CampusX Data Science Mentorship Program

Week 1: Basics of Python Programming

1. **Session 1: Python Basics**
    - Short info about DSMP (3:38 - 6:57)
    - About Python (7:00 – 24:30)
    - Python Output/print function (24:30 – 38:37)
    - Python Data Types (38:37 – 51:25)
    - Python Variables (51:25 – 1:04:49)
    - Python comments (1:04:49 – 1:09:09)
    - Python Keywords and Identifiers (1:09:09 – 1:22:38)
    - Python User Input (1:22:38 – 1:35:00)
    - Python Type conversion (1:35:00 – 1:47:29)
    - Python Literals (1:47:29 – 2:10:22)
2. **Session 2: Python Operators + if-else + Loops**
    - Start of the session (00:00:00 – 00:09:02)
    - Python Operators (00:09:02 – 00:43:00)
    - Python if-else (00:43:00 – 01:14:50)
    - Python Modules (01:14:50 – 01:24:48)
    - Python While Loop (01:24:48 – 01:48:03)
    - Python for loop (01:48:03 – 02:11:34)
3. **Session 3: Python Strings**
    - Introduction (00:00:00 – 00:09:08)
    - Solving Loop problems (00:00:08 – 00:47:10)
    - Break, continue, pass statement in loops (00:47:10 – 01:06:42)
    - Strings (01:06:42 – 1:14:15)
    - String indexing (01:14:15 – 01:18:14)
    - String slicing (01:18:14 – 01:27:06)
    - Edit and delete a string (01:27:06 – 01:32:14)
    - Operations on String (01:32:14 – 01:47:24)

- Common String functions (01:47:14 – 02:22:53)

4. **Session on Time complexity**
    - Start of the Session (00:00:00 – 00:11:22)
    - PPT presentation on Time Complexity (Efficiency in Programming and Orders of Growth) (00:11:22 – 01:12:30)
    - Examples (01:12:30 – 01:42:00)
5. **Week 1 Interview Questions**

Week 2: Python Data Types

1. **Session 4: Python Lists**
    - Introduction
    - Array vs List
    - How lists are stored in a memory
    - Characteristics of Python List
    - Code Example of Lists
        i. *Create and access a list*
        ii. *append(), extend(), insert()*
        iii. *Edit items in a list*
        iv. *Deleting items from a list*
        v. *Arithmetic, membership and loop operations on a List*
        vi. *Various List functions*
        vii. *List comprehension*
        viii. *2 Ways to traverse a list*
        ix. *Zip() function*
        x. *Python List can store any kind of objects*
    - Disadvantages of Python list
2. **Session 5: Tuples + Set + Dictionary**
    - **Tuple**
        i. *Create and access a tuple*
        ii. *Can we edit and add items to a tuple?*

iii. *Deletion*
iv. *Operations on tuple*
v. *Tuple functions*
vi. *List vs tuple*
vii. *Tuple unpacking*
viii. *Zip () on tuple*

- **Set**
  i. *Create and access a set*
  ii. *Can we edit and add items to a set?*
  iii. *Deletion*
  iv. *Operations on set*
  v. *set functions*
  vi. *Frozen set (immutable set)*
  vii. *Set comprehension*
- **Dictionary**
  i. *Create dictionary*
  ii. *Accessing items*
  iii. *Add, remove, edit key-value pairs*
  iv. *Operations on dictionary*
  v. *Dictionary functions*
  vi. *Dictionary comprehension*
  vii. *Zip() on dictionary*
  viii. *Nested comprehension*

3. Session 6: Python Functions

- Create function
- Arguments and parameters
- args and kwargs
- How to access documentation of a function
- How functions are executed in a memory
- Variable scope
- Nested functions with examples
- Functions are first class citizens
- Deletion of function
- Returning of function
- Advantages of functions
- Lambda functions

- Higher order functions
- map(), filter(), reduce()

4. **Paid Session on Career QnA**
5. **Array Interview Questions**
6. **Week 2 Interview Questions**

Week 3: Object Oriented Programming (OOP)

1. **Session 7: OOP Part1**
    - What is OOP?
    - What are classes and Objects?
    - Banking application coding
    - Methods vs Functions
    - Class diagram
    - Magic/Dunder methods
    - What is the true benefit of constructor?
    - Concept of 'self'
    - Create Fraction Class
    - __str__, __add__, __sub__ , __mul__ , __truediv__
2. **Session 8: OOP Part2**
    - Revision of last session by solving problems
    - How objects access attributes
    - Attribute creation from outside of the class
    - Reference Variables
    - Mutability of Object
    - Encapsulation
    - Collection of objects
    - Static variables and methods
3. **Session 9: OOP Part3**
    - Class Relationship
    - Aggregation and aggregation class diagram
    - Inheritance and Inheritance class diagram
    - Constructor example
    - Method Overriding
    - Super keyword
    - Super constructor
    - Practice questions on Inheritance

- Types of Inheritance (Single, Multilevel, Hierarchical,Multiple)
- Hybrid Inheritance
- Code example and diamond problem
- Polymorphism
- Method Overriding and Method Overloading
- Operator Overloading

4. **Session on Abstraction**
    - What is Abstraction?
    - Bank Example Hierarchy
    - Abstract class
    - Coding abstract class (BankApp Class)
5. **Session on OOP Project**
6. **Week 3 Interview Questions**

Week 4: Advanced Python:

1. **Session 10: File Handling + Serialization & Deserialization**
    - How File I/O is done
    - Writing to a new text file
    - What is open()?
    - append()
    - Writing many lines
    - Saving a file
    - Reading a file -> read() and readline()
    - Using context manager -> with()
    - Reading big file in chunks
    - Seek and tell
    - Working with Binary file
    - Serialization and Deserialization
    - JSON module -> dump() and load()
    - Serialization and Deserialization of tuple, nested dictionary and custom object
    - Pickling
    - Pickle vs JSON
2. **Session 11: Exception Handling**
    - Syntax Error with Examples
    - Exception with Examples

- Why we need to handle Exception?
- Exception Handling (Try-Except-Else-Finally)
- Handling Specific Error
- Raise Exception
- Create custom Exception

3. **Session 12: Decorators and Namespaces**
    - Namespaces
    - Scope and LEGB rule
    - Hands-on local, enclosing, global and built-in scope
    - Decorators with Examples
4. **Session on Iterators**
    - What are iterators
    - What are iterables
    - How for loop works in Python?
    - Making your own for loop
    - Create your own range function
    - Practical example to use iterator
5. **Session on Generator**
    - What is a generator?
    - Why to use Generator?
    - Yield vs Return
    - Generator Expression
    - Practical Examples
    - Benefits of generator
6. **Session on Resume Building**
7. **Session on GUI Development using Python**
    - GUI development using tkinter
8. **Week 4 Interview Questions**

Week 5: Numpy

1. **Session 13: Numpy Fundamentals**
    - Numpy Theory
    - Numpy array
    - Matrix in numpy
    - Numpy array attributes
    - Array operations

- Scalar and Vector operations
- Numpy array functions
    i. *Dot product*
    ii. *Log, exp, mean, median, std, prod, min, max, trigo, variance, ceil, floor, slicing, iteration*
    iii. *Reshaping*
    iv. *Stacking and splitting*

2. **Session 14: Advanced Numpy**
    - Numpy array vs Python List
    - Advanced, Fancy and Boolean Indexing
    - Broadcasting
    - Mathematical operations in numpy
    - Sigmoid in numpy
    - Mean Squared Error in numpy
    - Working with missing values
    - Plotting graphs
3. **Session 15: Numpy Tricks**
    - Various numpy functions like sort, append, concatenate, percentile, flip, Set functions, etc.
4. **Session on Web Development using Flask**
    - What is Flask library
    - Why to use Flask?
    - Building login system and name entity recognition with API

Week 6: Pandas

1. **Session 16: Pandas Series**
    - What is Pandas?
    - Introduction to Pandas Series
    - Series Methods
    - Series Math Methods
    - Series with Python functionalities
    - Boolean Indexing on Series
    - Plotting graphs on series
2. **Session 17: Pandas DataFrame**
    - Introduction Pandas DataFrame
    - Creating DataFrame and read_csv()

- DataFrame attributes and methods
- Dataframe Math Methods
- Selecting cols and rows from dataframe
- Filtering a Dataframe
- Adding new columns
- Dataframe function – astype()

3. **Session 18: Important DataFrame Methods**
    - Various DataFrame Methods
    - Sort, index, reset_index, isnull, dropna, fillna, drop_duplicates, value_counts, apply, etc.
4. **Session on API Development using Flask**
    - What is API?
    - Building API using Flask
    - Hands-on project
5. **Session on Numpy Interview Question**

