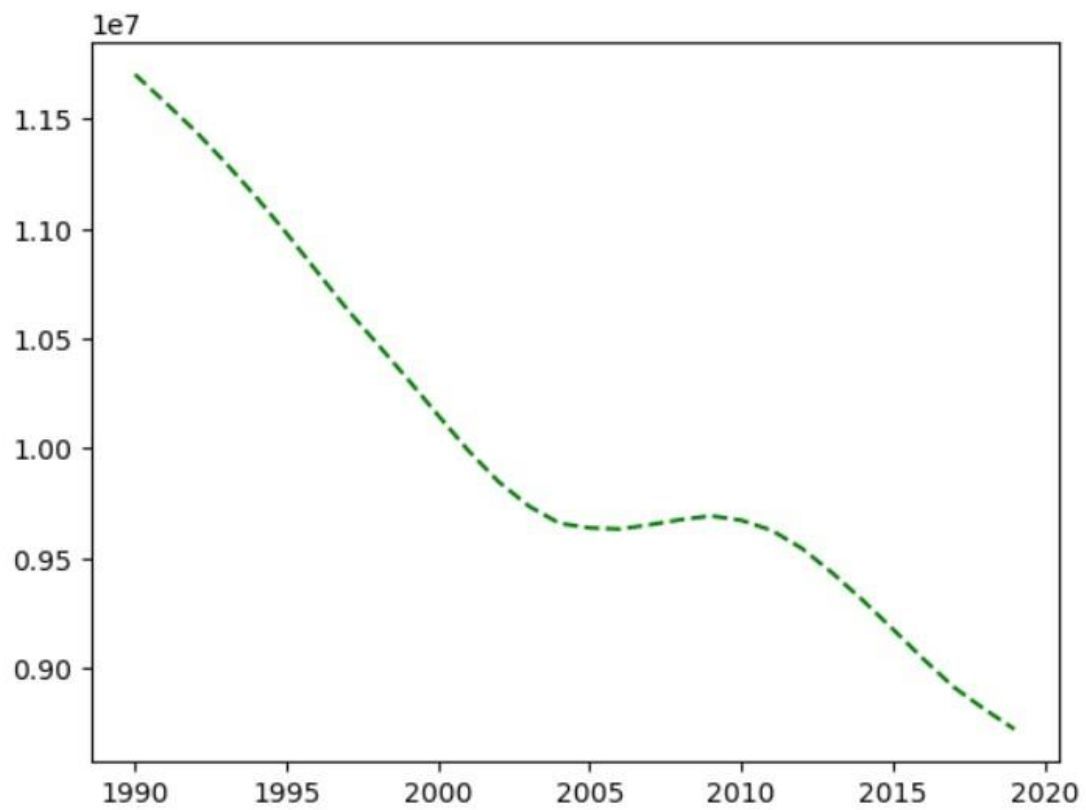
```
plt.plot(matplot["year_id"], matplot["val"], 'k-')
```

```
[<matplotlib.lines.Line2D at 0x26189bd33d0>]
```



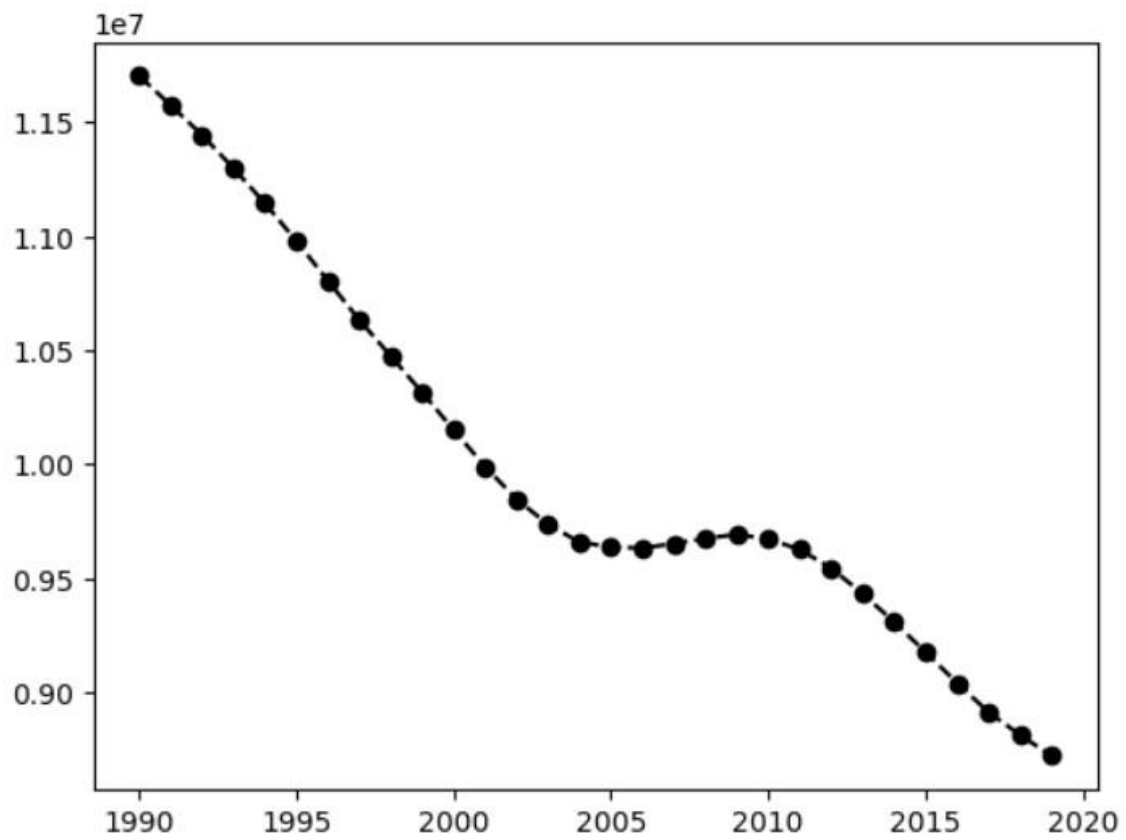
```
plt.plot(matplot["year_id"], matplot["val"], 'g--')
```

```
[<matplotlib.lines.Line2D at 0x26189479870>]
```



