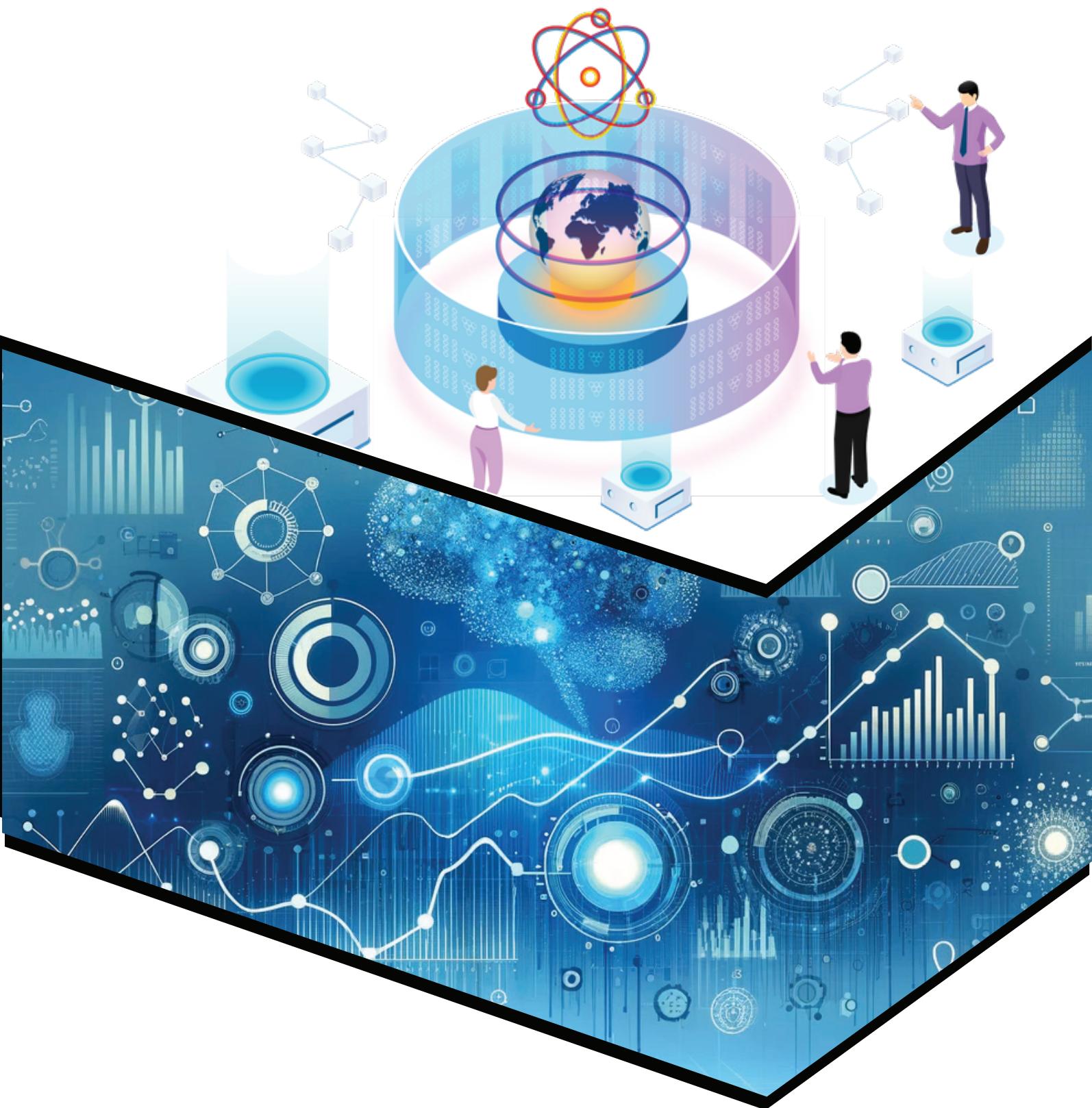


# Data Science



## What is Data Science?

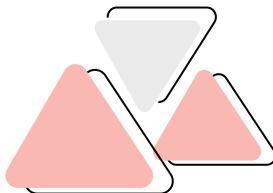
Data science is a field where data scientists apply scientific methods and algorithms to extract insights from structured and unstructured data. It includes data collection, cleaning, exploration, modeling, mining and analysis. They combine techniques from statistics and computer science to analyze complex data. Data scientists apply machine learning algorithms and statistical models to make predictions based on the given data. This field covers almost all industries, such as healthcare, finance, marketing, information technology, etc. It makes better business decisions, improves success rates and creates new products and services.



## Why Data Science?



Data science is a rapidly growing field that is in high demand and offers high-paying job opportunities. It is the best career choice for those who are interested. Data scientists apply strategic decision-making through data-driven insights. They need continuous learning to stay at the forefront of the industry and grow. Its effective work enables professionals to solve real-world problems and improve the standard of living. By combining statistical analysis, machine learning, and artificial intelligence, data scientists play a crucial role in converting ample amounts of data into actionable insights. It ultimately helps shape the future of business and society.



# Advantages of Data Science



## Decision Making

It empowers management and officers to make more informed decisions.



## Quick Action

Data scientists directly take actions based on the trends that are necessary to achieve the goals in favor of the company.



## Best Recruiter

Recruit the right, talented, skilled and knowledgeable candidate for the organization.

01

02

03

06

## Advantages



## Identifier

Identify and refine the target audiences.

05

04



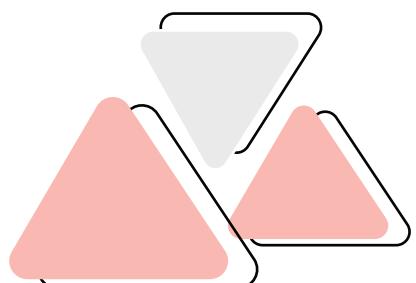
## Challenges Taker

Data scientists focus on taking challenges for the staff to adopt the best one and focus on the favorable elements that matter.



## Best Analyst

They excel in identifying the opportunities.



