Final Exam

We know that you've demonstrated your real skills and abilities in producing your final project the past two weeks. In fact, when you go to show the work you've done to your family, friends, colleagues, employers, or advisors, you'll likely show off the great notebooks you've created.

This isn't to say that the final exam isn't important, it is. The exam matters both for your grade in this course (for certificate learners) and for your own self-confidence about your understanding of the material. So we want you to take pride in finishing the course strong by completing the final exam. Good luck!

Notice that next to each submit button, it says "You have used 0 of 2 attempts."

These questions are **NOT** unlimited to retake like the quizzes!

Please use your attempts carefully.

Data Science Analysis

1/1 point (graded)

Which of the following data would you use the analysis technique classification?

- (•) predicting the weather ✔
- ( ) predict the price of a stock
- ( ) simulating sales of a new product
- ( ) predicting the score on a test

Submit

You have used 2 of 2 attempts

Data Science Problem

1/1 point (graded)

What are the ingredients to form a data science problem? (Select 3.)

- [x] define what it is you're trying to tackle
- [x] assess the situation with respect to the problem
- [ ] assess the population the problem refers to
- [x] define your goals and objectives

✔

Submit You have used 2 of 2 attempts

Python

1/1 point (graded)

What will be printed by the following code snippet?

```
x = [1, 2, 3]
y = x
x[1] = 42
print(y)
```

- [ ] [1, 2, 3]

- [x] [1, 42, 3]
- [ ] [42, 2, 3]
- [ ] [42, 1, 3]

Submit You have used 2 of 2 attempts

NumPy Slicing Equivalence

1/1 point (graded)

Given the following code:

```
arr = np.array([[1,2,3,4],[5,6,7,8],[9,10,11,12]])
```

Which of the two following commands produce the same result?

- [x] arr[0:1,1:3]
- [ ] arr[2,1:3]
- [ ] arr[1:2,1:3]

- [x] arr[:1,1:3]

✔

Submit You have used 2 of 2 attempts

NumPy Slicing Element Equivalence

1/1 point (graded)

You are given the following lines of code:

```
arr = np.array([[1,2,3],[4,5,6],[7,8,9]])
slice = arr[:2,1:3]
slice[0,0]
```

What element in arr is equivalent to slice[0,0]?

- [ ] arr[0,0]
- [x] arr[0,1] ✔
- [ ] arr[2,2]
- [ ] arr[0,2]

Submit You have used 2 of 2 attempts

NumPy Function

1/1 point (graded)

What is the result of the following lines of code?

```
a=np.array(["cat","dog","fish"])
b=np.array(["dog","fish","rabbit"])
print(np.setdiff1d(b,a))
```

- ['cat']
- ['rabbit'] ✔
- ['dog' 'fish']
- ['cat' 'dog' 'fish' 'rabbit']

Submit

You have used 2 of 2 attempts

Broadcasting

0/1 point (graded)

What is the output of the following broadcasting call?

```
A = np.array([1,2,3,4])
B = np.array([[1,2],[3,4]])
A + B
```

- ◉ array([[2, 4], [6, 8]]) ✖
- ○ array([[2, 4, 3, 4], [4, 5, 3, 4]])
- ○ array([[1, 2], [3, 4]])
- ○ Value Error

Submit You have used 2 of 2 attempts

Pandas Function

1/1 point (graded)

Which of the following statistics does the `describe` function show?

- ☑ mean()
- ☑ min()

- [ ] mode()
- [x] max()
- [x] std()

✔

Submit You have used 2 of 2 attempts

Pandas Timestamp

1/1 point (graded)

What is the function call to find cells in a dataframe df with timestamp on 2007-02-04, given the dataframe has a parsed time column labelled ‘parsed_time’?

- (•) df['parsed_time'] == '2007-02-04' ✔
- ( ) df['parsed_time'] = '2007-02-04'
- ( ) '2007-02-04' >= df['parsed_time'] >= '2007-02-04'
- ( ) '2007-02-05' > df['parsed_time'] > '2007-02-03'

Submit You have used 2 of 2 attempts

Pandas Series Data Access

1/1 point (graded)

Suppose the pandas.core.series.Series `ser` has indices 'apple' and 'orange'. Which of the following is NOT a correct way to access data at a certain location in the `ser` Series?

