FILTER 01

Methodology:

- Perform a basic data analysis describing the dataset, summary statistics, basic data correlation, visualizations, etc.
    - Describe the data domain. A complete and deep explanation.
    - How the data was recollected, limitations of the study, disadvantages, etc.
    - Describe each of the variables
    - Basic summary statistics.
    - Boxplots. Interpretation.
    - Skew of the data. Interpretation.
    - Histograms. Interpretation.
    - Quartiles and interpretation.
    - Correlation. Interpretation.
    - Scatterplots. Interpretation.
- KNN
    - Normalization (both methods)
    - Which feature were selected and why.
    - Compute distances (At least two metrics).
    - Training and testing set.
    - Determine the optimal K.
    - KNN classification outputs.
    - Frequency table and interpretation.
- Conclusions and limitations.

Considerations:

- All the previous points are present in the report.
- Report must be in English.
- All the code must be printed
- Quality of the report and presentation.
- Reproducibility.
- Clear description of each step. Ex. How the students managed missing values; which features did the student think are more important for this analysis.
- Originality.
- Similar reports are eliminated (without further questions).
- Students should use at least 2 distance metrics. Then the “Class” Package and its function KNN should be replaced with a different package that allows the specification of different distance measures.
- The interpretation of the results is a key point to evaluate. Students should give clear and deep explanations of each point.
- Each graphic or table should be fully explained
- Tables and graphics must be referenced and have their corresponding caption.

- Aesthetics of the report. An adequate size of graphics and tables. Nice merging of graphics, tables and text.
- First page with the title, second page with an abstract, third page an index, last page the bibliography.