Midterm Exam - DSC 10, Winter 2026

Full Name:

PID:

Section: ◯ 9am ◯ 11am

Instructions:

- This exam consists of 11 questions.
- 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!

Person to your left:

Person to your right:

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) (2 pts) Which column of the `pizza` DataFrame would be an appropriate index?

- ◯ `"store"`
- ◯ `"kind"`
- ◯ `"rating"`
- ◯ `"neighborhood"`
- ◯ None of these

(b) (2 pts) Suppose you want to see if there is a relationship between `num_ingredients` and `price`. Which visualization will *best* enable you investigate this?

- ◯ Overlaid histograms of `num_ingredients` and `price`.
- ◯ A bar chart of the averages of `num_ingredients` for each value of `price`.
- ◯ A bar chart of the averages of `price` for each value of `num_ingredients`.
- ◯ A scatter plot of `price` vs. `num_ingredients`.

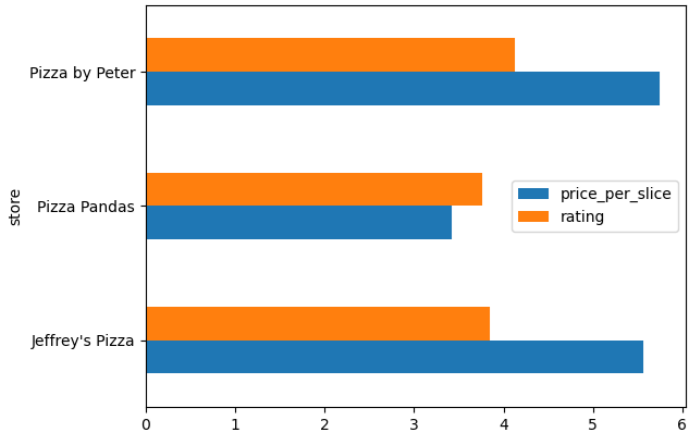
(c) (4 pts) Fill in the blanks to create a horizontal bar chart showing both the average `price_per_slice` and average `rating` for each store side by side. On a DataFrame of just the 10 rows shown in the data description, it would look like the image below.

```
(pizza._____(i)______
      .groupby("store")._____(ii)_____
      .plot(__(iii)__));
```

(i):

(ii):

(iii):



PID: _____

Question 2 (7 pts)

Fill in the blanks below so that the expression evaluates to a `float` that is the highest rating of any pizza slice that has more than 10 ingredients.

```
pizza[_____(a)_____]._____(b)_____
```

(a) (3 pts)

(b) (4 pts) What can go in blank (b)? **Select all that apply.**

- ☐ `get("rating").apply(max)`
- ☐ `get("rating").max()`
- ☐ `sort_values("rating", ascending=False).get("rating").iloc[0]`
- ☐ `sort_values("rating", ascending=True).get("rating").iloc[-1]`
- ☐ None of the above

Question 3 (4 pts)

First, the following code takes the first 10 rows of the `pizza` DataFrame and stores them into `pizza10` (thus `pizza10` consists of exactly the 10 rows shown on the Data Description page):

```
pizza10 = pizza.take(np.arange(10))
```

Next, consider the code below.

```
positions = np.arange(pizza10.shape[0])
result = positions[pizza10.get("rating") > 4].sum()
```

Hint: while `positions` is an array, the behavior of the code in the last line above is analogous to a query from a DataFrame.

(a) (2 pts) What does `result` represent?

- ◯ The number of pizzas in `pizza10` with rating > 4
- ◯ The sum of the ratings in `pizza10` that are greater than 4
- ◯ The sum of the index positions from `pizza10` for pizzas with rating > 4
- ◯ A Series of positions from `pizza10` for pizzas with rating > 4
- ◯ The average rating of pizzas from `pizza10` that have a rating > 4

(b) (2 (pts)] What does `result` evaluate to?

- ◯ 4
- ◯ 11
- ◯ 15
- ◯ 23
- ◯ 46

Question 4 (5 pts)

The Best Pizza Neighborhood Contest is coming up! For this contest, we want to find the neighborhood with the highest average rating for a slice of pizza. Fill in the blanks below so that the expression evaluates to the **name** of this neighborhood (as a string).

```
pizza.get([___(a)___]).groupby(___(b)___)
                      .___(c)___.sort_values(__(d)__, ascending=False)
                      .____(e)____
```

(a)

(b)

(c)

(d)

(e)

Question 5 (6 pts)

The DSC 10 staff is planning their quarterly pizza party, and they want to make sure that there are options for everyone. They're trying to find the total number of gluten-free pizza slice options in the San Diego area. Select all lines of code that correctly evaluate to the integer corresponding to the total number of gluten-free pizza slice options.

- ☐ `pizza.set_index("store").get("gluten_free").sum()`
- ☐ `pizza.get(["gluten_free", "store"]).groupby("store").sum()`
- ☐ `pizza[pizza.get("gluten_free")].get("gluten_free").count()`
- ☐ `pizza[pizza.get("gluten_free")].shape[0]`
- ☐ `pizza.set_index("store").get("gluten_free").count()`
- ☐ `pizza.get("gluten_free").shape[0]`

PID: _____

Question 6 (8 pts)

At Pizza by Peter, head chefs Bianca and Ella make each pizza, either together or alone. For any given pizza, there's a chance that the chefs mess up and the final product isn't suitable to serve. The probabilities are given below:

- If Bianca makes a pizza alone (tossing the dough and baking it), there's a $\frac{1}{4}$ chance of the pizza not being suitable to serve.
- If Ella makes a pizza alone (tossing the dough and baking it), there's a $\frac{1}{3}$ chance of the pizza not being suitable to serve.
- If Bianca tosses the dough and Ella bakes it, there's a $\frac{1}{6}$ chance of the pizza not being suitable to serve.
- If Ella tosses the dough and Bianca bakes it, there's a $\frac{1}{5}$ chance of the pizza not being suitable to serve.

a) (4 pts) Suppose that Ella makes one pizza alone, Bianca makes one pizza alone, and their suitability to serve when they each work alone is independent of each other. What is the probability that at least one of the pizzas is not suitable to serve? If it is possible to solve, express your final answer as a single fraction or decimal.

