

"Empowering Tomorrow's Solutions Today: Welcome to the Data Engineering Syllabus by TTS Nashik."

✚ *Here are some key points about what you can gain through the Data Engineering syllabus offered by TTS Nashik:*

Comprehensive Foundation: Acquire a solid understanding of the fundamental principles of data engineering, covering key concepts, methodologies, and best practices.

Hands-on Experience: Engage in practical, hands-on projects that allow you to apply theoretical knowledge to real-world scenarios, ensuring a well-rounded skill set.

Cutting-edge Technologies: Stay ahead in the industry by learning and implementing the latest tools and technologies in data engineering, including but not limited to data warehousing, ETL processes, and big data frameworks.

Problem-solving Skills: Develop the ability to analyze complex data challenges and implement effective solutions, enhancing your problem-solving skills crucial for success in the field.

Collaborative Learning Environment: Benefit from a collaborative and interactive learning environment, fostering teamwork and communication skills essential for thriving in data engineering projects.

Industry-Relevant Curriculum: Our syllabus is meticulously designed to align with industry trends and demands, ensuring that graduates are well-prepared for the rapidly evolving field of data engineering.

Practical Applications: Gain practical insights into deploying scalable and efficient data pipelines, ensuring the extraction, transformation, and loading (ETL) processes meet industry standards.

Data Governance and Security: Understand the importance of data governance and security, learning how to implement robust measures to safeguard sensitive information in compliance with industry regulations.

Career Readiness: Equip yourself with the skills and knowledge necessary for a successful career in data engineering, with a focus on professional development and readiness for real-world challenges.

Mentorship Opportunities: Benefit from mentorship programs, connecting with experienced professionals in the field who can provide guidance and insights, accelerating your learning and career growth.

Networking Opportunities: Expand your professional network through industry events, workshops, and networking sessions, opening doors to potential career opportunities and collaborations.

Continuous Learning: Foster a mindset of continuous learning, with access to resources and support to stay updated on emerging technologies and industry advancements throughout your career.

By enrolling in the Data Engineering syllabus at TTS Nashik, you can expect to gain a well-rounded skill set, practical experience, and the confidence to excel in the dynamic and rapidly evolving field of data engineering.

TTS Nashik

Tools to be covered!

MICROSOFT EXCEL



■ Program Curriculum

Basic Excel

This module provides a structured approach to learning Excel, gradually building from basic concepts to excel features.

- Introduction to Excel
- Workbook, worksheets, cells and ranges
- Basic Spreadsheet Operations
- Formatting Data
- Changing cell alignments
- Basic Formulas
- Using basic functions (SUM, AVERAGE, COUNT)
- Introduction to common functions (IF, ELSE)
- Sorting and Filtering
- Removing Duplicates
- Data Validation
- Setting up data validation rules
- Creating dropdown lists
- Error alert messages
- Flash Fill
- Password based files
- Set restrictions in excel
- Page layouts
- Page formatting
- Printing options

Advance Excel

Since we need to handle huge amounts of data to arrange and visualize, we will be implementing data handling techniques with Excel. And we will explore the different miscellaneous functions of Excel in detail.

- Charts and Graph
- Creating basic charts (bar, line, pie)
- Formatting and customizing charts
- Adding data labels and titles
- Conditional Formatting
- Highlighting cells based on conditions
- Using color scales and icon sets
- Managing conditional formatting rules
- Advanced Excel Features
- PivotTables and PivotCharts
- Creating PivotTables
- Analyzing data with PivotTables
- Creating PivotCharts from PivotTables
- Introduction to Macros
- Recording simple macros
- Running and editing macros
- Basic macro security considerations
- Applying learned skills to a practical project
- Integrating multiple Excel features
- Presenting findings using charts and tables

Python Programming Language



■ Program Curriculum

Python Basics

We will go through the basics of python with all essential beginner friendly concepts of Python programming like datatypes, loops, data structures and functions, followed by assessments and assignments

- Why Python
- Python IDE – Jupyter Notebook, PyCharm
- Anaconda Navigator
- Variables & Names
- Data Types in Python
- Integer basics
- String Basics
- List
- Tuple
- Dictionaries
- Conditional Statements
- For and While Loop
- Functions
- Numbers and Math Functions
- Common Errors in Python
- Functions as Arguments
- List Comprehension
- File Handling
- Debugging in Python
- Class and Objects
- Filters and Map
- Python PIP
- Read Excel Data in Python
- Iterators

Python Libraries

Since we need to handle huge amounts of data, we will be implementing data handling techniques with Pandas library. As well as we need visualization library like Matplotlib and Seaborn for plotting graphs and charts. And also be going with Numpy library.

