DATA C8
Fall 2025

FOUNDATIONS OF DATA SCIENCE

FINAL

8:10–11:00AM MONDAY, DECEMBER 15TH 2025

PRINT Your Name: _____

PRINT Your Student ID: _____

PRINT Your Exam Room: _____

PRINT the Name of Person to your Left: _____

PRINT the Name of Person to your Right: _____

PRINT Your GSI's Name (Write N/A if in Self-Service): _____

INSTRUCTIONS

You have **170 minutes** to complete the exam. There are **7 questions** and **20 pages** on this exam, including this cover page.

Question	1	2	3	4	5	6	7	Total
Points	19	6	16	18	18	18	5	100

- This exam is closed book, closed computer and closed calculator, except the Reference Sheet provided for you.
- You may only have with you: a pencil, an eraser and your student ID, unless you have pre-approved accommodations.
- If you need to use the restroom, bring your phone, exam, reference sheet and student ID to the front of the room.
- For written questions:
  - answers written outside the boxes provided will not be graded;
  - if your answer is ambiguous or you provide multiple answers, the worst interpretation will be graded.
- For coding questions:
  - blank spaces may include multiple arguments or functions per blank, but your solution must use every blank available;
  - you may assume the `datascience` and `numpy` libraries are imported, as seen in class;
  - the use of **any code** which has not been taught in this offering of the course is not allowed and will result in zero credit.
- For multiple choice questions, see question types and instructions below.

Questions with **circular bubbles**: you may select only **1 choice**. Questions with **square boxes**: you may select **1 or more choices**.

○ Unselected option (completely unfilled)

● Single option selected (completely filled)

■ You may select multiple squares

■ as long as they are completely filled

You must fill in the bubbles **completely**. Ticks, crosses, or other check marks will **not** receive credit.

HONOR CODE

"As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others."

SIGN Your Name: _____

This page intentionally left blank

The exam begins on the next page.

1. [19.0 points] General

Read the instructions for each part carefully and answer accordingly.

(a) **[4.0 pts]** What does the following Python expression output to? Write the output in the box below. If you believe the code produces an error, write *Error*.

```
np.average(make_array(True, False))
```

(b) For each of the following set of hypotheses, *select all of the appropriate test statistics.*

i. **[2.0 pts]** Null Hypothesis: The coin is fair.
Alternative Hypothesis: The coin is biased towards tails.

- ☐ TVD
- ☐ number of heads
- ☐ number of tails
- ☐ number of tails − expected number of tails
- ☐ expected number of tails − number of tails
- ☐ |number of tails − expected number of tails|
- ◯ None of the above

ii. **[1.0 pt]** Null Hypothesis: The six-sided die is fair.
Alternative Hypothesis: The six-sided die is not fair.

- ☐ TVD
- ☐ Proportion of 1's
- ☐ Sum of the dice rolls
- ☐ Average of the dice rolls
- ◯ None of the above

iii. **[1.0 pt]** Null Hypothesis: The coin is fair.
Alternative Hypothesis: The coin is not fair.

- ☐ TVD
- ☐ number of heads
- ☐ number of tails
- ☐ number of heads − expected number of heads
- ☐ expected number of heads − number of heads
- ☐ |number of heads − expected number of heads|
- ◯ None of the above

(c) **[2.0 pts]** In a greenhouse, the heights of a certain type of plant have a mean of 170 cm and a standard deviation of 10 cm. What is the minimum percentage of plants whose heights are between 150 cm and 190 cm?

◯ 50% ◯ 68% ◯ 75% ◯ 95% ◯ Not enough information has been given to determine an answer.

(d) The following table displays cross-classified data from Data C8 students' responses to the Fall 2025 Mid-Semester Feedback Form.

Lecture Score	Regular Lab	Self-Service Lab
Extremely Unhelpful	3	5
Unhelpful	12	1
Average	25	6
Helpful	15	9
Extremely Helpful	22	2

i. **[1.0 pt]** What is the name of this kind of table?

ii. **[1.0 pt]** How many *rows* were in the original table used to construct this table? *Write a number.*

iii. **[1.0 pt]** How many *columns* were in the original table used to construct this table? *Write a number.*

(e) **[3.0 pts]** In 2023, researchers affiliated with the University of California, San Francisco (UCSF) analyzed survey data obtained from 10,280 particpants in the Adolescent Brain Cognitive Development (ABCD) Study. They found that adolescents who left their phone ringer on before bed more commonly reported "trouble falling or staying asleep." Speaking on the implications of their work, the researchers wrote:

Within this study, there are several strengths and limitations worth noting. The large, diverse, and population-based sample is a major strength, which gives the study great external validity.

