

# Assignment-1

[Specification](#)[Make Submission](#)[Check Submission](#)[Collect Submission](#)

The assignment data was extracted from the Wikipedia entry on All Time Olympic Games Medals ([https://en.wikipedia.org/wiki/All-time\\_Olympic\\_Games\\_medal\\_table](https://en.wikipedia.org/wiki/All-time_Olympic_Games_medal_table)) , with some minor modification to make things interesting. The dataset is split into two CSV files one of **summer games** and the other for the **winter games** and combined total. Here you can find the datasets: Summer Games ([https://github.com/mysilver/COMP9321-Data-Services/blob/master/assignments/Olympics\\_dataset1.csv](https://github.com/mysilver/COMP9321-Data-Services/blob/master/assignments/Olympics_dataset1.csv)) and Winter Games ([https://github.com/mysilver/COMP9321-Data-Services/blob/master/assignments/Olympics\\_dataset2.csv](https://github.com/mysilver/COMP9321-Data-Services/blob/master/assignments/Olympics_dataset2.csv)) . Use the datasets to answer the following questions:

- **Question 1: (based on the both datasets)**

Merge the two datasets Olympics\_dataset1.csv and Olympics\_dataset2.csv and display the first five rows (do not concatenate the datasets and ignore the first row of each of the datasets). And rename columns of the result dataframe as follow (remove the rest of columns from the result dataset):

*Country, summer\_rubbish, summer\_participation, summer\_gold, summer\_silver, summer\_bronze, summer\_total, winter\_participation , winter\_gold, winter\_silver, winter\_bronze, winter\_total*

Remove the "Totals" row from the dataframe

- **Question 2: (based on the dataframe created in Question-1)**

Rename country to only keep the name of the country without the abbreviations (Afghanistan (AFG) --> Afghanistan) . Set the index as the country name, remove below columns and then display the first 5 countries in the Dataframe. :

*summer\_rubbish, summer\_total, winter\_rubbish, winter\_total*

- **Question 3: (based on the dataframe created in Question-2)**

Remove the rows with NaN fields and display the last 10 rows.

- **Question 4: (based on the dataframe created in Question-3)**

Calculate and display which country has won the most **gold medals** in summer games? (Just print the country name)

- **Question 5 : (based on the dataframe created in Question-3)**

Calculate and display which country had the biggest difference between their summer and winter gold medal?

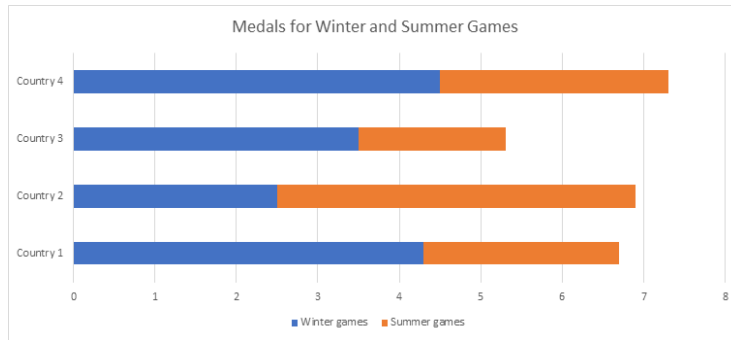
- **Question 6: (based on the dataframe created in Question-3)**

Sort the countries in descending order, according to the number of total of medals (summer and winter) and display the first and last 5 rows.

- **Question 7: (based on the dataframe created in Question-3)**

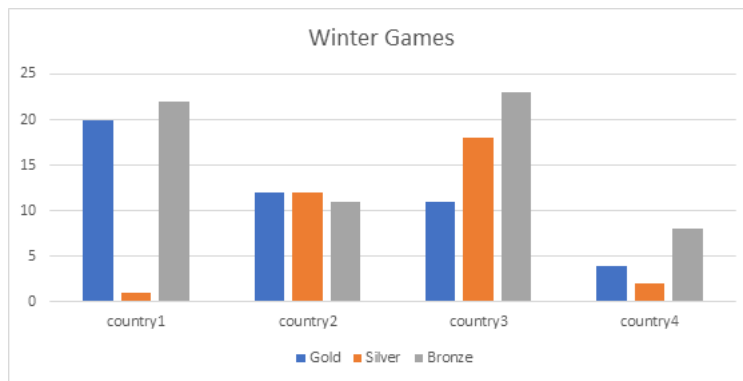
Plot a bar chart of the top 10 countries ordered by the number of total of medals (summer and winter). For each country use a stacked bar chart showing for each county the total medals for winter and summer games.