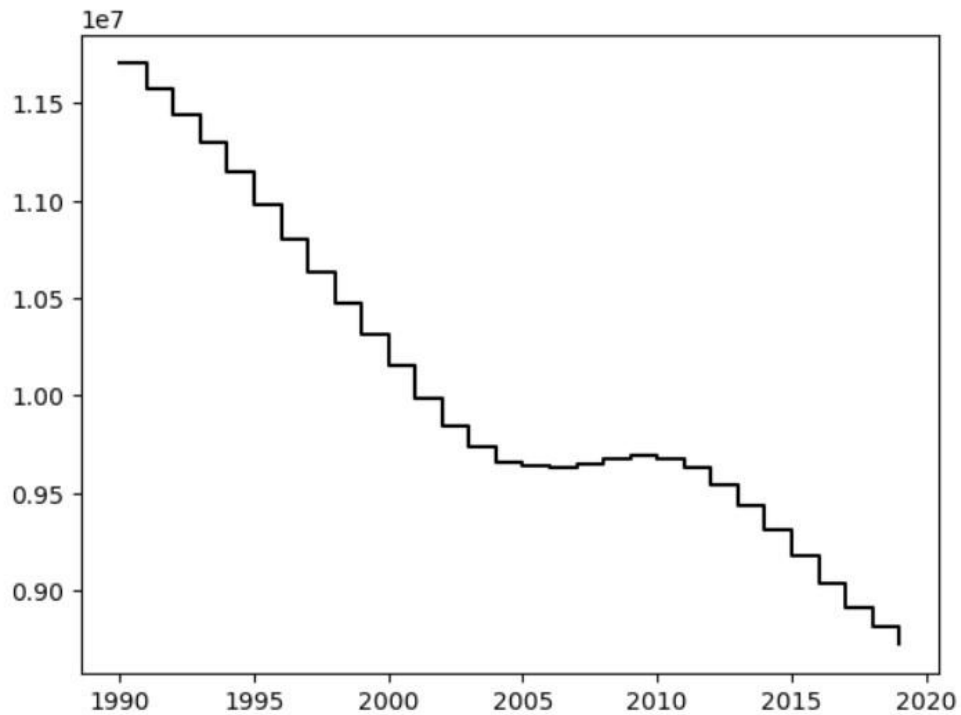
```
plt.plot(matplot["year_id"], matplot["val"], 'ko--')
```

```
[<matplotlib.lines.Line2D at 0x26189ea7d60>]
```



```
plt.plot(matplot["year_id"], matplot["val"], 'k-', drawstyle='steps-post', label='steps-post')
```

```
[<matplotlib.lines.Line2D at 0x2618a7fd870>]
```



5 - zapisywanie rysunków do pliku

```
: plt.savefig('figpath.png', dpi=400, bbox_inches='tight')
```

```
<Figure size 640x480 with 0 Axes>
```

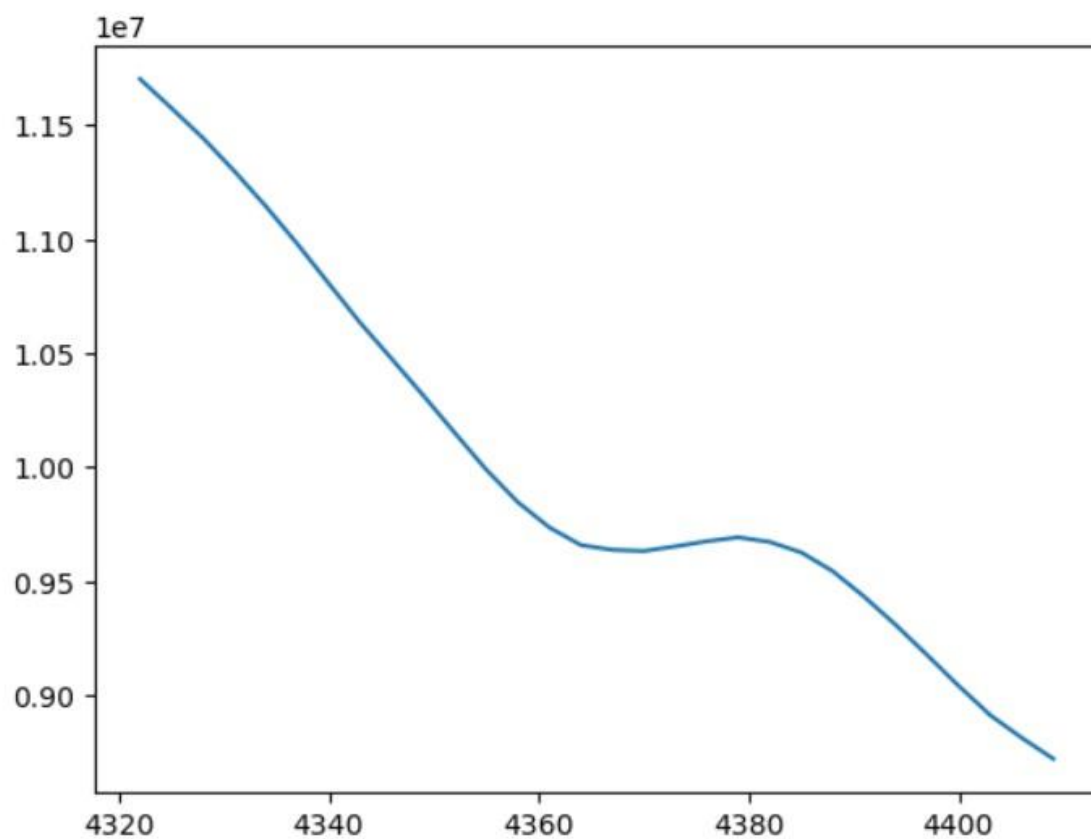
6 - wykresy liniowe

```
liniowy_series = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]  
liniowy_series
```