Week 7: Advanced Pandas

1. **Session 19: GroupBy Object**
    - What is GroupBy?
    - Applying builtin aggregation fuctions on groupby objects
    - GroupBy Attributes and Methods
    - Hands-on on IPL dataset
2. **Session 20: Merging, Joining, Concatenating**
    - Pandas concat method
    - Merge and join methods
    - Practical implementations
3. **Session on Streamlit**
    - Introduction to Streamlit
    - Features of Streamlit
    - Benefits of Streamlit
    - Flask vs Streamlit
    - Mini-project on Indian Startup Funding Dataset using Streamlit – Part 1
4. **Session on Pandas Case Study (Indian Startup Funding)**
    - Data Analysis on Indian Startup Funding Dataset and display results on the Dashboard made by Streamlit – Part 2
5. **Session on Git:**

- What is Git?
- What is VCS/SCM?
- Why Git/VCS is needed?
- Types of VCS
- Advantages
- How Git works?
- Installing git
- Creating and Cloning repo
- add, commit, add ., gitignore
- seeing commits (log -> oneline)
- Creating versions of a software

6. **Session on Git and GitHub:**
    - Nonlinear Development (Branching)
    - Merging branches
    - Undoing changes
    - Working with a remote repo

Week 8: Advanced Pandas Continued

1) **Session 21: MultiIndex Series and DataFrames**
    - About Multiindex objects
    - Why to use Multiindex objects
    - Stacking and unstacking
    - Multiindex DataFrames
    - Transpose Dataframes
    - Swaplevel
    - Long vs wide data
    - Pandas-melt
2) **Session 22: Vectorized String Operations | Datetime in Pandas**
    - Pivot table
    - Agg functions
    - Vectorized String operations
    - Common functions
    - Pandas Datetime
3) **Session on Pandas Case Study – time Series analysis**

4) **Session on Pandas Case Study – Working with textual data**

Week 9: Data Visualization

1) **Session 23: Plotting Using Matplotlib**
    - Get started with Matplotlib
    - Plotting simple functions, labels, legends, multiple plots
    - About scatter plots
    - Bar chart
    - Histogram
    - Pie chart
    - Changing styles of plots
2) **Session 24: Advanced Matplotlib**
    - Colored Scatterplot
    - Plot size, annotations
    - Subplots
    - 3D plots
    - Contour plots
    - Heatmaps
    - Pandas plot()
3) **Session on Plotly (Express)**
    - About Plotly
    - Disadvantages
    - Introduction about Plotly Go, Plotly Express, Dash
    - Hands-on Plotly
4) **Session on Plotly Graph Objects (go)**
5) **Session on Plotly Dash**
    - Basic Introduction about Dash
6) **Making a COVID-19 dashboard using Plotly and Dash**
7) **Deploying a Dash app on Heroku**
8) **Session on Project using Plotly**
    - Project using Indian Census Data with Geospatial indexing Dataset

Week 10: Data Visualization Continued

1) **Session 25: Plotting Using Seaborn- part 1**
    - Why seaborn?
    - Seaborn roadmap
    - Main Classification plotting
    - Relational plots
    - Distribution plots
    - KDE plot
    - Matrix plot
2) **Session 26: Plotting Using Seaborn- Part 2**
    - Categorical Plots
    - Stripplot
    - Figure level function — catplot
    - Swarmplot
    - Categorical Distribution Plots
    - Boxplot
    - Violinplot
    - Categorical Estimate Plot — for central tendency
    - Barplot
    - Pointplot
    - Countplot
    - Faceting
    - Doubt - error bar issue
    - Regression Plots
    - Regplot
    - Lmplot
    - Residual Plot
    - FacetGrid
    - Pairplot and Pairgrid
    - JointGrid and Jointplot
    - Utility Function in Seabom - load_dataset
    - Blog idea
3) **Session on Open-Source Software – Part 1**
4) **Session on Open-Source Software – Part 1**

Week 11: Data Analysis Process - Part1

1) **Session 27: Data Gathering | Data Analysis**
    - Data Analysis Process
    - Import Data from various sources (CSV, excel, JSON, text, SQL)
    - Export data in different file formats
    - Gather Data through API or Web Scraping
2) **Session 28: Data Assessing and Cleaning**
    - Data assessing
    - Types of unclean data
    - Write summary of data
    - Types of Assessment
    - Manual and Automatic Assessment
    - Data Quality Dimension
    - Data Cleaning
3) **Session on ETL using AWS RDS**
    - Introduction about Extraction, transform and Load pipeline
    - Fetch data from AWS
    - Apply transformation on the data
    - Upload transformed data into AWS RDS
4) **Session on Advanced Web Scraping using Selenium**
    - Introduction to Selenium and Chromedriver
    - Automated Web scraping on Smartprix website

Week 12: Data Analysis Process – Part 2

1) **Session on Data Cleaning Case Study – Smartphone dataset**
    - Quality issues
    - Tidiness issues
    - Data Cleaning
2) **Session 29: Exploratory Data Analysis (EDA)**
    - Introduction to EDA
    - Why EDA?

- Steps for EDA
- Univariate Analysis
- Bivariate Analysis
- Feature Engineering

3) **Session on Data Cleaning – Part 2**
    - Data Cleaning on Smartphone Dataset – Continued
4) **Session on EDA Case Study – Smartphone Dataset**

Week 13: SQL Basics

1) **Session 30: Database Fundamentals**
    - Introduction to Data and Database
    - CRUD operations
    - Properties of database
    - Types of Database
    - DBMS
    - Keys
    - Cardinality of Relationship
    - Drawbacks of Database
2) **Session 31: SQL DDL Commands**
    - Xammp Software
    - Types of SQL commands
    - DDL commands
3) **Session on Tableau – Olympics Dataset (Part 1)**
    - Download and Install Tableau
    - Getting Started with Tableau Desktop
    - Overview of Olympic datasets
    - Create Dashboards using Tableau

Week 14: SQL Continued – Part 1

1) Session 32: SQL DML commands
    - MySQL workbench
    - INSERT
    - SELECT
    - UPDATE
    - DELETE
    - Functions in SQL

2) **Session 33: SQL Grouping and Sorting**
    - Sorting Data
    - ORDER BY
    - GROUP BY
    - GROPU BY on multiple columns
    - HAVING clause
    - Practice on IPL Dataset
3) **Session on Tableau – Part 2**
    - Tableau Basics
        i. *Importing Data*
        ii. *Measures and Dimensions*
        iii. *Sheet, dashboard, story*
        iv. *Levels of Granularity*
        v. *Different types of charts*
        vi. *Datetime*
    - Hierarchical level of granularity
    - Common filters
    - Calculated fields and Table Calculations
    - Working with Geographical columns
    - Dashboard and Interactive filters
    - Blending and Dual axis chart
    - Connecting to a remote database

Week 15: SQL Continued - Part 2

1) **Session 34: SQL Joins**
    - Introduction to SQL joins
    - Types of Joins (Cross, inner, left, right, full outer)
    - SQL hands-on on joins
    - SET operations
    - SELF join
    - Query execution order
    - Practice questions
2) **Session on SQL Case Study 1 – Zomato Dataset**
    - Understanding Dataset through diagram
    - Solving SQL Questions on Zomato Dataset
3) **Session 35: Subqueries in SQL**

- What is Subquery
- Types of Subqueries
- Independent and Correlated subquery

4) **Session on Making a Flight Dashboard using Python and SQL**
    - How to Connect MYSQL through Python
    - Run SQL queries with Python
    - Creating a dynamic dashboard with Streamlit on Flights dataset
5) **Session on SQL Interview Questions – Part 1**
    - Database Server Vs Database Client
    - Database Engines
    - Components of DBMS
    - What is Collation?
    - COUNT(*) vs COUNT(col)
    - Dealing with NULL values
    - DELETE Vs TRUNCATE
    - Anti joins
    - Non-equi joins
    - Natural joins
    - All and Any operators
    - Removing Duplicate Rows
    - Metadata Queries

Week 16: Advanced SQL

1) **Session 36: Window Functions in SQL**
    - What are Window functions?
    - OVER(), RANK(), DENSE_RANK(), ROW_NUMBER(), FIRST_VALUE(), LAST_VALUE()
    - Concept of Frames
    - LAG(), LEAD()
2) **Session 37: Windows Functions Part 2**
    - Ranking
    - Cumulative sum and average
    - Running average
    - Percent of total
3) **Session 37: Windows Functions Part 3**
    - Percent Change

- Quantiles/Percentiles
- Segmentation
- Cumulative Distribution
- Partition by multiple columns

4) **Session on Data Cleaning using SQL | Laptop Dataset**
    - Basic level Data Cleaning and Data exploration using SQL
    - Why to use SQL for Data Cleaning
    - String Data types
    - Wildcards
    - String Functions
    - Data Cleaning
5) **Session on EDA using SQL | Laptop Dataset**
    - EDA on numerical and categorical columns
    - Plotting
    - Categorical – Categorical Analysis
    - Numerical – Numerical Analysis