- ser.loc['apple']
- ser.loc['apple','orange'] ✔
- ser.loc[['apple']]
- ser.loc[['apple','orange']]

Submit

You have used 2 of 2 attempts

Pandas Grouping

1/1 point (graded)

Suppose you have a pandas DataFrame called `df` which has a column called 'time'. Which function call will allow you to group the dataframe by 'time'?

- (•) df.groupby(['time']) ✔
- ( ) groupby(df['time'])
- ( ) df.aggregate(['time'])
- ( ) df['time'].group()

Submit You have used 2 of 2 attempts

Conceptual Visualization

1/1 point (graded)

Which is an example of conceptually driven data visualization?

- (•) Physicists use a visualization to teach students the well-known relationship between force and acceleration. ✔
- ( ) Doctors try to explore the relationship between a drug and the effect it has on their patients using data visualization.

- Realtors visualize a data set containing rental listings and the amount of interest they attract.

Submit You have used 2 of 2 attempts

Data Visualization Qualities

1/1 point (graded)

What are the qualities of good data visualization, according to Andy Kirk?

- [x] Trustworthy
- [x] Accessible
- [x] Elegant
- [ ] Meaningful

✔

Submit You have used 2 of 2 attempts

Correlation Visualization

1/1 point (graded)

Which graphing method should you use to visualize the correlation between two arrays?

- Histogram
- Barplot
- Scatter plot ✔
- Line plot

Submit

You have used 2 of 2 attempts

Machine Learning Approaches

1/1 point (graded)

What is true between supervised and unsupervised approaches?

- In supervised approaches, the target is unavailable. In unsupervised approaches, the target is unavailable.
- In supervised approaches, the target is provided. In unsupervised approaches, the target is provided.

- In supervised approaches, the target is unavailable. In unsupervised approaches, the target is provided.
- In supervised approaches, the target is provided. In unsupervised approaches, the target is unavailable. ✔

Submit You have used 2 of 2 attempts

Machine Learning Decision Trees

1/1 point (graded)

In a decision tree, which nodes do NOT have test conditions?

- Root nodes
- Internal nodes
- Leaf nodes ✔

Submit You have used 2 of 2 attempts

Machine Learning Function

0/1 point (graded)

True or False?: Not specifying the parameter `random_state` in the `train_test_split` function for every run will output the same result.

- ◉ True ✖
- ○ False

Submit You have used 2 of 2 attempts

Machine Learning Prediction

1/1 point (graded)

When is a prediction task referred to as simple linear regression?

- ◉ When there is only one input variable. ✔
- ○ When there are two input variables.
- ○ When there are more than two input variables.

Submit You have used 2 of 2 attempts

Data Retrieval

0.5/1 point (graded)

What are the 2 components of data retrieval mentioned in this class?

- [x] The way you store specific data in a data management system.
- [x] The way you specify how to get the desired data out of the relational data store.
- [ ] The internal processing that occurs within the data management system to compute or evaluate that specified retrieval request.
- [ ] How to handle the data once it has been retrieved.

Submit You have used 2 of 2 attempts

Split Function

1/1 point (graded)

Assume the code below:

```
text = 'New York-based'
```

What would be the output of `text.split()`?

- ['New', ' ', 'York-based']
- ['New', 'York', '-', 'based']
- ['New', 'York', 'based']
- ['New', 'York-based'] ✔

Submit You have used 2 of 2 attempts

JSON

1/1 point (graded)

Given the following code

```
import json
status = ['like','excited','wow','dislike']
```

What is the correct way to use json to print the first three elements in the list `status`?

- print(json.dumps(status[4], indent = 1))
- print(json.dumps(status[3], indent = 1))
- print(json.dumps(status[1:4], indent = 1))
- print(json.dumps(status[0:3], indent = 1)) ✔

Submit You have used 2 of 2 attempts

Distribution

1/1 point (graded)

If you find a peak distribution when plotting word frequency, what does this tell you about the vocabulary which produced that distribution?

- There are many unique words
- There is a large vocabulary
- There is a focused topic ✔

Submit You have used 2 of 2 attempts

Congratulations on completing your final exam for DSE 200X!