DATA ANALYST SYLLABUS

1. Introduction to Data Analyst

- Types of Data
  Qualitative/ Categorical
  Quantitative/ Numerical
- Exploratory Data Analysis
  Learn the Structure of the Data
  Uncover Patterns or Errors
  Find relationships and insights
- EDA Steps
  - Data Preparation
    Sources of Data
    Data Collection
    Data Cleaning and Wrangling
    Data Sanity Check
  - Data Exploration (Learn about each variable, Compute Summary Statistics, Find Correlation and Trends, Visualize the data)
  - Hypothesis Generation and Further analysis
- Metrics and Analysis
  - Understanding different types of metric
  - User journeys and investigating abnormal behaviours
- Data Exploration
  - Frequency Distribution
  - Measure of Central Tendency (Mean, Median, Mode)
  - Measure of Dispersion (Variance and Standard Deviation)
  - Z-score
- Data Visualization
  - Types of Charts
  - Selecting Charts
    - By Use-Case
    - By Data Types
  - Styling Charts
  - Explainability of the Chart

2. Data Preparation using Excel

- Introduction to Excel
- Formatting Cells
- Keyboard shortcuts
- Copy Paste in Excel
- Functions
- Filters
- Sorting
- Loading and Cleaning Data
  - Gathering Raw Data
  - Removing Duplicates
  - Fill Options
  - Data Validation
- Numerical Data Types
  - Operators
  - Range, Average, Count, Rounding
  - Variance
  - Summarising
- Handling Text Data Types
  - Categorical Variables
  - Cases and Spaces
  - Cleaning Strings
- Working with Dates and TimeStamps
  - Operators
  - Aggregations
  - Time Between
- Logical Functions
  - AND, OR, NOT, IF
  - Combining logical functions
  - Aggregate Logical Functions
- Data Protection
  - Protecting Sheets & Cells

3. Data Analysis & Visualisation Using Excel

- Summary Statistics
  - Measure of Central Tendencies
  - Range, Variance
- Referencing
  - LOOKUP functions
  - Index Match
- Summarising with Pivot tables
  - Introduction to pivot tables
  - Slicer, Multiple Pivot tables
- Data Visualisation
  - Charts

- o Formatting Charts
- o Building Dashboards
- Using Pivot Tables
- What If-Analysis
  - o Scenario Analysis
  - o Sensitivity Analysis
  - o Growth Rate
  - o What-If Analysis in Excel
- Forecasting
  - o Seasonality
  - o Reducing Bias
  - o Confidence Intervals
  - o Moving Averages
  - o Weighted Averages
  - o Techniques in Excel

4. Data Analysis using SQL

- Introduction to Databases
- Difference between SQLite, MySQL, PostgreSQL etc
- Database Terminologies (Tables/ Relations, Record/Raws, Schema, Field, Unique Identifiers, Primary Key, Relationships, Foreign Keys, Constraints)
- Overview of SQL (DDL, DML, Queries)
- Data Types (NULL, INTEGER, REAL, FLOT, NUMERIC, TEXT, CHAR / VARCHAR, BLOB)
- DDL (Create DB, Create Table, etc)
- Anatomy of a Query
- SELECT, FROM, WHERE Clause
- Aliasing
- Operators
  - o Relational Operators ($<$, $<=$,$>$,$>=$, $+$, $<>$, !=)
  - o Logical Operators (AND, OR, NOT)
  - o LIKE, BETWEEN
- GROUPBY, DISTINCT, HAVING Clause
- SORT, LIMIT
- Order of Queries

5. Advanced Data Analysis Using SQL

- Numerical Data Types
  - o Numeric Types, Operators
  - o Range, Average and Mean
  - o Variance, Rounding and Summarising
- Exploring Distributions
- Summarising
  - o Correlation function
  - o Median/Percentile Discrete and Continous
- Character Data Types
  - o Data Types, Categorical Variables
  - o Grouping, Counting and Ordering

- o Cases and Spaces
- o Searching in Strings, Trimming Spaces
- o Splitting, Concatenating, Full Text Search
- Working with Dates and Timestamps
  - o Types and Formats
  - o Comparisons and Operations
  - o Components and Aggregation
  - o Aggregating with date/time series
  - o Time between Events
    - ▪ Lead and Log
    - ▪ Average time between events
    - ▪ Change in time series
- Working with Arrays
  - o Arrays in PostgreSQL (CREATE, INSERT etc)
  - o Accessing Arrays
  - o Searching Arrays
  - o Array Functions and Operators

6. Data Manipulation with SQL

- Joins and Set Operations
  - o Relationships between Tables
  - o Inner Joins
  - o Outer Joins
  - o Joins on Join
  - o Cross Joins
  - o Self Joins
  - o Equi and Non Equi Joins
  - o Set Operators
- Data Manipulation Techniques
  - o Case Statements
  - o Subqueries
  - o Correlated Subqueries
  - o Nested Subqueries
  - o Common Table Expression
  - o Window Functions

7. Database Design with SQL

- Introduction to OLTP and OLAP
- Storing Data
  - o Types of Data
  - o Data Warehouses
  - o Data Lakes
  - o ETL and ELT
- Data Modelling and Schema Design
  - o Conceptual Data Model
  - o Logical Data Model
  - o Physical Data Model

