Exploratory Data Analysis

Exploratory Data Analysis (EDA), also known as Data Exploration, is a step in the Data Analysis Process, where a number of techniques are used to better understand the dataset being used. By conducting EDA, you can turn an **almost** useable dataset into a completely useable dataset.

Components of EDA

1. Understanding your variables
2. Cleaning your dataset
3. Analyzing relationships between variables

Cleaning of Dataset include-Removing Redundant variables, Variable Selection, Removing Outliers, Removing Rows with Null Values

Visualizing High Dimensional Data

The basic steps in this process are:

1. Scale the data (important here because our features are heterogenous)

2. Fit the PCA transformation (learn the transformation from the data)

3. Apply the transformation to the data to create the derived features

4. Use the derived features to look for patterns in the data and explore the coefficients