	measure_name	location_id	location_name	sex_id	sex_name	age_group_id	age_group_name	year_id	val	upper	lower
4322	Number of Smokers	51	Poland	3	Both	29	15+ years	1990	1.170448e+07	1.201222e+07	1.138875e+07
4325	Number of Smokers	51	Poland	3	Both	29	15+ years	1991	1.157647e+07	1.188570e+07	1.125896e+07
4328	Number of Smokers	51	Poland	3	Both	29	15+ years	1992	1.144506e+07	1.175827e+07	1.112616e+07
4331	Number of Smokers	51	Poland	3	Both	29	15+ years	1993	1.129936e+07	1.162928e+07	1.097590e+07
4334	Number of Smokers	51	Poland	3	Both	29	15+ years	1994	1.114488e+07	1.147066e+07	1.080890e+07
4337	Number of Smokers	51	Poland	3	Both	29	15+ years	1995	1.097960e+07	1.132458e+07	1.062839e+07
4340	Number of Smokers	51	Poland	3	Both	29	15+ years	1996	1.080558e+07	1.115800e+07	1.045887e+07
4343	Number of Smokers	51	Poland	3	Both	29	15+ years	1997	1.063264e+07	1.098329e+07	1.027121e+07
4346	Number of Smokers	51	Poland	3	Both	29	15+ years	1998	1.047442e+07	1.083031e+07	1.012083e+07
4349	Number of Smokers	51	Poland	3	Both	29	15+ years	1999	1.031418e+07	1.068428e+07	9.960286e+06
4352	Number of Smokers	51	Poland	3	Both	29	15+ years	2000	1.014975e+07	1.051794e+07	9.794037e+06
4355	Number of Smokers	51	Poland	3	Both	29	15+ years	2001	9.988362e+06	1.032469e+07	9.655642e+06
4358	Number of Smokers	51	Poland	3	Both	29	15+ years	2002	9.845932e+06	1.014307e+07	9.532336e+06
4361	Number of Smokers	51	Poland	3	Both	29	15+ years	2003	9.735058e+06	1.002336e+07	9.443244e+06
4364	Number of Smokers	51	Poland	3	Both	29	15+ years	2004	9.657597e+06	9.922015e+06	9.384474e+06
4367	Number of Smokers	51	Poland	3	Both	29	15+ years	2005	9.636582e+06	9.878019e+06	9.380627e+06
4370	Number of Smokers	51	Poland	3	Both	29	15+ years	2006	9.631582e+06	9.875923e+06	9.394114e+06
4373	Number of Smokers	51	Poland	3	Both	29	15+ years	2007	9.651840e+06	9.889532e+06	9.426812e+06
4376	Number of Smokers	51	Poland	3	Both	29	15+ years	2008	9.674385e+06	9.914297e+06	9.444482e+06

```
w liniowy_series = pd.Series(liniowy_series["val"])
w liniowy_series.plot()
```

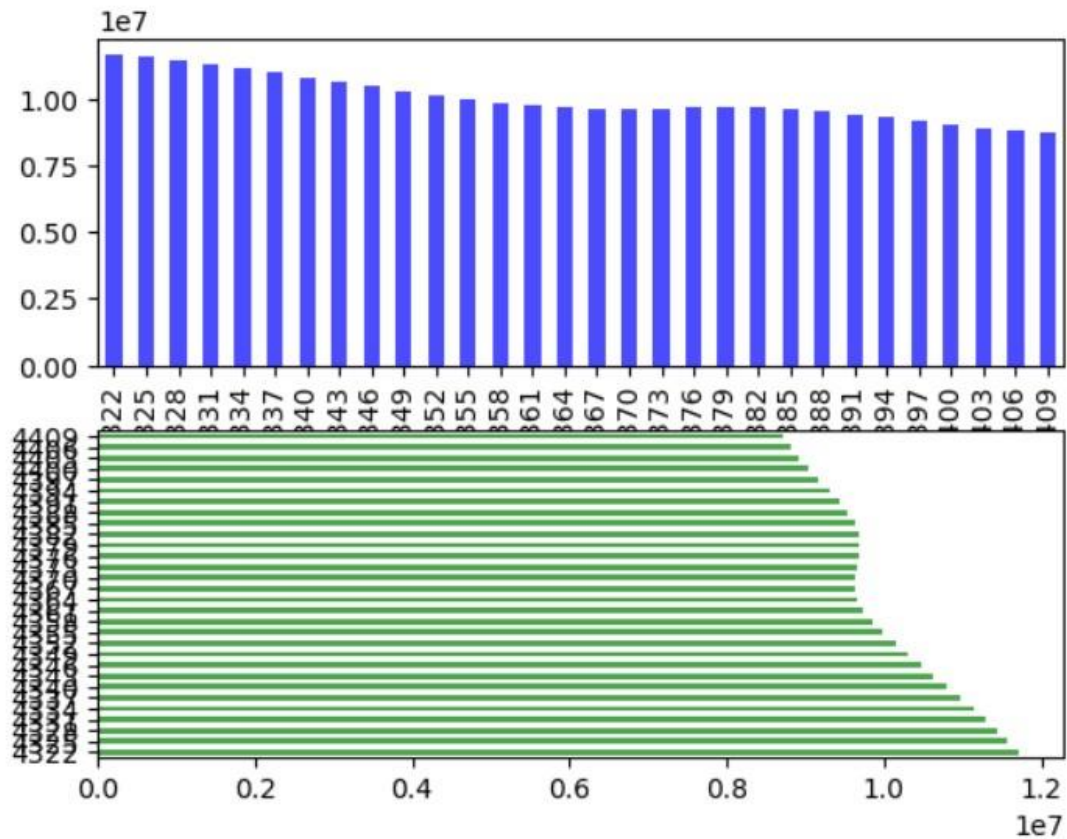
<Axes: >



7 - wykresy kolumnowe

```
fig, axes = plt.subplots(2, 1)
data = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]
data = pd.Series(liniowy_series["val"])
data.plot.bar(ax=axes[0], color='b', alpha=0.7)
data.plot.barh(ax=axes[1], color='g', alpha=0.7)
```

<Axes: >

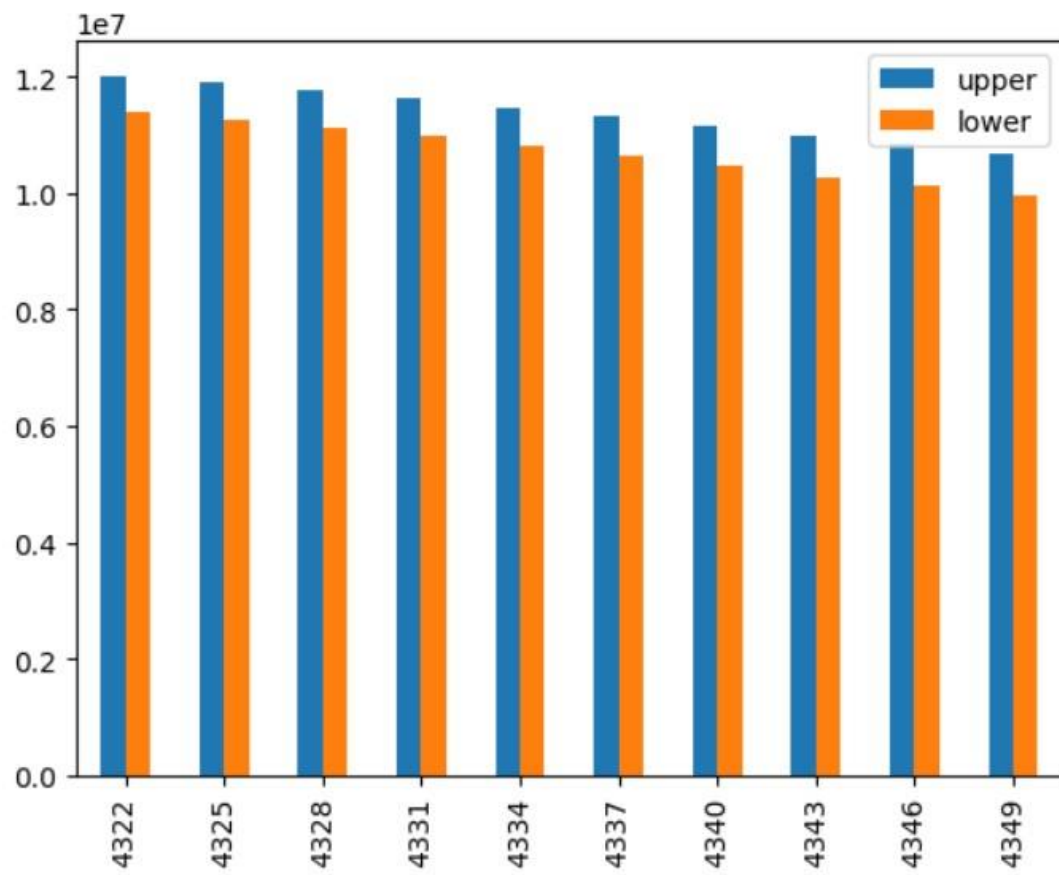


```
w_kolumnowy = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]
w_kolumnowy = w_kolumnowy[["upper", "lower"]]
w_kolumnowy = w_kolumnowy.head(10)
w_kolumnowy
```

