Data Science - Midterm Project

Campus Classroom Analysis

1 Project Overview

For this midterm project, you will work with data collected about classrooms on campus. The dataset contains information such as classroom numbers, area measurements, seating capacity, WiFi quality, air conditioning status, and other features. This project will test your ability to clean, analyze, and derive meaningful insights from real-world data.

2 Project Tasks

2.1 Data Cleaning

Your first task is to clean the dataset. This may include:

- Handling missing values (imputation or removal)
- Correcting inconsistent terminology (e.g., "TRUE" vs. "yes" or "-1" vs. "-1.Floor")
- Standardizing units of measurement
- Removing outliers or correcting erroneous entries
- Converting data types where necessary

2.2 Statistical Analysis

Perform statistical analysis on the numerical features:

- Calculate central tendency measures (mean, median)
- Determine dispersion metrics (standard deviation, range)
- Find quartiles
- Identify correlations between numerical features

For categorical features:

- Report value counts
- Create visualizations of distributions (bar charts, pie charts)
- Analyze patterns in categorical data

2.3 Data Storytelling

The key component of this project is to tell a meaningful story using the data. This requires creativity and analytical thinking. Some approaches include:

- Identifying classroom utilization patterns
- Discovering relationships between room features and WiFi quality
- Analyzing the distribution of resources across campus buildings
- Exploring whether larger rooms have better amenities

You may use any appropriate method to develop your data story, including:

- Correlation analysis
- Clustering to identify similar classrooms
- Visualizations that highlight interesting patterns
- Comparative analysis between different classroom types

3 Deliverables

Submit the following:

- Python script(s) containing your code for data cleaning and analysis
- A 2-3 page PDF report describing your:
  - Data cleaning approach
  - Statistical analysis methods and results
  - Data story with supporting visualizations
  - Conclusions and insights

4 Submission Guidelines

- Compress your Python script(s) and PDF report into a single ZIP file
- Upload the ZIP file to the LMS
- Naming convention: DS_Midterm_StudentID.zip
- Submission deadline: 15.04.2025

Important: Failure to adhere to the submission guidelines, including file formatting, will result in grade deductions.