# HSNC UNIVERSITY, MUMBAI

## School of Applied Sciences
## Syllabus of BSc Data Science and Business Analytics

## Board of Faculty of Science & Technology

## Board of Studies in the Subject of Data Science & Business Analytics

1. Name of the Chair Person: Dr. Maqsood Khan, Ex Registrar, University of Mumbai, Ex Director, Sydenham Management Institute, Ex Professor and Director, NL Dalmia Institute of Management studies, Ex CEO, HCOI Ministry of Minority Affairs, GOI.

2. Name of the Co – Chairperson: Dr. Nidhi Singh, Academic Head of Data Science and Business Analytics.

Teachers from the college:
3. Beenarani Karutharan, Assistant Professor, Department of Computer science
4. Mrs.Mrunal M Hardikar , Assistant Professor, Department of Mathematics

External Professor:
5. Rosemary Gosling, Director of External Studies(Retd) for the London School of Economics and Political Science (LSE)
6. Dr. Santosh Bothe, Founder and Director AiSense( Start up funded by BIRAC, Govt. of India), Principal, Saraswati College, Shegaon, Affliated to SGBU Amravati University.

External experts:
7. Prof. Parag Mahulikar is Ex- Dean and Senior Professor of Marketing at IES Management College and Research Centre, Bandra, India and Management Consultant.
8. Dr. Alok Deepak Dabade, Assistant Professor, Department of Statistics, University of Mumbai.
9. Dr. Sujata Suvarnapathki, Assitant Professor, Department of Statistics, Ramnarayan Ruia Autonomous College, Matunga, Mumbai.
10. Subhash kumar, Assistant Professor, MCA, MPHIL IT department, St.Xavier's college, Mumbai.
11. Industry Expert: Vinayak Deshpande, Managing Director, Sankhya Analytical Research Pvt. Ltd.
12. Industry Expert: Nishad Kapadia, MCA, Technical trainer,Data and Solution Architect Project Manager, TeraData, Mumbai.
13. Industry Expert: Gaurav Tiwari MSc, Working in Kotak Mahindra bank Ltd.,
14. Industry Expert: Praveena Premanand Menon, MSc in Big Data Analytics.
15. Alumni: Rushabh Maru, Harsh Mahapadi

**Program Outcomes (POs)**

After successfully completing the course, the students will be able to:

1. Apply principles of mathematics, statistics, and computing to analyse and interpret complex data.
2. Demonstrate proficiency in programming languages and computational tools relevant to data science.
3. Perform data acquisition, cleaning, pre-processing, and management using appropriate methodologies.
4. Analyze structured and unstructured data using suitable statistical and analytical techniques.
5. Design and develop effective data visualizations to communicate insights clearly.
6. Apply machine learning and data mining techniques for predictive and prescriptive analysis.
7. Utilize contemporary tools, technologies, and frameworks in data science applications.
8. Identify, formulate, and solve real-world problems using data-driven approaches.
9. Understand and adhere to ethical principles, data privacy, and security standards.
10. Communicate technical information effectively to both specialized and non-specialized audiences.
11. Function effectively as an individual as well as a member of multidisciplinary teams.

# Part –I

## Outline of Choice Based Credit System as outlined by University Grants Commission:

**R. **** : The Definitions Of The Key Terms Used In The Choice Based Credit System And Grading System Introduced From TheAcademicYear 2026-2027 Are As Under:**

1. **Core Course:** A course, which should compulsorily be studied by a candidate as a core requirement is termed as a core course.

2. **Elective Course:** Generally, a course which can be chosen from a pool of courses and which may be very specific or specialized or advanced or supportive to the discipline/subject of study or which provides an extended scope or which enables an exposure to some other discipline/subject/domain or nurtures the candidate's proficiency/skill is called an Elective Course.

2.1 **Discipline Specific Elective (DSE) Course:** Elective courses may be offered by the main discipline/subject of study is referred to as Discipline Specific Elective. The University/Institute may also offer discipline related Elective courses of interdisciplinary nature (to be offered by main discipline/subject of study).

2.2 **Dissertation/Project:** An elective course designed to acquire special/advanced knowledge, such as supplement study/support study to a project work, and a candidate studies such a course on his own with an advisory support by a teacher/faculty member is called dissertation/project. A Project/Dissertation work would be of 6 credits. A Project/Dissertation work may be given in lieu of a discipline specific elective paper.

2.3 **Generic Elective (GE) Course:** An elective course chosen generally from an unrelated discipline/subject, with an intention to seek exposure is called a Generic Elective.

P.S.: A core course offered in a discipline/subject may be treated as an elective by other discipline/subject and vice versa and such electives may also be referred to as Generic Elective.

3. **Choice Base Credit System:** CBCS allows students to choose inter- disciplinary, intra-disciplinary courses, skill oriented papers (even from other disciplines according to their learning needs, interests and aptitude) and more flexibility for students.

4. **Honours Program:** To enhance employability and entrepreneurship abilities among the learners, through aligning Inter Disciplinary / Intra Disciplinary courses with Degree Program. Honours Program will have 40 additional credits to be undertaken by the learner across three years essentially in Inter / Intra Disciplinary course.

A learner who joins Regular Undergraduate Program will have to opt for Honours Program in the first year of the Program. However, the credits for honors, though divided across three years can be completed within three years to become eligible for award of honours Degree.

5. **Program:** A Program is a set of course that are linked together in an academically meaningful way and generally ends with the award of a Degree Certificate depending on the level of knowledge attained and the total duration of study, B.Sc. Programs.

6. **Course:** A 'course' is essentially a constituent of a 'program' and may be conceived of as a composite of several learning topics taken from a certain knowledge domain, at a certain level. All the learning topics included in a course must necessarily have academic coherence, i.e. there must be a common thread linking the various components of a course. A number of linked courses considered together are in practice, a 'program'.

7. **Bridge Course:** Bridge course is visualized as Pre semester preparation by the learner before commencement of regular lectures. For each semester the topics, whose knowledge is considered as essential for effective and seamless learning of topics of the Semester, will be specified. The Bridge Course can be conducted in online mode. The Online content can be created for the Bridge Course Topics.

8. **Module and Unit:** A course which is generally an independent entity having its own separate identity, is also often referred to as a 'Module' in today's parlance, especially when we refer to a 'modular curricular structure'. A module may be studied in conjunction with other learning modules or studied independently. A topic within a course is treated as a Unit. Each course should have exactly 3 Units.

9. **Self-Learning: 20% of the topics will be marked for Self-Learning.** Topics for Self Learning are to be learned independently by the student, in a time- bound manner, using online and offline resources including online lectures, videos, library, discussion forums, fieldwork, internships etc.

Evaluative sessions (physical/online), equivalent to the credit allocation of the Self Learning topics, shall be conducted, preferably, every week for each course.

Learners are to be evaluated real time during evaluative sessions. The purpose of evaluative sessions is to assess the level of the students' learning achieved in the topics earmarked for Self-Learning.

The teacher's role in these evaluative sessions will be that of a Moderator and Mentor, who will guide and navigate the discussions in the sessions, and offer concluding remarks, with proper reasoning on the aspects which may have been missed by the students, in the course of the Self-Learning process.

The modes to evaluate self-learning can be a combination of the various methods such as written reports, handouts with gaps and MCQs, objective tests, case studies and Peer learning. Groups can be formed to present self- learning topics to peer groups, followed by Question and Answer sessions and open discussion. The marking scheme for Self-Learning will be defined under Examination and Teaching.

