

Data Science

- Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning.
- The accelerating volume of data sources, and subsequently data, has made data science is one of the fastest growing field across every industry. As a result, it is no surprise that the role of the data scientist was dubbed the “sexiest job of the 21st century” by Harvard Business Review (link resides outside of IBM). Organizations are increasingly reliant on them to interpret data and provide actionable recommendations to improve business outcomes.

The data science lifecycle involves various roles, tools, and processes, which enables analysts to glean actionable insights. Typically, a data science project undergoes the following stages:

- **Data ingestion:** The lifecycle begins with the data collection--both raw structured and unstructured data from all relevant sources using a variety of methods. These methods can include manual entry, web scraping, and real-time streaming data from systems and devices. Data sources can include structured data, such as customer data, along with unstructured data like log files, video, audio, pictures, the Internet of Things (IoT), social media, and more.
- **Data storage and data processing:** Since data can have different formats and structures, companies need to consider different storage systems based on the type of data that needs to be captured. Data management teams help to set standards around data storage and structure, which facilitate workflows around analytics, machine learning and deep learning models. This stage includes cleaning data, deduplicating, transforming and combining the data using ETL (extract, transform, load) jobs or other data integration technologies. This data preparation is essential for promoting data quality before loading into a data warehouse, data lake, or other repository.
- **Data analysis:** Here, data scientists conduct an exploratory data analysis to examine biases, patterns, ranges, and distributions of values within the data. This data analytics exploration drives hypothesis generation for a/b testing. It also allows analysts to determine the data's relevance for use within modeling efforts for predictive analytics, machine learning, and/or deep learning. Depending on a model's accuracy, organizations can become reliant on these insights for business decision making, allowing them to drive more scalability.
- **Communicate:** Finally, insights are presented as reports and other data visualizations that make the insights—and their impact on business—easier for business analysts and other decision-makers to understand. A data science programming language such as R or Python includes components for

generating visualizations; alternately, data scientists can use dedicated visualization tools.

(Syllabus)

(Excel,SQL,Power Pivot,Power Query,Power BI,Python with Data Science,R Language)

MS EXCEL

EXCEL INTRODUCTION

- What is Excel
- Excel Interface
- An overview of the screen, navigation and basic spreadsheet concepts
- Shortcut Keys

CUSTOMIZING EXCEL

- Customizing the Toolbar
- Customizing the Ribbon

FORMATTING AND PROOFING

- Formatting Cells with Number formats, Font formats, Alignment, Borders, etc
- Freeze Panes
- Conditional formatting

ADVANCED PASTE SPECIAL TECHNIQUES

- Paste Formulas, Formats,
- Operation- Add, Multiply, Subtract ,Divide
- Transpose Tables

FORMULA REFERENCE/Cell Reference

- Relative
- Absolute
- Mixed

EXCEL BUILT IN FUNCTIONS(100+ important function)

- Logical Functions
- Math Functions

- Text Functions
- Date and Time Functions
- Statistical Functions
- Information Functions
- Database Functions
- Lookup and Reference Functions
- Financial Functions

WORKGROUP COLLABORATIONS

- Protecting a workbook
- Protecting a worksheet
- Locking /Unlocking cells in a worksheet
- Creating a shared workbook
- Tracking changes to a workbook
- Accepting and protecting a workbook

PRINTING WORKBOOKS

- Setting Up Print Area
- Customizing Headers & Footers
- Designing the structure of a template
- Print Titles –Repeat Rows / Columns

FORMULA TOOLS

- Tracing formula precedents
- Tracing cell dependents
- Error checking

SORTING AND FILTERING

- Filtering on Text, Numbers & Colors
- Sorting Options
- Advanced Filters on different criteria(s)

USING RANGES

- Ranges in Ms Excel
- Naming Range

WORKING WITH THE WEB AND EXTERNAL DATA

- Inserting a Hyperlink
- Importing Data from an Access Database or Text File
- Importing Data from the web and other sources
- Working with Existing Data Connections

SUMMARIZING DATA

- Adding subtotals to a list
- Nesting subtotals
- Text to column

DATA VALIDATION

- Number, Date and Time Validation
- Text and List Validation
- Custom validations based on formula for a cell