What facet of data science is alluded to by the above statement?

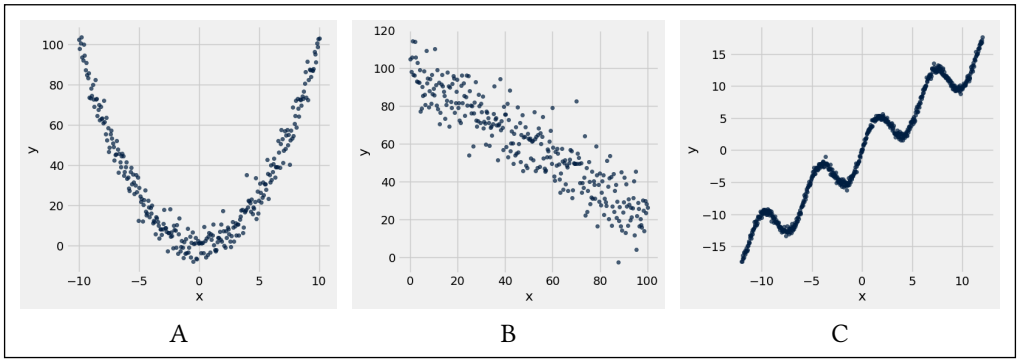
◯ Exploration ◯ Inference ◯ Prediction

(f) **[3.0 pts]** In what cases is it appropriate to establish causality from an observational study? Explain in *one sentence.*

Initials:

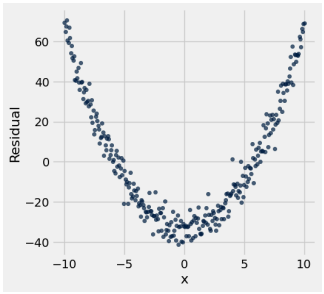
2. [6.0 points] Regression Warmup

This question focuses on the following three visualizations.



(a) Linear regression was performed on each of the three datasets, and below are the **residual plots** which resulted. Match these to the visualizations above. If you believe the residual plot cannot be produced using linear regression, select *Invalid.*

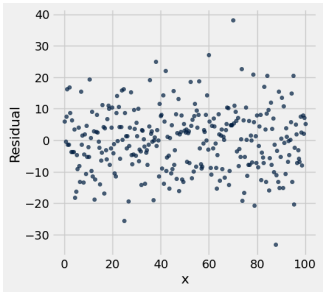
i. [1.0 pt] Residual 1



Select the matching visualization.

◯ A ◯ B ◯ C ◯ Invalid

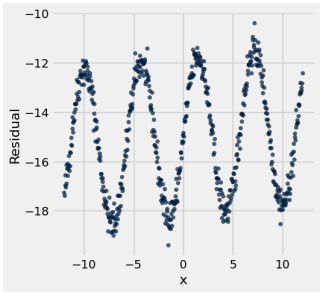
ii. [1.0 pt] Residual 2



Select the matching visualization.

◯ A ◯ B ◯ C ◯ Invalid

iii. [1.0 pt] Residual 3



Select the matching visualization.

◯ A ◯ B ◯ C ◯ Invalid

(b) [1.5 pts] For which of the visualizations is linear regression appropriate? *Select all that apply.*

☐ A

☐ B

☐ C

(c) Determine the possible approximate correlation coefficients for Plots A, B and C.

i. [0.5 pts] Plot A.

◯ -1 ◯ -0.75 ◯ 0 ◯ 0.75 ◯ 1 ◯ Not enough information

ii. [0.5 pts] Plot B.

◯ -1 ◯ -0.75 ◯ 0 ◯ 0.75 ◯ 1 ◯ Not enough information

iii. [0.5 pts] Plot C.

◯ -1 ◯ -0.75 ◯ 0 ◯ 0.75 ◯ 1 ◯ Not enough information

3. [16.0 points] OSki Trip

The Oski Ski Club is looking forward to a weekend ski trip! They are not exactly the best at planning, so they need your help to figure out which weekend will be perfect for their chilly escape.