The topics stipulated for self-learning can be increased or reduced as per the recommendations of the Board of Studies and Academic Council from time to time. All decisions regarding evaluation need to be taken and communicated to the stakeholders preferably before the commencement of a semester. Some exceptions may be made in exigencies, like the current situation arising from the lockdown, but such ad hoc decisions are to be kept to the minimum possible.

10. **Credit Point:** Credit Point refers to the 'Workload' of a learner and is an index of the number of learning hours deemed for a certain segment of learning. These learning hours may include a variety of learning activities like reading, reflecting, discussing, attending lectures / counseling sessions, watching especially prepared videos, writing assignments, preparing for examinations, etc. Credits assigned for a single course always pay attention to how many hours it would take for a learner to complete a single course successfully. A single course should have, by and large a course may be assigned anywhere between 2 to 8 credit points wherein 1 credit is construed as corresponding to approximately 30 to 40 learning hours.

11. **Credit Completion and Credit Accumulation:** Credit completion or Credit acquisition shall be considered to take place after the learner has successfully cleared all the evaluation criteria with respect to a single course. Thus, a learner who successfully completes a 4 CP (Credit Point) course may be considered to have collected or acquired 4 credits. Learner level of performance above the minimum prescribed level (viz. grades / marks obtained) has no bearing on the number of credits collected or acquired. A learner keeps on adding more and more credits as he completes successfully more and more courses. Thus, the learner 'accumulates' course wise credits.

12. **Credit Bank:** A Credit Bank in simple terms refers to stored and dynamically updated information regarding the number of Credits obtained by any given learner along with details regarding the course/s for which Credit has been given, the course-level, nature, etc. In addition, all the information regarding the number of Credits transferred to different programs or credit exemptions given may also be stored with the individual's history.

13. **Credit Transfer:** (performance transfer) When a learner successfully completes a program, he/she is allowed to transfer his/her past performance to another academic program having some common courses and Performance transfer is said to have taken place.

14. **Course Exemption:** Occasionally, when two academic programs offered by a single university or by more than one university, may have some common or equivalent course-content, the learner who has already completed one of these academic programs is allowed to skip these 'equivalent' courses while registering for the new program. The Learner is 'exempted' from 'relearning' the common or equivalent content area and from re-appearing for the concerned examinations. It is thus taken for granted that the learner has already collected in the past the credits corresponding to the exempted courses.

Part-II

**O***** The fees for transfer of credits or performance will be based on number of credits that a learner has to complete for award of the degree.**

**The Scheme of Teaching and Examination:**

The performance of the learners shall be evaluated in two components: Internal Assessment with 15 marks by way of continuous evaluation and by Semester End Examination with 60 marks by conducting the theory examination.

**INTERNAL ASSESSMENT: - It is defined as the assessment of the learners on the basis of continuous evaluation as envisaged in the credit-based system by way of participation of learners in various academic and correlated activities in the given semester of the programme.**

## 1. For Theory Courses

| Sr. No. | Particulars | Marks |
|---|---|---|
| 1 | Self-Learning Evaluation | 15 Marks |

## 2. For Courses with Practical

Each practical course can be conducted out of 25 marks

**The semester end examination (external component) of 60% for each course will be as follows:**

i) **Duration** – **2 Hours**

ii) **Theory Question**

**Paper Pattern**: -

1. There shall be three questions each of 20 marks. On each unit there will be one question.
2. All questions shall be compulsory with internal choice within the questions. (Each question will be of 30 marks with options.)
3. Question may be subdivided into sub-questions a, b, c… and the allocation of marks depend on the weightage of the topic.

The marks will be given for all examinations and they will be converted into grade (quality) points. The semester-end, final grade sheets and transcripts will have only credits, grades, grade points, SGPA and CGPA.

## 3. Project:

• Project which can in the following forms
− Case Studies
− Videos

− Blogs
− Research paper (Presented in Seminar/Conference)
− Field Visit Report
− Presentations related to the subject (Moot Court, Youth Parliament, etc.)
− Internships (Exposition of theory into practice)
− Open Book Test
− any other innovative methods adopted with the prior approval of Director Board of Examination and Evaluation.

## 4. Self-Learning Evaluation:

**− 20% OF THE TOPICS OF CURRICULUM ARE LEARNED BY THE STUDENT THROUGH SELF LEARNING USING ONLINE / OFFLINE ACADEMIC RESOURSE SPECIFIED IN THE CURRICULUM.**

**− HENCE 20% OF THE LECTURES SHALL BE ALLOCATED FOR EVALUATION OF STUDENTS ON SELF LEARNING TOPICS**

**− The identified topics in the syllabus shall be learnt independently by the students in a time bound manner preferably from online resources. Evaluative sessions shall be conducted by the teachers and will carry 10 Marks.**

**CLUB the self-learning topics into 3-4 GROUPS OF TOPICS ONLY FOR EVALUATION.**

**- PRESCRIBE TIME DURATION (IN DAYS) FOR COMPLETION OF EACH GROUP OF TOPIC AND EARMARK SELF LEARNING EVALUATION LECTURES IN THE IMETABLE. HENCE EACH GROUP OF TOPIC CAN BE ASSIGNED 3 REGULAR LECTURES FOR THIS EVALUATION FOR THE ENTIRE CLASS**

**3 Sub Topics**
Each evaluative session shall carry 5 Marks.
**4 Sub Topics**
Each evaluative session shall carry 5 Marks

**EVALUATION OF SELF LEARNING TOPICS CAN COMMENCE IN REGULAR LECTURES ASSIGNED FOR SELF LEARNING EVALUATION IN THE TIMETABLE**

**3 Evaluative sessions**

Each evaluative session shall carry 5 Marks
**4 Evaluative sessions**

Each evaluative session shall carry 5 Marks

## Methods for Evaluation of Self-learning topics:

− Seminars/presentation (PPT or poster), followed by Q&A − Objective questions /Quiz / Framing of MCQ questions.

− Debates

− Group discussion

− You-Tube videos (Marks shall be based on the quality and viewership)

− Improvisation of videos

− Role Play followed by question-answers

**TEACHERS CAN FRAME OTHER METHODS OF EVALUATION ALSO PROVIDED THAT THE METHOD, DULY APPROVED BY THE COLLEGE EXAMINATION COMMITTEE, IS NOTIFIED TO THE STUDENTS AT LEAST 7 DAYS BEFORE THE COMMENCEMENT OF THE EVALUATION SESSION AND IS FORWARDED FOR INFORMATION AND NECESSARY ACTION AT LEAT 3 DAYS BEFORE THE COMMENCEMENT OF THE EVALUATION SESSION.**

− Viva Voce

− Any other innovative method

## SEMESTER END EXAMINATION: -

**It is defined as the examination of the learners on the basis of performance in the semester end theory / written examinations.**

**B. Semester End Examination- 60 Marks**

1) Duration – These examinations shall be of 2 Hours duration.

2) Question Paper Pattern: -

i. There shall be four questions each of 15 marks.

ii. All questions shall be compulsory with internal choice within the questions.

iii. Question may be sub-divided into sub-questions a, b, c, d, e & f only and the allocation of marks depends on the weightage of the topic.

THE MARKS OF THE INTERNAL ASSESSMENT SHOULD NOT BE DISCLOSED TO THE STUDENTS TILL THE RESULTS OF THE CORRESPONDING SEMESTER IS DECLARED.

