BHARTIYA VIDYA BHAVAN'S

SARDAR PATEL INSTITUTE OF TECHNOLOGY

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400058-India

CLASS: B.E. I.T **ACADEMIC YEAR: 2022-23**

COURSE: DATA ANALYTICS

STUDENT NAME: SMRITHI AGRAWAL **ROLL NO: 2019140001**

EXPERIMENT 1

AIM:

To perform Exploratory Data Analysis such as number of data samples, number of features, number of classes, number of data samples per class, removing missing values, conversion to numbers, using seaborn library to plot different graphs.

THEORY:

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it. The EDA approach is precisely that--an approach--not a set of techniques, but an attitude/philosophy about how a data analysis should be carried out.

Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

1. maximize insight into a data set;
2. uncover underlying structure;
3. extract important variables;
4. detect outliers and anomalies;
5. test underlying assumptions;
6. develop parsimonious models; and
7. determine optimal factor settings.

EDA is not identical to statistical graphics although the two terms are used almost interchangeably. Statistical graphics is a collection of techniques--all graphically based and all focusing on one data characterization aspect. EDA encompasses a larger venue; EDA is an approach to data analysis that postpones the usual assumptions about what kind of model the data follow with the more direct approach of allowing the data itself to reveal its underlying structure and model. EDA is not a mere collection of techniques; EDA is a philosophy as to how we dissect a data set; what we look for; how we look; and how we interpret. It is true that EDA

heavily uses the collection of techniques that we call "statistical graphics", but it is not identical to statistical graphics per se

COLAB LINK:
https://colab.research.google.com/drive/1mgykKE_4BWg5rsmNP3EdvyRPpOdn_wn8?usp=sharing

GITHUB LINK: https://github.com/smrithi1912/DataAnalytics-Lab

DATASET: https://worldhappiness.report/ed/2021/#appendices-and-data