DATA 8 Foundations of Data Science

Fall 2023

FINAL EXAM

INSTRUCTIONS

You have 2 hours and 50 minutes to complete the exam.

- The exam is closed book, closed notes, closed computer, closed calculator, except the provided reference sheet.
- Mark your answers on the exam itself in the spaces provided. We will not grade answers written on scratch paper or outside the designated answer spaces.
- If you need to use the restroom, bring your phone and exam to the front of the room.

For questions with **circular bubbles**, you should select exactly *one* choice.

○ You must choose either this option

○ Or this one, but not both!

For questions with **square checkboxes**, you may select *multiple* choices.

□ You could select this choice.

□ You could select this one too!

Preliminaries

You can complete and submit these questions before the exam starts.

The exam is worth 140 points.

The sections are as follows:

True or False - 30 points

Community - 12 points

Merch - 40 points

Spotify - 30 points

Bears - 28 points

There is also a Just For Fun section, worth 0 points, and a Last Words section, where you can state any assumptions you made on the exam, also worth 0 points.

(a) What is your full name?

(b) What is your student ID number?

(c) Who is your lab GSI? You may write *Unknown* if you don't know their name.

(d) Sign here to confirm that all work on this exam is your own (or type your name if online).

1. **(30.0 points)** **True or False**

   (a) **(2.0 pt)** If a scatterplot has a correlation coefficient of 1, all of the points must lie perfectly on a straight line.

   ● True

   ○ False

   (b) **(2.0 pt)** When building a classifier, ensuring that you have a large and diverse training set is a good way to mitigate overfitting.

   ● True

   ○ False

   (c) **(2.0 pt)** According to the Central Limit Theorem, if a sample is large, and drawn at random from the population with replacement, then the probability distribution of the *sample mean* is roughly normal.

   ● True

   ○ False

   (d) **(2.0 pt)** If we use linear regression to predict y-values based on our x-values, where both x and y are standardized, the estimate of the intercept could be negative.

   ○ True

   ● False

   (e) **(2.0 pt)** If you are sampling a numerical attribute that can only take on values of 0 or 1, the SD of your sample could have a value of 0.5.

   ● True

   ○ False

   (f) **(2.0 pt)** If we use linear regression to predict y-values based on our x-values, the average of our residuals will always be zero, regardless of whether x and y are standardized.

   ● True

   ○ False

   (g) **(2.0 pt)** If you use k-nearest neighbors on a data set that has only 2 possible categories for class (e.g. 0 or 1) and a k of 4, there is guaranteed to be a unique class that has the majority among the k nearest neighbors in the training set.

   ○ True

   ● False

   (h) **(2.0 pt)** The total variation distance can only be applied to categorical distributions in which there are 3 or more unique categories.

   ○ True

   ● False

(i) **(2.0 pt)** The recommended way to estimate a classifier's accuracy on the population is to evaluate its accuracy on the training set.

○ True

● False

(j) **(2.0 pt)** For any distribution, the percent of data that lies within 3 SDs of the average is at least 80%.

● True

○ False

(k) **(2.0 pt)** When conducting a randomized control experiment, random assignment of treatment and control serves as a way to simulate data from the null hypothesis.

○ True

● False

(l) **(2.0 pt)** You can have two individuals whose distance is zero if calculated using only 1 numerical attribute, but whose distance is greater than zero if calculated using 2 numerical attributes.

● True

○ False

(m) **(2.0 pt)** Chebychev's Rule allows us to model subjective beliefs about events that involve randomness.

○ True

● False

(n) **(2.0 pt)** Modern neural networks are powerful machine learning models for classifying images because their features are *learned* (as opposed to being inputted as columns from the training set).

● True

○ False

(o) **(2.0 pt)** If a scatterplot has a correlation coefficient of 0, there is no way that all of the points lie on a straight line.

● True

All students were given credit for this problem

● False

2. (12.0 points) Community

Writers for the upcoming *Community* movie are writing the script as having three acts.

For each act, they will randomly select a theme for it to be about. The themes are randomly chosen from the following distribution generated from a public poll from X (formerly Twitter):

- 60% chance of paintball fight
- 40% chance of multiverse

Note: Assume each act is sampled with the same set of probabilities regardless of what is picked for the other acts.

