

zero-semester

July 7, 2023

0.0.1 TASK 1

Write a Python program to calculate the area of a rectangle given its length and width.

```
[29]: length = 10
      width = 10
      area = length * width
      print("area: ", area)
```

area: 100

0.0.2 TASK 2

Write a Python program to create a list of numbers, calculate the sum of the list, and find the maximum and minimum values.

```
[30]: l = [1,4,3,-5,0,10,-20,40]
```

```
[31]: print(sum(l))
```

33

```
[32]: print(max(l))
```

40

```
[33]: print(min(l))
```

-20

0.0.3 TASK 3

Load a CSV file using Pandas, identify and fill missing values, remove duplicates, and plot a histogram of a numeric column.

```
[34]: import pandas as pd
```

```
[35]: # LOAD CSV FILE
      data = pd.read_csv("Sample.csv", encoding="cp852", delimiter=",", index_col=0)
      df = pd.DataFrame(data=data)
```

```
# IDENTIFY MISSING VALUES
for index, row in df.iterrows():
    if pd.isna(row['Company']):
        print("EMPTY")
    else:
        print(row['Company'])

# IDENTIFY MISSING VALUES OTHER WAYS
df.isna()
df.isnull()
```

```
EMPTY
Cardinal Slant-D« Ring Binder, Heavy Gauge Vinyl
Cardinal Slant-D« Ring Binder, Heavy Gauge Vinyl
EMPTY
G.E. Longer-Life Indoor Recessed Floodlight Bulbs
Angle-D Binders with Locking Rings, Label Holders
SAFCO Mobile Desk Side File, Wire Frame
EMPTY
Xerox 198
```

```
[35]:      Company   Name    age
      id
2      True  False  False
3      False False  False
3      False False  False
5      True  False   True
6      False False  False
7      False  True   True
8      False False  False
9      True  False  False
10     False  True   True
```

```
[36]: # FILL MISSING VALUES
df.fillna("FILLED")
# OR
df.fillna(method='pad')
```

```
[36]:      Company      Name    age
      id
2      NaN      Barry French  12.0
3  Cardinal Slant-D« Ring Binder, Heavy Gauge Vinyl  Barry French  34.0
3  Cardinal Slant-D« Ring Binder, Heavy Gauge Vinyl  Barry French  34.0
5  Cardinal Slant-D« Ring Binder, Heavy Gauge Vinyl  Carlos Soltero  34.0
6  G.E. Longer-Life Indoor Recessed Floodlight Bulbs  Carlos Soltero  65.0
7  Angle-D Binders with Locking Rings, Label Holders  Carlos Soltero  65.0
```

8	SAFCO Mobile Desk Side File, Wire Frame	Carl Jackson	47.0
9	SAFCO Mobile Desk Side File, Wire Frame	Monica Federle	54.0
10	Xerox 198	Monica Federle	54.0