Week 17: Descriptive Statistics

1) **Session 38: Descriptive Statistics Part 1**
    - What is Statistics?
    - Types of Statistics
    - Population vs Sample
    - Types of Data
    - Measures of central tendency
    - Measure of Dispersion
    - Coefficient of variation
    - Graphs for Univariate Analysis
    - Frequency Distribution table
    - Graphs for bivariate Analysis
    - Categorical – Categorical Analysis
    - Numerical – Numerical Analysis
    - Categorical – Numerical Analysis
2) **Session on Datetime in SQL**
    - Remaining topics of EDA using SQL (numerical - categorical,missing values, ppi, price_bracket, one hot encoding)
    - Temporal Data types

- Creating and Populating Temporal Tables
- DATETIME Functions
- Datetime Formatting
- Type Conversation
- DATETIME Arithmetic
- TIMESTAMP VS DATETIME
- Case Study – Flights

Week 18: Descriptive Statistics continued

1) **Session 39: Descriptive Statistics part 2**
    - Quantiles and Percentiles
    - Five Number Summary
    - Boxplots
    - Scatterplots
    - Covariance
    - Correlation
    - Correlation vs Causation
    - Visualizing multiple variables
2) **Session 40: Probability Distribution Functions (PDF, CDF, PMF)**
    - Random Variables
    - Probability Distributions
    - Probability Distribution Functions and its types
    - Probability Mass Function (PMF)
    - Cumulative Distribution Function (CDF) of PMF
    - Probability Density Function (PDF)
    - Density Estimation
    - Parametric and Non-parametric Density Estimation
    - Kernel Density Estimate (KDE)
    - Cumulative Distribution Function (CDF) of PDF.
3) **Session on SQL Datetime Case Study on Flights Dataset**
4) **Session on Database Design | SQL Data Types | Database Normalization**
    - Different SQL Datatypes (Numeric, Text, Datetime, Misc)
    - Database Normalization
    - ER Diagram

Week 19: Probability Distributions

1) **Session 41: Normal Distribution**
    - How to use PDF in Data Science?
    - 2D density plots
    - Normal Distribution (importance, equation, parameter, intuition)
    - Standard Normal Variate (importance, z-table, empirical rule)
    - Properties of Normal Distribution
    - Skewness
    - CDF of Normal Distribution
    - Use of Normal Distribution in Data Science
2) **Session 42: Non-Gaussian Probability Distributions**
    - Kurtosis
    - Excess Kurtosis and Types of kurtosis
    - QQ plot
    - Uniform Distribution
    - Log-normal distribution
    - Pareto Distribution
    - Transformations
        i. *Mathematical Transformation*
        ii. *Function Transformer*
        iii. *Log Transform*
        iv. *Reciprocal Transform / Square or sqrt Transform*
        v. *Power Transformer*
        vi. *Box-Cox Transform*
        vii. *Yeo-Johnson Transformation*
3) **Session on views and User Defined Functions in SQL**
    - What are views?
    - Types of views
    - User Defined Functions (Syntax, Examples, Benefits)
4) **Session on Transactions and Stored Procedures**
    - Stored Procedures
    - Benefits of using stored procedures
    - Transactions (Commit, rollback, savepoint)
    - ACID properties of a Transaction

Week 20: Inferential Statistics

1) **Session 43: Central Limit Theorem**
    - Bernouli Distribution
    - Binomial Distribution
        i. *PDF formula*
        ii. *Graph of PDF*
        iii. *Examples*
        iv. *Criteria*
        v. *Application in Data Science*
    - Sampling Distribution
    - Intuition of Central Limit Theorem (CLT)
    - CLT in code
    - Case study
    - Assumptions of making samples
2) **Session on Central Limit Theorem Proof**
3) **Session 44: Confidence Intervals**
    - Population vs Sample
    - Parameter vs Estimate
    - Point Estimate
    - Confidence Interval
        i. *Ways to calculate CI*
        ii. *Applications of CI*
        iii. *Assumptions of z-procedure*
        iv. *Formula and Intuition of z-procedure*
        v. *Interpreting CI*
        vi. *T-procedure and t-distribution*
        vii. *Confidence Intervals in code*

Week 21: Hypothesis Testing

1) **Session 45: Hypothesis Testing (Part 1)**
    - Key idea of hypothesis testing
    - Null and alternate hypothesis
    - Steps in Hypothesis testing
    - Performing z-test
    - Rejection region and Significance level

- Type-1 error and Type-2 Error
- One tailed vs. two tailed test
- Applications of Hypothesis Testing
- Hypothesis Testing in Machine Learning

2) **Session 46: Hypothesis Testing (Part 2) | p-value and t-tests**
   - What is p-value?
   - Interpreting p-value
   - P-value in the context of z-test
   - T-test
   - Types of t-test
     i. *Single sample t-Test*
     ii. *Independent 2-sample t-Test*
     iii. *Paired 2 sample t-Test*
     iv. *Code examples of all of the above*
3) **Session on Chi-square test**
   - Chi – square distribution (Definition and Properties)
   - Chi-square test
   - Goodness of fit test (Steps, Assumptions, Examples)
   - Test for Independence (Steps, Assumptions, Examples)
   - Applications in machine learning
4) **Session on ANOVA**
   - Introduction
   - F-distribution
   - One-way ANOVA
     i. *Steps*
     ii. *Geometric Intuition*
     iii. *Assumptions*
     iv. *Python Example*
   - Post – Hoc test
   - Why t-test is not used for more than 3 categories?
   - Applications in Machine Learning

Week 22: Linear Algebra

1) **Session on Tensors | Linear Algebra part 1(a)**
   - What are tensors?
   - 0D, 1D and 2D Tensors

- Nd tensors
- Rank, axes and shape
- Example of 1D, 2D, 3D, 4D, 5D tensors

2) **Session on Vectors | Linear Algebra part 1(b)**
    - What is Linear Algebra?
    - What are Vectors?
    - Vector example in ML
    - Row and Column vector
    - Distance from Origin
    - Euclidean Distance
    - Scalar Addition/Subtraction (Shifting)
    - Scalar Multiplication/Division [Scaling]
    - Vector Addition/Subtraction
    - Dot product
    - Angle between 2 vectors
    - Equation of a Hyperplane
3) **Linear Algebra Part 2 | Matrices (computation)**
    - What are matrices?
    - Types of Matrices
    - Matrix Equality
    - Scalar Operation
    - Matrix Addition, Subtraction, multiplication
    - Transpose of a Matrix
    - Determinant
    - Minor
    - Cofactor
    - Adjoint
    - Inverse of Matrix
    - Solving a system of Linear Equations
4) **Linear Algebra Part 3 | Matrices (Intuition)**
    - Basis vector
    - Linear Transformations
    - Linear Transformation in 3D
    - Matrix Multiplication as Composition
    - Test of Commutative Law
    - Determinant and Inverse

- Transformation for non-square matrix?
- Why only square matrix has inverse?
- Why inverse is possible for non-singular matrices only?

Week 23: Linear Regression

1) **Session 48: Introduction to Machine Learning**
    - About Machine Learning (History and Definition)
    - Types of ML
        i. *Supervised Machine Learning*
        ii. *Unsupervised Machine Learning*
        iii. *Semi supervised Machine Learning*
        iv. *Reinforcement Learning*
    - Batch/Offline Machine Learning
    - Disadvantages of Batch learning
    - Online Machine Learning
        i. Importance
        ii. When to use and how to use
        iii. Learning Rate
        iv. Out of core learning
        v. Disadvantages
    - Batch vs Online learning
    - Instance based learning
    - model-based learning
    - Instance vs model-based learning
    - Challenges in ML
        i. Data collection
        ii. Insufficient/Labelled data
        iii. Non-representative data
        iv. Poor quality data
        v. Irrelevant features
        vi. Overfitting and Underfitting
        vii. Offline learning
        viii. Cost
    - Machine Learning Development Life-cycle
    - Different Job roles in Data Science
    - Framing a ML problem | How to plan a Data Science project

2) **Session 49: Simple Linear regression**
    - Introduction and Types of Linear Regression
    - Simple Linear Regression
    - Intuition of simple linear regression
    - Code example
    - How to find m and b?
    - Simple Linear Regression model code from scratch
    - Regression Metrics
        i. *MAE*
        ii. *MSE*
        iii. *RMSE*
        iv. *R2 score*
        v. *Adjusted R2 score*
3) **Session 50: Multiple Linear Regression**
    - Introduction to Multiple Linear Regression (MLR)
    - Code of MLR
    - Mathematical Formulation of MLR
    - Error function of MLR
    - Minimizing error
    - Error function continued
    - Code from scratch
4) **Session on Optimization the Big Picture**
    - Mathematical Functions
    - Multivariable Functions
    - Parameters in a Function
    - ML models as Mathematical Function
    - Parametric Vs Non-Parametric ML models
    - Linear Regression as a Parametric ML model
    - Loss Function
    - How to select a good Loss Function?
    - Calculating Parameters from a Loss Function
    - Convex And Non-Convex Loss Functions
    - Gradient Descent
    - Gradient Descent with multiple Parameters
    - Problems faced in Optimization
    - Other optimization techniques

