Take Home Exam

Start Assignment

Due 27 Feb by 16:00 **Points** 0 **Submitting** a website url or a file upload
Available 24 Jan at 10:50 - 27 Feb at 16:00

Project description:

This exam is to be made in teams of 2 students. Please create the group yourself, if you cannot find a team member, please email us and we will try to find you a team.

Deadline: **27th of February 16:00 PM** (CET time)

Dataset: **Health factors dataset** on different countries. See appendix A for variable descriptions.

You are asked to load, explore and report findings on the datasets in a Google Colab file. You will use and structure this file as a report following the guidelines provided in the course manual (see Assessment). The course manual is published in Canvas. Note that aside from code cells you should use **markdown (https://www.markdownguide.org/cheat-sheet/)** cells for headers and other text elements of your report, such as the introduction and conclusion.

For this assignment you are given the freedom to explore a question of your choice, in doing so however, please make sure to complete the steps described below:

1. **Introduction: (5 pt)**

- Formulate a research question that allows for explorations and comparisons. For example: How have vaccinations influenced hospitalizations in different European Union (EU) countries? You can divide this in two steps: 1. How vaccinations have influenced hospitalizations in a country (i.e Netherlands or Italy); 2. How does one country compare to the other countries in the EU? (5 pt)

2. **Data preparation: (20 pt)**

- Load the dataset using **this link for health factors data (https://raw.githubusercontent.com/NHameleers/dtz2025-datasets/master/CountryHealthFactors.csv)** . Copy the link and load it as you did in the practicals. (5 pt)
- Select rows and columns relevant to your research question. (15 pt)

3. **Explore and clean the data by: (20 pt)**

- Exploring data using descriptive statistics or visualisations to get to know the dataset and spot possible issues (such as outliers or typos)(10 pt)
- Identify and report issues with missing data. (5 pt)
- Resolve issues with missing data and clean other data inconsistencies. Report also if you found no issues and how you verified this. (5 pt)

4. **Describe and visualise: (50 pt)**

- Provide a description of the population for the reader (e.g. countries or continents) relevant to your question using a table with descriptive statistics (i.e. means, medians, standard deviations) and where possible visualisation. (15pt)

- Make the report interactive: Create at least one interactive visualisation using input from the user. (20pt)
- Turn your interactive report into an application using GitHub, Voila and Binder.** (15pt)

5. **Conclusion: (5 pt)**

- Summarise the work and the main findings related to the initial research question. (5 pt)

* Clarity of the report and presentation will be checked with each step of the solution.

** Info on this will follow in lecture 13.

Exam Rules:

Please observe some of the rules:

1. Deadline for sending the notebook: **16:00 (CET) 27/02/2024**
2. The exam requires work in teams of 2 students. If you cannot find a team, you can be alone.
3. If working in pairs, please define clear sub questions to be answered by each of you. Include a contribution statement at the end of your file. On the 28th of February, you will be presenting your team-work.
4. As with our practical sessions, you should answer the questions using Python and Google Colab.
5. Plagiarism of the solution or the explanatory text for any of the sections (1-6) invalidates your exam.
6. Please apply a double submission: You send us your GitHub link and your Colab Python file link by email. Additionally, submit your links in Canvas.
7. Mail to: v.urovi@maastrichtuniversity.nl (mailto:v.urovi@maastrichtuniversity.nl) , with cc to niels.hameleers@maastrichtuniversity.nl (mailto:niels.hameleers@maastrichtuniversity.nl) .
8. Subject: Introduction to Programming in Python <Insert Student Names>
9. Submit your file links in Canvas.
10. Ensure that you share the link to the Colab file in ways that we can execute the cells and give you feedback.
11. Consult the course manual to identify what is expected and what are the evaluation criteria. Ensure that the first field of **your notebook contains your names and student ids.**
12. To the best of your abilities ensure that your notebook runs top to bottom without errors (Runtime → Run all) If for some reason the notebook has last-minute errors, you should show the notebook to teachers and demonstrate what does work.
13. We assume that you will be present on the 28th! If for very good reason you cannot be present, please contact us before the latest on the **27th of February at 12:00** to receive additional instructions
14. **We wish you best of luck!**

Appendix A: Health factors dataset variables description

Variable name	Description
Adult Mortality	Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
infant deaths	Number of Infant Deaths per 1000 population

Alcohol	Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
percentage expenditure	Expenditure on health as a percentage of Gross Domestic Product per capita(%)
Hepatitis B	Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
Measles	Measles - number of reported cases per 1000 population
BMI	Average Body Mass Index of entire population
under-five deaths	Number of under-five deaths per 1000 population
Polio	Polio (Pol3) immunization coverage among 1-year-olds (%)
Total expenditure	General government expenditure on health as a percentage of total government expenditure (%)
Diphtheria	Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
HIV/AIDS	Deaths per 1 000 live births HIV/AIDS (0-4 years)
GDP	Gross Domestic Product per capita (in USD)
Population	Population of the country
thinness 1-19 years	Prevalence of thinness among children and adolescents for Age 10 to 19 (%)
thinness 5-9 years	Prevalence of thinness among children for Age 5 to 9(%)
Income composition of resources	Human Development Index in terms of income composition of resources (index ranging from 0 to 1)

Schooling Number of years of Schooling(years)