Resources / Assignments (/COMP9321/19T3/resources/36342)
/ Week 4 (/COMP9321/19T3/resources/36343) / Assignment-1

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) , 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) and Winter Games (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 (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
  Display the first five rows of 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 rows in the Dataframe :

  summer_rubbish, summer_total , 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 name had the biggest difference between their summer and winter gold medal? (Just print the country name and the difference)
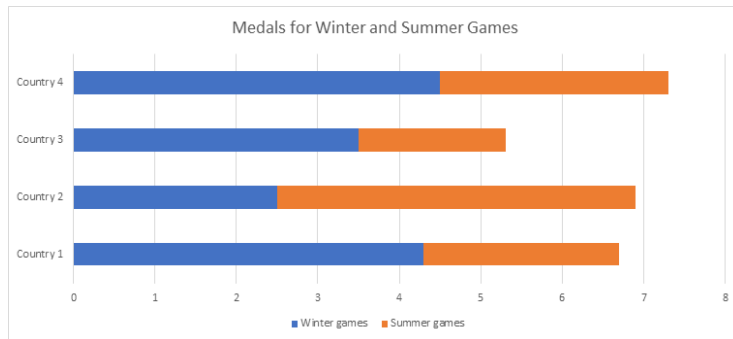- **Question 6: (based on the dataframe created in Question-3)**
  Sort the countries in descending order, according to the number of total of medals (summer and winter) and display the first and last 5 rows of the dataframe (including a column showing the total number of medals) .
- **Question 7: (based 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
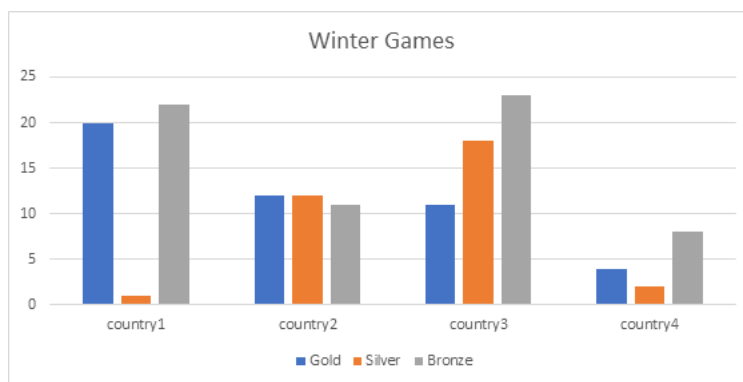
summer games.
See example chart below:

- **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 county you need to show the gold, silver and bronze medals for winter games. See example below of the chart:
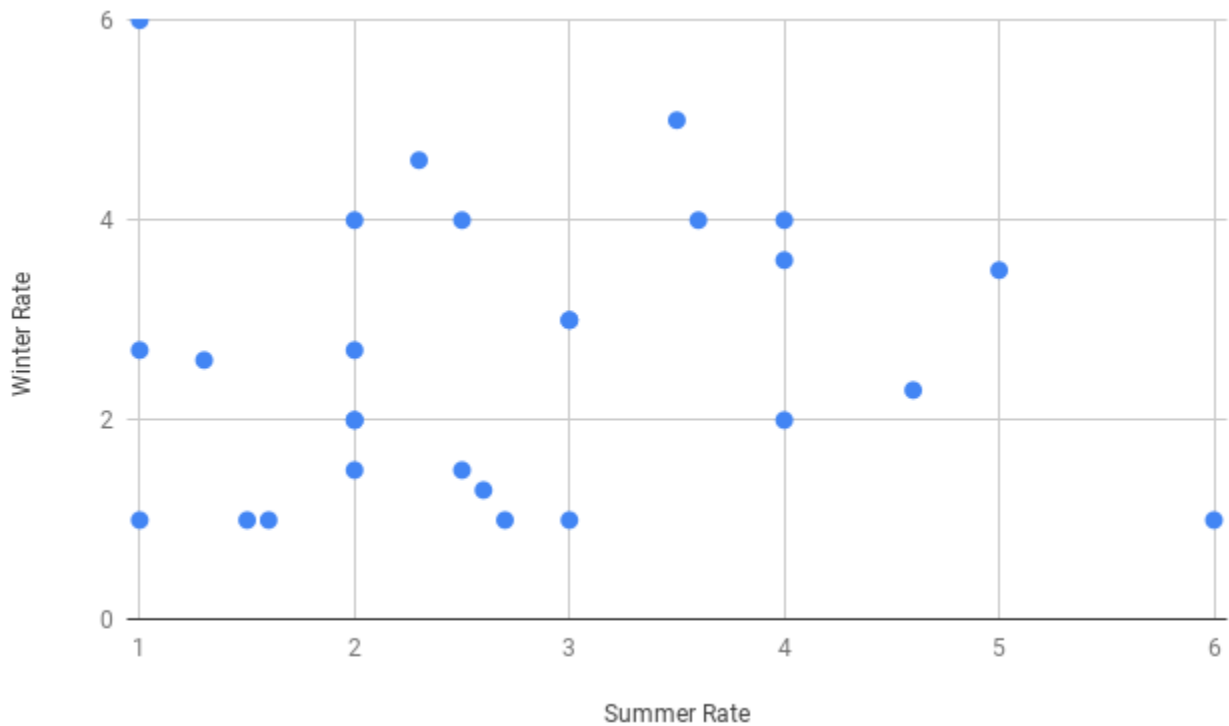