◯ Not enough information.

b) (4 pts) Let A be the event that **Ella worked on the pizza** (either alone or with Bianca), and B be the event that the pizza is **suitable to serve**.

Suppose that for a randomly selected pizza,

- $P(A) = \frac{1}{4}$, and
- $P(B) = \frac{3}{5}$.

Using **only** this information and the information from the 4 bullet points at the top of the question, what is P(A or B)? If it is possible to solve, express your final answer as a single fraction or decimal.

◯ Not enough information.

Question 7 (8 pts)

Recall from Question 3 that the DataFrame `pizza10` consists of exactly the 10 rows shown on the Data Description page. Now consider the following code:

```
neighborhood_counts = (pizza10.groupby('neighborhood')
                                        .count().get(['rating']))
```

(a) (2 pts) What type of object is `neighborhood_counts`?

- ◯ A DataFrame with 3 rows and 1 column
- ◯ A DataFrame with 10 rows and 1 column
- ◯ A Series with 3 elements where the index contains neighborhood names
- ◯ A Series with 3 elements where the index is 0, 1, 2
- ◯ An array with 3 elements

(b) (6 pts) We want to apply a function to the neighborhood names (which are currently in the index). Fill in the blanks to complete the code below so that `neighborhood_firstchar` is a Series where each element is the first letter of the entire string consisting of each neighborhood name. Note that your answer to (ii) may need to contain multiple methods sequentially.

```
def first_letter(s):
    ____(i)____

neighborhood_firstchar = neighborhood_counts._____(ii)_____
```

(i)

(ii)

Question 8 (7 pts)

A student at UCSD runs a food Instagram account, @ucsdfoodeater. They go around San Diego trying pizza at different restaurants and keep notes on what kinds of pizza they have tried and where.

They keep their notes in a DataFrame called `notes`; in the `notes` DataFrame, the **index** is the pizza kind, and the `restaurants_tried` column is a list of the restaurants where they've tried that kind.

```
notes = bpd.DataFrame().assign(
    kind=["Cheese", "Pepperoni", "Veggie", "Margherita",
          "Hawaiian", "Supreme", "BBQ Chicken", "White"],
    restaurants_tried=[
        ["Jeffrey's Pizza", "Pizza Pandas"],
        ["Jeffrey's Pizza", "Pizza by Peter"],
        ["Pizza Pandas"],
        ["Pizza by Peter"],
        ["Jeffrey's Pizza"],
        ["Jeffrey's Pizza"],
        ["Regents Pizzeria"],
        ["Pizza on Pearl"]
    ]
).set_index("kind")
```

The `notes` DataFrame has 8 rows. Note that some pizza kinds in `notes` do not appear in `pizza`, and some kinds in `pizza` do not appear in `notes`.

(a) (2 pts) Recall again that `pizza10` was created in Question 3, containing exactly the 10 rows shown on the Data Description page. What would the following expression evaluate to?

```
pizza10.merge(notes, left_on="kind", right_index=True).shape[0]
```

◯ 0
◯ 6
◯ 9
◯ 10
◯ Cannot be determined

(b) (5 pts) Which of the following expressions evaluate to the same value as part (a)? **Select all that apply.**

☐ `pizza10.merge(notes.reset_index(), on="kind").shape[0]`
☐ `pizza10.merge(notes, on="kind").shape[0]`
☐ `notes.merge(pizza10, left_index=True, right_on="kind").shape[0]`
☐ `notes.merge(pizza10, left_index=True, right_on="store").shape[0]`
☐ `notes.merge(pizza10.reset_index(), on="kind").shape[0]`

Question 9 (5 pts)

DoorDash is having a special promotion! This promotion allows Ray to order four random pizza slices from the `pizza` DataFrame, but he only wants to eat the best slices—according to the ratings. He decides to pick the first two of those random slices and compare their ratings to the other two slices. He will eat only if the sum of the ratings of his two slices is greater than the sum of the ratings of the other two slices.

Fill in the blanks in the code below so that `prob_eats` evaluates to an estimate of the probability that Ray gets to eat.

```
repetitions = 1000
count_eats = 0

for i in np.arange(repetitions):
  promo_ratings = np.random.choice(__(a)__, 4, replace=False)
  ray_sum_of_ratings = _____(b)_____
  other_sum_of_ratings = _____(c)_____
  if ray_sum_of_ratings > other_sum_of_ratings:
    count_eats = ____(d)____
prob_eats = _____(e)_____
```

(a)

(b)

(c)

(d)

(e)

PID: _____

Question 10 (4 pts)

Pizza by Peter awards loyalty points: every third slice you purchase from them earns 2^i points, starting from the 1st slice. For example, the 1st, 4th, 7th and 10th slices that a customer purchases each earn 2^i loyalty points where i is 1, 4, 7 and 10 respectively. Without using the `+` operator, write a **one-line expression** that evaluates to the **total** loyalty points from those four slices.

Question 11 (2 pts)

The management of Pizza by Peter is rebranding. The new name is the output of the following line of code. Write the new name as a string.

```
(pizza.get("kind").iloc[6].replace("a", "o") + "!").upper()
```