- Introduction to Pandas
- Series Data Structure - Querying and Indexing
- DataFrame Data Structure - Querying, Indexing and loading
- Merging data frames
- Group by operation
- Date/Time functionality
- Example: Manipulating DataFrame
- Introduction to Matplotlib
- Basic Plotting
- Scatter Plots and Bar Charts
- Advanced Matplotlib Techniques
- Subplots, figures, histograms, box plots
- 3D plots, polar plots
- Introduction to Seaborn
- Categorical scatter plots, bar plots, violin plots
- Heatmaps, clustermaps, pair plots, facet grids
- Applying Matplotlib and Seaborn to a dataset
- Reviewing key concepts and best practices
- Introduction to NumPy
- Creating NumPy arrays
- Array indexing and slicing
- Array Operations
- Aggregation Functions
- Array Manipulation
- Reshaping arrays
- Concatenation and splitting
- Indexing and Slicing
- NumPy and Integration with Pandas
- Understanding the efficiency of NumPy operations

SQL Database



■ Program Curriculum

SQL

We will continue with SQL beyond basics like embed documents, dealing with types of data, schema types and Data relationships in SQL, etc.

- Overview of SQL in data engineering
- Installation of MySQL workbench, MySQL Shell, PHPMysqlAdmin
- Introduction to relational databases
- DDL and DML statements
- Managing the database on MySQL workbench, MySQL Shell, PHPMysqlAdmin etc.
- Basic SQL syntax and querying
- Using SQL queries
- Retrieving and Filtering Data
- SELECT statements for data retrieval
- Filtering data using WHERE
- Sorting results with ORDER BY
- Creating index on tables
- Optimizing queries for better performance
- Joins and Subqueries
- Inner and outer joins for data combining
- Using aliases for table and column names
- Subqueries for advanced filtering
- Aggregation and Analysis
- Aggregating data using GROUP BY
- Filtering aggregated data with HAVING
- Basic statistical functions
- Import and export

Probability and Statistics with NumPy



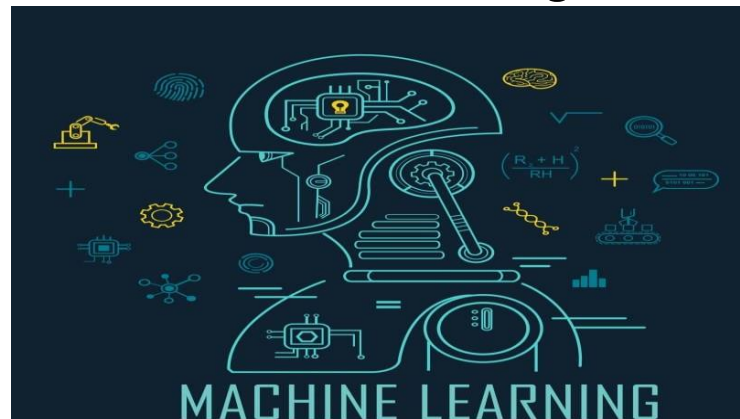
■ Program Curriculum

Probability and Statistics with Numpy

We will continue with statistics and probability and we will deal with descriptive and inferential statistics along with Hypothesis testing and a lot of other relevant statistics methods

- Inferential statistics - sample vs population
- CLT and its proof
- Chi-squared distribution and its properties Point and Interval Estimators
- Estimation technique – MLE
- Interval Estimator of μ with unknown σ
- Examples of estimators
- Hypothesis Testing – I
- Hypothesis Testing – II
- Hypothesis Testing - III
- Assessment
- Handling missing data
- Techniques to impute missing values
- Encoding the data
- Outlier detection and correction
- Meaningful data transformation
- Assessment

Machine Learning



■ Program Curriculum

Machine Learning

We are going to explore the need for machine learning and its types, Algorithms when to use and how to use essential mathematical intuition along with Evaluation metrics. We will see in detail about regression algorithms.

- Introduction to machine learning
- Expert systems and 6 Jars
- Supervised Learning - Regression and Classification
- Evaluation metrics and measuring accuracy
- Introduction to regression & Classification
- Interpreting models
- Regularisation - Ridge and Lasso
- Evaluation metrics - TP, FP and AUC
- Classification using logistic regression
- Classification using KNN
- SVM
- Introduction to decision trees
- Building, pruning and interpreting trees
- Ensemble techniques - Bagging and boosting
- Random forests and Boosted trees - Gradient boosting
- Comparison of supervised techniques
- Do's and Dont's while training ML models
- Handling imbalanced data
- Undersampling and Oversampling
- Other methods - ROSE, SMOTE, etc.
- Introduction to unsupervised learning
- Market Basket Analysis
- K means algorithm
- Assessment

PowerBI



Power BI

■ Program Curriculum

POWER BI

This syllabus provides a structured approach to learning Power BI, covering data preparation, modeling, visualization, and advanced features.