- **Question 9: (based on the dataframe created in Question-3)**
  Assume that there countries are ranked based on a new ranking scheme for the summer games. In the new ranking scheme, a Gold, Silver, and Bronze medals have 5, 3, and 1 points respectively. And countries are ranked based on the points 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*). Print both country names and rates.

  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. (if per *participation is 0, the rate should be 0;)

- **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 the 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 counties, and use a default color (Gray) for countries which are not listed in the dataset. You chart must also have legends and labels showing the name of countries beside the points inside the chart.

Submission Guideline:

Due Date: Monday the 28th of October 2019 17:59

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 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 reposiroty 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 16 days ago (Wednesday 09 October 2019, 01:59:37 PM), last modified 2 days ago (Wednesday 23 October 2019, 05:47:15 PM).

Comments

🔖 🔍 (/COMP9321/19T3/forums/search?forum_choice=resource/36346)

💬 (/COMP9321/19T3/forums/resource/36346)

💬 Add a comment

Usman Haidar (/users/z5062617) a day ago (Thu Oct 24 2019 13:34:40 GMT+1100 (澳大利亚东部夏令时间))

Hi for question 7: "Plot a bar chart of the top 10 countries ordered by the number of total of medals (summer and winter)"

Does this mean ascending or descending order?

Thanks

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) a day ago (Thu Oct 24 2019 13:44:55 GMT+1100 (澳大利亚东部夏令时间))

Hi, **TOP** means descending.

Reply

Mohammad Rahman (/users/z3119277) a day ago (Thu Oct 24 2019 10:37:59 GMT+1100 (澳大利亚东部夏令时间))

Hi Mohammadali,

I figured out that you cann't simply see the graph from remote ssh like as I did logging from my mac to cse server.

There are some ways you can but need to install some packages and config as described

here (https://www.pyimagesearch.com/2015/08/24/resolved-matplotlib-figures-not-showing-up-or-displaying/)

I tested my code on piano lab computer and graphs just pop up fine. I think that is what expected from us.

Please make it clear.

