

Practical Work 2: Part 1 Data cleaning and preprocessing

1. Dataset Overview

- Import the basic libraries: `pandas`, `numpy`, and `matplotlib`.
- Import the dataset `Autos.csv` using Python into a `pandas DataFrame`.
(Download the file from the link https://elearning-facsci.univ-annaba.dz/pluginfile.php/89586/mod_folder/content/0/autos.csv?forcedownload=1)
- Display the first 15 rows of the dataset.
- What types of columns do you observe?
- How many rows and columns does the dataset contain?
- What do you observe about the data?
- Why do we analyze this dataset

2. Understanding the Variables:

- List all columns in the dataset. Based on their names, briefly infer the meaning of each column.
- Identify which columns contain numerical values and which contain textual or categorical values.
- Are there columns that seem irrelevant for analysis? If yes, which ones and why?

3. Checking and Handling Missing Values

- Identify the columns that contain missing values and report the number of missing entries in each.
- Propose appropriate strategies for handling missing values. Justify your choice for each affected column.
- Apply the chosen strategies and explain how you filled or removed the missing values.

4. Detecting and Handling Outliers

- Identify columns where outliers may be present, especially numerical attributes such as price or mileage.
- Use statistical and graphical methods (e.g., interquartile range method or scatter plots to detect potential outliers.
- Explain your method for handling outliers and apply it to the dataset.

5. Data Cleaning

- Convert categorical columns containing repeated labels (e.g., car brands) into an appropriate format if needed.\n"
- Standardize the formatting of data across columns: ensure consistent units (e.g., price in euros) and clean textual strings.
- Check for and correct typographical or entry errors in text-based columns."Identify potential *outliers* in the `price` column using graphical methods (e.g., boxplots).
- Propose a strategy to handle these outliers (removal, transformation, etc.).