**CSE 5345/7345**
**Fall 2018**
**Midterm Exam - Overview Sheet**

Time: 2 hr 30 minutes
Open book, notes and computer

Read over the entire exam before jumping in. Pace yourself so you do not run out of time.

## Part 1. Multiple Choice (25 points)

Take the Canvas quiz : MidtermQuiz with 25 questions

Note that Questions 7-14 may have multiple answers.

## Part 2. Live Coding with Jupyter Notebook

- Download the Jupyter Notebook MidTerm.ipynb from Canvas
  - A pdf version is also available
- Download Olympics.csv which is needed for the Panda questions
- Add your code to MidTerm.ipynb
- Submit to Canvas Assignment Midterm Link

**Notes:**

Q9. Beautiful Soup

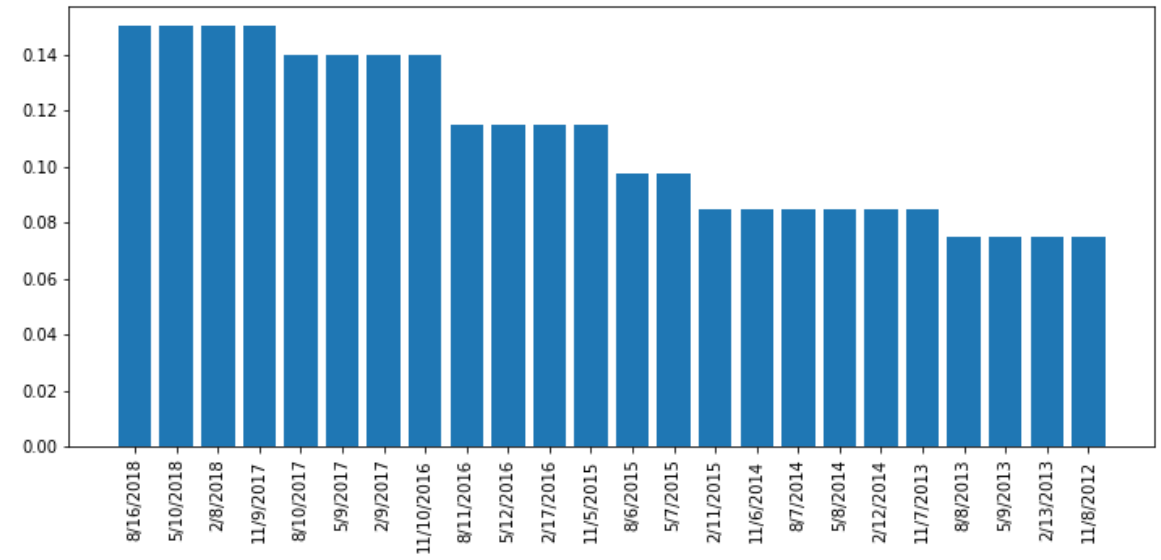
Using Jupyter, load the page at: https://www.nasdaq.com/symbol/nvda/dividend-history

You will find a list of dividend payouts and a variety of dates.

A) Extract all the dividends (**Cash Amount** - column 3) and the corresponding **Declaration Date** (Column 4) listed on the website.

B) Load the data into a DataFrame and **display with two columns**: **Declaration Date** and **Cash Amount** sorted by ascending date. There will be 24 rows of data.
   a. To obtain full credit for this assignment you should **NOT hardcode 48 ids** into your code. Use id fields to navigate to the parts of the page you need and generate the Ids based on the naming pattern of the ids. You may assume there are 24 pairs of ids.

C) Plot the dividends over time and generate the following graph.



## Q10 Panda Exercise 1

Create a DataFrame that looks like the following:

| | one | two | three | four |
|---|---|---|---|---|
| NY | 0 | 1 | 2 | 3 |
| TX | 4 | 5 | 6 | 7 |
| CA | 8 | 9 | 10 | 11 |
| OK | 12 | 13 | 14 | 15 |

Modify so that all values < 5 are replaced with zeros and all other values remain the same, as in:

| | one | two | three | four |
|---|---|---|---|---|
| NY | 0 | 0 | 0 | 0 |
| TX | 0 | 5 | 6 | 7 |
| CA | 8 | 9 | 10 | 11 |
| OK | 12 | 13 | 14 | 15 |

**Q11: Panda Exercise 2**

# Q in one line, print the product of the two values circled below

| | one | two | three | four |
|---|---|---|---|---|
| NY | 0 | 0 | 0 | 0 |
| TX | 0 | 5 | 6 | 7 |
| CA | 8 | 9 | 10 | 11 |
| OK | 12 | 13 | 14 | 15 |

**Q12 Panda EXERCISE 3**

# create the following dataframe

| | a | b | c | d | e |
|---|---|---|---|---|---|
| 0 | 10 | 12 | 13 | 14 | 15 |
| 1 | 10 | 20 | 30 | 40 | 50 |
| 2 | 2 | 3 | 44 | 15 | 66 |

Then display the max values for each column as:

```
a    10
b    20
c    44
d    40
e    66
dtype: int64
```

And display the max values for rows as:

```
0    15
1    50
2    66
dtype: int64
```

**Q13 Olympics 1**

Download the Olympics.csv and do the following :

- Display the count of the number of Hungarians (HUN) who won medals between 2000 and 2008 in the Discipline of Swimming or Fencing

**Q14 Olympics 2**

- In which events did Jesse Owens win medals - report as a DataFrame with column headings of: Year, City and Event