5) **Session on Differential Calculus**
    - What is differentiation?
    - Derivative of a constant
    - Cheatsheet
    - Power Rule
    - Sum Rule
    - Product Rule
    - Quotient Rule
    - Chain Rule
    - Partial Differentiation
    - Higher Order Derivatives
    - Matrix Differentiation

Week 24: Gradient Descent

1) **Session 51: Gradient descent from scratch**
    - What is Gradient Descent?
    - Intuition
    - Mathematical Formulation
    - Code from scratch
    - Visualization 1
    - Effect of Learning Rate
    - Adding m into the equation
    - Effect of Loss function
    - Effect of Data
2) **Session 52 (part 1): Batch Gradient Descent**
    - Types of Gradient Descent
    - Mathematical formulation
    - Code from scratch
3) **Session 52 (part 2): Stochastic Gradient Descent**
    - Problems with Batch GD
    - Stochastic GD
    - Code from scratch
    - Time comparison
    - Visualization

- When to use stochastic GD
- Learning schedules
- Sklearn documentation

4) **Session 52 (part 3): Mini-batch Gradient Descent**
    - **Introduction**
    - **Code**
    - **Visualization**
5) **Doubt Clearance session on Linear Regression**

Week 25: Regression Analysis

1) **Session on Regression Analysis (Part 1)**
    - What is Regression Analysis?
    - Why Regression Analysis is required?
    - What's the Statistic connection with Regression analysis?
    - Inference vs Prediction
    - Statsmodel Linear Regression
    - TSS, RSS, and ESS
    - Degree of freedom
    - F-statistic and Prob(F-statistic)
2) **Session on Regression Analysis (Part 2)**
    - F -test for overall significance
    - R-squared (Goodness of fit)
    - Adjusted R-squared
    - T – Statistic
    - Confidence Intervals for Coefficients
3) **Session on Polynomial Regression**
    - Why we need Polynomial Regression?
    - Formulation of Polynomial Regression
    - Polynomial Regression in python
4) **Session on Assumptions of Linear Regression**
    - Assumptions of Linear Regression
        i. *Linearity*
        ii. *Normality of Residuals*
        iii. *Homoscedasticity*
        iv. *No Autocorrelation*
        v. *No or little Multicollinearity*

- What happen when these assumptions failed?
- How to check each of these assumptions?
- What to do when an assumption fails?
- Standard Error

5) **Session 53: Multicollinearity**
    - What is multicollinearity?
    - When is Multicollinearity bad?
    - Multicollinearity (Mathematically)
    - Perfect and non-perfect Multicollinearity
    - Types of multicollinearity
    - How to detect Multicollinearity
    - Correlation
    - VIF (Variance Inflation Factor)
    - Condition Number
    - How to remove Multicollinearity

Week 26: Feature Selection

1) **Session 54: feature Selection Part 1**
    - What is Feature Selection?
    - Why to do Feature Selection?
    - Types of Feature Selection
    - Filter based Feature Selection
        i. *Duplicate Features*
        ii. *Variance Threshold*
        iii. *Correlation*
        iv. *ANOVA*
        v. *Chi-Square*
        vi. *Advantages and Disadvantages*
2) **Session 55: Feature Selection Part 2**
    - Wrapper method
    - Types of wrapper method
    - Exhaustive Feature Selection/Best Subset Selection
    - Sequential Backward Selection/Elimination
    - Sequential Forward Selection
    - Advantages and Disadvantages
3) **Session on Feature Selection part 3**

- Embedded Methods
    i. *Linear Regression*
    ii. *Tree based models*
    iii. *Regularized Models*
- Recursive Feature Elimination
- Advantages and Disadvantages

Week 27: Regularization

1) **Session on Regularization Part 1 | Bias-Variance Tradeoff**
    - Why we need to study Bias and Variance
    - Expected Value and Variance
    - Bias and Variance Mathematically
2) **Session on Regularization Part 1 | What is Regularization**
    - Bias Variance Decomposition
    - Diagram
    - Analogy
    - Code Example
    - What is Regularization?
    - When to use Regularization?
3) **Ridge Regression Part 1**
    - Types of Regularization
    - Geometric Intuition
    - Sklearn Implementation
4) **Ridge Regression Part 2**
    - Ridge Regression for 2D data
    - Ridge Regression for nD data
    - Code from scratch
5) **Ridge Regression Part 3**
    - **Ridge regression using Gradient Descent**
6) **Ridge Regression Part 4**
    - 5 Key Understandings
        i. *How do the coefficients get affected?*
        ii. *Higher values are impacted more*
        iii. *Bias variance tradeoff*
        iv. *Contour plot*
        v. *Why is it called Ridge?*

7) **Lasso Regression**
    - Intuition
    - Code example
    - Lasso regression key points
8) **Session on Why Lasso Regression creates Sparsity?**
9) **ElasticNet Regression**
    - Intuition
    - Code example
10) **Doubt Clearance session on regularization**

Week 28: K Nearest Neighbors

1) **Session on K nearest Neighbors Part 1**
    - KNN intuition
    - Code Example
    - How to select K?
    - Decision Surface
    - Overfitting and Underfitting in KNN
    - Limitations of KNN
2) **Session on coding K nearest Neighbors from scratch**
3) **Session on How to draw Decision Boundary for Classification problems**
4) **Session on Advanced KNN Part 2**
    - KNN Regressor
    - Hyperparameters
    - Weighted KNN
    - Types of Distances (Euclidean and Manhattan)
    - Space and Time Complexity
    - KD-Tree
5) **Classification Metrics Part 1**
    - Accuracy
    - Accuracy of multi-classification problems
    - How much accuracy is good?
    - Problem with accuracy
    - Confusion matrix
    - Type 1 error
    - Type 2 error
    - Confusion matrix of multi-classification problems

- When accuracy is misleading

6) **Classification Metrics Part 2**
    - Precision
    - Recall
    - F1 score
    - Multi class Precision and Recall
    - Multi class F1 score

Week 29: PCA

1) **Session on Curse of Dimensionality**
2) **PCA part 1**
    - Introduction
    - Geometric Intuition of PCA
    - Why is Variance important?
3) **PCA part 2**
    - Mathematical Problem formulation
    - What is covariance and covariance matrix?
    - Matrices as Linear Transformation
    - EigenVectors and Eigenvalues
    - Step by step solution of PCA
    - How to transform points?
    - PCA step-by-step code in python
4) **PCA Part 3**
    - Practical example on MNIST dataset
    - PCA demo with sklearn
    - Visualization
    - What is explained variance
    - Find optimum number of Principal components
    - When PCA does not work?
5) **Session on Eigen vectors and Eigen Values**
    - What are Matrices?
    - What are Eigen Vectors and Eigen Values?
    - Intuition – Axis of rotation
    - How to calculate Eigen Vectors and Eigen Values
    - Properties
    - Eigen Vectors in PCA

6) **Session on Eigen Decomposition and PCA variants**
    - Types of PCA variants
    - What are some special matrices?
        i. *Diagonal Matrix*
        ii. *Orthogonal Matrix*
        iii. *Symmetric Matrix*
    - Matrix as Linear Transformation Visualization tool
    - Matrix Composition
    - Matrix Decomposition
    - Eigen decomposition
    - Eigen decomposition of Symmetric Matrix (Spectral Decomposition)
    - Advantages of Eigen decomposition
    - Kernel PCA
    - Code example of Kernel PCA

7) **Session on eigen Singular Value Decomposition (SVD)**
    - Intuition of Non-Square Matrix
    - Rectangular Diagonal Matrix
    - What is SVD
    - Applications of SVD
    - SVD - The intuition of the mathematical equation
    - Relationship with Eigen Decomposition
    - Geometric Intuition of SVD
    - How to calculate SVD
    - SVD in PCA

Week 30: Model Evaluation & Selection

1. **ROC Curve in Machine Learning**
    a. ROC AUC Curve and it's requirements
    b. Confusion matrix
    c. True Positive Rate (TPR)
    d. False Positive Rate (FPR)
    e. Different cases of TPR & FPR
2. **Session on Cross Validation**
    a. Why do we need Cross Validation?

b. Hold-out approach
c. Problem with Hold-out approach
d. Why is the Hold-out approach used?
e. Cross Validation
    i. Leave One Out Cross Validation (LOOCV)
        1. Advantages
        2. Disadvantages
        3. When to use
    ii. K-Fold Cross Validation
        1. Advantages
        2. Disadvantages
        3. When to use
    iii. Stratified K-Fold CV

3. **Session on Data Leakage**
    a. What is it and what is the problem
    b. Ways in which Data Leakage can occur
    c. How to detect
    d. How to remove Data Leakage
    e. Validation set
4. **Session on Hyperparameter Tuning**
    a. Parameter vs Hyperparameter
    b. Why the word “hyper” in the term
    c. Requirements
        i. Grid Search CV
        ii. Randomized Search CV
    d. Can this be improved?

