Quiz 2 - DSC 10, Spring 2024

Full Name:

PID:

Quiz Time: ◯ 2PM ◯ 2:30PM ◯ 3PM ◯ 3:30PM ◯ 4PM

Instructions:

- This quiz consists of 4 questions. You have a total of 20 minutes to complete it.
- 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**.
- If your answer is a string, make sure to put it in quotes. If your answer is a float, make sure to include a decimal point.
- No aids are allowed (no notes, no calculators, and no computers).

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

Signature:

Version A

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

An art museum records information about its collection in a DataFrame called `art`. The columns of `art` are as follows:

- `"title" (str)`: the name of the art piece.
- `"artist" (str)`: the name of the artist.
- `"year" (int)`: the year the art piece was produced.
- `"price" (float)`: the selling price of the art piece in dollars.

Question 1

Write an expression that evaluates to the number of art pieces made in 1950 that cost less than $10,000.

Question 2

a) Fill in the blanks in the code below to find the name of the artist in `art` who has the lowest mean price of art pieces.

```
art.groupby(___(x)___).___(y)___.sort_values(by="price").index[0]
```

`(x)`: _____ , `(y)`: _____

b) Fill in the blanks in the code below to find the name of the artist in `art` who made the most art pieces in a single year.

```
(art.groupby(___(x)___).___(y)___.reset_index()
    .sort_values(by="title", ascending=False)
    .get("artist").iloc[0])
```

`(x)`: _____ , `(y)`: _____

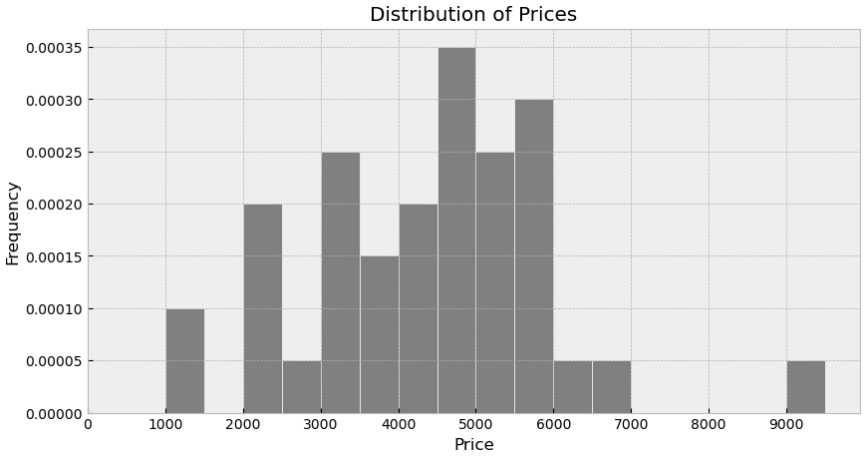
Question 3

a) Which of the following correctly plots a density histogram showing the distribution of `"price"` in `art`? Select all that apply.

- ☐ `art.get(["price"]).plot(kind="hist", density=True)`
- ☐ `art.get(["price"]).plot(kind="hist")`
- ☐
```
(art.drop(columns=["artist", "year", "price"])
    .plot(kind="hist", density=True))
```
- ☐ `art.plot(kind="hist", y="price")`
- ☐ `art.plot(kind="hist", y="price", density=True)`
- ☐ `art.plot(kind="hist", x="price", density=True)`

PID: _____

b) The density histogram below shows the distribution of `"price"` in `art`. If the museum has 100 art pieces in total, how many pieces cost at least \$3,000 but less than \$4,500?



Question 4

a) Fill in the return statement of the function `is_expensive`, which takes as input the price of an art piece (as a float, in dollars) and returns `True` if the price is more than 20 million dollars. Otherwise, it returns `False`.

```
def is_expensive(price):
    return ___(a)___
```

`(a)`:

b) Write one line of code to add a new column called `exp` to the `art` DataFrame, which categorizes if each art piece is worth more than 20 million dollars, using Boolean values. You must use the `is_expensive` function you wrote above. Make sure to modify `art`!

c) Next, we make a new DataFrame called `expensive` as follows.

```
expensive = art[art.get(exp)]
merged = art.merge(expensive, on="artist")
```

Van Gogh is one artist represented in `art` and exactly half of his pieces in `art` are worth over 20 million dollars. If Van Gogh's art appears in 72 rows of the `merged` DataFrame, how many rows does Van Gogh actually have in the original `art` DataFrame?

Before submitting your quiz, make sure your PID is on the front page and on the top right corner of page 3.