

PROPOSED SYLLABUS & COURSE FRAMEWORK
ONE YEAR DIPLOMA IN DATA ANALYTICS

ABOUT THE COURSE

Data Analytics is a field of study that deals with understanding and analysing huge amount of data through cutting-edge tools and techniques to identify trends, hidden associations, useful information for improved business decisions. It combines domain expertise, programming skills, along with knowledge of basic concepts of statistics and mathematics in order to derive meaningful insights from the data. Data science, AI, and machine learning are becoming increasingly essential for businesses. Organizations those want to stay competitive in the age of big data, regardless of industry or size, must build and execute data analytics capabilities quickly or risk being left behind. There are four different streams of data analytics- Descriptive analytics describes what happened at a certain time period. Diagnostic analytics is focussed on why something happened. Predictive analytics is concerned with what is expected to occur in the near future. Finally, prescriptive analytics makes a strategy recommendation. Businesses use data analytics based tools to monitor, track, and record performance measures for informed decision-making, efficient operations, better customer support and effective marketing strategies. Data Analytics has tremendous scope in various industries e.g. healthcare, aviation and airlines, cyber security, genomics, automotive, software development, e-commerce, transport sector, supply chain management, risk analysis and management, marketing, etc.

SCOPE

Diploma in Data Analytics at School of Open Learning is designed for students and business professionals. It provides theoretical and practical knowledge on data handling, python programming, database management, statistical techniques, machine learning and related business applications. This program is designed with a perfect balance between theoretical and technology based learning. Capstones projects have been included in the curriculum to encourage practical applications of data management tools and techniques to aid managerial decision making. This program aims to prepare the students for the industry by providing analytical knowledge and hands-on experience with respect to applications of key data analytics techniques in today's business environment.

Duration: 1 year

Total Credits: 44

Mode of Class: Contract/Online

Medium of Teaching: English

Faculty: Faculties from Reputed Institutes/Universities & Industry Experts

SCHEME OF SYLLABUS

Semester 1

Credits: 22

S. No.	Course Title	No. of Credits	Exam Pattern			
			CIA	Practical & Viva	Semester end Exam	Total
1	Descriptive Analytics	4	25	25	50	100
2	Analytics using Spreadsheets	4	25	25	50	100
3	Optimization in Analytics	4	25	25	50	100
4	Database Management System	4	25	25	50	100
5	Machine Learning	4	25	25	50	100
6	Case study on Data Analytics	2	50		-	50

Semester 2

Credits: 22

S. No.	Course Title	No. of Credits	Exam Pattern			
			CIA	Practical & Viva	Semester end Exam	Total
1	Advanced Statistical Techniques	4	25	25	50	100
2	Data Visualization using Excel	4	25	25	50	100
3	Introduction to Python	4	25	25	50	100
4	Business Forecasting and Quality Control	4	25	25	50	100
5	Project based on Applications of Data Analytics	6	The project work shall be carried out under the supervision of faculty members of the department and a project report will be submitted for evaluation at the end of semester. It shall carry 150 marks, where 100 marks will be for the project report; 25 marks on the basis of CIA; 25 marks on the basis of presentation and viva-voce at the end of the semester.			150

*CIA (Continuous Internal Assessment) =
Attendance/ Assignments / Projects/ Presentations/ Written Test etc.

DETAILED SYLLABUS

Semester 1

<p>1. Descriptive Analytics</p>	<p>Data- Types and sources, quantitative and qualitative, attributes, Big Data and its utility in business analytics, Scales of measurement: nominal, ordinal, interval and ratio. Measures of Central Value: Mean, Median and Mode, Measures of Dispersion: Absolute and Relative measures of dispersion - Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation, Box and whisker plot, Moments, Shape of the distribution: Skewness and Kurtosis.</p> <p>Correlation- Meaning and implication. Correlation vs Causation, Types of correlation, Methods for Simple correlation - Scatter diagram, Karl Pearson's coefficient of correlation, Spearman's Rank correlation coefficient,</p> <p>Probability- Meaning and need, Theorems of addition and multiplication, conditional probability, Bayes' theorem, Random Variable- discrete and continuous. Probability Distribution: Meaning, Expectation and variance of Binomial, Poisson and Normal distribution, z-score, Chebyshev and empirical rule, Central limit theorem.</p>
<p>2. Analytics using Spreadsheets</p>	<p>Introduction to Excel About Microsoft Excel – uses of Excel, Excel Spreadsheets, Window bar, Task bar, Formula Bar, Task Pane, File Tab, Selecting Columns and Rows, Width and Height, Auto-Filling Column and Row, Hiding/Unhiding, Inserting and deleting.</p> <p>Ranges and Tables Range name syntax rule, Managing range names, Deleting name with error value, Editing names, Applying name, Using name in formula, Viewing name in workbook, Using name for range intersection, Copying formula with names; Tables- Difference Between Range and Table. Create table, Manage table, Header, Propagation of formula in a table, Table resize, Remove duplicate, Convert to range, Table style.</p> <p>Cleaning Data/Conditional Formatting Cleaning data – Eliminating unwanted characters and numbers, Extracting data values (text, date and time values), Formatting Data (with text functions, dates and time values), Finding workday after particular days, Number of work days</p>