	upper	lower
4322	12012219.84	1.138875e+07
4325	11885703.34	1.125896e+07
4328	11758271.45	1.112616e+07
4331	11629281.89	1.097590e+07
4334	11470659.42	1.080890e+07
4337	11324575.92	1.062839e+07
4340	11158004.92	1.045887e+07
4343	10983286.74	1.027121e+07
4346	10830307.69	1.012083e+07
4349	10684277.63	9.960286e+06

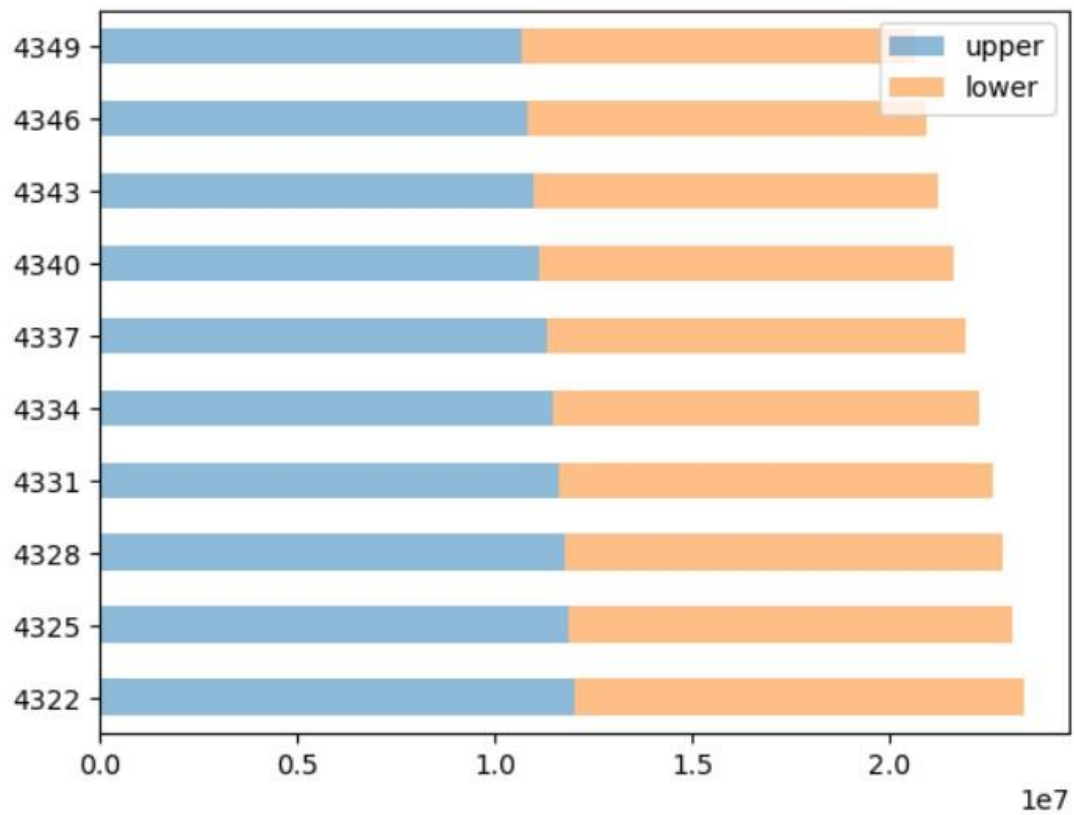
```
w_kolumnowy.plot.bar()
```

<Axes: >



```
: w_kolumnowy.plot.barh(stacked=True, alpha=0.5)
```

```
: <Axes: >
```



