

CHRIST COLLEGE (AUTONOMOUS), IRINJALAKUDA



DEGREE OF B.VOC.IT

BACHELOR OF VOCATION IN INFORMATION TECHNOLOGY

(CHOICE BASED CREDIT AND SEMESTER SYSTEM FOR UNDERGRADUATE CURRICULUM)

UNDER THE FACULTY OF VOCATIONAL STUDIES

SYLLABUS

(FOR THE STUDENTS ADMITTED FROM THE ACADEMIC YEAR 2018 – '19 ONWARDS)

BOARD OF VOCATIONAL STUDIES IN IT (UG)

CHRIST COLLEGE (AUTONOMOUS), IRINJALAKUDA - 680125, KERALA, INDIA

JUNE, 2018

ADMISSION CRITERIA

Candidates who have passed (Eligible for Higher Studies) the HSE/VHSE of the Kerala State Board of Higher Secondary Examination or any other examination recognized as equivalent thereto with Mathematics/Statistics/Computer Science/Computer Application are eligible for admission.

COURSE STRUCTURE

Credit Distribution

<i>Semester</i>	<i>Common course</i>		<i>General course</i>	<i>Total for Gen. & Comm.</i>	<i>Skill Component courses</i>	<i>Total for Skill</i>	<i>Total</i>
	<i>English</i>	<i>Second Language</i>					
I	4	4	4	12	3+6+3+6*	18	30
II	4	4	4	12	4+3+7*+4**	18	30
III	4	-	4+4	12	3+6+5*+4*	18	30
IV	4	-	4+4	12	6+4*+4*+4**	18	30
V			4+4+4	12	3+4+5*+6*	18	30
VI					30***	30	30
Total	16	8	36	60		120	180

***Practical **Mini project ***Major project/Internship**

(1 credit = 15h: 30 credit = 450 hours: 1 semester = 90 days: 18 weeks of 5 working days consisting of six hours. For internship/I.V./field-work/self-learning, the credit weightage for equivalent hours shall be 50% of that for lectures/practical)

MARK DISTRIBUTION AND INDIRECT GRADING SYSTEM

Indirect Grading System based on a 7 -point scale is used to evaluate the performance of students. Mark system is followed for the evaluation of answer scripts. After external and internal evaluations, mark lists are submitted to the office of the Controller of Examinations. All other calculations, including grading, will be done by the college using the appropriate software. Indirect Grading System in 7 point scale is followed. Each course is evaluated by assigning marks with a letter grade (A+, A, B, C, D, E or F) to that course by the method of indirect grading.

Mark Distribution

<i>Sl. No.</i>	<i>Course</i>	<i>Marks</i>
1	English	400
2	Additional Language	200
3	General courses	900

4	Skill Component courses	2100
Total Marks		3600

An aggregate of E grade with 40 % of marks (after external and internal put together) is required in each course for a pass (Except for project*) and also for awarding a degree/diploma. Appearance for Internal Assessment and End Semester Evaluation are compulsory and no grade shall be awarded to a candidate if she/he is absent for Internal Assessment / End Semester Evaluation or both. A student who fails to secure a minimum grade for a pass in a course is permitted to write the examination along with the next batch. After the successful completion of a semester, Semester Grade Point Average (SGPA) of a student in that semester is calculated using the formula given below. For the successful completion of a semester, a student should pass all courses. However, a student is permitted to move to the next semester irrespective of SGPA obtained.

Note: *For project/Internship the minimum for a pass shall be 50% of the total marks assigned to the respective examination.

Seven point Indirect Grading System

<i>% of Marks</i>	<i>Grade</i>	<i>Interpretation</i>	<i>Grade Point Average</i>	<i>Range of Grade points</i>	<i>Class</i>
90 and above	A+	Outstanding	6	5.5 - 6.0	First Class with distinction
80 to below 90	A	Excellent	5	4.5 - 5.49	
70 to below 80	B	Very good	4	3.5 - 4.49	First Class
60 to below 70	C	Good	3	2.5 - 3.49	
50 to below 60	D	Satisfactory	2	1.5 - 2.49	Second Class
40 to below 50	E	Pass/Adequate	1	0.5 - 1.49	Pass
Below 40	F	Failure	0	0 - 0.49	Fail

$$\text{SGPA} = \frac{\text{Sum of the credit points of all courses in a semester}}{\text{Total credits in that semester}}$$

The Cumulative Grade Point Average (CGPA) of the student is calculated at the end of a programme. The CGPA of a student determines the overall academic level of the student in a programme and is the criterion for ranking the students. CGPA can be calculated by the following,

$$\text{CGPA} = \frac{\text{Total credit points obtained in six semesters}}{\text{Total credits acquired (180)}}$$

SGPA and CGPA shall be rounded off to two decimal places.

CREDIT AND MARK DISTRIBUTION IN EACH SEMESTER**Total Credits: 180; Total Marks: 3600**

<i>Semester</i>	<i>Course</i>	<i>Credit</i>	<i>Marks</i>
I	Transactions: Essential English Language Skills - A01	4	100
	Malayalam-Bhashayum Sahithyavum I - MAL1A01(2) or Communication Skills in Hindi - A07	4	100
	Basic numerical skills - BC3A11	4	100
	Business analytics	3	100
	Python programming, Basic electronics, Introduction to IOT	6	100
	Database management system - Mongo DB	3	100
	Practical 1 - Python programming & RDBMS	6	100
	Total	30	700
II	Ways with words: Literatures in English - A02	4	100
	Malayalam-Bhashayum Sahithyavum II / Literature in Hindi	4	100
	Basic statistics and probability - ST1CO1	4	100
	Microsoft excel with VBA & SSPS	4	100
	Introduction to data science	3	100
	Practical 2- Microsoft excel for data science & SPSS	7	100
	Mini project	4	100
	Total	30	700
III	Writing for Academic and Professional Success - A03	4	100
	Probability distribution - ST1CO2	4	100
	Problem solving using C - BCS2B02	4	100
	Computer network for IOT	3	100
	Electronic devices and circuit design for IOT, IOT with Raspberry Pi	6	100
	Practical 3 - Python for data science, Computer network for IOT	5	100
	Practical 4 - Basics of IOT and IOT with Raspberry Pi	4	100
	Total	30	700
IV	Zeitgeist: Readings on Contemporary Culture - A04	4	100
	Statistical inferences and applications - ST1CO3	4	100
	Micro controllers and applications - EL6B14	4	100
	Data science with R and Python programming, Embedded Linux OS and Android programming	6	100
	Practical 5 - Data science with R programming & Data visualization using Qlik & Power Bi	4	100

	Practical 6 - Android and LINUX	4	100
	Mini project	4	100
	Total	30	700
V	Artificial intelligence - CS14703	4	100
	Embedded system - CS14601	4	100
	Numerical analysis and optimization techniques - MCA10301	4	100
	Machine learning using python	3	100
	Big data & Cloud platform for IOT	4	100
	Practical 7 - Hadoop & NO SQL Database	5	100
	Practical 8 - Machine learning using Python & Data visualization using Tableau	6	100
	Total	30	700
VI	Major Project	30	100
	Total	30	100
	Grant Total	180	3600

CODES AND INSTRUCTIONAL DETAILS

<i>Semesters</i>	<i>Code</i>	<i>Course Title</i>	<i>Instructional hours/week</i>	<i>Instructional hours/Sem</i>
I	GEC1 EG01	Transactions: Essential English Language Skills - A01	4	60
	GEC1 ML02 or GEC1 HD02	Malayalam-Bhashayum Sahithyavum I - MAL1A01(2) or Communication Skills in Hindi - A07	4	60
	GEC1 BM03	Basic numerical skills - BC3A11	4	60
	SDC1 BA01	Business analytics	3	45
	SDC1 PP02	Python programming, Basic electronics, Introduction to IOT	6	90
	SDC1 DB03	Database management system - Mongo DB	3	45
	SDC1 DB05P	Practical 1 - Python programming & RDBMS	6	90
II	GEC2 EG04	Ways with words: Literatures in English - A02	4	60
	GEC2 ML 05 or GEC2 HD05	Malayalam-Bhashayum Sahithyavum II / Literature in Hindi	4	60
	GEC2 ST 06	Basic statistics and probability	4	60
	SDC2 ME06	Microsoft excel with VBA & SPSS	4	60

	SDC2 DS07	Introduction to data science	3	45
	SDC2 DS08P	Practical 2- Microsoft excel for data science & SPSS	7	105
	SDC2 MP09	Mini project	4	60
III	GEC3 EG07	Writing for Academic and Professional Success - A03	4	60
	GEC3 ST08	Probability distribution	4	60
	GEC3 PC09	Problem solving using C	4	60
	SDC3 CN11	Computer network for IOT	3	45
	SDC3 ED12	Electronic devices and circuit design for IOT, IOT with Raspberry Pi	6	90
	SDC3 PC13P	Practical 3 - Python for data science & Computer network for IOT	5	75
	SDC3 IR14P	Practical 4 - Basics of IOT and IOT with Raspberry Pi	4	60
IV	GEC4 EG10	Zeitgeist: Readings on Contemporary Culture - A04	4	60
	GEC4 ST11	Statistical inferences and applications	4	60
	GEC4 MC12	Micro controllers and applications	4	60
	SDC4 DS15	Data science with R and python programming, Embedded Linux OS & Android programming	6	90
	SDC4 DS17P	Practical 5 - Data science with R programming & Data visualization	4	60
	SDC4 AL18P	Practical 6 - Android and LINUX	4	60
	SDC4 MP19	Mini project	4	60
V	GEC5 AR13	Artificial intelligence	4	60
	GEC5 ES14	Embedded system	4	60
	GEC5 OT15	Optimization techniques	4	60
	SDC5 ML20	Machine learning using python	3	45
	SDC5 BC21	Big data & Cloud platform for IOT	4	60
	SDC5 HN22P	Practical 7 - Hadoop & NO SQL Database	5	75
	SDC5 ML23P	Practical 8 - Machine learning using Python & Data visualization using Tableau	6	90
VI	SDC6 PR24Pr	Major Project	-	900

Commented [1]:
Leyon Varghese Aug 30, 8:40 PM

EVALUATION SCHEME FOR THEORY COURSES

The evaluation scheme for each course contains two parts: *viz.*, internal evaluation and external evaluation.

1. INTERNAL EVALUATION

20% of the total marks in each course are for internal evaluation. The Department shall send only the marks obtained for internal examination to the CoE of the College.

Table 1 : Components of Evaluation (Assuming total marks is 100)

<i>Sl. No</i>	<i>Components</i>	<i>Marks</i>
1	Attendance (Refer Table 1a for the matrix)	5
2	Assignment (Refer Table 1b for the matrix)	5
3	Test paper : I & II (Refer Table 1c for the matrix)	5 + 5
Total Marks		20

Table 1a : Percentage of attendance and eligible internal marks

<i>% of attendance</i>	<i>Marks</i>
Above 90	5
85 - 89	4
80 - 84	3
76 - 79	2
75	1

A candidate shall be permitted to appear for the Semester-End examinations only if he/she secure not less than 75% attendance in the total number of working hours in each semester.

Table 1b : Assignment

Assignments can be home work, problem solving, group discussions, quiz, literature survey, seminar, team project, software exercises, etc. As decided by the faculty handling the course, and regularity in the class. Assignments of every semester shall preferably be submitted in Assignment Book, which is a bound book similar to laboratory record.

Table 1c : Pattern of test papers and eligible internal marks

<i>Duration</i>	<i>Pattern</i>	<i>Total number of questions</i>	<i>Number of questions to be answered</i>	<i>Marks for each question</i>	<i>Marks</i>
1.5 Hours	One word	4	4	1	4
	Short answer	5	4	2	8
	Paragraph	5	3	6	18
	Essay	2	1	10	10
Total Marks*					40

**Eligible internal marks. 90% and above = 5, 80 to below 90% = 4.5, 70 to below 80% = 4, 60 to below 70% = 3.5, 50 to below 60% = 3, 40 to below 50% = 2, 35 to below 40% = 1, below 35% = 0*

2. EXTERNAL EVALUATION

External evaluation carries 80% marks. External examinations will be conducted at the end of each semester.

Table 1 : Pattern of question paper

<i>Duration</i>	<i>Pattern</i>	<i>Total number of questions</i>	<i>Number of questions to be answered</i>	<i>Marks for each question</i>	<i>Marks</i>
3 Hours	One word	10	10	1	10
	Short answer	12	10	2	20
	Paragraph	8	5	6	30
	Essay	4	2	10	20
Total Marks					80

EVALUATION SCHEME FOR PRACTICAL COURSES

Practical evaluation will be conducted at the end of 1st 2nd 3rd 4th and 5th semesters. The evaluation scheme for each course contains two parts: viz., internal evaluation and external evaluation.

Table 1 : Internal evaluation for practical

<i>Sl. No</i>	<i>Criteria</i>	<i>Marks</i>
1	Evaluation in the lab and Rough Record	10
2	Test Paper	4
3	Viva voce	1
4	Regularity (Refer Table 1a for the matrix)	5
Total Marks		20

Note: No candidate will be permitted to attend the end-semester practical examination unless he/she produces certified record of the laboratory.

Table 1a : Percentage of attendance and eligible internal marks

<i>% of attendance</i>	<i>Marks</i>
Above 90	5
85 - 89	4
80 - 84	3
76 - 79	2
75	1

Table 2 : External evaluation for practical

<i>Sl. No</i>	<i>Type of question</i>	<i>Question number</i>	<i>Nature of question</i>	<i>Marks</i>
1	Theory/ Algorithm/Flow diagram			20

2	Implementation			30
3	Result/Output			10
4	Record			10
5	Viva voce			10
Total Marks				80

EVALUATION SCHEME FOR PROJECT/INTERNSHIP

There will be iv/internship evaluation at the end of 2nd and 4th semesters

Table 1 : Mark distribution for mini project (Internal - 20% and External - 80%)

<i>Sl. No</i>	<i>Criteria</i>	<i>Marks</i>
1	Punctuality	20
2	Content and relevance of the report	30
3	Viva voce	20
4	Project presentation	30
Total Marks		100

The major project/internship evaluation will be conducted at the end of 6th semester.

Internship and the major project should be carried out in the industry, not necessarily with industry partner. The major idea for internship is to implement the things learned and to get a real life experience. **The Evaluation process for this alone, follows 100% external assessment.**

Every student will be assigned an internal guide, allotted from the parent department concerned or an expert available in the college appointed by the principal. The student has to make regular discussions with the guide while choosing the subject/area and throughout the life time of the project. At least three reviews should be conducted to evaluate the progress of work. An evaluation team is constituted for conducting the evaluation. The team consist of external examiner, representative from the industry and a faculty of the concerned department. Students should submit a report of their work. A valid certificate from the organization should be produced as a proof that the work is carried out in the respective organization. Students are required to demonstrate the working model of their work (if possible) to the panel of examiners. A viva will be conducted based on the report and students are supposed to clarify the queries regarding their work.

Table 2 : Mark distribution for major internship assessment.

<i>Sl. No</i>	<i>Criteria</i>	<i>Marks</i>
1	Punctuality	20
2	Content and relevance of the report	30
3	Viva voce	20
4	Project presentation	30
Total Marks		100

SEMESTER I

SDC1BA01 – BUSINESS ANALYTICS

Number of Contact Hours: 45 Hrs.

Course Outline

Module 1[8 T]

Introduction to business analytics: IT in business (CRM, Fraud Detection); Excel-Formatting of Excel Sheets, Excel Formula Function, Data Filter and Sort, Charts and graphs, Table formula, pivot tables, Business Analytics vs data science, Introduction to Business Analytics applications-Data visualization tools, Business Intelligence, Self-service analytics, Big data.