USING MACROS

- Macro options
- Creating Macros
- Editing and Deleting Macros

DATA ANALYSIS TOOL

- Goal Seek
- Scenario Manager
- Data Tables

PIVOT TABLES

- Introduction to PivotTables (collapsing and expanding fields, sorting data)
- How to automatically update PivotTables
- Text Filters-how text filters are used to filter a PivotTable based on text.
- Grouping and Pivot Charts-How to group data in a PivotTable and create a chart based on a PivotTable
- Value Field Settings and Show Values As-how to summarize data as a Percentage of a Row or Column.
- Creating Calculated Fields and Calculated Items
- Using Slicers, Filters, and Timelines to quickly analyse data

CHARTS AND SLICERS

- Creating a chart with the 2D or 3D
- Create column chart, pie chart
- Various Charts i.e. Bar Charts / Pie Charts / Line Charts
- Moving a chart one to another sheet
- Formatting category & value Axis data
- Formatting a data series
- Changing a charts source data
- Using SLICERS, Filter data with Slicers

ARRAY FUNCTIONS

- What are the Array Formulas, Use of the Array Formulas?
- Basic Examples of Arrays (Using ctrl+shift+enter).
- Array with if, len and mid functions formulas.
- Array with Lookup functions.
- Advanced Use of formulas with Array.

LOOKUP FUNCTIONS

- Vlookup / Hlookup
- Row , Rows , Column , Columns
- Index and Match
- Reverse Lookup using Match()+Index() Function

SQL

1. Introduction to SQL

(What is SQL?, Purpose of SQL, Who should learn SQL?, What are the subsets of SQL?, Data Definition Language, Data Manipulation Language, Data Control Language, and SQL vs. NoSQL)

2. Introduction to Databases and RDMBS

(What is a Database?, Database Objects, Database Tables, Table Records, Types of Database Management Systems, Relational Database Management Systems, and SQL/Relational Databases vs. No SQL Databases)

3. Install a Database Engine

(Download MS SQL Server or Oracle or MySQL Database Engine, and Install. Launch SQL Server Management Studio, Select New Query, and launch SQL Query. Type SQL Commands and Execute.)

4. SQL Syntax

(Focus on SQL Syntax, SQL keywords, SQL is not case sensitive, SQL Comments, SQL Commands, and writing SQL Statements.)

5. SQL Data Types

(SQL Numeric data types, Date and Time data types, Character and String data types, and Miscellaneous data types.)

6. SQL Operators

(SQL Arithmetic Operators, Comparison Operators, Logical Operators,

7-SQL Database

SQL CREATE Database,SQL DROP Database,SQL RENAME Database,SQL SELECT Database

8-SQL Table

What is TableSQL, CREATE TABLE,SQL DROP TABLE,SQL DELETE TABLE,SQL RENAME TABLE,SQL TRUNCATE TABLE,SQL COPY TABLE, SQL ALTER TABLE

9-SQL Select

SELECT Statement,SQL SELECT UNIQUE,SQL SELECT DISTINCT,SQL SELECT COUNT,SQL SELECT TOP,SQL SELECT FIRST,SQL SELECT LAST,SQL SELECT RANDOM,SQL SELECT IN,SQL SELECT Multiple,SQL SELECT DATE,SQL SELECT SUM,SQL SELECT NULL

10-SQL Clause

Group by ,Group by with Having etc.

11-SQL Order By

ORDER BY Clause,ORDER BY ASC,ORDER BY DES.

12-SQL Insert

INSERT Statement,INSERT INTO Values,INSERT INTO Table by select.

13-SQL Update

UPDATE Statement,UPDATE Specific record,SQL UPDATE DATE

14-SQL Delete

DELETE Statement,SQL DELETE TABLE,SQL DELETE ROW,SQL DELETE All Rows,DELETE Duplicate Rows,SQL DELETE DATABASE,SQL DELETE VIEW,SQL DELETE JOIN

15-SQL Join

SQL JOIN,SQL Outer Join,SQL Left Join,SQL Right Join,SQL etc.

16-SQL Keys

Primary Key,Foreign Key,Unique Key etc.

17- SQL Functions

Sum(),Average(),Count(),Max(),Min() etc.