8 - seaborn

```
import seaborn as sns
```

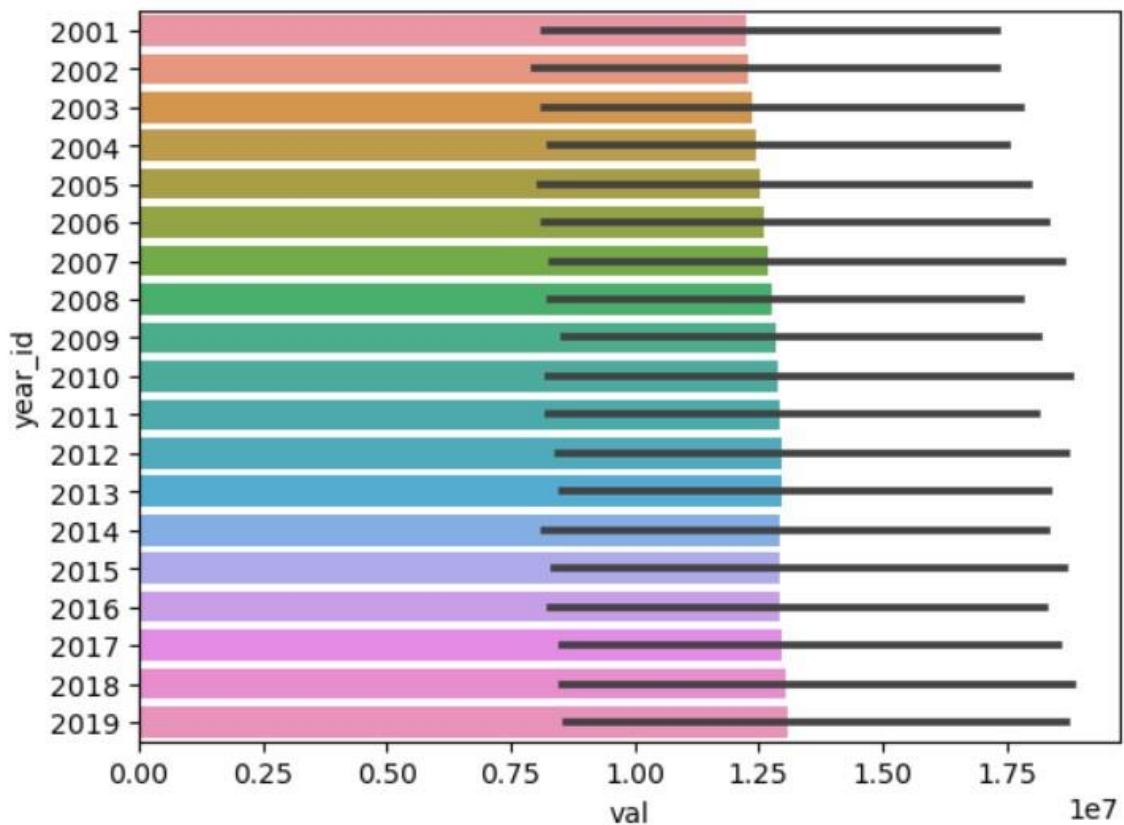
```
sb = df  
sb = sb[sb["year_id"] > 2000]  
sb
```

	measure_name	location_id	location_name	sex_id	sex_name	age_group_id	age_group_name	year_id	val	upper	lower
33	Number of Smokers	1	Global	1	Male	29	15+ years	2001	8.653357e+08	8.709454e+08	8.597397e+08
34	Number of Smokers	1	Global	2	Female	29	15+ years	2001	2.043228e+08	2.078411e+08	2.011226e+08
35	Number of Smokers	1	Global	3	Both	29	15+ years	2001	1.069659e+09	1.076377e+09	1.062829e+09
36	Number of Smokers	1	Global	1	Male	29	15+ years	2002	8.697218e+08	8.753433e+08	8.642284e+08
37	Number of Smokers	1	Global	2	Female	29	15+ years	2002	2.051251e+08	2.086136e+08	2.019433e+08
...
20965	Number of Smokers	522	Sudan	2	Female	29	15+ years	2018	2.435999e+05	3.286166e+05	1.752508e+05
20966	Number of Smokers	522	Sudan	3	Both	29	15+ years	2018	2.610672e+06	2.833943e+06	2.409108e+06
20967	Number of Smokers	522	Sudan	1	Male	29	15+ years	2019	2.439150e+06	2.656579e+06	2.236450e+06
20968	Number of Smokers	522	Sudan	2	Female	29	15+ years	2019	2.500800e+05	3.345384e+05	1.816686e+05
20969	Number of Smokers	522	Sudan	3	Both	29	15+ years	2019	2.689230e+06	2.918332e+06	2.480656e+06

13281 rows × 11 columns

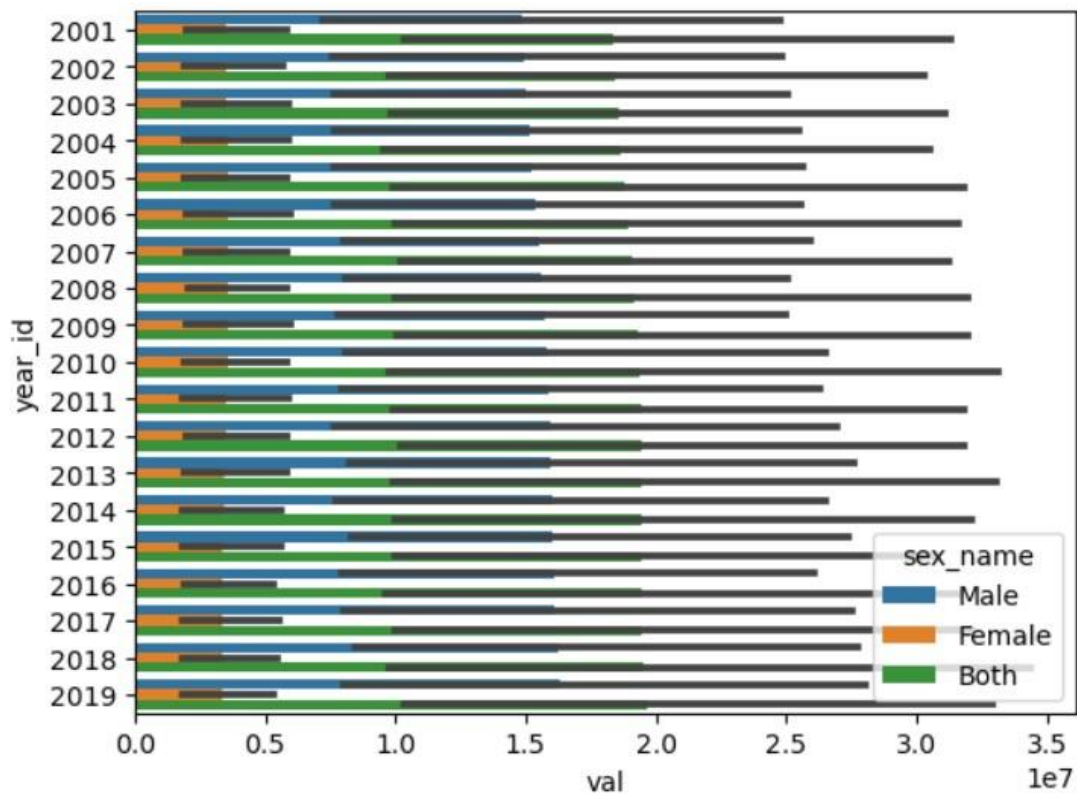
```
sns.barplot(x='val', y='year_id', data=sb, orient='h')
```

```
<Axes: xlabel='val', ylabel='year_id'>
```



```
: sns.barplot(x='val', y='year_id', hue= "sex_name", data=sb, orient='h')
```

```
: <Axes: xlabel='val', ylabel='year_id'>
```



9 - histogramy i wykresy gęstości

```
histogram = df[df["year_id"] > 2000]
histogram = histogram["val"]
histogram
```

```
33      8.653357e+08
34      2.043228e+08
35      1.069659e+09
36      8.697218e+08
37      2.051251e+08
```

