PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Semester : __V__   Subject : __SAIDS__   Academic Year: 2023 - 2024

Module 1 :- Exploratory Data Analysis

1) Elements of Structured Data

What is Data?

Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things.

What is structured data

Structured data is quantitative data in the form of numbers and values.

Example: It includes names, dates, addresses, credit card numbers, stock information etc.
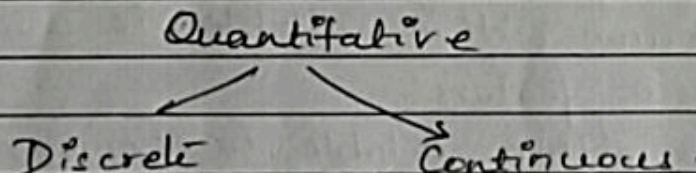
* Structured data is highly organized.
* Data organised in Relational database Management System (RDBMS).

What is unstructured data?

Unstructured data is qualitative data in the form of text files, audio files, video files.

Structured $\longrightarrow$ Quantitative data

Quantitative data is numerical information. (numbers).

```
                Quantitative
                /          \
         Discrete       Continuous.
```

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Semester: Ⅴ   Subject: Statistics for AIDS   Academic Year: 2023-2024

Discrete Data :-

Discrete data can only take certain values.

Example :-

The number of students in a class.

The result of rolling 2 dice.

Continuous Data :

Continuous Data can take any value (within a range).

Example:

A person's height

Time in a race.

Qualitative data - unstructured data

Qualitative data is descriptive information (it describes something).

==Rectangular data :-==

Rectangular data is the general term for a two-dimensional matrix with rows indicating records (cases) and columns indicating features (variables).

Key Terms of Rectangular Data:

Data frame :- Rectangular data (like a spreadsheet) is the basic data structure for statistical and machine learning models.

Feature :- A column within a table is commonly referred to as a feature.

Records :- A row within a table is commonly referred to as a record.

Semester: ⇧   Subject: Statistics for AIDS   Academic Year: 2023-20 24.

Example for Rectangular Data:

	A	B	C		csv
					name, gender, date
1	name	gender	date	format.	Dezik, Male, 1951-07-22
2	Dezik	Male	1951/07/22	→	Dezik, Male, 1951-07-29
3	Dezik	Male	1951/07/29		Tsygan, Male, 1951-07-22
4	Tsygan	Male	1951/07/22		Lisa, Female, 1951-07-29
5	Lisa	Female	1951/07/29		Chizhik, Male, 1951-08-15
6	Chizik	Male	1951/08/15		

It reads the data from a file which has .csv format.

Example:

The following code can be executed using python:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('salary.csv')   → upload salary.csv in
df.                                 colab Notebook.
```

By executing the above code, the rectangular data of salary will be displayed.

```
df.info()
```

By executing this function, it displays the entire information about the data frame.

Semester : **V**   Subject : Statistics for AIDS   Academic Year: 2023-2024

Non-Rectangular Data Structures :-

There are other data structures besides rectangular data.

* Time series data records successive measurements of the same variable.

* Spatial data structures, which are used in mapping and location analytics, are more complex and varied than rectangular data structures

* In the object representation, the focus of the data is an object (eg. a house) and its spatial coordinates.

* Graph data structure are used to represent physical, social and abstract relationships.

For example, a graph of a social network such as Facebook or LinkedIn, may represent connections between people on the network.

Example for non-rectangular data:

The data format used for non-rectangular data can be JSON, XML etc.

JSON	XML
{ "name": "Darth Vader",	\<note\>
"species": "Human",	\<from\> Teacher \</from\>
"homeworld": Tatooine",	\<to\> Student \</to\>
"films": [	\<heading\>Almost there\</heading\>
"Revenge of the Sith",	\<body\> Iti the final chaptu!\</body\>
"Return of the Jedi",	\</note\>
]	

3