

Your name: \_\_\_\_\_

Your student ID: \_\_\_\_\_

Your Berkeley email: \_\_\_\_\_

Your room location: \_\_\_\_\_

Student ID of the person to your left: \_\_\_\_\_

Student ID of the person to your right: \_\_\_\_\_

---

You have 50 minutes. There are 4 questions of varying credit. (37 points total)

|           |    |   |   |    |   |       |
|-----------|----|---|---|----|---|-------|
| Question: | HC | 1 | 2 | 3  | 4 | Total |
| Points:   | 1  | 7 | 9 | 20 | 0 | 37    |

For questions with **circular bubbles**, you may select only one choice.

- ☐ Unselected option (Completely unfilled)
- ☒ Don't do this (it will be graded as incorrect)
- ☒ Only one selected option (completely filled)

For questions with **square checkboxes**, you may select one or more choices.

- ☒ You can select
- ☒ multiple squares
- ☒ (Don't do this)

Anything written outside the answer boxes or ~~crossed-out~~ will not be graded. If you write multiple answers, your answer is ambiguous, or the bubble/checkbox is not entirely filled in, we will grade the worst interpretation. For coding questions with blanks, you may write at most one statement per blank, unless otherwise stated.

---

|                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------------------------|
| As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules of this exam. |
|----------------------------------------------------------------------------------------------------------------------------------------|

Honor Code (HC): I have read and agree to the honor code above.

(1 point) Sign your name: \_\_\_\_\_

## Q1 What Would Python Do? (WWPD)

(7 points)

Q1.1 (1 point) [True/False] In Data 6, a “variable type” and a “data type” refer to different concepts.

☐ True

☐ False

Q1.2 (1 point) [True/False] If a variable type is numerical, then we must store that variable in Python as an **int**.

☐ True

☐ False

For the following three questions (1.3 to 1.5), consider the function below, which takes in an integer **miles** and calculates the equivalent distance in kilometers:

```
1 def to_kilometers(miles):  
2     print(miles * 1.609)  
3 distance = to_kilometers(10)
```

Q1.3 (1 point) After running the code block, what is *printed*?

☐ None

☐ 16.09

☐ TypeError

☐ 10

☐ 16

☐ None of the above

Q1.4 (1 point) After running the code block, what does **distance** evaluate to?

☐ None

☐ 16.09

☐ TypeError

☐ 10

☐ 16

☐ None of the above

Q1.5 (1 point) Assume the code block has been fully run. What would **distance + 5** evaluate to?

☐ None

☐ 21.09

☐ TypeError

☐ 5

☐ 21

☐ None of the above

Q1.6 (1 point) What does the following Python code evaluate to? Write your answer in the box below.

```
1 min(5, 2) * max(abs(-10), abs(2)) // 10
```

Q1.7 (1 point) If this code block was a Jupyter notebook cell, what would be displayed?

```
1 def data(six):
2     result = six**2
3     return result
4
5 data(10)
```

☐ 100

☐ None

☐ 10

☐ Error

## Q2 A little bit of everything

(9 points)

Q2.1 (2 points) Recall that **print** displays arguments separated by spaces (' '). Further, recall that '\ ' acts as an escape character.

When the below code is run, what is printed?

```
1 course_nums = make_array(6, 8, 100, 101)
2 next = "DATA" + str(course_nums.item(-2))
3 print("I think I'll take DATA", course_nums.item(1), "and then", next)
```

☐ I think I\\'ll take DATA 8 and then DATA 100

☐ I think I'll take DATA 8 and then DATA 100

☐ I think I'll take DATA 8 and then DATA100

☐ I think I'll take DATA8 and then DATA 100

☐ None of these

Q2.2 (2 points) What does **arr** evaluate to after the following lines of code fully execute? Write your answer in the box below.

```
1 arr = make_array(19, 20, 22, 21, 30)
2 arr = arr - 10
3 arr[0] = arr[0] + 5
```

Q2.3 (1 point) Assume that all lines of code above have been fully executed. How can I calculate the average of the values in **arr**? Select all that apply.

☐ `sum(arr) / len(arr)`

☐ `max(arr) / min(arr)`

☐ `arr / 5`

☐ `sum(arr) / 5`

Q2.4 (1 point) Assume that all lines of code above have been fully executed. How can I access the last item in `arr`? Select all that apply.

☐ `arr.item(-1)`

☐ `arr.item(4)`

☐ `arr.item(0)`

☐ `arr.item(len(arr) - 1)`

Q2.5 (1 point) What is the unit of analysis for the following research question: Do families in Los Angeles County go to the beach more often than families in other counties in California?