You are given a table called **`availability`** that has a list of weekends where every member of the club is available and which locations have the best ski conditions for each respective weekend, with the following columns. An excerpt is shown below.

- **Weekend** (`string`): the Friday of the weekend every member is available, in the form "YY-MM-DD".
- **Month** (`string`): the corresponding month for a given weekend.
- **Desired Location** (`string`): the best ski location for a given weekend.

Weekend	Month	Desired Location
25-12-12	December	Mammoth
26-02-20	February	Palisades
25-12-26	December	Heavenly
26-01-09	January	Northstar
25-12-05	December	Big Bear
25-12-19	December	Mammoth
26-03-27	March	Palisades

(a) [3.0 pts] First, we want to make sure that the weekend we select is a ski season weekend. You are given a function called `is_ski_season` that takes in a single date as a string and returns a boolean for whether or not that date is a part of the ski season.

Fill in the blank `[A]` to create a table `availability_with_ski`. This table will consist of the columns in the `availability` table, plus an additional column, **Ski Season**, that signifies if the date is in the ski season. *Select all that apply.*

```
availability_with_ski = availability.with_column("Ski Season", _____[A]_____).where("Ski Season", True)
```

- ☐ `availability.apply(is_ski_season, "Weekend")`
- ☐ `availability.apply(is_ski_season, 1)`
- ☐ `availability.select("Weekend").apply(is_ski_season)`
- ☐ `availability.apply(is_ski_season, 0)`
- ☐ `availability.column("Weekend").apply(is_ski_season)`
- ☐ `is_ski_season(availability.column("Weekend"))`

The Oski Ski Club now needs to find a hotel to stay at for the weekend. You are given a table **`hotels`** which is a table of prices for each hotel on the specific weekends when there is a room available. Assume each hotel-weekend combination appears at most once in the table. The first few rows are shown below:

- **Hotel** (`string`): name of the hotel.
- **Date** (`string`): the Friday of the weekend of the hotel stay, in the form "YY-MM-DD".
- **Price ($)** (`int`): price for the entire stay at a given hotel.
- **Rating** (`float`): rating (out of 5.0 stars) for a given hotel.
- **Location** (`string`): Location of the hotel.

Hotel	Date	Price ($)	Rating	Location
Oski's Snow Lodge	26-01-02	350	4.7	Heavenly
Snowy Wonderland	25-12-19	800	2.2	Palisades
Golden Bear Cabin	25-12-19	150	4.3	Heavenly
Mouse Mountain Resort	26-01-02	310	4.5	Big Bear
Mouse Mountain Resort	26-03-06	225	4.5	Big Bear

(b) Help the club produce a table called `final_options` which contains information on the hotels available at their desired location during their available weekends. The table should contain the columns **Hotel**, **Weekend**, **Location**, **Price ($)** and **Rating**, in that order.

```
possible_hotels = availability.[B]
available_hotels = [C].with_column("Matching Dates", [D])
final_options = available_hotels.[E].[F]
```

i. [2.0 pts] Fill in blank `[B]`.

ii. [0.5 pts] Fill in blank `[C]`.

iii. [2.0 pts] Fill in blank `[D]`.

iv. [0.5 pts] Fill in blank `[E]`.

v. [1.0 pt] Fill in blank `[F]`.

(c) The club is ready to book their trip! Here are some extra criteria they would like to meet.

- It wants the trip to be on the earliest available weekend.
- It has a budget of $750.
- It wants to stay at the highest-rated hotel within that budget.

Based on these extra criteria, write a line of code that assigns `selected_hotel` to the name of the club's most ideal hotel (`string`). You may assume the code in **part (b)** was implemented correctly.

```
selected_hotel = final_options.[G].[H]
                              .[I].[J]
                              .item(0)
```

i. [0.5 pts] Fill in blank `[G]`.

ii. [0.5 pts] Fill in blank `[H]`.

iii. [0.5 pts] Fill in blank `[I]`.

iv. [0.5 pts] Fill in blank `[J]`.

(d) The Oski Ski Club now has access to a table called `filtered_hotels` with only the hotels that are available on every weekend. This new table has the same columns as `hotels`. Using `filtered_hotels`, visualize how the average price across all hotels changes over the weekend dates.

```
filtered_hotels.[K].[L]
```

i. [1.0 pt] Fill in blank `[K]`.

ii. [1.0 pt] Fill in blank `[L]`.

