

# CP 321 - Data Visualization

## Assignment 2

Check dropbox for due date

### Tools/Software Requirement

- An IDE for writing Python code
- Use Python and Pandas library to complete this assignment.
- This assignment may require some concepts that are not covered in the class. Students are expected to research those topics on their own and use them to complete the assignment. Self-learning will be a goal for all the assignments and the project in this course.

**Tasks:** All the tasks must be complete programmatically. Manually editing the dataset is not allowed. I encourage you to use jupyter notebook for this assignment. Each task is 2 points.

1. Download the dataset csv file (**Assignment2Data.csv**) and load it in your program. The dataset contains information about badminton players in the Waterloo Region.
2. Display the head of the dataset.
3. Display 5 random rows from the dataset.
4. Write code to find any record (row) where the number of wins is more than the matches.
5. Delete the rows found in task 4.
6. Create a column called **Lost** and populate it with the number of matches lost by each player.
7. Create a column called **Win percentage** and populate it with win percentage for each player
8. Create a column called **Category**. Populate this column using the following rules:
  - a. **Professional:** Win percentage greater than 90
  - b. **Target:** Win percentage greater than 75 but less than 91
  - c. **Development:** Win percentage less than 76
9. Display the player ID of the player with highest win percentage
10. Display the player ID of the player with lowest win percentage
11. Group the players based on their **Category** and find the best player in each category (**Professional, Target, Development**)
12. Find and list which age group is doing the best in Waterloo region. The age groups are: 20 and below, between 21 and 25, between 26 and 30, 30 and above.
13. Find win percentage for each city.
14. Find the combined win percentage for all the players above 20 years of age.
15. Display the dataset sorted based on **Matches**.

**Submission Details.**

1. Create a folder named: FIRSTNAME\_LASTNAME.
2. Place your code and the dataset file in that folder.
3. Compress the folder into a zip file. Please make sure that your folder is appropriately named (See step 1 above) before creating the zip file. Do not rename the zip file after creating it.
4. Upload the zip file in the appropriate dropbox on MLS.
5. Redownload the file you submitted and confirm if you uploaded the correct file on MLS.