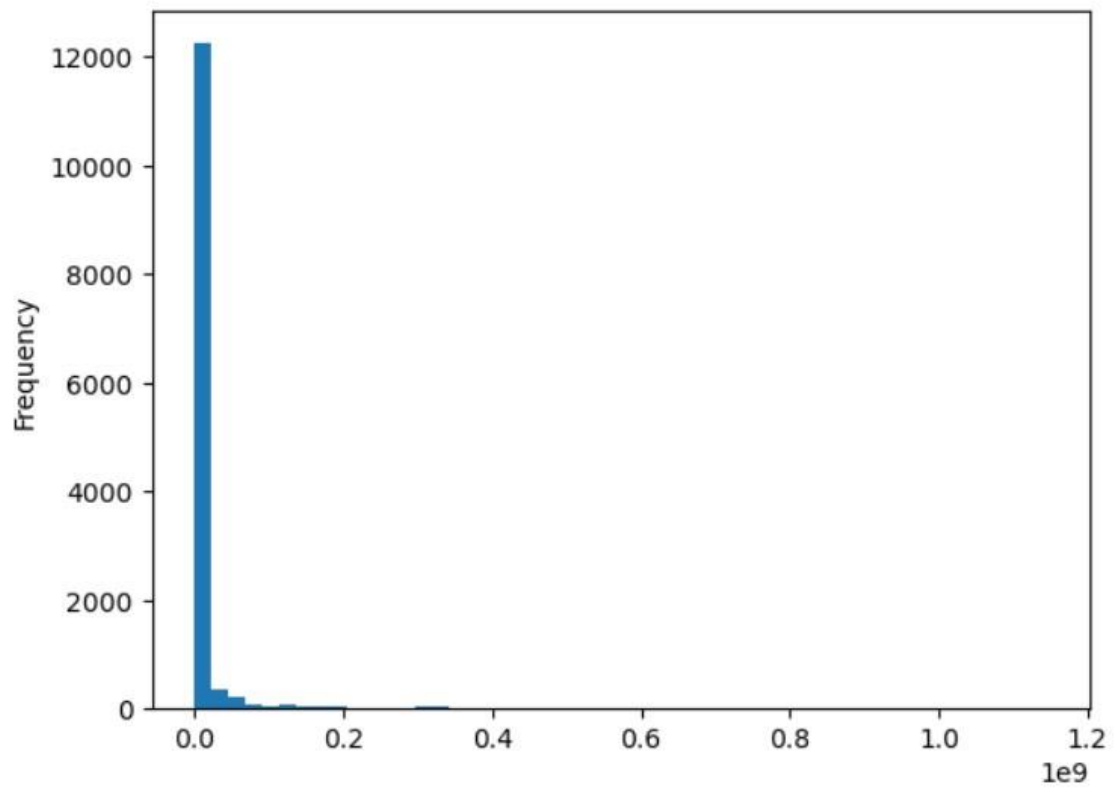
...

```
20965    2.435999e+05
20966    2.610672e+06
20967    2.439150e+06
20968    2.500800e+05
20969    2.689230e+06
```

```
Name: val, Length: 13281, dtype: float64
```

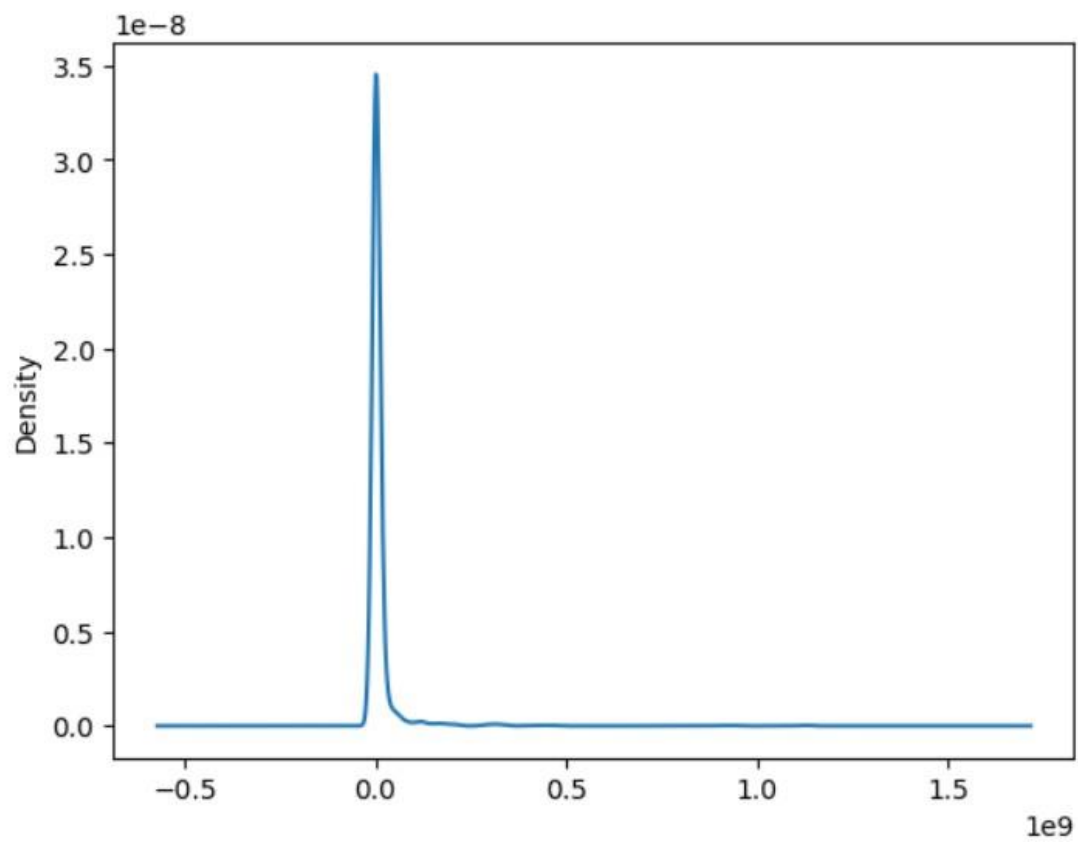
```
: histogram.plot.hist(bins=50)
```

```
: <Axes: ylabel='Frequency'>
```




```
: histogram.plot.kde()
```

```
: <Axes: ylabel='Density'>
```