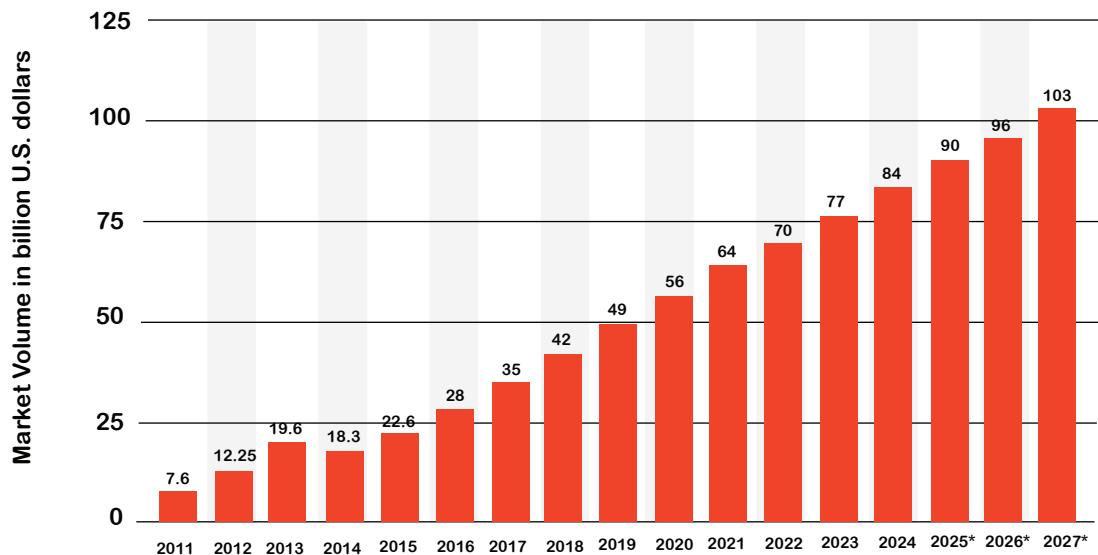
## Job Opportunities & Salary

Job Profile	Average Salary
Data Scientist	12 - 18 LPA
Data Analyst	6 - 8 LPA
Machine Learning Engineer	15 - 20 LPA
Business Analyst	6 - 10 LPA
Data Engineer	8 - 12 LPA
Data Architect	18 - 25 LPA
Statistician	6 - 10 LPA
Business Intelligence Analyst	8 - 12 LPA

## Future Scope & Career Growth

Job opportunities are increasing day by day for data-driven insights in various industries. The career growth of data science is robust. It offers various job roles such as data analyst, data scientist, machine learning engineer, data architect and more. Each job profile offers high salaries and secure jobs. In this field, scientists need to continuously learn new tools and methodologies.

## Big Data Market Size Revenue Forecast Worldwide From 2011 to 2027 (in billion U.S. dollars)



## Data Science Platform Market Size, 2022 To 2032 [ USD BILLION ]



# Big Data

## What is Big Data



Big data refers to data that contains several varieties, including increasing velocity and volumes. It is also known as the three V's Volume, Velocity, and Variety. This data, which can be structured, semi-structured, or unstructured, is often too large for traditional data processing tools. Big data technologies like Hadoop and Spark are used to handle, store, and analyze this data. It enables the extraction of valuable insights that impacts better decision-making and innovation across various industries.

## Why Big Data



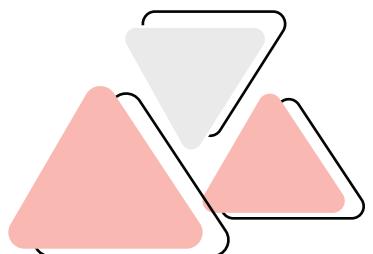
Big data is so important because it allows organizations to analyze big datasets to show valuable insights, optimize decision making and more. By using big data, companies can grow their customer experiences with the help of personalized marketing to improve operational efficiency through automation and real-time monitoring. They can also easily gain a competitive edge by quickly responding to changes in the market. In the healthcare sector, big data provides better diagnoses and personalized treatments to improve patient results. Consequently, big data transforms data into actionable insights to increase growth and efficiency across various industries.

## Job Opportunities & Salary

After learning Big Data, you will be eligible to find a job in several roles, such as Big Data Engineer, Big Data Admin, Big Data Analyst, Big Data Architect, etc. See the below table to see all the information in detail.

Big data analytics in India is to be worth US \$ 16 BN in 2025.

Job Profile	Average Salary
Big Data Engineer	8 - 12 LPA
Big Data Admin	10 - 12 LPA
Big Data Analyst	6 - 8 LPA
Big Data Architect	18 - 25 LPA
Big Data Scientist	12 - 18 LPA
Big Data Modeler	10 - 14 LPA
Big Database Manager	12 - 18 LPA
Big Database Developer	10 - 15 LPA
Big Data Consultant	12 - 18 LPA
Data Visualization Analyst	7 - 10 LPA
BI (Business Intelligence) Developer	8 - 12 LPA



## Machine Learning (ML) & Artificial Intelligence (AI)

### Machine Learning

ML is a subset of artificial intelligence (AI) in which machine learning engineers develop algorithms and statistical models that enable computers to learn and make decisions or predictions based on data. Engineers do not have a program to perform specific programs. In fact, machine learning systems are created to improve their performance over time as they are exposed to more data.

### Artificial Intelligence

AI is a branch of computer science that is focused on creating systems that can solve problems requiring human intelligence. These tasks include learning, reasoning, problem-solving, language understanding, etc. AI is created to develop machines that can act or simulate human perceptual processes.