**HSNC University, Mumbai**

**(2026-2027)**

Ordinances and Regulations

With Respect to

Choice Based Credit System (CBCS)

For the Programmes Under

**The Faculty of Science and Technology**

For the Course

**Data Science & Business Analytics**

**Curriculum – First Year Undergraduate Programmes (NEP)**

**Semester-I and Semester -II**

**(2026-2027)**

# Curriculum Structure

(Under NEP Guidelines)

## SEM 1 and SEM 2

| SEMESTER 1 | | |
|---|---|---|
| **Course Type** | **Course Name** | **Credits** |
| Major | Applied Probability and Statistics | 3 |
| Major | Foundation of Data Analytics using Excel | 3 |
| Minor | Discrete Mathematics for Data Science & Business Analytics | 3 |
| GE | Principles of Management/ An Overview of Indian Theatre | 2 |
| VAC | Contemporary India: Values and Issues - I | 2 |
| IKS | Indian Knowledge system | 1 |
| Ability Enhancement Course (AEC) | Communications Skills in English-I | 2 |
| Vocational | Linux Fundamentals for Data Science | 1 |
| Discipline Specific practical | Applied Probability and Statistics Practical | 1 |
| | Foundation of Data Analytics using Excel Practical | 1 |
| | Discrete Mathematics for Data Science & Business Analytics Practical | 1 |
| **Total Credits:20** | | |
| **SEMESTER 2** | | |
| **Course Type** | **Course Name** | **Credits** |
| Major | Applied Inference and Testing | 3 |
| Major | Data Analysis using R | 3 |

| | | |
|---|---|---|
| Minor | Linear Algebra for Data Science & Business Analytics | 3 |
| GE | Cyber Law/Performing Arts | 2 |
| VAC | Contemporary India: Values and Issues - II | 2 |
| IKS | Indian Knowledge system | 1 |
| Ability Enhancement Course (AEC) | Communications Skills in English-II | 2 |
| Vocational | Analyzing Social Media Networks with NodeXL | 1 |
| Discipline Specific practical | Applied Inference and Testing Practical | 1 |
| | Data Analysis using R Practical | 1 |
| | Linear Algebra for Data Science & Business Analytics Practical | 1 |
| **Total Credits:20** | | |

| | | | |
|---|---|---|---|
| **Course Name:** Applied Probability and Statistics | | **Course Code:** | |
| **Session Per Week (1 session is 60 minutes)** | | **3** | |
| **Credits** | | **3** | |
| | | **Hours** | **Marks** |
| **Evaluation System** | **Theory Examination** | **2** | **60** |
| | **Internal** | **-** | **15** |
| | **Practical Examination** | **2** | **25** |

**Course Outcomes (COs)**

After successful completion of the course, students will be able to:
**CO1:** Understand and interpret various measures of central tendency such as mean, median, mode, and partition values.
**CO2:** Apply appropriate measures of dispersion, skewness, and kurtosis to analyze the variability and shape of data distributions.
**CO3:** Compute and interpret correlation coefficients and analyze relationships between variables.
**CO4:** Develop and apply regression models (linear, quadratic, exponential) for data analysis and prediction.
**CO5:** Evaluate statistical measures and select appropriate techniques for real-world data analysis.
**CO6:** Analyze datasets using statistical methods and interpret results for decision-making.

**Applied Probability and Statistics** **(Total Hours : 45 Lectures)**

| Unit | Content | No. of Lectures |
|---|---|---|
| **1** | 1.1 Probability Basic definitions of probability, Events, Properties of probability.<br>1.2 Conditional probability - Multiplication rule, Independence, Law of total probability, Bayes' theorem<br>1.3 Random Variables - Random variable, Discrete random variable, Probability mass function, Cumulative density function<br>1.4 Expectation and Variance - Expectation of a discrete random variable, Variance, and standard deviation of a discrete random variable. Properties of Expectation and variance. | **15** |
| **2** | Measures of Dispersion, Skewness & Kurtosis:<br>2.1 Concept of dispersion. Requirements of good measure.<br>2.2 Absolute and Relative measures of dispersion: Range, Quartile Deviation, Mean absolute deviation, Standard deviation, Variance and Combined variance.<br>2.3. Raw moments and central moments and relations between them.<br>2.4 Concept of Skewness and Kurtosis: Measures of Skewness: Karl Pearson's, Bowley's and Coefficient of skewness based on moments. Measure of Kurtosis, Box Plot. | **15** |
| **3** | Correlation and regression analysis:<br>3.1 Scatter Diagram, Product moment correlation coefficient and its properties. Spearman's Rank correlation. (With and without ties)<br>3.2 Concept of linear regression. Principle of least squares. Fitting a straight line by method of least squares.<br>3.3 Relation between regression coefficients and correlation coefficient.<br>3.4 Concept and use of coefficient of determination ($R^2$).<br>3.5 Fitting a quadratic curve by method of least squares.<br>3.6 Fitting of exponential curves. | **15** |

Teaching of the unit will be done through teaching mode and through self-learning mode. Evaluation of self-learning topics to be undertaken before the concluding lecture instructions of the respective UNIT.

## Self-Learning topics (Unit wise)

| **Sub Unit** | **Topics** |
|---|---|
| **1** | Probability Basic definitions of probability, Events, Properties of probability. Random variable, Expectation and Variance - Expectation of a discrete random variable, Variance, and standard deviation of a discrete random variable. |
| **2** | Range, Quartile Deviation, Mean absolute deviation, Measures of Skewness, Box Plot |
| **3** | Scatter Diagram, Concept of linear regression and correlation coefficient. |

**Reference Books**


| |
|---|
| Medhi J.: Statistical Methods, An Introductory Text, Second Edition, New Age InternationalLtd. |
| Agarwal B.L.: Basic Statistics, New Age InternationalLtd. |
| Spiegel M.R.: Theory and Problems of Statistics, Schaum's Publications series. TataMcGraw Hill. |
| Kothari C.R.: Research Methodology, Wiley EasternLimited. |
| David S.: Elementary Probability, Cambridge UniversityPress. |
| Hoel P.G.: Introduction to Mathematical Statistics, Asia PublishingHouse. |
| Hogg R.V. and Tannis E.P.: Probability and Statistical Inference. McMillanPublishing Co. Inc. |
| Pitan Jim: Probability, Narosa PublishingHouse. |
| Goon A.M., Gupta M.K., Dasgupta B.: Fundamentals of Statistics, Volume II: The World Press Private Limited,Calcutta. |


| **Course Name: Applied Probability and Statistics Practical** | **Course Code:** |
|---|---|
| **Sessions Per Week (1 session is 60 minutes)** | **2** |
| **Credits** | **1** |

| **Suggestive list of Practical: Applied Probability and Statistics** | |
|---|---|
| 1 | Study of Basic Probability and Conditional Probability |
| 2 | Analysis of Discrete Random Variables Using PMF and CD |
| 3 | Computation of Expectation, Variance, and Standard Deviation of a Discrete Random Variable |
| 4 | Measures of Dispersion: Range, Quartile Deviation, Mean Deviation, and Standard Deviation |
| 5 | Study of Skewness and Kurtosis Using Different Methods and Box Plot Representation |
| 6 | Correlation and Regression Analysis Including Curve Fitting |

## Foundations of Data analytics using Excel (Total Hours : 45 Lectures)

| Course Name: Foundations of Data analytics using Excel | | Course Code: | |
|---|---|---|---|
| **Session Per Week(1 session is 60 minutes)** | | **3** | |
| **Credits** | | **3** | |
| | | **Hours** | **Marks** |
| **Evaluation System** | **Theory Examination** | **2** | **60** |
| | **Internal** | **-** | **15** |
| | **Practical Examination** | **2** | **25** |

**Course Outcomes (COs)**

After successful completion of the course, students will be able to:

**CO1:** Understand the fundamentals of Exploratory Data Analysis (EDA) and classify different types of data used in analysis.

