

Introduction to Data Science

III INTRODUCTION TO DATA —

In our day to day life we use the term 'data' which refers to the information stored in computer system. This data can be raw, structured, or unstructured. Data is raw, unorganized facts that need to be processed. When data is processed, organized, structured or



- What is Data Science?
- Applications of Data Science
- ▶ Career in Data Science
- Data Visualization

presented in a given context so as to make it useful, it is called information. Data can be in the form of images, sound, videos, numbers, text etc. Data is gathered and translated for some purpose, usually analysis. In a computer, the data is stored in binary form i.e. in the form of 0's and 1's.

Data can be of two types:

QUALITATIVE DATA

Qualitative data cant be measured or counted in the form of numbers. These types of data are sorted by category, not by number. These data consist of audio, images, symbols, or text. The gender of a person, i.e., male, female, or others, is qualitative data.

Qualitative data tells about the perception of people. This data helps market researchers understand the customers' tastes and then design their ideas and strategies accordingly.

QUANTITATIVE DATA

Quantitative data can be measured by numerical variables. For example: scores in exams e.g. 65, 88, 97 etc., weight of a person. Quantitative data is further classified into discrete data and continuous data.

Discrete data: Discrete data contains only a limited number of possible values. It only contains finite

values, the subdivision of which is not possible. It includes only those values which are separate and can only be counted in whole numbers or integers, which means that the data can not be split into fractions or decimals.



Continuous data: Continuos data are in the form of fractional numbers. It can be the version of an android phone, the height of a person, the length of an object, etc. Continuous data represents information that can be divided into smaller levels. The continuous variable can take any value within a range.



Healthcare: Identifying patients at risk of readmission by analyzing patient demographics, medical history, lab results and vital signs. Using deep learning algorithms to analyze medical images such as CT scans and MRI to assist in the diagnosis of diseases.



Finance: Identifying suspicious transactions and patterns using machine learning algorithms and anomaly detection techniques.



Retail: Analyzing customer data to make personalized product and content recommendations. Forecasting sales and demand using time series analysis and machine learning methods. Optimizing inventory levels using forecasting and optimization methods.



Manufacturing: Identifying equipment that is likely to fail using predictive models and sensor data. Identifying bottlenecks and inefficiencies in production processes using techniques such as process mining and simulation.



Transport: Data Science helps in creating the systems for self-driving cars using sensory data. Data Science allows extensive analysis of fuel consumption patterns, driver monitoring, and path selection helps in optimizing the industry.



E-commerce: Analyzing customer browsing and purchase history to make personalized product recommendations. Optimizing prices using techniques such as dynamic pricing and optimization algorithms to maximize revenue or profit.

Data science is a rapidly growing field, and it is being applied in many other industries and fields as well.

CAREER IN DATA SCIENCE —

Data science continues to rise as one of the most in-demand career paths in technology today. Data scientists may work in various settings, including research organizations, businesses, or government agencies. Beyond data analysis, mining, and programming, data scientists program code and combine it with statistics to transform data.

There are plenty of data science jobs to choose from. All of them are integral to making key business decisions.

the original data collected by a researcher for a specific research purpose.

SECONDARY

Secondary data is a second hand data collected by other parties and have already undergone statistical analysis.

DATA ANALYSIS —

Data analysis is the process of cleaning, analyzing, interpreting, and visualizing data, with the goal of discovering valuable information for business decisions. The purpose of data analysis is to extract useful information from data and taking the decision based upon the data analysis.

A simple example of data analysis can be seen whenever we make a decision in our daily lives by evaluating what has happened in the past or what will happen if we make that decision. Basically, this is the process of analysing the past or future and making a decision based on that analysis. We gather past memories and experiences or dreams of our future. So that is nothing but data analysis.

While there are several different ways of collecting and interpreting this data, most data-analysis processes follow the same six general steps.

1. Specify Data Requirements

2. Collect Data

3. Clean and Process the Data

4. Analyse the Data

5. Interpretation

6. Report

III DATA VISUALIZATION —■

Data visualization is one of the steps of the data science process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made. Data visualization is a graphical representation of quantitative information and data by using visual elements like graphs, charts, and maps. Data visualization convert large and small data sets into visuals, which is easy to understand and process for decision making.

For best results, we need to ensure that the data is correct and complete, ample data sets are available to make predictions more accurate and to construct the algorithm that can have an impact.

BENEFITS OF DATA VISUALIZATION

- The ability to understand information quickly, improve insights and make faster decisions.
- An increased understanding of the next steps that must be taken to improve the organization.
- An improved ability to maintain the audience's interest with information they can understand.

- ▶ AData refers to raw information that consists of basic facts and figures.
- Data is gathered and translated for analysis.
- ▶ Data Science is about finding patterns in data, through analysis, making future predictions.
- Data Science is used in many industries in the world today, e.g. banking, consultancy, healthcare and manufacturing.
- Data visualization means the graphical or visual representation of data and information using visual elements like graphs, charts, maps, tables.

PRACTICE TIME	PRA			PRACTICE	TIME	
---------------	-----	--	--	----------	------	--

A Tick ✓ the correct answer

a. Video	b. Messages	c. Both a and b	

2. Which of the following is a use of data visualization?

1. Which of the following is considered data?

- a. Healthcare b. Sales and Marketing c. Politics
- 3. "How is today's weather?", is what kind of data?

a. Qualitative	b. Quantitative	c. Both a and b	

4.is the job role of a data scientist.

a.	To extract consumable information from structured and unstructured data.	
b.	Creating a predictive model based on data studied.	
c.	Build algorithms through which stakeholders can easily consume raw data.	