(a) (3.0 pt) What is the probability that three acts are multiverse, paintball fight and multiverse, in that order?

○ $(2 \times 0.4) \times (0.6)$

○ $(2 \times 0.4) + 0.6$

● $0.4^2 \times 0.6$

○ $0.6 \times 0.4 \times 0.6$

○ $1 - (0.6 \times 0.4 \times 0.6)$

Donald Glover, an actor from the original *Community* TV show, hasn't yet confirmed whether he will return for the movie.

Suppose we know the following conditional probabilities:

(b)

- If the third act has a paintball fight, there is a 20% chance Glover will return for the movie
- If the third act has a multiverse theme, there is a 50% chance Glover will return for the movie

i. (3.0 pt) What is the chance that the third act has a paintball fight and Glover does **not** return for the movie?

○ 0.4×0.8

○ 0.8

○ $1 - (0.2 + 0.5)$

○ $0.6 \times 0.8 + 0.4 \times 0.5$

● None of the above.

ii. (3.0 pt) Suppose the script has now been finalized and Glover announces that he will be returning for the movie.

What is the probability that the third act will be a paintball fight?

○ $\frac{0.2 \times 0.6}{0.2 \times 0.6 + 0.5 \times 0.8}$

○ 0.6×0.2

○ $\frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.4 \times 0.2}$

● $\frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.4 \times 0.5}$

○ $0.6 \times 0.2 + 0.4 \times 0.5$

○ None of the above.

iii. (3.0 pt) Suppose that before the script is finalized, there is a leak on social media that indicates the chance of the third act being a paintball fight is 90%.

The script then gets finalized and Glover announces that he will be returning for the movie.

Given the new information in the leak, what is our updated uprobability that the third act will be a paintball fight?

○ $\frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.5 \times 0.8}$

○ $\frac{0.9 \times 0.2}{0.9 \times 0.2 + 0.1 \times 0.2}$

○ 0.9×0.2

○ $\frac{0.9 \times 0.2}{0.9 \times 0.2 + 0.4 \times 0.5}$

○ $0.9 \times 0.2 + 0.1 \times 0.5$

● None of the above.

3. (40.0 points) Merch

Ernest and Mollie watched *Barbie* on opening day and were surprised that in the weeks after they saw several people wearing shirts that say "I am Kenough".

(a) **(3.0 pt)** Suppose that Mollie wants to randomly sample movies and use their merchandise sales to create a **95%** confidence interval for the **population mean** of 2023 merchandise sales.

If she knows that the population SD is $10 million, what is the minimum sample size she needs to create a confidence interval that has a width of $2 million?

Please draw a box around your final answer.

400

$$2,000,000 = 4 \cdot \frac{10,000,000}{\sqrt{samplesize}}$$

$$500,000 = \frac{10,000,000}{\sqrt{samplesize}}$$

$$\sqrt{samplesize} = 20$$

$$samplesize = 400$$

(b) **(2.0 pt)** Suppose that Mollie uses bootstrapping to create a 95% confidence interval using a sample size smaller than the one from part (a). Ernest states that the interval is guaranteed to be wider than $2 million.

Is Ernest's statement true or false?

Note: Assume your answer in part (a) is correct.

○ True

● False

(c) **(3.0 pt)** Suppose that Mollie uses the sample size from part (a) and constructs a 95% confidence interval of [35.1, 78.9].

What is the probability that the true population mean of merchandise sales is outside of this interval?

○ 2.5%

○ 5%

○ 10%

○ 95%

○ There is not enough information to answers because we don't know the endpoints of the confidence interval.

● None of the above. There is no chance involved in whether our confidence interval contains the true parameter.

(d) **(3.0 pt)** Suppose that Mollie wants to create a 95% confidence interval for the population 75th percentile of merchandise sales.

Which of the following methods could be used to create such an interval?

Select all that apply.