Module 2[8 T]

Statistical Techniques: Types of Data, Data summarization, Spreadsheets, Frequency table, Frequency Distribution, Histogram, Measures of Central Tendency, Skewness and Kurtosis, Basic Probability, Conditional Probability, Sampling methods, Point and interval estimation, Central Limit theorem, Null and alternative hypothesis, Level of Significance, P value, Types of errors, Hypothesis Testing, Predictive Analytics-Linear and Multilinear Regression, Forecasting/Time series

Module 3[10 T]

Data mining Techniques: Market Basket Analysis-basic concepts, frequent itemset mining methods, Apriori, FP Growth, Pattern Evaluation Methods-Lift, Chi-square, Classification- Decision Tree induction, Bayes Method, Rule based classification, Clustering- Partitioning methods, Hierarchical methods, Density based methods, Grid based methods

Module 4[9 T]

Applications of R and SAS: Reading and writing data in R, Vectors, Frames and subsets, R code debugger, Managing and manipulating data in SAS, Creating charts in SAS, Linear regression in SAS, Orientation on Big Data and Hadoop: Awareness, The four V's, Applications of Big Data OLTP and OLAP

Module 5[10 T]

Web and Mobile Analytics: Text Analytics, Click Analytics, Sentiment Analytics, Google Analytics, CASE STUDIES: Population Census, Marketing, Banking, Retail, Industrial, Mining Patterns, Making models, Model selection and validation.

References

1. Business Analytics by James Evans
2. Business Analytics: Data Analysis & Decision Making by S. Christian Albright and Wayne L. Winston
3. Introduction to Business Analytics Using Simulation by Jonathan Pinder

SDC1PP02 – PYTHON PROGRAMMING, BASIC ELECTRONICS AND INTRODUCTION TO IOT

Number of Contact Hours: 45 Hrs.

Number of Credits: 6

Course Outline

PART A - PYTHON PROGRAMMING

Module 1[8 T]

Introduction to digital computer – Von Neumann concept – A simple model of computer, acquisition of data, storage of data, processing of data, output of processed data. Details of functional units of a computer. Storage – primary storage and secondary storage. Introduction to programming languages: - types of programming languages - high level language, assembly language and machine language, System software - Operating systems – objectives of operating systems, compiler, assembler and interpreter.

Module 2[9 T]

Problem Solving strategies – Problem analysis – formal definition of problem – Solution – Top-down design – breaking a problem into sub problems- overview of the solution to the sub problems by writing step by step procedure (algorithm) - representation of procedure by flowchart - Implementation of algorithms – use of procedures to achieve modularity. Examples for algorithms and flow charts - At least 10 problems Starting with non-numerical examples, and numeric problems like factorial, largest among three numbers, largest among N, Fibonacci.

Module 3[10 T]

Introduction to Python – Variables, Expressions and Statements, Evaluation of Expressions, Operators and Operands, Order of precedence, String Operations, Control statements, Boolean Expressions and logical operators, conditional and alternative executions, Nested Conditionals, Recursion. Iteration - Multiple Assignment, The While Statement, Tables, Two Dimensional Tables. Encapsulation and generalization, Local Variables.

Module 4[9 T]

Functions, Calling functions, Type conversion and coercion, composition of functions, Mathematical functions, User-defined Functions, Parameters and Arguments. Strings and Lists – string traversal and comparison with List operations with Examples, Tuples and dictionaries – Operations and Examples.

Module 5[9 T]

Files and exceptions - Text files, Directories, Python – Modules, Python – JSON, Python- PIP, Lambda. Classes/Objects-Create class, Create Object, The `__init__()` Function, Python Object Oriented Principles, Object Methods, Self-Parameter, Python Exception Handling, Python MySQL, Python MongoDB

References

1. Downey, A. et al., How to think like a Computer Scientist: Learning with Python, John Wiley, 2015
2. Goel, A., Computer Fundamentals, Pearson Education
3. Lambert K. A., Fundamentals of Python - First Programs, Cengage Learning India, 2015
4. Rajaraman, V., Computer Basics and C Programming, Prentice-Hall India
5. Dromy, R. G., How to solve it by Computer, Pearson India
6. Sprankle, M., Problem Solving & Programming Concepts, Pearson India
7. Website URL - <https://www.w3schools.com/python>

PART - B BASIC ELECTRONICS AND INTRODUCTION TO IOT

Number of Contact Hours: 45 Hrs.

Module 1 [9 T]

Semi conductor diode and applications: p-n junction diode, characteristics and parameters, diode approximations, DC load line analysis, Half-wave rectifier, Two diode-full wave rectifier, Bridge rectifier, Capacitor filter circuit, Zener diode voltage regulators: Regulator circuit with no load, Loaded Regulator. Numerical examples as applicable.

Bipolar Junction Transistors: BJT operation, BJT Voltages and Currents, BJT amplification, Common Base, Common Emitter and Common Collector Characteristics, Numerical examples as applicable.

Module 2 [10T]

Digital Electronics: Introduction, Switching and Logic Levels, Digital Waveform

Number Systems: Decimal Number System, Binary Number System, Converting Decimal to Binary, Hexadecimal Number System: Converting Binary to Hexadecimal, Hexadecimal to Binary, Converting Hexadecimal to Decimal, Converting Decimal to Hexadecimal, Octal Numbers: Binary to Octal Conversion. Complement of Binary Numbers. Boolean Algebra Theorems, De Morgan's theorem.

Digital Circuits: Logic gates, NOT Gate, AND Gate, OR Gate, XOR Gate, NAND Gate, NOR Gate, X-NOR Gate. Algebraic Simplification, NAND and NOR Implementation: NAND Implementation, NOR Implementation.

Module 3 [8T]

Computer architecture: Computer generations, Processor and memory, memory organization (hierarchy), input output organization, Design of control unit-hardwired & micro-programmed control unit, programming the basic computer (machine, assembly, high level languages), parallel processing.

Module 4 [9T]

Introduction to IoT - Defining IoT, Characteristics of IoT, Physical design of IoT, Logical design of IoT, Functional blocks of IoT, Communication models & APIs.

IoT & M2M Machine to Machine, Difference between IoT and M2M, Software define Network.

Network & Communication aspects Wireless medium access issues, MAC protocol survey, Survey routing protocols, Sensor deployment & Node discovery, Data aggregation & dissemination.

Module 5 [9T]

Challenges in IoT - Design challenges, Development challenges, Security challenges, Other challenges.
Domain specific applications of IoT Home automation, Industry applications, Surveillance applications,
Other IoT applications.

Developing IoTs Introduction to Python, Introduction to different IoT tools, developing applications through IoT tools, developing sensor-based application through embedded system platform, Implementing IoT concepts with python.

References

1. David A. Bell, "Electronic Devices and Circuits", Oxford University Press, 5th Edition, 2008.
2. D.P. Kothari, I. J. Nagrath, "Basic Electronics", McGraw Hill Education (India) Private Limited, 2014.
3. Vijay Madiseti, Arshdeep Bahga, "Internet of Things: A Hands-On Approach"
4. Waltenegus Dargie, Christian Poellabauer, "Fundamentals of Wireless Sensor Networks: Theory and Practice"
5. Computer system architecture-M MORis Mano, Third edition.

SDC1DB03 – DATABASE MANAGEMENT SYSTEM: MONGO DB

Number of Contact Hours: 45 Hrs.

Course Outline

Module 1 [7 T]

Database System concepts and applications Introduction to databases, File Systems vs. DBMS, Advantages and Disadvantages of using DBMS Approach, Database administrators and user, Data Models, Schemas, and Instances, Types of Data Models, Three Schema Architecture and Data Independence, Database Languages and Interfaces.

Module 2 [8 T]

Entity-Relationship Model - Conceptual Data Models for Database Design Entity Relationship Models, Concept of Entity, Entity Sets, Relationship Sets, Attributes, Domains, Constraints, Keys, Strong and Weak Entities, Concepts of EER. Relational Data Model Relations, Domains and Attributes, Tuples, Keys. Integrity Rules, Relational Algebra and Operations, Relational Calculus and Domain Calculus, Relational Database Design using ER to Relational Mapping.

Module 3 [10 T]

Relational Database Design - Relational database design Anomalies in a Database, Normalization Theory, Functional Dependencies, First, Second and Third Normal Forms, Relations with more than one Candidate Key, Good and Bad Decompositions, Boyce Codd Normal Form, Multivalued Dependencies and Fourth Normal Form, Join Dependencies and Fifth Normal Form.

Module 4 [10 T]

SQL Concepts: Basics of SQL, DDL, DML, DCL, Tables – Create, Modify and Delete table structures, Rename and Drop tables, Defining constraints – Primary key, foreign key, unique, not null, check, IN operator Select Command, Logical Operators, Functions - aggregate functions, Built-in functions –numeric, date, string functions, set operations, sub-queries, correlated sub-queries, Use of group by, having, order by, join and its types, Exist, Any, All. View - Creation, Renaming the column of a view, destroys view.

Module 5 [10 T]

Transaction Management and Concurrency Control - Transaction Properties (ACID), states, Commit, Rollback; Concurrency Control Lost update problems, Locks, two phase locking. Introduction to MongoDB and its Features, Database, Collection and Documents, Various Data Types in MongoDB, Introduction to mongo shell.

References

1. Abraham Silberschatz, Henry F Korth, S.Sudharshan, Database System Concepts, 6th Edition
2. W. Gilmore, Beginning PHP and PostgreSQL 8: From Novice to Professional, Goels Computer Hut (2007), ISBN: 9788181286000
3. Karl Seguin, The little MongoDB Book
4. CJ Date, Introduction to Database Systems, Addison Wesley.

MongoDB Online References:

- i. http://mdslab.unime.it/sites/default/files/mongodb_tutorial.pdf
- ii. <https://pdfs.semanticscholar.org/presentation/066b/600837053e0172ea3e514efaaa3bd8485e46.pdf>

SDC1DB05P – PRACTICAL I: PYTHON PROGRAMMING & DBMS

Number of Contact Hours: 45 Hrs.

Course Outline

PART A: PYTHON PROGRAMMING

List of Exercises:

1. Program to demonstrate basic data type in python
2. Program to demonstrate operators in python.
3. A cashier has currency notes of denominations 10, 50, and 100.If the amount to be withdrawn is input through the keyboard using input () function in hundreds, find the total number of currency notes of each denomination the cashier will have to give to the withdrawer.
4. Program to demonstrate list and tuple in python.
5. Write a program in Python, A library charges a fine for every book returned late. For first 5 days the fine is 50 paisa, for 6-10 days fine is one rupee and above 10 days fine is 5 rupees. If you return the book after 30 days your membership will be cancelled. Write a program to accept the number of days the member is late to return the book and display the fine or the appropriate message

6. Write a program to calculate overtime pay of 10 employees. Overtime is paid at the rate of Rs.12.00 per hour for every hour worked above 40 hours. Assume that employee do not work for fractional part of an hour.
7. Two numbers are entered through the keyboard; write a program to find the value of one number raised to the power of another.
8. Write a function that receives marks received by a student in 3 subjects and returns the average and percentage of these marks. Call this function from main () and print the result in main
9. Write a program to read a file and display its contents.
10. Write a program to demonstrate database connectivity in python.
11. Write a program to compute distance between two points taking input from the user (Pythagorean Theorem)
12. Write a program add.py that takes 2 numbers as command line arguments and prints its sum.
13. Write a Program for checking whether the given number is an even number or not.
14. Using a for loop, write a program that prints out the decimal equivalents of $1/2$, $1/3$, $1/4$.
15. Write a program using a while loop that asks the user for a number, and prints a countdown from that number to zero
16. Write a Python program to print Fibonacci series.
17. Write a program to count the numbers of characters in the string and store them in a dictionary data structure.
18. Write a program to use split and join methods in the string and trace a birthday with a dictionary data structure.
19. Write a program combine lists that combines these lists into a dictionary.
20. Write a program to count frequency of characters in a given file. Can you use character frequency to tell whether the given file is a Python program file, C program file or a text file.
21. Write a program to print each line of a file in reverse order.
22. Write a program to compute the number of characters, words and lines in a file.
23. Write a function ball collide that takes two balls as parameters and computes if they are colliding. Your function should return a Boolean representing whether or not the balls are colliding.
24. Hint: Represent a ball on a plane as a tuple of (x, y, r), r being the radius If (distance between two balls centers) \leq (sum of their radii) then (they are colliding) Find mean, median, mode for the given set of numbers in a list.
25. Write a function nearly equal to test whether two strings are nearly equal. Two strings a and b are nearly equal when a can be generated by a single mutation on b.
26. Write a function dups to find all duplicates in the list.
27. Write a function unique to find all the unique elements of a list.
28. Write a function cumulative product to compute cumulative product of a list of numbers.
29. Write function to compute gcd and lcm of two numbers.
30. Write a function reverse to reverse a list. Without using the reverse function.

31. Write a program that defines a matrix and prints.
32. Write a program to perform addition of two square matrices.
33. Write a program to perform multiplication of two square matrices.
34. Write a Python Program to connect to MySQL database.
35. Write a Python Program to insert and delete records.
36. Write a Python Program to search and update records.

PART B: DBMS

List of Exercises:

1. In this session you need to create database for an Employee management system of an ABC organization. The details about different tables are given below. According to that you can proceed further and create tables using MySQL. Create the following tables with the specified constraints:

Department:

Department name - Not NULL unique

Department number - Primary Key

ManagerId - Refers to employee-id of employee table.

Manager

date of joining - Not NULL.

Employee:

First name - Not NULL

Middle initials

Last name - Not NULL

Employee id - Primary Key

Date of Birth - Not NULL

Address

Gender - M or F

Salary - Range of 5000 to 25000

Date of Joining

Department number - Refers to Department Number of Department table.

Department location:

Department number - Refers to Department number of department table.

Department location - Not NULL.

Department number & Department location are combined Primary Key

Project:

Project name-Not NULL.

Project number-Primary Key.

Project location-Not NULL.

Department number-Refers to department number of Department table.

Works-on:

Employee-id - Not NULL refers to employee-id of employee table.

Project number- Not NULL refers to Project number of Project table.

Hours - Not NULL.

Employee-id & Project number are combined primary key.

Dependent:

Employee-id - Refer to employee table employee id field

Dependent name - Gender - M or F

Date of Birth - Not NULL

Relationship - Not NULL

Now enter a few sets of meaningful data and answer the following queries.

1. List the department wise details of all the employees.
 2. Find out all those departments that are located in more than one location.
 3. Find the list of projects.
 4. Find out the list of employees working on a project.
 5. List the dependents of the employee whose employee id is 001
2. These sessions is similar to the previous one, but in this session, assume that you are developing a prototype database of the College library management system, for that you need to create the following tables:

_ Book Records

_ Book details

_ Member details and

_ Book issue details

Book Records:

Accession Number

ISBN Number

Books:

ISBN Number

Author

Publisher

Price

Members:

Member Id

Member Name

Maximum Number of books that can be issued

Maximum Number of days for which book can be issued

Book Issue:

Member Id

Accession Number

Issue Date

Return Date

You must create constraints, including referential integrity constraints, as appropriate. Please note accession number is unique for each book. A book, which has no return date, can be considered as issued book.

Enter suitable data into the tables. Now answer the following:

1. Insert data in all the three tables (use insert).
2. Insert appropriate description associated with each table and the column (use comment).
3. Display the structure of the tables.
4. Display the comments that you have inserted.
5. Using SELECT statement, write the queries for performing the following function:
 - (a) Get the list of all books (No need to find number of copies).
 - (b) Get the list of all members.
 - (c) Get the Accession number of the books which are available in the library.
 - (d) On return of a book by a member calculate the fine on that book.
 - (e) List of books issued on 01-Jan-2005.
 - (f) Get the list of all books having price greater than Rs. 500/-
 - (g) Get the list of members who did not have any book issued at any time.
 - (h) Get the list of members who have not returned the book.
 - (i) Display member ID and the list of books that have been issued to him/her from time to time.
 - (j) Find the number of copies of each book (A book accession number would be different but ISBN number would be the same).
 - (k) Find the number of copies available of a book of given ISBN number.
 - (l) Get the member ID and name of the members to whom no more books can be issued, because they have already got as many books issued as the number for which they are entitled.
3. This session is based on Lab 2 where you have created a library management system. In this session you have different query specification. You must create appropriate forms, reports, graphs, views and data filtering, use of multilevel report, etc. to answer these queries.
 1. Get the list of ISBN-Number, Book name, available copies of the books of which available copies are greater than zero.
 2. Get the list of ISBN-Number, Book name, Total copies, available copies of the book of which available copies are greater than zero. List should be displayed in alphabetical order of book name.
 3. Get the list of ISBN number, Book name, Author, total copies, cost (cost is price total copies). List should be displayed in descending order of cost.
 4. Get the list of books issued to each member.
 5. Write query to know the maximum and average price of the books.