(e) For the following two scenarios, pick the table and visualization that are most appropriate. Assume that you can perform any necessary methods on the tables to obtain the desired visualization.

i. Visualize the distribution of the ratings of all the hotels.

1. [0.5 pts] Table

- ◯ `availability`
- ◯ `hotels`
- ◯ `final_options`

2. [1.0 pt] Visualization

- ◯ Line Plot
- ◯ Scatter Plot
- ◯ Bar Chart
- ◯ Histogram
- ◯ Overlaid Line Plot
- ◯ Overlaid Scatter Plot
- ◯ Overlaid Bar Chart
- ◯ Overlaid Histogram

ii. Visualize a distribution of the desired locations for each month.

1. [0.5 pts] Table

- ◯ `availability`
- ◯ `hotels`
- ◯ `final_options`

2. [1.0 pt] Visualization

- ◯ Line Plot
- ◯ Scatter Plot
- ◯ Bar Chart
- ◯ Histogram
- ◯ Overlaid Line Plot
- ◯ Overlaid Scatter Plot
- ◯ Overlaid Bar Chart
- ◯ Overlaid Histogram

4. [18.0 points] Dylan's Bikes

Dylan likes bike riding and, as a civil engineer, wants to investigate the trip durations of bike rides in Berkeley. He thinks that the mean trip duration in Berkeley is not equal to the national mean trip duration of 20 minutes. To test his belief, he collects the duration of 100 random bike ride trips in Berkeley and stores them in a single-column table, **`trips`**.

(a) Write code to simulate and visualize an empirical distribution of the mean duration of all bike rides in Berkeley. The empirical distribution should be stored in `trip_means`.

```
trip_means = _____[A]_____

for i in np.arange(10000):
    simulated_trips = _____[B]_____._____[C]_____
    one_mean = np.mean(_____[D]_____._____[E]_____)
    _____[F]_____

_____[G]_____._____[H]_____._____[I]_____
```

i. **[2.0 pts]** Fill in the blank `[A]`.

ii. **[0.5 pts]** Fill in the blank `[B]`.

iii. **[0.5 pts]** Fill in the blank `[C]`.

iv. **[0.5 pts]** Fill in the blank `[D]`.

v. **[0.5 pts]** Fill in the blank `[E]`.

vi. **[2.0 pts]** Fill in the blank `[F]`.

vii. [0.5 pts] Fill in the blank `[G]`.

viii. [0.5 pts] Fill in the blank `[H]`.

ix. [2.0 pts] Fill in the blank `[I]`.

(b) After completing the above code, Dylan finds that the 5^{th} percentile of `trip_means` is 4.7 minutes and the 95^{th} percentile of `trip_means` is 18.5 minutes.

i. [4.0 pts] Which of the following statements are true? *Select all that apply.*

- ☐ If we increase the number of bootstrap resamples, the accuracy of our estimate for the mean trip duration of all bike rides on campus will be higher.
- ☐ If Dylan collects 100 more bike ride durations, he would expect about 90 of them to be in the interval $[4.7, 18.5]$.
- ☐ If we repeat our bootstrapping process to compute 100 confidence intervals using the same percentiles, approximately 90 of them will contain the mean duration of all bike rides in Berkeley.
- ☐ If we use `trip_means` to compute a 95% confidence interval, both the values 4.7 and 18.5 will be within our confidence interval.
- ◯ None of these statements are true.

ii. [2.0 pts] Based on Dylan's initial hypothesis, what conclusions can he make at a **5%** p-value cutoff?

- ◯ The data is consistent with the hypothesis that the mean ride duration in Berkeley is equal to the national mean ride duration.
- ◯ The data is consistent with the hypothesis that the mean ride duration in Berkeley is not equal to the national mean ride duration.
- ◯ There is insufficient evidence to make any conclusion.

(c) Dylan would like to make some statements about the properties of a confidence interval. Choose if the following statements are True or False.

i. [1.0 pt] The confidence level refers to the probability that a previously generated interval contains the true parameter.

◯ True ◯ False

ii. [1.0 pt] The larger the sample size, the larger the width of the confidence interval.

◯ True ◯ False

iii. [1.0 pt] The larger the confidence level, the larger the width of the confidence interval.