□ Chebychev's Inequality

■ Bootstrapping

□ Central Limit Theorem

□ Randomized Control Experiment

□ None of the above

Mollie suspects that a movie's Rotten Tomatoes score might have a relationship with the amount of merchandise it sells within the first month of theatrical release.

She randomly samples movies released in 2023 from Rotten Tomatoes and collects them into a table called `movies`. The first few rows are shown here:

Name	Critics	Audience	Sales
Barbie	88	83	202.3
Hunger Games	64	89	51.8
The Flash	63	83	36.9

. . . (57 rows omitted)

The table has the following columns:

(e)

- *Name*: (string) the movie's name
- *Critics*: (int) the movie's Rotten Tomatoes Tomatometer score (a percentage from 0 to 100)
- *Audience*: (int) the movie's Rotten Tomatoes Audience score (a percentage from 0 to 100)
- *Sales*: (float) the amount of movie's merchandise sold (in millions of dollars)

Note: The table has exactly 60 rows in it.

i. Mollie wants to fit a regression line to predict *Sales* from *Critics*, so she writes the following partially completed code:

```
def su(array):

    return (array - np.mean(array)) / np.std(array)

def intercept(x, y):

    correlation = _________
                     (a)

    slope = correlation * _________
                             (b)

    return _________
              (c)
```

The `intercept` function returns the intercept of the regression line.

Note: Both functions take in arrays as input.

A. (3.0 pt) Write a Python expression to fill in blank (a).

Note: Both functions take in arrays as input.

```
np.mean(su(x) * su(y))
```

B. (3.0 pt) Write a Python expression to fill in blank (b).

```
np.std(y) / np.std(x)
```

C. (3.0 pt) Write a Python expression to fill in blank (c).

```
np.mean(y) - slope * np.mean(x)
```

ii. (3.0 pt) Mollie fits a regression line to predict *Sales* from *Critics* and gets a slope of `-2.1`.

Which of the following would she expect to happen with the regression line's predictions?

Select all that apply.

□ The regression line will tend to overestimate *Sales* for movies with a below average *Critics* score.

□ The regression line will tend to underestimate *Sales* for movies with a below average *Critics* score.

□ The regression line will tend to overestimate *Sales* for movies with an above average *Critics* score.

□ The regression line will tend to underestimate *Sales* for movies with an above average *Critics* score.

■ None of the above.

iii. (3.0 pt) Ernest thinks the true slope of the regression line in the population is `0` and that the value observed in the sample above is due to chance. He bootstraps the data in `movies` to generate a confidence interval for the true slope.

Which of the following statements are true?

Select all that apply.

□ Every bootstrapped estimate of the slope will be negative.

■ The size of the bootstrap resamples will all be exactly `60`.

□ All `60` movies in the original sample will appear in every bootstrap resample.

□ The bootstrap process is equivalent to permuting the rows of the dataset repeatedly.

□ None of the above.

iv. (3.0 pt) Ernest constructs a 90% confidence interval for the true slope and finds it to be [-4.5, -1.1].

Assuming a p-value cutoff of 5%, which of the following can Ernest conclude based on his confidence interval?

Select all that apply.

□ The true slope in the population is `0`.

□ The true slope in the population is not `0`.

□ The true slope in the population is less than `0`.

■ None of the above.

v. (3.0 pt) Mollie's sister, Anna, argues that Ernest should have made a confidence interval for the correlation coefficient instead.

Which of the following statements are true?

Select all that apply.

□ The correlation coefficient should be used instead because it is unitless.

□ The correlation coefficient should be used instead because the magnitude of the slope could be affected by the units of the x-axis and y-axis.

□ It doesn't matter which value is used since the slope is equal to the correlation coefficient.

■ It doesn't matter which value is used since a slope of 0 implies the correlation coefficient is 0 as well.

□ None of the above.

(f) Rather than using the critics' scores, Mollie thinks it's a better idea to use the audience scores to predict merchandise sales.

Suppose she knows the following:

- the *Audience* column has a mean of 70 and a standard deviation of 10
- the *Sales* column has a mean of 100 and a standard deviation of 50
- the correlation between the *Audience* and *Sales* columns is 0.4

i. **(3.0 pt)** If Mollie wants to predict *Sales* from *Audience*, what would be the **intercept** of her regression line?

