Midterm Exam - DSC 10, Spring 2026

Full Name:

PID:

Exam Time: ◯ 10AM ◯ 11AM

Instructions:

- This exam consists of 8 questions, worth a total of 96 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 tear off the last page of this exam packet and read the data description.

Question 1 (16 pts)

Recall that each value in the `"height"` column of `mascots` is a string with two heights separated by `" - "`. Heights are formatted like `"5ft 6in"` or `"6ft"`.

In this problem, you will need to fill in the blanks in the provided code to add three new columns to `mascots`:

- `"min"` (`int`) The smallest height that fits into the mascot costume, measured in **inches**.
- `"max"` (`int`) The largest height that fits into the mascot costume, measured in **inches**.
- `"range"` (`int`): The difference between the smallest and largest heights that fit into the mascot costume, measured in **inches**.

After the blanks have been correctly filled in, the first five rows of `mascots` should appear as follows.

	school	mascot	type	height	team	colors	min	max	range
0	UC San Diego	King Triton	other	5ft 6in - 5ft 11in	The Tritons	[blue, gold]	66	71	5
1	University of Minnesota	Goldy the Gopher	animal	5ft 7in - 6ft	The Golden Gophers	[red, gold]	67	72	5
2	University of Notre Dame	The Leprechaun	human	4ft - 7ft	The Fighting Irish	[blue, gold, green]	48	84	36
3	Rutgers University	Sir Henry	human	5ft 1in - 5ft 8in	The Scarlet Knights	[red]	61	68	7
4	Loyola University Maryland	Iggy the Greyhound	animal	5ft 7in - 6ft 2in	The Greyhounds	[green, grey]	67	74	7

Important: For the rest of the exam, we'll assume that `mascots` contains these three additional columns.

a) (8 pts) To start, complete the implementation of the function `inches` which takes as input a single height and returns that height in inches, as an int. Example inputs and outputs are given below.

```
>>> inches("6ft 11in")
83

>>> inches("5ft")
60

def inches(h):
    parts = h.split(" ")
    output = __(a)__
    if len(parts) == 2:
        output = __(b)__
    return output
```

(a):

(b):

b) (8 pts) Fill in the blanks below so that the code adds the three required columns to `mascots`. Feel free to use the `inches` function defined in part (a).

```
def min_height(s):
    return __(c)__

def max_height(s):
    return __(d)__

mascots = mascots.assign(min = mascots.get("height")
                                        .apply(min_height))
mascots = mascots.assign(max = mascots.get("height")
                                        .apply(max_height))
mascots = mascots.assign(range = __(e)__)
```

(c):

(d):

(e):

Question 2 (10 pts)

Which of the following expressions evaluate to the name of the school with the widest height range for its mascot costume (or any such school, in the case of a tie)? **Select all that apply.**

Hint: `.groupby` orders rows in ascending order of the values in the index.

☐ `mascots.set_index("school").get("range").max()`

☐ `mascots.set_index("school").sort_values(by="range").index[-1]`

☐ `mascots.groupby("range").max().get("school").iloc[-1]`

☐ `mascots.groupby("school").max().index[-1]`

☐
```
mascots[mascots.get("range") ==
        mascots.get("range").max()].get("school").iloc[0]
```

☐ None of the above.

PID: _____

Question 3 (10 pts)

Sofia defines the variables below. Note that `hundred` is a DataFrame containing the first 100 rows of `mascots`.

```
hundred = mascots.take(np.arange(100))
has_state = hundred.get("school").str.contains("State")
is_animal = hundred.get("type") == "animal"
```

a) (2 pts) What is the data type of `is_animal`?

◯ string ◯ boolean ◯ array ◯ Series ◯ DataFrame

b) (4 pts) Sofia randomly selects one row from `hundred`. Which expression evaluates to the conditional probability that she selects a school with an `"animal"` mascot, given that she selects a school with `"State"` in its name?

◯ `(has_state & is_animal).sum() / 100`

◯ `(has_state & is_animal).sum() / is_animal.sum()`

◯ `(has_state & is_animal).sum() / has_state.sum()`

◯ `is_animal.sum() / 100`

◯ `is_animal.sum() / has_state.sum()`

c) (4 pts) The distribution of mascot `"type"` in `hundred` is shown in the table below.

`"human"`	`"animal"`	`"other"`
40	50	10

Now, Sofia randomly selects **two rows** from `hundred`, independently and **with replacement**. What is the probability that Sofia selects at least one school with a `"human"` mascot? Give your answer as a **mathematical expression**, which you do not need to simplify.

Question 4 (14 pts)

Ella creates two histograms as follows.