18-String Functions

19-Math Function

POWER PIVOT

1. Intro to Power Pivot

2. Data Modeling Concept
3. Relationships in Pivot Tables
4. Calculated Item and Calculated Field in traditional Pivot Tables
5. Calculated Column vs Calculated Field (Measures)
6. Intro to Data Analysis Expressions (DAX)
7. Important DAX Formulas and Functions
8. Practical Case Studies using Power Pivot
9. Usage of KPI

POWER QUERY

Section 1: Intro to Power Features

1. Versions and Compatibility
2. Intro to ETL Concepts
3. Around Get & Transform Data Tab

Section 2: Working with Power Query

4. Power Query Editor and User Interface
5. Manage Columns and Rows
6. Transform Table, Column etc
7. Transform Text, Number and Date Columns

8. Column from Example vs Flash Fill

Section 3: Combine with Power Query

9. Append Queries

10. Case Studies on Append Queries

- a) Many Tables Data from Same Worksheet
- b) Data from Different Sheets of Same Workbook
- c) Data from Difference Workbook
- d) Data from Multiple Sheets of Different Workbooks

11. Concept of Merge

12. Different Types of Joining Queries

- a) Left Outer Join
- b) Right Outer Join
- c) Full Join
- d) Inner Join
- e) Left Anti Join
- f) Right Anti Join

13. Fuzzy Matching

14. Practical Case Studies

Section 4: Importing of Data into Power Query

15. Importing of Excel Files

16. Importing of Text Files

17. Connecting to the Folder

18. Connecting to the Web Pages

Section 5: Transformation Techniques

19. Concept of Split with delimited, positions, characters

20. Concept of Unpivot and Pivot
21. Concept of Group by
22. Using Standard transformation features
23. Practical Case Studies
24. Diff b/w Transform Tab vs Add Column Tab
25. Duplicate vs Reference Query

Section 6: Mash-up Language

26. Intro to M-Formula
27. Understanding the "let..in.." statements
28. Writing Custom Functions
29. Invoking Custom Functions
30. Practical Case Studies

POWER BI(Syllabus)

Introduction to Power BI

- An Overview of the Workflow in Power BI Desktop
- Download and Installing Power BI Desktop
- Getting to know the Interface
- Power BI Free Vs. Pro Vs. Premium
- Power BI Key Components
- Introducing the Different Views of the Data Model
- Taking a Closer Look at the Query Editor Interface

Power BI Vs Excel Reporting

- When we use which reports

Power BI overview Concept

- Dataset
- Reports
- Dashboard
- Workspace
- Tile

Connecting Power BI Desktop to our Data Sources

- Excel
- CSV
- Web
- PDF
- MS Access
- SQL Developer
- XML
- Connecting to a folder
- Others

Power Query

- When to Use Power Query
- Import and analysis millions of Rows
- Best things to happen in Power BI
- Type of Connection
- Automate your data processing
- Query Editor Tool
- Pivoting and Un pivoting Columns
- Duplicating Vs Referencing Queries
- Transformations

Editing Rows

- First Header Row
- Remove Top Rows
- Filtering- Remove-Replace
- Techniques

Editing columns

- Remove or Keep Columns
- Navigation to the Columns
- Change Columns Header Name

Understanding Append Queries

- Append Queries as existing (With two or multiple tables)
- Append Queries as New
- Formatting Data
 - Change data Type
 - Detect Data type for multiple columns
- Remove error while loading
- Remove null values
- Splitting Columns
- Delimiters

NETWORK NET(By : Avinash Sir)

- Delimiters positions
- Creating a New Group for our Queries
- Others
 - Sorting
 - Performing Basic Mathematical Operations in Query Editor
 - Improving Performance and Loading Data into the Data Model
 - Columns distribution, Quality, Profile

Joins

- Outer Join
- Left, Right, Full, Anti
- Inner join

Data Table Vs Lookup Table

- Primary key Vs, Foreign Key
- Relationship Vs. Merged Table
- Connecting Lookup to Lookups
- Relationship Cardinality
- Introducing and Understanding Relationships
- One to One Relationship
- One to Many Relationship
- Many to Many Relationship