10 - wykresy punktowe lub wykresy bitowe

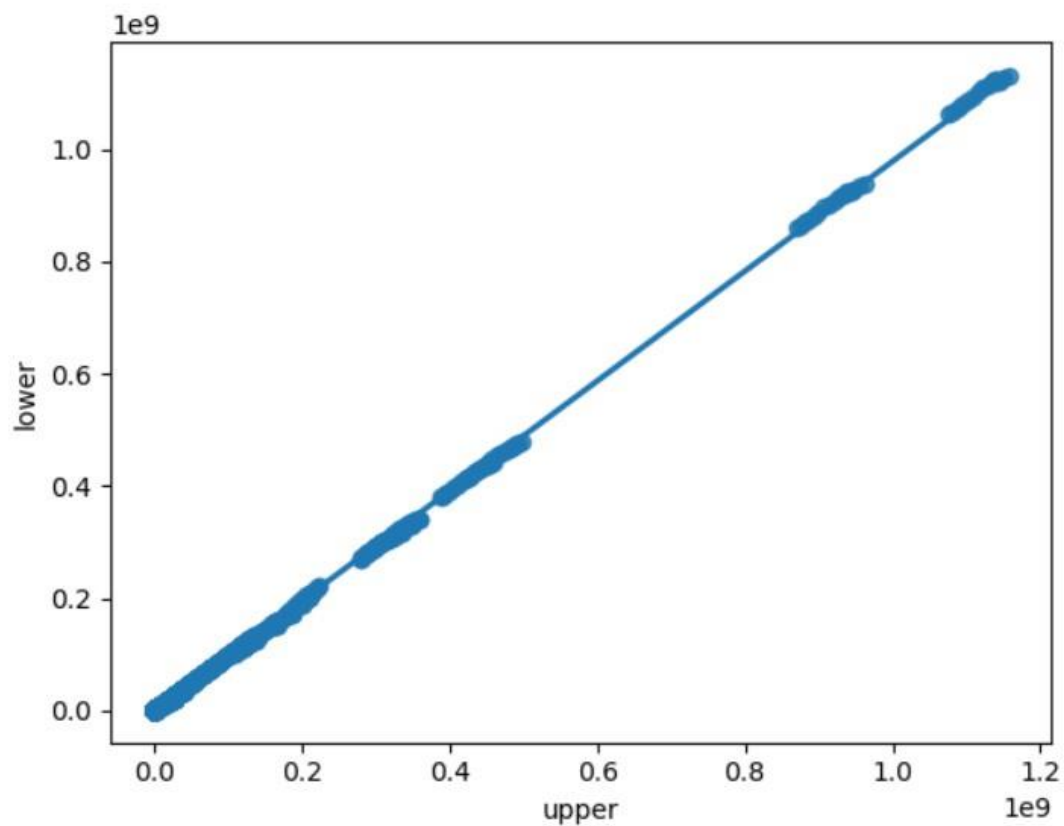
```
punkt = df
punkt = punkt[punkt["year_id"] > 2000]
punkt
```

	measure_name	location_id	location_name	sex_id	sex_name	age_group_id	age_group_name	year_id	val	upper	lower
33	Number of Smokers	1	Global	1	Male	29	15+ years	2001	8.653357e+08	8.709454e+08	8.597397e+08
34	Number of Smokers	1	Global	2	Female	29	15+ years	2001	2.043228e+08	2.078411e+08	2.011226e+08
35	Number of Smokers	1	Global	3	Both	29	15+ years	2001	1.069659e+09	1.076377e+09	1.062829e+09
36	Number of Smokers	1	Global	1	Male	29	15+ years	2002	8.697218e+08	8.753433e+08	8.642284e+08
37	Number of Smokers	1	Global	2	Female	29	15+ years	2002	2.051251e+08	2.086136e+08	2.019433e+08
...
20965	Number of Smokers	522	Sudan	2	Female	29	15+ years	2018	2.435999e+05	3.286166e+05	1.752508e+05
20966	Number of Smokers	522	Sudan	3	Both	29	15+ years	2018	2.610672e+06	2.833943e+06	2.409108e+06
20967	Number of Smokers	522	Sudan	1	Male	29	15+ years	2019	2.439150e+06	2.656579e+06	2.236450e+06
20968	Number of Smokers	522	Sudan	2	Female	29	15+ years	2019	2.500800e+05	3.345384e+05	1.816686e+05
20969	Number of Smokers	522	Sudan	3	Both	29	15+ years	2019	2.689230e+06	2.918332e+06	2.480656e+06

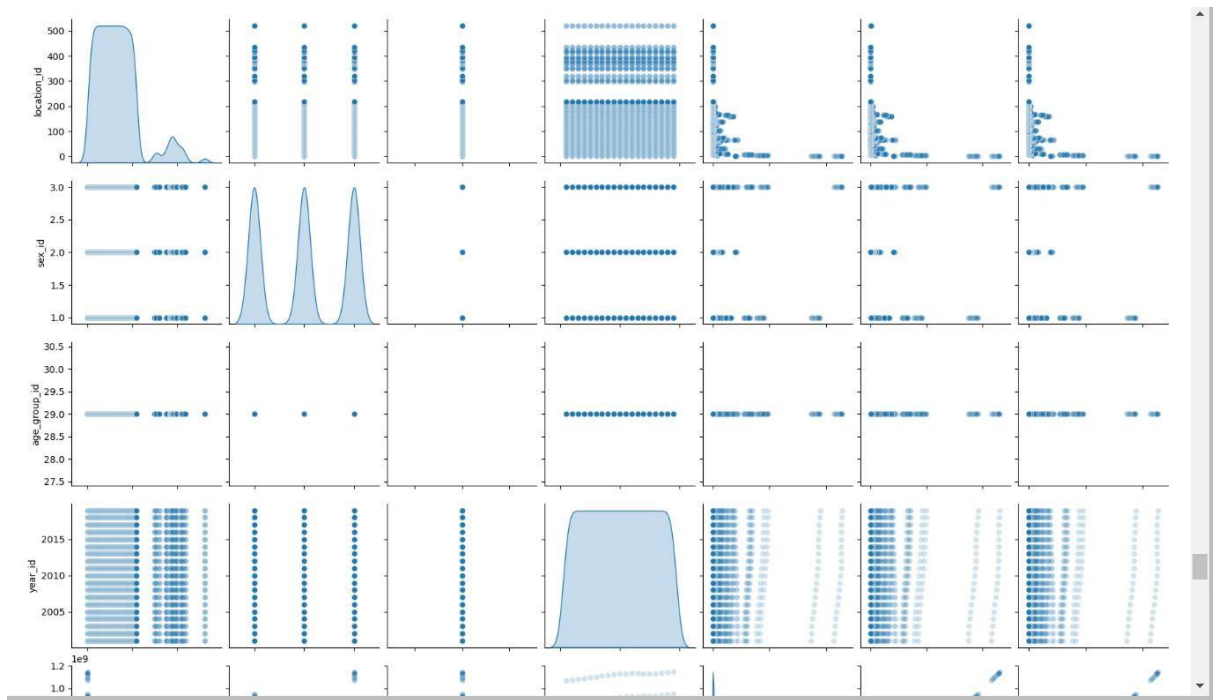
13281 rows × 11 columns

```
: sns.regplot(x='upper', y='lower', data=punkt)
```

```
: <Axes: xlabel='upper', ylabel='lower'>
```



```
: sns.pairplot(punkt, diag_kind='kde', plot_kws={'alpha': 0.2})
```