See example chart below:



• **Question 8 : (based on the dataframe created in Question-3)**

Plot a bar chart of the countries (United States, Australia, Great Britain, Japan, New Zealand). For each country you need to show the gold, silver and bronze medals for winter games. See example below of the chart:



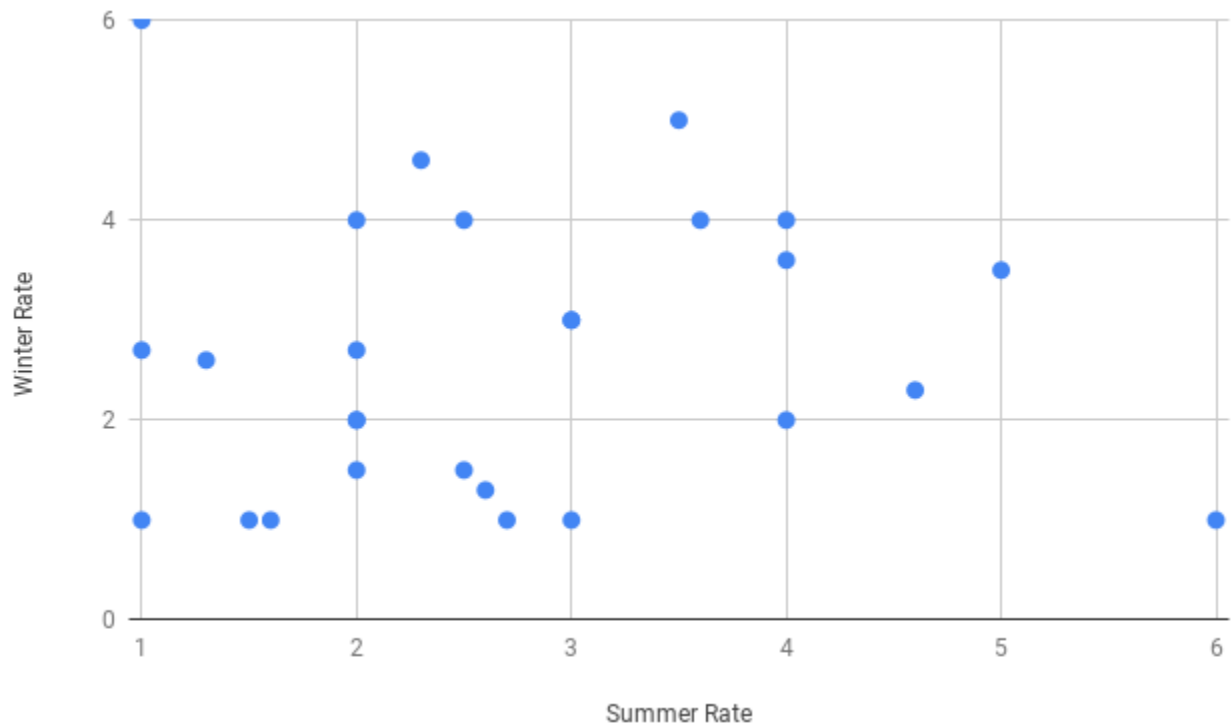
• **Question 9: (based on the dataframe created in Question-3)**

Assume that there countries are ranked based on a new ranking scheme for the summer games. In the new ranking scheme, a Gold, Silver, and Bronze medals have 5, 3, and 1 points respectively. And countries are ranked based on the points eared per *participation* (total points divided by total number of *participations* in games). Based on this scheme, rank the countries and print the name of top 5 countries having the best rate (points per *participation* ).

Example: Imagine that a country has 1 gold medal, and 1 silver (5 + 3 points = totally 8 points); and this country had 10 *participation* in the summer games. The rate of points per *participation* will be  $8 / 10 = 0.8$  for this country.

• **Question 10: (based on the dataframe created in Question-3)**

Based on the raking scheme in Question 9, also calculate the points per *participation* for each country in the Winter Games. Next plot a scatter chart with x = "points per *participation* for summer games" , and y = "points per *participation* for winter games" . Here is an example of such a chart:



however, you also need to ink bubbles based on their continents (e.g, Asia red, Africa blue, ..). You may use the Country-Continent dataset (<https://github.com/mysilver/COMP9321-Data-Services/blob/master/assignments/Countries-Continents.csv>) to colour the countries. Your chart must also have legends and labels showing the name of countries beside the points inside the chart.

