Quiz 2

● Graded

Student

Rey Stone

Total Points

11 / 20 pts

Question 1

Question 1

2 / 2 pts

Question 1 (2 pts)

Structure, Granularity, Scope, Temporality and Faithfulness

✓ **– 0 pts** Correct

Question 2

Question 2a

0 / 2 pts

Question 2a (2 pts)

We accepted two possible answers - without access to the entire dataset it was not possible to determine whether there were cases where the same food was served in different categories (i.e. scrambled eggs for breakfast and dinner)

Option 1:
Food item in a particular month.
OR
Option 2:
Food item in a particular category in a particular month (if the same food was served in different categories)

✓ **– 2 pts** Incorrect

Question 3

Question 2b

4.5 / 6 pts

Question 2b (6 pts)

Possible answers

```
(menu.groupby("Category")
    .agg({"Calories":"max"})
    .rename(columns={"Calories":"Max Calories"})
)
```

```
(menu.groupby("Category")
    .max()[["Calories"]]
    .rename(columns={"Calories":"Max Calories"})
)
```

```
(menu.groupby("Category")[["Calories"]]
    .max()
    .rename(columns={"Calories":"Max Calories"})
)
```

✓ **− 1 pt** Incorrect aggregation function or setup

✓ **− 0.5 pts** Incorrect or missing rename column

Question 4

Question 3

2.5 / 5 pts

Question 3 (5 pts)

□ $\sum_{i=1}^{n} 2 = 2\bar{x}$

● $\sum_{i=1}^{n} x_i = n\bar{x}$

□ $\sum_{i=1}^{n} (x_i - \bar{x}) = n\bar{x}$

● $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

□ The function $f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$ is minimized when $c = 0$

● The function $f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$ is minimized when $c = \bar{x}$

✓ **− 1.5 pts** Missing one of the correct options

✓ **− 1 pt** Marked one of the incorrect options

Question 5

Question 4a

1 / 1 pt

Question 4a (1 pt)

Correct answer: $L(m) = m^3(1-m)^4$

✓ **− 0 pts** Correct

Question 6

Question 4b

1 / 4 pts

Question 4b (4 pts)

Correct answer: $\frac{3}{7}$

Solution:

Since m is the probability of landing on heads, **we showed in HW 2 that the maximum of $L(m)$ will be equal to the proportion of heads in the sequence above.**

Here's a reminder of why (note the work shown below was not necessary to receive credit on the quiz if you just explained the reasoning given above).

The maximum of $L(m)$ occurs at the same value of m as the maximum of $\log(L(m))$ (recall: logs turn products into sums and are easier to work with !)

$$\log(L(m)) = \log(m^3(1-m)^4)$$
$$= 3\log(m) + 4\log(1-m).$$

To find the max, we find where the derivative equals 0. We start by taking the derivative:

$$\implies \tfrac{d}{dm}\left(\log(L(m)\right) = \tfrac{3}{m} - \tfrac{4}{1-m}$$

Now we solve for where the derivative equals 0:

$$\tfrac{d}{dm}\left(\log(L(m)\right) = 0$$

$$\implies \tfrac{3}{m} - \tfrac{4}{1-m} = 0$$

$$\implies \tfrac{3}{m} = \tfrac{4}{1-m}$$

$$\implies 3(1-m) = 4m$$

$$\implies 3 - 3m = 4m$$

$$\implies 3 = 7m$$

$$\implies \boxed{m = \frac{3}{7}}$$

✓ **– 3 pts** Rewrote using properties of logarithms, but incorrect or missing steps taking derivative and setting equal to zero.

CSCI 3022 **Quiz 2** **Sp25**

Write **clearly** and **in the box**:

Name: Rey Stone	**Student ID:**

Quiz Rules:

DO NOT TURN THIS PAGE OVER UNTIL THE QUIZ BEGINS.

- All cell phones must be stored in your backpack. If you have a cell phone anywhere on your body or at your desk during this quiz you will receive a 0 on this quiz.
- You are allowed a two-sided 8.5" x 11" crib sheet with hand-written (not typed) notes
- You are allowed a calculator
- You are allowed to use the Data Wrangling with Pandas Cheatsheet from the Canvas Modules
- **No tablets, smartphones, smartwatches or any other electronic devices allowed.**
- No collaboration with other students is allowed during this quiz.
- **Show all work and simplify your answers!**
- You have **15 minutes** for this quiz.

Once the quiz begins you can use this extra space for your work if you need more space.