	<p>between two given Dates; Sorting – sort the Data by Text, Numbers, Dates or times, by cell colour, font colour, Sort by custom list & rows (more than one column and row); Filtering – Filter by Selected Values, (Text, Date, Numbers, Cell Colour, Font Colour), Clear Filter, Advanced Filtering, Filters with Slicers; Conditional Formatting – Highlight cell rules, data bars, colour scales, icon sets, clear rules, and manage rules.</p> <p>Functions Some mathematical and statistical functions Sum(), Count(), Average(), Time(), Date(), Left(), Right(), Now(), Count(), Counts(), Mean(), Median() etc., Creating Functions using Excel</p>
<p>3. Optimization in Analytics</p>	<p>Introduction to Operations Research (OR), Linear Programming Problems (LPP), Geometry of linear, Programming, Sensitivity and Post-optimal analysis, Duality and its economic interpretation. Integer Programming Problem. Network models and project planning, Non-linear Programming – KKT conditions, Introduction to Multicriteria Decision Making (MCDM), MCDM methods and applications: Goal programming, Analytic Hierarchy Process (AHP), Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), Data Envelopment Analysis (DEA) method.</p>
<p>4. Database Management System</p>	<p>Introduction- Meaning of Data, Information and Knowledge, Introduction to database, Features and applications of a database, Advantages and limitations of a database, Importance/need for using a database, Types of database, Introduction to Database Management System (DBMS), Operations performed by DBMS: creating a database, populating with data and data management, Types of users in DBMS, Components of DBMS, Entity types, entity set, attribute and key, relationships, relation types, E-R diagrams, database design using ER diagrams</p> <p>DBMS Terminology- Basic terminology: Relations/table, tuples, attributes, relationships, Primary Key, Foreign Key, Candidate Key, Types of Constraints: Domain, Key, Entity Integrity, Referential Integrity</p> <p>Structured Query Language (SQL)- Introduction to SQL, Types of SQL commands: DDL, DML, SQL Commands: create table, select, insert into, delete, update, alter table, drop, SQL Constraints: Not NULL, Unique, Default, Like, Between, In, Distinct, Any, All, SQL queries- Queries using Select-From-Where, Queries using Joins</p> <ul style="list-style-type: none"> • Queries using Aggregate functions and grouping • Nested queries

	Transaction Management- Transaction, Concurrency Control, Distributed Databases.
5. Machine Learning	Introduction- Introduction to machine learning, Applications of Machine Learning, Key elements of Machine Learning, Supervised vs Unsupervised Learning. Regression: Linear regression with one variable, linear regression with multiple variables, gradient descent, logistic regression, over-fitting, regularization. performance evaluation metrics, validation methods. Classification/ Clustering- Decision trees, Naive Bayes classifier, k-nearest neighbour classifier, perceptron, Multilayer perceptron, neural networks, back-propagation algorithm, Approaches for clustering, distance metrics, K-means clustering. Principle component analysis. Applications in Hospitality industry and Retail market.
6. Case study on Data Analytics	As per course and industry requirements

Semester 2

1. Advanced Statistical Techniques	Theory of estimation- Estimation of population mean, confidence intervals for the parameters of a normal distribution (one sample)., Testing of Hypothesis; Null and Alternate Hypothesis, Concept of Level of Significance and Confidence interval, Type 1 and Type 2 Error, Parametric tests- Z test for single mean, Z test for proportion, t test for single mean and paired t test, One-Way ANOVA Test, Non-Parametric Tests; Chi-square test (Goodness of fit), Sign test for median, Sign test for symmetry, Wilcoxon two-sample test. Introduction to basic principles of design of experiments, treatment, plot and block.
2. Data Visualization using Excel	Advance Functions-IF, Nested IF, VLOOKUP, Nested VLOOKUP, Reverse VLOOKUP, INDEX HLOOKUP AND MATCH. Data Visualization- Basics of Pivot Table – Creating Pivot Table, Nesting in Pivot Table, Filters, Pivot table tools, Fields and Areas, Analyze and Design, expanding and collapsing or changing Data Field, Timeline; Displaying Pivot charts, Work with Visualization – Bar/Pie/Donut/Line/Scatter charts._Pivot Table for Data Analysis- Grouping in Pivot Tables, Custom calculation in Pivot table- how to add fields and add a calculated item in worksheet; Add a slicer in Pivot table.
3. Introduction to Python	Elementary Programming- Writing a Simple Program, Reading Input from the Console, Identifiers, Variables, Assignment Statements, and Expressions, Simultaneous Assignments, Named Constants, Numeric Data Types and Operators, Evaluating Expressions and Operator Precedence, Augmented Assignment Operators, Type Conversions and Rounding. Common Python Functions,