**CO2:** Perform data entry, cleaning, and pre-processing in Microsoft Excel to handle missing values, duplicates, and inconsistencies.

**CO3:** Apply basic and intermediate Excel functions (such as SUM, IF, VLOOKUP/XLOOKUP) for data manipulation and analysis.

**CO4:** Analyze datasets using descriptive statistics and summarize data effectively using Pivot Tables and Pivot Charts.

**CO5:** Create meaningful visualizations (charts, graphs, dashboards) in Excel to identify patterns, trends, and outliers.

**CO6:** Apply statistical tools such as correlation and basic regression using Excel for data interpretation.

**CO7:** Utilize advanced Excel features (INDEX-MATCH, What-if analysis tools) for deeper data analysis and decision-making.

**CO8:** Interpret analytical results and communicate insights effectively through reports and dashboards.

**CO9:** Conduct an end-to-end EDA process, including data collection, cleaning, analysis, visualization, and presentation of findings.

| Unit | Content | No. of lectures |
|---|---|---|
| 1 | **Fundamentals of EDA and Excel Basics**<br>• Introduction to Exploratory Data Analysis (EDA) and its importance in data analysis<br>• Types of data: structured, unstructured, categorical, numerical<br>• Overview of Microsoft Excel interface and tools for data analysis<br>• Data entry, formatting, and organization in Excel<br>• Data cleaning techniques: handling missing values, duplicates, and inconsistencies<br>• Sorting, filtering, and conditional formatting<br>• Data validation and error checking<br>• Introduction to descriptive statistics (mean, median, mode, standard deviation) | **15** |

| | | |
|---|---|---|
| 2 | **Data Analysis and Visualization in Excel**<br>• Data summarization using Pivot Tables and Pivot Charts<br>• Grouping and aggregation techniques<br>• Data visualization principles<br>• Creating charts in Excel: bar charts, line charts, pie charts, histograms, scatter plots<br>• Identifying patterns, trends, and outliers<br>• Using Excel ToolPak for statistical analysis<br>• Correlation analysis and basic regression using Excel | **15** |
| 3 | **Advanced EDA Techniques and Case Study**<br>• Advanced data cleaning and transformation techniques<br>• Working with large datasets efficiently<br>• Basic Excel functions: SUM, AVERAGE, COUNT, IF, VLOOKUP/XLOOKUP<br>• Advanced Excel functions: INDEX-MATCH, nested IF, logical and text functions<br>• What-if analysis: Goal Seek, Scenario Manager, Data Tables<br>• Dashboard creation and interactive visualizations<br>• Interpreting insights and storytelling with data<br>• End-to-end case study: data collection, cleaning, analysis, and visualization<br>• Report preparation and presentation of findings | **15** |

**Self – Learning Topics (Unit wise)**

| **Sub Unit** | **Topics** |
|---|---|
| **1** | Introduction to descriptive statistics (mean, median, mode, standard deviation) |
| **2** | Creating charts in Excel: bar charts, line charts, pie charts, histograms, scatter plots |
| **3** | Data collection, cleaning, analysis, and visualization |

**Reference Books**


| |
|---|
| Data Analysis with Excel, Manisha Nigam (Indian author), BPB Publications (India) |
| Exploratory Data Analysis in Business and Economics – Thomas Cleff |
| Advancing into Analytics – George Mount |
| Exploratory Data Analysis – John W. Tukey |
| Data Visualization with Excel Dashboards and Reports Indian Adaptation. |
| Excel Statistics-A Quick Guide, Nel J. Salkind, Sage Publications. |
| Zeynep Tufekci, How social media took us from Tahrir Square to Donald Trump. |
| Timnit Gebru et al, Datasheets for Datasets. |
| Rachelle Hampton, The Black Feminists Who Saw the Alt-Right Threat Coming |

| Course Name: Foundations of Data analytics using Excel Practical | Course Code: |
|---|---|
| **Sessions Per Week(1 session is 60 minutes)** | **2** |
| **Credits** | **1** |

| Suggestive list of Practicals: Foundations of Data analytics using Excel Practicals | |
|---|---|
| 1 | Data Entry and Cleaning in Excel- Load dataset, handle missing values, duplicates. |
| 2 | Descriptive Statistics in Excel - Calculate mean, median, mode, variance, standard deviation, create summary tables for numerical and categorical data and use Excel ToolPak for statistical functions. |
| 3 | Sorting, Filtering, and Conditional Formatting |
| 4 | Data Summarization Using Pivot Tables and Pivot Charts |
| 5 | Correlation and Basic Regression Analysis |
| 6 | What-if Analysis - Goal Seek, Scenario Manager, Data Tables. |

## Discrete Mathematics for Data Science & Business Analytics (Total Hours : 45 Lectures)

| Course Name: Discrete Mathematics for Data Science & Business Analytics | | Course Code: | |
|---|---|---|---|
| Session Per Week (1 session is 60 minutes) | | 3 | |
| Credits | | 3 | |
| | | Hours | Marks |
| Evaluation System | Theory Examination | 2 | 60 |
| | Internal | - | 15 |
| | Practical Examination | 2 | 25 |

### Course Objectives

1. To develop logical and mathematical thinking for data-driven problem solving
2. To understand discrete structures relevant to data science and business analytics
3. To apply combinatorial and graph-theoretic concepts in real-world scenarios
4. To build a foundation for advanced courses in data science and analytics

### Course Outcomes

After completing the course the students will be able to:

**1.** Understand and apply set theory and logic in data-related problems
**2.** Analyze relations and functions in mathematical and real-world contexts
**3.** Apply combinatorial techniques to solve counting and probability problems
**4.** Model and analyze problems using graph theory concepts
**5.** Interpret real-world scenarios using discrete mathematical tools.