- Introduction to Power BI
- Introduction to business intelligence and Power BI
- Key features and components of Power BI
- Installing Power BI Desktop
- Configuring the initial settings
- Connecting to Data Sources
- Importing data from various sources (Excel, databases, online services)
- Cleaning and transforming data in Power BI
- Data Modeling in Power BI
- Understanding tables, relationships, and fields
- Introduction to DAX functions and formulas
- Calculated columns and measures
- Writing basic DAX expressions
- Aggregating and summarizing data
- Creating different types of visualizations (charts, tables, maps)
- Formatting and customizing visuals
- Using slicers and filters
- Implementing drill-down features
- Creating hierarchical structures in visuals
- Working with Power BI accounts
- Using advance charts by Power BI account
- Publishing reports to Power BI Service
- Sharing and collaborating on reports
- Power BI and Excel Integration
- Applying Power BI concepts to a real-world dataset
- Building a comprehensive report and dashboard

Tableau



■ Program Curriculum

Tableau

This curriculum provides a structured approach to learning Tableau, covering basic to advanced features for data visualization and analysis.

- Introduction to Tableau
- Overview of Tableau
- Key features and capabilities
- Installing Tableau Desktop
- Installing Tableau Public
- Account and pricing of Tableau
- Introduction to dashboard of tableau
- Importing and Data source management and connections
- Creating simple charts (bar, line, scatter)
- Customizing colors, labels, and formatting
- Applying filters to data
- Building a data story
- Combining data from multiple sources
- Handling the data relationships
- Optimizing the performance of large datasets
- Integrating tableau with SQL, Excel, Python and R Language
- Building interactive dashboards
- Maps and Geographic Analysis
- Advanced Tableau Features
- Exploring tableau for sharing the data publicly

JIRA

 Jira Software

■ Program Curriculum

JIRA

This syllabus provides a structured approach to learning Jira Software for work management, covering both basic and advanced features.

- Introduction to Jira
- Introduction to Jira as a project management tool
- Understanding Jira's role in work management
- Jira Interface and Navigation
- Navigating through Jira projects and boards
- Understanding the Jira dashboard and project layout
- Creating a Jira Account
- Setting up a Jira account
- Configuring user profiles and preferences
- Creating Projects and Issues
- Managing Issues
- Agile Development with Jira
- Scrum and Kanban Boards
- Configuring screens for data entry
- Jira Reporting

Linux in Data Engineering



■ Program Curriculum

Linux in Data Engineering

Linux OS knowledge is essential in data engineering for managing and optimizing server infrastructure, executing command-line data processing tasks efficiently, and ensuring seamless integration with various data tools and services.

- Understanding the Linux operating system
- Basic command-line navigation and file manipulation
- File system hierarchy and permissions
- User and group management
- Package management with apt/yum
- Process management and monitoring
- System performance tuning and optimization
- Basic networking concepts and configuration
- Text processing tools (grep, awk, sed)
- Working with CSV, JSON, and other common data formats
- Basics of Linux security

Cloud Computing in Data Engineering



■ Program Curriculum

Cloud Computing in Data Engineering

- Definition and characteristics of cloud computing
- Service models: IaaS, PaaS, SaaS
- Deployment models: Public, private, hybrid, and multi-cloud
- Key cloud computing providers (AWS, Azure, Google Cloud)
- Virtualization and containerization
- Understanding virtual machines (VMs) and containers
- Creating and managing VMs and containers in the cloud
- Overview of cloud storage services (AWS S3, Azure Storage Services)
- Overview of cloud-based ETL (Extract, Transform, Load) services
- Introduction to serverless architecture
- Cloud Functions (e.g., AWS Lambda, Azure Functions)
- Security best practices in cloud computing
- Real-world examples of successful cloud-based data engineering implementations

PROJECTS

- Final Project using integrations

"As we embark on this learning journey together, we encourage you to explore and engage actively with the content. Feel free to ask questions, share insights, and apply the knowledge gained to real-world scenarios. Let's make this learning experience both informative and enjoyable. Best wishes for a successful and enriching learning experience with DATA ENGINEERING!!"