

Exploratory data analysis (EDA)

Process of investigating, organizing and analyzing data sets. And summarizing their main characteristics often employing data wrangling and visualization methods.

6 main practices of EDA

1. Discovering

Familiarize with the data to start conceptualizing how to use it.

2. Structuring

Process of taking raw data and organizing or transforming it to be more easily visualize, explained, or modeled.

3. Cleaning

Process of removing errors that may distort your data or make it less useful (missing values, misspellings, duplicate entries, or extreme outliers)

4. Joining

Process of augmenting or adjusting data by adding values from other datasets.

5. Validating

Process of verifying that the data is consistent and high quality.

6. Presenting

Making your cleaned data set or data visualizations available to others for analysis or further modeling.