6. Get the list of all existing members and the number of days for which a member is allowed to keep the book. Also find out the members who have got the maximum number of books issued.
7. Get the list of member codes of those members who have more than two books issued.
8. Find the details of the books presently issued to a member.
9. Create the history of issue of a book having a typical accession number.
10. To set the width of the book name as 35.

4. Create the following table and perform the necessary tasks defined below one by one. You must use the query tools/ SQL/ Reports/ Forms/Graphs/Views/ using client/server wherever needed.

1. Create the following table named customer

Column name type size

Customer id Character 10

Name Character 25

Area Character 3

Phone Numeric 7

Insert the appropriate data into table and do the following.

- Update Phone numbers of all customers to have a prefix as your city
- STD Code
- Print the entire customer table
- List the names of those customers who have e as second letter in their names.
- Find out the Customer belonging to area abc
- Delete record where area is NULL.
- Display all records in increasing order of name.
- Create a table temp from customer having customer-id, name, and area fields only
- Display area and number of records within each area (use GROUP by clause)
- Display all those records from customer table where name starts with a or area is abc.
- Display all records of those where name starts with a and phone exchange is 55.

5. Answer the following queries using Library system as created earlier. You must create a view to know member name and name of the book issued to them, use any inbuilt function and operators like IN, ANY, ALL, EXISTS.

- a. List the records of members who have not been issued any book using EXISTS operator.
- b. List the members who have got issued at least one book (use IN /ANY operator).
- c. List the books which have maximum Price using ALL operator.
- d. Display Book Name, Member Name, Issue date of Book. Create a view of this query of the currently issued books.

6. Create a table of Employee (emp-number, name, dept, salary) and Department (dept number, dept name). Insert some records in the tables through appropriate forms having integrity checks. Add some records in employee table where department value is not present in department table. Now answer the following query:

- a. Display all records from employee table where department is not found in department table.
- b. Display records from employee table in a report format with proper headings. This report must also contain those records where department number does not match with any value of department table.
- c. Display those employee records who have salary less than the salary of person whose empcode= A100.
- d. Create another table: SalesData (RegionCode, City, Salesperson-Code, SalesQty).
- e. Display records where salesperson has achieved sales more than average sales of all sales persons of all the regions.

7. Create the following tables:

Order party: (Order number, Order date, customer code)

Order: Order number, Item code, Quantity

The key to the second table is order-number + item-code

Create a form for data entry to both the tables.

8. Implement student information system

9. Implement hotel billing system

10. SQL scripts to display various reports like Result of an Examination, Salary Report, Sales Report, Sales reports grouped on Sales person or item, etc.

11. Create a STUDENT table with following fields:

Field Name Type Width

regno character 10

name character 20

paper1 numeric 3

paper2 numeric 3

paper3 numeric 3

paper4 numeric 3

paper5 numeric 3

total numeric 3

result character 6

grade character 1

Enter the regno, name and marks in 5 papers of at least 10 students. Write a program to process the records to update the table with values for the fields total (paper1+paper2+paper3+paper4+paper5), result('passed' if total is greater than or equal to 50% of the total; 'failed' otherwise), and grade ('A' if mark obtained is greater than or equal to 90% of the total mark, 'B' if mark obtained is greater than or equal to 75% of the total mark, 'C' if mark obtained is greater than or equal to 60% of the total mark, 'D' if mark obtained is greater than or equal to 50% of the total mark, and 'F' if mark obtained is less than 50% of the total mark).

Display a report in descending order of the total mark, showing the data entered into the table along with the total marks, result and grade.

12. Generate a Hospital information system that can generate the following reports:

- Patients who belongs to in-patient category

- Patients who belongs to out-patient category

For this purpose, create a table named HOSPITAL having the following structure.

Field Name Type Width

patientid character 10
 name character 20
 age numeric 3
 doctor character 20
 patienttype character 15
 consultcharge numeric 6
 bloodtestcharge numeric 6
 xraycharge numeric 6
 othercharges numeric 6
 totalamount numeric 6

Enter the records of at least 10 patients. Write a SQL program to display the report in the ascending order of patient name.

13. Create A Table EMP And DEPT Using the Following Information.

a. DEPT:

COLUMN NAME DATATYPE (SIZE)

-----	-----
DEPTNO	NUMBER (2)
DNAME	VARCHAR2 (14)
LOC	VARCHAR2 (14)

b. EMP:

COLUMN NAME DATATYPE(SIZE)

-----	-----
EMPNO	NUMBER (4)
ENAME	VARCHAR2 (10)
JOB	VARCHAR2 (9)
MGR	NUMBER (4)
HIREDATE	DATE
SAL	NUMBER (7, 2)
COMM	NUMBER (7, 2)
DEPTNO	NUMBER (2)

a) Check the Default Size of a Number, Char and Date Data types.

b) Describe the Structure of the Table and EMP Table.

14. Design a Hotel Bill calculating system that generates hotel bills for thecustomers.

15. Design a Hostel Accounting system that generates the Hostel Due Report.

16. Design an Electricity Bill Report generating system that generates electricity bills details of customers for a month.

17. Generate a Library Information System that generates report of the books available in the library.

18. Programs involving multiple tables.

19. Consider the schema for College Database:

STUDENT (USN, SName, Address, Phone, Gender)

SEMSEC (SSID, Sem, Sec)

CLASS (USN, SSID)

SUBJECT (Subcode, Title, Sem, Credits)

IAMARKS (USN, Subcode, SSID, Test1, Test2, Test3, FinalIA)

Write SQL queries to

1. List all the student details studying in fourth semester 'C' section.
2. Compute the total number of male and female students in each semester and in each section.
3. Create a view of Test1 marks of student USN '1BI15CS101' in all subjects.
4. Calculate the FinalIA (average of best two test marks) and update the corresponding table for all students.

5. Categorize students based on the following criterion:

If FinalIA = 17 to 20 then CAT = 'Outstanding'

If FinalIA = 12 to 16 then CAT = 'Average'

If FinalIA < 12 then CAT = 'Weak'

Give these details only for 8th semester A, B, and C section students.

20. Consider the schema for Company Database:

EMPLOYEE (SSN, Name, Address, Sex, Salary, SuperSSN, DNo)

DEPARTMENT (DNo, DName, MgrSSN, MgrStartDate)

DLOCATION (DNo, DLoc)

PROJECT (PNo, PName, PLocation, DNo)

WORKS_ON (SSN, PNo, Hours)

Write SQL queries to

1. Make a list of all project numbers for projects that involve an employee whose last name is 'Scott', either as a worker or as a manager of the department that controls the project.
2. Show the resulting salaries if every employee working on the 'IoT' project is given a 10 percent raise.
3. Find the sum of the salaries of all employees of the 'Accounts' department, as well as the maximum salary, the minimum salary, and the average salary in this department
4. Retrieve the name of each employee who works on all the projects controlled by department number 5 (use NOT EXISTS operator). For each department that has more than five employees, retrieve the department number and the number of its employees who are making more than Rs. 6,00,000.

SEMESTER II

SDC2ME06 – MICROSOFT EXCEL FOR VBA & SPSS

Number of Contact Hours: 60 Hrs.

Course Outline

Module 1 [10 T]

Introduction to Spreadsheets, Understanding Microsoft Excel, Workbook, Cells, Range in Excel, Selecting Cells, Auto Fill, Cell Referencing- Using Absolute and Relative References, creating new worksheet, Formulas and **Functions**---Mathematical Functions: SUM, PRODUCT, SQRT, ROMAN and ROUND -Statistical Functions: AVERAGE, MEDIAN, MODE, STDEV, CORREL and FORECAST ,Financial Functions: DB,SLN, SYD, PMT, NPER, IPMT, PMT,(Annuity Functions, Investment analysis functions, Bond Functions, Depreciation Functions)PV, NPV, XNPV, IRR, MIRR and XIRR Database Functions: DMAX, DMIN,DAVERAGE,DCOUNT and DSUM

Defining Names in Excel, Sorting Data, Using Excel Tables, Filtering Data in Excel, Understand Charts, Chart Design Options, Sensitivity Analysis using Excel-Scenario Manager, Other Sensitivity Analysis Features, Goal Seek, Data Tables, What-If Analysis, Introduction to functions such as the IF, nested IF, VLOOKUP and HLOOKUP in Excel.

Module 2 [10 T]

Quick Analysis--Quick Analysis with TOTALS, Sum, Average, Count, %Total, Running Total, Sum of Columns, Sorting, Filtering. Subtotals with Ranges-Subtotals, Nested Subtotals, Introduction to charts in Excel. Constructing various Line, Bar and Pie charts. Using the Pivot chart features of Excel, Understanding and constructing Histograms and Scatterplots. Importing Data into Excel .VBA-OVERVIEW,VBA-EXCEL,VBA-EXCEL TERMS-Modules, Procedure , VBA – MACRO COMMENTS VBA-MESSAGEBOX, VBA-INPUTBOX, VBA-VARIABLES, Data Types VBA – CONSTANTS, VBA – OPERATORS-- The Arithmetic Operators, The Comparison Operators ,The Logical Operators, The Concatenation Operators VBA , VBA – DECISIONS- If Statement, If Else Statement, If Elseif - Else statement, Nested If Statement, Switch Statement,

Module 3 [15 T]

VBA – LOOPS , For Loop, For Each Loops , While Wend Loops, Do While Loops, Do Until Loops, Loop Control Statements, Exit For, Exit Do, VBA – STRINGS -- Instr , InString Reverse ,LCase, UCase, Left, Right ,Mid , Ltrim , Rtrim, Trim , Len , VBA-DATE-TIME FUNCTION ,VBA-ARRAYS , VBA – USER-DEFINED FUNCTIONS - Function Definition Calling a Function, VBA – SUB PROCEDURE- Calling Procedures, VBA- EVENTS- Worksheet Events ,Workbook Events, Worksheet Events.

Module 4 [10 T]

Introduction to SPSS - Data analysis with SPSS: general aspects, workflow, critical issues - SPSS: general description, functions, menus, commands - SPSS file management Input and data cleaning - Defining

variables - Manual input of data - Automated input of data and file import

Module 5 [15 T]

Data manipulation - Data Transformation - Syntax files and scripts - Output management Descriptive analysis of data - Frequencies - Descriptives - Explore - Crosstabs, Charts, Statistical tests - Means - T-test - One-way ANOVA - Non parametric tests - Normality tests Correlation and regression - Linear correlation and regression - Multiple regression (linear) Multivariate analysis - Factor analysis - Cluster analysis

References

1. Field A., Discovering Statistics Using SPSS, Fourth Edition, SAGE, 2013
2. Sabine Landau and Brian S. Everitt, A handbook of statistical analyses using spss
3. Excel 2010 power programming with VBA-John Walkenbach
4. Excel 2007 with VBA-John Green, Rob Bovey

SDC2DS07 – INTRODUCTION TO DATA SCIENCE

Number of Contact Hours: 45 Hrs.

Course Outline

Module 1 [8 T]

Introduction: What is Data Science, Big Data and Data Science hype and getting past the hype Why now. Datafication, Current landscape of perspectives, Skill sets needed, Statistical Inference, Populations and samples, Statistical modeling, probability distributions, Intro to R

Module 2 [8 T]

Exploratory Data Analysis and the Data Science Process, Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA, The Data Science Process Case Study: Real Direct (online real estate _rm), Three Basic Machine Learning Algorithms, Linear Regression, k-Nearest Neighbors (k-NN), k-means

Module 3 [10 T]

One More Machine Learning Algorithm and Usage in Applications, Motivating application: Filtering Spam, Why Linear Regression and k-NN are poor choices for Filtering Spam, Naive Bayes and why it works for Filtering Spam, Data Wrangling: APIs and other tools for scrapping the Web, Feature Generation and Feature Selection (Extracting Meaning From Data),

Motivating application: user (customer) retention, Feature Generation (brainstorming, role of domain expertise, and place for imagination), Feature Selection algorithms, Filters; Wrappers; Decision Trees; Random Forests

Module 4 [9 T]

Recommendation Systems: Building a User-Facing Data Product, Algorithmic ingredients of a Recommendation Engine, Dimensionality Reduction, Singular Value Decomposition, Principal Component Analysis, Exercise: build your own recommendation system, Mining Social-Network Graphs,

Social networks as graphs, Clustering of graphs, Direct discovery of communities in graphs, Partitioning of graphs, Neighborhood properties in graphs

Module 5 [10 T]

Data Visualization, Basic principles, ideas and tools for data visualization, Examples of inspiring (industry) projects Exercise: create your own visualization of a complex dataset, Data Science and Ethical Issues, Discussions on privacy, security, ethics, A look back at Data Science, Next-generation data scientists

References

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly. 2014.
2. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2014. (free online)
3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. ISBN 0262018020. 2013. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. ISBN 1449361323. 2013.
4. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. ISBN 0387952845. 2009. (free online)
5. Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science. Mohammed J. Zaki and Wagner Miera Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press. 2014.
6. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, Third Edition. ISBN 0123814790. 2011.

SDC2DS08P – PRACTICAL II: MICROSOFT EXCEL FOR DATA SCIENCE AND SPSS

Number of Credits: 7

Course Outline

PART A: MICROSOFT EXCEL

PART B: SPSS

List of Exercises: PART A

1. Using the data given, get the sum of all the figures within the range.

	A	B	C	D	E	F	G
1		Mon	Tue	Wed	Thu	Fri	TOTAL
2	Breakfast	3,560	3,186	2,952	3,395	3,436	
3	Lunch	20,163	21,416	19,912	19,681	18,628	

4	Bar	9,873	12,172	12,642	12,711	18,846	
5	Snacks	2,405	3,544	2,694	3,120	3,712	
6	TOTALS						

2. Enter the data given below into a worksheet.

	A	B	C	D	E
1	Stationery Supplies Ltd				
2					
3	Date	Sales Person	Item	Receipt No	Amount
4	21-Nov	Carl	Toys	1238	1,782.10
5	26-Nov	Carl	Stationery	1255	4,853.55
6	26-Nov	Carl	Toys	1395	51.35
7				Carl's Total	
8	21-Nov	John	Cards	1141	91.15
9	24-Nov	John	Books	1982	442.60
10	21-Nov	John	Toys	1885	561.50
11	26-Nov	John	Toys	1875	62.75
12				John's Total	
13	22-Nov	Judy	Books	1032	234.50
14	26-Nov	Judy	Sports goods	1920	472.60
15				Judy's Total	
16	25-Nov	Mary	Toys	1774	364.15
17				Mary's Total	
18	22-Nov	Susan	Electronics	1160	52.95
19	23-Nov	Susan	Cards	1075	81.60
20	23-Nov	Susan	Others	1745	132.95
21	24-Nov	Susan	Sports goods	1662	2,580.10
22				Susan's Total	
23					
24				Grand Total	

- i. Calculate the totals for each salesperson and get the grand total.:
- ii. Format the worksheet as follows:
Make all the Totals bold, two decimal places, comma, center the title across columns A-E and make it size 16, bold and Italic.
- iii. Put a double border round the whole table and a single line border inside the table.
- iv. Save the worksheet as **Stationery Analysis**.

