Data Science Course Curriculum by 10Pie

👉 Whether you want to start your IT career or looking for guidance on how to plan your future career path, 10pie is your learning hub for tech career knowledge.

Here's a Data Science course syllabus at a glance:

SL No.	Module Name	Topics Covered	Data Science Projects
1	**Data Science Foundations**	- Importance of data in decision-making - The Data Science Lifecycle	1. Traffic Pattern Analysis: Optimize traffic flow and reduce congestion using traffic data. 2. Predicting House Prices: Predict house prices based on features like the number of bedrooms, bathrooms, and location.
2	**Python for Data Science**	- Python Libraries and Frameworks - Advanced Python Concepts	1. COVID-19 Data Visualization: Load a dataset of COVID-19 cases using Matplotlib and Seaborn to create informative visualizations. 2. Spam Classification: Train a Scikit-learn model to classify emails as spam or not spam.
3	**Statistical Inference and Modeling**	- Probability - Hypothesis Testing - Regression Analysis	1. Coin Flip Simulation: Simulate 10,000 coin flips and calculate the probability of getting a certain number of heads. 2. Credit Risk Assessment: Use logistic regression to predict the probability of a customer defaulting on a loan based on credit information.
4	**Machine Learning Fundamentals**	- Supervised Learning - Unsupervised Learning - Model Selection and Evaluation	1. Credit Card Approval: Predict credit card approval based on credit score, income, and debt-to-income ratio using logistic regression. 2. Titanic Survival Prediction: Predict Titanic passenger survival based on demographic and travel information.

5	Deep Learning	- Neural Network Architectures - Libraries and Frameworks - Advanced Deep Learning Topics	1. Image Classification with CNNs: Build a CNN model to classify images into different categories using the CIFAR-10 dataset. 2. Chatbot with Seq2Seq RNNs: Build a chatbot that responds to user queries using a sequence-to-sequence RNN model.
6	Natural Language Processing (NLP)	- Text Preprocessing and Representation - NLP Applications - Libraries and Frameworks	1. Language Translation: Use the Transformers library to build a machine translation model. 2. News Navigator: Implement a Named Entity Recognition (NER) model to extract named entities from news articles.
7	Big Data and Distributed Computing	- Big Data Ecosystem - Spark Programming - Scalable Machine Learning	1. Twitter Sentiment Analysis: Analyze Twitter tweets in real time using Spark Streaming. 2. Customer Purchase Prediction: Build a machine learning model using Spark MLlib to predict customer purchases.
8	Data Engineering and Pipelines	- Data Ingestion and Extraction - Data Transformation and Orchestration - Data Quality and Governance	1. Weather Data Ingestion: Ingest weather data from APIs and web scraping using Apache Airflow. 2. Data Quality Guard: Create a data quality pipeline using Apache Airflow to detect anomalies.

Module 1: Data Science Foundations

- Importance of data in decision-making
- The Data Science Lifecycle (problem definition, data collection, preprocessing, EDA, feature engineering, model building and evaluation, deployment, and monitoring)

⭐ **Hands-on projects to practice:**

- **Traffic Pattern Analysis:** Optimize traffic flow and reduce congestion using traffic data.
- **Predicting House Prices:** Predict house prices based on features like the number of bedrooms, bathrooms, and location.
- **Weather Forecasting:** Analyze weather data to predict weather patterns and temperatures.

Module 2: Python for Data Science

Python Libraries and Frameworks

- NumPy for numerical computing
- Pandas for data manipulation and analysis
- Matplotlib and Seaborn for data visualization
- Scikit-learn for machine learning

Advanced Python Concepts

- List comprehensions and generator expressions
- Functional programming (lambda, map, filter, reduce)
- Object-oriented programming principles
- Decorators and context managers

⭐ **Hands-on projects to practice:**

- **COVID-19 Data Visualization:** Load a dataset of COVID-19 cases using Matplotlib and Seaborn to create informative and attractive visualizations(line plots, bar charts, and heatmaps).
- **Spam Classification:** Train a Scikit-learn model to classify emails as spam or not spam.
- **Web Scraper:** Use list comprehensions and generator expressions to build a web scraper that extracts data from a website.
- **Machine Learning Model Deployment:** Use Scikit-learn and Flask to deploy a machine learning model as a web application.