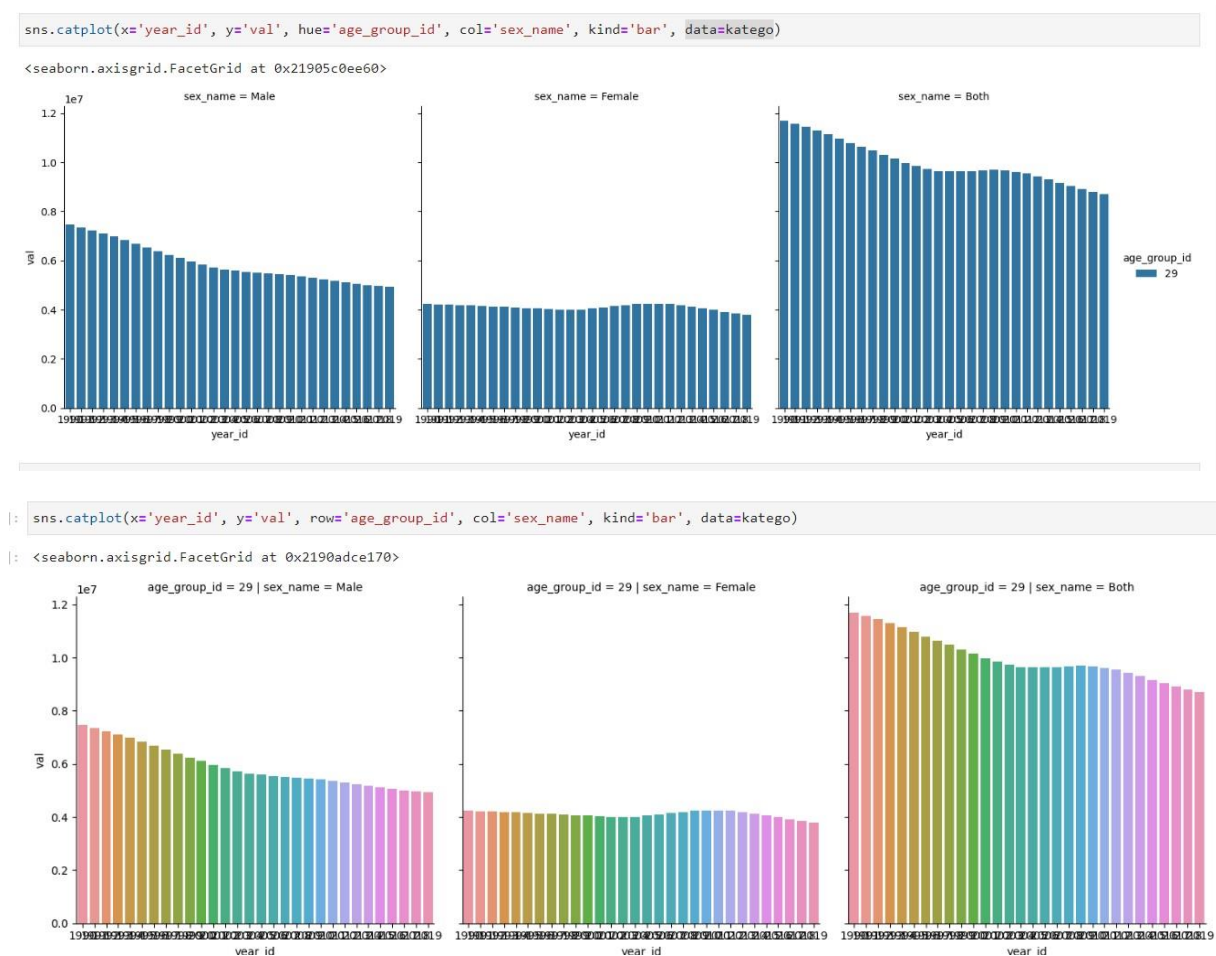


11 - dane kategoryczne

```
katego = df
katego = katego[(katego["location_name"] == "Poland")]
katego
```

	measure_name	location_id	location_name	sex_id	sex_name	age_group_id	age_group_name	year_id	val	upper	lower
4320	Number of Smokers	51	Poland	1	Male	29	15+ years	1990	7.468990e+06	7.679963e+06	7.261637e+06
4321	Number of Smokers	51	Poland	2	Female	29	15+ years	1990	4.235491e+06	4.465004e+06	3.987365e+06
4322	Number of Smokers	51	Poland	3	Both	29	15+ years	1990	1.170448e+07	1.201222e+07	1.138875e+07
4323	Number of Smokers	51	Poland	1	Male	29	15+ years	1991	7.358603e+06	7.564003e+06	7.154446e+06
4324	Number of Smokers	51	Poland	2	Female	29	15+ years	1991	4.217866e+06	4.454244e+06	3.968376e+06
...
4405	Number of Smokers	51	Poland	2	Female	29	15+ years	2018	3.848206e+06	4.116964e+06	3.595334e+06
4406	Number of Smokers	51	Poland	3	Both	29	15+ years	2018	8.811197e+06	9.135481e+06	8.490369e+06
4407	Number of Smokers	51	Poland	1	Male	29	15+ years	2019	4.938503e+06	5.166168e+06	4.737903e+06
4408	Number of Smokers	51	Poland	2	Female	29	15+ years	2019	3.780665e+06	4.072287e+06	3.504634e+06
4409	Number of Smokers	51	Poland	3	Both	29	15+ years	2019	8.719168e+06	9.075511e+06	8.370563e+06

90 rows × 11 columns



3. Wnioski

Na podstawie otrzymanego wyniku można stwierdzić, że biblioteka Matplotlib umożliwia tworzyć wiele wizualnych wykresów. Biblioteka Pandas pozwala uzyskać wiele kolumn, oraz etykiety wierszy i kolumn. Do tworzenia wykresów statycznych umożliwia biblioteka Seaborn.