3. Using the information given in the table below, calculate the total amount payable by the company to the employees.

	A	B	C	D	E
1	Services Company Ltd				
2	Overtime Details				
3	Date	Name	Hours Worked	Rate	Amount
4	26-Nov	Kennedy	5	70	350.00
5	26-Nov	Kennedy	5	100	500.00
6	26-Nov	Mary	5	100	500.00
7	26-Nov	Lewis	4	100	400.00
8	30-Nov	Judy	3	100	300.00
9	30-Nov	Kennedy	6	70	420.00
10	30-Nov	Lewis	5	100	500.00
11	30-Nov	Kennedy	4	70	280.00
12	30-Nov	Judy	5	100	500.00
13	30-Nov	Lewis	5	100	500.00
14	02-Dec	Judy	4	70	280.00
15			Total Amount		

4. A Payroll consists of Basic Pay, Allowances, Gross Salary, Deductions and Net Salary. The Allowances are 23% of the Basic Pay while the Deductions are 12% of the Gross Salary. In the given worksheet, indicate in each cell what will be inserted, that is – a value or a formula. In the case of a formula, write down the formula in the cell.

	A	B	C	D	E	F
1	Stationery Supplies Ltd					
2						
3	Name	Basic Pay	Allowances	Gross Salary	Deductions	Net Salary

4	Lewis					
5	Francis					
6	Edwin					
.						.
.						.
.						.
13	Totals					

5. Assume you are the Accountant of Stationery Supplies Ltd. Below is the current payroll in the workbook OLD PAYROLL.

OLD PAYROLL.

	A	B	C	D	E	F
1	Stationery Supplies Ltd.					
2						
3	Name	Basic Pay	Allowances	Gross Salary	Deductions	Net Salary
4	Lewis	15,791	3,137	18,928	1,256	17,672
5	Francis	15,537	3,061	18,598	776	17,822
6	Edwin	15,506	3,051	18,557	999	17,558
7	Bernard	15,417	3,025	18,442	1,099	17,343
8	George	15,008	2,902	17,910	718	17,192
9	Albert	14,969	2,890	17,859	846	17,013
10	Edward	14,651	2,795	17,446	760	16,686
11	Cornell	14,618	2,785	17,403	663	16,740
12	John	14,553	2,765	17,318	558	16,760
13	Carl	14,508	2,752	17,260	706	16,554
14						
15	Totals	150,558	29,163	179,721	8,381	171,340

The following salary review is given to you in the workbook INCREMENT.

INCREMENT

	A	B	C
1	Name	Current Pay	% increase

2	Lewis	15,791	19%
3	Francis	15,537	19%
4	Edwin	15,506	22%
5	Bernard	15,417	18%
6	George	15,008	21%
7	Albert	14,969	17%
8	Edward	14,651	15%
9	Cornell	14,618	25%
10	John	14,553	19%
11	Carl	14,508	20%
12			
13	<u>Allowances</u>		
14	20%		

Required:

Using formulas, you are required to update the payroll with the changes in a blank worksheet. This new blank sheet is in the workbook NEW PAYROLL.

6. You are in charge of a young and growing business. You have identified the various factors (sources of revenue and expenses) that influence the business as shown in the table below. Use the figures provided and the layout to create a financial projection model for the business for the next six years. The parameters are given on Sheet 2.

SHEET 1:

INCOME AND EXPENSES PROJECTIONS						
	1999	2000	2001	2002	2003	2004
Sales	10,000					
% Growth over the previous year		20%	30%	20%	10%	10%
Materials						
Wages						
Other benefits						
Others						

Total Cost of Goods Sold						
Salary: Office						
Salary: Sales						
Other Benefits						
Advertising & Promotions						
Depreciation						
Miscellaneous						
Total General & Admin. Expenses						
Total Operating Costs						
Interest on Loans						
Pre-tax Income						
Tax						
Profit						

SHEET 2:

Parameters		Description
Sales	10,000	Starts at 10,000 and grows by a percentage
Materials	17%	17% of Sales
Wages	14%	14% of Sales

Other benefits	2.1%	2.1% of Sales
Others	8%	Starts at 100, then grows by 8% yearly
Salary: Office	10%	Starts at 1,000, then grows by 10% annually
Salary: Sales	8%	8% of Sales
Other Benefits	17%	17% of Total Salary
Advertising & Promotions	2.5%	2.5% of Sales
Depreciation	20	Fixed at 20 every year
Miscellaneous	10	Starts at 10 and grows by a fixed amount of 10 annually
Interest on Loans	10	A fixed amount of 10 each year
Tax	52%	52% of Pre-tax Income

Exercise Instructions.

- i. Open the worksheet named **Income and Expenses Projections.xls**.
 - ii. Rename Sheet1 as **Projections** while Sheet 2 should now be **Parameters**.
 - iii. Calculate the Sales for the year 2000 using the percentage given in cell **C5**.
 - iv. Copy the formula across to the Year 2004.
 - v. Calculate the different items that make up the Total Operating Costs using the parameters in the Parameters sheet.
(You should enter the formula for the Year 1999 and copy down to the year 2004. Use Absolute Referencing effectively).
Hint: $Total\ Cost\ of\ Goods\ Sold = Materials + Wages + Other\ Benefits + Others$
 - vi. Calculate the Total Operating Costs:
 $Total\ Cost\ of\ Goods\ Sold + Total\ General\ and\ Administrative\ Expenses$.
 - vii. Calculate the Interest on Loans:
 - viii. Calculate the Pre-tax Income.
 $Sales - Total\ Operating\ Cost - Interest\ on\ Loans$.
 - ix. Calculate the Tax.
 - x. Calculate the Profit:
 $Pre-tax\ Income - Tax$.
 - xi. Format the worksheet as follows:
Make all the Totals bold, zero decimal places, comma, center the heading between A1:G1 and make it size 16, bold.
 - xii. Save the file as **C:\Exams\Creative.xls**
7. From the data given in the table below, create a Pie Chart to show the distribution of the total amount amongst the various salesmen.

	A	B	C	D	E	F
1	ABC Company Sales Performance Report					
2						
3	Salesman	Qtr1	Qtr2	Qtr3	Qtr4	Total
4	Albert	148	156	171	140	615

5	Carl	122	131	153	118	524
6	Cornell	211	243	246	250	950
7	Edwin	129	150	92	218	589
8	Francis	311	270	247	322	1,150

8.

	A	B	C	D	E	F	G	H	I
1	XYZ COMPANY SALES PERFORMANCE								
2									
3									
4	SALES PERSON	ANNUAL TARGET	QTR1	QTR2	QTR 3	QTR 4	TOTAL SALES	AVERAGE SALES	COMM ISSION
5									
6	ALBERT	750	148	256	133	154	X		
7	MICHAEL	650	187	143	258	143	X		
8	CARL	800	233	200	216	152	X		
9	GEORGE	700	256	145	136	259	X		
10	LUCY	1,000	249	212	215	124	X		
11									
12	TOTAL	X	X	X	X	X	X		
13									
14	COMMIS SION	6%							
15									

- Given the table above, write formulas or describe how you would calculate the Total Sales in column G and Row 12.
- How would you calculate the Average Sales. Write the formula as it should appear in Excel and show the method of duplicating it to the other cells.
- How would you calculate the Commission. Write the formula as it would appear in Excel and explain the method of duplicating it to the other cells.
- Explain how you can insert two rows above Row 10.
- What does this sign (#####) mean when seen in cell(s). What should you do when you see this sign.
- How does one delete an entire row or column and all its cells.

- vii. You are asked to compare QTR1 and QTR3 sales for all salespersons in the above table using a chart:
- What range of cells do you need to select and how would you select it.
 - What type of graph would you use.
- viii. How can you edit a chart once it is created, lets say, to change the series from columns to rows, to change the legend, etc.

9. Create the worksheet below from scratch.

	A	B	C	D	E	F
1	TOWN / DAY	Mon	Tue	Wed	Thu	Fri
2	Mombasa	30	29.5	31	28.5	32
3	Kisumu	31	33	30	30	32
4	Nakuru					
5	Nairobi	24	23.5	22	23	24.5
6						
7						
8	TOWN / DAY	Mon	Tue	Wed	Thu	Fri
9	Mombasa					
10	Kisumu					
11	Nakuru					
12	Nairobi					

Questions

- Use a formula to calculate the temperature for Nakuru, given that the temperature for Nakuru is $\frac{3}{4}$ that of Mombasa.
- Create two blank rows below Row 1.
- Type "**Temp. in degrees Celsius**" in Cell A2 and "**Temp. in degrees Fahrenheit**" in Cell A9.
- Calculate the temperature in F using the conversion factor given as $F=(C+19)*9/5$.
- Insert a blank row before Nakuru and enter the row heading Kericho in both temperature versions.
- Calculate the temperature for Kericho in both versions given that it is $\frac{4}{5}$ that of Kisumu.
- Type "**Average Temp. (C)**" in Cell G3 and "**Average Temp. (F)**" in Cell G11 respectively. Calculate the average temperatures for all the towns.
- Format all cells containing the temp. values to zero decimal places with no commas.
- Format all cells containing Average Temp. to Arial Black, Italics, size 12.
- Centre the Titles across columns A and G.
- Put a double border round the whole table and a single line border inside the table. Shade the column for Average temperature gray.
- Use the Average values (C) in column G to create a 3-D Exploded Pie Chart to show distribution of temperature for the towns.

The title should be “**Average Temp. (C)**”. Use the text in column A as the legend. In the data labels, select **Show Value**.

- xiii. Move the chart created above to Sheet3. Do not insert it as an object.
- xiv. Move the left top corner of the chart in Sheet3 to cell A7. Resize the chart to fit into the range A7:h20.
- xv. Save the worksheet as **Weather**.

10. Use the worksheet given to answer the questions that follow:

Expenses for the Month of January vs. Budget		
	Budget	Savings
Salaries and Wages	156675.00	
Rent	4300.00	
Electricity	1000.00	
Telephone	200.00	
Advertisements	20000.00	
Freight and clearing	15650.00	
Security	3800.00	

Questions

- i. Insert a new column between Budget and Savings column.
- ii. Enter the title ‘Actual’ in cell C3.
- iii. Enter the following figures in the new column.

	Actual
Salaries and Wages	145200
Rent	4300
Electricity	1207
Telephone	142
Advertisements	18550
Freight and clearing	13400
Security	3800

- iv. Calculate the savings in cells D4:D10.

- iii. Calculate the amount of tax deducted from each employee, given that the tax rate is 12% of the gross pay. Tax rate is found in cell A16 of the worksheet.
- iv. Format the text orientation in the range **A4:G4** to 0 degrees.
- v. Adjust the column width such that all the headings are visible.
- vi. In cell **H4**, enter the title **Net Pay** and calculate the Net pay for all employees.
- vii. Format the range **G4:H13** to **zero (0)** decimal places.
- viii. Format the title **Net Pay** to match the other titles.
- ix. Set the range A1 to H16 as **Print Area**.
- x. Using the Names in column B, Basic Pay in column D, and Allowances in column E, insert a **Clustered Column Pie Chart** on the same sheet to show comparison of the salaries for the employees. The Chart Title should be **Employee Details**, the Y-axis should be **Employee Names** and the X-axis should be **Thousands (Kshs)**.
- xi. Move the chart so that the top left corner is on cell A18.
- xii. Change the Chart Title to **Employees' Salary Details**.
- xiii. Change the Chart Type to **Clustered bar with a 3-D visual effect**.
- xiv. Increase the Chart Title **Font size to 14**.
- xv. Change the text direction for the title of the X-axis to 0 degrees, and for the Y-axis to -90 degrees.
- xvi. Resize the chart such that the bottom left corner is on cell A55, while the bottom right corner is on cell I55 so that all the details are clearly visible.
- xvii. Save the worksheet as **Salary Details (Reviewed Copy)**.

12. Use the data given in the table below:

SHEET 1:

Brick Court Investments					
Overtime calculations for the month of November					
Date	Name	Hours Worked	Day of Week	Rate	Amount
27-Oct	Albert	4	Sun	100	400
8-Nov	Albert	3	Fri	50	150
18-Nov	Lewis	4	Mon	50	200
9-Nov	Albert	4	Sat	50	200
12-Nov	Albert	3	Tue	50	150
27-Oct	Carl	5	Sun	100	500
15-Nov	Albert	4	Fri	50	200
19-Nov	Albert	1	Tue	50	50
19-Nov	Albert	4	Tue	50	200
5-Nov	Bernard	4	Tue	50	200
7-Nov	Bernard	5	Thu	50	250
20-Nov	Albert	2	Wed	50	100
21-Nov	Albert	4	Thu	50	200
28-Oct	Carl	3	Mon	50	150
11-Nov	Lewis	4	Mon	50	200
22-Nov	Lewis	2	Fri	50	100
26-Oct	Bernard	2	Sat	50	100

28-Oct	Bernard	4	Mon	50	200
30-Oct	Bernard	3	Wed	50	150
2-Nov	Bernard	1	Sat	50	50
9-Nov	Bernard	4	Sat	50	200
14-Nov	Bernard	1	Thu	50	50

SHEET 2:

Date	Name	Hours Worked	Day of Week	Rate	Amount
			Average		
			Maximum		

Questions

- i. Find the name **Lewis** and replace it with **Martin**.
- ii. Sort the data according to name in **Ascending** order.
- iii. Insert a row at each change of name.
- iv. Type the heading '**Average**' in cell **E13**.
- v. Adjust the width of column E to view the contents in full.
- vi. Copy the heading in E13 to cells: E22, E38, E50, E61, E74, E87, E98 and E109.
- vii. Use the **AVERAGE** function to calculate the average sales of all salesmen.
- viii. Insert four blank rows above row 1.
- ix. Insert a picture (Logo) into the worksheet.
- x. Resize the picture to Height 0.62 and Width 3.76 inches.
- xi. Position the picture such that the bottom left corner of the picture is within cell A4.
- xii. Cut Bernard's details (A18:F26) and paste them in Sheet 2, starting from Cell A2.
- xiii. In cell E11 type the title 'Maximum'.
- xiv. Use a function to extract the maximum amount that Bernard has ever earned.
- xv. Save the file as **Overtime analysis**.

13. The following is a simple payroll:

	A	B	C	D	E	F	G	H	I
1	Name	Hours Worked	Hourly Rate	Basic Pay	Gross Pay	Tax Deductions	NSSF Contributions	Allowances	Net Pay
2	John	8	200						
3	Peter	12	450						
4	Sam	22	300						
5	Njogu	30	286						
6	Mary	16	220						
7	Sally	45	468						
8	Jane	15	150						
9	Tina	3	280						
10									
11									

Required:

Write formulae using cell names for the following expressions. State where the formula is placed.

- Basic Pay = Hours Worked * Hourly Rate.
- Allowances are allocated at 10% of the Basic Pay.
- Gross Pay = Basic Pay + Allowances.
- Tax Deduction is calculated at 20% of the Gross Pay.
- Net Pay = Gross Pay – Tax Deductions.

14. From the data given in the table below, create a Pie Chart to show the distribution of the total amount amongst the various salesmen.

	A	B	C	D	E	F
1	ABC Company Sales Performance Report					
2						
3	Salesman	Qtr1	Qtr2	Qtr3	Qtr4	Total
4	Albert	148	156	171	140	615
5	Carl	122	131	153	118	524
6	Cornell	211	243	246	250	950
7	Edwin	129	150	92	218	589
8	Francis	311	270	247	322	1,150

15.

<u>Group Name</u>
Buttersworth Co
Marleyville Tea
Marleyville Tea
Marleyville Tea
Postcards Store
Buttersworth Co
Postcards Store
Buttersworth Co
Buttersworth Co
Examineers Inc
Postcards Store
Buttersworth Co
FASST Trax
Hehehe
Randles Roadshack
Hehehe
Marleyville Tea
FASST Trax
Randles Roadshack
Examineers Inc
Randles Roadshack
Creepy Lada

*You have a list of repeating group names.

*Using a pivot table, count how many times each group name appears in the list.

*Keep pivot table in this tab.

