CSE3505 – FOUNDATIONS OF DATA ANALYTICS

Course Plan – Theory & Lab

S.No. (Lecture Session)	Title (Lecture)	Level of Delivery	Lab
1.	Analytics – What and Why? Analytics life cycle	Concept	
2.	Introduction to R, R Studio (GUI): R Windows Environment Introduction to R basic data types	C-I-P	
3.	Introduction to array, matrix	C-I-P	Session 1: Installation of R and RStudio Understanding R packages, their installation and management R program structure R Basic data types
4.	Business Analytics, Lending Analytics	Concept	
5.	Introduction to Data frame	C-I-P	
6.	Exploring data frame	C-I-P	Session 2: R control structure Data frame Working with different file types
7.	Recommendation Analytics, Healthcare Analytics	Concept	
8.	Functions and loops	C-I-P	
9.	Reading datasets, Working with different file types .txt,.csv	C-I-P	Session 3: Excel and R integration with R connector
10.	Financial Analytics, Sports Analytics		
11.	Combining Datasets in R	C-I-P	
12.	Extracting datasets	C-I-P	Session 4: Extracting and working on datasets
13.	Preparing datasets: Data Cleaning	C-I-P	
14.	Data imputation	C-I-P	
15.	Data conversion Analysis	C-I-P	Session 5: Preparing datasets-Cleaning, Missing value imputation, Data conversion
16.	Time management	Activity/Discussion	
17.	Basic statistics: mean, median	C-I-P	

18.	standard deviation, variance Summary Statistics - Summarizing data with R	C-I-P	Assessment - 1
19.	Work management & Prioritization	Activity/Discussion	
20.	correlation, covariance	C-I-P	
21.	Outliers	C-I-P	Session 6: Basic statistics with R
22.	Quality Adherence	Activity/Discussion	
22.	Correlation	C-I-P	
23.	Regression	C-I-P	Session 7: Analysing and Interpreting Datasets
24.	Team work	Activity/Discussion	
26.	Professionalism	Activity/Discussion	
27.	Effective Communication	Activity/Discussion	Assessment - 2
28.	Procedures, Guidelines, Purpose & Scope of documentation	Activity/Discussion	
29.	Structure of documents	Activity/Discussion	
30.	Tools for preparing Document	Activity/Discussion	

C – Concept; I – Implementation; P - Project

Resources for the course

- Every Lecture will be supported by the Presentation Slides uploaded in the LMS and Website
- Text Books
    1. Trevor Hastie and Rob Tibshirani, “An Introduction to Statistical Learning with Applications in R”, Springer, 2017.
    2. Mark van der Loo, Edwin de Jonge, “Learning R Studio for R Statistical Computing”, Packt Publishing, 2012.
    3. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. “Mining of Massive Datasets”. Cambridge University Press. 2014. Reference Books
    1. Hadley Wickham and Garrett Grolemund, “R for Data Science: Import, Tidy, Transform, Visualize, and Model Data”, O’Reilly, 2017.
    2. Grolemund, Garrett. “Hands-on programming with R”, O’ Reilly Media, Inc., 2014.
    3. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze, “Introduction to Information Retrieval”, Cambridge University Press, First South Asian Edition, 2008.
    4. Trevor Hastie, Robert Tibshirani, Jerome Friedman, “The Elements of Statistical Learning”, Springer, Second Edition, 2011.
    5. https://www.sscnasscom.com/qualification-pack/SSC/Q2101/

Theory Assessment

- Internal Assessment 60 Marks
    - ✓ Quiz (20 marks)
        - ➢ One quiz at the end of 3 modules
        - ➢ Every quiz shall have 10 questions

- Each quiz will have a weight of 10

- ✓ Challenging Assignment (10 marks)
- ✓ CAT-1 (15 marks)
- ✓ CAT-2 (15 marks)

Lab Assessment

- Internal Assessment 60 Marks
  - ✓ Regular Submission (30 marks)
  - ✓ Assessment - 1 (15 marks)
  - ✓ Assessment - 2 (15 marks)

Project Assessment

- Review I – 20 marks
- Review II – 30 marks
- Review III – 50 marks