Advanced Visualisations 2022: Project

Blerina Sinaimeri and Lorenzo Madeddu

Deadline: Only one member of your team should submit the work on luiss.learn by **23:59 (CET) on Monday, May 11.**

1 Description of the project

The project should be carried out in groups of 2-3 students. The group will analyze a dataset using the libraries of python we have seen.

Dataset choice: You can choose the dataset you want. The dataset you use does not need to be too large. However, it must contain interesting data and some structure. Some links when you can check for a dataset are the followings: http://www-personal.umich.edu/~mejn/netdata/ and http://snap.stanford.edu/data/.

Please notify us by email the dataset you chose and wait for our confirmation before starting the analysis. If you cannot chose a dataset let us know by email and we will assign one to you.

2 Expected outcome

The project will be valued **14 points** and will make for 40% of your final grade. Your team is required to submit :

(1) A folder containing the code related to the project and a pdf containing the results (figures, values, etc.) **OR** a single Jupyter notebook containing both code and results.

(2) A small report of at most 5 pages (figures excluded).

if we sent the notebook we can include in it the small report

(3) A file pdf/ppt containing the presentation of the results.

We detail now what we expect in each of these points:

(1) The code folder:

You should submit one or more files that allow to reproduce all the results you have claimed. The dataset should **not** be included but it has to be available from a link. **If you have a doubt whether to include or not a file just ask us!** Your code should allow for the following points:

Plots You should make at least **two** different types of plots that answer different questions. These plots should be done using Seaborn and/or Plotly libraries and at least one of them allow for some animation/interaction as you have seen in class. The plots must enable you to extract interesting insight from the data. Feel free to use even more complex constructs that were not covered in the course.

Network analysis You should be able to extract a network from your dataset. Use the networkX package to describe different structural properties of your network (is it connected? bipartite? scale-free? what is the diameter? etc). You should compute at least **two** of different measures among the ones seen in class, to determine the importance of the nodes. Be prepared to explain your choice and comment your results.

Network Visualisation Use one of the tools seen in class (as for example Gephi) to visualise your network. Experiment with different layouts for your visualization. Which one do you prefer and why?

+ visualize some clustering of the network

(2) The report file:

It should not be more than 5 pages and it should contain: (i) the description of the dataset, its format, if you have perfomed any cleaning or modifications, (ii) the discussion of the results of your plots and network analysis and eventually the reasoning behind the choices you made (if any). Here you can include also all the figures and information that are **not** explicitly included in the presentation file.

Important: If you prefer you can also decide to include all the comments and the information of the report into your jupyter notebook. In that case you do not need to include also a report file or any figures (if they already appear in the notebook).

(3) The presentation file:

This should be a file pdf/ppt that contains your presentation of **10 mins** which is to be given by one or both of the members of your group. I suggest the presentation should have no more than 10 slides. The presentation should focus on :

1. a description of the dataset and in case you cleaned it, how you did this. You must conveniently clean the dataset.
2. a description of the questions you plan to address.
3. plots/tables/network visualisation that show how the visualisations you made helped you to answer what questions.

Finally don't forget that your presentation should "tell a story"!!

Good luck!