GEN AI COURSE CONTENT

S.no	Core Course	Course Content
1	**Fundamentals of Data Analyst**	1. What is Data Science 2. Need for Data Scientists 3. Foundation 4. Business Intelligence 5. Data Analysis 6. Data Mining 7. Machine Learning 8. Analytics 9. Data Science 10. Types of Analytics 11. DS across industries 12. DS across functions 13. Analytics Project Lifecycle 14. Data 15. Basis of Data Categorization 16. Types of Data 17. Data Collection
2	**Excel**	1. Introduction to Excel 2. Columns & Rows 3. Functionality Using Ranges. 4. Excel Basic & Advanced Formulae's 5. Data Analysis Using Excel 6. Pivot & Power Pivot 7. Spreadsheet Tools 8. Data Cleaning 9. Data Validation 10. Data Visualization Using Excel
3	**DATA BASE**	**Introduction to RDBMS** **Sub Language Commands** • Data Definition Language (DDL) • Data Retrieval Language (DRL) • Data Manipulation Language (DML) • Transaction Control Language (TCL) • Database Security and Privileges (DCL) **Introduction to SQL Database Object** • Oracle Pre Defined Datatypes • DDL Commands

		• Create, Alter (add, modify, rename, drop)Columns, Rename, truncate, drop • DML-Insert, update, delete • DQL-SELECT Statements using WHEREclause • Comparison and Conditional Operators • Arithmetic and Logical Operators • Set Operators (UNION, UNION ALL, INTERSECT, MINUS) • Special Operators – IN (NOT IN), BETWEEN (NOT BETWEEN), LIKE (NOT LIKE), IS NULL (IS NOT NULL)
4	**Statistical**	**Descriptive & Inferential Statistics** 1. Turning Data into Information 2. Probability Distributions 3. Sampling Distributions 4. Confidence Intervals 5. Hypothesis Testing 6. Comparing Two Groups 7. Analysis of Variance (ANOVA) **Prediction Analytics** 1. Simple Linear Regression 2. Multiple Linear Regression 3. Model Adequacy Checking 4. Transformations 5. Diagnostics for Leverage and 6. influence 7. Polynomial Regression 8. Dummy Variables 9. Variables Selection and Model Building 10. Generalized Linear Models Autocorrelation **Applied Multivariate Analysis** 1. Measures of Central Tendency, Dispersion and Association 2. Multivariate Normal Distribution 3. Sample Mean Vector and Sample Correlation 4. Principal Components Analysis(PCA) 5. Factor Analysis 6. Discriminant Analysis **7. MANOVA**

5	R Programming	**R Programming**
		1. R Basics
		2. Numbers, Attributes
		3. Creating Vector
		4. Mixing Objects
		5. Explicit Coercion
		6. Formatting Data Values
		7. Matrices, List, Factors, Data Frames,
		8. Missing Values, Names
		9. Reading and Writing Data
		10. Using Dput/DDump
		11. Interface to the Outside world
		12. Sub setting R objects
		13. Vectorized Operations
		14. Dates and Times
		15. Managing Data Frames with the DPLYR
		16. package
		17. Control Structures
		18. Functions
		19. Lexical /Dynamic Scoping
		20. Loop Functions
		21. Debugging
		Data Visualization in R
		1. Storytelling with Data
		2. Principle tenets
		3. Elements of Data Visualization
		4. Infographics vs Data Visualization
		5. Data Visualization & Graphical functions in R
		6. Plotting Graphs
		7. Customizing Graphical Parameters to improvise the plots
		8. Various GUIs
		9. Spatial Analysis
		10. Other Visualization concepts
6	**Python**	1. Python Overview
		2. About Interpreted Languages
		3. Advantages/Disadvantages of Python pydoc
		4. Starting Python
		5. Interpreter PATH
		6. Using the Interpreter
		7. Running a Python Script
		8. Python Scripts on UNIX/Windows, Editors and IDEs
		9. Using Variables

10. Keywords
11. Built-in Functions
12. StringsDifferent Literals
13. Math Operators and Expressions
14. Writing to the Screen
15. String Formatting
16. Command Line Parameters and Flow Control

