



Steve Nouri 

@SteveNouri

Data Cleaning is super important and here are 8 essential steps that you need to follow:



1- Missing Values:

- Identify missing values in the dataset
- Decide on an appropriate method to handle missing values, such as imputation or deletion
- Impute missing values with appropriate values based on the method selected, such as mean, median, or mode

2- Duplicates:

- Identify and remove duplicate records in the dataset
- Verify that all duplicates have been removed

3- Outliers:

- Identify and handle outliers or extreme values in the data
- Decide on an appropriate method to handle outliers, such as deletion, transformation, or imputation
- Impute or delete outliers based on the method selected

4- Data Format:

- Verify that all data is in the correct format
- Convert data into the appropriate format, such as converting dates into a common format
- Handle inconsistent data formats

5- Data Validity:

- Verify that all data is valid and consistent
- Check for errors and correct them, such as typos or incorrect values
- Ensure that the data follows the appropriate rules and constraints

6- Data Consistency:

- Ensure that the data is consistent across all records and variables
- Check for and handle inconsistent values or data types
- Correct any discrepancies in the data


7- Data Standardization:

- Standardize variable names and values for consistency
- Ensure that all categorical variables are in a standard format, such as converting all uppercase to lowercase
- Convert all variables into a common scale, such as converting all weights into kilograms

8- Data Transformation:

- Transform data to make it suitable for analysis, such as scaling, normalization, or aggregation and to meets the requirements and assumptions of the analysis method



Steve Nouri 

@SteveNouri

Thanks for reading!

If you enjoyed this post, dont forget to
Click Follow.