Please draw a box around your final answer.

-40

$$100 - (0.4 \cdot \frac{50}{10}) \cdot 70 = -40$$

ii. **(3.0 pt)** For a movie that has an audience score of 80, what would the regression line above predict as the merchandise sales?

○ 200

○ 150

○ 140

● 120

○ 110

○ None of the above

iii. **(2.0 pt)** What are the units for the slope in the above regression?

○ Dollars per Percent

○ Millions of Dollars

○ Millions of Dollars per Tomato

○ Dollars per Ounce of Ketchup

● None of the above

4. (30.0 points) Spotify

Barbara and Jeanine were quarantined for a week with 13 other friends due to unforseen circumstances.

They are curious to understand what songs each friend listened to during the quarantine period.

To evaluate this, they randomly sample song “plays” by the 15 people during quarantine and put that into `spotify` table. Here are the first few rows:

Username	Artist	Song	Genre	Duration
barbz23	Olivia Rodrigo	Vampire	Pop	3.14
jea9	The Weeknd	Popular	R&B	2.78
ronnieboi	Doja Cat	Paint the Town Red	Hip-Hop	3.05

. . . (328 rows omitted)

The table has the following columns:

- *Username*: (string) the spotify username of the person who played the song
- *Artist*: (string) the song’s artist
- *Song*: (string) the song’s name
- *Genre*: (string) the song’s genre
- *Duration*: (float) the number of minutes the song was played on that occasion

Note: There is a row for each time a song was played, so many rows will be repeated. For example, if Jeanine listened to the song Vampire 3 times, then there will be 3 rows in the table for those “plays”.

(a) **(3.0 pt)** Which of the following Python expression returns the name of the artist with the most plays in the table?

Hint: Each row of the table is equivalent to a single play.

Select all that apply.

□ `spotify.sort('Duration', descending=True).column(1).item(0)`

■ `spotify.group('Artist').sort(1, descending=True).column(0).item(0)`

□ `spotify.sort('Duration', descending=True).column('Artist').item(0)`

□ `spotify.select('Artist','Duration').group(0, max).sort(1, descending=True).column(0).item(0)`

□ None of the above.

(b) **(3.0 pt)** Write a Python expression that returns a table with more than 3 columns that displays the average play duration for each unique combination of artist and song.

```
spotify.pivot('Artist','Song','Duration',np.average)
or
spotify.pivot('Song', 'Artist', 'Duration', np.average)
```

(c) **(3.0 pt)** Write a Python expression that returns the name of the artist that has the largest number of unique songs in the table.

```
spotify.group(['Artist', 'Song']).group('Artist').sort('count',
descending=True).column(0).item(0)
```

Can also use `make_array()` for the first `group` argument.

(d) While looking at a table of song plays is helpful, Barbara notices that the table doesn't contain the names of people who played the songs.

She creates a separate table called `accounts` that contains their friends' Spotify accounts. The first few rows are shown here:

Identifier	DisplayName
jmarsdenofficial	James Marsden
margarita23	Inez De Leon
ken_the_og	Ken Hyun

. . . (12 rows omitted)

The table has the following columns:

- *Identifier*: (string) the account's ID in Spotify's database
- *DisplayName*: (string) the account's display name (first and last name)

Barbara notices that one of the friends, `'Todd Gregory'`, tends to skip Pop songs after listening to them for just a few seconds.

She writes the following partially completed code, which assigns `result` to an array containing the average play duration for every unique Pop song that Todd played.

```
combined = ______(a)______
todd_pop_songs = _____(b)_____
result = todd_pop_songs.______(c)______
```

Recall: The `spotify` table has columns *Username*, *Artist*, *Song*, *Genre* and *Duration*.

i. **(3.0 pt)** Write a Python expression to fill in blank (a).

```
spotify.join(‘Username’, accounts, ‘Identifier’)
```

or

```
accounts.join(‘Identifier’, spotify, ‘Username’)
```

ii. **(3.0 pt)** Write a Python expression to fill in blank (b).

```
combined.where('DisplayName', 'Todd Gregory').where('Genre', 'Pop')
```

or

```
combined.where('Genre', 'Pop').where('DisplayName', 'Todd Gregory')
```

iii. **(3.0 pt)** Write a Python expression to fill in blank (c).

```
.select('Song', 'Duration').group('Song', np.average).column(1)
```

(e) Jeanine notices that average play durations for `'Pop'` songs are typically lower than those for `'Hip-Hop'` songs across all 15 friends.