◯ True ◯ False

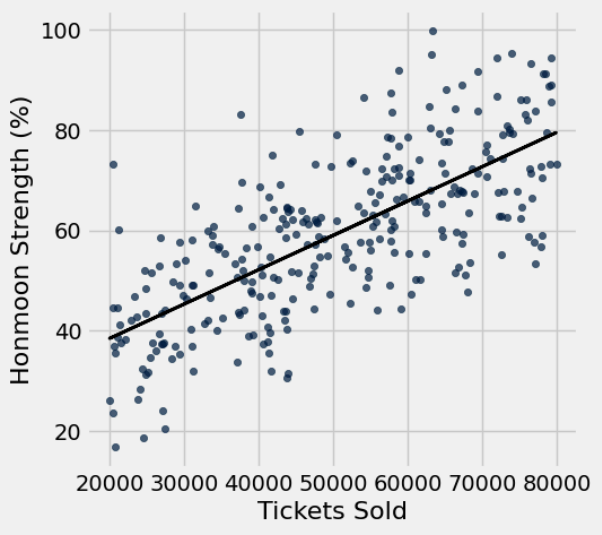
5. [18.0 points] KPop Demon Hunters

Tim and Marissa are huge Huntrix fans! They love going to their concerts and listening to their favorite songs like "Golden". They're interested in predicting the strength of the Honmoon (a magical forcefield that keeps demons from invading the human world) using the number of tickets sold in a show, and have collected data in the **`shows`** table; an excerpt is shown below.

- **Date** (`string`): the date of the show.
- **Tickets Sold** (`int`): the total number of tickets sold for that show.
- **Honmoon Strength (%)** (`float`): the strength of the Honmoon on a scale from 0 to 100.

Date	Tickets Sold	Honmoon Strength (%)
2025-01-01	47384	53.0
2025-01-02	78019	59.0
2025-01-03	64425	73.4
2025-01-04	58756	72.6
2025-01-05	38888	48.0

(a) Marissa plots the following scatter plot between **Tickets Sold** and **Honmoon Strength (%)** along with the regression line.



For the following statements, select True or False:

i. [2.0 pts] A regression line appears appropriate to fit to the data, given the scatter plot.

◯ True ◯ False

ii. [1.0 pt] It would be appropriate to predict the approximate Honmoon Strength in percent for a show that sold 100,000 tickets.

◯ True ◯ False

iii. [1.0 pt] Using the average number of Tickets Sold as an input, the regression line will predict the average percentage of Honmoon Strength.

◯ True ◯ False

(b) Choose the required quantities that Tim and Marissa need to obtain each of the following components of the regression line.

A Correlation Coefficient, r	**B** Number of Rows in the table	**C** Date of each Row
D SD of Tickets Sold	**E** SD of Honmoon Strength	**F** Mean of Tickets Sold
G Mean of Honmoon Strength	**H** Median of Tickets Sold	**I** Median of Honmoon Strength

i. [1.0 pt] Slope of the regression line in original units.

☐ **A** ☐ **B** ☐ **C** ☐ **D** ☐ **E** ☐ **F** ☐ **G** ☐ **H** ☐ **I** ◯ None of the above

ii. [1.0 pt] Intercept of the regression line in original units.

☐ **A** ☐ **B** ☐ **C** ☐ **D** ☐ **E** ☐ **F** ☐ **G** ☐ **H** ☐ **I** ◯ None of the above

iii. [1.0 pt] Slope of the regression line in standard units.

☐ **A** ☐ **B** ☐ **C** ☐ **D** ☐ **E** ☐ **F** ☐ **G** ☐ **H** ☐ **I** ◯ None of the above

iv. [1.0 pt] Intercept of the regression line in standard units.

☐ **A** ☐ **B** ☐ **C** ☐ **D** ☐ **E** ☐ **F** ☐ **G** ☐ **H** ☐ **I** ◯ None of the above

(c) Tim and Marissa decide that just using number of **Tickets Sold** is not good enough to predict the **Honmoon Strength**, so they now want to use more variables in the hopes of improving their predictions. They add two more columns to the `shows` table:

- **Ramen Eaten** (`int`): The total number of ramen cups that the Huntrix girls ate before the show.
- **Demons Slayed** (`int`): The total number of demons that the Huntrix girls slayed on the day of the show.