Module 3: Statistical Inference and Modeling

Probability Distributions

- Discrete distributions (Bernoulli, Binomial, Poisson)
- Continuous distributions (Normal, Exponential, Gamma)
- Joint and conditional probability

Hypothesis Testing

- One-sample and two-sample tests
- ANOVA and Chi-square tests
- Non-parametric tests

Regression Analysis

- Linear regression (simple and multiple)
- Logistic regression for classification
- Regularization techniques (Ridge, Lasso, Elastic Net)

⭐ **Hands-on projects to practice:**

- **Coin Flip Simulation:** Simulate 10,000 coin flips and calculate the probability of getting a certain number of heads.
- **Website Conversion Rate:** Use a one-sample t-test to determine if a website's conversion rate is significantly different from an industry benchmark.
- **Energy Consumption Prediction:** Use simple linear regression to predict energy consumption based on a single feature (e.g., number of occupants).
- **Credit Risk Assessment:** Use logistic regression to predict the probability of a customer defaulting on a loan based on credit information.

Module 4: Machine Learning Fundamentals

Supervised Learning

- Linear and logistic regression
- Decision trees and random forests
- Support Vector Machines (SVMs)
- Ensemble methods (bagging and boosting)

Unsupervised Learning

- K-means clustering

- Hierarchical clustering
- Principal Component Analysis (PCA)
- Anomaly detection techniques

Model Selection and Evaluation

- Train-validation-test split
- Cross-validation techniques
- Performance metrics (accuracy, precision, recall, F1-score)
- ROC curves and AUC

⭐ **Hands-on projects to practice:**

- **Credit Card Approval:** Predict credit card approval based on credit score, income, and debt-to-income ratio using logistic regression.
- **Wine Quality Prediction:** Use ensemble methods (bagging and boosting) to predict wine quality based on features like chemical composition and sensory data.
- **Gene Expression Analysis:** Use hierarchical clustering to identify patterns in gene expression data.
- **Titanic Survival Prediction:** Predict Titanic passenger survival based on demographic and travel information.

Module 5: Deep Learning

Neural Network Architectures

- Feedforward neural networks
- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)
- Autoencoders and Generative Adversarial Networks (GANs)

Deep Learning Libraries and Frameworks

- TensorFlow and Keras
- PyTorch

Advanced Deep Learning Topics

- Transfer learning and fine-tuning
- Attention mechanisms
- Reinforcement learning
- Interpretability and explainability

⭐ **Hands-on projects to practice:**

- **Image Classification with CNNs:** Build a CNN model to classify images into different categories (e.g., animals, vehicles, buildings) using the CIFAR-10 dataset.
- **Sentiment Analysis with RNNs:** Develop an RNN model to classify movie reviews as positive or negative using the IMDB dataset.
- **Generative Adversarial Networks (GANs) for Face Generation:** Build a GAN model to generate new face images using the CelebA dataset.
- **Chatbot with Seq2Seq RNNs:** Build a chatbot that responds to user queries using a sequence-to-sequence RNN model trained on the Cornell Movie Dialog Corpus.

Module 6: Natural Language Processing (NLP)

Text Preprocessing and Representation

- Tokenization and normalization
- Stemming and lemmatization
- Bag-of-words and TF-IDF
- Word embeddings (Word2Vec, GloVe, FastText)

NLP Applications

- Sentiment analysis (using lexicon-based and machine-learning approaches)
- Text classification (using Logistic Regression, SVM, and Deep Learning)
- Named Entity Recognition (NER) (using Conditional Random Fields (CRFs) and Deep Learning)
- Machine translation
- Text generation

NLP Libraries and Frameworks

- NLTK and spaCy
- Gensim for topic modeling
- Transformers for state-of-the-art NLP models

⭐**Hands-on projects to practice:**

- **Language Translation:** Use the Transformers library to build a machine translation model to translate sentences from one language to another using the WMT dataset.
- **Topic Tracker:** Apply topic modeling using Gensim to extract underlying topics from a dataset of news articles.
- **News Navigator:** Implement a Named Entity Recognition (NER) model to extract named entities (e.g., people, organizations, locations) from news articles.
- **Word Wizard:** Use Word2Vec, GloVe, and FastText to create word embeddings and calculate text similarity between sentences.