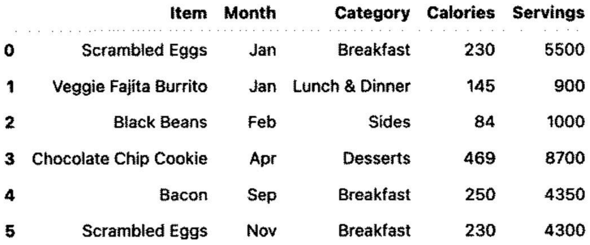
1. (2 pts) What are the 5 key data properties to consider when doing EDA as outlined in the lecture slides, lecture demos & HW?

1. structure 2. granularity 3. scope 4. temporality 5. faithfulness

2. You decide to study the food served at the Alferd Packer Grill (APC) on CU's campus.

   You gather data on all the menu items served at the APC **over the last 12 months** and put the data into a DataFrame called `menu`.

   Here are the first 6 rows of the `menu` DataFrame:



	Item	Month	Category	Calories	Servings
0	Scrambled Eggs	Jan	Breakfast	230	5500
1	Veggie Fajita Burrito	Jan	Lunch & Dinner	145	900
2	Black Beans	Feb	Sides	84	1000
3	Chocolate Chip Cookie	Apr	Desserts	469	8700
4	Bacon	Sep	Breakfast	250	4350
5	Scrambled Eggs	Nov	Breakfast	230	4300

   The `menu` DataFrame contains 5 columns:

   - **Item**: The name of the item on the menu
   - **Month**: The month of the year that the item was served
   - **Category**: The category of food (either Breakfast, Lunch & Dinner, Soups, Sides or Desserts)
   - **Calories**: The number of calories per serving of the item.
   - **Servings**: The number of servings served to students in the month.

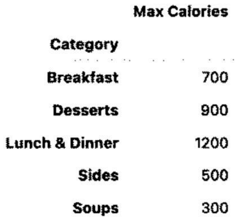
   (a) (2 pts) Based on the description of the columns, what is the granularity of the `menu` DataFrame?

   Answer to part (a)

   It lists an available menu item | meal at APC.

   (b) (6 pts) Write Python code to use an aggregation function on the DataFrame `menu` to construct a new DataFrame that is indexed by `Category` and that has one column whose value is equal to the the maximum `Calories` of all menu items in that Category. Name the aggregated column `Max Calories`.

   For example, the output of your code should return the following DataFrame:



	Max Calories
Category	
Breakfast	700
Desserts	900
Lunch & Dinner	1200
Sides	500
Soups	300

   Write your code directly in the box provided below.

```
import numpy as np
max_calories menu.groupby ("Category")["Max. Calories"] .agg(np.max)

max_calories #RUN DF
```

CONTINUED ON NEXT PAGE

3. (5 pts)

Let $x_1, x_2, \ldots, x_n$ be a fixed list of numbers and let $\bar{x}$ denote the mean (i.e. the average) of those numbers. Which of the following statements are always true? (For full credit you must select all that apply)

□ $\sum_{i=1}^{n} 2 = 2\bar{x}$ ☑ $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

☑ $\sum_{i=1}^{n} x_i = n\bar{x}$ ☑ The function $f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$ is minimized when $c = 0$

□ $\sum_{i=1}^{n} (x_i - \bar{x}) = n\bar{x}$ □ The function $f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$ is minimized when $c = \bar{x}$

4. (5 pts)

I have a **biased** coin that lands on **heads** with an **unknown probability** m.

If you toss this coin 7 times, the probability that you get the exact sequence TTTHTHH is a function of m.

As you learned in HW2, that function is called the **likelihood** of the sequence TTTHTHH, denoted $L(m)$.

(a) What is $L(m)$ for the sequence TTTHTHH?

Answer to part (a)

$(m)^3 \cdot (1-m)^4$

(b) What is the value of m that maximizes the likelihood of this sequence? **Explain/justify your reasoning.**
(Hint: You don't have to do any Calculus to find this value if you apply the concepts you learned when doing a similar problem on HW 2. Or you're welcome to do Calculus if you prefer).

Answer to part (b)

$\frac{3}{4}$

$\ln(m^3 \cdot (1-m)^4)$

$= \ln(m)^3 + \ln(1-m)^4$

$= 3\ln(m) + 4\ln(1-m)$

$= 3\ln(m) = -4\ln(1-m)$

$= \frac{3\ln(m)}{-4} = \ln(1-m)$

$-\frac{3}{4} = \frac{\ln(1-m)}{\ln(m)}$

$-\frac{3}{4} = \ln(1-m-m)$

$-\frac{3}{4} = \ln(1-2m)$ $m = \frac{3}{4}$

END OF QUIZ

If you finish early please close your quiz and wait until time is up.