Date	Tickets Sold	Ramen Eaten	Demons Slayed	Honmoon Strength (%)
2025-01-01	47384	9	2106	53.0
2025-01-02	78019	18	3000	59.0
2025-01-03	64425	12	3000	73.4

i. [3.0 pts] Tim writes out the form of the multiple linear regression equation he will use to predict **Honmoon Strength (%)** using **Tickets Sold**, **Ramen Eaten**, and **Demons Slayed** as predictors.

$$\textbf{Honmoon Strength } (\%) = a \cdot \textbf{Tickets Sold} + b \cdot \textbf{Ramen Eaten} + c \cdot \textbf{Demons Slayed} + d$$

In one sentence, interpret b in the context of the problem.

ii. [5.0 pts] Tim splits his data into `train` and `test` sets, and would like to use the `minimize` function in the `datascience` library to help him determine the exact numbers he should use in the multiple linear regression equation on the previous page. What piece of the `shows` table should he have `minimize` work with?

◯ The entire `shows` table

◯ `train`

◯ `test`

iii. [2.0 pts] Tim stores the model predictions in an array called `predictions`. Use this array to write a line of code that calculates the root mean square error (RMSE) for his multiple linear regression model.

```
rmse = [A]
```

Fill in blank `[A]`.

6. [18.0 points] Gob Ears!

Miriam is currently at the Big Game watching Cal and Stanford play! She wants to predict whether Cal or Stanford will win. Luckily, she collected data from past Big Games in a table called **`go_bears`**. An excerpt is shown below, along with the result of running `go_bears.group("Winner")`.

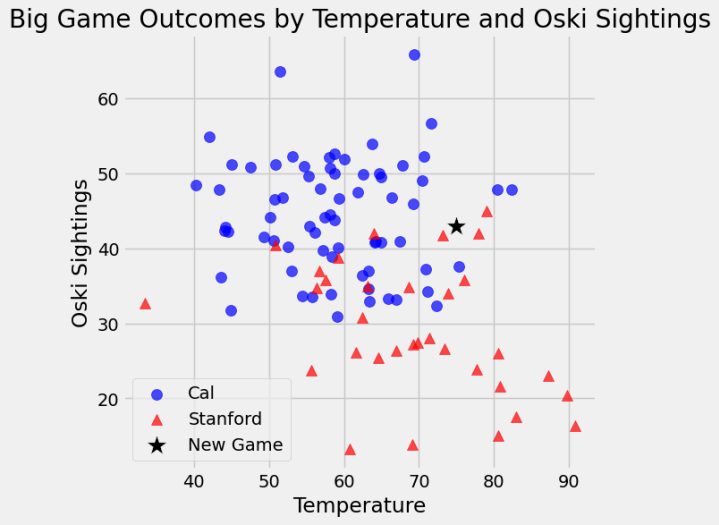
- **Winner** (`string`): The winner of the game, either *Cal* or *Stanford*.
- **Temperature** (`float`): The temperature in Fahrenheit during the game.
- **Oski Sightings** (`int`): The total number of reported Oski sightings at the game.
- **Cal Score** (`int`): The total number of points that Cal scored during the game.

Winner	Temperature	Oski Sightings	Cal Score
Stanford	80.6	26	12
Cal	50.8	46	35
Cal	50.2	44	40
Cal	58.8	43	24
Stanford	71.4	28	21

Winner	count
Cal	69
Stanford	31

`go_bears.group("Winner")`

Miriam plans to use k-nearest neighbors classification to predict whether Cal or Stanford will win. She creates the following scatterplot of **Temperature** against **Oski Sightings**, with the shape of each point representing the type of **Winner**, and measures the current temperature and number of Oski sightings at the game, adding it to the plot as a star.



(a) Based on the scatterplot above, match the k values to the corresponding classification for the point representing the current game (the star).

i. [2.0 pts] $k = 5$

◯ Cal ◯ Stanford

ii. [2.0 pts] $k = 15$

◯ Cal ◯ Stanford

(b) [2.0 pts] What is the minimum possible value for k that Miriam can choose to ensure that her model always classifies the winner as Cal?

(c) [4.0 pts] To help her classify, Miriam defines a `distance` function below, which takes arrays of two predictor variables `p1` and `p2` and calculates a distance between a new point `row` and each row in the training data.

```
def distance(p1, p2, row):
    arr = np.array(row)
    v1 = arr.item(0)
    v2 = arr.item(1)
    distances = _____[A]_____
    return distances
```

Fill in the blank `[A]`, such that the `distance` function returns an **array** of Euclidean distances.

