ICT Department

Data Wrangling & Exploratory Data Analysis

Hands-On Term Project Guidelines for Data Wrangling and Exploratory Analysis Course

Project Overview

The project involves data wrangling and exploratory analysis on a dataset related to your submitted approved topic. It will be assessed based on the quality and comprehensiveness of the data-wrangling process, the depth of the exploratory analysis, and the clarity of the write-up.

Project Milestones and Activities

1. Topic Selection and Dataset Acquisition (10 marks)

- Activity: Choose a topic of interest and acquire a relevant dataset.

- Instructions: Ensure the dataset is sufficiently complex (e.g., multiple variables, at least 500 records). Submit a brief description of the chosen topic and dataset.

2. Data Cleaning (20 marks)

- Activity: Perform data cleaning on the acquired dataset.

- Instructions: Handle missing values, correct data types, remove duplicates, and address inconsistencies. Document each step taken in the cleaning process.

- Submission: Cleaned dataset and a log of cleaning steps.

3. Data Transformation (20 marks)

- Activity: Perform data transformation.

- Instructions: Create new variables, normalise or standardise data, and apply any necessary transformations to prepare the data for analysis.

- Submission: Transformed dataset and a log of transformation steps.

4. Exploratory Data Analysis (EDA) (30 marks)

- Activity: Conduct an exploratory data analysis.

Instructions: Use visualisations and summary statistics to uncover patterns, trends, and relationships in the data. Include at least five different types of visualisations (e.g., histograms, scatter plots, box plots, and correlation heatmaps).

- Submission: EDA scripts and visualisations.

5. Final Write-Up (20 marks)

- Activity: Write a comprehensive report on the project. The write-up should include:

- Introduction: Briefly describe the chosen topic and the importance of the analysis.

- Dataset Description: Detail the metadata, including the source, the variables, and initial observations.

- Data Cleaning: Summarize the data cleaning process, explaining the issues encountered and how they were resolved.

- Data Transformation: Describe the transformations applied and the rationale behind them.

- Exploratory Data Analysis: Discuss the findings from the EDA, supported by visualisations and summary statistics.

- Conclusion: Summarize the key insights and potential next steps or recommendations.

- Submission: Final report (PDF or Word document).

Marking Rubric

- Topic Selection and Dataset Acquisition: 10 marks

- Data Cleaning: 20 marks

- Data Transformation: 20 marks

- Exploratory Data Analysis: 30 marks

- Final Write-Up: 20 marks

Submission Guidelines

- Submit all project files (scripts, datasets) in a compressed folder (Name the folder with Reg/ID number, then fullname e.g. BICT2B0929_REUBENMOYO). Upload your compressed folder to https://drive.google.com/drive/folders/1hUt4dYeYX_6jD0SZDHFdn4hNJfwasPtF?usp=drive_link

- Ensure the final write-up is comprehensive and clearly explains each project step.

- Adhere to the deadlines to ensure timely feedback and guidance.

- Ensure all code is well-documented and reproducible. All submissions MUST follow this guideline. Careless submissions will not be considered for marking. ZOBAKASHA WONT BE MARKED

- Changing the topic is not allowed. Any submission contrary to the approved topic will be automatically disqualified.

- Use of AI tools is highly discouraged and will be penalised harshly. For instance, all documents and codes produced by AI authoring tools will NOT be marked.

- The deadline is not negotiable.

Submission Deadline is 10 June 2024