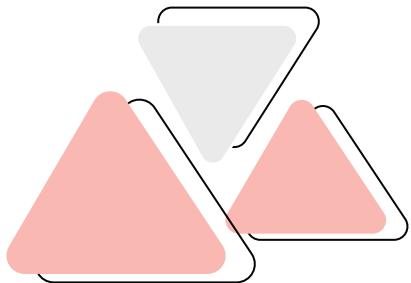
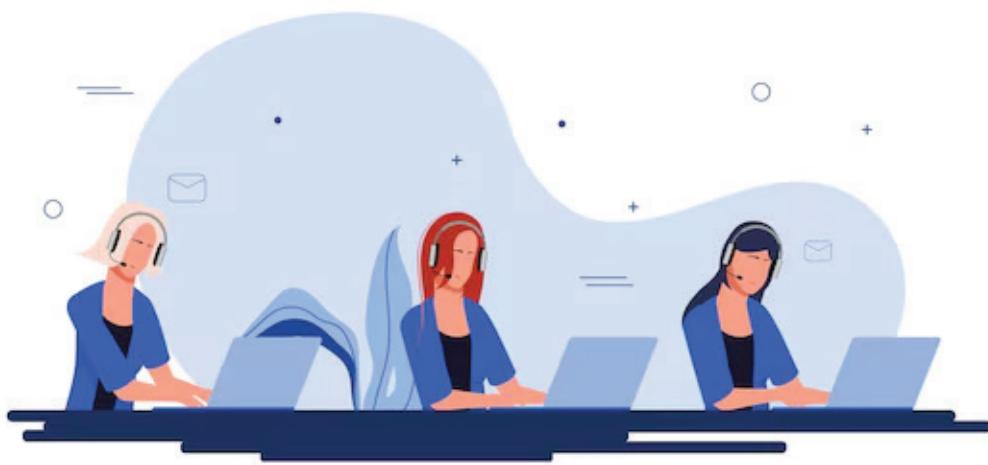
## Why AI & ML



AI and ML are transformative technologies that are used to increase efficiency, automate tasks, and provide data-driven insights by bulk information. They enable personalization in services, advance healthcare through improved diagnostics and treatment, and drive innovation in several industries. By automating repeated processes and optimizing operations AI and ML create new opportunities, improve quality of life through assistive technologies, and offer a competitive advantage to businesses. However, their adoption requires careful management to address ethical concerns, data security, and the need for specialized skills.

# Job Opportunities & Salary

Job Profile	Average Salary
Machine Learning Engineer	8 - 20 LPA
Business Intelligence (BI) Developer	8 - 12 LPA
Algorithm Specialists	10 - 15 LPA
Data Scientist	12 - 18 LPA
AI Research Scientist	15 - 30 LPA
Data Engineer	8 - 16 LPA
AI Product Manager	10 - 20 LPA
Computer Vision Engineer	10 - 18 LPA
NLP (Natural Language Processing) Engineer	10 - 18 LPA



# CURRICULUM

## Programming with Python

- Introduction
- Setting Up the Environment
- Programming Fundamentals
- Inbuilt Functions
- Modules
- **Project 1**
- String & String Formatting
- Data Structures
- Conditional Statements
- Loop & Iterator
- **Project 2**
- Understanding Exception Handling
- User Defined Functions & Modules
- File Handling

## Programming with Python

- Data Manipulation Using Numpy
- Numpy Arrays
- Array Constructor
- Understanding Arrays
- Numpy.Range()
- LINPAC
- Creating 2-D Array
- Indexing Arrays
- Indexing in 1-D Arrays
- Indexing in 2-D Arrays
- Slicing Arrays
- Slicing in 1-D Arrays
- Slicing in 2-D Arrays
- Indexing
- Slicing
- Vectorization in Arrays
- Broadcasting in Arrays
- Scalar Vectorization
- Array Comparison
- New Axis