Thanks

Reply

> **Mohammadali Yaghoubzadehfard (/users/z5138589)** a day ago (Thu Oct 24 2019 10:46:32 GMT+1100 (澳大利亚东部夏令时间))
>
> Hi,
>
> Your are fine then. We will not use ssh!
>
> Reply

> **Mohammad Rahman (/users/z3119277)** a day ago (Thu Oct 24 2019 11:23:53 GMT+1100 (澳大利亚东部夏令时间))
>
> Thanks a lot
>
> Reply

Jinpeng Jiao (/users/z5143932) 2 days ago (Thu Oct 24 2019 00:44:12 GMT+1100 (澳大利亚东部夏令时间))

Hi Mohammad Ali,

Just wanna know do we return the results of each function or just print them , if returning result needed, how can i return the result of Q8 and Q10.Thanks

Reply

> **Mohammadali Yaghoubzadehfard (/users/z5138589)** a day ago (Thu Oct 24 2019 08:09:29 GMT+1100 (澳大利亚东部夏令时间))
>
> Hi,
>
> Just print or show the graph (pop up)
>
> Reply

Rakan Frijat (/users/z3218473) 2 days ago (Wed Oct 23 2019 20:23:56 GMT+1100 (澳大利亚东部夏令时间))

Hi there, I am getting an error when running my code on CSE machine. Essentially, it's not allowing me to run code df.columns.to_list().

The reason for the above is that I am running a newer version of pandas on my laptop compared to CSE. When I update pandas on CSE, my code works fine.

Are we able to submit our code based on the latest pandas version? In the specs, there was no specification that we had to use an older pandas version.

Reply

Mohammad Rahman (/users/z3119277) 2 days ago (Wed Oct 23 2019 12:41:43 GMT+1100 (澳大利亚东部夏令时间))

Anybody able to see graphs popup on cse machine?

Reply

Victor Tse (/users/z5075018) 2 days ago (Wed Oct 23 2019 14:25:54 GMT+1100 (澳大利亚东部夏令时间))

If you're connecting to CSE from an ssh session, can you try setting up X11 forwarding first?

Reply

Mohammad Rahman (/users/z3119277) 2 days ago (Wed Oct 23 2019 14:50:07 GMT+1100 (澳大利亚东部夏令时间)), last modified 2 days ago (Wed Oct 23 2019 15:25:16 GMT+1100 (澳大利亚东部夏令时间))

I'm connecting from mac with ssh.

How do we set up x11 forwaing on cse machine? It needs config changes?

I am afraid we are allowed to do that.

Reply

Victor Tse (/users/z5075018) a day ago (Thu Oct 24 2019 10:10:35 GMT+1100 (澳大利亚东部夏令时间))

Correct me if I'm wrong, but I think it's more so that you need to enable X11 forwarding on your Mac, so that your Mac can visually display what the ssh session is trying to show you. No configurations needed on the CSE machine.

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) 2 days ago (Wed Oct 23 2019 12:51:45 GMT+1100 (澳大利亚东部夏令时间))

Take a look at the code:
https://github.com/mysilver/COMP9321-Data-Services/blob/master/Week4_Visualization/activity_1.py
(https://github.com/mysilver/COMP9321-Data-Services/blob/master/Week4_Visualization/activity_1.py)

Reply

Mohammad Rahman (/users/z3119277) 2 days ago (Wed Oct 23 2019 14:43:13 GMT+1100 (澳大利亚东部夏令时间)), last modified 2 days ago (Wed Oct 23 2019 14:48:09 GMT+1100 (澳大利亚东部夏令时间))

My problem is not creating graph but showing up in cse machine.

I copied my code along with the csv files to a cse machine and logged in using ssh zid@login.cse.unsw.edu.au.

Then I ran the code from the command line as

python3 zid.py

It works fine except none of the graphs shows up.

I can save the graphs as png files.

I am not sure, do we need another program to install on linux to see graph.

They show pefectly on mac when runs from cmd.