- 5. Data can be visualized using
 - a. Graphs c. Both a and b

4.	Mention any four important applications where data science can be used.
5.	Write a short note on how data science is helping sports team.
6	
6.	Differentiate between discrete and continuous data.
	ACTIVITY ————————————————————————————————————

- A Each student should write down the marks he/she had received in the examination for the subjects studied in the previous class. Use these marks to plot on paper.
 - 1. A bar graph to display marks of each subject.
 - 2. A line graph to display marks of each subject.
 - 3. A pie chart to show the percentage contribution of marks of each subject to the total marks obtained.



TERM II - SKILL SHEET Chapter (6, 7, 8, 9, 10)



D. ANSWER IN ONE WORD

- 1. Range() function is used with this loop
- 2. Default frame rate for new Animate document ________
- 3. Extension of an external Javascript file
- 4. Al technology used in Search Engines _____
- 5. This loop is also known as indefinite loop
- 6. A place where all the symbols are stored in Animate

E. WRITE SHORT NOTES ON THE FOLLOWING

- 1. Membership Operators in Python
- 2. Role of Data Analyst
- 3. Tweened animation
- 4. Computer Vision (CV) in AI
- 5. The object model in Javascript
- 6. Big Data

F. Answer the following questions

- 1. Differentiate between For loop and While loop.
- 2. Write any three applications of Natural Language Processing (NLP).
- Differentiate between unstructured data and structured data.
- 4. What is data visualization?
- 5. What are the three types of symbols available in Animate CC?
- 6. What is window.confirm() method in Javascript?
- 7. What is user-defined function in Python?



Fill in the blanks

 protocol is used for transferring files over the network. Ais a network device that connects two or more networks. is a network device that connects two or more networks. is a network device that connects two or more networks. attributes data from two or more tables. Theattribute of the frame tag tells the browser which HTML page to load into the file. A link that connects two different web pages is called aslink. list is a list of terms followed by their descriptions. Unordered list is displayed with symbol like 		- miles
 retrieves data from two or more tables. The attribute of the frame tag tells the browser which HTML page to load into the factor of the frame tag tells the browser which HTML page to load into the factor of the frame tag tells the browser which HTML page to load into the factor of the frame tag tells the browser which HTML page to load into the factor of the	1.	protocol is used for transferring files over the network.
 Theattribute of the frame tag tells the browser which HTML page to load into the A link that connects two different web pages is called aslink. list is a list of terms followed by their descriptions. 	2.	A is a network device that connects two or more networks.
 5. A link that connects two different web pages is called aslink. 6list is a list of terms followed by their descriptions. 	3.	retrieves data from two or more tables.
6list is a list of terms followed by their descriptions.	4.	Theattribute of the frame tag tells the browser which HTML page to load into that frame.
· ·	5.	A link that connects two different web pages is called as link.
7. Unordered list is displayed with symbol like	6.	list is a list of terms followed by their descriptions.
	7.	Unordered list is displayed with symbol like

Definition | Bullet | Router | External | Relationship | TCP/IP | SRC

Answer the following questions

- 1. State the types of computer networks based on the geographical area covered by them.
- 2. What is the difference between HTTP and HTTPS protocols?
- 3. Write the different types of wireless technologies. Explain any one.
- 4. What are Reports? Name the different views of Reports.
- 5. Is it possible to extract information based on more than one condition in a single query? If yes, how?
- 6. In how many categories are the link states in HTML divided? Explain them.
- 7. How can you change the type of bullets used in an unordered list?
- 8. What are the different attributes of a <TABLE> tag?

(E) Concept Based Questions

- 1. Jayant wants to set up a network for his organization. He wants to use a topology in which all nodes are to be connected in a circular path. Which topology should he use?
- 2. Saumya wants to retrieve data from two tables stored in a database. Using which feature of Access can she perform the given task.
- 3. Naveen created a simple web page. Now he wants to add a logo, link a few pages and add images. Can you suggest him the tags and their syntax?

** WORKSHEET 2 .** CHAPTER - 6, 7, 8, 9, 10

Fill in the blanks

	Hints Digital Camera NLP Infinite Data Engineer Loopelse Motion Guide
1.	If the test condition is always true, the loop will be executed number of times.
2.	Moving an object in a predefined path such as curves or circles is called as
3.	When we use else with for loop, this structure is known as .
4.	Email service providers usetechnology to filter spam emails.
5.	device can be used for image acquisition.
6.	prepares and manages large amounts of data.

Answer the following questions

- 1. What are the four elements of a while loop in Python?
- 2. How while loop is different from for loop?
- 3. Write steps to hide a layer in Animate CC.
- 4. What is frame-by-frame animation? Explain.
- 5. What is the importance of Javascript?
- 6. Explain the different types of comments in Javascript.
- 7. What do you understand by Computer Vision? Write its three main processing components.
- 8. Explain the characteristics of Big Data.

E Concept Based Questions

- 1. Radhika wants to apply shape tweening on a symbol. But when she tried to do so, the shape tween option got disabled. What mistake is Radhika doing?
- 2. Your teacher asked you to write a program in which she wants the statement by which the program jumps out of the loop unconditionally. By which statement can you do so?