Barbara argues that any differences observed in the sample are only due to chance.

Recall: The `spotify` table has columns *Username*, *Artist*, *Song*, *Genre* and *Duration*.

i. **(3.0 pt)** Which of the following is an alternative hypothesis that Jeanine could use to assess her claims?

Select all that apply.

■ `'Pop'` song plays have a lower *Duration* on average than `'Hip-Hop'` song plays.

□ `'Pop'` song plays have have the same *Duration* distribution as `'Hip-Hop'` song plays.

□ All `'Pop'` song plays have a lower *Duration* than all `'Hip-Hop'` song plays.

■ `'Hip-Hop'` song plays have a higher *Duration* on average than `'Pop'` song plays.

□ None of the above.

ii. **(3.0 pt)** Which of the following test statistics could Jeanine use to assess her claims?

Select all that apply.

□ The total variation distance between the *Duration* distribution of `'Pop'` song plays and the *Duration* distribution of `'Hip-Hop'` song plays.

■ The mean *Duration* among `'Hip-Hop'` song plays minus the mean *Duration* among `'Pop'` song plays.

■ The mean *Duration* among `'Pop'` song plays minus the mean *Duration* among `'Hip-Hop'` song plays.

□ The mean *Duration* among `'Pop'` song plays.

□ The mean *Duration* among `'Hip-Hop'` song plays plus the mean *Duration* among `'Pop'` song plays.

□ None of the above.

iii. **(3.0 pt)** Jeanine chooses a test statistic such that large values favor the alternative.

She simulates the test statistic many times and stores these in an array called `test_stats`. Suppose the observed value of the test statistic is `12.1`.

Write a Python expression that returns the p-value for this hypothesis test.

```
np.sum(test_stats >= 12.1)/len(test_stats)
or
np.mean(test_stats >= 12.1)
```

iv. (3.0 pt) Jeanine use a p-value cutoff of **5%** and finds that this corresponds to a simulated test statistic of 10.2.

Given the information in part (iii), Which of the following can she conclude?

Select all that apply.

□ The data are consistent with the null hypothesis.

■ The data are consistent with the alternative hypothesis.

□ There is a 5% chance that the null hypothesis is true.

□ There is a 5% chance that the alternative hypothesis is true.

■ ’Pop’ song plays had a lower duration on average than ’Hip-Hop’ songs.

□ There is not enough information to make a conclusion of any kind.

5. (28.0 points) Bears

The Cal Bears and UCLA Bruins are currently members of the *Pac-12* sports conference, but need to move to another conference, such as the *ACC* or *Big Ten*, in 2024.

To predict which conference the schools might move to, Prof. Lawrence and Prof. Strauss collect a random sample of schools across the country and put this in a table called `schools`.

The first few rows are shown here:

Name	Conference	Public	Score	Ranking	Ratio
Penn State	Big Ten	True	986.0	60	15
Illinois	Big Ten	True	443.3	35	19
Miami	ACC	False	480.0	67	13

. . . (67 rows omitted)

The table contains the following columns:

- *Name*: (string) the name of school
- *Conference*: (string) the name of the school's current sports conference
- *Public*: (bool) whether the school is public (`True`) or private (`False`)
- *Score*: (float) the school's Director's Cup score, which measures sports performance
- *Ranking*: (int) the school's US News Ranking, which measures academic performance
- *Ratio*: (int) the school's student-to-faculty ratio

Note: There are schools from over 10 different conferences in the table.

(a) (2.0 pt) Prof. Lawrence would like to understand how *Score* varies across different conferences.

Which of the following would be most appropriate to visualize this information?