☐ Individuals in California

☐ Neighborhoods

☐ Families in California

☐ Counties

☐ Families

Q2.6 (2 points) In *Weapons of Math Destruction*, how was the college ranking system “self-reinforcing”? Please limit your answer to 2-3 sentences.

### Q3 Olé, Olé, Olé, Olé!

(20 points)

Consider a World Cup dataset containing performance statistics for participating teams. We have created **worldcup**, a table with key performance metrics for teams in the tournament (Table 1). Each row in **worldcup** represents one team’s aggregate statistics.

| Team          | Goals Scored | Goals Conceded | Performance | Top Scorer         |
|---------------|--------------|----------------|-------------|--------------------|
| Mexico        | 10           | 3              | [4, 0, 1]   | Julián Quiñones 16 |
| Norway        | 13           | 11             | [3, 0, 2]   | Erling Haaland 9   |
| France        | 16           | 2              | [6, 0, 0]   | Kylian Mbappé 10   |
| Egypt         | 8            | 7              | [1, 2, 1]   | Mohamed Salah 10   |
| United States | 11           | 8              | [2, 1, 2]   | Folarin Balogun 20 |
| Cabo Verde    | 4            | 5              | [0, 3, 1]   | Ryan Mendes 20     |

Table 1: The first six rows of the World Cup Statistics table.

(Question 3 continued...)

Variable descriptions:

- **Team:** Name of the country (string).
- **Goals Scored:** Total number of goals scored by the team across all matches (integer).
- **Goals Conceded:** Total number of goals scored against the team across all matches (integer).
- **Performance:** Record in group stage matches formatted as [wins, draws, losses] (array of 3 integers).
- **Top Scorer:** Name and jersey number of the team's leading goal scorer (string).

Q3.1 (1 point)

What is the variable type of **Team**?

- |                                            |                                           |
|--------------------------------------------|-------------------------------------------|
| <input type="radio"/> Discrete Numerical   | <input type="radio"/> Ordinal Categorical |
| <input type="radio"/> Continuous Numerical | <input type="radio"/> Nominal Categorical |

Q3.2 (1 point) What is the variable type of **Goals Scored**?

- |                                            |                                           |
|--------------------------------------------|-------------------------------------------|
| <input type="radio"/> Discrete Numerical   | <input type="radio"/> Ordinal Categorical |
| <input type="radio"/> Continuous Numerical | <input type="radio"/> Nominal Categorical |

Q3.3 (1 point) What is the variable type of **Top Scorer**?

- |                                            |                                           |
|--------------------------------------------|-------------------------------------------|
| <input type="radio"/> Discrete Numerical   | <input type="radio"/> Ordinal Categorical |
| <input type="radio"/> Continuous Numerical | <input type="radio"/> Nominal Categorical |

Q3.4 (1 point) Consider the code below. What is the data type of **goals**, i.e., what does **type(goals)** evaluate to?

```
goals = worldcup.column("Goals Scored")
```

- |                                                             |                                                  |
|-------------------------------------------------------------|--------------------------------------------------|
| <input type="radio"/> <code>datascience.tables.Table</code> | <input type="radio"/> <code>numpy.ndarray</code> |
| <input type="radio"/> <code>int</code>                      | <input type="radio"/> None of the above          |

Q3.5 (2 points) Assume the code in the previous part has executed correctly. Which of the below lines of code will assign **max\_goals** to the maximum number of goals scored by any team? Select all that apply.

- ☐ `max_goals = max(goals)`
- ☐ `max_goals = worldcup.sort("Goals Scored", descending=True).column("Goals Scored").item(0)`
- ☐ `max_goals = worldcup.sort("Goals Scored").column("Goals Scored").item(0)`
- ☐ `max_goals = max(worldcup.select("Goals Scored"))`
- ☐ `max_goals = worldcup.sort("Goals Scored").column("Goals Scored").item(-1)`
- ☐ `max_goals = worldcup.sort("Goals Scored").goals.item(-1)`



(Question 3 continued...)

Q3.11 (2 points) The column **Jersey Number** now contains each top scorer's jersey number as an integer. Gilberto notices that a lot of the top scorers have jersey numbers that end in 0. He wants to calculate the total number of teams whose top scoring player has a jersey number ending in 0. Which of the following expressions will correctly calculate this number?

- ☐ `sum(worldcup.column('Jersey Number') % 10 == 0)`
- ☐ `sum(worldcup.column('Jersey Number') // 10 == 0)`
- ☐ `np.count_nonzero(worldcup.column('Jersey Number') % 10 == 0)`
- ☐ `sum(worldcup.column('Jersey Number') % 10)`

Q3.12 (1 point) Gilberto decides goal differences (goals scored minus goals conceded) might be a better variable for measuring team performance than jersey numbers. Fill in the code below to create a new table **team\_performance** with the columns **Team** and **Goal Difference**.

```
team_performance = Table().with_columns(  
    'Team', worldcup.column('Team'),  
    'Goal Difference', _____  
)
```

Q3.12

Q3.13 (1 point) What is the *domain* of the **Goals Difference** column?

