# Assignment 5

- Please write your code in a Jupyter Notebook file (.ipynb) and submit it on Blackboard
- Make sure to label your answer of each question clearly and add comments to make it readable.
- You are allowed to discuss with other students (up to three) or the instructor. Please put all the names of students that you discussed with. However individual students must write their own solutions.
- Copying a program, or letting someone else copy your program, is a form of academic dishonesty . Any referred material must be cited properly.
- Maximally leverage Piazza to benefit other students by your questions and answers.
- Try to be updated by checking notifications in both Piazza and Blackboard class webpage.

## Exercise 1. Use UIC2016MensBasketBall dataset. (50 pts)

### 1. DataFrame from CSV

The *UIC2016MensBasketBall.csv* records the game results of the 2016 Men's Basketball. Implement Python code to load the exact data into a data frame **df2016**.

### 2. Column Names

Assign column headers to:

*Date*

*Opponent*

*UIC Score*

*Opp Score*

*UIC Field Goal Percentage*

*Opp Field Goal Percentage*

*UIC 3 point Field Goal Percentage*

*Opp 3 point Field Goal Percentage*

*UIC Rebound*

*Opp Rebound*

*UIC Assists*

*Opp Assists*

**3. Missing data**

Fill the missing data cell with a hyphen symbol '-'.

**4. Data Types**

What are the data types used in this data frame? Provide both Python code and the answer.

**5. Count**

How many school opponents did Men's Basketball played in 2016? Provide both Python code and the answer.

**6. Filter**

List all the games that UIC scored more than 65 points. Provide both Python code and the answer.

**7. Win-Los-Tie**

Calculate numbers of games of Wins, Losses and Ties, accordingly.

## Exercise 2. Use customer-savings dataset (20 pts)

Data columns in customer-savings.txt (separated by comma):

| Customer ID | Customer Name | Customer Surname | Gender | Age | Region | Job Classification | Date joined | Balance |
|---|---|---|---|---|---|---|---|---|
| 100000001 | Simon | Walsh | Male | 21 | England | White Collar | 05.Jan.15 | 113810.15 |
| 100000003 | Liam | Brown | Male | 46 | England | White Collar | 07.Jan.15 | 101536.83 |

Answer the following questions:

1. Load data and add column names.
   (Hint: To load data from .txt with pandas, you can use *data = pd.read_csv('test.txt', sep=",")* )
2. What's the average balance for male and female? Provide both Python code and the answer
3. What's the average balance for while collar and blue collar in England? Provide both Python code and the answer

## Exercise 3: Use customer-status.csv and sales.csv (30 pts)

Data columns in customer-status.csv (separated by comma):

| Account Number | Name | Status |
|---|---|---|
| 527099 | Sanford and Sons | bronze |

Data columns in sales.csv (separated by comma):

| **Account Number** | **Name** | **SKU** | **Quantity** | **Unit Price** | **Ext Price** | **Date** |
|---|---|---|---|---|---|---|
| 163416 | Purdy-Kunde | S1-30248 | 19 | 65.03 | 1235.57 | 2014-03-01 16:07:40 |
| 527099 | Sanford and Sons | S2-82423 | 3 | 76.21 | 228.63 | 2014-03-01 17:18:01 |

Perform the following merge operations on the two dataframes, and report the total number of rows of the combined dataframes:

1. Inner join on column “Account Number”.
2. Full outer join on column “Account Number”.
3. Left join on column “Account Number”, using customer-status.csv as the base.
4. Left join on column “Account Number”, using sales.csv as the base.