| Unit | Content | No. of Lectures |
|---|---|---|
| **Unit I: Foundations of Discrete Mathematics** | **Set Theory:** Number systems (brief overview), Sets, subsets, power set<br>**Set operations:** Union, Intersection, Difference, Complement, Cartesian products<br>**Relations and Functions:** Relations: definition and types (reflexive, symmetric, transitive)<br>**Functions:** definition and types (one-one, onto, bijective)<br>**Functions in Data Science:** Horizontal and vertical line tests, Composite and inverse functions (basic idea), Exponential functions, Logarithmic Functions<br>**Real-World Contexts**<br>i. Use of set operations in data grouping and segmentation<br>ii. Boolean logic in decision-making and filtering conditions<br>iii. Exponential growth in data and user trends<br>iv. Functional mapping in data transformation processes | 15 |
| **Unit 2: Combinatorics and Discrete Structures** | **Counting Techniques:** Basic counting principles, Permutations and combinations<br>**Advanced Counting Principles:** Pigeonhole principle, Inclusion-Exclusion principle, Summation Techniques Basic summations, Simple manipulation of sums<br>**Recurrence Relations:** Introduction and formulation, Solution of simple recurrence relations<br>**Real-World Contexts**<br>i. Counting possible outcomes in business decision scenarios<br>ii. Use of combinatorics in product selection and arrangements<br>iii. Probability in risk assessment and decision-making<br>iv. Handling overlapping data using inclusion-exclusion principles | 15 |
| **Unit 3: Graph Theory and Network Analysis** | **Basics of Graph Theory:** Graph representation (adjacency matrix and list), Types of graphs<br>**Graph Traversal Algorithms:** Breadth-First Search (BFS), Depth-First Search (DFS)<br>**Shortest Path Algorithms:** Single-source shortest path Dijkstra's algorithm (concept and steps)<br>**Spanning Trees:** Minimum spanning tree, Prim's algorithm, Kruskal's algorithm<br>**Real-World Contexts**<br>i. Graphs in social networks and connectivity analysis | |

| | | |
|---|---|---|
| | ii. Use of traversal algorithms in search and exploration problems<br>iii. Shortest path concepts in route optimization<br>**iv.** Network design using spanning tree concepts | |

Self-Learning topics (Unit wise)

| Unit | Topics |
|---|---|
| 1 | **Set Theory:** Number systems (brief overview)**,** Sets, subsets, power set<br>**Set operations:** Union, Intersection, Difference, Complement, Cartesian products<br>i. Use of set operations in data grouping and segmentation<br>ii. Boolean logic in decision-making and filtering conditions<br>iii. Exponential growth in data and user trends<br>Functional mapping in data transformation processes |
| 2 | **Counting Techniques:** Basic counting principles, Permutations and combinations<br>**Real-World Contexts**<br>i. Counting possible outcomes in business decision scenarios<br>ii. Use of combinatorics in product selection and arrangements<br>iii. Probability in risk assessment and decision-making<br>Handling overlapping data using inclusion-exclusion principles |
| 3 | **Basics of Graph Theory:** Graph representation (adjacency matrix and list)**,** Types of graphs<br>**Real-World Contexts**<br>i. Graphs in social networks and connectivity analysis<br>ii. Use of traversal algorithms in search and exploration problems<br>iii. Shortest path concepts in route optimization<br>Network design using spanning tree concepts |

| **Course Name: Discrete Mathematics for Data Science & Business Analytics Practical** | **Course Code:** |
|---|---|
| **Sessions Per Week(1 session is 60 minutes)** | **2** |
| **Credits** | **1** |

**Suggestive List of Practical:**

Practical should be based on:

1. Set Theory and Data Interpretation
2. Logical Reasoning and Boolean Analysis
3. Functions and Data Mapping
4. Combinatorial Problem Solving
5. Basic Probability and Decision Analysis
6. Graph Theory and Network Analysis

**Reference Books:**


1. Pope, J. Discrete Mathematics for Data Science. 1st Edition. CRC Press / Routledge, 2023.
2. Epp, Susanna S. Discrete Mathematics with Applications. 4th Edition. Cengage Learning, 2011.
3. Rosen, Kenneth H. Discrete Mathematics and Its Applications. 8th Edition. McGraw-Hill Education, 2019.
4. Liu, C. L. Elements of Discrete Mathematics. 2nd Edition. McGraw-Hill, 1985.
5. Lipschutz, Seymour, and Marc Lipson. Schaum's Outline of Discrete Mathematics. 3rd Edition. McGraw-Hill, 2017.
6. Grimaldi, Ralph P. Discrete and Combinatorial Mathematics. 5th Edition. Pearson, 2003.
7. Kolman, Bernard, Robert C. Busby, and Sharon Ross. Discrete Mathematical Structures. 6th Edition. Pearson, 2008.

# SEMESTER II

**Applied Inference and Testing** **(Total Hours: 45 Lectures)**

| **Course Name: Applied Inference and Testing** | | **Course Code:** | |
|---|---|---|---|
| **Session Per Week(1 session is 60 minutes)** | | **3** | |
| **Credits** | | **3** | |
| | | **Hours** | **Marks** |
| **Evaluation System** | **Theory Examination** | **2** | **60** |
| | **Internal** | **-** | **15** |
| | **Practical Examination** | **2** | **25** |

**Course Outcomes (COs)**

After successful completion of the course, students will be able to:

**CO1:** Understand probability distributions such as Binomial, Poisson, Normal, and Exponential and apply them in real-life problems.

**CO2:** Explain the concept of estimation and compute point and interval estimates for population parameters.

**CO3:** Understand the principles of hypothesis testing, including errors, level of significance, and critical regions.

**CO4:** Perform large sample tests for means and proportions and interpret the results.

**CO5:** Apply small sample tests using t-distribution, chi-square test, and F-test in practical situations.

**CO6:** Analyze variance using ANOVA techniques and draw conclusions for decision-making.

**Applied Inference and Testing** **Total Hours: 45 Lectures**

| Unit | Content | No. of Lectures |
|---|---|---|
| 1 | 1.1 Probability distributions:<br>Binomial and Poisson random variables - Bernoulli trials, Binomial distribution, Expectation and variance of a binomial random variable, Poisson distribution.<br>Normal Distribution and Exponential distribution.<br>1.2 Central Limit theorem (statement only).<br>Estimation:<br>1.3 Concept of Parameter, statistic, estimator, and estimate.<br>1.4 Properties of good estimator (Only names), unbiasedness and standard error of an estimator.<br>(Development of critical region is not expected.)<br>1.5 Point and Interval estimate of single proportion, difference of two proportions. | 15 |
| 2 | 2.1 Testing of hypothesis:<br>Concept of hypothesis, Simple Hypothesis and composite hypothesis, Null and alternate hypothesis, Types of errors, Critical region, Level of significance.<br>2.2 Large sample tests: | 15 |

| | | |
|---|---|---|
| | Sampling distribution of sample means and sample proportion.<br>Test of significance for specified value of mean of Normal population.<br>Test of significance for difference between means of two independent Normal populations with equal variances and inequal variances. | |
| 3 | Statistical Tests:<br>3.1 Student's t distribution: Estimation of mean and difference between two means using t distribution.<br>3.2 Exact sample tests using t distribution: Independents and Dependent samples (Paired t test)<br>3.3 Applications of Chi-Square: Testing for association, Testing for variance, Testing for goodness of fit.<br>3.4 Applications of F: Test procedure for testing equality of variances of two independent Normal populations<br>i. Mean is known<br>ii. Mean is unknown.<br>3.5 Concept of ANOVA, One way and Two-way ANOVA | 15 |

**Self – Learning Topics (Unit wise)**

| **Sub Unit** | **Topics** |
|---|---|
| **1** | Concept of Parameter, statistic, estimator, and estimate, difference of two proportions. |
| **2** | Test of significance for difference between means of two independent Normal populations with equal variances |
| **3** | Applications of F: Test procedure for testing equality of variances of two independent Normal populations<br>i. Mean is known |