	<p>Strings and Characters, Introduction to Objects and Methods, Formatting Numbers and Strings</p> <p>Selections and Loops- Boolean Types, Values, and Expressions, Generating Random Numbers, if Statements, Case Study: Guessing Birthdays, Two-Way if-else Statements, Nested if and Multi-Way if-elif-else Statements, Common Errors in Selection Statements, Logical Operators, Conditional Expressions, Operator Precedence and Associativity. The while Loop, The for Loop, Nested Loops, Minimizing Numerical Errors, Keywords break and continue.</p> <p>Functions and Lists- Defining a Function, Calling a Function, Functions with/without Return Values, Positional and Keyword Arguments, Passing Arguments by Reference Values, Modularizing Code, The Scope of Variables, Default Arguments, Returning Multiple Values, Function Abstraction and Stepwise Refinement. List Basics, Copying Lists, Returning a List from a Function, Passing Lists to Functions, Returning a List from a Function, Searching Lists, Sorting Lists, Processing Two-Dimensional Lists, Passing Two-Dimensional Lists to Functions, Multidimensional Lists.</p> <p>Objects and Classes- Defining Classes for Objects, UML Class Diagrams, Immutable Objects vs. Mutable Objects, Hiding Data Fields, Class Abstraction and Encapsulation, Object-Oriented Thinking.</p>
<p>4. Business Forecasting and Quality Control</p>	<p>Forecasting: Definition, types, qualitative (grass roots, market research and Delphi method) and quantitative approach (simple moving average method, weighted moving average and single exponential smoothing method), forecast error, Mean square error.</p> <p>Quality Control: Statistical Quality control: Variations in process (common & assignable causes), Variable measures (mean and range chart), Attribute measures (proportion of defects and no. of defects) using control tables control charts, OC Curve, 2 card Kanban Production Control system. Brief Exposure to the concept of Total Quality Management, Lean Management and Six Sigma.</p>
<p>5. Project based on Applications of Data Analytics</p>	<p>Capstone Projects as per course and industry needs such as- Fraud detection, Spam detection/ Filtering, Medical Diagnostic Testing, Opinion Mining, Recommender Systems, Portfolio Optimization, Quality Management for Manufacturing and other Industries, Customer Segmentation, Risk Analytics, Ranking and Selection, Market Segmentation, Market-Basket Analytics and similar topics.</p>

Readings:

<https://medium.com/analytics-vidhya/linear-regression-with-gradient-descent-derivation-c10685ddf0f4>.

<https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>

Online course module on DataCamp

[Pivot tables in Spreadsheets](#)

[Conditional Formatting in Spreadsheets](#)

MySQL : Reference Manual

Online course module on DataCamp – [Data Analysis in Excel](#)

Asllani, A. (2014). *Business analytics with management science models and methods*. India: Pearson FT Press.

Bowerman, B. L. (2019). *Business statistics in practice*: Boston: McGraw-Hill/Irwin.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software* (Vol. 2). New York: Springer.

Ehrgott, M. (2005). *Multicriteria optimization* (Vol. 491). Berlin, Heidelberg: Springer Science & Business Media.

Evans, J.R. (2021). *Business Analytics: Methods, Models and Decisions*, India: Pearson.

Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data*: Cambridge university press.

Hanke, J. E., & Wichern, D. W. (2005). *Business forecasting*: Pearson Educación.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

James, R. (2013). *Business analytics: Methods, models and decisions*. India: Pearson FT Press.

Liang, Y. D. (2013). For Introduction to Programming Using Python. *displays*, 8(8), 8.

Paczkowski, W. R. (2022). *Business Analytics: Data Science for Business Problems*: Springer Nature.

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2002). *Database system concepts* (Vol. 5): McGraw-Hill New York.

Taylor, B. W., Bector, C., Bhatt, S., & Rosenbloom, E. S. (2013). *Introduction to management science*: Pearson Boston, MA, USA.

Winston, W. (2016). *Microsoft Excel data analysis and business modeling*: Microsoft press.