(d) Now Miriam writes the `predict` function. You may assume the `distance` function in **part (c)** has been implemented correctly. Help her complete the `predict` function, which returns the predicted class of the `label` column as a string.

```
1  def predict(p1, p2, new_row, label, k):
2      distances = _____[B]_____
3      tbl_with_distances = go_bears.with_column("Distance", distances)
4      top_k_neighbors = tbl_with_distances._____[C]_____
5      result = top_k_neighbors._____[D]_____._____[E]_____._____[F]_____.item(0)
6      return result
```

i. [0.5 pts] Fill in blank `[B]`.

ii. [1.0 pt] Fill in blank `[C]`.

iii. [0.5 pts] Fill in blank `[D]`.

iv. [0.5 pts] Fill in blank `[E]`.

v. [0.5 pts] Fill in blank `[F]`.

(e) **[3.0 pts]** The temperature during the 2025 Big Game was 62.0 degrees and there were 30 Oski sightings at the game. Let the name `game_2025` contain the Row object with this information. Write *one line of code* to predict the winner of the game using the `go_bears` table as a training set, with a 7-nearest neighbors classifier. Assume the functions defined in **part (c)** and **(d)** work correctly.

(f) **[2.0 pts]** Instead of performing kNN classification, Miriam decides to perform kNN *regression* to predict Cal's score. Select the line(s) from the code in **part (d)** that we need to change for this task.

- ☐ Line 1
- ☐ Line 2
- ☐ Line 3
- ☐ Line 4
- ☐ Line 5
- ☐ Line 6

7. [5.0 points] Bayesian Burgers

Seizing on the consistently disappointing Impossible Oski Burgers at Silver Bear Cafe, Cafe 8 rolls out a revised burger under new head chef Mariel. Prakrat hears about Cafe 8's burger reinvention and is skeptical. He wants to determine whether the new Cafe 8 burger is *good* or *bad* before deciding to hike all the way up to Cafe 8.

(a) [1.0 pt] Traumatized by over two years of sloppy burgers, Prakrat assigns the probability $\mathbb{P}[\text{Good}] = 0.1$. What does this probability represent?

- ◯ The posterior probability that the burger is good.
- ◯ The prior probability representing Prakrat's belief of the probability that the burger is good.
- ◯ The likelihood of the burger being good tonight.
- ◯ The actual probability, under any circumstance, that the burger is good tonight.

He then learns from Cyrus that Cafe 8 is more likely to play music when the burgers are good. More formally:

$$\mathbb{P}[\text{Cafe 8 plays music} \mid \text{Good}] = 0.9 \qquad \text{and} \qquad \mathbb{P}[\text{Cafe 8 plays music} \mid \text{Bad}] = 0.2$$

Feel free to use this box below to draw a diagram to help with the following questions. This will **not** be graded.

(b) [1.0 pt] What is the probability that Cafe 8 plays music? You may write an unsimplified math expression. Please show your work and how you obtained your final answer in the box below.

For **parts (c)** and **(d)**, no credit will be awarded for only filling in the bubble. You must show your work and how you obtained your final answer.

(c) [2.0 pts] Prakrat hears from a friend that there is currently music playing in Cafe 8! Given the information on the prior page, and using a "more likely than not" binary classifier, would Prakrat conclude the burger is *Good* or *Bad*?

◯ Good

◯ Bad

(d) [1.0 pt] Prakrat now feels (slightly) more confident in Cafe 8's food. He revises his beliefs and sets $\mathbb{P}[\text{Good}] = 0.25$. Given that there is still music playing, what is Prakrat's conclusion now, also using the "more likely than not" binary classifier?

◯ Good

◯ Bad

Congratulations!

You have now completed the Final Exam. If you have not been told otherwise, you may bring all of your testing materials (reference sheet and this test paper), as well as your student ID, to the front of the room. Once you have been checked off, you may leave quietly.

- Make sure you have written your initials on **each page** of the exam, otherwise you may lose points.
- Make sure you have filled in bubbles and squares completely, and that you have **not** used a checkmark or cross.
- Double check that you have not skipped over any questions.

Below, you may draw and caption your favorite Data 8 experience or staff member!