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) 2 days ago (Wed Oct 23 2019 16:37:58 GMT+1100 (澳大利亚东部夏令时间))

Have you added the following lines of code:

```
import matplotlib.pyplot as plt
plt.show()
```
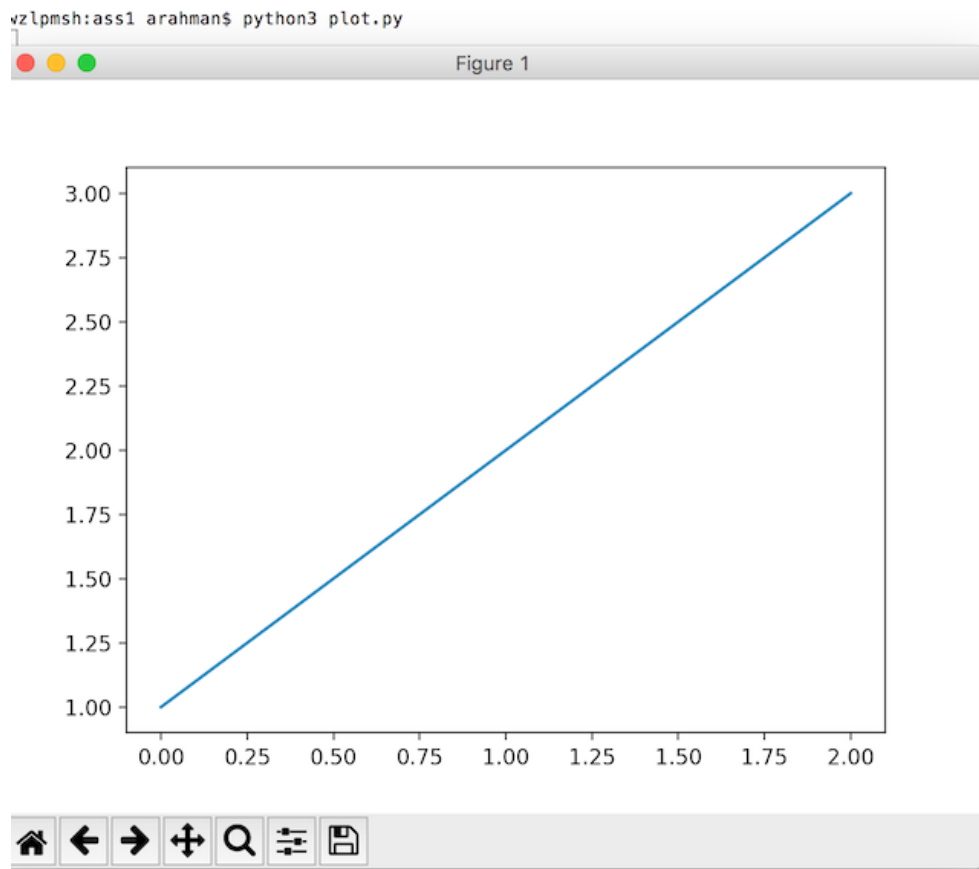
Reply

Mohammad Rahman (/users/z3119277) 2 days ago (Wed Oct 23 2019 17:24:13 GMT+1100 (澳大利亚东部夏令时间))

Yes, as I said it works on my mac from command line.

For example, this code snippet pops up a graph on mac but not cse server.

```
import matplotlib.pyplot as plt
plt.plot([1,2,3])
plt.show()
```

```
vzlpmsh:ass1 arahman$ python3 plot.py
```


but nothing shows on cse server

```
weill % python3 plot.py
weill %
```

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) 2 days ago (Wed Oct 23 2019 19:06:20 GMT+1100 (澳大利亚东部夏令时间))

I just check the whole code on cse machine. The sample code from the github, i mean.

It works. It means you need to inspect your code because cse machines are just fine

Reply

Mohammad Rahman (/users/z3119277) 2 days ago (Wed Oct 23 2019 21:05:57 GMT+1100 (澳大利亚东部夏令时间)), last modified 2 days ago (Wed Oct 23 2019 21:15:30 GMT+1100 (澳大利亚东部夏令时间))

Did you login from remote using ssh?

I logged into the cse server using ssh from my mac.

Then run the code from cmd. Even the code snippet I pasted above does not show any graph.

Also, I just copied the tutorial code as it is to the cse server and run the activity but same case.

Probably, we are missing something here.

