Data 6 Summer 2023

Final Exam Review

Week 6 Tutoring Worksheet

August 2023

1. Arithmetic, Arrays, and Functions

Evaluate the following code cells just as Python would in a Jupyter notebook. If evaluating the expression causes an error, write "Error".

a. `1 + 1.0`

b. `not 3`

c.
```
a = make_array(1, 2, 5)
b = a * 2
c = a.item(-1)
b - a + c
```

d. `int("3.14")`

e. `make_array(3, 6) + np.arange(2, 8)`

f. `4 > False`

g. `min(sum(make_array(11, 13, 15)), sum(np.arange(12, 15, 2) - 1))`

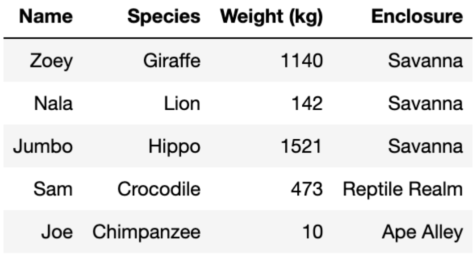
h. `False and 1 / 0`

i. `len(np.arange(10))`

j. `min(max(min(abs(-5), -1), 0)), 2`

2. Table Methods

The Berkeley Zoo needs your help caring for its dozens of animals. The table `animals` below lists each animal currently at the Zoo.



Name	Species	Weight (kg)	Enclosure
Zoey	Giraffe	1140	Savanna
Nala	Lion	142	Savanna
Jumbo	Hippo	1521	Savanna
Sam	Crocodile	473	Reptile Realm
Joe	Chimpanzee	10	Ape Alley

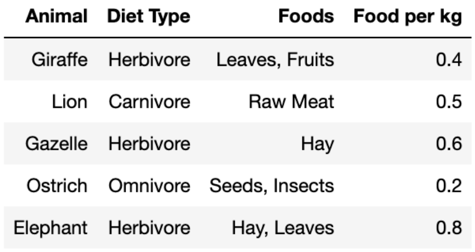
```
(... 72 rows omitted)
```

a. A young Zoo visitor needs your help finding her favorite animal, "Alejandro" (an alpaca).

   i. Write an expression that evaluates to the name of the enclosure where the visitor can find Alejandro.

   ii. The visitor is also curious about the heaviest and lightest animals in the park. Write an expression to find the species of the five heaviest animals in the `animals` table.

```
diets
```



Animal	Diet Type	Foods	Food per kg
Giraffe	Herbivore	Leaves, Fruits	0.4
Lion	Carnivore	Raw Meat	0.5
Gazelle	Herbivore	Hay	0.6
Ostrich	Omnivore	Seeds, Insects	0.2
Elephant	Herbivore	Hay, Leaves	0.8

```
(... 13 rows omitted)
```

b. Zookeeper Timothy is in charge of feeding the animals in the Savanna. However, he's forgotten what he's supposed to feed each animal.

   i. To help him, create a table of Savanna animals and their diets by joining the `animals` and `diets` tables. Call this table `savanna_animals`.

   ii. Create a new column in the `savanna_animals` table called "Daily Nutrition", which is the total weight (in kilograms) of food each animal should be given each day. Calculate this weight by multiplying the animal's weight by their "Food per kg" ratio.

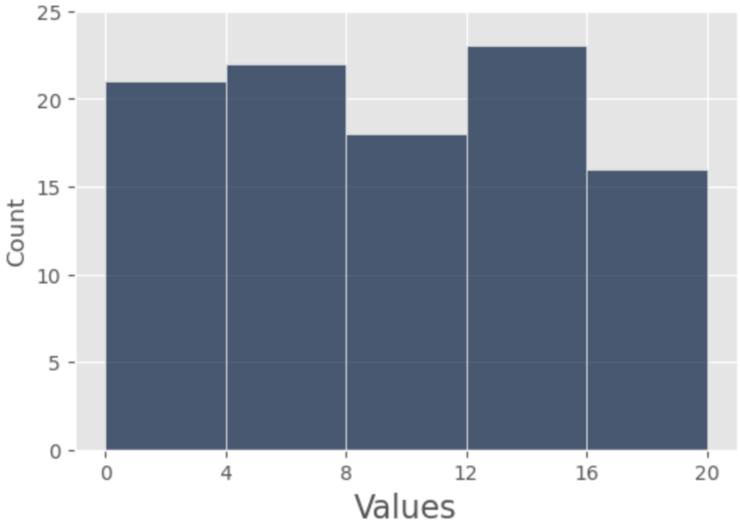
c. Zoo manager Amanda wants to find some statistics about the animals that she can publish on the Berkeley Zoo website.

   i. Write an expression to create a table with the average weight for each species of animal at the Zoo.