Week 31: Naive Bayse

1. **Crash course on Probability Part 1**
    a. 5 important terms in Probability
        i. Random Experiment
        ii. Trials
        iii. Outcome
        iv. Sample Space
        v. Event
    b. Some examples of these terms

c. Types of events
d. What is probability
e. Empirical vs Theoretical probability
f. Random variable
g. Probability distribution of random variable
h. Mean of 2 random variable
i. Variance of Random variable

2. **Crash course on Probability Part 2**
    a. Venn diagrams
    b. Contingency table
    c. Joint probability
    d. Marginal probability
    e. Conditional probability
    f. Intuition of Conditional Probability
    g. Independent vs Dependent vs Mutually Exclusive Events
    h. Bayes Theorem
3. **Session 1 on Naive Bayes**
    a. Intuition
    b. Mathematical formulation
    c. How Naive Bayes handles numerical data
    d. What if data is not Gaussian
    e. Naive Bayse on Textual data
4. **Session 2 on Naive Bayes**
    a. What is underflow in computing
    b. Log Probabilities
    c. Laplace Additive Smoothing
    d. Bias Variance Trade off
    e. Types
        i. Gaussian Naive Bayes
        ii. Categorical Naive Bayes
        iii. Multinomial Naive Bayes
5. **Session 3 on Naive Bayes**
    a. Probability Distribution related to Naive Bayes
        i. Bernoulli Distribution
        ii. Binomial Distribution
        iii. Categorical distribution / Multinoulli distribution

iv. Multinomial Distribution
v. Why do we need these distributions?

b. Categorical Naive Bayes
c. Bernoulli Naive Bayes
d. Multinomial Naive Bayes
e. Out of Core Naive Bayes

6. **End to End Project | Email Spam Classifier**

Week 32 : Logistics Regression

1. **Session 1 on Logistic Regression**
    a. Introduction
    b. Some Basic Geometry
    c. Classification Problem
        i. Basic Algorithm
        ii. Updation in Basic Algorithm
    d. Sigmoid Function
    e. Maximum Likelihood
    f. Log Loss
    g. Gradient Descent
    h. summary
2. **Session on Multiclass Classification using Logistic Regression**
    a. What is Multiclass Classification
    b. How Logistic Regression handles Multiclass Classification Problems.
    c. One vs Rest (OVR) Approach
        i. Intuition
        ii. Code
    d. SoftMax Logistic Regression Approach
        i. SoftMax Function
        ii. Code
    e. When to use what?
    f. Tasks
3. **Session on Maximum Likelihood Estimation**
    a. Recap
    b. Some Examples
        i. Example 1 - Coin Toss

ii. Example 2 - Drawing balls from bag
iii. Example 3 - Normal Distribution

c. Probability Vs Likelihood
d. Maximum Likelihood Estimation
e. MLE for Normal Distribution
f. MLE in Machine Learning
g. MLE in Logistic Regression
h. Some Important Questions

4. **Session 3 on Logistic Regression**
    a. Maximum Likelihood in Logistic Regression
    b. FAQ on MAE (Maximum Likelihood Estimation)
        i. Is MLE a general concept applicable to all ML algorithms?
        ii. How is MLE related to the concept of loss functions?
        iii. Why does the loss function exist, why don't we maximize likelihood?
        iv. Why study about maximum likelihood at all?
    c. An interesting task for you
    d. Assumptions of Logistics Regression
    e. Odds and Log(Odds)
    f. Another interpretation of Logistic Regression
    g. Polynomial Features
    h. Regularization in Logistic Regression
5. **Logistic Regression Hyperparameters**

Week 33 : Support Vector Machines (SVM)

1. **SVM Part 1 - Hard Margin SVM**
    a. Introduction
    b. Maximum Margin Classifier
    c. Support Vectors
    d. Mathematical Formulation
    e. How to solve this?
    f. Prediction
    g. Coding Example
    h. Problems with Hard Margin SVM

2. **SVM Part 2 | Soft Margin SVM**
    a. Problems with Hard Margin SVM
    b. Slack Variable
    c. Soft Margin SVM
    d. Introduction of C
    e. Bias-Variance Trade Off
    f. Code Example
    g. Relation with Logistic Regression
3. **Session on Constrained Optimization Problem**
    a. Problem with SVC
    b. Kernel's Intuition
    c. Coding examples of Kernel
    d. Types of Kernels
    e. Why is it called Trick?
    f. Mathematics if SVM
4. **Session on SVM Dual Problem**
    a. SVM in n Dimensions
    b. Constrained Optimization Problems
    c. Karush Kuhn Tucker Conditions
    d. Concept of Duality
    e. SVM Dual Problem
    f. Dual Problem Derivation
    g. Observations
5. **Session on Maths Behind SVM Kernels**
    a. SVM Dual Formulation
    b. The Similarity Perspective
    c. Kernel SVM
    d. Polynomial Kernel
        i. Trick
        ii. What about the other Polynomial terms
    e. RBF Kernel
        i. Local Decision Boundary
        ii. Effect of Gamma
    f. Relationship Between RBF and Polynomial K...
    g. Custom Kernels
        i. String kernel

ii. Chi-square kernel
iii. Intersection kernel
iv. Hellinger's kernel
v. Radial basis function network (RBFN) kernel
vi. Spectral kernel

Extra Sessions - Feature Engineering

1. **Session on Handling Missing Values Part - 1**
    a. Feature Engineering
        i. Feature Transformation
        ii. Feature Construction
        iii. Feature Extraction
        iv. Feature Selection
    b. Types of Missing Values
        i. Missing Completely at random
        ii. Missing at Random
        iii. Missing Not at Random
    c. Techniques for Handling Missing Values
        i. Removing Missing Values
        ii. Imputation
    d. Complete Case Analysis
2. **Session 2 on Handling Missing Data**
    a. Univariate Imputation - Numerical Data
        i. Mean Imputation
        ii. Median Imputation
    b. Univariate Imputation - Arbitrary Value & End Distribution Value
    c. Univariate Imputation - Categorical Data
        i. Mode Imputation
        ii. Missing Category Imputation
    d. Univariate Imputation - Random (Numerical + Categorical)
    e. Missing Indicator
3. **Session 3 on Handling Missing Data -** Multivariate Imputation
    a. KNN Imputer
        i. Steps in KNN Imputation
        ii. Advantages and Disadvantages in KNN Imputation
    b. Iterative Imputer

i. MICE
ii. When To Use
iii. Advantages and Disadvantages
iv. Demonstration of MICE algorithm
v. Steps involved in Iterative Imputer
vi. Important parameter in Iterative Imputer

Week 34 : Decision Trees

1. **Session 1 on Decision Tree**
    a. Introduction
    b. Intuition behind DT
    c. Terminology in Decision Tree
    d. The CART Algorithm - Classification
    e. Splitting Categorical Features
    f. Splitting Numerical Features
    g. Understanding Gini Impurity?
    h. Geometric Intuition of DT
2. **Session 2 on Decision Tree**
    a. CART for Regression
    b. Geometric Intuition of CART
    c. How Prediction is Done
    d. Advantages & Disadvantages of DT
    e. Project Discussion - Real Estate
3. **Session 3 on Decision Tree**
    a. Feature Importance
    b. The Problem of Overfitting
    c. Why Overfitting happens
    d. Unnecessary nodes
    e. Pruning & its types
        i. Pre-pruning
        ii. Post Pruning
    f. Cost Complexity Pruning

