Math 446, Fall 2024, USC Instructor: Steven Heilman

Name: _____ USC ID: _____ Date: _____

Signature: _____. Discussion Section: _____
(By signing here, I certify that I have taken this test while refraining from cheating.)

Exam 1

This exam contains 6 pages (including this cover page) and 4 problems. Enter all requested information on the top of this page.

You may *not* use your books, notes, or any calculator on this exam.

You are required to show your work on each problem on this exam. The following rules apply:

- You have 50 minutes to complete the exam, starting at the beginning of class.
- **Organize your work**, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.
- **Mysterious or unsupported answers will not receive full credit**. A correct answer, unsupported by calculations, explanation, or algebraic work will receive no credit; an incorrect answer supported by substantially correct calculations and explanations might still receive partial credit.
- If you need more space, use the back of the pages; clearly indicate when you have done this. Scratch paper appears at the end of the document.

Do not write in the table to the right. Good luck![a]

Problem	Points	Score
1	10	
2	10	
3	10	
4	10	
Total:	40	

[a] November 6, 2024, © 2024 Steven Heilman, All Rights Reserved.

1. Suppose we run the commands

```
import pandas as pd
data = {
    "state": ["Ohio", "Ohio", "Ohio", "Nevada", "Nevada", "Nevada"],
    "year": [2000, 2001, 2002, 2001, 2002, 2003],
    "pop": [1.5, 1.7, 3.6, 2.4, 2.9, 3.2]
}
frame = pd.DataFrame(data)
```

   (a) (5 points) What is the output of the following commands?

```
frame2 = frame.reindex(index = [3, 2, 5])
frame2
```

   (b) (5 points) What is the output of the following commands?

```
frame3 = frame2.set_index("year")
frame3
```

2. (10 points) What is the output of the following program? Explain your reasoning.

```
import re
data = '''
"data-testid="bar-chart--results-bar" style="width:51%"
role="progressbar" aria-valuenow="51" class="jsx-4201391551
jsx-842384122 labeled-bar df white"><span data-testid=
"bar-chart--results-bar-percent" class="jsx-4201391551 jsx-842384122"
'''
search_string = r'jsx([\w-]{3})'

found_strings = re.findall(search_string, data)
found_strings
```

3. (10 points) Suppose we have a list of strings of the following form

```
strings = ['Blue Horse', 'Purple Cat', 'White Dog', 'Yellow Duck']
```

Write a Python function that removes the spaces from this list. That is, the output should be

```
strings = ['BlueHorse', 'PurpleCat', 'WhiteDog', 'YellowDuck']
```

4. (10 points) Suppose we have a Pandas DataFrame named `df` with the following entries

product_id	product_name	units_sold	unit_price	sale_date	region
4	widget_a	150	2.5	2023-01-10	east
3	widget_b	200	3.0	2023-01-12	east
2	widget_c	250	1.5	2023-01-14	west
1	widget_d	300	4.0	2023-01-10	south
0	widget_e	100	5.0	2023-01-15	east

That is, the `index` of `df` is named `product_id`, so the command `df.index` returns `Index([4, 3, 2, 1, 0], dtype='int64', name='product_id')`

Answer the following questions.

- What is the output of `df.drop(index = [4, 1, 0])`
- What is the output of `df.reindex(np.arange(6), method = "ffill")`
- What is the output of `df[2]["units_sold"]` ?
- Write a single line of Python code that returns a DataFrame containing only the rows of `df` where sales occurred in the `east` region.
- Write a single line of Python code to compute the total sales for each row of `df` (i.e. compute `units_sold` multiplied by `unit_price`) and create a new column of `df` called `total_sales` that contains the total sales of each row of `df`.

(Scratch paper)