- Dimension Modelling
  - Fact Tables, Dimension Tables
  - Star Schema
  - Snowflake Schema
- Normalization and Denormalization
- Database Views
- Scalability
  - Partitioning (Vertical Partitioning, Horizontal Partitioning)
  - Sharding
- Data Integration
  - Data Sources, Transformation, Unified Data Model
  - Update Cadence
  - ETL

8. Introduction to Data Visualisation

- Data Visualisation
  - Why is visualisation important?
  - Visualisation Framework
  - Chart Types
  - Trend Visualisation
  - Tips and Tricks for Visualisation
- Data Story Telling
  - What is Data Storytelling?
  - Biases- When do they appear?
  - Biases- Formal Terminology

9. Data Preparation & Exploration with Tableau

- Introduction to Tableau
- Loading Data in Tableau
  - Data Sources and Loading Data Types
  - Joins and Relationships
  - Fields in Data and their types
  - Dimensions of Measures
  - Data Roles
  - Navigating UI elements
- Combining Data
  - Unions, Joins, Relationships
- Filtering and Sorting
  - Types of filters
  - Filtering on Dimensions, Measures
  - Sorting and Filtering through Selections
- Aggregation
  - Aggregating Measures
  - Aggregating Dimensions
  - Scatter Plot and Aggregations
- Calculated Fields and Table Calculations
  - Functions/ Operators

- o Formatting Numbers
- o Type Conversions
- o Level of Details Expression (LOD)
- o Table Calculations

10. Data Visualisation

- Chart Types with Tableau
- Exploratory Analysis using Visualisation of Trends
    - o Reference Lines, Trend Lines and Forecasting
    - o BarCharts and Line Charts for discrete and continuous data
    - o Discrete Time Analysis
    - o Quick Tables
    - o Formatting through Colours
    - o Bubble Chart
- Mapping your Data
    - o Geographic Data Types
    - o Geocoding
    - o Creating Maps
- Dashboards and Stories
    - o Introduction to Dashboards
    - o Introduction to Stories (Story Points,etc)
    - o Creating Dashboards and Stories
    - o Building a KPI Dashboard
    - o Updating the Tooltip
- Analysis
    - o Seasonality Analysis
    - o YoY Analysis
    - o YTD Calculation
    - o Calculating Growth
    - o CAGR Analysis
    - o Moving Rolling Calculations
    - o Cohort Analysis

11. Essentials of Python for Data Analysis

- Python Environment setup
- Python Data Types
    - o Variables
    - o Python is Dynamically Typed
    - o Rules for Naming variables and Naming Conventions
    - o Overview of Python Data Types
        - ▪ Numeric
        - ▪ Sequence
        - ▪ Set
        - ▪ Dictionary
    - o Literals or constants
    - o Type Conversion
- Operators and Expression
    - o Arithmetic Operators

- o Expressions
- o Operator Precedence
- o Arithmetic Operators on All Data Types
- o Assignment Operators
- o Relational Operator
- o Logical Operators
- o Boolean
- o Special Operators
- o Mathematical
- Conditional Statements
- Loops & Control Flow
- String
- Lists
- Tuple

12. Advanced Data Analysis using Python

- NumPy
  - o Introduction to NumPy and importing NumPy
  - o Array Creation
  - o NumPy Attributes
  - o Creating Different Types of Arrays
  - o Accessing elements of an Array
  - o NumPy Slicing
  - o Reshaping & Flattening an Array
  - o Data Types in NumPy
  - o Operators
  - o Data Analysis using NumPy
- Pandas
  - o Datasets in Python
  - o DataFrame
  - o Accessing Data
  - o Filtering
  - o Traversing Data Frame
  - o Sorting
  - o Merging-Data/ Joins in Pandas
    - Inner Joins
    - Outer Joins
    - Self Joins
    - Merging on indexes
    - Filtering Joins
- Concatenating Data Vertically
- Data Integrity Check
- Reshaping Data

13. Data Visualisation using Python

- Visualisation
  - o Introduction to MatPlotlib

- o Line Plots, Scatter Plots, Histograms, BarPlots and Vertical BarPlots
- o Customising Plots
- o Introduction to SeaBorn
- o Relational Plots and SubPlots
- o Customising Scatter Plots
- EDA Using Pandas
  - o Feature Engineering
  - o Summary Statistics
  - o Data Validation and Cleanup
  - o Group Summary Statistics
  - o Pivot Tables
  - o Explicit Indexes
  - o Checking for Missing Values
  - o Handling Outliers
  - o Patterns over date time
  - o Correlation
  - o Updating CSV Files

14. Interview Prep & Portfolio Building

- Data Analysis Process Interview questions
- Technical interview questions
- SQL interview questions
- Building your resume
- Building your data Portfolio
- Graduation Test & Projects

BONUS CONTENT

Descriptive and Inferential Statistics

- Z-core
- Central Limit Theorem, p-value
- Hypothesis testing
- Power Analysis

◆◆ NEWGEN CORPORATE TRAINING CENTER ◆◆

OFFLINE / ONLINE BATCH

Address: NEWGEN CORPORATE TRAINING CENTER, Bremen Chowk, Office No.217, West Avenue,

Above Atithi Hotel, Aundh, Pune – 411027

URL: http://www.newgensoftech.com

Mail Id: balkrishna8588@gmail.com

Insta Id: @Newgen_Softech