Notice that this pivot table is a list comprised of each group name only occurring once.

Then, within the pivot table, put the list in decreasing order based on the number of times the name appears in the list.

16.

<u>Group Segment</u>	<u>Revenue</u>	<u>Expenses</u>
Medium	8249.328	594.8782
Medium	6557.873	83.33939

Small	6180.036	551.2668
Medium	9915.454	336.8577
Small	4047.669	481.0083
Small	4797.088	962.3551
Small	6537.8	931.9697
Medium	3151.423	848.2135
Big	9504.782	827.5632
Medium	9378.14	566.7977
Big	6772.133	436.88
Small	1854.632	930.5483
Big	8303.943	680.7504
Small	6741.46	418.0317
Small	4151.44	336.4056
Small	7972.302	917.1406
Big	10331.58	440.4744
Small	3695.193	217.6563
Small	7904.294	764.5839
Medium	849.2599	94.51704
Big	7157.628	915.9245
Medium	3712.565	432.5486

There are preset 3 segments that all groups are put into:

Big

Medium

Tiny

Using a pivot table, create a table displaying Revenue for each segment.

Keep pivot table in this tab.

Large \$ amounts (over \$100 or so) are typically shown with zero decimals.

17.

<u>Group Segment</u>	<u>Revenue</u>	<u>Expenses</u>	<u>Member Months</u>
Medium	8249.328	594.8782	13
Medium	6557.873	83.33939	9
Small	6180.036	551.2668	51
Medium	9915.454	336.8577	97
Small	4047.669	481.0083	92
Small	4797.088	962.3551	63
Small	6537.8	931.9697	43
Medium	3151.423	848.2135	64
Big	9504.782	827.5632	28

Medium	9378.14	566.7977	41
Big	6772.133	436.88	75
Small	1854.632	930.5483	21
Big	8303.943	680.7504	13
Small	6741.46	418.0317	96
Small	4151.44	336.4056	20
Small	7972.302	917.1406	43
Big	10331.58	440.4744	94
Small	3695.193	217.6563	98
Small	7904.294	764.5839	13
Medium	849.2599	94.51704	57
Big	7157.628	915.9245	20
Medium	3712.565	432.5486	4
Small	7596.44	134.8226	63
Small	2198.658	1037.294	13

Make a pivot table showing each segments revenue and expenses. Make sure to use both row headers and column headers. Have "Sum of Member Months" be the left most column.

Create a column adjacent to the pivot table that represents the Revenue and Expenses PMPM's.

PMPM's are typically represented with dollar signs (\$) and to two decimal places. Keep the table in this tab

18.

<u>Subdivision</u>	<u>Group Segment</u>	<u>Revenue</u>	<u>Expenses</u>
Bigger	Medium	8249.328	594.8782
Bigger	Medium	6557.873	83.33939
Pretty Small	Small	6180.036	551.2668
Bigger	Medium	9915.454	336.8577
Very Small	Small	4047.669	481.0083
Pretty Small	Small	4797.088	962.3551
Pretty Small	Small	6537.8	931.9697
Not as Big	Medium	3151.423	848.2135
Not Real Big	Big	9504.782	827.5632
Not as Big	Medium	9378.14	566.7977
Not Real Big	Big	6772.133	436.88
Pretty Small	Small	1854.632	930.5483
More than 500	Big	8303.943	680.7504
Pretty Small	Small	6741.46	418.0317

Very Small	Small	4151.44	336.4056
Pretty Small	Small	7972.302	917.1406
Not Real Big	Big	10331.58	440.4744
Pretty Small	Small	3695.193	217.6563
Pretty Small	Small	7904.294	764.5839
Bigger	Medium	849.2599	94.51704
More than 500	Big	7157.628	915.9245
Bigger	Medium	3712.565	432.5486
Very Small	Small	7596.44	134.8226
Very Small	Small	2198.658	1037.294

Create a column entitled "Net Income/Loss." This column will represent Revenue minus Expenses.

Make a pivot table representing Each Segment and Subdivision by Revenue, Expenses, and Net Income. Use row headers and column headers.

Keep pivot table in this tab

19. How To Access The Excel Range And Show The Value Using Message Box
20. How To Enter Data into a Cell
21. How To Change The Background Color Of A Particular Range
22. How To Change The Font Color And Font Size Of A Particular Range
23. How To Change The Text To Upper Case Or Lower Case
24. How To Copy Data From One Range To Another Range
25. How To Select And Activate Worksheet
26. How To Get The Active Sheet Name And Workbook Name
27. How To Add New Worksheet And Rename A Worksheet and Delete Worksheet
28. How To Create New Workbook, Add Data, Save And Close The Workbook
29. How To Hide And Unhide Rows And Columns
30. How To Insert And Delete Rows And Columns
31. How To Set The Row Height And Column Width
32. How To Merge and UnMerge Cells
33. How To Compare Two Values – A Simple Example On If Condition
34. How To Print 1000 Values – A Simple Example On For Loop

List of Exercises: PART B

1. 8 Children selected at random from an elementary school were rated on a 7- point scale with regard to some of their personality traits. The rating work was first done by their family members and then by three teachers. Can you conclude that no significant difference exists between these two different sets of rating?

CHILDREN	A	B	C	D	E	F	G	H
RATING OF SCORES OF FAMILY MEMBERS	6	18	14	10	20	17	18	8
RATING OF SCORES OF TEACHERS	3	15	16	12	13	11	8	9

2. For the following data perform median test.

Group I	6	6	14	13	15	6	8	7	10	14	10	14
Group 2	7	2	12	13	8	6	4	2	2	12		

3. There are two independent Samples A and B. Test whether that both samples have been drawn from the same population using Mann- Klitney U Test

Sample A	Sample B
8	9
6	7
10	11
5	8
	12

4. On a railway reservation window, there was long queue of Men (M) and women (W) standing in the order in which they have come as depicted in the following displayed matter. Confirm whether or not they were standing in a random order.

M	W	M	W	M	M	M	W	M	W	M	W	W	M	M	M	M	W	M	W	M
W	M	M	W	W	W	M	W	M	W	M	W	M	M	M	W	M	M	W	M	M
M	M	W	M	W	M	M														

5. Ten students have obtained the following scores on tests in history and Hindi. find the coefficient of correlation of and also Express these scores through scatter diagram.

Individuals	A	B	C	D	E	F	G	H	I	J
Scores in History(X)	13	12	10	8	7	6	6	4	3	1
Scores in Hindi(Y)	7	11	3	7	2	12	6	2	9	6

6. A Group of five students obtained the following scores on two Achievement test X & Y

Student	A	B	C	D	E
Scores in X test	10	11	12	9	8
Scores in Y test	12	18	20	10	10

- Determine both the regression equation
- If a student scores 15 in test X , predict his probable scores in test y.

7. Find the multiple regression equation for the following data.

Y	10	17	48	27	55	26	9	16
X1	2	3	4	1	5	6	7	8
X2	1	2	5	2	6	4	3	4

8. Two groups of 10 students each got the following scores on an attitude Scales:

Groups I	10	9	8	7	7	8	6	5	6	4
Groups II	9	8	6	7	8	8	11	12	6	5

Test the significance of the difference of means.

9. Apply K.S Test to the following data(N>20)

Scores	≥ 20	15-19	10-14	5-9	≤ 4
f	6	14	55	20	5

10. Two groups of students were rated on a 9-point scale for their habit of cleanliness. Do the 2 groups differ significantly in terms of their habit of cleanliness.

Scores on Scale	9	8	7	6	5	4	3	2	1
Group-A (f)	2	6	5	4	6	8	7	9	3
Group-B	4	11	8	5	7	10	6	4	5

11. The aim of an experimental study was to determine the effect of three different techniques of training on the learning of a particular skill. 3 groups, each consisting of seven students of a class IX, assigned randomly. We give training through different techniques. The scores obtained on a performance test we recorded as follows perform ANOVA.

Group-I	3	5	3	1	7	3	6
Group-II	4	5	3	4	9	5	5
Group-III	5	5	5	1	7	3	7

1. Use sign test to see if there is a difference b/w the no. of days until collection of an account receivable before & after a new collection policy.

Before	30	28	34	35	40	42	33	38	34	45	28	27	25	41	36
--------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

After	32	29	33	32	37	43	40	41	37	44	27	33	30	38	36
-------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

13. By using Kruskal-Wallis test whether the 3 methods are equally effective, in an inspection procedure

Method A : 80,83,79,85,90,68

Method B : 82,84,60,72,86,67,91

Method C : 93,65,77,78,88

14. Two random samples were drawn from two normal population and their values are:

A: 66 67 75 76 82 84 88 90 92

B: 64 66 74 78 82 85 87 92 93 95 97

Test whether the two populations have the same variance.

15. Find Mean

Marks	15-25	25-35	35-45	45-55	55-65	65-75	75-85
No of Students	3	5	12	15	9	9	7

16. Find Median

Marks	10-20	20-30	30-40	40-50	50-60	60-70
No of students	13	15	21	20	19	12

17. Find Mode

Size	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	10	14	16	14	11	13	17

18. Find Range

size	10-20	20-30	30-40	40-50	50-60
Frequency	5	10	12	8	4

19. Find Q.D

Size	5	8	10	12	19	20	22
F	3	10	15	20	8	7	6

20. Find M.D

class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	18	16	15	12	10	5	2	2

21. Find S.D

Size	2	3	4	5	6	7	8
f	1	2	3	5	3	2	1

22. Calculate the Karl Pearson's coefficient of Skewness.

x	12.5	17.5	22.5	27.5	32.5	37.5	42.5	47.5
y	28	42	54	108	129	61	45	33

23. Compute Coefficient of Kurtosis for the following data.

X	4.5	14.5	24.5	34.5	44.5	54.5	64.5	74.5	84.5	94.5
Y	1	5	12	22	17	9	4	3	1	1

24. Below are given the gain in weights (in Kg) of pigs fed on two diets A & B

Diet A	25	32	30	34	24	14	32	24	30	31	35	25
Diet B	44	34	22	10	47	31	40	30	32	35	18	21

Test if the two diets differ significantly as regards their effect on increasing weights.

25. The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data

Item	1	2	3	4	5	6	7	8	9	10
In '000 Hrs	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of bulbs is 4000 hrs.

26. To the question "should women have competition with men in getting job", the replies received from a sample of 200 subjects of both sexes we as follow

	Yes	No	Total
Males	34	46	80
Females	72	48	120

Sl. No	Job Role	Qualification Packs	NSQF Level
1	Data Analyst	To be aligned	5 (To be aligned)
2	Data Librarian	To be aligned	5 (To be aligned)
3			
4			
5			

SEMESTER III

SDC3CN11 – COMPUTER NETWORKS FOR IOT

Number of Contact Hours: 45 Hrs.

Course Outline

Module 1 [3 T]

Introduction: Overview, Objectives, Networking Connectivity Network Extension, Network Topologies, Protocols, Programs and Processes, Protocol Layering Concepts, The OSI Model Layers: The Physical Layer, The Data Link Layer, The Network Layer, The Transport Layer, The Session Layer, The Presentation Layer The Application Layer

Module 2 [4 T]

LAN Architecture: Transmission Methods and Media LAN Protocols, Ethernet, Token Ring, Token Bus, Fiber Distributed Data Interface (FDDI), Wireless LANs, LAN Protocols. Internetworking: Relays, Repeaters, Bridges, Routers, Gateways, Hubs, Switches, Network Management and SNMP, Network layer: forwarding & routing algorithms (Link, DV), IP-addresses, DNS, NAT, and routers.

Module 3 [10 T]

Data & Network Layer Protocols in IOT- PHY/MAC Layer (3GPP MTC, IEEE 802.11, IEEE 802.15), Wireless HART, Z-Wave, Bluetooth Low Energy, Zigbee Smart Energy, DASH7, Network Layer: IPv4, IPv6, 6LoWPAN, 6TiSCH, ND, DHCP, ICMP, RPL, CORPL, CARP.

Module 4 [12 T]

Transport & Session layer Protocols in IOT-Transport Layer: TCP, MPTCP, UDP, DCCP, SCTP-TLS, DTLS, Session Layer-HTTP, CoAP, XMPP, AMQP, MQTT, SERVICE LAYER PROTOCOLS & SECURITY, Service Layer: oneM2M, ETSI M2M, OMA, BBF, Security in IoT Protocols, MAC 802.15.4, 6LoWPAN, RPL, Application Layer

Module 5 [15 T]

Mobile Networking: roaming and handoffs, mobile IP, and ad hoc and infrastructure less networks Wireless Sensor Networks: Introduction, WSN Coverage & Placement, Topology Management in Wireless Sensor Networks, Medium Access Control in Wireless Networks, Routing in Wireless Sensor Networks, Congestion and Flow Control, Security of Wireless Sensor Networks, Hardware Design of Sensor Node ,Case studies, e.g. sensor body-area-network and control of a smart home

References

1. Data Communications and Networking by Behrouz A. Forouzan
2. Daniel Minoli, "Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications", ISBN: 978-1-118-47347-4, Willy Publications

3. Vijay Madiseti and ArshdeepBahga, "Internet of Things (A Hands-on Approach)", 1st Edition, VPT, 2014
4. Walteneus Dargie, Christian Poellabauer, "Fundamentals of Wireless Sensor Networks, Theory and Practice", Wiley Series on wireless Communication and Mobile Computing, 2011
5. Kazem Sohraby, Daniel manoli, "Wireless Sensor networks- Technology, Protocols and Applications", Wiley InterScience Publications 2010.
6. Bhaskar Krishnamachari, "Networking Wireless Sensors", Cambridge University Press, 2005
7. C.S Raghavendra, Krishna M.Sivalingam, Taiebznati , "Wireless Sensor Networks", Springer Science 2004.
8. Mobile Communications, Second Edition by Jochen Schiller

SDC3CN12 – ELECTRONIC DEVICES AND CIRCUIT DESIGN FOR IOT, IOT WITH RASPBERRY Pi

Number of Credits: 6

Number of Contact Hours: 90 Hrs.

Course Outline

PART - A- ELECTRONICS DEVICES AND CIRCUIT DESIGN FOR IOT

Module 1 [9 T]

Digital IC's:

Monolithic IC technology-planar process-Bipolar junction transistor, FET fabrication, CMOS technology, Digital IC'Ss, logic families - DTL, HTL, RTL, TTL, ECL, PMOS, I²L, performance criteria -comparison, application, advantages

Module 2 [10 T]

Digital Circuits:

Introduction to Operational Amplifiers: Ideal OPAMP, Inverting and Non-Inverting OPAMP circuits, OPAMP applications: voltage follower, addition, subtraction, integration, differentiation; Numerical examples as applicable.

Half adder, Full adder.

Flip-Flops: Introduction to Flip-Flops, NAND Gate Latch/ NOR Gate Latch, RS Flip-Flop, Gated Flip-Flops: Clocked RS Flip-Flop.

Microcontrollers: Introduction to Microcontrollers, 8051 Microcontroller Architecture

Module 3 [9 T]

Communication System:

Introduction to communication, elements of communication system- information source, transmitter, channel, receiver, destination, Need for modulation, Electromagnetic spectrum and typical applications, Terminologies in communication

Elements of analog communication, amplitude modulation (AM) techniques

Theory of angle modulation techniques-frequency modulation, phase modulation, comparison of frequency and phase modulation

Module 4 [8 T]

Digital communication

Pulse modulation techniques-introduction, pulse analog modulation techniques- pulse amplitude modulation (PAM), pulse width modulation, pulse position modulation

Digital modulation techniques – introduction, Basic digital modulation schemes-amplitude shift keys (ASK), frequency shift keying (FSK), phase shift keying (PSK)

Module 5 [9 T]

Sensors

Definition, Types, Basic principle and applications of Resistive, Inductive, Capacitive, Piezoelectric and their Dynamic performance. Fiber optic sensors, Bio-chemical sensors, Hall-Effect, Photoemissive, Photo Diode/ Photo Transistor, Photovoltaic, LVDT, Strain Gauge Digital transducers: Principle, Construction, Encoders, Absolute and incremental encoders, Silicon micro transducers.

