

My Data-Driven Analysis

Student My Name

Begin by changing title, author in the first cell of the notebook.

1 Introduction:

In this assignment, you will have the opportunity to explore a data-driven question or problem of your choice. The aim of this assignment is for you to engage in a data-driven project from start to finish. It is a great way to develop practical skills in data analysis and apply what you have learned in class about Python.

The goal is to follow a comprehensive data analysis process, including data collection, data preparation, data exploration, data analysis, and to draw conclusions based on your findings, as well as suggested next steps.

Project Structure:

Your project should be structured into three main sections:

2 Motivation (Data-Driven Question or Problem):

Describe the data-driven question or problem that you will explore.

This should be a real-world issue or question that you find interesting.

Explain why this question or problem is important or relevant. What motivated you to choose it?

Any text / image

```
# Any code
```

3 Solution (Data Analysis Process):

A tentative roadmap is the following:

Data Collection:

Explain where and how you obtained the data for your analysis. Provide details on the data source, format, and any challenges you encountered during data collection. The code implementing the collection should be here. Either generate the data dynamically in the notebook as a first step (if the process is light enough), or provide files storing the collected data (additional code used to generate the data should be provided).

Any text / image

```
# Any code
```

Data Preparation:

Describe briefly the steps you took to clean and preprocess the data. This may include handling missing values, outliers, and data transformation.

Any text / image

```
# Any code
```

Data Exploration:

Present your exploratory data analysis findings. Use descriptive statistics, visualizations, and summary tables to gain insights into the data.

Any text / image

```
# Any code
```

Data Analysis or Light Modeling:

Conduct the data analysis required to address your data-driven question. You may apply basic statistical methods, create indicators from your data, or any other relevant tools. Explain your approach, methodology, and any assumptions made during analysis.

Any text / image

```
# Any code
```

4 Findings (Conclusions, Impact and Next Steps):

4.1 Conclusions:

Summarize the key findings and insights obtained from your analysis. What did you discover in relation to your data-driven question or problem?

If applicable, highlight any metric or indicators that emerged from your analysis.

4.2 Impact:

Discuss briefly the real-world implications of your findings. How can the insights from your analysis be applied to address the original question or problem?

4.3 Next steps:

Provide recommendations or next steps for further investigation or action.

5 Instructions

5.1 Submission Requirements:

Your assignment should be submitted as a well-structured report, including text, figures, tables, and Python code using a Jupyter Notebook.

Document the data sources and code used in your analysis. Make your report reader-friendly and ensure that your analysis and findings are well-explained and justified.

Submit both the Jupyter Notebook and a rendered html file (using Quarto) before Oct 20 23:59.

Additionally, prepare for your live presentation on Oct 22, where you will demonstrate the Solution and discuss the Motivation and Findings. During the Solution demonstration show something that you learned about Python.

We will take time to setup in class a working development environment based on [miniconda](#) providing you a Python distribution + `conda` allowing you to manage the install of Python packages, create and manage Python isolated environments etc. We will also install a text editor `VS Code`, allowing with extensions ([Python/Jupyter](#)) to run Jupyter Notebooks within VS Code. Then we will install Quarto (both [cli](#) + VS Code extension), allowing you to produce a document (.html, .pdf or slides) from a Jupyter Notebook within VS Code.

In the online version of Wes McKinney (creator of pandas) book *Python for Data Analysis*, 3E, a [section \(1.4 Installation and Setup\)](#) is dedicated to installing miniconda and using conda to manage environment and install packages.

As a soft deadline to kick off your project, consider submitting me via email your idea of project and data source before Sep 20.

If you lack inspiration, consider looking at Open Data initiatives like for example the one by [French Government](#), or any other Area/Region/Country office of statistics/International Organization etc.

5.2 Grading Criteria:

Your assignment will be graded based on the following criteria:

- Clarity and relevance of the data-driven question or problem. (10%)
- The Solution:
 - Quality of data collection and preparation.
 - Depth and effectiveness of data exploration.
 - Richness of the data analysis and its results. Alignment with the initial question or problem.
 - Appropriateness of the modeling techniques (if used).
 - Quality of the code (correct, readable, documented)
 - Altogether (70%)
- Findings. (10%)
- Your presentation. (10%)

5.3 Additional Tips:

I encourage you to select data sets and questions that genuinely interest you to drive your motivation.