4. **Session on Decision Tree Visualization**
    a. dtreeviz Demo - Coding

Week 35 : Ensemble Methods - Introduction

1. **Introduction to Ensemble Learning**
    a. Intuition
    b. Types of Ensemble Learning
    c. Why it works?
    d. Benefits of Ensemble
    e. When to use Ensemble
2. **Bagging Part 1 - Introduction**
    a. Core Idea
    b. Why use Bagging?
    c. When to use Bagging?
    d. Code Demo
3. **Bagging Part 2 - Classifier**
    a. Intuition through Demo app
    b. Code Demo
4. **Bagging Part 3 - Regressor**
    a. Core Idea
    b. Intuition through demo web app
    c. Code Demo
5. **Random Forest : Session 1**
    a. Introduction to Random Forest
    b. Bagging
    c. Random Forest Intuition
    d. Why Random Forest Works?
    e. Bagging vs. Random Forest
    f. Feature Importance
6. **Random Forest : Session 2**
    a. Why Ensemble Techniques work?
    b. Random Forest Hyperparameters
    c. OOB Score
    d. Extremely Randomized Trees

e. Advantages and Disadvantages of Random Forest

Week 36 : Gradient Boosting

1. **Gradient Boosting : Session 1**
    a. Boosting
    b. What is Gradient Boosting
    c. How
    d. What
    e. Why
2. **Gradient Boosting : Session 2**
    a. How Gradient Boosting works?
    b. Intuition of Gradient Boosting
    c. Function Space vs. Parameter Space
    d. Direction of Loss Minimization
    e. How to update the function
    f. Iterate
    g. Another perspective of Gradient Boosting
    h. Difference between Gradient Boosting and Gradient Descent
3. **Gradient Boosting : Session 3 (Classification - 1)**
    a. Classification vs. Regression
    b. Prediction
4. **Gradient Boosting for Classification - 2 | Geometric Intuition**
    a. Geometric Intuition
5. **Gradient Boosting for Classification - 3 | Math Formulation**
    a. Step 0 : Loss Function
    b. Step 1 : Minimise Loss Function to get $F_0(x)$
    c. Step 2 :
        i. Pseudo Residuals
        ii. Training Regression Tree
        iii. Compute Lambda for all leaf nodes
        iv. Update the Model
    d. Step 3 : Final Model
    e. Log(odds) vs Probability

Capstone Project:

1. **Session 1 on Capstone Project | Data Gathering**

   a. Project overview in details
   b. Gather data for the project
   c. Details of the data
2. **Session 2 on Capstone Project | Data Cleaning**
   a. Merging House and Flats Data
   b. Basic Level Data Cleaning
3. **Session 3 on Capstone Project | Feature Engineering**
   a. Feature Engineering on Columns:
      i. additionalRoom
      ii. areaWithType
      iii. agePossession
      iv. furnishDetails
      v. features : luxury Score
4. **Session 4 on Capstone Project | EDA**
   a. Univariate Analysis
   b. PandasProfiling
   c. Multivariate Analysis
5. **Session 5 on Capstone Project | Outlier Detection and Removal**
   a. Outlier Detection And Removal
6. **Session 6 on Capstone Project | Missing Value Imputation**
   a. Outlier Detection and Removal on area and bedroom
   b. Missing Value Imputation
7. **Session 7 on Capstone Project | Feature Selection**
   a. **Feature Selection**
      i. Correlation Technique
      ii. Random Forest Feature Importance
      iii. Gradient Boosting Feature Importance
      iv. Permutation Importance
      v. LASSO
      vi. Recursive Feature Elimination
      vii. Linear Regression with Weights
      viii. SHAP (Explainable AI)
   b. **Linear Regression -** Base Model
      i. One-Hot Encoding
      ii. Transformation
      iii. Pipeline for Linear Regression

    c. SVR
8. **Session 8 on Capstone Project | Model Selection & Productionalization**
    a. **Price Prediction Pipeline**
        i. Encoding Selection
            1. Ordinal Encoding
            2. OHE
            3. OHE with PCA
            4. Target Encoding
        ii. Model Selection
    b. Price Prediction Web Interface -Streamlit
9. **Session 9 on Capstone Project | Building the Analytics Module**
    a. geo map
    b. word cloud amenities
    c. scatterplot -> area vs price
    d. pie chart bhk filter by sector
    e. side by side boxplot bedroom price
    f. distplot of price of flat and house
10. **Session 10 on Capstone Project | Building the Recommender System**
    a. Recommender System using TopFacilities
    b. Recommender System using Price Details
    c. Recommender System using LocationAdvantages
11. **Session 11 on Capstone Project | Building the Recommender System Part 2**
    a. Evaluating Recommendation Results
    b. Web Interface for Recommendation (Streamlit)
12. **Session 12 on Capstone Project | Building the Insights Module**
13. **Session 13 on Capstone Project | Deploying the application on AWS**

XGBoost (Extreme Gradient Boosting)

1. **Introduction to XGBoost**
    - Introduction
    - Features
        - Performance
        - Speed

- Flexibility

2. **XGBoost for Regression**
    a. Regression Problem Statement
    b. Step-by-Step Mathematical Calculation
3. **XGBoost for Classification**
    a. Classification Problem Statement
    b. Step-by-Step Mathematical Calculation
4. **The Complete Maths of XGBoost**
    a. Prerequisite & Disclaimer
    b. Boosting as an Additive Model
    c. XGBoost Loss Function
    d. Deriving Objective Function
    e. Problem With Objective Function and Solution
        i. The Taylor series
        ii. Applying Taylor Series
        iii. Simplification
    f. Output Value for Regression
    g. Output Value for Classification
    h. Derivation of Similarity Score
    i. Final Calculation of Similarity Score

MLOps Curriculum

Week 1: Introduction to MLOps and ML-DLC

Introduction to MLOps: Understanding what is MLOps and why is it an important field
Maintainable Code Development: Understanding version control and using tools like Git
Challenges in ML Model Deployment: Overview of ML life cycle stages, challenges in model deployment, approaches to deploying ML models

MLOps Best Practices: Industry standards and guidelines (high level overview of the next 8 weeks)

1. Session 1: Introduction to MLOps
    a. Reality of AI in the market
    b. Introduction
        i. Standard ML Cycle
        ii. What is DevOps?
        iii. What is MLOPs?
    c. Machine Learning Lifecycle
    d. Introduction to Version Control
        i. Key aspects of Version Control
        ii. Types of Version Control Systems
    e. Next two weeks plan
2. Session 2: Version Control
    a. Using GitHub for Version Control
        i. What is GitHub?
        ii. Setting Up GitHub
        iii. Creating a Repository
        iv. Cloning a Repository
        v. Making Changes
        vi. Committing Changes
        vii. Pushing Changes
        viii. Branching
        ix. Pull Requests
        x. Collaborating with Others
    b. Revisiting ML Cycle
        i. ML Pipeline Example
    c. Industry Trivia
3. Doubt Clearance Session 1
    a. Create an Account on GitHub
    b. Github using GUI: VsCode, Github Desktop
    c. Assignment on GitHub Fundamentals
    d. Git Push using CLI

e. Solving Error: Head is not in Sync
f. 55:10: Touch command
g.

Week 2: ML Reproducibility, Versioning, and Packaging

ML Reproducibility: Ensuring consistent results in ML experiments.
Model Versioning: Tools like MLflow.
Packaging and dependency management: Developing deployable ML packages with dependency management. Example - DS-cookie-cutter
Data Versioning and Management: Tools like DVC (Data Version Control). [Optional]

4. Session 3: Reproducibility
    a. Story
    b. Industry Tools
    c. Cookiecutter
        i. Step 1: Install the Cookiecutter Library and start a project
        ii. Step 2: Explore the Template Structure
        iii. Step 3: Customize the Cookiecutter Variables
        iv. Step 4: Benefits of Using Cookiecutter Templates in Data Science
5. Session 4: Data Versioning Control
    a. Introduction
    b. Prerequisites
    c. Setup
        i. Step 1: Initialize a Git repository
        ii. Step 2: Set up DVC in your project
        iii. Step 3: Add a dataset to your project
        iv. Step 4: Commit changes to Git
        v. Step 5: Create and version your machine learning pipeline
        vi. Step 6: Track changes and reproduce experiments
6. Doubt Clearance Session 2
    a. Assignment Solution on DVC: 10:19

b. Doubt Clearance
    i. DVC with G-Drive 42:50
    ii. DVC Setup Error: 48:45
    iii. Containerization with Virtual Environment 49:40
    iv. Create Version and ML Pipeline: 56:50
    v. DVC Checkout 57:50
    vi. How to which ID(commit) to go to - through commit messages? 1:00:00
    vii. What is Kubernetes?
    viii. Not able to understand by reading documentation 1:04:30
    ix. Getting no of commits 11k+ 1:09:40

Week 3: End-to-end ML lifecycle management

Setting up MLFlow: Understand MLFlow and its alternatives in depth.
Life cycle components: Projects, model registry, performance tracking.
Best Practices for ML Lifecycle

7. Session 5 - ML Pipelines and Experimentation Tracking
    a. Doubts
        i. DVC Track by Add
        ii. Git clone with SSH vs HTTPS
    b. Recap
    c. Pipelines + DVC + Experimentation Tracking
    d. MLFlow
8. Session 6 on MLOPs
    a. Recap of Pipelines - Credit Card Example
    b. Writing dvc.yaml File
    c. Reproducibility after Data Changes
    d. Reproducibility after Params Changes
    e. ML end-2-end Pipeline
    f. Tools for different Stages of Pipeline
9. Doubt Clearance Session 3
    a. Assignment Solution
    b. File not found error - Joblib/Data

c. Models Not found error
d. Get familiar With Terminal - DVC –help
e. DVC Repro vs DVC Exp