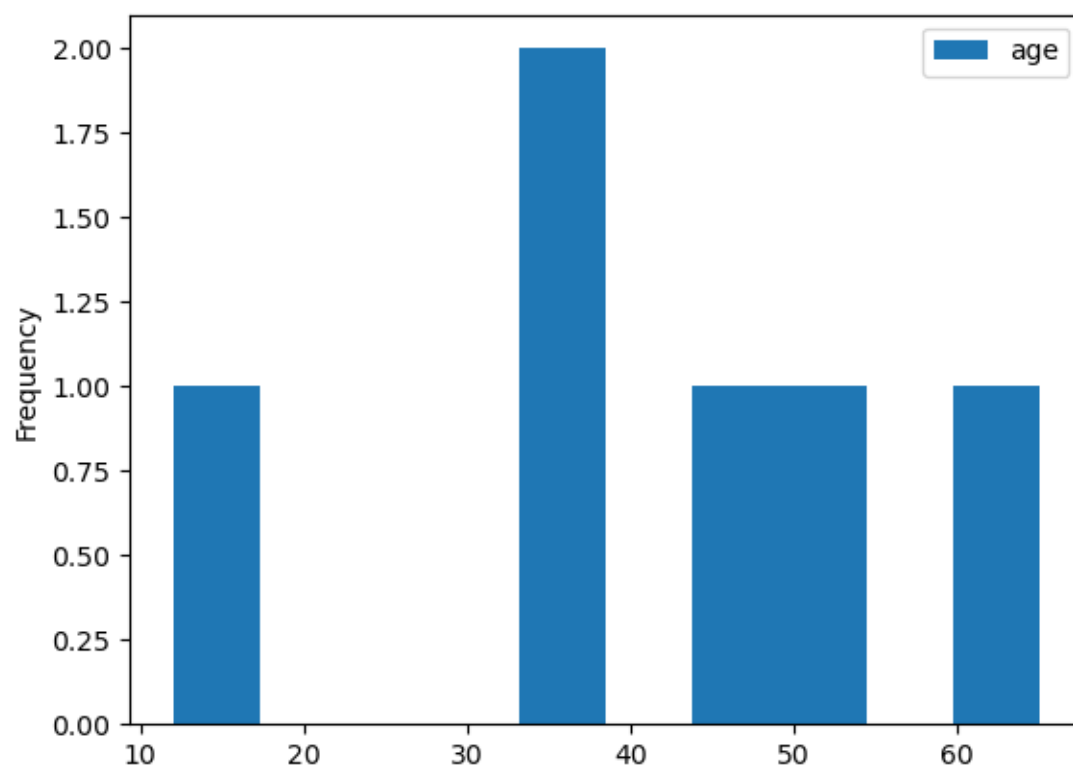
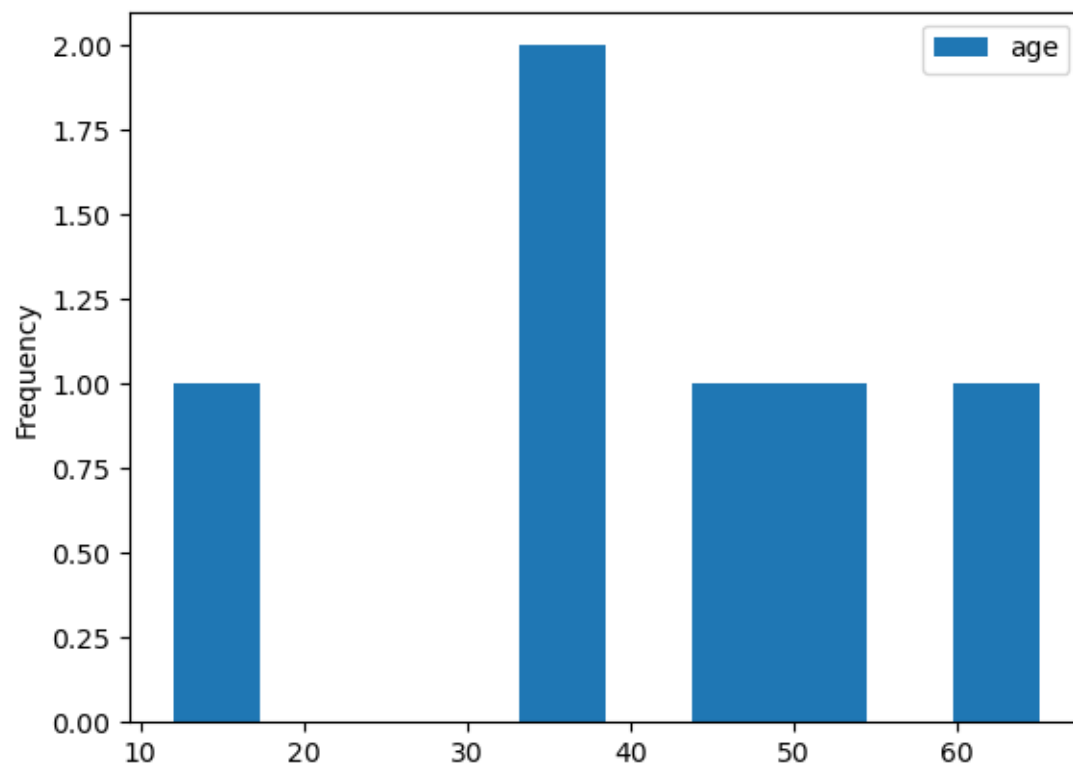
```
[37]: # REMOVE DUPLICATES
new_df = df.drop_duplicates()
new_df
```

```
[37]:
```