Module 7: Big Data and Distributed Computing

Big Data Ecosystem

- Hadoop (HDFS, MapReduce, Hive, Spark)
- Apache Spark, Spark Streaming, and Kafka
- NoSQL databases (MongoDB, Cassandra, HBase)

Spark Programming

- RDDs and DataFrames
- Spark SQL and Datasets
- Spark MLlib for machine learning
- Spark Streaming for real-time data processing

Scalable Machine Learning

- Distributed training and inference
- Hyperparameter tuning at scale
- Model serving and deployment

⭐**Hands-on projects to practice:**

- **Twitter Sentiment Analysis:** Analyze Twitter tweets in real time using Spark Streaming, Spark MLlib, and MongoDB.
- **Customer Purchase Prediction:** Build a machine learning model using Spark MLlib to predict customer purchases based on transaction data.
- **Scalable Recommendation System:** Build a scalable recommendation system using Apache Spark, Spark MLlib, and TensorFlow Serving.

Module 8: Data Engineering and Pipelines

Data Ingestion and Extraction

- Batch and streaming data sources
- APIs and web scraping
- Data lakes and data warehouses

Data Transformation and Orchestration

- ETL pipelines with Apache Airflow
- Data transformation with Apache Beam
- Containerization and orchestration (Docker, Kubernetes)

Data Quality and Governance

- Data profiling and anomaly detection
- Data lineage and provenance
- Privacy and security considerations (laws like GDPR, CCPA, HIPAA)

⭐**Hands-on projects to practice:**

- **Weather Data Ingestion:** Ingest weather data from APIs and web scraping using Apache Airflow, and load into a data warehouse using Apache Beam.
- **Data Quality Guard:** Create a data quality pipeline using Apache Airflow to detect anomalies and perform data profiling, with data lineage and provenance using Apache Atlas and Apache Beam.
- **ETL Flow:** Build a scalable ETL pipeline by packaging it with Docker and managing it with Kubernetes, using Apache Beam to move and prepare batch data (e.g., CSV files) for a PostgreSQL database.

- **Privacy Shield:** Implement data privacy and security considerations in a data pipeline using Apache Airflow and Apache Beam, with access control and encryption using Apache Ranger and Apache Knox.

B.Sc Data Science Syllabus

The BSc (Hons) in Data Science is a 3-year undergraduate (UG) program that provides students with a strong foundation in Data Science principles and practices.

The average fees for the BSc (Hons) in Cyber Security course range from INR 30,000 to 4,00,000 per annum, depending on the college and location.

Semester	Name	Topics Covered
I	Fundamentals of Data Science	Introduction to Data Science, Linear Algebra, Basic Statistics, Programming in C, Communication Skills in English, Python Programming, Introduction to Geospatial Technology
II	Programming for Data Science	Probability and Inferential Statistics, Discrete Mathematics, Data Structures and Program Design in C, Computer Organization and Architecture, Machine Learning, Advanced Python Programming for Spatial Analytics, Image Analytics
III	Data Management and Analytics	Programming in C Lab, Microsoft Excel Lab, Research Proposal, Natural Language Processing, Genomics, Data Warehousing and Multidimensional Modeling
IV	Advanced-Data Science Techniques	Data Structure Lab, Exploratory Data Analysis, Programming in R Lab, Research Publication
V	Machine Learning and Big Data	Machine Learning II, Introduction to Artificial Intelligence, Big Data Analytics, Data Visualizations, Programming in Python Lab

VI	Capstone and Practical Experience	Elective papers, Grand Viva, Major Project

B.Tech Data Science syllabus

The B.Tech in Data Science is a 4-year undergraduate program that equips students with the knowledge and skills to analyze and interpret complex data to make informed decisions. You must complete 12th grade with a minimum of 45-60% marks, including Mathematics.

