Question 1 in HW0

I’m aiming to gain a strong command of Python coding, particularly as it pertains to processing large datasets in this course. I am also curious about the concepts, algorithms, and especially the applications of machine learning in various industries, including finance. At the start of this course, I need to establish a clear framework of the mathematics and statistics knowledge required so that I can be well-prepared.

homework0_pa_template

September 1, 2024

0.1 HW0: Introduction to CSE 6740

0.2 Programming Assignment Template

Please fill in your name and GTID below

Name: Yuanting Fan

GTID: 904047984 Section 1:

[6]:
```
# Import necessary libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Section 1: Basic Python and Numpy Exercises

# Exercise 1: Create a 1D numpy array with values from 0 to 9
array_1d = np.arange(10)
print("1D Array:", array_1d)

# Exercise 2: Reshape the array to 2x5
array_2d = array_1d.reshape(2, 5)
print("Reshaped to 2x5 Array:\n", array_2d)
```

```
1D Array: [0 1 2 3 4 5 6 7 8 9]
Reshaped to 2x5 Array:
 [[0 1 2 3 4]
 [5 6 7 8 9]]
```

Section 2:

[7]:
```
# Section 2: Basic DataFrame Operations with Pandas

# Exercise 3: Create a simple DataFrame with random numbers
df = pd.DataFrame(np.random.rand(4, 3), columns=['A', 'B', 'C'])
print("DataFrame:\n", df)

# Exercise 4: Add a new column 'D' which is the sum of 'A' and 'B'
df['D'] = df['A'] + df['B']
```

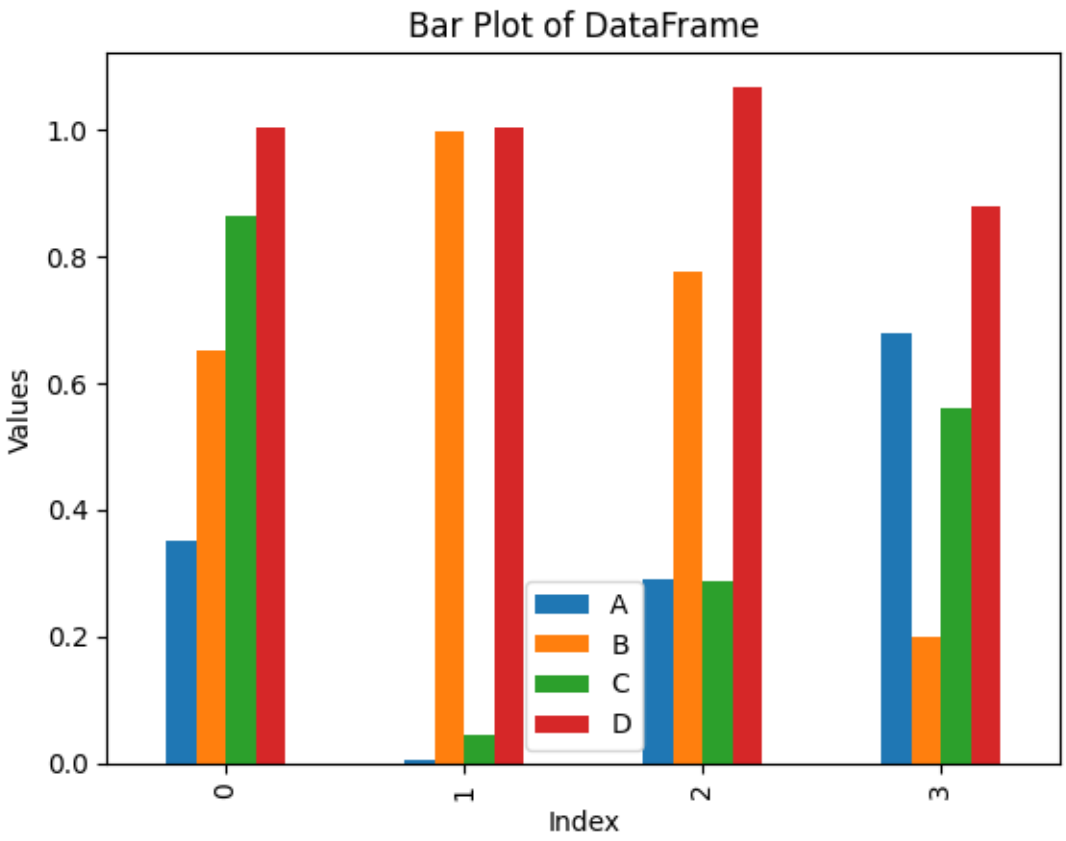
```
print("DataFrame with new column 'D':\n", df)
```

```
DataFrame:
          A         B         C
0  0.350754  0.652420  0.864793
1  0.005473  0.996719  0.045342
2  0.290593  0.776563  0.286791
3  0.678686  0.200899  0.560675
DataFrame with new column 'D':
          A         B         C         D
0  0.350754  0.652420  0.864793  1.003174
1  0.005473  0.996719  0.045342  1.002192
2  0.290593  0.776563  0.286791  1.067157
3  0.678686  0.200899  0.560675  0.879585
```

Section 3

[8]:
```
# Section 3: Visualization with Matplotlib

# Exercise 5: Plot the DataFrame
df.plot(kind='bar')
plt.title('Bar Plot of DataFrame')
plt.xlabel('Index')
plt.ylabel('Values')
plt.show()
```


Bar Plot of DataFrame
Values
1.0
0.8
0.6
0.4
0.2
0.0
A
B
C
D
0
1
2
3
Index

[ ]: