National University of Science and Technology

POLITEHNICA Bucharest

Faculty of Automation and Computers

Master Program

Advanced Analytics for Business

Visual Analytics Techniques

First semester, 2024-2025

Homework

Your title here

Teachers: Monica Drăgoicea Student(i): ...

Fadi Hosni E-mail: ....

Contents

[1. Introduction – brief description of your project (0.25 points) 4](#_Toc187053254)

[2. Data Sourcing & Acquisition (0.25 points) 4](#_Toc187053255)

[3. Data Wrangling (0.5 points) 5](#_Toc187053256)

[4. Data Preparation (1 point) 5](#_Toc187053257)

[5. Research Methodology (1 point) 5](#_Toc187053258)

[6. First Exploratory Data Analysis (0.5 points) 5](#_Toc187053259)

[7. Investigate relationships in data 6](#_Toc187053260)

[8. First step to build a decision 6](#_Toc187053261)

[9. Final conclusions (0.25 points) 6](#_Toc187053262)

[10. References (0.25 points) 6](#_Toc187053263)

|  |
| --- |
| This project must demonstrate that you have acquired basic skills in developing a data-driven project. In today's world, every organization collects and processes vast amounts of data to enhance operations, drive improvements, and foster innovation. Your project should showcase the following abilities:   * **Data Sourcing & Acquisition**:   + Ability to identify and access relevant data sources for your chosen research problem. * **Data Wrangling & Preparation**:   + Proficiency in inspecting, cleaning, transforming, and validating data for analysis. * **Data Visualization**:   + Skill in creating effective and informative visualizations (charts, graphs, etc.) to present data clearly and facilitate user understanding. * **Research Methodology**:   + Ability to define a clear research problem, formulate 1-2 focused research questions, and identify key dependent and independent variables. * **Data Analysis & Interpretation**:   + Proficiency in analyzing data, drawing meaningful insights, and interpreting the results in a clear and concise manner. * **Effective Communication**:   + Ability to present your findings effectively to an audience with varying levels of subject matter expertise, using a combination of written reports and compelling visualizations.   Key Considerations:   * Focus on Practical Application: The project should demonstrate a practical application of data analysis techniques to address a real-world problem or challenge. * Data-Driven Approach: Emphasize the importance of data-driven decision making throughout the project lifecycle. * Professional Presentation: Ensure your project is well-organized, professionally presented, and includes clear documentation of your methodology and findings. |

# Introduction – brief description of your project (0.25 points)

In this section you will present a brief description of your project, i.e. a data-driven problem. To solve it, you must apply all the steps included in SAS Visual Analytics methodology: Access, Investigate, Prepare, Analyze, and Report.

The data set that you will work on may be selected from GDELT Event database[[1]](#footnote-1), Kaggle[[2]](#footnote-2), and Economic Tracker[[3]](#footnote-3).

**NB**: You won’t always find pre-engineered data files readily available. It may be necessary to merge multiple databases or data files to answer your question. You might also need to apply specific transformations in SAS Viya using SAS Data Explorer to create a workable dataset. This content is explained in Visual Analytics 1 lecture accessible in SAS Skills Builder for Students.

Your problem must address resilience, sustainability, and responsibility issues. Please check here some examples of thinking about these challenges: “10 Big Data-Driven Sustainability Use Cases You Should Know”[[4]](#footnote-4).

**Please adapt your proposal to include a brief description of how your problem can be framed as a User–Data–Task Triangle design problem** (see Lecture 2).

Nevertheless, if you have already selected a research topic for “Research”, this is an excellent opportunity to reflect on it and begin drafting your work.

In each case, overall, you must connect your problem to the 17 Sustainable Development Goals defined by United Nations[[5]](#footnote-5).

This section should be written as a concise abstract summarizing your contributions, similar to how you would draft an abstract for a research paper or article. Throughout this section, use references purposefully to support your perspective. Minimize the use of website references, prioritizing scientific references that add meaningful support to your arguments.

# Data Sourcing & Acquisition (0.25 points)

After you have identified the general problem that you want to address, you may check whether you can find relevant data sources to support your quest. There is a general truth that even though you can find large amounts of data on existing data repositories, these data are not ready-to-use. Therefore, a lot of pre-processing is needed before you can use these datasets in your project.