Working on our Data Model Data Relationship View

- Query Editor vs. Data Model
 - Feature of Query Editor (Structure + Prepare Data)
 - Data Model – Add Relationship, Add Measures (Analysis Data)
- Active properties
- Understanding the differences between the M-Language and DAX
- DAX – Basic information
- Calculated Columns vs. Measures
- Introducing DAX-Measures
- DAX-Measures

DAX (Data Analysis Expression)

- Math & Stats Functions
 - Sum, Average, Max/Min, Divide, Count/Counta, Countrows, Distinctcount

Functions: SUMX, AVERAGEX, MAXX/MINX, RANKX, COUNTX

- Logical Function
 - IF, IFERROR, AND, OR, NOT, SWITCH

NETWORK NET(By : Avinash Sir)

- Text Function
 - Concatenate, Format, Left/Mid/Right,
 - Upper/Lower, Proper, Len, Search/Find,
 - Replace, Rept, Substitute, Trim
- Filter Function
 - Calculate, Filter, All, Allexcept, Related, Distinct, Values
- Date Function
 - Datediff, Yearfrace, Year/Month/Day, Hour/Minute/Second,
 - Today/Now, Weekday/Weeknum

Working in the Report View to Visualize our Results

- Understanding the Interface of the Report View
- Creating our First Visualizations
- Editing Interactions and Adding Tooltips
- Adding Color Saturation
- Understanding Hierarchies and Adding Drill-Down
- Formatting charts and Sorting
- Essential Formatting setting: Color, Data Labels, Axis, Placement, legends, Value
- Table View vs. Matrix View
- Conditional Formatting in Table View
- Basic Charts: Line, Bar Chart, Column Chart, Tree map, Pie, Doughnut, Area
- Tooltips – On hover Text
- Color Saturation in Column Chart
- Slicers: Vertical list Vs. Horizontal tiles
- Text Cards
- Adding Tree maps and Tables
- Applying different Filter Types
- Creating Multi-Row Cards
- Understanding Combined Visualizations and Waterfalls

Advance Visualization:

- KPI & Gauge Chart
- Forecast, Trent-line
- Map Visuals
- Custom Visuals – Downloadable Charts
- Use Drill down/up feature to specify chart behavior – granular view
- Show/Expand next level
- Filter level: Report level vs Page level vs Visual level

Power BI Service

- Introducing Power BI Service
- Connecting to Data
- Gateway Refresh and Data Flow

- Understanding Gateways
- Gateway Types
- Reports & Dashboard
- Sharing & Collaboration Tool

Data Science Using Python

GETTING STARTED

• History & need of Python • Application of Python • Advantages of Python • Disadvantages of Python • Installing Python • Program structure • Interactive Shell • Executable or script files. • User Interface or IDE

PYTHON FUNDAMENTALS

• Working with Interactive mode • Working with Script mode • Python Character Set • Python Tokens, Keywords, Identifiers, Literals, Operators • Variables and Assignments • Input and Output in Python

OPERATORS

• Arithmetic Operators • Relational Operators • Logical Operators • Membership Operators • Identity Operators • Bitwise Operators • Assignment Operators • Operators Precedence • Evaluating Expression • Type Casting

DATA HANDLING

• Data Types Numbers → Strings → Lists → Tuples → Dictionary → Set → Frozenset → Bool → Mutable and Immutable

STRING MANIPULATION

• Introduction to Python String • Accessing Individual Elements • String Operators • String Slices • String Functions and Methods

LIST MANIPULATION

- Introduction to Python List • Creating List • Accessing List • Joining List • Replicating List • List Slicing

TUPLES

- Introduction to Tuple • Creating Tuples • Accessing Tuples • Joining Tuples • Replicating Tuples • Tuple Slicing

DICTIONARIES

- Introduction to Dictionary • Accessing values in dictionaries • Working with dictionaries • Properties

- SET AND FROZENSET** • Introduction to Set and Frozenset • Creating Set and Frozenset • Accessing and Joining • Replicating and Slicing

PROGRAM CONTROL FLOW

- **Conditional Statements** → The if Statement → The if-else Statement → The if-elif Statement → Nested if Statements → Python Indentation

- Looping and Iteration** → The For Loop → The While Loop → Loop else Statement → Nested Loops → Break and Continue