Here is the B.Tech Data Science syllabus semester-wise.

Semester	Course Name	Topics Covered
1	Problem-Solving Using C	Basics of C programming, algorithms, and problem-solving techniques
	Data Structures	Linear and non-linear data structures, algorithms for data manipulation
	Python for Data Science	Python programming, data structures, and libraries for data analysis
2	Analytical Mathematics	Advanced calculus, differential equations, and applications
	Data Structures	Linear and non-linear data structures, algorithms for data manipulation
3	Applied Linear Algebra	Vector spaces, linear transformations, and matrix theory

	Design and Analysis of Algorithms	Algorithm design techniques, complexity analysis, and optimization
	Database Management Systems	Database design, SQL, and data modeling
	Java Programming	Object-oriented programming concepts and Java applications
	R for Data Science	Statistical computing and graphics using R
4	Discrete Mathematics	Set theory, combinatorics, graph theory, and logic
	Data Wrangling	Techniques for data cleaning, transformation, and preparation
	Data Handling and Visualization	Techniques for data visualization and presentation
5	Probability and Statistics	Probability theory, statistical inference, and data analysis
	Business Intelligence and Analytics	BI tools, data analysis, and decision-making processes

	Predictive Modeling and Analytics	Techniques for predictive modeling and analysis
	Artificial Intelligence	Introduction to AI concepts and applications
6	Machine Learning	Supervised and unsupervised learning techniques
	Data Warehousing and Data Mining	Concepts of data warehousing and mining techniques
	Modern Software Engineering	Software development methodologies and practices
7	Text Analytics and Natural Language Processing	Techniques for analyzing and processing text data
	Big Data and Analytics	Big data technologies and analysis techniques
	Time Series Analysis and Forecasting	Methods for analyzing time series data
	Deep Learning	Neural networks and deep learning architectures

| 8 | Project & Viva-Voce | Comprehensive project presentation and evaluation |
| | Capstone Project | Final project demonstrating cumulative knowledge and skills |

BCA Data Science syllabus

The BCA in Data Science is a 3-year undergraduate program that equips students with the knowledge and skills to analyze and interpret complex data to make informed decisions. You must complete your 12th grade with a minimum of 45-60% marks, including Mathematics.

- This course is designed to provide you with knowledge in both computer applications and data science, bridging the gap between the two fields.

Here are some key details about the BCA Data Science program:

Semester	Course Name	Topics Covered
1	Problem-Solving Using C	Basics of C programming, algorithms, and problem-solving techniques
	Data Structures	Linear and non-linear data structures, algorithms for data manipulation
	Computer Essentials for Data Science	Basics of computer systems and applications
	Statistics and Probability	Statistical methods and probability theory

	Database Management Systems	Database concepts, SQL, and data modeling
	Data Structure and Algorithm	Data structures and algorithm design
3	Introduction to Data Mining	Data mining techniques and applications
	Python Programming	Python programming for data science
	Object Oriented Programming using C++	OOP principles and C++ programming
4	Data Modelling and Visualization	Techniques for data modeling and visualization
	R Programming for Data Sciences	R programming for statistical analysis
	Machine Learning	Introduction to Machine Learning Algorithms
	Big Data Analytics	Techniques and tools for big data analysis
	Natural Language Processing	Techniques for processing and analyzing natural language data

	Information and Data Security	Data security principles and practices
	Project	Capstone project demonstrating cumulative knowledge and skills
6	Minor Project	Smaller scale project for practical experience

M.sc Data Science Syllabus

The M.Sc in Data Science is a 2-year program focused on advanced data analysis, machine learning, and big data technologies.

Designed for graduates with a relevant background, the program typically requires 50-60% in a bachelor's degree and may include entrance exams or interviews.