## Submission Guideline:

**Due Date:** Friday the 25th of October 2019 23:59

Submit your script named "YOUR\_ZID .py" (z2123232.py) which contains your code.

Define a function for each question separately and name them like: *question\_1()* , *question\_2()* , etc. Here is an example:

```
import pandas as pd
...
def question_1()
    print("-----Question 1 -----")
    ...
def question_2()
    print("-----Question 2 -----")
    ...
...
def question_10()
    print("-----Question 10 -----")
    ...
if __name__ == "__main__":
    question_1()
    question_2()
    ...
    question_10()
```

you can download the code template from : <https://raw.githubusercontent.com/mysilver/COMP9321-Data-Services/master/assignments/z1111111.py>  
(<https://raw.githubusercontent.com/mysilver/COMP9321-Data-Services/master/assignments/z1111111.py>)

if you do not follow this structure, you will be penalised.

---

## FAQ:

- **Can I pass variables to functions?**  
YES
  - **Shall I skip the first row (team, summer games) in the datasets?**  
YES
  - **Can we create our own functions besides the question functions (e.g., question\_1)?**  
YES
  - **Can I call another function inside the question functions? e.g., calling question\_1 inside question\_2**  
YES
  - **What should I do if my charts are not shown automatically?**  
Look at the lab sample codes; if still need a help, ask your tutor during the tutorials.
  - **How should I print my dataframe?**  
Use:  
`print(df.to_string())`
  - **Is it okay that the graph for Q8 does not pop up until the graph for Q7 is closed or should they both pop up at the same time?**  
This is fine
  - **In Question 8, does the order of countries in the plot matter? i.e. should country 1 always be United States, then Australia, Great Britain, and so on or can we plot in alphabetical order of country names. So, country 1 = Australia, country 2 = GB, etc**  
No, the order is not important
  - **Do the charts in q8 and q9 need to look the same (colors, legend position, grid) as the examples shown? or would it be fine to just use the default plotting from pandas?**  
The default colours/fonts are fine
  - **How are our submissions marked?**  
They are marked manually by tutors, by running the following command: `python3 z{YOUR_ZID}.py`
  - **What python packages can I use in my assignment?**  
You can only use pandas and matplotlib to do the assignment.
  - **What version of python should I use?**  
Python 3+
  - **How I can submit my assignment?**  
Go to the assignment page click on the "Make Submission" tab; pick your files which must be named "YOUR\_ZID.py". Make sure that the files are not empty, and submit the files together.
  - **When is Assignment 1 due?**  
the Assignment is due on Friday the 25th of October 2019 23:59
  - **Can I submit my file after deadline?**  
Yes, you can. But 20% of your assignment will be deducted as a late penalty per day. In other words, if you be late for more than 4 days, you will not be marked.
- 

## Plagiarism

This is an *individual assignment*. The work you submit must be your own work. Submission of work partially or completely derived from any other person or jointly written with any other person is not permitted. The penalties for such offence may include negative marks, automatic failure of the course and possibly other academic discipline. Assignment submissions will be examined manually.

Do not provide or show your assignment work to any other person - apart from the teaching staff of this course. If you knowingly provide or show your assignment work to another person for any reason, and work derived from it is submitted, you may be penalized, even if the work was submitted without your knowledge or consent. Pay attention that it is also your duty to protect your code artifacts. If you are using any online solution to store your code artifacts (e.g., GitHub) then make sure to keep the repository private and do not share access to anyone.

**Reminder:** Plagiarism is defined as (<https://student.unsw.edu.au/plagiarism>) using the words or ideas of others and presenting them as your own. UNSW and CSE treat plagiarism as academic misconduct, which means that it carries penalties as severe as being excluded from further study at UNSW. There are several on-line sources to help you understand what plagiarism is and how it is dealt with at UNSW:

- Plagiarism and Academic Integrity (<https://student.unsw.edu.au/plagiarism>)
- UNSW Plagiarism Procedure (<https://www.gs.unsw.edu.au/policy/documents/plagiarismprocedure.pdf>)

Make sure that you read and understand these. Ignorance is not accepted as an excuse for plagiarism. In particular, you are also responsible for ensuring that your assignment files are not accessible by anyone but you by setting the correct permissions in your CSE directory and code repository, if using one (e.g., Github and similar). Note also that plagiarism includes paying or asking another person to do a piece of work for you and then submitting it as your own work.