References

1. Ramakant A.Gayakwad- “Op Amps and Linear integrated circuits” – PHI 2008
2. D.Roy Choudhury ,Shail B Jain “–Linear Integrated circuits”, New Age international (P)Ltd,2010
3. Kennedy’s Electronic communication system – George Kennedy, Bernard Davis, S R M Prasanna,5e, Mc Graw Hill Education
4. David A. Bell, “Electronic Devices and Circuits”, Oxford University Press, 5th Edition, 2008.
5. D.P. Kothari, I. J. Nagrath, “Basic Electronics”, McGraw Hill Education (India) Private Limited, 2014.
6. Muhammad Ali Mazidi, “The 8051 Microcontroller and Embedded. Systems. Using Assembly and C.” Second Edition, 2011, Pearson India.
7. Doebelin, E.O. and Manic, D.N., Measurement Systems: Applications and Design, McGraw–Hill (2004).
8. Sawhney, A.K. and Sawhney, P., A Course in Electrical and Electronic Measurements and Instrumentation, Dhanpat Rai (2008).

PART - B - IOT WITH RASPBERRY Pi

Module 1 [8 T]

Preparing our IoT Projects -Creating the sensor project, Creating the actuator project, Creating a controller, Creating a camera. The HTTP Protocol-HTTP basics, Adding HTTP support to the sensor, Adding HTTP support to the actuator, Adding HTTP support to the controller.

Module 2 [8 T]

The UPnP Protocol-Introducing UPnP, Creating a device description document, Creating the service description document, Providing a web interface, Creating a UPnP interface, Implementing the Still Image service, Using our camera. The CoAP Protocol-Making HTTP binary, Adding CoAP to our sensor, Adding CoAP to our actuator, Using CoAP in our controller

Module 3 [10 T]

The MQTT Protocol-Publishing and subscribing, Adding MQTT support to the sensor, Adding MQTT support to the actuator, Adding MQTT support to the controller, The XMPP Protocol-XMPP basics, Adding XMPP support to a thing, Providing an additional layer of security, Adding XMPP support to the actuator, Adding XMPP support to the camera, Adding XMPP support to the controller, Connecting it all together.

Module 4 [10 T]

Using an IoT Service Platform-Selecting an IoT platform, The Clayster platform, Interfacing our devices using XMPP, Creating our control application.

Creating Protocol Gateways, understanding protocol bridging, Using an abstraction model, The basics of the Clayster abstraction model, Understanding the CoAP gateway architecture.

Module 5 [9 T]

Security and Interoperability-Understanding the risks, Modes of attack, Tools for achieving security, The need for interoperability.

References

1. Learning Internet of Things, Peter Waher.
2. Raspberry Pi IoT Projects, Prototyping Experiments for Projects, John C Shovic

SDC3PC13P– PRACTICAL 3: PYTHON FOR DATA SCIENCE & COMPUTER NETWORKS FOR IOT

Number of Credits: 5

Course Outline

PART A: PYTHON FOR DATA SCIENCE

List of Exercises:

- 1 Program to stack two arrays vertically and horizontally.
- 2 Program to remove items from one array that exist in another.
- 3 Program to extract all numbers between a given range from a numpy array.
- 4 Program to swap two columns in a 2d numpy array.
- 5 Program to create a 2D array containing random floats between 5 and 10.
- 6 Program to print numpy array with three places.
- 7 Program to print the full numpy array without truncating.
- 8 Program to import a dataset with numbers and texts keeping the text intact in python numpy.

- 9 Program to extract a particular column from 1D array of tuples.
- 10 Program to compute the mean, median, standard deviation of a numpy array.
- 11 Program to find the percentile scores of a numpy array.
- 12 Program to insert values at random positions in an array.
- 13 Program to find the position of missing values in numpy array.
- 14 Program to find the correlation between two columns of a numpy array.
- 15 Program to get the second largest value of an array when grouped by another array.
- 16 Program to remove rows from a data frame that are present in another data frame.
- 17 Program to create a new column that contains the row number of nearest column by euclidean distance.
- 18 Program to find and cap outliers from a series or data frame column.
- 19 Program to get the last n rows of a data frame with row sum > 100.
- 20 Program to find the position of the nth largest value greater than a given value.
- 21 Program to filter every nth row in a data frame.
- 22 Program to format all the values in a data frame as percentages.
- 23 Program to format or suppress scientific notations in a pandas data frame.
- 24 Program to change the order of columns of a data frame.
- 25 Program to count the number of missing values in each column.
- 26 Program to rename a specific column in a data frame.
- 27 Program to compute the autocorrelations of a numeric series.

PART B: COMPUTER NETWORKS FOR IOT

List of Exercises:

1. Implement functions that send and receive Ethernet packet.
2. Implement functions that send and receive ARP packets.
3. Implement an end-host version of the Internet Protocol
4. Implement the sending and receiving of UDP datagrams.
5. Implement the Microprotocol Blast
6. Implement TCP/IP Connectivity
7. Implement UDP Connectivity
8. Implement Distance Vector routing algorithm
9. Implement Zigbee protocol
10. Implement Bluetooth protocol
11. Implement z-wave protocol
12. Implement 6LowPAN protocol
13. Implement WiFi protocol
14. Implement Cellular protocol
15. Implement NFC protocol
16. Implement Sigfix protocol
17. Implement MQTT protocol
18. Implement TR-069 protocol

19. Implement JSON-LD protocol
20. Implement Homekit protocol
21. Implement DNS-SD protocol
22. Implement EPC protocol
23. Cost–Benefit Analysis for a Networking Project:
24. Simulate network using Hubs
25. Simulate Server-Based LAN Design
26. Simulate IP Subnet Addressing
27. Simulate Internetworking using routers
28. Simulate Dynamic routing using RIP
29. Simulate routing using OSPF
30. Simulate routing using PING
31. Simulate routing using IGMP
32. Simulate routing using SNMP
33. Simulate static routing
34. Simulate packet tracer
35. Simulate access lists

SDC3IR14P– PRACTICAL 4: BASICS OF IOT WITH RASPBERRY PI

Number of Credits: 4

Course Outline

PART A: BASICS OF IOT

List of Exercises:

1. Plot the input and output characteristics of an NPN transistor in common emitter configuration and to calculate the dynamic input resistance, dynamic output resistance and common emitter current gain β .
2. To plot the V-I characteristics of silicon PN junction diodes, and calculate their static and dynamic resistances.
3. Study and plot the zener diode characteristics.
4. Design and setup an RC integrator and study its response to pulse and square waves
5. Design and set up an RC differentiator circuit and study its response to pulse and square waves
6. Design and set up fullwave rectifier (FWR) with and without filters.
7. Design and set up an RC-coupled common emitter (CE) amplifier using bipolar junction transistor and to plot its frequency response.
8. Design, set up and study clipping circuits using series and shunt clippers
9. Study clamping circuits using diodes
10. Design and setup an RC phase shift oscillator using op-amp for a frequency of 1kHz
11. Design and setup a non inverting amplifier and an inverting amplifier

12. Design and set up an astablemultivibrator using BJT to generate a square wave for a given frequency, amplitude and duty cycle.
13. Become familiar with the basics of LTSpice for circuit simulation.
14. Become familiar with the basis of Operational amplifiers

PART B: IOT WITH RASPBERRY Pi

List of Exercises:

1. Configure Raspberry Pi.
2. Implement WAP for LED blink using Raspberry Pi.
3. Implement Zigbee Protocol using Raspberry Pi.
4. To demonstrate the basic linux commands on Raspberry pi.
5. To create a database & Store the value in Raspberry Pi.
6. To Setup RPi first time without using screen, mouse, keyboard

SEMESTER IV

SDC4DS15 – DATA SCIENCE WITH R AND PYTHON PROGRAMMING, EMBEDDED LINUX OS & ANDROID PROGRAMMING

Number of Contact Hours: 90 Hrs.

Course Outline

PART -A- DATA SCIENCE WITH R AND PYTHON PROGRAMMING

Module 1[8 T]

Introduction: History and overview of R, Getting Started with R, R Nuts and Bolts, Getting Data In and Out of R. Using the readr Package, Interfaces to the Outside World, Subsetting R Objects, Vectorized Operations, Dates and Times, Managing Data Frames with the dplyr package, Control Structures, Functions. Scoping Rules of R, Coding Standards for R, Loop Functions, Regular Expressions, Data Manipulation (dplyr, reshape2 packages), String Operations (stringr package), Packaging, Debugging and Object Oriented Programming.

Module 2[8 T]

Data Visualization (ggplot2 package), Domain specific languages: HTML, LaTeX, Using R in statistics: Clustering, Regression and Classification. High performance functions with Rcpp, R's C interface, Data Analytics Case Study

Module 3[10 T]

Introduction – The Ascendance of Data, what is Data Science, Introduction to NumPy: Understanding Data Types in Python, The Basics of NumPy Arrays, Computation on NumPy Arrays: Universal Functions,

Aggregations: Min, Max, and Everything in Between, Computation on Arrays: Broadcasting, Comparisons, Masks, and Boolean Logic, Fancy Indexing, Sorting Arrays, Structured Data: NumPy's Structured Arrays.

Module 4[9 T]

Data Manipulation with Pandas, Installing and Using Pandas, Introducing Pandas Objects, Data Indexing and Selection, Operating on Data in Pandas, Handling Missing Data, Hierarchical Indexing, Combining Datasets: Concat and Append, Combining Datasets: Merge and Join, Aggregation and Grouping, Pivot Tables, Vectorized String Operations, Working with Time Series, High-Performance Pandas: eval () and query ()

Module 5[10 T]

Visualization with Matplotlib, General Matplotlib Tips, Two Interfaces for the Price of One, Simple Line Plots, Simple Scatter Plots, Visualizing Errors, Density and Contour Plots, Histograms, Binnings, and Density, Customizing Plot Legends, Customizing Color bars, Multiple Subplots, Text and Annotation, Customizing Ticks, Customizing Matplotlib: Configurations and Stylesheets, Three-Dimensional Plotting in Matplotlib, Geographic Data with Basemap, Visualization with Seaborn. Machine learning, Neural Networks, Natural Language Processing, Networks and Analysis

References

1. R Programming for Data Science, by Roger D. Peng
2. Using R for Introductory Statistics, by John Verzani, Chapman & Hall/CRC, 2004, ISBN 1584884509
3. Advanced R, by Hadley Wickham, ISBN 9781466586963.
4. Big data:The black book ,Dream Tech Press, 2017 Edition,ISBN 9789351199311
5. Data Science From Scratch First Principle with Python
6. R Programming For Data Science by Roger D. Peng
7. Python Data Science Handbook Essential Tools For Working With Data by Jake VanderPlas

PART -B- EMBEDDED LINUX OS & ANDRIOD PROGRAMMING

Module 1[7 T]

Introduction, Processor Basics, The Linux Kernel - An Introduction, Kernel Initialization, System Initialization. Bootloaders, Device Driver Basics, File Systems, MTD Subsystem, BusyBox. Development environment and Tools: Basics regarding Kernel Space and its interface to User Space, Shell and basic shell commands, Basic IP Filters, Linux Memory Model, Linux Scheduling and priorities.

Module 2[8 T]

Comparing Linux to other RTOS, Porting considerations from other RTOS (VxWorks) to Linux, Tools chain Overview, POSIX Threads, Synchronization and Communication, C Signals, /proc and /sys for Kernel 2.6, dynamic Modules and Procs, Kernel Debugging, Embedded Linux Applications, Development Tools, Porting Linux, User Mode Linux

Module 3[10 T]

Introducing the android computing platform, History of android, android software stack, Developing end user application using Android SDK, Android java packages, Setting up the development environment, Installing android development tools (ADT), Fundamental components, Android virtual devices, Running on real device, Structure of android application, Application life cycle. Understanding android resources - String resources, Layout resources, Resource reference syntax, Defining own resource IDs - Enumerating key android resources, string arrays, plurals, Color resources, dimension resources, image resources, Understanding content providers - Android built in providers, exploring databases on emulator, architecture of content providers, structure of android content URIs, reading data using URIs, using android cursor, working with where clause, inserting updates and deletes, implementing content, AsyncTask class- Create Services, Services Using AsyncTask.

Module 4[10 T]

Create Broadcast Receivers, Understanding intents basics of intents, available intents, exploring intent composition, Rules for Resolving Intents to Their Components, ACTION PICK, GET CONTENT, pending intents User interfaces development in android - building UI completely in code, UI using XML, UI in XML with code, Android's common controls - Text controls, button controls, checkbox control, radio button controls, image view, date and time controls, map view control, understanding adapters, adapter views, list view, grid view, spinner control, gallery control, styles and themes, Understanding layout managers - linear layout manager, table layout manager, relative layout manager, frame layout manager, Constraint layout manager, Absolute layout manager, grid layout manager. Android menus - creating menus, working with menu groups, responding to menu items, icon menu, sub menu, context menu, dynamic menus, loading menu through XML, popup menus, option menus, Fragments in Android structure of fragment, fragment life cycle, fragment transaction and back stack, fragment manager, saving fragment state, persistence of fragments, communications with fragments, startActivity() and setTargetFragment(), using dialogs in android, dialog fragments, working with toast.

Module 5[10 T]

Implementing action bar - tabbed navigation action bar activity, implementing base activity classes, tabbed action bar and tabbed listener, debug text view layout, action bar and menu interaction, list navigation action bar activity, spinner adapter, list listener, list action bar, standard navigation action bar activity, action bar and search view, action bar and fragments. Android Advanced Concepts-Drag and Drops, Notification, Sending SMS and Email, Phone Calls, Location Based Services- Google Map, Persisting data - Files, saving state and preferences - saving application data, creating, saving and retrieving shared preferences, preference framework and preference activity, preference layout in XML, native preference controls, preference fragments, preference activity, persisting the application state, including static files as resources, Working with file system, SQLite – SQL Lite types, database manipulation using SQL Lite, SQL and

database centric data model for Android, Android database classes, Introduction to API, Publishing Android Application

References

1. Embedded Linux Primer: A Practical Real-world Approach by Christopher Hallinan
2. Building Embedded Linux Systems by Karim Yaghmour
3. Linux: Embedded Development by Alex Gonzalez and Chris Simmonds
4. Pro Android 4, SatyaKomatineni& Dave MacLean, Apress.
5. Professional Android 4 Application Development, Retomeier, Wrox.
6. Programming Android, ZigurdMednieks, Laird Dornin, G. BlakeMeike, and Masumi Nakamura, O'Reilly.
7. <https://developer.android.com>

SDC4DS17P– PRACTICAL 5: DATA SCIENCE WITH R PROGRAMMING AND DATA VISUALISATION USING QlikView AND Power BI

Number of Credits: 4

Course Outline

PART A: DATA SCIENCE WITH R PROGRAMMING PART B: DATA VISUALISATION USING QlikView AND Power BI

List of Exercises: PART A

36. Iris Data Set

This is probably the most versatile, easy and resourceful dataset in pattern recognition literature. The data has only 150 rows & 4 columns.

Problem: Predict the class of the flower based on available attributes.

37. Loan Prediction Dataset

This dataset provides you a taste of working on data sets from insurance companies – what challenges are faced there, what strategies are used, which variables influence the outcome, etc. This is a classification problem. The data has 615 rows and 13 columns.

Problem: Predict if a loan will get approved or not.

38. Bigmart Sales Data Set

Retail is another industry which extensively uses analytics to optimize business processes. Tasks like product placement, inventory management, customized offers, product bundling, etc. are being smartly handled using data science techniques. As the name suggests, this data comprises of transaction records of a sales store. This is a regression problem. The data has 8523 rows of 12 variables.

Problem: Predict the sales of a store.

39. Boston Housing Data Set

This is another popular dataset used in pattern recognition literature. The data set comes from the real estate industry in Boston (US). This is a regression problem. The data has 506 rows and 14 columns.