## Using Pandas

- Pandas And Data Manipulation
- Dropping/Deleting Columns
- Data Frame Indexing
- Series
- Series. Apply ()
- Creating Series
- Applying Method To A Series
- Data Frame & Basic Functionality
- Printing Columns
- DataFrame. Head ()
- Creating DataFrame
- Descriptive Statistical Function
- Data Frame Manipulation
- Mean()
- Correlation
- Indexing & Missing Values
- Grouping & Reshaping

## Wielding Big Data Using Pyspark

- Introduction to Big Data
- What is Cluster Computing?
- Understanding Hadoop Architecture
- What is In-memory Computation?
- Understanding Apache Spark Architecture
- Difference Between Hadoop & Spark
- What is Spark?
- Why Use Pyspark?
- Setting up Databricks and Forming a Cloud Cluster
- Introduction To Data Frames
- Basics of Data Frames
- Understanding Dataframe Operations
- Group By Operations
- Aggregate Operations
- Introduction to Handling Missing Data
- Triggering SQL Query in Pyspark
- Dates And Timestamp
- Milestone Project

## Importing & Cleaning Data

- How to host a Database On a Cloud/Local System?
- Python SQL Connectivity
- CRUD Operation On Database Tables Using Python
- Importing Data Into Pandas Dataframe
- Dealing with Missing Values
- Processing & Cleaning Data Using Pandas Methods

## Data Visualization Using Matplotlib

- Understanding Data Visualization
- Data Visualization Using the `matplotlib.pyplot` Interface
- Types of Charts
- Creating Bar Chart using Matplotlib
- Creating a Stacked Bar Chart using Matplotlib
- Creating Histogram Chart using Matplotlib
- Creating Scatter plots using Matplotlib
- How to add data to an `Axes` object?
- How to Customize Your Plots?
- How to Customize Data Appearance?
- How to Create a grid of subplots?
- Area Plot for Indexed Data

## Data Visualization Using Matplotlib

- Using Seaborn
- Difference Between Seaborn & Matplotlib
- Showing Seaborn Plots
- Using Seaborn Using Matplotlib Defaults
- Set `xlim` or `ylim` in Seaborn
- What are Box and Whiskers?

## Statistical Exploration

- Understanding Graphical Exploratory Data Analysis
- Plotting and Computing Simple summary statistics
- What is Quantitative Exploratory Data Analysis?
- Features of a Dataset with numbers
- Understanding Probabilistically-- Discrete variables
- Understanding Statistical Inference Rests upon Probability
- Thinking probabilistically-- Continuous variables

## Machine Learning

- Introduction to Machine Learning
- History of ML
- Examples of ML
- Artificial Intelligence vs Machine Learning vs Deep Learning
- Machine Learning Life Cycle
- Types of Machine Learning
- Features and Labels

## Supervised Learning

- › What is Supervised Learning?
- › Types of Supervised Learning
- › Regression vs Classification

## Regression

- › Understanding Regression?
- › What is Simple Linear Regression?
- › Understanding Simple Linear Regression Algorithm
- › Example of Simple Linear Regression
- › What is Multiple Linear Regression?
- › Understanding Multiple Linear Regression Algorithm
- › Example of Multiple Linear Regression
- › What is Polynomial Regression?
- › Understanding the Polynomial Regression Algorithm
- › Example of Polynomial Regression
- › What is Backward Elimination?
- › Understanding the Backward Elimination Algorithm
- › Example of Backward Elimination

## Classification

- › Understanding Classification Algorithm
- › What is Naive Bayes?
- › Understanding the Naive Bayes Algorithm
- › Naive Bayes Example
- › What is K-Nearest Neighbours (K-NN)?
- › Understanding The K-Nearest Neighbours (K-NN) Algorithm
- › K-Nearest Neighbours (K-NN) Example
- › What is a Support Vector Machine (SVM)?
- › Understanding the Support Vector Machine (SVM) Algorithm
- › Example of Support Vector Machine (SVM)
- › Understanding Decision Tree Classification
- › Understanding the Decision Tree Classification Algorithm
- › Example of Decision Tree Classification
- › What is Random Forest Classification?
- › Understanding the Random Forest Classification Algorithm
- › Random Forest Classification Example
- › What is Logistic Regression?
- › Understanding the Logistic Regression Algorithm
- › Example of Logistic Regression
- › Classification Model Selection