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW staff and students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW.

Resource created 3 days ago (Wednesday 09 October 2019, 01:59:37 PM), last modified about 11 hours ago (Saturday 12 October 2019, 12:34:42 AM).

## Comments

 [Q \(/COMP9321/19T3/forums/search?forum\\_choice=resource/36346\)](/COMP9321/19T3/forums/search?forum_choice=resource/36346)

 [\(/COMP9321/19T3/forums/resource/36346\)](/COMP9321/19T3/forums/resource/36346)

 Add a comment



Wenda Yan (/users/z5187136) 35 minutes ago (Sat Oct 12 2019 10:34:36 GMT+1100 (澳大利亚东部夏令时间))

Hi

One more question, in q6 it asks us to show the first and last 5 rows. However I have created a new column named 'Total' for solving this question. So when I print the result, do I need to drop the column what I created before?

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) 3 minutes ago (Sat Oct 12 2019 11:07:00 GMT+1100 (澳大利亚东部夏令时间))

Hi,

You need to print the total number of medals as well.

Reply



Wenda Yan (/users/z5187136) about an hour ago (Sat Oct 12 2019 10:10:25 GMT+1100 (澳大利亚东部夏令时间))

Hi

Can I replace the ',' in columns with ''. So I can change the data type to calculate something.

For example, exchange '1,022' to '1022'

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about an hour ago (Sat Oct 12 2019 10:13:17 GMT+1100 (澳大利亚东部夏令时间))

Yes, you need to convert those columns to numeric columns

Reply



Wenda Yan (/users/z5187136) about 11 hours ago (Sat Oct 12 2019 00:39:07 GMT+1100 (澳大利亚东部夏令时间))

Hi

I wanna make sure one thing. Will there always be a space between the squares and last character?

For example: Afghanistan (AFG) --> Afghanistan. There is a space between 'n' and '('.

Thanks

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 3 hours ago (Sat Oct 12 2019 08:34:51 GMT+1100 (澳大利亚东部夏令时间))

Hi,

You need to check the dataset and names one by one to make sure.

Reply



Ruiling Wang (/users/z5202464) about 11 hours ago (Sat Oct 12 2019 00:13:26 GMT+1100 (澳大利亚东部夏令时间))

For Question 1, is it necessary to remove the first row (Team, Summer Games, Team Winter Games) before we print the first five rows of the data frame? I am confused about that.

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 11 hours ago (Sat Oct 12 2019 00:16:19 GMT+1100 (澳大利亚东部夏令时间))

You need to skip the first row

Reply



Ruiling Wang (/users/z5202464) about 10 hours ago (Sat Oct 12 2019 01:06:51 GMT+1100 (澳大利亚东部夏令时间))

so Is the print() function carried after rename columns finished?

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 3 hours ago (Sat Oct 12 2019 08:37:04 GMT+1100 (澳大利亚东部夏令时间))

You need to print the dataframe at the end of each function

You will be marked manually based on your output.

Reply



Wenda Yan (/users/z5187136) about 12 hours ago (Fri Oct 11 2019 23:32:59 GMT+1100 (澳大利亚东部夏令时间))

About Question1

The specification said that we should use to\_string() to print our dataframe. However I find that the result of print is not clear, the it can not print all the names of columns. Are we still need to use this or the print function which the tutor had used.

By the way, Can I pass variables in these question functions? For example, if the q2 wanna use the dataframe from q1. I can not directly call the q1 function in q2 as q1 will print the five rows of dataframe again. So can I create some variables in main function and pass the to questions?

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 11 hours ago (Sat Oct 12 2019 00:19:53 GMT+1100 (澳大利亚东部夏令时间))

It means that you have not skipped the first row of the dataset. If you skip it, then the column names will be fine, and print will be as expected

- Yes you can create and pass variables

Reply



Saurabh Acharya (/users/z5130134) about 12 hours ago (Fri Oct 11 2019 23:09:49 GMT+1100 (澳大利亚东部夏令时间))

In the Winter Games dataset, there isn't a 'Rubish' column like the Summer Games dataset has. Is this intentional? Question 1 asks us to rename one of the columns to 'winter\_rubbish', however that dataset doesn't have a 'Rubish' column to begin with, hence my confusion.

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 11 hours ago (Sat Oct 12 2019 00:31:18 GMT+1100 (澳大利亚东部夏令时间))

Thanks for noting. This is a mistake. Please ignore "winter\_rubbish"

Reply