Sequences and File Operations

- Lists
- Tuples
- Indexing and Slicing
- Iterating through a Sequence
- Functions for all Sequences
- Using Enumerate()
- Operators and Keywords for Sequences
- The xrange() function
- List Comprehensions
- Generator Expressions
- Dictionaries and Sets

| 7 | Machine Learning Introduction | 1. ML Fundamentals
2. ML Common Use Cases
3. Understanding Supervised and Unsupervised Learning Techniques
4. Clustering
5. Similarity Metrics
6. Distance Measure Types: Euclidean, Cosine Measures
7. Creating predictive models
8. Understanding K-Means Clustering
9. Understanding TF-IDF, Cosine Similarity and their application to Vector Space Model
10. Case study
11. Implementing Association rule mining
12. Case study
13. Understanding Process flow of Supervised Learning Techniques
14. Decision Tree Classifier
15. How to build Decision trees
16. Case study
17. Random Forest Classifier
18. What is Random Forests
19. Features of Random Forest
20. Out of Box Error Estimate and Variable Importance
21. Case study
22. Naive Bayes Classifier
23. Case study
24. Project Discussion
25. Problem Statement and Analysis |

26. Various approaches to solving a Data Science Problem
27. Pros and Cons of different approaches and algorithms
28. Linear Regression
29. Case study
30. Logistic Regression
31. Case study
32. Text Mining
33. Case study
34. Sentimental Analysis
35. Case study

Power BI Or Tableau
- Power BI Introduction
- Data Visualization, Reporting
- Business Intelligence (BI), Traditional BI, Self-Serviced BI
- Cloud Based BI, On Premise BI
- Power BI Products
- Power BI Desktop (Power Query, Power Pivot, Power View)
- Flow of Work in Power BI Desktop
- Power BI Report Server, Power BI Service, Power BI Mobile Flow

Power Query
- Data Transformation, Benefits of Data Transformation
- Shape or Transform Data using Power Query
- Overview of Power Query / Query Editor, Query Editor User Interface The
- Ribbon (Home, Transform, Add Column, View Tabs)
- The Queries Pane, The Data View / Results Pane, The Query Settings Pane, FormulaBar
- Saving the Work
- Datatypes, Changing the Datatype of a Column Filter
- in Power Query
- Auto Filter / Basic Filtering
- Filter a Column using Text Filters
- Filter a Column using Number Filters
- Filter a Column using Date Filters
- Filter Multiple Columns
- Remove Columns / Remove Other Columns
- Name / Rename a Column

- Reorder Columns or Sort Columns
- Add Column / Custom Column Split
- Columns
- Merge Columns
- PIVOT, UNPIVOT Columns
- Transpose Columns
- Header Row or Use First Row as Headers
- Keep Top Rows, Keep Bottom Rows Keep
- Range of Rows
- Keep Duplicates, Keep Errors
- Remove Top Rows, Remove Bottom Rows, Remove Alternative Rows
- Remove Duplicates, Remove Blank Rows, Remove Errors

Data Modeling
- Data Modeling Introduction
- Relationship, Need of Relationship
- Relationship Types / Cardinality in General
- One-to-One, One-to-Many (or Many-to-One), Many-to-Many
- AutoDetect the relationship, Create a new relationship, Edit existing relationships
- Make Relationship Active or Inactive
- Delete a relationship

DAX
- What is DAX, Calculated Column, Measures
- DAX Table and Column Name Syntax
- Creating Calculated Columns, Creating Measures Calculated
- Columns Vs Measures
- DAX Syntax & Operators

DAX Operators
- Types of Operators
- Arithmetic Operators, Comparison Operators, Text Concatenation Operator, Logical
- Operators

DAX Functions Types
- Date and Time Functions
- Text Functions
- Logical Functions
- Math & Statistical Functions
- Filter Functions
- Time Intelligence Functions

- Date and Time Functions
- YEAR, MONTH, DAY
- WEEKDAY, WEEKNUM
- FORMAT (Text Function) à Month Name, Weekday Name
- DATE, TODAY, NOW
- HOUR, MINUTE, SECOND, TIME
- DATEDIFF, CALENDAR
- Creating Date Dimension Table
- Text Functions
- LEN, CONCATENATE (&)
- LEFT, RIGHT, MID UPPER, LOWER
- TRIM, SUBSTITUTE, BLANK

Logical Functions
- IF
- IFERROR SWITCH
- Math & Statistical Functions
- INT
- ROUND, ROUNDUP, ROUNDDOWN
- DIVIDE
- EVEN, ODD
- POWER, SIGN
- SQRT, FACT
- SUM, SUMX
- MIN, MINX
- MAX, MAXX
- COUNT, COUNTX
- AVERAGE, AVERAGEX
- COUNTROWS, COUNTBLANK

Filter Functions
- CALCULATE
- ALL
- RELATED
- Report View
- Report View User Interface
- Fields Pane, Visualizations pane, Ribbon, Views, Pages Tab, Canvas Visual

Interactions
- Interaction Type (Filter, Highlight, None)
- Visual Interactions Default Behavior, Changing the Interaction
- Grouping and Binning Introduction