○ Scatterplot

○ Pivot Table

● Overlaid Histogram

○ Line Graph

● Bar Chart

Credit was given for both Overlaid Histogram and Bar Chart

(b) **(2.0 pt)** Prof. Lawrence wants to see the 50th percentile of *Score* for every combination of *Conference* and *Public*.

Which of the following functions could be used to visualize this information?

Select all that apply.

□ `scatter`

■ `pivot`

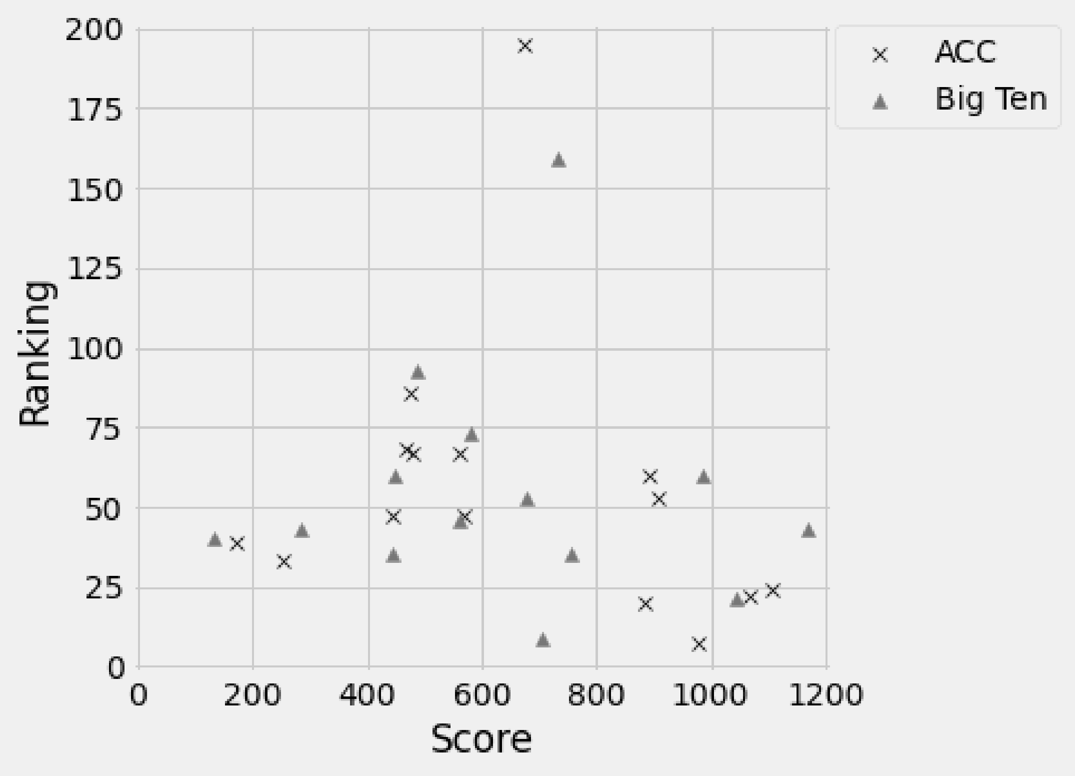
□ `hist`

■ `group`

□ `barh`

□ None of the Above

Prof. Lawrence creates the following chart showing *Score* against *Ranking* only for schools in the dataset that are part of the `'ACC'` or `'Big Ten'` conferences.



(c) **i.** **(2.0 pt)** Prof. Lawrence uses the chart above to build a k-nearest-neighbor classifier with $k = 4$ to predict the conference of schools outside the training set.

UCLA has a *Score* of `1000.25` and *Ranking* of `15`.

What would this nearest neighbor classifier predict as UCLA's *Conference*?

● `'ACC'`

○ `'Big Ten'`

○ There is no majority class.

ii. **(2.0 pt)** Cal has a *Score* of `833.25` and *Ranking* of `15`.

What would a nearest neighbor with $k = 5$ predict as as Cal's *Conference*?

○ `'ACC'`

● `'Big Ten'`

○ There is no majority class.

iii. **(2.0 pt)** Stanford has a *Score* of `1412` and *Ranking* of `3`.

