

# Introduction to Medical Informatics

Data Science  
Lecture: 9



# Contents of Lecture



*What is Data Science?*



*Data Analysis Process*



*Exploratory Data Analysis*

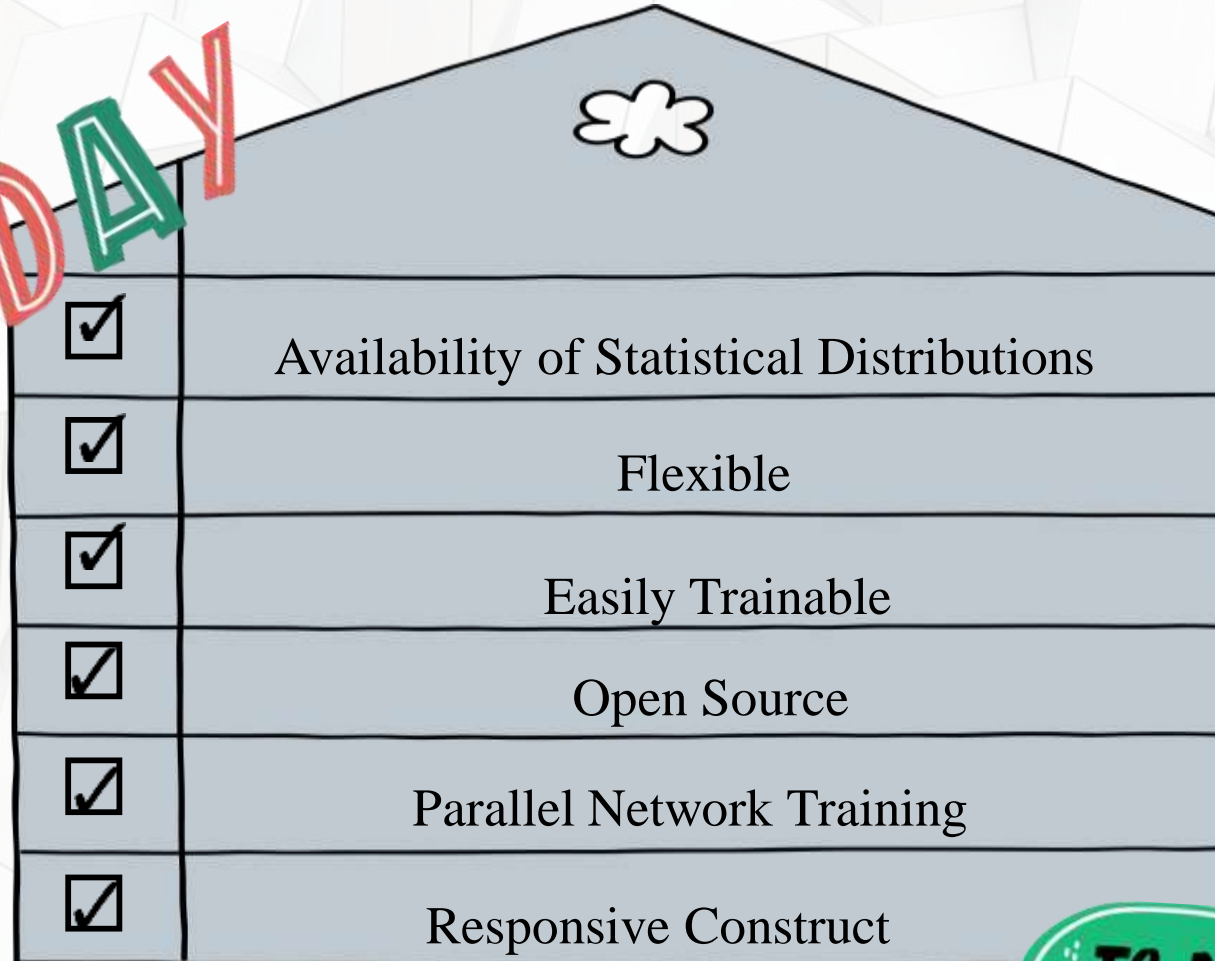
# What is Data Science

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science combines multiple fields, including statistics, scientific methods, artificial intelligence (AI), and data analysis, to extract value from data.

- Data science encompasses preparing data for analysis, including cleansing, aggregating, and manipulating the data to perform advanced math & statistics data analysis.