- Using grouping, Creating Groups on Text Columns
- Using binning, Creating Bins on Number Column and Date Columns
- Sorting Data in Visuals
- Changing the Sort Column, Changing the Sort Order
- Sort using column that is not used in the Visualization
- Sort using the Sort by Column button
- Hierarchy Introduction, Default Date Hierarchy
- Creating Hierarchy, Creating Custom Date Hierarchy
- Change Hierarchy Levels
- Drill-Up and Drill-Down Reports
- Data Actions, Drill Down, Drill Up, Show Next Level

Visualizations

- Visualizing Data, Why Visualizations
- Visualization types, Create and Format Bar and Column Charts
- Create and Format Stacked Bar Chart Stacked Column Chart Create
- and Format Clustered Bar Chart, Clustered Column Chart
- Create and Format 100% Stacked Bar Chart, 100% Stacked Column Chart CreateandFormat Pie and Donut Charts
- Create and Format Scatter Charts
- Create and Format Table Visual, Matrix Visualization
- Line and Area Charts
- Create and Format Line Chart, Area Chart, Stacked Area Chart
- Combo Charts
- Create and Format Line and Stacked Column Chart, Line and Clustered ColumnChart
- Create and Format Ribbon Chart, Waterfall Chart, Funnel Chart
- Power BI Service
- Power BI Service Introduction, Power BI Cloud Architecture
- Creating Power BI Service Account, SIGN IN to Power BI Service Account

		• Publishing Reports to the Power BI service, Import / Getting the Report to PBI ServiceMy • Workspace / App Workspaces Tabs • DATASETS, WORKBOOKS, REPORTS, DASHBOARDS • Working with Datasets, Creating Reports in Cloud using Published Datasets
8	**GEN AI**	• Overview of Generative AI • Evolution from statistical ML to deep learning with neural networks • Transformers have led to a variety of powerful AI models. • Understand Vector database and semantic search • Word to numeric representation using word embeddings • Retrieval Augmented Generation (RAG) allows fine-tuning or building AI models on specific data sets. Using GPT-3 for querying data sources • To work with Gen AI, you need models, cloud services, and frameworks. • LLM based application using OpenAI API • Creating separate keys for different projects and managing them securely • Using LLM to generate creative restaurant names based on cuisine preferences. • Implementing Simple Sequential Chain concept in Gen AI Course • Using sequential chains to get both restaurant name and menu items in Gen AI Course • Using response data to display restaurant name and menu items in UI • Practice coding while watching the video • Gen AI has a reasoning component for answering questions beyond its knowledge • Initializing environment variables and using AI tools in an agent. • Using conversation chain to optimize cost and memory • Optimize OpenAI token cost by limiting conversational buffer memory • Importance of building a specialized tool over using Chat GPT • Using semantic search for finding relevant text chunks • Utilizing Vector databases for faster search and retrieval • Installing and loading Lang chain library in Jupyter notebook. • Text splitting and merging for efficient processing • Using Lang chain provides a simple API for text splitting. • Text splitting process using separators and size constraints

- Gen AI Course explains recursive text splitter and merging for chunk optimization
- Demonstration of using libraries to convert text into vectors
- Using meditation and yoga to improve mental health with vector representation
- Gen AI performs semantic search to provide results based on context, not just keywords
- Retrieval QA with Sources Chain
- Creating open API embeddings for document retrieval
- Using a Vector database and creating a retrieval chain for text analysis.
- Assembling the final project with pre-built individual pieces
- Loading environment variables and creating basic UI in Gen AI module
- Saving data in memory and displaying progress
- Setting up Vector database and creating retrieval QA chain in Python
- Automated news research tool for equity analysts
- Utilizing vector databases for storing embeddings in large projects
- Utilizing Google Palm and SQL Database Chain in Lang Chain Frame
- Introduction to Vector Database and Google Palm for AI applications
- Explaining the t-shirt records and database setup
- Setting up a SQL database object with URI
- Understanding the importance of correct database mapping for accurate results
- Learning to run and manage queries in DB Chain
- Summing of stock quantities for white color t-shirts
- Storing data into arrays and using Hugging Face embedding for generating embeddings
- Vector store converts input questions into embeddings for similarity matching.
- Import SQL prompt template to generate queries.
- Enhanced querying functionality with additional parameters
- Creating a main function in Python and running SQL queries in a database chain
- Using Streamlet to ask and answer questions efficiently Crafted by Merlin AI.

9	Projects (Elective 1 only)	• Facial Recognition • Social Media Analytics • Facial Detection • Data Acquisition and Productization • Object Detection • Handwriting Recognition • Sales Prediction • Stock Market Prediction • Data Security