**Reference Books**


| |
|---|
| Medhi J. : Statistical Methods, An Introductory Text, Second Edition,New Age InternationalLtd. |
| Agarwal B.L. : Basic Statistics, New Age InternationalLtd. |
| Spiegel M.R. : Theory and Problems of Statistics, Schaum's Publications series. TataMcGrawHill. |
| Kothari C.R. : Research Methodology, Wiley EasternLimited. |
| David S. : Elementary Probability, Cambridge UniversityPress. |
| Hoel P.G. : Introduction to Mathematical Statistics, Asia PublishingHouse. |
| Hogg R.V. and Tannis E.P. : Probability and Statistical Inference. McMillanPublishing Co. Inc. |
| Pitan Jim : Probability, Narosa PublishingHouse. |
| Goon A.M., Gupta M.K., Dasgupta B. : Fundamentals of Statistics, Volume II : The World Press Private Limited,Calcutta. |
| Gupta and Kapoor: Fundamentals of Applied Statistics, S. Chand |
| Gupta and Kapoor: Fundamentals of Mathematical Statistics, S. Chand |

| **Course Name: Applied Inference and Testing Practical** | **Course Code:** |
|---|---|
| **Sessions Per Week (1 session is 60 minutes)** | **2** |
| **Credits** | **1** |

| **Suggestive list of Practical: Applied Inference and Testing Practical** | |
|---|---|
| 1 | Computation of Probabilities using Binomial, Poisson, Normal and Exponential |
| 2 | Estimation of Population Parameters using Point and Interval Estimation |
| 3 | Testing of Hypothesis for Population Mean and Proportion (Large Sample Tests) |
| 4 | Testing Difference Between Two Means using Large Sample Tests |
| 5 | Application of t-test for Independent and Paired Samples |
| 6 | Application of Chi-Square, F-test and ANOVA Techniques |

## Data Analysis Using R (Total Hours: 45 Lectures)

| **Course Name: Data Analysis Using R** | | **Course Code:** | |
|---|---|---|---|
| **Session Per Week(1 session is 60 minutes)** | | **3** | |
| **Credits** | | **3** | |
| | | **Hours** | **Marks** |
| **Evaluation System** | **Theory Examination** | **2** | **60** |
| | **Internal** | **-** | **15** |
| | **Practical Examination** | **2** | **25** |

| Unit | Content | No. of Lectures |
|---|---|---|
| 1 | 1 Fundamentals of R<br>1.1 Introduction to R features of R, installation of R, Starting and ending R session, getting help in R ,<br>1.2 Value assigning to variables.<br>1.3 Basic Operations: +, -, *, ÷, ^, sqrt.<br>1.4 Numerical functions: log 10, log , sort, max, unique, range, length, var, prod, sum, summary, dim, sort, five num etc.<br>1.5 Reading and writing data: From and to CSV files and HTML.<br>1.6 Data Type: Vector, list, matrices, array and data frame 1.7 Variable Type: logical, numeric, integer, complex, character and factor.<br>1.8 Operations on matrices.<br>1.9 Control statements: if, if-else, if-else-if, while loop, for loop. 1.10 Defining functions and Printing outputs. | 15 |
| 2 | Data Handling<br>2.1 Data Manipulation: Selecting random N rows, removing duplicate row(s), dropping a variable(s), Renaming variable(s), sub setting data, creating a new variable(s), selecting of random fraction of row(s), appending of row(s) and column(s), simulation of variables.<br>2.2 Data Processing: Data import and export, setting working directory, checking structure of Data :Str(), Class(),, Changing type of variable (for eg as.factor, as.numeric)<br>2.3 Introduction to dplyr and data.table packages<br>2.4 Data Visualisation using simple functions and ggplot: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot eg plot(), correlation plot. | 15 |
| 3 | Statistical Computing:<br>3.1 Descriptive Statistics: Averages, Positional Averages, Dispersion, Skewness , Kurtosis, Correlation Curve Fitting and Regression<br>3.2 Some Probability Distributions: Plotting of pmf/pdf, cdf, computation of probabilities of Binomial, Poisson, Normal, Exponential.<br>3.3 Statistical Tests for t, Chi-square, F and ANOVA<br>3.4 Operations Research Techniques: | 15 |

| | | |
|---|---|---|
| | A Concept and Mathematical Formulation of Linear Programming Problem, Transportation Problems and Assignment Problems. Numerical problems of all above using lpSolve. | |

**Self-Learning topics (Unit wise)**

| Sub Unit | Topics |
|---|---|
| 1.1 | Introduction to R features of R, installation of R, Starting and ending R session, getting help in R |
| 1.8 | Operations on matrices. |
| 2.4 | Data Visualisation using simple functions and ggplot: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot eg plot(), correlation plot. |
| 3.1 | Descriptive Statistics: Averages, Positional Averages, Dispersion, Skewness , Kurtosis, Correlation Curve Fitting and Regression |

| **Course Name: Data Analysis using R Practical** | **Course Code:** |
|---|---|
| **Sessions Per Week (1 session is 60 minutes)** | **2** |
| **Credits** | **1** |

| Suggestive list of Practical: Data Analysis using R Practical | |
|---|---|
| 1 | Introduction to R and Basic Operations |
| 2 | Data Import, Export and Data Structures in R |
| 3 | Control Structures and User-Defined Functions in R |
| 4 | Data Manipulation and Processing using R |
| 5 | Data Visualization using Base R and ggplot2 |
| 6 | Statistical Analysis and Operations Research using R |

Reference Books:

1. Crawley, M. J. (2006). Statistics - An Introduction Using R. John Wiley, London.
2. Purohit, S. G.; Gore, S. D. and Deshmukh, S. R. (2015). Statistics Using R (Second Edition). Narosa Publishing House, New Delhi.
3. Shahababa, B. (2011). Biostatistics with R. Springer, New York.
4. Verzani, J. (2005). Using R for Introductory Statistics. Chapman and Hall/CRC Press, New York.
5. Matloff, N. (2011). The Art of R Programming. No Starch Press, San Francisco.
6. Wickham, H. and Grolemund, G. (2017). R for Data Science. O'Reilly Media, New York.
7. Dalgaard, P. (2008). Introductory Statistics with R (Second Edition). Springer, New York.
8. Grolemund, G. (2014). Hands-On Programming with R. O'Reilly Media, New York.
9. Taha, H. A. (2017). Operations Research: An Introduction (Tenth Edition). Pearson, New Delhi.
10. Venables, W. N. and Ripley, B. D. (2002). Modern Applied Statistics with S (Fourth Edition). Springer, New York.

## Linear Algebra for Data Science & Business Analytics (Total Hours: 45 Lectures)

| | | | |
|---|---|---|---|
| **Course Name: Linear Algebra for Data Science & Business Analytics** | | **Course Code:** | |
| | | | |
| **Session Per Week (1 session is 60 minutes)** | | **3** | |
| **Credits** | | **3** | |
| | | **Hours** | **Marks** |
| **Evaluation System** | **Theory Examination** | **2** | **60** |
| | **Internal** | **-** | **15** |
| | **Practical Examination** | **2** | **25** |

**Course Objectives**

1. To introduce fundamental concepts of matrices and vector spaces
2. To develop understanding of linear systems and transformations
3. To build a foundation for machine learning and data analysis
4. To relate algebraic concepts with real-world data problems

**Course Outcomes**

After completing the course the students will be able to:

1. Perform matrix operations and solve systems of equations
2. Understand and apply eigenvalues and quadratic forms
3. Analyze vector spaces, basis, and dimension
4. Apply concepts of rank and linear independence
5. Interpret linear transformations in data contexts

| Unit | Content | No. of Lectures |
|---|---|---|
| **Unit I: Matrices and Linear Systems** | **Matrices and Operations:** Definition and types of matrices<br>**Basic operations**: addition, multiplication, transpose, properties of matrices<br>**Determinants and Applications:** Determinants of matrices (up to higher order)**,** Properties of determinants, Applications in solving systems and evaluating invertibility<br>**Special Matrices:** Inverse of a matrix, Trace of a matrix, Partition of matrices, Idempotent matrices<br>**Matrix Forms:** Row echelon and reduced row echelon forms, Canonical forms<br>**Systems of Linear Equations;** Solving using Gaussian elimination, Interpretation of solutions (unique, infinite, no solution)<br>**Eigenvalues and Eigenvectors:** Characteristic equation, Eigenvalues and eigenvectors, Basic properties<br>**Quadratic Forms:** Definition, Positive definite and semi-definite matrices (basic idea)<br>**Real-World Contexts**<br>i. Representation of datasets using matrices<br>ii. Solving linear systems in business decision models<br>iii. Eigenvalues in data variance and dimensionality reduction (intuitive idea)<br>iv. Use of matrices in transformations and scaling of data | 15 |
| **Unit 2: Vector Spaces** | **Vector Spaces:** Definition and examples<br>**Subspaces:** Definition and properties<br>**Linear Independence:** Concept and identification<br>**Basis and Dimension:** Basis of a vector space Dimension, Rank and Dimension, Relation between rank and dimension<br>**Rank using Gaussian Elimination:** Finding rank through row reduction<br>**Real-World Contexts**<br>i. Feature representation in datasets as vectors<br>ii. Understanding redundancy and independence in variables<br>iii. Dimension as number of features in data<br>iv. Rank as measure of useful information in datasets | 15 |

| | | |
|---|---|---|
| **Unit 3: Linear Transformations** | **Linear Transformations:** Definition and examples<br>**Matrix Representation:** Representation of linear transformations using matrices<br>**Linear Mapping:** Concept and interpretation<br>**Kernel and Image:** Definition **,** Finding basis for kernel and image<br>**Rank-Nullity Theorem:** Statement and interpretation<br>**Linear Isomorphism:** Definition and basic properties<br>**Real-World Contexts**<br>i. Transformations of data (scaling, rotation, projection)<br>ii. Kernel as loss of information<br>iii. Image as transformed feature space<br>**iv.** Applications in dimensionality reduction (basic intuition) | |

**Self-Learning topics (Unit wise)**

| Sub Unit | Topics |
|---|---|
| 1.1 | **Matrices and Operations:** Definition and types of matrices<br>**Basic operations**: addition, multiplication, transpose, properties of matrices<br>**Determinants and Applications:** Determinants of matrices (up to higher order)**,** Properties of determinants<br>**Real-World Contexts**<br>i. Representation of datasets using matrices<br>ii. Solving linear systems in business decision models<br>iii. Eigenvalues in data variance and dimensionality reduction (intuitive idea)<br>Use of matrices in transformations and scaling of data |
| 1.2 | **Real-World Contexts**<br>i. Feature representation in datasets as vectors<br>ii. Understanding redundancy and independence in variables<br>iii. Dimension as number of features in data<br>Rank as measure of useful information in datasets |
| 1.5 | **Real-World Contexts**<br>i. Transformations of data (scaling, rotation, projection)<br>ii. Kernel as loss of information<br>iii. Image as transformed feature space<br>Applications in dimensionality reduction (basic intuition) |

| Course Name: Linear Algebra for Data Science & Business Analytics Practical | Course Code: |
|---|---|
| **Sessions Per Week (1 session is 60 minutes)** | **2** |
| **Credits** | **1** |

## Suggestive List of Practical:

Practical should be based on:

1. Matrix Operations and Data Representation
2. Solving Linear Systems and Interpretation
3. Eigenvalues and Matrix Analysis
4. Vector Spaces and Linear Independence
5. Rank and Dimension Analysis
6. Linear Transformations and Mapping

**Reference Books:**


1. Lay, David C., Lay, Steven R., and McDonald, Judi J. Linear Algebra and Its Applications. 5th Edition. Pearson, 2016.
2. Strang, Gilbert. Introduction to Linear Algebra. 5th Edition. Wellesley-Cambridge Press, 2016.
3. Anton, Howard, and Rorres, Chris. Elementary Linear Algebra. 11th Edition. Wiley, 2014.
4. Kolman, Bernard, and Hill, David R. Introductory Linear Algebra with Applications. 9th Edition. Pearson, 2008.
5. Lipschutz, Seymour, and Lipson, Marc. Schaum's Outline of Linear Algebra. 6th Edition. McGraw-Hill, 2017.

## Multi-Disciplinary Elective

| Year | Sem. | Papers | Paper Code | Course Title Network security and Data visualization techniques using Software | No of Credits | No of Lectures Hours | Total Credits |
|---|---|---|---|---|---|---|---|
| 1 | I | ME1 | US-FDS-ME1 | MIS and EXCEL for Financial functions | 2 | 30 | 2 |
| | II | ME2 | US-FDS-ME2 | Data analysis using R/EXCEL | 2 | 30 | 2 |

## MS-FDS-ME1: MIS and EXCEL for Financial functions (Total Hours: 30 Lectures)

| Unit | Content | No. of Lectures |
|---|---|---|
| I | Introduction to MIS and Excel Basics<br><br>Understanding Management Information Systems (MIS) and its importance<br>Introduction to Microsoft Excel: Interface, Navigation, and Data Entry<br>Basic Excel Functions: Sum, Average, Count, and IF statements<br>Data Management in Excel<br><br>Sorting and Filtering Data<br>Data Validation and Conditional Formatting<br>Working with Tables and PivotTables<br>Advanced Excel Functions for Financial Analysis<br><br>Financial Functions: NPV, IRR, PMT, and more<br>Time Value of Money (TVM) calculations<br>Analyzing data with Goal Seek and Solver<br>Data Visualization and Charts<br><br>Creating Charts: Line, Bar, Pie, and Scatter plots<br>Customizing Charts and Adding Trendlines<br>Interactive Dashboards in Excel | 15 |

| | | |
|---|---|---|
| II | Advanced Data Analysis Techniques<br><br>Using Excel's Data Analysis Tool Pak<br>Introduction to Power Query for data transformation<br>Scenario Analysis and Sensitivity Analysis<br>Financial Modelling and What-If Analysis<br><br>Building Financial Models in Excel<br>Performing What-If Analysis with Data Tables<br>Monte Carlo Simulation using Excel<br>Macros and Automation in Excel<br><br>Introduction to Excel Macros and VBA (Visual Basic for Applications)<br>Creating and Running Macros<br>Automation and Efficiency Tips<br>Integrating External Data and Collaborating in Excel<br><br>Importing data from external sources (e.g., SQL databases)<br>Collaborating on workbooks using Excel Online and SharePoint<br>Data security and protection<br>Project Work and Case Studies<br><br>Practical projects involving MIS and Financial Data Analysis<br>Analyzing real-world business scenarios using Excel<br>Troubleshooting and Best Practices | 15 |

## **US-FDS-ME2: Data analysis using R**/EXCEL (**Total Hrs: 30 Lectures)**

| Unit | Content | No. of Lectures |
|---|---|---|
| I | Statistical Computations<br>1.1Some Probability Distributions: Plotting of pmf/pdf, cdf, computation of probabilities of Binomial, Poisson, Normal, Exponential<br>1.2 Operations Research Techniques:<br>Concept and Mathematical Formulation of Linear Programming Problem, Transportation Problems and Assignment Problems. Numerical problems of all above using Solver.<br>1.3Concept and Numerical Problems with Excel add-in MegaStat, Real analysis tool pak | 15 |
| II | Data Visualisation using simple functions and ggplot: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot eg plot(), correlation plot. | 15 |