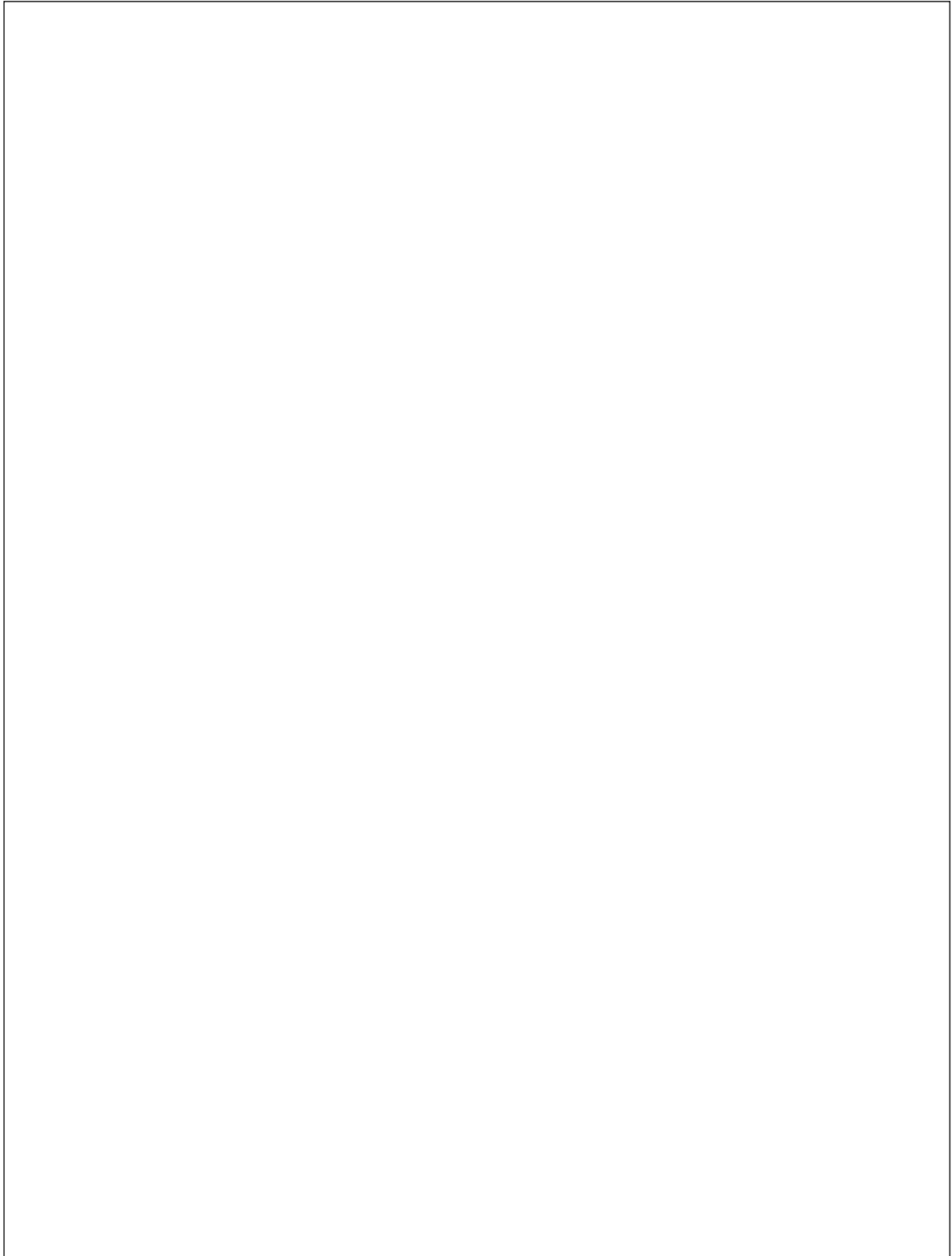
Q3.14 (3 points) Using the new table **team\_performance** calculate the *percentage* of teams who have a **Goal Difference** of 4 or greater and assign it to **percent\_above\_four**. Note that depending on your approach, you may not need both lines below. If you only use one line, cross out the unused line in the box below.

```
teams_above_four = _____  
Q3.15  
percent_above_four = _____  
Q3.16
```

**Q4** *Just for fun!*

**(0 points)**

Q4.1 Draw something fun, or write a message for the staff! Or leave this blank!

A large, empty rectangular box with a thin black border, intended for a drawing or message.