	Company	Name	age
id			
2	NaN	Barry French	12.0
3	Cardinal Slant-D« Ring Binder, Heavy Gauge Vinyl	Barry French	34.0
5	NaN	Carlos Soltero	NaN
6	G.E. Longer-Life Indoor Recessed Floodlight Bulbs	Carlos Soltero	65.0
7	Angle-D Binders with Locking Rings, Label Holders	NaN	NaN
8	SAFCO Mobile Desk Side File, Wire Frame	Carl Jackson	47.0
9	NaN	Monica Federle	54.0
10	Xerox 198	NaN	NaN

```
[38]: # PLOT HISTOGRAM
df.plot.hist()
# OR
df.plot.hist(column=['age'])
```

```
[38]: <Axes: ylabel='Frequency'>
```



0.0.4 TASK 4

Analyze a dataset using Pandas, calculate the mean and median of a numerical column, and find the mode of a categorical column.

```
[39]: data = {'name': ["Johar", "Jhon", "Jake", "Joe"],
            'age': [45, 90, 25, 75],
            'job': ['teacher', 'engineer', 'teacher', 'doctor']}

df = pd.DataFrame(data)
median = df['age'].median()
mean = df['age'].mean()
mode = df['job'].mode()

print("Median: {}\nMean: {}\nMode: \n{}".format(median, mean, mode))
```

```
Median: 60.0
Mean: 58.75
Mode:
0    teacher
Name: job, dtype: object
```

0.0.5 TASK 5

Plot a line chart showing the trend of a numerical variable over time and create a bar chart to compare the values of different categories.

```
[40]: pip install matplotlib
```

