Exploratory Data Analysis With Python -I

Programming Perspective of Machine Learning

EDA Objectives

1. To load a data set
2. To get initial information about the data set
3. To get descriptive statistics

We will be using Pandas and Numpy Libraries from Python

Titanic Data Set

Sr. No	Task
1	To load a CSV file
2	To Print the header of CSV file
3	To print dimensional size of the data set
4	To print names of variables or features
5	To print types of variables or features
6	To print descriptive statistics of numeric variables
7	To print descriptive statistics of non- numeric variables
8	To print counts of distinct values for categorical variables
9	To print percentage of distinct values for categorical variables

Task	Command
To load a CSV file	df= pd.read_csv()
To Print the header of CSV file	df.head()
To print dimensional size of the data set	df.shape()
To print names of variables or features	df.columns()
To print types of variables or features	df.info()
To print descriptive statistics of numeric variables	df.describe()
To print descriptive statistics of non- numeric variables	df.describe(include= 'object')
To print counts of distinct values for categorical variables	df['Sex'].value_counts()
To print percentage of distinct values for categorical variables	df['Sex'].value_counts(normalize= True)