**Exploratory Data Analysis (EDA)**

**Data analysis** is the procedural practice of carefully applying mathematical and statistical or logical methods to describe, illustrate, clean and evaluate data for reporting.

Data analysis is very useful in critical decision-making for companies and businesses e.g when analyzing sales reports or revenue reports when coming up with new marketing strategies.

On the other hand, **exploratory data analysis (EDA)** is often practiced by data scientists to critically analyze, investigate and summarize the main features of data while engaging in data visualization.

Data scientists often use various tools such as python, R, Power BI, tableau, excel, SPSS, and so on to successfully carry out EDA.

**Some Examples of EDA**.

* When analyzing rational customer behavior while buying a particular commodity.
* Analyzing sales reports.
* Analyzing cost-benefit analysis.
* Employee performance analysis.

**The Focus of EDA.**

The overall objective of EDA is to travel, examine and study data sets in place of confirming the statistical hypothesis.

EDA employs data visualization tactics such as graphs and charts to represent findings after successful analysis.

Some other objectives of EDA are:

1. Identifying and removing outliers.
2. Recognizing trends in time and space data series.
3. Reveal trends related to data sets.
4. Creating and testing hypotheses through experimentation.
5. Recognizing new data sources.

**Steps Involved in a successful EDA exercise.**

1. **Data collection** – data can be gathered from surveys, social media, customer reviews, etc.
2. **Revealing valuables and comprehending them** – data features that affect the outcome and have greater effects on the results are identified.
3. **Data Cleaning** – checking of anomalies, null and extraneous values in a bid to minimize the computational power during execution of analysis.
4. **Recognizing related variables** – correlation matrices help in illustrating the relationship between data aiding in easy comprehension of the relationship between data.
5. **Choosing the right statistical methods** – mathematical and statistical techniques give relative information whereas data visualizations such as graphs give a concise output.
6. **Visualizing and analyzing results** – findings are careful studies together with the visualizations to make correct interpretations. Patterns in the data spread and their correlation aid in modifying suitable changes to parameters for easier comprehension.

**Examples of EDA tools.**

1. **Python.**

Besides being a widely used programming language, python is crucial in EDA as it helps in gathering data, describing data, outlier management, visualizations, etc. Python libraries such as Numpy, Pandas, Matplotlib, Seaborn, etc are useful in EDA, especially for beginners.

Alternatively, python packages like D-tale, AutoVizm PandasProfiling, etc automize EDA saving a lot of time and effort.

1. **R.**

This language is used by statisticians and data scientists in extensive EDA. Similar to python, R is an open-source programming language for statistical computing and graphical visualization. R libraries such as ggplot, leaflet, and lattice are efficient in EDA. Data Explorer Smart EDA and GGally are also useful automated libraries of R or quick EDA.

**Merits of EDA.**

1. **Useful in gaining insights into data trends and patterns.**

Helps data scientists in recognizing key trends through data visualizations like graphs and histograms. Suh is used in compies to formulate new strategies e,h marketing strategies.

1. **Enhanced variable comprehension**.

Using EDA, data scientists can cite various statistical information like mean, minimum, maximum quartiles, variance, etc.

1. **Saves a lot of time and effort.**

EDA helps data scientists in identifying mistakes, abnormalities, or missing values in a data set. This helps save a lot of time and effort by evading errors when applying machine learning models during reporting and visualization.

1. **Making data-focused decisions.**

EDA ais businesses and companies to comprehend their data. With EDA, extraction key indicators are simplified are useful in drawing realistic conclusions, proving a great asset during decision-making.