Problem: Predict the median value of owner occupied homes.

40. Time Series Analysis Dataset

Time Series is one of the most commonly used techniques in data science. It has wide ranging applications – weather forecasting, predicting sales, analyzing year on year trends, etc. This dataset is specific to time series and the challenge here is to forecast traffic on a mode of transportation. The data has ** rows and ** columns.

Problem: Predict the traffic on a new mode of transport.

41. Wine Quality Dataset

This is one of the most popular datasets along data science beginners. It is divided into 2 datasets. You can perform both regression and classification tasks on this data. It will test your understanding in different fields – outlier detection, feature selection, and unbalanced data. There are 4898 rows and 12 columns in this dataset.

Problem: Predict the quality of the wine.

42. Turkiye Student Evaluation Dataset

This dataset is based on an evaluation form filled out by students for different courses. It has different attributes including attendance, difficulty, score for each evaluation question, among others. This is an unsupervised learning problem. The dataset has 5820 rows and 33 columns.

Problem: Use classification and clustering techniques to deal with the data

43. Heights and Weights Dataset

This is a fairly straightforward problem and is ideal for people starting off with data science. It is a regression problem. The dataset has 25,000 rows and 3 columns (index, height and weight).

Problem: Predict the height or weight of a person.

44. Black Friday Dataset

This dataset comprises of sales transactions captured at a retail store. It's a classic dataset to explore and expand your feature engineering skills and day to day understanding from multiple shopping experiences. This is a regression problem. The dataset has 550,069 rows and 12 columns.

Problem: Predict purchase amount.

45. Human Activity Recognition Dataset

This data set is collected from recordings of 30 human subjects captured via smartphones enabled with embedded inertial sensors. This is a multi-classification problem. The data set has 10,299 rows and 561 columns..

Problem: Predict the activity category of a human.

46. Text Mining Dataset

This dataset is originally from the Siam Text Mining Competition held in 2007. The data comprises of aviation safety reports describing problem(s) which occurred in certain flights. It is a multi-classification and high dimensional problem. It has 21,519 rows and 30,438 columns.

Problem: Classify the documents according to their labels.

47. Trip History Dataset

This dataset comes from a bike sharing service in the United States. This dataset requires you to exercise your pro data munging skills. The data is provided quarter-wise from 2010 (Q4) onwards. Each file has 7 columns. It is a classification problem.

Problem: Predict the class of user.

48. Million Song Dataset

This data set puts forward a regression task. It consists of 5,15,345 observations and 90 variables.

Problem: Predict release year of the song.

49. Census Income Dataset

It's an imbalanced classification and a classic machine learning problem. You know, machine learning is being extensively used to solve imbalanced problems such as cancer detection, fraud detection etc. It's time to get your hands dirty. The data set has 48,842 rows and 14 columns..

Problem: Predict the income class of US population.

50. Movie Lens Dataset

This dataset is one of the most popular & quoted datasets in the data science industry. It has dimension 1 million ratings from 6,000 users on 4,000 movies.

Problem: Recommend new movies to users.

51. Twitter Classification Dataset

Working with Twitter data has become an integral part of sentiment analysis problems. The dataset is 3MB in size and has 31,962 tweets.

Problem: Identify the tweets which are hate tweets and which are not.

52. Identify your Digits Dataset

This dataset allows you to study, analyze and recognize elements in the images. That's exactly how your camera detects your face, using image recognition! It's your turn to build and test that technique. It's a digit recognition problem. This data set has 7,000 images of 28 X 28 size, totalling 31MB.

Problem: Identify digits from an image.

53. Urban Sound Classification

This practice problem is meant to introduce you to audio processing in the usual classification scenario. This dataset consists of 8,732 sound excerpts of urban sounds from 10 classes.

Problem: Classify the type of sound from the audio.

54. Vox Celebrity Dataset

Audio processing is rapidly becoming an important field in deep learning hence here's another challenging problem. This dataset is for large-scale speaker identification and contains words spoken by celebrities, extracted from YouTube videos. It's an intriguing use case for isolating and identifying speech recognition. The data contains 100,000 utterances spoken by 1,251 celebrities.

Problem: Figure out which celebrity the voice belongs to.

55. ImageNet Dataset

ImageNet offers variety of problems which encompasses object detection, localization, classification and screen parsing. All the images are freely available. You can search for any type of image and build

your project around it. This image engine has more than 15 million images of multiple shapes sizing up to 140GB.

Problem: Problem to solve is subjected to the image type you download.

56. Chicago Crime Dataset

The ability to handle large datasets is expected of every data scientist these days. Companies no longer prefer to work on samples when they have the computational power to work on the full dataset. This dataset provides you a much needed hands-on experience of handling large data sets on your local machines. The problem is easy, but data management is the key! This dataset has 6M observations. It's a multi-classification problem.

Problem: Predict the type of crime.

57. Age Detection of Indian Actors Dataset

This is a fascinating challenge for any deep learning enthusiast. The dataset contains thousands of images of Indian actors and your task is to identify their age. All the images are manually selected and cropped from the video frames resulting in a high degree of variability in terms of scale, pose, expression, illumination, age, resolution, occlusion, and makeup. There are 19,906 images in the training set and 6,636 in the test set.

Problem: Predict the age of the actors.

58. Recommendation Engine Dataset

This is an advanced recommendation system challenge. In this practice problem, you are given the data of programmers and questions that they have previously solved, along with the time that they took to solve that particular question. As a data scientist, this will help online judges to decide the next level of questions to recommend to a user.

Problem: Predict the time taken to solve a problem given the current status of the user.

59. VisualQA Dataset

VisualQA is a dataset containing open-ended questions about images. These questions require an understanding of computer vision and language. There is an automatic evaluation metric for this problem. The dataset has 265,016 images, 3 questions per image and 10 ground truth answers per question.

Problem: Use deep learning technique to answer open-ended questions about images.

60. Implement law of large numbers

61. You are a Data Scientist working for a consulting firm. One of your colleagues from the Auditing department has asked you to help them assess the financial statement of organisation X.

You have been supplied with two vectors of data: monthly revenue and monthly expenses for the financial year in question. Your task is to calculate the following financial metrics: -

- profit for each month
- profit after tax for each month (the tax rate is 30%)
- profit margin for each month - equals to profit after tax divided by revenue
- good months - where the profit after tax was greater than the mean for the year

- bad months - where the profit after tax was less than the mean for the year
- the best month - where the profit after tax was max for the year
- the worst month - where the profit after tax was min for the year.

All results need to be presented as vectors. Results for dollar values need to be calculated with \$0.01 precision, but need to be presented in Units of \$1,000 (i.e. 1k) with no decimal points. Results for the profit margin ratio need to be presented in units of % with no decimal points. Note: Your colleague has warned you that it is okay for tax for any given month to be negative (in accounting terms, negative tax translates into a deferred tax asset)

62. You have been approached as an analytics consultant by a movie reviews website. They are writing up an article analysing movie ratings by critics and audience as well as movie budgets for the years 2007-2011.

This is the first time they are doing this analysis and they don't know exactly what they need. They have asked you to look into the data and provide them with 5 graphs that tell a story about the data. However, there is one chart the CEO specifically requested - a diagram showing how the correlation between audience and critic ratings has evolved throughout the years by genre.

63. The movie reviews website was very happy with your deliverable for the previous assignment and now they have a new request for you. The previous consultant had created a chart for them (using above data). However, the R code used to create the diagram has since been lost and cannot be recovered. Your task is to come up with the code that will re-create the same chart making it look as close as possible to the original.

A new dataset has been supplied.

64. You have been supplied data for two more additional in-game statistics:

- Free Throws
 - Free Throw Attempts
- You need to create three plots that portray the following insights:
- Free Throw Attempts per game
 - Accuracy of Free Throws
 - Player playing style (2 vs 3 points preference) excluding Free Throws* *Each Free Throw is worth 1 point.

The data has been supplied in the form of vectors. You will have to create the two matrices before you proceed with the analysis

65. You are employed as a Data Scientist by the World Bank and you are working on a project to analyse the World's demographic trends.

You are required to produce a scatterplot illustrating Birth Rate and Internet Usage statistics by Country.

The scatterplot needs to also be categorised by Countries' Income Groups. You have received an urgent update from your manager.

You are required to produce a second scatterplot also illustrating Birth Rate and Internet Usage statistics by Country.

However, this time the scatterplot needs to be categorised by Countries' Regions.

Additional data has been supplied in the form of R vectors

66. The World Bank was very impressed with your deliverables on the previous assignment and they have a new project for you.

You are required to produce a scatterplot depicting Life Expectancy (y-axis) and Fertility Rate (x-axis) statistics by Country.

The scatterplot needs to also be categorised by Countries' Regions.

You have been supplied with data for 2 years: 1960 and 2013 and you are required to produce a visualisation for each of these years.

List of Exercises: PART B

1. Program to get data from excel files.
2. Program to load data from multiple sources.
3. Program to remove unwanted columns from tables.
4. Program to combine multiple table.
5. Implement and format area chart
6. Implement and format Donut chart.
7. Implement and format clustered column chart.
8. Implement and format multi-row card.
9. Implement and format filled map .
10. Implement R script.
11. Implement slicer.
12. Implement filters on measures.
13. Implement page level filters.
14. Implement report level filters
15. Implement drill through filters.
16. Create a Dashboard and implement Dashboard actions.
17. Program to add web content to Dashboard.
18. Program to upload excel files Dashboard.
19. Program to pin workbook to Dashboard.
20. Program to get insights of a Dataset.
21. Display data from a QVD file.
22. Import data from Excel to QlikView.

23. Implement aggregate functions in the dataset .
24. Implement match ,rank ,peek, rangesum functions.
25. Implement List Box to filter the data present in Charts.
26. Implement Multi Box to filter the data present in Charts
27. Implement table Box to filter the data present in Charts
28. Create Bar Chart in QlikView.
29. Create Block Chart in QlikView.
30. Create Funnel Chart in QlikView
31. Create grid Chart in QlikView
32. Create line Chart in QlikView
33. Create Mekko Chart in QlikView
34. Create Pie Chart in QlikView
35. Create Radar Chart in QlikView
36. Create Scatter Chart in QlikView.
37. Create Dashboard in QlikView.
38. Implement rotating table .
39. Implement mapping/ lookup table.
40. Create a Master calendar in QlikView.

SDC4AL18P– PRACTICAL 6: ANDROID AND LINUX

Number of Credits: 4

Course Outline

PART A: ANDRIOD

List of Exercises:

1. Simple User Interface to display a message
2. Android program to add two numbers
3. Build a sample android application to display a message on button click
4. Program to demonstrate the use of string.xml file
5. Android program to demonstrate Activity life cycle
6. Android program to change the background color of your Activity
7. Program to implement the use of Event handling on Button click
8. Program to Start another activity from your own activity using Intent
9. Program to understand basic logic operations
10. Android program to create login screen with two Activity
11. Program to demonstrate passing data using Intent

12. Program to Dial a number the using Implicit Intent
13. Program to demonstrate the use of content provider
14. Program to launch the existing camera application to capture an image and display the result in the form of bitmap.
15. Program to change the image displayed on the screen using Frame layout
16. Program to demonstrate Toast in an application
17. Program to Demonstrate Dialogue box in an application
18. Program to demonstrate Radio Group application
19. Program to demonstrate the use of Checkbox control
20. Program to demonstrate the use of Date time Control
21. Create an Android Application to implement custom Styles and Themes
22. Program to get IP Address
23. Create a Calculator APP in Android
24. Create your own Android application using List View
25. program to Display Images in Grid view
26. Program to Demonstrate the use of Spinner Control
27. Android program to create option menu
28. Program to demonstrate the use of Context menu
29. Create an Android Application to create a pop-up menu
30. Program to demonstrate the use of Submenu
31. Program to Start a service from your own activity
32. Create an Android Application to Play music using Service
33. Create an Android Application to download a file by Service using AsyncTask Class
34. Android Application to Display Battery low notification using broadcast Receiver Class
35. Program to demonstrate the use of fragment
36. Program to setup Action bar with search option
37. Create an Android Application to send SMS
38. Create a Sample app with location Finder
39. Create a sample Sample app with Notification
40. Create a sample App with Location based services
41. Create a sample App to work with Drag and Drop
42. Program to demonstrate sending Email using Android App
43. Create an Android Application to store and display data using Shared Preferences
44. Create an Android Application to store, Filter and Display data using SQLite
45. Android login App with Api

PART B: LINUX

List of Exercises:

1. Write a shell script program to display list of user currently logged in.
2. Write a shell script program to display "HELLO WORLD".

3. Write a shell script program to develop a scientific calculator.
4. Write a shell Script program to check whether the given number is even or odd.
5. Shell script Program to search whether element is present is in the list or not.
6. Shell script program to check whether given file is a directory or not.
7. Shell script program to count number of files in a Directory.
8. Shell script program to copy contents of one file to another.
9. Create directory, write contents on that and Copy to a suitable location in your home directory.
10. Use a pipeline and command substitution to set the length of a line in file to a variable.
11. Write a program using sed command to print duplicated lines of Input.
12. Write a grep/egrep script to find the number of words character, words and lines in a file.
13. Write an awk script to develop a Fibonacci series.
14. Write a perl script to compute the power of a given number.
15. Write an awk script to display the pattern of given string or number.
16. Write a perl script to check a number is prime or not.
17. Write an egrep script to display list of files in the directory.
18. Write a shell script program to display the process attributes.
19. Write a shell script to change the priority of processes.
20. Write a shell script to change the ownership of processes.
21. Write a program to send back a process from foreground.
22. Write a program to retrieve a process from background.
23. Write a program to create a Zombie process.
24. Write a program to create a child process and allow the parent to display "parent" and the child to display "child" on the screen.
25. Write a shell script program to check variable attributes of file and processes.
26. Write a shell script program to check and list attributes of processes.
27. Shell Script program to implement read, write, and execute permissions.
28. Shell Script program for changing process.
29. To execute programs using gdb to utilize its various features like breakpoints, conditional breakpoints.
30. Write a shell script program to include verbose Debug option for debugging.
31. Write a shell script program to include xtrace Debug option for debugging.
32. Write a shell script program to include verbose and trace Debug option
33. Create and compile your own cross-compiling toolchain for the uClibc C library
34. Create and set upserial communication, compile and install the U-Boot bootloader, use basic U-Boot commands, set up TFTP communication with the development workstation.
35. Get the kernel sources and patch them
36. Cross-compile a kernel for an ARM target platform.
37. Create a tiny yet full featured embedded system
38. configure and boot an embedded Linux system relying on block storage
39. Implement flash and flash file systems usage and their integration on the target
40. Build a Linux system with libraries and make it work on the board.

41. Compile and run your own ncurses application on the target.
42. Use strace to diagnose program issues. Use gdbserver and a cross-debugger to remotely debug an embedded application
43. Learn how to handle real-time processes and practice with the different real-time modes. Measure scheduling latency.

Sl. No	Job Role	Qualification Packs	NSQF Level
1	Senior Data Analyst	To be aligned	6 (To be aligned)
2	Data Curator	To be aligned	6 (To be aligned)
3			
4			
5			

SEMESTER V

SDC5ML20 – MACHINE LEARNING USING PYTHON

Number of Contact Hours: 45 Hrs.