In this section you must address the data access step. Follow the content of the lectures regarding SAS Visual Analytics methodology: Access. You have received important insight on how to find datasets, identify analysis tables that will be used in Visual Analytics and load those tables into CAS. Please explain all these activities in detail.

You must use the specific components in SAS Viya: SAS Data Explorer – Manage Data.

# Data Wrangling (0.5 points)

In this section you must address the data investigation step, following the content of the lectures. Align your presentation to the SAS Visual Analytics methodology: Investigate, i.e. you inspect the tables to determine whether any changes are needed for data items due to data inconsistencies or data quality issues, as well as identify any new data items that need to be calculated.

You must use the specific components in SAS Viya: SAS Information Catalog - Discover Information Assets (Lecture 3).

Is there any cross-tabulation needed? Is there a need to create new data items? These are two important observations to address in the next step regarding data preparation.

# Data Preparation (0.5 points)

In this section you must address the data preparation step, including data cleaning. Here, you correct any data quality issues and create any new calculated items needed for analysis.

Follow the content of the lectures 3, 4 and 5.

Use the specific components SAS Viya: SAS Data Explorer – Manage Data, SAS Studio - Develop Code and Flows.

Present visually the data that you will use further, after you applied all the required transformations. Use SAS Information Catalog - Discover Information Assets to present a first brief descriptive statistics overview on your cleaned data. Are there any missing values left?

# Research Methodology (0.5 points)

Your data is ready now, and you must set up a plan for analysis. Please better consider the kinds of questions you would like to explore. You can only ask questions that can be answered by the data you have available.

Please review Module 2, “Prepare Data for Analysis” in the “Data Literacy in Practice” course, to validate:

* The questions (and possible sub-questions) you would like to answer to
* The dependent variables based on which your answer will be formulated
* The independent variables that will drive you to the explanation on the dependent variable(s).

# First Exploratory Data Analysis (0.5 points)

In this section you must address this first data analysis step, using various charts and graphical representations, to understand the story that the data is telling, to explore and guide the business user with smart visualizations. You must explore data to discover meaning, to explore patterns in the data and making sense of these patterns.

Use the SAS Visual Analytics methodology: Analyze. In this section you must address the presentation of the data analysis process in a report that will catch the eye of the user.

Please review Module 3, “Exploratory Data Analysis” in the “Data Literacy in Practice” course.

# Investigate relationships in data (0.5 points)

Use Lectures 6, 7, 8 to profile your data with interactive graphs.

Please review also Module 4, “Investigating relationships in data using visualizations” in the “Data Literacy in Practice” course.

# First step to build a decision (0.5 points)

In this section you will try to evaluate the most important underlying features for a target variable using automatic explanations and classification decision trees.

Follow lectures 9 and 10.

This will help you to decide whether your choice of dependent and independent variables is valid, and it may lead you to a good answer for your questions.

# Final conclusions (0.25 points)

In this section, you should formulate final conclusions about your work approach, what was easy, what was more challenging, and how it helps you better understand how to use what you've learned to contribute to society, bring value through the knowledge you've acquired, etc.

Please review Module 5, “Contextualizing and communicating” in the “Data Literacy in Practice” course, to empower the presentation of results for your target audience.

You should present some thoughts on how advanced data analysis can contribute to achieving the Sustainable Development Goals (SDGs) formulated by the United Nations: https://sdgs.un.org/goals.

The presentation style should also highlight the main thread of the topic. The presentation of each step must be consistent and should reflect the problem formulated in the first section. The datasets should reflect the approached problem.

# References (0.25 points)

Present the bibliography you consulted on preparing this work. It is not enough to mention only the course and a few web pages; it is necessary to consult additional materials.

Some good starting links for research:

<https://scholar.google.com/>

<https://ieeexplore.ieee.org/Xplore/home.jsp> (free from UPB network)

<https://ieeexplore.ieee.org/search/searchresult.jsp?topic=climateChange> (IEEE Climate Collection)

<https://www.sciencedirect.com/> (free from UPB network)

If you need access to more references, just ask me.

1. <https://www.gdeltproject.org/> [↑](#footnote-ref-1)
2. <https://www.kaggle.com/> [↑](#footnote-ref-2)
3. <https://tracktherecovery.org/> [↑](#footnote-ref-3)
4. <https://www.iunera.com/kraken/sustainability/10-big-data-sustainability-use-cases/> [↑](#footnote-ref-4)
5. <https://sdgs.un.org/goals> [↑](#footnote-ref-5)