Week 4: Containerisation and Deployment Strategies

Containerization: Introduce docker internals and usage.
Distributed infrastructure: Introduce Kubernetes and basic internal components.
Online vs. Offline Model Deployment: Kubernetes for online, and batch processing for offline. A/B testing.
Canary Deployment and Blue/Green Strategies: Kubernetes, AWS CodeDeploy. [Optional]

10. Session 7 Continous Integration
    a. Philosophy Behind CI/CD
    b. Setting UP Github Actions for CI/CD
        i. Workflow setup
        ii. Integrating CML for Version Control
        iii. Update settings in GitHub Actions
            1. Setting Secret Tokens
11. Session 8 Containerisation - Docker
    a. Containerization
    b. Docker

Week 5: DAGs in MLOps

Understanding DAGs in MLOps: Dependency management in ML pipelines.
Building and Managing DAGs: Apache Airflow, Kubeflow Pipelines.
Continuous Integration/ Development: Discuss tools like Github Actions

12. Session 9 - Continuous Deployment
    a. Recap of Continous Integration
    b. Continuous Delivery/ Deployment
        i. Credit Card Project Code Example

ii. FastApi, pydantic, joblib, uvicorn
iii. How to write app.py

c. Multi Container

13. Doubt Clearance Session 4
    a. Assignments Solution

Week 6: Monitoring, Alerting, Retraining, and Rollback

Continuous Monitoring in MLOps: Prometheus, Grafana.

Alerting Systems

Automated Retraining Strategies: Kubeflow Pipelines

Rollback Design Patterns in MLOps: Feature flags, canary releases.

14. Session 10 - Introduction to AWS
    a. Introduction to AWS Machine Learning and MLOps Services
        i. Pre-requisite
        ii. IAM
    b. AWS Sagemaker
    c. Amazon S3
    d. AWS Lambda
    e. Amazon ECR and ECS
    f. AWS CodePipeline and AWS CodeBuild
    g. Components for SageMaker
15. Session 11 - Deployment on AWS
    a. Code Demo with Credit Card Project
    b. Project Code explained
        i. CI.yml file
            1. Configuring AWS Credentials on GitHub secrets
            2. Creating ECR Repository Getting URI
            3. AWS Region
        ii. Pulling ECR repo to EC2 machine
        iii. How to create Workflow YML files
        iv. Connecting to EC2
    c. Self Runner
    d. AWS Actions

Week 7: Scaling and Efficiency in MLOps

AutoML and Hyperparameter Tuning: AutoML tools (e.g., Google AutoML, Sagemaker Autopilot and Azure AML).
Data Consistency and availability: How offline development becomes a online nightmare
Resource Management and Cost Optimization: Discuss AWS Elastic Kubernetes Service (EKS) Autoscaling, Azure Kubernetes Service (AKS) Autoscale.

16. Session 12 - Distributed Infrastructure
    a. Understanding Distributed Computing
    b. Node, Communication, and Concurrency
    c. Why Distributed Computing?
    d. Docker and Microservices
        i. Fundamentals of Microservices
        ii. Data Bricks Architecture as Microservice
        iii. Uber Microservices Architecture
17. Session 13 on - Kubernetes Internals
    a. What is Kubernetes?
    b. Need of Kubernetes: Container orchestration system
    c. Key Concepts
        i. Pods
        ii. Nodes
        iii. Cluster
        iv. Control Panel
            1. API Server
            2. Resource Manager
            3. Database
        v. Deployment
            1. Deploying with Kubectl

18. MLOps Doubt Clearance Session 6
    a.

Week 8: Final Project

Final Project: Implementing a Full MLOps Pipeline for a Real-World Use Case **Locally**

19. Session 14 - Deployment on Kubernetes
    a. Deployment with Kubectl
        i. Explained Deployment yml file
        ii. Deployment Demo
    b. Deployment Strategies
    c. Service and Load Balancing

20. Session 15 - Seldon Deployments
    a. Introduction to Sheldon
    b. Key Features of Sheldon
    c. Sheldon vs Competitors
    d. Pre-Requisite: Kubectl and helm
    e. Deployment
        i. Step 1: Install Seldon Core using Helm
        ii. Step 2: Define a Simple Machine Learning Model
        iii. Step 3: Push your model to S3 or Google store
        iv. Step 4: Define a Seldon Deployment
        v. Step 5: Deploy the Seldon Deployment
    f. Seldon vs EKS vs K8s
    g. Kubeflow Pipelines
        i. KubeFlow Pipeline Overview with MNIST Dataset
    h. Apache Airflow
21. MLOps Doubt Clearance Session 7
    a. TBD

Week 9: ML Technical Debt

Understanding and Managing ML Technical Debt: Identifying and addressing technical debt in ML projects.

22. Session 16 - Monitoring & Alerting
    a. TBD
23. Session 17 - Rollout & Rollback Strategies
    a. TBD
24. Session on MLOps Interview Questions
25. Session 18 - ML Technical Debt
    a. TBD
26. MLOps Doubt Clearance Session 8

Feature Engineering

1. **Session on Encoding Categorical Features - 1**
    a. Feature Engineering Roadmap
    b. What is Feature Encoding
    c. Ordinal Encoding
        i. Code examples in Python
        ii. Handling Rare Categories
    d. Label Encoding
        i. Code Example using Sklearn LabelEncoder
    e. One Hot Encoding
        i. Code Examples using Sklearn OneHotEncoder
        ii. Handling unknown Category
    f. LabelBinarizer
2. **Session on Sklearn ColumnTransformer & Pipeline**
    a. What is ColumnTransformer
    b. Code implementation of ColumnTransformer
        i. OHE
        ii. Ordinal
    c. SKLearn Pipelines
        i. Implementing multiple transformations in Pipeline
            1. Missing value imputation
            2. Encoding Categorical Variables

        a. Handeling rare Categories
    3. Scaling
    4. Feature Selection
    5. Model building
    6. Prediction

3. **Session on Sklearn Deep Dive**
    a. Estimators
    b. Custom Estimators
    c. Mixins
    d. Transformers
    e. Custom Transformer
    f. Composite Transformars
    g. Column transformer
    h. Feature Union
    i. Pipeline

4. **Session 2 on Encoding Categorical Features**
    a. Count and Frequency Encodes
        i. CountEncoder Library
    b. Binary Encoder
    c. Target Encoder
5. **Session 1 on Discretization**
    a. Remainig topics of last session
        i. Weight of Evidence
        ii. Advice on when to use which Encoder
    b. What is Discretization?
    c. Why learn Discretization?
        i. Reducing Overfitting
        ii. Handling Non-Linear Relationships
        iii. Handling Outliers
        iv. Better Interpretability
        v. Model Compatibility
    d. Disadvantages of Discretization

6. **Session 2 on Discretization**
    a. Types of Discretization
        i. Uniforn Binning
        ii. Quantile Binning
        iii. K-Means Binning
        iv. Decision Tree Based tiening
        v. Custom Binning
        vi. Threshold Binning (Binarization)
7. **Session 1 on Handling Missing Data**
    a. Missing Values
    b. The missingo library
    c. Why missing velues occur?
    d. Types of missing values
    e. How missing visual impact ML?
    f. How to handle missing values?
        i. Removing
        ii. Imputing
    g. Removing Missing Data

8. **Session 2 on Handling Missing Data**
    a. Removing Missing Values
    b. Missing Indicator
        i. When to use?
        ii. When to not use?
    c. Simple Imputer -
        i. Mean & Median
        ii. Most Frequent
        iii. Constant
    d. How la select the best
9. **Session 3 on Handling Missing Data - Multivariate Imputer**
    a. NaN Euclidean distance
    b. KNN Imputer
        i. Code and Hyperparameter
        ii. When to use and not to use.
        iii. Advantages and Disadvantages

c. Iterative Imputer
    i. Code sed Parameters
    ii. When to use and not to use.
    iii. Advantages and Disadvantages
d. Coding Framework to compare different techniques

10. Session on Feature Scaling

a. What is Feature Scaling
    i. Why do we need feature scaling?
    ii. Which algorithms are affected if the features are not scaled?
    iii. Which algorithms are not affected?
b. Standardization

11. Session 2 on Feature Scaling

a. Minmax Scaling
b. Standardization vs Minmax Scaling
c. Robust Scaler
d. Max Absolute Scaler
e. L2/L1 Normalization
f. Comparison
g.

12. Session 1 on Outlier Detection

a. What are Outliers
b. Types of Outliers
c. Impact of Outliers
d. How to deal with outliers
e. Outlier Detection Techniques
    i. Univariate
        1. Z-Score
        2. IQR and BoxPlot
        3. Problem with Univariate Techniques
    ii. Multivariate Outlier Detection
        1. Isolation Forest

13. Session 2 on Outlier Detection

a. Isolation Forest
b. Calculation of Analoly Score

c. Outlier Detection using KNN
d. Local vs Global Outliers
e. Local Outlier Factor(LOF)

14. Session 3 on Outlier Detection

a. Local Outlier Factor(LOF)
b. DBSCAN - Code visualization
c. How to access the accuracy?
d. When to use which algo for outlier detection?