Course Outline

Module 1 [8 T]

Machine learning and Python – Getting started, Introduction to NumPy, SciPy, and Matplotlib, Installing Python, Chewing data efficiently with NumPy and intelligently with SciPy, Learning NumPy, Indexing, Handling non-existing values, Comparing runtime behaviors, Learning SciPy, Our first machine learning application, Reading in the data, Preprocessing and cleaning the data, Choosing the right model and learning algorithm, The Iris dataset, The first step is visualization, Building our first classification model, Evaluation – holding out data and cross-validation

Module 2 [8 T]

Learning How to Classify with Real-world Example, Building more complex classifiers, A more complex dataset and a more complex classifier, Learning about the Seeds dataset, Features and feature engineering, Nearest neighbor classification, Binary and multiclass classification, Clustering – Finding Related Posts, Measuring the relatedness of posts, Preprocessing – similarity measured as similar number, of common words, Converting raw text into a bag-of-words, Counting words, Normalizing the word count vectors, Removing less important words, Stemming, Installing and using NLTK, Extending the vectorizer with NLTK's stemmer, Stop words on steroids, Clustering, KMeans Clustering posts, Solving our initial challenge, Another look at noise, Tweaking the parameters, Topic Modeling, Latent Dirichlet allocation

(LDA), Building a topic model, Comparing similarity in topic space, Modeling the whole of Wikipedia, Choosing the number of topics

Module 3 [10 T]

Classification: Detecting Poor Answers, Tuning the instance, Tuning the classifier, Fetching the data Slimming the data down to chewable chunks, Preselection and processing of attributes, Defining what is a good answer, Creating our first classifier, Starting with the k-nearest neighbor (KNN) algorithm, Engineering the features, Training the classifier, Measuring the classifier's performance, Designing more features, Deciding how to improve, Bias-variance and its trade-off, Fixing high bias, Fixing high variance, High bias or low bias, Using logistic regression, A bit of math with a small example, Applying logistic regression to our post classification problem, Classification II – Sentiment Analysis, Fetching the Twitter data, Introducing the Naive Bayes classifier, Getting to know the Bayes theorem, Being naïve, Using Naive Bayes to classify, Accounting for unseen words and other oddities, Accounting for arithmetic underflows, Creating our first classifier and tuning it Solving an easy problem first, Using all the classes, Tuning the classifier's parameters, Cleaning tweets, Taking the word types into account, Determining the word types.

Module 4 [10T]

Predicting house prices with regression, Multidimensional regression, Cross-validation for regression, Penalized regression, L1 and L2 penalties, Using Lasso or Elastic nets in scikit-learn, P greater than N scenarios, An example based on text, Setting hyperparameters in a smart way, Rating prediction and recommendations, Improved recommendations, Using the binary matrix of recommendations, Looking at the movie neighbors, Combining multiple methods, Basket analysis Obtaining useful predictions Analyzing supermarket shopping baskets Association rule mining More advanced basket analysis. Fetching the music data, converting into a wave format, looking at music, Decomposing music into sine wave components, Using FFT to build our first classifier Increasing experimentation agility, Training the classifier, Using the confusion matrix to measure accuracy in multiclass problems, an alternate way to measure classifier performance using Receiver Operating Characteristics.

Computer Vision – Pattern Recognition, introducing image processing, Loading and displaying images, Basic image processing, Thresholding, Gaussian blurring, Filtering for different effects, adding salt and pepper noise, Putting the center in focus, Pattern recognition, Computing features from images, writing your own features, Classifying a harder dataset, Local feature representations.

Module 5 [9 T]

Dimensionality Reduction, Sketching our roadmap, Selecting features, Detecting redundant features using filters, Correlation, Mutual information, Asking the model about the features using wrappers, Other feature selection methods, Feature extraction, About principal component analysis (PCA), Sketching PCA, Applying PCA, Limitations of PCA and how LDA can help, Multidimensional scaling (MDS), Learning about big data, Using jug to break up your pipeline into tasks, About tasks, Reusing partial results, Looking under the hood, Using jug for data analysis.

References

1. Building Machine Learning Systems with Python - Willi Richert, Luis Pedro Coelho
2. <http://scikit-learn.org/stable/>
3. Introduction to machine learning with python - Andreas C. Müller and Sarah Guido

SDC5BC21- BIG DATA & CLOUD PLATFORM FOR IOT

Number of Contact Hours: 60 Hrs.

Course Outline

Module 1[8 T]

Overview of big Data-big data stack-Business motivation and drivers for everything-big data analysis techniques-Introducing technologies for handling big data. Big data management- big data storage concepts and technology –polygot persistence -big data processing concepts. Understanding Hadoop Ecosystem- MapReduce Fundamentals and HBase -processing data with MapReduce- MapReduce execution and implementing MapReduce program

Module 2[8 T]

Data Science-utilizing data structures and algorithm-Integrating R and Hadoop for statistics-Predictive analytics with Mahout-Introduction to Mahout. Understanding Analytics and big Data- Analytical Approaches and tools to analyze data- Exploring Hive- Analyzing Data with Pig.

Module 3[15 T]

Cloud Computing definition, private, public and hybrid cloud. Cloud types; IaaS, PaaS, SaaS. Benefits and challenges of cloud computing, public vs private clouds, role of virtualization in enabling the cloud; Business Agility: Benefits and challenges to Cloud architecture. Application availability, performance, security and disaster recovery; next generation Cloud Applications. Technologies and the processes required when deploying web services; Deploying a web service from inside and outside a cloud architecture, advantages and disadvantages

Module 4[14T]

Reliability, availability and security of services deployed from the cloud. Performance and scalability of services, tools and technologies used to manage cloud services deployment; Cloud Economics: Cloud Computing infrastructures available for implementing cloud-based services. Economics of choosing a Cloud platform for an organization, based on application requirements, economic constraints and business needs (e.g Amazon, Microsoft and Google, Salesforce.com, Ubuntu and Redhat) Service creation environments to develop cloud-based applications. Development environments for service development; Amazon, Azure, Google App. Internet of Things Promises–Definition, Scope, IoT-An Architectural Overview– Building an architecture, Main design principles and needed capabilities, An IoT architecture outline, standards considerations. M2M and IoT Technology Fundamentals- Devices and gateways, Local

and wide area networking, Data management, Business processes in IoT, Everything as a Service (XaaS), M2M and IoT Analytics, Knowledge Management

Module 5[15 T]

Sensors for IoT Applications–Structure of IoT– IoT Map Device. The IoT Era - Scope of the Internet of Things - Components of IoT-Enabled Things - IoT World Forum Reference Model - ITU-T IoT Reference Model - IoTivity - Cisco IoT System - ioBridge - SDN and NFV over IoT Deployment. IOT PROTOCOLS Security Requirements in IoT Architecture - Security in Enabling Technologies - Security Concerns in IoT Applications. Security Architecture in the Internet of Things - Security Requirements in IoT. Cloud services and IoT – offerings related to IoT from cloud service providers – Cloud IoT security controls – An enterprise IoT cloud security architecture – New directions in cloud enabled IoT computing.

References

1. Black book –BIG DATA, ISBN:978-93-5119-931-1.
2. Thomas Erl with Wajid Khattak and Paul Buhler, Big Data Fundamentals: Concepts: Drivers and Techniques.
3. Hurwitz, Alan Nugent, Fern Halper and Marcia Kaufman, Big Data For Dummies, ISBN:9781118504222.
4. Robert D. Schneider , Hadoop for Dummies, ISBN: 978-1-118-25051-8.
5. Alex Holmes, Hadoop in Practice, ISBN :9781617290237
6. Gautam Shroff, “Enterprise Cloud Computing Technology Architecture Applications”, Cambridge University Press; 1 edition, [ISBN: 978- 0521137355], 2010.
7. Toby Velt, Anthony Velt, Robert Elsenpeter, “Cloud Computing, A Practical Approach” McGraw-Hill Osborne Media; 1 edition [ISBN: 0071626948], 2009.
8. Dimitris N. Chorafas, “Cloud Computing Strategies” CRC Press; 1 edition [ISBN: 1439834539],2010.
9. Peter Waher, “Learning Internet of Things”, PACKT publishing, BIRMINGHAM – MUMBAI
10. Bernd Scholz-Reiter, Florian Michahelles, “Architecting the Internet of Things”, ISBN 978-3-642-19156-5 e-ISBN 978-3-642-19157-2, Springer
11. Daniel Minoli, “Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications”, ISBN: 978-1-118- 47347-4, Willy Publications
12. Vijay Madiseti and ArshdeepBahga, “Internet of Things (A Hands-on- Approach)”, 1st Edition, VPT, 2014.

SDC5HN22P- PRATICAL I: HADOOP & NO SQL DATABASE

Number of Contact Hours: 75 Hrs.

Course Outline

PART A: HADOOP

List of Exercises:

1. Lab on Manipulating files in HDFS pragmatically using the File System API. Alternative Hadoop File Systems: IBM GPFS, MapR-FS, Lustre, Amazon S3 etc.
2. Lab on Write an Inverted Index MapReduce Application with custom Partitioner and Combiner Custom types and Composite Keys Custom Comparators Input Formats and Output Formats Distributed Cache MapReduce Design Patterns Sorting Joins.
3. Lab on Writing a streaming MapReduce job in Python YARN and Hadoop 2.0.
4. Lab on Importing data from an RDBMS to HDFS using Sqoop.
5. Lab on Exporting data from HDFS to an Other data integration tools: Flume, Kafka, Informatica, Talend etc.

PART B: NO SQL

List of Exercises:

1. Create a sample Database Student and create collection using MongoDB.
2. Create a MongoDB array with 5 fields.
3. Program to use find () in MongoDB
4. Write a program to create Cursor in MongoDB
5. Write a program to return the documents from a sample collection in ascending order
6. Consider a sample collection and count the number documents in a collection, delete a particular document
7. Consider a sample collection and update a particular document.
8. Write a program for Mongoddb authentication with Kerberos
9. Create a sample database and perform createIndex(), getindexes (), dropindex () operations
10. Consider a sample collection and retrieve documents based on a particular pattern.

SDC5ML23P- PRACTICAL VIII: MACHINE LEARNING PYTHON AND DATA VISUALISATION USING TABLEAU

Number of Credits-6

Course Outline

PART A: MACHINE LEARNING PYTHON

List of Exercises:

1. Plotting Cross Validated Predictions.
2. Concatenating multiple feature methods.
3. Pipelining: Chaining a PCA and a logistic regression.
4. Isotonic regression.
5. Inputting missing values before building an estimator.
6. Face completion with a multi-output estimators.
7. Selecting Dimensionality reduction with pipeline Grid search cv.

8. Multilabel classification
9. The Johnson-Lindenstrauss bound for embedding with random projections.
10. Comparison of kernel ridge regression and SVR,
11. Outlier detection on a real data set.
12. Faces recognition example using eigenfaces and SVMs.
13. Demonstrate how to generate a dataset and bicluster it using the Spectral Co-Clustering algorithm.
14. Demonstrate how to generate a checkerboard dataset and bicluster it using the Spectral Biclustering algorithm.
15. Demonstrate the Spectral Co-clustering algorithm on the twenty newsgroups dataset.
16. Comparison of Calibration of Classifiers.
17. Develop a program to display how well calibrated the predicted probabilities are and how to calibrate an uncalibrated classifier.
18. Develop a program to demonstrate the Probability calibration of classifiers.
19. Develop a program to demonstrate the Probability Calibration for 3-class classification.
20. Develop a program showing how the scikit-learn can be used to recognize images of hand-written digits.
21. Normal and Shrinkage Linear Discriminant Analysis for classification
22. Develop a program to implement Plot classification probability.
23. Develop a program to illustrate the nature of decision boundaries of different classifiers.
24. Develop a program to implement Linear and Quadratic Discriminant Analysis with covariance ellipsoid.
25. Develop a program to implement Feature agglomeration using clustering.
26. Develop a program to demonstrate the mean-shift clustering algorithm.
27. Develop a program to Demonstrate the k-means assumptions.
28. Develop a program to implement K-means Clustering.
29. Develop a program to implement Spectral clustering for image segmentation.
30. Develop a program to demonstrate Color Quantization using K-Means.
31. Develop a program to implement A demo of K-Means clustering on the handwritten digits data.
32. Develop a program to implement Sparse inverse covariance estimation.
33. Develop a program to implement The Digit Dataset.
34. Develop a program to plot randomly generated classification dataset.
35. Develop a program to demonstrate the Principal Component Analysis applied to the Iris dataset.
36. Develop a program to demonstrate Image denoising using dictionary learning.
37. Develop a program to demonstrate Decision Tree Regression with AdaBoost.
38. Develop a program to implement Pixel importance with a parallel forest of trees.
39. Develop a program to implement Feature importances with forests of trees.
40. Develop a program to implement Isolation Forest example.
41. Develop a program to Plot the decision boundaries of a Voting Classifier.

42. Develop a program to implement Prediction Intervals for Gradient Boosting Regression.
43. Develop a program for the Illustration of the effect of different regularization strategies for Gradient Boosting.
44. Develop a program to exercise regarding the use of classification techniques on the Digits dataset.
45. Develop a program to exercise Cross-validation with an SVM on the Digits dataset.
46. Develop a program to implement Cross-validation on diabetes Dataset Exercise.
47. Develop a program to demonstrate Recursive feature elimination.
48. Develop a program to implement Pipeline Anova SVM.
49. Develop a program to implement Gaussian process classification (GPC) on iris dataset.
50. Develop a program to implement the Comparison of kernel ridge and Gaussian process regression
51. Develop a program to Illustrate the Gaussian process classification (GPC) on the XOR dataset.
52. Develop a program to implement Logistic function.
53. Develop a program to illustrate the metric and non-metric MDS on generated noisy data.
54. Develop a program to Plot Validation Curves.
55. Demonstrate the resolution of a regression problem using a k-Nearest Neighbor and the interpolation of the target using both barycenter and constant weights.
56. Develop a program to implement the usage of Nearest Neighbors classification. It will plot the decision boundaries for each class.
57. Develop a program to implement the Visualization of MLP weights on MNIST.
58. Develop a program to implement Clustering text documents using k-means.
59. Develop a program to implement the Decision Tree Regression.
60. Develop a program to Perform binary classification using non-linear SVC with RBF kernel.

PART B: DATA VISUALISATION USING TABLEAU

List of Exercises:

1. To load and display dataset on tableau.
2. To join tables from single and multiple databases.
3. Implement Data Preparation using Blending
4. Implement Data Preparation using Metadata Grid.
5. Implement Data Preparation using Pivot.
6. Implement Data Preparation using Union.
7. Implement Data Preparation using Data Interpreter.
8. To build groups, hierarchies and sets.
9. Implement Sorting using measure.
10. Implement Sorting using header and legends.
11. Implement sorting using pill(use-case).
12. Implement Interactive Filter,
13. Implement Quick Filter.

14. Implement context filtering.
15. To plot a graph to show the Data in histogram.
16. To plot a bullet graph.
17. To plot a dual axis graph.
18. To plot an area graph with dual axis.
19. To show Data in TreeMap.
20. To build a Box plots.
21. To build a Gantt Bar Charts.
22. To build a Paretos.
23. To build a Sparklines.
24. Implement Grouping using Data Window and Visual Grouping.
25. Implement using all Functions (Number, String, Date, Logical, Aggregate).
26. Create quick table calculations.
27. Implement LOD calculations; types of LOD calculations.
28. Implement Ad-hoc calculations.
29. Implement mapping using Pan & Zoom.
30. Implement mapping using Filtering.
31. Implement mapping using Map layering.
32. Implement mapping using Custom territories.
33. Implement mapping using Lasso & Radial selection.
34. To modify locations within Tableau.
35. To import and manage custom geocoding.
36. Create a custom territory, building a polygon map.
37. To use a background image map
38. To use Geographic search.
39. Implement analytics using Reference Lines and Reference Bands.
40. Implement analytics using Trend Lines. and Trend Model.
41. Implement analytics using Forecasting using precise range.
42. Implement analytics using Drag & Drop Analytics and Box Plot.
43. Create a dashboard for publishing & sharing options.
44. Implement an interactive dashboard using actions (Filter, Highlight, URL).
45. Implement Device Designer to create a device.

Sl. No.	Job Role	Qualification Packs	NSQF Level
1	Junior Data Associate Alias – Data Scientist, Data Warehousing Engineer, Business Intelligence Engineer	SSC/Q0401	7
2			
3			

4			
5			

SEMESTER VI
PROJECT

Number of Contact Hours: 900 Hrs.