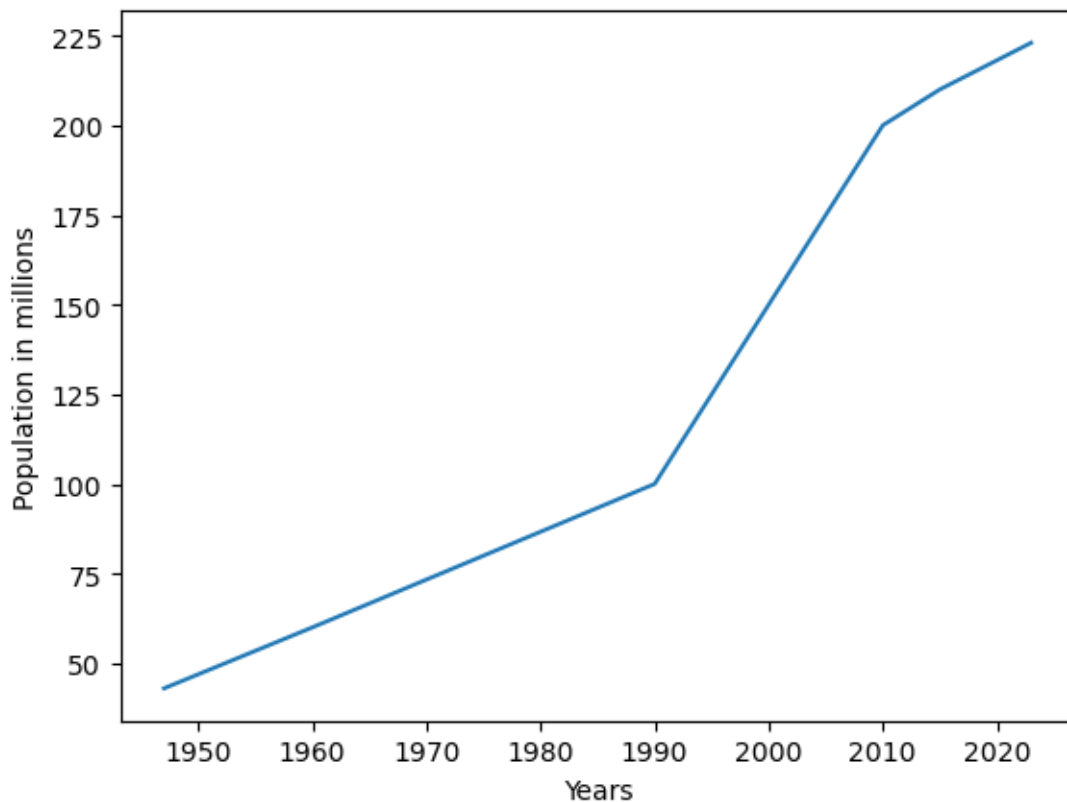
```
Requirement already satisfied: matplotlib in d:\apps\anaconda\files\lib\site-
packages (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in d:\apps\anaconda\files\lib\site-
packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in d:\apps\anaconda\files\lib\site-
packages (from matplotlib) (1.23.5)
Requirement already satisfied: packaging>=20.0 in
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (23.0)
Requirement already satisfied: pillow>=6.2.0 in d:\apps\anaconda\files\lib\site-
packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
```

```
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: importlib-resources>=3.2.0 in
d:\apps\anaconda\files\lib\site-packages (from matplotlib) (5.2.0)
Requirement already satisfied: zipp>=3.1.0 in d:\apps\anaconda\files\lib\site-
packages (from importlib-resources>=3.2.0->matplotlib) (3.11.0)
Requirement already satisfied: six>=1.5 in d:\apps\anaconda\files\lib\site-
packages (from python-dateutil>=2.7->matplotlib) (1.12.0)
Note: you may need to restart the kernel to use updated packages.
```

```
[41]: import matplotlib.pyplot as plt
import numpy as np

years = [1947, 1960, 1990, 2000, 2010, 2015, 2023]
population = [43, 60, 100, 150, 200, 210, 223]

plt.plot(years, population)
plt.xlabel("Years")
plt.ylabel("Population in millions")
plt.show()
```



```
[42]: animals = ['cat', 'lion', 'jaguar', 'crane']
total = [12, 30, 5, 50]

plt.bar(animals, total, color='orange', width=0.5, edgecolor="green"### TASK 5

Plot a line chart showing the trend of a numerical variable over time and
↳ create a bar chart to compare the values of different categories.### TASK 5

Plot a line chart showing the trend of a numerical variable over time and
↳ create a bar chart to compare the values of different categories.)
plt.xlabel("animal")
plt.ylabel("total")
plt.show()
```

Cell In[42], line 6

Plot a line chart showing the trend of a numerical variable over time and
↳ create a bar chart to compare the values of different categories.### TASK 5

SyntaxError: invalid syntax

0.0.6 TASK 6

Create a NumPy array, perform basic arithmetic operations, and calculate the mean and standard deviation of the array.

```
[43]: import numpy as np
arr1 = np.array([20, 20, 4, 1, 0, 20])
arr2 = np.array([12, 3, 9, 10, 3, 25])

print(arr1+arr2)
print(np.add(arr1, arr2),end="\n\n")

print(arr1-arr2)
print(np.subtract(arr1, arr2),end="\n\n")

print(arr1*arr2)
print(np.multiply(arr1, arr2),end="\n\n")

print(arr1/arr2)
print(np.divide(arr1, arr2),end="\n\n")

print(arr1**2)
print(np.power(arr1, 2),end="\n\n")

print(arr1%arr2)
```

```

print(np.mod(arr1, arr2),end="\n\n")

# mean and standard deviation
print(np.argmax(arr1),end="\n\n")
print(np.argmax(arr2),end="\n\n")

counts = np.bincount(arr1)
mode = np.argmax(counts)
print(mode)

```

[32 23 13 11 3 45]

[32 23 13 11 3 45]

[8 17 -5 -9 -3 -5]

[8 17 -5 -9 -3 -5]

[240 60 36 10 0 500]

[240 60 36 10 0 500]

[1.66666667 6.66666667 0.44444444 0.1 0. 0.8]

[1.66666667 6.66666667 0.44444444 0.1 0. 0.8]

[400 400 16 1 0 400]

[400 400 16 1 0 400]

[8 2 4 1 0 20]

[8 2 4 1 0 20]

0

5

20