Semester	Course Name	Topics Covered
	Introduction to Data Science	Data science lifecycle, data types, data collection, and preprocessing
	Programming for Data Science	Python/R programming, data manipulation, libraries (NumPy, Pandas)
	Probability and Statistics	Probability theory, random variables, descriptive and inferential statistics
1	Machine Learning I	Supervised learning algorithms, regression, classification, and decision trees

2	Data Visualization	Data visualization principles, tools, interactive visualizations
	Machine Learning II	Unsupervised learning, clustering algorithms, dimensionality reduction
	Big Data Technologies	Hadoop, Spark, streaming data processing, NoSQL databases
	Data Mining	Data mining process, association rule mining, anomaly detection
3	Natural Language Processing	Text preprocessing, sentiment analysis, named entity recognition
	Deep Learning	Neural networks, deep learning architectures, CNNs, RNNs
	Computer Vision	Image processing, object detection, facial recognition
4	Capstone Project	Comprehensive data science project, applying learned concepts

M.Tech Data Science Syllabus

The M.Tech in Data Science is a 2-year postgraduate program focused on advanced data analysis, machine learning, and big data technologies.

It's designed for graduates with a relevant background and typically requires 50-60% in a bachelor's degree, along with qualifying in entrance exams like GATE, followed by an interview.

Semester	Course Name	Topics Covered
1	Mathematical Foundation for Data Science	Probability theory, statistics, random processes, linear algebra, matrices
	Data Structures and Algorithms	Algorithm analysis, data structures (lists, trees, graphs), sorting, searching
	Machine Learning	Supervised and unsupervised learning algorithms, model evaluation
	Big Data Management	Hadoop ecosystem, NoSQL databases, distributed processing frameworks
2	Data Visualization	Data visualization principles, tools (Tableau, D3.js, Matplotlib)
	Elective I: Natural Language Processing	Text processing, sentiment analysis, speech recognition
	Elective II: Deep Learning	Neural networks, deep learning architectures, CNNs, RNNs
	Elective III: Big Data Analytics	Big data analytics tools, predictive modeling, anomaly detection

	Research Methodology	Research design, data collection methods, quantitative and qualitative analysis
3	Seminar	Literature survey, research presentation, peer review
4	Dissertation	Comprehensive research project, thesis writing and defense

Diploma Data Science Course Syllabus

The Diploma in Data Science is a comprehensive program designed to provide practical skills in data analysis, machine learning, and data management.
 Typically lasting 6 months to 1 year, it is suitable for those seeking a focused introduction to data science.
Admission usually requires a basic understanding of mathematics and computer science, with entry based on academic qualifications or entrance tests.

Semester	Course Name	Topics Covered
	Introduction to Data Science	Overview of data science, data types, data collection
	Programming for Data Science	Python programming basics, data structures, control structures
	Probability and Statistics	Probability theory, random variables, descriptive statistics

	Machine Learning I	Supervised learning algorithms, regression, classification
	Data Visualization	Data visualization principles, creating visualizations
	Machine Learning II	Unsupervised learning, clustering algorithms, dimensionality reduction
	Big Data Technologies	Introduction to Hadoop and Spark, NoSQL databases
2	Capstone Project	Applying learned concepts to a data science problem, project presentation

Data Science course subjects and topics to learn

If you want to start a career in data science, below are the topics you need to learn:

- Programming (Python or R)
- Statistics and mathematics
- Data wrangling, manipulation, and management
- Data visualisation
- Machine learning and deep learning

1. Programming (Python or R)

Python and R are often a minimum requirement in entry-level data science roles. Python ranks first as a programming language as per TIOBE and PYPL Index. R is a top option for many data scientists for data manipulation, processing, and so on.

Also, tech Giants like Google, Microsoft, and Netflix heavily rely on Python and R for data science tasks.

Hence, learning these languages will increase your chances of employability, be it internships or placements. You can also learn SAS, SQL, or Julia.

2. Statistics and mathematics

As a data scientist, you should know how to collect, present, and interpret data. Therefore, you should learn different concepts like mean, median, mode, etc., in statistics. You must understand statistical techniques.

You should also cover areas like calculus, linear algebra, matrices, probability, and other important mathematical concepts.