- **The Range Function** → Introduction to range() → Types of range() function → Use of range() function

INTRODUCTION TO FUNCTIONS

- Built-In Functions → Introduction to Functions → Using a Functions → Python Function Types → Structure of Python Functions → E.g. - map, zip, reduce, filter, any, chr, ord, sorted, globals, locals, all, etc

- User Defined Functions** → Structure of a Python Program w.r.t. UDF → Types of Functions → Invoking UDF → Flow of Execution → Arguments and Parameters → Default Arguments, Named Arguments → Scope of Variables → Lambda function • Recursion Function → Use of recursion function

MODULES AND PACKAGES

- Built-in Modules → Importing Modules in Python Programs → Working with Random Modules → E.g. - builtins, os, time, datetime, calendar, sys, etc.
- User Defined Functions → Structure of Python Modules

FILE OPERATIONS

- Text and Bytes files → Opening a file → Reading and Writing Files → Other File tools

MS Excel files and Other → Introduction to MS Excel files, CSV, JSON files.

INTRODUCTION TO DATA SCIENCE AND Data Analytics

- Why Analytics? • Traditional Data Management • Analytical tools • Types of Analytics • Hind sight, ore sight and insight • Dimensions and measures • Why learn Python for data analysis? • Using the IPython notebook

LIBRARIES FOR DATA Analytics/Science

Numpy, Pandas, Matplotlib and Seaborn

NUMPY

- Creating NumPy arrays • Indexing and slicing in NumPy • Downloading and parsing data • Creating multidimensional arrays • NumPy Data types • Array tributes • Indexing and Slicing • Creating array views copies • Manipulating array shapes I/O

PANDAS

- Using multilevel series • Series and Data Frames • Grouping, aggregating • Merge DataFrames • Generate summary tables • Group data into logical pieces • Manipulate dates • Creating metrics for analysis • Data wrangling • Merging and joining • Analytics Vidhya dataset- Loan Prediction Problem • Data Mugging using Pandas • Building a Predictive Model

MATPLOTLIB

• Scatter plot • Bar charts, histogram • Stack charts • Legend title
Style • Figures and subplots • Plotting function in pandas • Labelling
and arranging figures • Save plots

Seaborn

Getting Started, Using Seaborn with Matplotlib, Customizing Seaborn
Plots, Changing the figure Size, Scaling the plots, Setting the Style
Temporarily, Color Palette, Diverging Color Palette, Sequential Color
Palette, Multiple plots with Seaborn, Using Matplotlib, Using
Seaborn, Creating Different Types of Plots, Relational
Plots, Categorical Plots, Distribution Plots, Regression Plots. More
Gaphs in Seaborn

Other - (Only Overview)

- ✓ *Introduction to Machine Learning With Python*
- ✓ *Introduction to Deep Learning*

➔ **Note :- More than 200 programs will be practiced during the course.**

Popular IDE/Editor for 'Python' language

- IDLE(BY DEFAULT)
- PyCharm
- VS Code
- Sublime Text3
- Atom
- Jupyter etc.

R Language

R Installation

RStudio IDE
R Advantage & Disadvantage
R Hadoop Integration

R Packages
List of R Packages

R Basic

Syntax
R Data Types
R Data Structures
R Variables
R Keywords
R Operators

R Statements

R If Statement
If-else Statement
else if Statement
R Switch Statement
R Next Statement
R Break Statement

R Loops

R For Loop
R Repeat Loop
R While Loop

R Functions

R Functions
R Built-in Functions

R Data Structures

R Vectors
R Lists
R Arrays
R Matrix
R Data Frame
R Factors

R Advance

Data Reshaping

Object-Oriented Programming

R Debugging

Data Interfaces

R CSV File

R Excel File

R Binary File

R JSON File

R XML File

R Database

Data Visualization

R Data Visualization

R Pie Charts

R Bar Charts

R Boxplot

R Histogram

R Line Graphs

R Scatterplots

R Regression

Linear Regression

Multiple Regression

Logistic Regression

Poisson Regression

R Statistics

Normal Distribution

Binomial Distribution

R Classification Time Series Analysis

R Random Forest

T-Test in R

Chi-Square Test

R vs Python

NETWORK NET