Exploratory Data Analysis - Workshop

Ben Sivan

Exploratory Data Analysis

Workshop

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process that involves visually and statistically summarizing, interpreting, and understanding the main characteristics of a dataset. EDA is particularly useful for several purposes:

- **Understanding the Data Distribution:** EDA helps you grasp the underlying structure and distribution of the data. This includes identifying patterns, trends, and outliers that may exist in the dataset.

- **Identifying Patterns and Relationships:** EDA helps uncover patterns, relationships, and associations between variables. This can be valuable for formulating hypotheses, guiding feature engineering, and gaining insights into the underlying data dynamics.

- **Data Cleaning and Preprocessing:** During EDA, you might identify missing values, inconsistencies, or errors in the data. Cleaning and preprocessing the data are essential steps to ensure the accuracy and reliability of subsequent analyses.

- **Hypothesis Generation:** Exploratory Data Analysis often involves forming hypotheses about the data. These hypotheses can then be tested using more formal statistical methods, contributing to a more rigorous analysis.

In summary, exploratory data analysis is a versatile and essential step in the data analysis process, providing a foundation for subsequent modeling and decision-making. It helps data analysts and scientists gain insights, identify patterns, and make informed decisions based on a deeper understanding of the data.

Data types:

- Categorical Nominal
- Categorical Ordinal
- Numerical Discrete
- Numerical Continuous