Advanced Visualisations 2024: Project

Professor: Blerina Sinaimeri

Deadline: Send a compressed file with the code and the report to bsinaimeri@luiss.it within **February 1, 2024**.

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. You can even create your own dataset. Some links when you can check for a database are the followings: http://www-personal. umich.edu/~mejn/netdata/ and http://snap.stanford.edu/data/, https://www.kaggle.com/

2 Expected outcome

The project will be valued **15 points** and will make for 50% of your final grade. Your team is required to submit :

(1) A zip 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 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. You should either add the dataset in the folder or make it available it from a link. The code for the following points:

Plots You should make at least two different types of plots that answer different questions. 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. The networks should have approximately 100 nodes. If your dataset is larger please select only a part of it. Experiment with different layouts for your visualization. Which one do you prefer and why? Use the networkX package to describe different structural properties of your network (is it connected?bipartite? what is the diameter? etc). Be prepared to explain why they are interesting.

Interaction Add some interaction to at least one of your visualisations.

(2) 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!