Midterm Exam - DSC 10, Spring 2025

Full Name:

PID:

Exam Time: ◯ 9AM ◯ 11AM

Instructions:

- This exam consists of 7 questions, worth a total of 77 points.
- Write your PID in the top right corner of each page in the space provided.
- Please write **clearly** in the provided answer boxes; we will not grade work that appears elsewhere. Completely fill in bubbles and square boxes; if we cannot tell which option(s) you selected, you may lose points.

  ◯ A bubble means that you should only **select one choice**.

  ☐ A square box means you should **select all that apply**.
- For full credit, your solutions must use methods of the course.
- You may use one page of double-sided handwritten notes. Aside from this, you may not refer to any other resources or technology during the exam. No calculators!

By signing below, you are agreeing that you will behave honestly and fairly during and after this exam.

Signature:

Version A

Please do not open your exam until instructed to do so.

Important: Before proceeding, make sure to rip off the last page of this exam packet and read the data description.

Question 1 (8 pts)

a) (4 pts) Which country has the highest `"Reciprocal Tariff"`? Write one line of code that evaluates to the name of this country.

b) (4 pts) How many countries have a `"Reciprocal Tariff"` above 30%? Write one line of code that evaluates to the number of such countries.

PID: _____

Question 2 (10 pts)

In `tariffs`, we use integers to represent percentages, but we could also use strings with the percent symbol %. For example, the integer `34` and the string `"34%"` both represent the same thing.

a) (4 pts) Fill in the functions `with_percent_symbol` and `without_percent_symbol` below. The function `with_percent_symbol` should take as input an integer and give as output a string with the percent symbol. The function `without_percent_symbol` should do the opposite. Example behavior is given below.

```
with_percent_symbol(34)
>>> '34%'

def with_percent_symbol(x):
    return __(a)__
```

(a): [ ]

```
without_percent_symbol("34%")
>>> 34

def without_percent_symbol(x):
    return __(b)__
```

(b): [ ]

b) (3 pts) Define the variable `y` as follows.

```
y = tariffs.get("Reciprocal Tariff").apply(with_percent_symbol)
```

Below, define `z` so that it evaluates to exactly the same Series as `y`. You may **not** use `with_percent_symbol` or `y` when defining `z`.

```
z =
```

[ ]

c) (3 pts) Determine the value of the following expression.

```
y.iloc[3] + " tax on goods from " + tariffs.get("Country").loc[3]
```

[ ]

Question 3 (10 pts)

Trump's administration set the reciprocal tariffs based on tariffs charged to the USA.

For each country in `tariffs`, the value in the `"Reciprocal Tariff"` column is simply half of the value in the `"Tariffs Charged to USA"` column, rounded **up** to the next integer.

In addition, if the `"Tariffs Charged to USA"` is less than 20 percent, then the `"Reciprocal Tariff"` is set to 10 percent, so that no country's reciprocal tariff is ever less than 10 percent.

a) (3 pts) Fill in the blanks in the function `reciprocate` which takes as input an integer representing the tariffs charged to the USA by a country, and returns an integer representing the reciprocal tariff that the US will impose on that country.

```
def reciprocate(charged):
    half = int((charged + 1) / 2)
    if __(a)__:
        return __(b)__
    else:
        return __(c)__
```

(a): (b):

(c):

b) (3 pts) Fill in the return statement of the function `reciprocate_2` which behaves the same as `reciprocate` but is implemented differently. You may **not** call the `reciprocate` function.

```
def reciprocate_2(charged):
    return __(d)__
```

(d):

c) (4 pts) Define the variable `ch` as follows.

```
ch = tariffs.get("Tariffs Charged to USA")
```

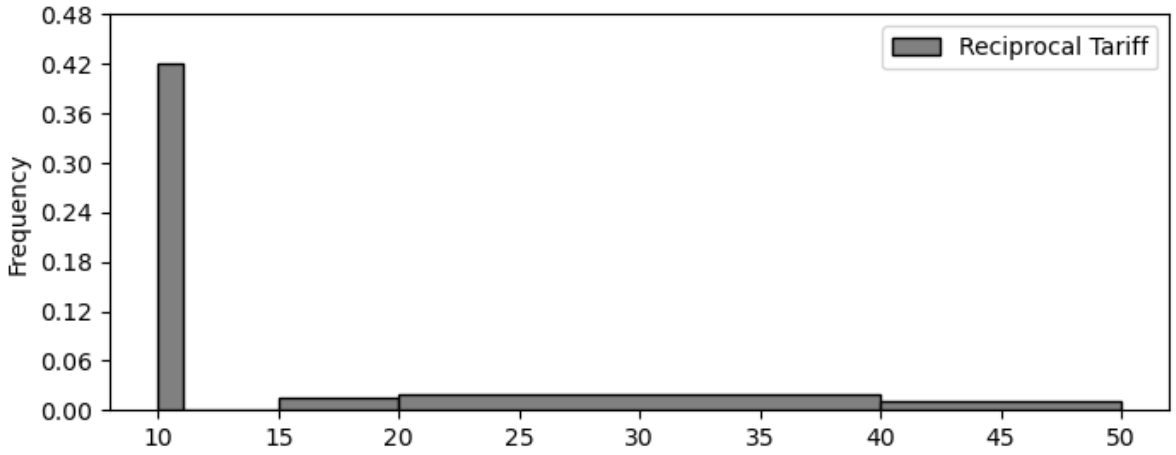
You want to check that `reciprocate` and `reciprocate_2` give the same outputs on all inputs in `ch`. Write an expression that evaluates to `True` if this is the case, and `False` otherwise.