   ii. The table `all_animal_diets` contains the same information as `savanna_animals`, except for all animals across all enclosures. Using this table, create a new table with each type of diet (herbivore, carnivore, omnivore) as its own column and each unique enclosure as its own row. The table should show the number of animals of each type of diet in each enclosure.

   iii. **BONUS:** Instead of finding the count of each combination of diet type and enclosure, find the most common food using the `most_common` function. (This function has already been defined and its details are not important)

3. Visualizations



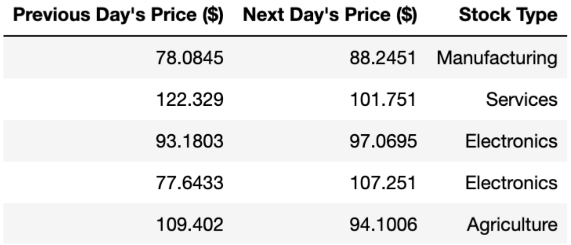
a. Given the histogram of values above, answer the following questions:

   i. **True or False?** The bin from 12 to 16 contains values equalling 16.

   ii. **True or False**? The bins `[0, 4)` and `[12, 16)` combined have more values than there are in the range `[4, 12)`

   iii. Within the `[4, 8)` bin, what percentage of values are less than 6?

`stonks`



Previous Day's Price ($)	Next Day's Price ($)	Stock Type
78.0845	88.2451	Manufacturing
122.329	101.751	Services
93.1803	97.0695	Electronics
77.6433	107.251	Electronics
109.402	94.1006	Agriculture

`(...  495 rows omitted)`

b. The table `stonks` above shows various stocks and their prices over two days.

i. What visualization type is most appropriate for visualizing the relationship between the previous day's stock price and the next day's stock price for each stock?

ii. How many variables can be encoded in a graph using only the information in the `stonks` table?

iii. Write an expression to generate an overlaid histogram visualizing the distribution of next day prices by stock type.

4. Control and Iteration

a. Convert the for-loop below into a while loop with the same functionality.

```
for i in np.arange(1, 6):
    if i % 2 == 0:
        print("Happy!")
    else:
        print("Sad")
```

b. In Group E of the 2023 Women’s World Cup, the Netherlands and United States advanced out of the group while Portugal and Vietnam did not. For each of the countries in Group E, print “Advanced” if they advanced and “Did Not Advance” if they didn’t.

```
countries = make_array("Portugal", "United States", "Vietnam", "Netherlands")

for ____________________:

        if ____________________________________:
            print("Advanced")
        else:
            print("Did Not Advance")
```

5. Random Sampling

Randomness is an everyday phenomenon that has given rise to its own academic discipline of *Statistics*. In Python, the `np.random` submodule provides useful tools for working with randomness, most having to do with **random sampling** (i.e. select an item at random from a collection of multiple items).

a. Write a line of code to generate a random number between 1 (inclusive) and 100 (inclusive).

b. Is the number generated in part (a) *truly* random? Why or why not?

c. When the code below is run multiple times (e.g. in a Jupyter notebook cell), will the output change? Why or why not?

```
np.random.seed(12)
np.random.randint(1, 11)
```

d. Andy has a collection of four playing cards to choose from: “Ace”, “Jack”, “Queen”, “King”.

   i. Write an expression to choose a random card (from the four options) four times.

   ii. Is Andy guaranteed to draw a “King” at least one time during these four draws?

   iii. Andy now draws a card at random from the four available cards 800 times. On average, how many times should we *expect* Andy to draw the “Jack”?

6. Dictionaries

Dictionaries are a way to store pairs of **keys** and **values**, like a student ID number (the key), and the student's schedule of classes (the value). Keys must be a simple data type like integers or strings, but the values can be any data type (even more dictionaries!). Unlike arrays, dictionary **key-value pairs don't have an inherent ordering**, and you can't find a particular value using an index.

The dictionary `pets` contains the names of pets for various owners.

```
pets = {
    ‘Rick’: ‘Fuzzball’,
    ‘Sally’: ‘Sparkie’,
    ‘Shawn’: [‘Bing’, ‘Bong’],
    ‘Leanne’: {‘Luke’: ‘Henry’, ‘Lisa’: ‘Hannah’}
}
```

a. Add James and his dog Lego to the `pets` dictionary.

b. Write an expression to find the name of Leanne's daughter Lisa's pet.

c. Use a for loop to iterate through all keys in the `pets` dictionary. For each owner, if their pet's name contains the letter ‘e’, print “Excellent!”. If their pet's name contains the letter ‘o’, print “Outstanding!”. For now, you don't have to worry about owners with multiple pets.

d. **BONUS:** Describe how you would complete the task described in (c), including owners with multiple pets.