Should the chart above be used to predict Stanford's *Conference*?

○ Yes

● No

iv. **(3.0 pt)** Prof. Lawrence's student, Atticus, thinks that the data should be standardized before building the k-nearest neighbors classifier.

Which of the following statements are true?

○ It doesn't matter if the data is standardized, since the set of nearest neighbors will be unchanged.

● It is important to standardize, since the mangnitude of the features affects how distance is calculated.

○ None of the above.

(d) Prof. Strauss now wants to use k-nearest-neighbors to predict a school's *Ranking* based on its *Score* and *Ratio* (i.e., he wants to predict a numerical value instead of a category).

i. To use the k-nearest-neighbors to perform this prediction, Prof. Strauss needs to first find the k nearest neighbors of the school with respect to *Score* and *Ratio*.

He writes a `neighbors()` function, which takes in the following arguments:

- `train`: A three-column table in which the first column is labeled *Score*, the second column is labeled *Ratio*, and the third column is labeled *Ranking*. Each row of the table represents a school in the training set.
- `new_school`: An array of length two containing a school's score and ratio. For example, `array([1200, 50])` corresponds to a school with a Director's Cup score of 1200 and a student-to-faculty ratio of 50.
- `k`: The value of k to use for k-nearest-neighbors.

The function returns a **table** containing the k neighbors in `train` that are closest to `new_school`. It is shown, partially completed, here:

```
def neighbors(train, new_school, k):
    score_diffs = ________(a)_________
    ratio_diffs = _________(b)_________
    distances = (________(c)_________) ** 0.5
    train_dist = train.with_column('Distance', distances)
    return ________(d)_________
```

A. (3.0 pt) Write a Python expression to fill in blank (a).

```
train.column(0) - new_school.item(0)
```

B. (3.0 pt) Write a Python expression to fill in blank (c).

Note: We did not have you fill out blank (b) to save you some time!

```
score_diffs ** 2 + ratio_diffs ** 2
```

C. (3.0 pt) Write a Python expression to fill in blank (d).

```
train_dist.sort('Distance', descending=False).take(np.arange(k))
```

ii. Now that he has a way to determine the k nearest neighbors, the last step is to create a prediction for a new school's US News Ranking.

To do this, Prof. Strauss will use the *geometric mean* of the k-nearest-neighbors' *Ranking* values. The geometric mean is calculated by multiplying all k of the *Ranking* values and then taking the k-th root of the multiple.

For example, if $k = 3$ and the 3 nearest neighbors have *Ranking* values of 40, 50 and 60, then the *geometric mean* is:

$$\sqrt[3]{40 \times 50 \times 60} = 49.32$$

Prof. Strauss writes the following partially completed code to generate the k nearest neighbors:

```
def prod(array):
    result = 1
    for value in array:
        result = result * value
    return result

def prediction(train, new_school, k):
    neighbors_rankings = ________(a)________
    return ________(b)_________
```

The `prediction()` function takes in the same arguments as the `neighbors()` function (i.e., `train`, `new_school` and `k`) and returns the *geometric mean* of the *Ranking* values from among the school's k nearest neighbors.

Note: The k-th root of a number is equivalent to raising the number to the power of $\frac{1}{k}$.

A. (3.0 pt) Write a Python expression to fill in blank (a).

```
neighbors(train, new_school, k).column('Ranking')
```

B. (3.0 pt) Write a Python expression to fill in blank (b).

Hint: You may use the `prod` function defined above.

```
prod(neighbors_rankings) ** 1/k
```

6. (0.0 points) Just for Fun :)

(a) Based on the exam questions, Prof. Sahai hasn't seen which of the following shows or movies?

○ Oppenheimer

○ Killers of the Flower Moon

● The Bear

○ Jury Duty

○ Community

(b) Draw a picture or meme describing your experience in Data 8!

7. (0.0 points) Last Words

(a) If there was any question on the exam that you thought was ambiguous and required clarification to be answerable, please identify the question (including the title of the section, e.g., Experiments) and state your assumptions. Be warned: We only plan to consider this information if we agree that the question was erroneous or ambiguous and we consider your assumption reasonable.