- Histogram A:

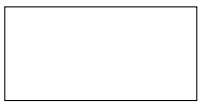
```
mascots.plot(kind="hist", y="max", density=True,
             bins=np.arange(48, 85, 8))
```

- Histogram B:

```
mascots.plot(kind="hist", y="max", density=True,
             bins=np.arange(48, 85, 4))
```

a) (3 pts) How many bars are in Histogram A? Give your answer as an **integer**.

b) (3 pts) What is the largest value that can be displayed in Histogram B? Give your answer as an **integer**.



c) (8 pts) Which of the following statements **must** be true? **Select all that apply.**

☐ The total area of Histogram A equals the total area of Histogram B.

☐ If no mascot has a maximum height of 70 or 71 inches, then in both histograms, the height of the bar that includes the interval [70, 71] is 0.

☐ The proportion of mascots with maximum height between 65 and 70 inches (inclusive) can be found by adding the areas of all bars that cover any part of the interval [65, 70].

☐ The tallest bar in Histogram B is the same height as the tallest bar in Histogram A.

☐ None of the above.

PID: _____

Question 5 (10 pts)

Define two DataFrames `single` and `double` as follows.

```
single = mascots.groupby("range").count()
double = mascots.groupby(["min", "max"]).count()
```

a) (3 pts) Which statement below evaluates to `True`?

◯ `single.shape[0] < double.shape[0]`
◯ `single.shape[0] == double.shape[0]`
◯ `single.shape[0] > double.shape[0]`

b) (4 pts) Which of the following statements are true? **Select all that apply.**

☐ If two schools are grouped together in `single`, they must also be grouped together in `double`.

☐ If two schools are grouped together in `double`, they must also be grouped together in `single`.

☐ If two schools are in different groups in `single`, they must also be in different groups in `double`.

☐ If two schools are in different groups in `double`, they must also be in different groups in `single`.

☐ None of the above.

c) (3 pts) Which statement below evaluates to `True`?

◯ `mascots.groupby("height").count().shape[0] < double.shape[0]`
◯ `mascots.groupby("height").count().shape[0] == double.shape[0]`
◯ `mascots.groupby("height").count().shape[0] > double.shape[0]`

Question 6 (12 pts)

Suppose Pranav randomly selects 4 different schools from among those included in `mascots`. Fill in the blanks below to implement a simulation that estimates the probability that **at least two** of the schools Pranav selected have a mascot of type `"animal"`.

```
mascot_types = np.array(mascots.get("type"))

repetitions = 10000
count_event = 0

for i in np.arange(repetitions):
    sample = np.random.choice(__(a)__)

    num_animals = np.count_nonzero(__(b)__)

    if __(c)__:
        count_event += 1

prob = __(d)__
```

a) (3 pts) What goes in blank `(a)`?

- ◯ `mascot_types, repetitions, replace=True`
- ◯ `mascot_types, repetitions, replace=False`
- ◯ `mascot_types, 4, replace=True`
- ◯ `mascot_types, 4, replace=False`

b) (3 pts) What goes in blank `(b)`?

c) (3 pts) What goes in blank `(c)`?

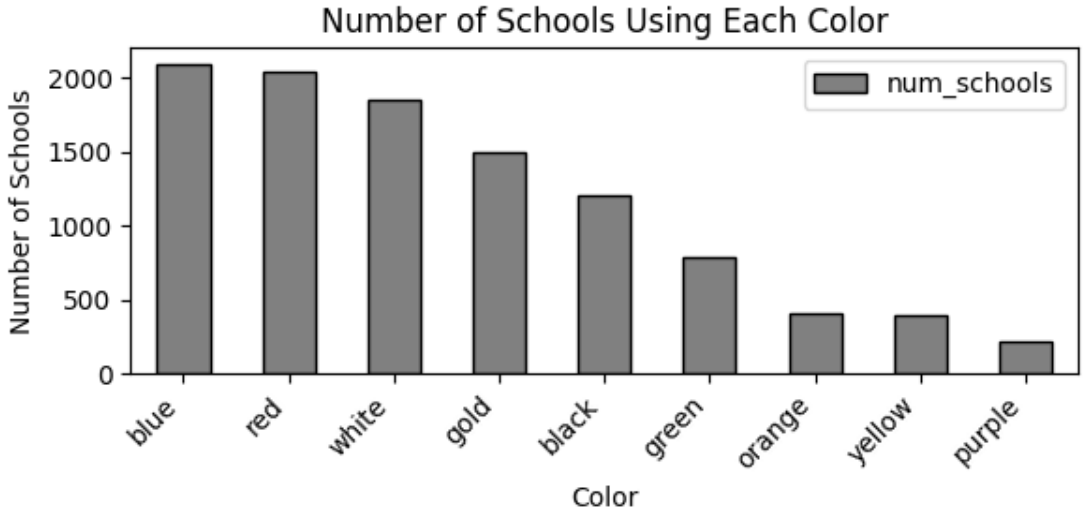
d) (3 pts) What goes in blank `(d)`?