PID: _____

Question 4 (9 pts)

Below is a density histogram displaying the distribution of reciprocal tariffs for each of the 50 countries on Trump's chart. It was plotted with the argument `bins=[10, 11, 15, 20, 40, 50]`.

Note that while the European Union is actually a group of many countries, it is counted as one country here.



a) (3 pts) How many countries have a reciprocal tariff of 10%?

b) (3 pts) Suppose we plotted the same histogram, except we changed the `bins` argument to `bins = [8, 15, 22, 30, 40, 50]`. What would be the height of the **leftmost** bar in this histogram? Give your answer as a number to two decimal places.

c) (3 pts) The European Union is not actually one country, but a group of 27 countries. Imagine we were to replace the row of `tariffs` corresponding to the European Union with 27 rows representing each of the member countries (all with a 20% reciprocal tariff), then plot a histogram of the reciprocal tariffs using `bins = [10, 11, 15, 20, 40, 50]`.

Let h_{new} be the height of the **rightmost** bar in this histogram, and let h_{old} be the height of the rightmost bar in the original histogram shown above. Express h_{new} in terms of h_{old}.

$h_{\text{new}} =$

Question 5 (14 pts)

Suppose we have another DataFrame called `trade_partners` that has a row for every country that the United States trades with. `trade_partners` is indexed by `"Country"` and has two columns:

- The `"Proportion"` column contains `float`s representing the proportion of US imports coming from each country.
- The `"Continent"` column contains the name of the continent where the country is located.

All countries in `tariffs` are included in `trade_partners` (including `"European Union"`), but not all countries in `trade_partners` are included in `tariffs`. The first three rows of `trade_partners` are shown at right.

Country	Proportion	Continent
Mexico	0.16	North America
China	0.14	Asia
Canada	0.13	North America

a) (3 pts) Write one line of code to merge `tariffs` with `trade_partners` and store the result in `merged`.

b) (3 pts) How many rows does `merged` have?

c) (4 pts) In which of the following DataFrames does the `"Proportion"` column sum to 1? Select all that apply.

- ☐ `trade_partners`
- ☐ `trade_partners.groupby("Continent").mean()`
- ☐ `trade_partners.groupby("Continent").sum()`
- ☐ `merged`
- ☐ None of the above.

d) (4 pts) Write one line of code that would produce an appropriate data visualization showing the **median reciprocal tariff for each continent**.

PID: _____

Question 6 (12 pts)

Most imported goods are transported to the US in shipping containers. The table to the right shows the probability that a randomly selected shipping container comes from a given location (continent or country). Note that the probability for each continent is the sum of the probabilities for each country in that continent.

For all parts of this question, you can leave your answer as an **unsimplified mathematical expression.**

Continent	Country	Probability
Asia (0.41)	China	0.14
	Japan	0.05
	Vietnam	0.04
	Other	0.18
Europe (0.24)	Germany	0.05
	Ireland	0.03
	Italy	0.02
	Other	0.14
North America (0.30)	Mexico	0.16
	Canada	0.13
	Other	0.01
Other (0.05)	*Other*	0.05

a) (3 pts) Suppose you randomly select **one** shipping container. You are told that it comes from Asia but not from Japan. What is the probability that it comes from China?

b) (3 pts) True or False: Selecting a shipping container from Vietnam and selecting a shipping container from Ireland are independent events.

◯ True ◯ False

c) (3 pts) Suppose you randomly select **two** shipping containers, with replacement. What is the probability that exactly one comes from Germany and the other comes from somewhere in North America?

d) (3 pts) Suppose you randomly select **three** shipping containers, with replacement. What is the probability that none of them comes from Asia?

Question 7 (14 pts)

The announcement of the tariffs affected many products, one of which was the Nintendo Switch 2, a new video game console. Due to the tariffs, preorders of the Nintendo Switch 2 were put on hold so pricing could be reconsidered. In this problem, we'll imagine a scenario in which Nintendo used this delay period to drum up excitement for their new product.

Suppose Nintendo arranges a contest to give away `k` of their new Switch 2 consoles. The contest is open to anyone and `n` people participate, with `n > k`. Everyone has an equal chance of winning, and nobody can win more than once. Jason and Ray both enter the contest, and they want to estimate the probability that they **both win**.

a) (6 pts) Fill in the blanks in the function `giveaway` so that it returns an estimate of the probability that Jason and Ray both win a Switch 2, when there are `n` participants and `k` prizes.

```
1 def giveaway(n, k):
2     count = 0
3     for i in np.arange(10000):
4         winners = np.random.choice(___(a)___)
5         if ___(b)___:
6             count = count + 1
7     return ___(c)___
```

(a):

(b):

(c):

b) (3 pts) If you implement `giveaway` correctly, what should `giveaway(100, 100)` evaluate to?

c) (5 pts) Suppose you modify the `giveaway` function as follows:

- Change line 2 to `results = np.array([])`.
- Change line 6 to `results = np.append(results, "WIN!")`.
- Leave lines 1, 3, 4, and 5 unchanged, including your code in blanks `(a)` and `(b)`.

Which of the following could be used to fill in blank `(c)`? Select all that apply.

- ☐ `len(results)/10000`
- ☐ `(results == "WIN!").sum()`
- ☐ `(results == "WIN!").mean()`
- ☐ `np.count_nonzero(results)`
- ☐ `np.random.choice(results)`
- ☐ None of the above.