This helps you write high-quality algorithms and machine learning models.

3. Data wrangling, manipulation, and management

These topics help you work with raw, real-world data and perform complex queries.

These tasks are foundational in data science as you must prepare the data to provide accurate business insights. Data wrangling deals with cleaning and organising data sets for easier analysis.

You are also expected to learn database management to extract data and transform it into suitable formats.

Data wrangling tools:

- Altair
- Alteryx
- Talend

Data manipulation tools:

- Pandas
- NumPy
- scikit-learn

Database management tools:

- MySQL

- MongoDB
- Oracle database

4. Data visualisation

Being able to present data is important to being a data scientist. You will need to master reporting and visualisation to present business insights to key stakeholders. So, learn how to create charts, graphs, dashboards, and tables.

 Learning the tools below will prepare you well in this area:

- Tableau
- Power BI
- QlikView/Qlik Sense
- Matplotlib
- Plotly

5. Machine learning and deep learning

As per Stanford University, machine learning is the most in-demand skill followed by NLP. With this skill, you can develop algorithms and models that make predictions and automate decision-making.
Students who learn these techniques can solve real-world problems. These skills are highly sought-after in the job market.

To begin, master the fundamentals of statistics and programming. Then, explore introductory courses on machine and deep learning.

Data Science course fees and duration 2024

What is the course fee of Data Science courses?

The fees of a data science course typically starts from INR 30,000 and can reach up to INR 3 lakhs. You can find various institutes that offer both online and offline data science courses. For instance, IIT Madras offers the following fee structure:
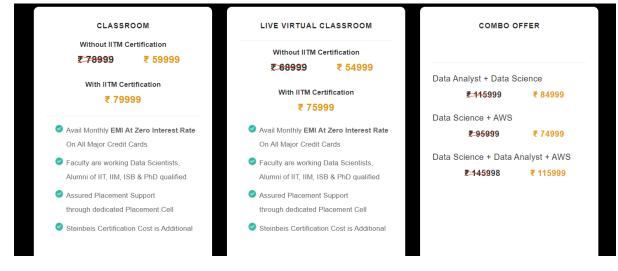
Fee Structure 🎥

✔ For details about application fees, check **Application Process** in Admissions page.

✔ Each term, <u>pay only for the courses you register for in that specific term.</u>

Goal	Total Credits	Total Fees INR
Foundation Only	32	₹32,000
Foundation + One Diploma	59	₹94,500
Foundation + Two Diplomas	86	₹1,57,000
BSc Degree	114	₹2,21,000 - ₹2,27,000
BS Degree	142	₹3,15,000 - ₹3,51,000

✔ Fee waivers depend on category of learner and family income.

Another example: [ExcelR](#) in Hyderabad

The course fees for data science varies on different factors:

1. Brand affiliation or partnerships with Microsoft, Google, NASSCOM, etc.
2. Opting for certification
3. Topics covered (advanced/foundational)
4. Learning format (instructor-led, real-time support)
5. Job placement assistance

Data science course duration

On average, a data science course spans from 6 months to 3 years, depending on the curriculum, projects, and student availability.

For instance, the data science course from IIT Madras is at least 2 years long, and can stretch up to 3 years.

ExcelR, a reputable choice among data science learners, provides a 6-month data science course. It also has various branches in different locations in India.

Who is eligible for Data Science courses?

If you want to enroll in any online training course for Data Science, there is no such criteria or eligibility. However, knowing the basics of computers and data science fundamentals will be helpful.

For academic courses in India: Students are eligible for Data Science courses after completing their 12th grade, with specific criteria depending on the course type:

- **Diploma in Data Science:** Open to any stream with 10+2 completion.
- **BTech in Data Science:** Requires 10+2 with Physics, Chemistry, and Mathematics, along with a minimum of 50% marks.
- **B.Sc/ BCA in Data Science:** Eligible for students who have completed 10+2 with Mathematics, also need at least 50% marks.
- **Postgraduate Courses:** A bachelor's degree in IT or related fields is necessary, with a minimum of 50% marks required.