# Features of Data Science

TODAY

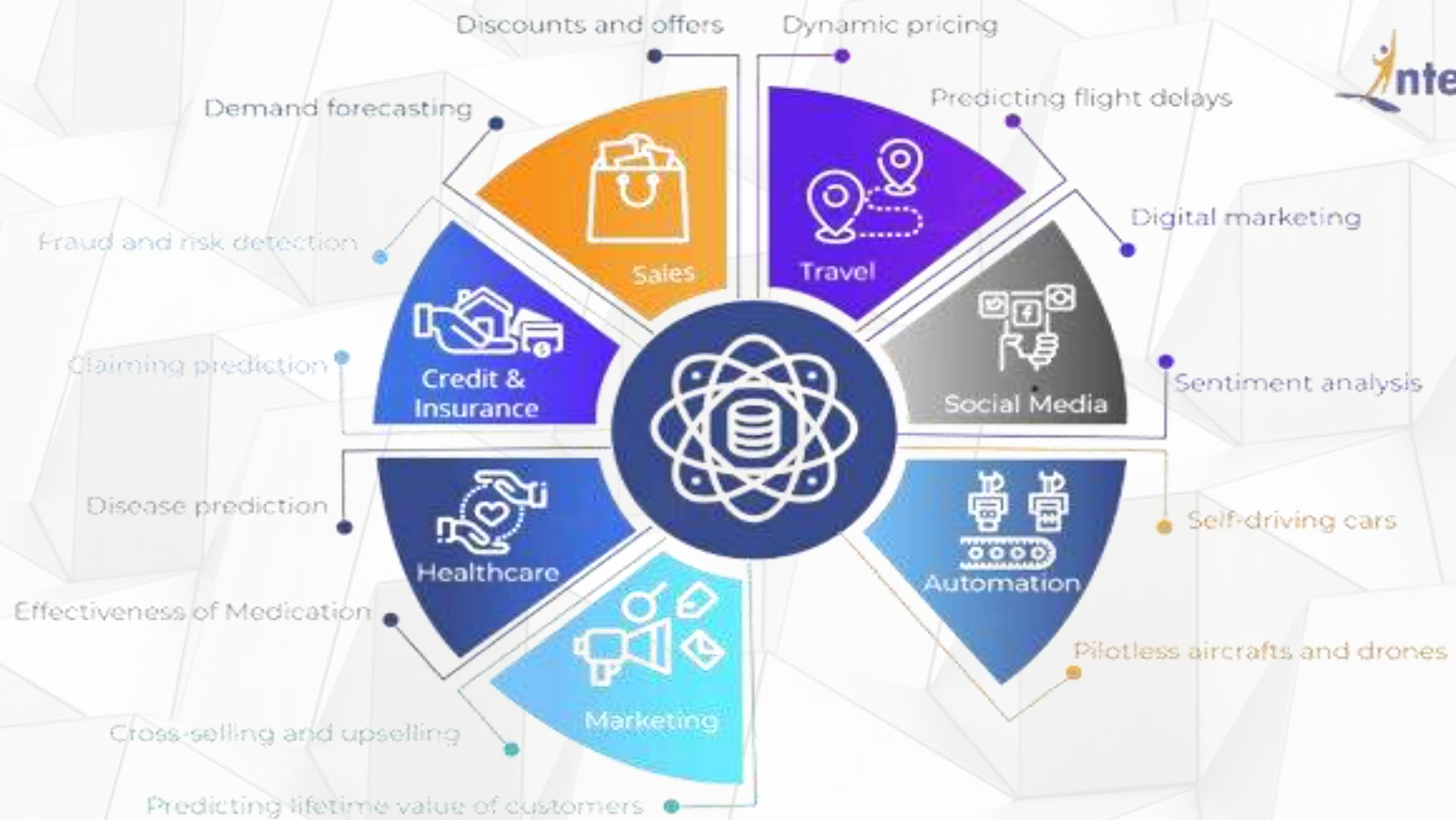


<input checked="" type="checkbox"/>	Availability of Statistical Distributions
<input checked="" type="checkbox"/>	Flexible
<input checked="" type="checkbox"/>	Easily Trainable
<input checked="" type="checkbox"/>	Open Source
<input checked="" type="checkbox"/>	Parallel Network Training
<input checked="" type="checkbox"/>	Responsive Construct

TO DO

03

# Different sectors where we using data science



# Data Analysis Process



04

# Data Analysis Process in Data Science

Data Analysis Process consists of the following phases:

## THE DATA ANALYSIS PROCESS

**Step 1:**

Define the question

**Step 2:**

Collect the data

**Step 3:**

Clean the data

**Step 4:**

Analyze the data

**Step 5:**

Visualize and share your findings

# Data Analysis Process in Data Science

## 1- Data Requirements Specification

The data required for analysis is based on a question or an experiment. Based on the requirements of those directing the analysis, the data necessary as inputs to the analysis is identified (e.g., Population of people).

## 2- Data Collection

Data Collection is the process of gathering information on targeted variables identified as data requirements.



# Data Analysis Process in Data Science

## 3- Data Cleaning

The processed and organized data may be incomplete, contain duplicates, or contain errors. Data Cleaning is the process of preventing and correcting these errors.

## 4- Data Analysis

Data that is processed, organized and cleaned would be ready for the analysis. Various data analysis techniques are available to understand, interpret, and derive conclusions based on the requirements

## 5- Communication

The results of the data analysis are to be reported in a format as required by the users to support their decisions and further action.



# Exploratory Data Analysis

# EDA (Exploratory Data Analysis)

Exploratory data analysis( EDA) is a method of analyzing and investigating the data sets to summaries their main characteristics.

- EDA build a robust understanding of the data, issue associated with either the info or process. it's a scientific approach to get the story of the data.

# Types of Exploratory Data Analysis (EDA)

There are four types of EDA in all:-

- 1 . Univariate Non-graphical
- 2 . Univariate graphical
- 3 . Multivariate Non-graphical
- 4 . Multivariate graphical

# Types of Exploratory Data Analysis (EDA) Tools

## 1- Python

- EDA can be done using python for identifying the missing value in a data set. Other functions that can be performed are the description of data, handling outliers, getting insights through the plots. Its high- level, built-in data structure and dynamic typing and binding make it an attractive tool for EDA

Analyzing a dataset is a hectic task that takes a lot of time. Python provides certain open- source modules that can automate the whole

# Types of Exploratory Data Analysis (EDA) Tools

## 1- R

The R language is used widely by data scientists and statisticians for developing statistical observations and data analysis.

R is an open-source programming language that provides a free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing.



# Thank You

Do you have any question!