## Vocational Courses Linked to Major/Minor

| Year | Sem. | Papers | Paper Code | Course Title | No of Credits | No of Lectures Hours | Total Credits |
|---|---|---|---|---|---|---|---|
| I | I | V1 | | **Linux Fundamentals for Data Science** | 1 | 30 | 1 |
| I | II | V1 | | **Analyzing Social Media Networks with NodeXL** | 1 | 30 | 1 |

## Linux Fundamentals for Data Science

**Course Objectives**

- To introduce students to the Linux environment and command-line interface
- To develop skills for handling datasets using Linux tools
- To enable basic data processing using shell commands

**Course Outcomes**

After completion of this course, students will be able to:

**CO1:** Understand and use basic Linux commands and file system
**CO2:** Manage and organize datasets using Linux directory structures
**CO3:** Apply Linux commands for text processing and data filtering
**CO4:** Use pipes and redirection for data manipulation workflows

# Linux Fundamentals for Data Science

| Unit | Content | No. of Lectures |
|---|---|---|
| **I: Linux for Data Handling** | **Linux Environment & Basics**<br>Introduction to Linux OS and shell, Terminal usage and basic commands<br>**File System & Dataset Organization**<br>File and directory operations, Organizing datasets<br>**Viewing and Understanding Data Files**<br>File viewing commands, Counting data, File type identification<br>**Text Processing for Data Science**<br>Searching data, Column extraction, Sorting, Removing duplicates<br>**Pipes and Redirection**<br>Output redirection, Input redirection, Pipes and command chaining<br>**File Permissions and Security**<br>File permissions, Ownership, Basic data security practices<br>**System Monitoring**<br>Disk usage, Memory usage, Process monitoring<br>**Working with Compressed Data**<br>Compression, Extraction of dataset files<br>**Basic Shell Scripting**<br>Writing simple shell scripts, Automating repetitive data tasks | 30 |

## Suggestive List of Practical:

1. Creation and organization of a dataset directory structure
2. Displaying and exploring dataset contents
3. Fetching the count of records in a dataset file
4. Extraction of specific columns from a CSV file
5. Search and filter dataset
6. Sort and remove duplicate entries
7. Use pipes to combine multiple commands
8. Compress and extract dataset files
9. Write a basic shell script for data filtering

## Reference Books:


1. The Linux Command Line by William Shotts, 2nd Edition
**2.** Linux Command Line and Shell Scripting Bible (Richard Blum), Wiley Publication
**3.** Mastering Linux Shell Scripting by Ebrahim, Mallatte, Second Edition, Packt Publication

## Analyzing Social Media Networks with NodeXL

| Course Name: Analyzing Social Media Networks with NodeXL | | Course Code: | |
|---|---|---|---|
| **Session Per Week(1 session is 60 minutes)** | | **2** | |
| **Credits** | | **1** | |
| | | **Hours** | **Marks** |
| **Evaluation System** | **Practical Examination** | **2** | **25** |

**Analyzing Social Media Networks with NodeXL** **(Total Hours: 30 Lectures)**

| Sem | Content | No. of Lectures |
|---|---|---|
| I | 1.1 Introduction to Social Media Networks (4 Hours)<br>• Basics understanding of social media networks and their importance in research and real-world applications<br>• Introduction to NodeXL: installation, setup, and basic functionalities<br>1.2 Data Collection and Preprocessing (6 Hours)<br>• Identifying relevant data sources<br>• Collecting data from social media platforms<br>• Data cleaning and preprocessing techniques (handling noise and irrelevant data)<br>1.3 Network Visualization Techniques (5 Hours)<br>• Creating visual representations using NodeXL<br>• Customizing graph layouts, colors, and labels<br>• Filtering and zooming techniques for large networks<br>1.4 Network Metrics and Structural Analysis (6 Hours)<br>• Degree centrality, betweenness centrality, clustering coefficient<br>• Hands-on exercises for calculation and interpretation<br>• Identifying influential nodes and analyzing network structures<br>1.5 Community Detection and Analysis (5 Hours)<br>• Introduction to community detection algorithms<br>• Identifying and visualizing communities using NodeXL<br>1.6 Temporal Analysis and Case Study (4 Hours)<br>• Analyzing temporal patterns in social media data<br>• Visualizing and interpreting network dynamics over time<br>• Case study integrating all concepts<br>Case study based on all the topics covered. | 30 |

**Reference Books**

| |
|---|
| Analyzing Social Media Networks with NodeXL, Insights from a connected world, Derek L. Hansen, Ben Shneiderman, Marc A Smith, and Itai Himelboim. |

| Suggestive list of Practical: Analyzing Social Media Networks with NodeXL Practical | |
|---|---|
| 1 | Orientation and layout |
| 2 | Labelling and visual attributes of edges |
| 3 | Labelling and visual attributes of vertices |
| 4 | Filtering and grouping |
| 5 | Understanding overall graph metrics |
| 6 | Understanding vertex specific graph metrics |
| 7 | Calculating and Visualizing network metrics |
| 8 | Grouping and Filtering |

**GENERAL ELECTIVE**

| Year | Sem. | Papers | Paper Code | Course Title | No of Credits | No of Lectures Hours | Total Credits |
|---|---|---|---|---|---|---|---|
| I | I | | | Principles of Management/ An Overview of Indian Theatre | 2 | 30 | 1 |
| I | II | | | Cyber Law/ An Overview of Indian Classical Music | 2 | 30 | 1 |

## Cyber Law (2 credits)

Course Outcomes:

After completion of course, students will:

- understand basic concepts of cyber law.
- become aware of IT Act and data protection laws.
- be able to identify cybercrimes and find their remedies.
- be able to apply ethical practices in data handling.
- be able to recognize legal responsibilities in data science.

| Unit | Content | No. of Lectures |
|---|---|---|
| Unit I: Foundations of Cyber Law & Data Protection | Introduction to Cyber Law: Meaning, scope, evolution, Importance in digital era<br>Cyber Space & Legal Issues: Cyberspace concept, jurisdiction challenges<br>IT Act 2000: Key features, amendments, digital signatures, e-governance<br>Data Protection & Privacy: Personal data, DPDP Act 2023 overview, Concept of Consent<br>Cyber Ethics: Ethical data usage, bias, accountability & responsibility, Basics of Digital Ethics | (15 Lectures) |
| **Unit II: Cyber Crimes, Security & Legal Compliance** | Cyber Crimes: Types such as hacking, phishing, identity theft, cyber stalking, data breaches, data misuse<br>Legal Provisions: IT Act sections, penalties, adjudication authorities & powers<br>IPR: Copyright, trademark, patents, plagiarism, data ownership<br>Cyber Security: Encryption, authentication, secure storage<br>Compliance & Case Studies: Data breaches, GDPR overview | (15 Lectures) |

**Reference books:**

1. Cyber Law Simplified by Vivek Sood, Tata McGraw-Hill
2. Cyber Laws and Information Technology by Suresh T. Viswanathan, 3rd Edition
3. Guide to Cyber Laws by Pavan Duggal
4. Data Privacy and GDPR Handbook by Alan Calder