15. Session on Feature Transformation

a. Why do we need transformations?
b. What are Feature Transformations
    i. Problems after transformation
c. Log transformation
    i. Algorithms Benefitted
    ii. When to use
    iii. When not to use
d. Square Root transformation
e. Reciprocal Transformation
f. Case study on Boston Housing Price
g. Square Transformation
h. Box-Cox Transform
i. Yeo Johnson Transform

Unsupervised Learning

1. KMeans Clustering

a. Session 1 on KMeans Clustering

i. Plan of Attack (Getting Started with Clustering)
ii. Types of ML Learning

iii. Applications of Clustering
iv. Geometric Intuition of K-Means
v. Elbow Method for Deciding Number of Clusters
    1. Code Example
    2. Limitation of Elbow Method
vi. Assumptions of KMeans
vii. Limitations of K Means

b. Session 2 on KMeans Clustering

i. Recap of Last class
ii. Assignment Solution
iii. Silhouette Score
iv. Kmeans Hyperparameters
    1. Number of Clusters(k)
    2. Initialization Method (K Means++)
    3. Number of Initialization Runs (n_init)
    4. Maximum Number of Iterations (max_iter)
    5. Tolerance (tol)
    6. Algorithm (auto, full, ..)
    7. Random State
v. K Means ++

c. Session 3 on KMeans Clustering

i. K-Means Mathematical Formulation (Loyd's Algorithm)
ii. K-Means Time and Space Complexity
iii. Mini Batch K Means
iv. Types of Clustering
    1. Partitional Clustering
    2. Hierarchical Clustering
    3. Density Based Clustering
    4. Distribution/Model-based Clustering

d. K-Means Clustering Algorithms from Scratch in Python

i. Algorithms implementation from Scratch in Python

2. Other Clustering Algorithms

a. **Session on DBSCAN**
    i. Why DBSCAN?
    ii. What is Density Based Clustering
    iii. MinPts & Epsilon
    iv. Core Points, Border Points & Noise Points
    v. Density Connected Points
    vi. DBSCAN Algorithm
    vii. Code
    viii. Limitations
    ix. Visualization
b. **Session on Hierarchical Clustering**
    i. Need of Other Clustering Methods
    ii. Introduction
    iii. Algorithm
    iv. Types of Agglomerative Clustering
        1. Min (Single-link)
        2. Max (Complete Link)
        3. Average
        4. Ward
    v. How to find the ideal number of clusters
    vi. Hyperparameter
    vii. Code Example
    viii. Benefits/Limitations
c. **Session - 1 on Gaussian Mixture Models (GMM)**
    i. The Why?
    ii. The What?
    iii. Geometric Intuition
    iv. Multivariate Normal Distribution
    v. Geometric Intuition 2D
    vi. EM(Expectation Minimization) Algorithm
    vii. Python Code
d. **Session - 2 on Gaussian Mixture Models**
    i. Recap of Session 1
    ii. Covariance Types: Spherical, Diagonal, Full, and Tied

iii. How to decide n_components?
    1. Akaike Information Criterion (AIC)
    2. Bayesian Information Criterion(BIC)
    3. Likelihood Formula for GMM
    4. Python Implementation.
    5. Why not Silhouette Score?
iv. Visualization
v. Assumptions
vi. Advantages & Disadvantages
vii. K Means vs GMM
viii. DBSCAN vs GMM
ix. Applications of GMM

e. Session on T-SNE

i. What is T-SNE?
ii. Why learn T-SNE?
iii. Geometric Intuition
iv. Mathematical Formulation
v. Code Implementation

2. Session 2 on T-SNE

i. Mathematical Formulation
ii. Some Questions!
    1. Why use probabilities instead of distances to calculate similarity?
    2. Why use Gaussian distribution to calculate similarity in high dimensions?
    3. How is variance calculated for each Gaussian distribution?
    4. Why use T-distribution in lower dimensions?
iii. Code Example
iv. Hyperparameters
    1. Perplexity
    2. Learning Rate
    3. Number of Iterations
v. Points of Wisdom

vi. Advantages & Disadvantages

1. LDA (Linear Discriminant Analysis)
    a. **Introduction**: Supervised dimensionality reduction
    b. **Algorithm Explanation**: Maximizing between-class variance
    c. **Applications**: Use cases in classification and visualization
    d. **Comparison**: Differences and similarities with PCA

2. Apriori
    a. **Introduction**: Principles of association rule mining
    b. **Key Concepts**: Support, Confidence, Lift
    c. **Algorithm Steps**: Candidate generation, Pruning
    d. **Applications**: Market Basket Analysis, Recommender Systems

3. UMAP (Uniform Manifold Approximation and Projection)
    a. **Introduction**: Overview of the algorithm
    b. **Key Concepts**: Topological Structures, Nearest Neighbors
    c. **Applications**: Dimensionality reduction, Visualization
    d. **Comparison**: Differences and similarities with t-SNE and PCA

Competitive Data Science

1. Adaboost
    a. **Introduction**: Overview and intuition of the algorithm
    b. **Components**: Weak Learners, Weights, Final Model
    c. **Hyperparameters**: Learning Rate, Number of Estimators

d. **Applications**: Use Cases in Classification and Regression

2. Stacking
    a. **Introduction**: Concept of model ensembling
    b. **Steps**: Base Models, Meta-Model, Final Prediction
    c. **Variations**: Different approaches and modifications
    d. **Best Practices**: Tips for effective stacking

3. LightGBM
    a. Introduction: Gradient boosting framework
    b. Key Features: Handling large datasets, Categorical feature support
    c. Parameters: Core hyperparameters and their tuning
    d. Applications: Common use cases and performance considerations

    Session 1 on Introduction to LightGBM
    a. Introduction and core features
    b. Boosting and Objective Function
    c. Histogram-Based Split finding
    d. Best-fit Tree (Leaf-wise growth strategy)
    e. Gradient-based One side sampling(GOSS)
    f. Exclusive Feature Bundling (EFB)

4. CatBoost
    a. **Introduction**: Categorical boosting algorithm
    b. **Handling Categorical Features**: Algorithmic approach
    c. **Key Parameters**: Learning Rate, Depth, Iterations
    d. **Practical Usage**: Tips and common practices

5. Advanced Hyperparameter Tuning
    a. **Strategies**: Bayesian Optimization
    b. **Libraries**: Optuna, Hyperopt
    c. **Practical Tips**: Efficient tuning, Avoiding overfitting
    d. **Evaluation**: Ensuring robust model performance

6. Participating in a real Kaggle Competition
    a. **Getting Started**: Understanding the problem, Exploring datasets

b. **Strategy**: Model selection, Preprocessing, Validation
c. **Collaboration**: Teamwork, Sharing, and Learning
d. **Submission and Evaluation**: Making effective submissions, Learning from feedback

Miscellaneous Topics

1. NoSQL
    a. **Introduction**: Overview of NoSQL databases
    b. **Types**: Document, Key-Value, Column-Family, Graph
    c. **Use Cases**: When to use NoSQL over SQL databases
    d. **Popular Databases**: MongoDB, Cassandra, Redis, Neo4j

2. Model Explainability
    a. **Introduction**: Importance of interpretable models
    b. **Techniques**: LIME, SHAP, Feature Importance
    c. **Application**: Applying techniques to various models
    d. **Best Practices**: Ensuring reliable and accurate explanations

3. FastAPI
    a. **Introduction**: Modern, fast web framework for building APIs
    b. **Features**: Type checking, Automatic validation, Documentation
    c. **Building APIs**: Steps and best practices
    d. **Deployment**: Hosting and scaling FastAPI applications

4. AWS Sagemaker
    a. **Introduction**: Fully managed service for machine learning
    b. **Features**: Model building, Training, Deployment
    c. **Usage**: Workflow from data preprocessing to model deployment
    d. **Best Practices**: Optimizing costs and performance

5. Handling Imbalanced Data
    a. **Introduction**: Challenges of imbalanced datasets
    b. **Techniques**: Resampling, SMOTE, Using different evaluation metrics
    c. **Algorithms**: Choice of algorithms for imbalanced data
    d. **Best Practices**: Ensuring robust and unbiased models

6. Online Machine Learning
    a. **Introduction**: Machine learning on streaming data
    b. **Algorithms**: Suitable algorithms for online learning
    c. **Challenges**: Handling concept drift, Scalability
    d. **Application**: Real-time learning and prediction use cases

Note: The schedule is tentative and topics can be added/removed from it in the future.