Reply

Hoang Phan (/users/z5019338) 2 days ago (Wed Oct 23 2019 22:59:01 GMT+1100 (澳大利亚东部夏令时间))

when you login using ssh from the command line on your mac, it won't open any other windows (which is what happens when you plot a graph using matplotlib).
If you want to see the plots, then maybe logging into Vlab might work

Reply

Mohammad Rahman (/users/z3119277) a day ago (Thu Oct 24 2019 10:34:19 GMT+1100 (澳大利亚东部夏令时间)), last modified a day ago (Thu Oct 24 2019 10:35:22 GMT+1100 (澳大利亚东部夏令时间))

You right.You cann't see the graph from remote ssh.

There are some ways you can but need to install some packages and config as described

here (https://www.pyimagesearch.com/2015/08/24/resolved-matplotlib-figures-not-showing-up-or-displaying/)

I tested my code on piano lab and graphs just pop up fine. I think that is what expected from us.

Reply

Chengcheng Jiang (/users/z5226733) 3 days ago (Tue Oct 22 2019 18:57:15 GMT+1100 (澳大利亚东部夏令时间))

Hi Mohammadali,

I have used the dataset 'Countries-Continents.csv' for my Q10. So does the testing environment include the Country-Continent dataset (https://github.com/mysilver/COMP9321-Data-Services/blob/master/assignments/Countries-Continents.csv) as well?

Thanks

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) 3 days ago (Tue Oct 22 2019 20:10:15 GMT+1100 (澳大利亚东部夏令时间))

Yes

Reply

Mohammad Rahman (/users/z3119277) 3 days ago (Tue Oct 22 2019 17:01:14 GMT+1100 (澳大利亚东部夏令时间))

Hi Mohammadali (/users/z5138589) ,

I was testing my code on cse machine but graphs do not pop up at all.

Is that normal?

I copied my code along with the csv files to a cse machine and logged in using ssh zid@login.cse.unsw.edu.au.

Then I ran the code from the command line as

python3 zid.py

It works fine except none of the graphs shows up.

I can save the graphs as png files.

I am not sure, do we need another program to install on linux to see graph.

They show pefectly on mac when runs from cmd.

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) a day ago (Thu Oct 24 2019 08:10:30 GMT+1100 (澳大利亚东部夏令时间))

Hi,
Use Vlab, and take a look at the example code on tutorials.

Reply

Mohammad Rahman (/users/z3119277) 3 days ago (Tue Oct 22 2019 21:11:22 GMT+1100 (澳大利亚东部夏令时间)), last modified 3 days ago (Tue Oct 22 2019 21:12:06 GMT+1100 (澳大利亚东部夏令时间))

Hi Mohammadali (/users/z5138589) ,

Any comment on above issue?

Reply

Zekai Dong (/users/z5199242) 3 days ago (Tue Oct 22 2019 16:12:23 GMT+1100 (澳大利亚东部夏令时间))

so the ass's due is Monday?

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) 3 days ago (Tue Oct 22 2019 16:34:01 GMT+1100 (澳大利亚东部夏令时间))

Yes

Reply

Mohammad Rahman (/users/z3119277) <u>3 days ago (Tue Oct 22 2019 14:38:34 GMT+1100 (澳大利亚东部夏令时间))</u>

So we are just uploading a single file **zid.py** for submission?

We don't need to upload the dataset files along with it?

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) <u>3 days ago (Tue Oct 22 2019 14:45:06 GMT+1100 (澳大利亚东部夏令时间))</u>

Yes

Reply

Sachin Krishnamoorthy (/users/z5232031) <u>3 days ago (Tue Oct 22 2019 12:51:23 GMT+1100 (澳大利亚东部夏令时间))</u>

Hello,

For Q10 there are many countries that overlap, and thus have overlapping country labels. Is this okay?

Thanks!

Reply

Mohammadali Yaghoubzadehfard (/users/z5138589) <u>3 days ago (Tue Oct 22 2019 13:04:18 GMT+1100 (澳大利亚东部夏令时间))</u>

Hi,

That is fine

Reply

Load More Comments