## Unsupervised Learning

- What is Unsupervised Learning?
- Understanding K-Means Clustering
- Understanding The K-Means Clustering Algorithm
- K-Means Clustering Example
- What is Hierarchical Clustering?
- Understanding the Hierarchical Clustering Algorithm
- Example of Hierarchical Clustering
- Understanding Fuzzy K-means
- Understanding The Fuzzy K-means Algorithm
- Fuzzy K-means Example

## Live Data Streaming

- Spark Streaming Using Python
- Data in Motion
- Documentation Example
- Converting the Terminal into a Stream For Live Data And Reading From That Live Stream
- **Assignment 1: Real-Time Stock Market Data Analysis & Visualization**
- Doubt Clearing Session

## Reporting with Microsoft Power BI

- Introduction to Power BI
- Setting up the Environment
- What is Power BI Desktop Workflow?
- Understanding the Interface of the Data Model
- Exploring the Query Editor Interface

## Working with the Query Editor

- Introduction to the Query Editor
- How to connect Power BI Desktop to Source Files?
- Keeping & Removing Rows
- Working with Filters
- How to Remove Empty Rows?
- Appending Queries
- Formatting Data
- Handling Formatting Errors
- Understanding Pivoting & Unpivoting Data
- How to Split Columns?
- Data Model - The Star Schema

## Understanding the Data Model: Data & Relationship View

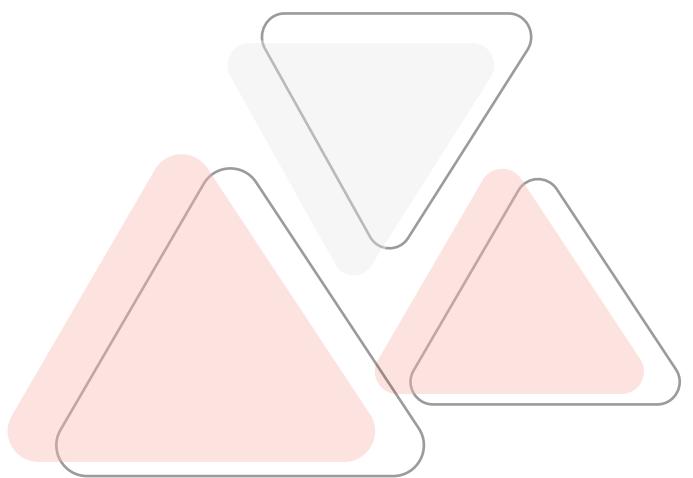
- Introduction to Data & Relationships
- Difference Between M-Language & DAX (Data Analysis Expressions)
- Basics of the DAX
- Understanding Operators
- DAX Functions
- Applying the DAX
- Difference Between Calculated Columns & Measures
- How to Create Measures with Measures?
- **Assignment 2:** How to Manipulate & Process the Above-Loaded Data

## How to Create Visuals in the Report View

- Introduction
- Introduction to Basic Visual Concepts
- Creating First Visuals
- Introduction to Tooltips & Interactions
- Hierarchies & Drill Mode
- Introduction to Data Colors
- What is Conditional Formatting?
- Formatting Report Pages
- What is a Slicer?
- What is Default Summarization & Sorting?
- How to Sync Slicers?
- What are Filter Types?
- How to Create Combined Visuals & Waterfalls?
- Using Custom Visuals
- **Assignment 3:** Visualizing the Above-Processed Data

## Deploying Project Using the Power BI Pro

- Introduction
- What are Workspaces?
- Working with Reports
- How to Create Dashboards?
- How to Refresh Data with Gateways?
- How to share data from "My Workspace"?
- How to Publish an App?



**T**point Tech

**G-13, 2nd Floor / Sector - 3 , Noida  
Contact us - 8448085412**