- ◯ `count_event`
- ◯ `count_event / repetitions`
- ◯ `count_event / 4`
- ◯ `np.mean(count_event)`
- ◯ `count_event.sum() / repetitions`

PID: _____

Question 7 (12 pts)

The bar chart below was produced from the data in `mascots`. It shows how many schools use each color.



a) (4 pts) If you knew the exact heights of each bar, how could you determine the total number of schools in `mascots`?

◯ Sum the heights of all bars.

◯ Sum the heights of all bars and divide by the number of colors (9).

◯ You could not determine the total number of schools from this information.

b) (4 pts) In the bar chart, the height of the blue bar is 2100 and the height of the gold bar is 1500. Use this information and the first five rows of `mascots` (provided with the data description) to determine the **maximum possible** number of schools in `mascots` whose school colors are blue and gold **only**. Give your answer as an **integer**.

c) (4 pts) Write one line of code that uses the `mascots` DataFrame to calculate the sum of the heights of all bars in the bar chart above.

Question 8 (12 pts)

Austin and Ray are contestants on Season One of *America's Next Top Mascot*, a new reality TV show. In each episode, contestants will compete for one of the mascot roles in the DataFrame below, called `antm`. The data includes the name of each mascot and the minimum and maximum heights, in inches, for the costume.

	mascot	min	max
0	Bucky Badger	48	66
1	The Blue Devil	48	71
2	Leo III	61	66
3	The Nittany Lion	61	71
4	Brutus Buckeye	66	71
5	Pistol Pete	66	71
6	Cosmo the Cougar	66	74
7	The Mountaineer	66	74
8	Sparky	71	76
9	Sammy the Slug	71	76

a) (4 pts) The *America's Next Top Mascot* costume director wants to identify every pair of costumes where one costume's maximum height exactly equals the other costume's minimum height. She calls these "single-height costume pairs" because there is only a single height that can wear both costumes.

Write one line of code that uses `.merge` to calculate the number of single-height costume pairs in `antm`. Your answer should be a **Python expression that evaluates to an integer**.

b) (4 pts) What is the number of single-height costume pairs? In other words, what number does a correct answer to part (a) evaluate to? Give your answer as an **integer**.

c) (4 pts) The last episode of the season will reveal the mascot roles to be featured next season. Next season's mascots are stored in a DataFrame called `antm2` which has the same column names as `antm`, but contains different data (i.e. a different set of mascots). We don't have all of `antm2` available, but do know some information about it. First, we know that `antm2.merge(antm2, on="min")` gives a DataFrame with 61 rows.

Additionally, we have the table at right, which shows the number of mascots in `antm2` with each minimum height requirement, except one value is missing.

How many mascots in `antm2` have a minimum height requirement of 66 inches? Give your answer as an **integer**.

min	count
48	2
61	2
66	
71	2

College Mascots

A college mascot is a costumed character that represents the school at athletic competitions and other major campus events. For example, you might see UCSD's mascot, King Triton, at basketball games and the upcoming Sun God Festival.

College students can apply to serve as the mascot for their school, which is quite competitive! Applicants usually need to fall within a certain height range to fit the costume. Throughout this exam, we will format heights in feet and inches, for example `"5ft 7in"` or `"6ft"`, omitting inches when the height is a whole number of feet, as in the second example. Recall that there are 12 inches in one foot.

The DataFrame `mascots` has one row for every college or university in the US with a named mascot. The columns are:

- `"school"` (`str`): The name of the school. This column has no repeated values.
- `"mascot"` (`str`): The name of the school's mascot.
- `"type"` (`str`): Mascot category, either `"human"`, `"animal"`, or `"other"`.
- `"height"` (`str`): The range of heights that fit into the mascot costume. Values consist of two heights, formatted as described above, separated by `" - "`.
- `"team"` (`str`): The general name of the school's athletic team.
- `"colors"` (`list`): A list of the school colors. Each school has at least one school color.

The first few rows of `mascots` are shown below, though `mascots` has many more rows than pictured.

	school	mascot	type	height	team	colors
0	UC San Diego	King Triton	other	5ft 6in - 5ft 11in	The Tritons	[blue, gold]
1	University of Minnesota	Goldy the Gopher	animal	5ft 7in - 6ft	The Golden Gophers	[maroon, gold]
2	University of Notre Dame	The Leprechaun	human	4ft - 7ft	The Fighting Irish	[blue, gold, green]
3	Rutgers University	Sir Henry	human	5ft 1in - 5ft 8in	The Scarlet Knights	[scarlet]
4	Loyola University Maryland	Iggy the Greyhound	animal	5ft 7in - 6ft 2in	The Greyhounds	[green, grey]

Throughout this exam, we will refer to `mascots` repeatedly. Assume that we have already run `import babypandas as bpd` and `import numpy as np`.