DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE
(AUTONOMOUS)
(Approved by AICTE & Affiliated to Anna University, Chennai)
Re-Accredited by NAAC with 'A' Grade
Accredited by NBA for AERO, BME, CSE, ECE, EEE, IT & MECH.
PERAMBALUR-621212, TAMILNADU, INDIA.
Website: www.dsengg.ac.in

COURSE PLAN (25-26 EVEN SEM)

Course Code/Name	U23AIT42/ DATA SCEINCE			
Year/Section/Department	II/C/AI&DS			
Credits Details	**L:3**	**T:0**	**P:0**	**C:3**
Total Contact Hours Required	45			

Syllabus:

UNIT I INTRODUCTION TO DATASCIENCE	**No. of Periods 8**
Need for data science–benefits and uses–facets of data–data science process–setting the research goal – retrieving data – cleansing, integrating, and transforming data – exploratory data analysis – build the models –presenting and building applications	
UNIT II / DESCRIPTIVEANALYTICS	**No. of Periods1 0**
Frequency distributions – Outliers –interpreting distributions – graphs – averages - describing variability– inter quartile range–variability for qualitative and ranked data-Normal distributions– z scores–correlation–scatter plots–regression –regression line–least squares regression line –standard error of estimate–interpretation of r^2–multiple regression equations–regression towards the mean.	
UNIT III / INFERENTIALSTATISTICS	**No. of Periods 9**
Populations–samples–random sampling–Sampling distribution-standard error of the mean -Hypothesis testing–z-test–z-test procedure–decision rule–calculations–decisions–interpretations-one-tailed and two-tailed tests –Estimation–point estimate–confidence interval-level of confidence – effect of sample size.	
UNIT IV / ANALYSIS OF VARIANCE	**No. of Periods 9**
t-test for one sample–sampling distribution of t–t-test procedure–t-test for two independent samples–p-value–statistical significance–t-test for two related samples. F-test– ANOVA –Two-factor experiments – three f-tests–two-factor ANOVA–Introduction to chi- square tests.	
UNIT V / PREDICTIVEANALYTICS	**No. of Periods 9**
Linear least squares–implementation–goodness of fit–testing a linear model– weighted re sampling. Regression using Stats Models – multiple regressions – nonlinear relationships–logistic regression–estimating parameters–Time series analysis–moving averages –missing values –serial correlation–autocorrelation. Introduction to survival analysis.	

Objective:

- To understand the techniques and processes of data science
- To apply descriptive data analytics
- To visualize data for various applications
- To understand inferential data analytics
- To analysis and build predictive models from data

Text Book:

1. ChiragShah,"AHands-onIntroductiontoDataScience",CambridgeUniversityPress, 2020.
2. Sinan Ozdemir, "PrinciplesofDataScience",PacktPublication,2016
3. JulioCesarRodriguezMartino,"Hands-onMachineLearningwithMicrosoft Excel", Packt Publication, 2019.

Website:

www.kaggle.com
towardsdatascience.com
www.statmethods.net
www.datacamp.com
www.analyticsvidhya.com
www.coursera.org

Online Mode of Study:

NPTEL details can be listed.

- Business Intelligence & Analytics - Course
- Data Analytics with Python - Course
- Data Science for Engineers - Course

Course Plan:

Topic Number	Topic	Reference Detail	Page Number	Mode of teaching	Number of Periods Required	Cumulative Period
	UNIT I INTRODUCTION TO DATASCIENCE					**8**
1	Need for Data Science – Benefits and Uses	T1: Chirag Shah, T2: Sinan Ozdemir/	1.1 – 1.3	BB	1	1
2	Facets of Data – Data Science Process	T1: Chirag Shah, T2: Sinan Ozdemir/	1.5 –1.10	PPT	1	2
3	Setting the Research Goal	T1: Chirag Shah, T2: Sinan Ozdemir/	1.10 1.13	BB	1	3
4	Retrieving Data – Cleansing, Integrating, and Transforming Data	T1: Chirag Shah, T3: Julio Cesar Rodriguez/	1.14 – 1.25	Video	2	5
5	Exploratory Data Analysis – Build the Models	T1: Chirag Shah, T3: Julio Cesar Rodriguez/	1.27-1.36	PPT	1	6
6	Presenting and Building Applications	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez/	1.37	BB	2	8
	UNIT II: DESCRIPTIVE ANALYTICS					**10**
7	Frequency Distributions, Outliers, Interpreting Distributions	T2: Sinan Ozdemir, T3:	2.1-2.13	BB	2	11

		Julio Cesar Rodriguez/				
8	Graphs, Averages, Describing Variability	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez/	2.14-2.34	PPT	2	13
9	Inter quartile Range, Variability for Ranked Data	T3: Julio Cesar Rodriguez/	2.34-2.37	Video	2	15
10	Normal Distributions, Z-Scores, Correlation, Scatter plots	T1: Chirag Shah, T2: Sinan Ozdemir/	2.37-2.56	BB	2	17
11	Regression – Least Squares Regression Line	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez/	2.60-2.62	PPT	1	18
	UNIT III / INFERENTIALSTATISTICS					**9**
12	Populations, Samples, Random Sampling	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	3.1-3.6	BB	1	19
13	Sampling Distribution, Standard Error of the Mean	T1: Chirag Shah, T3: Julio Cesar Rodriguez	3.12-3.16	Video	1	20
14	Hypothesis Testing – Z-Test – Procedure	T1: Chirag Shah, T2: Sinan Ozdemir	3.8-3.21	PPT	1	21
15	Hypothesis Testing – Calculations, Decision Rules	T1: Chirag Shah, T2: Sinan Ozdemir	3.21-3.23	BB	1	22

16	One-Tailed and Two-Tailed Tests – Applications	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	3.24-3.27	BB	1	23
17	Estimation – Point Estimate, Level of Confidence	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	3.30-3.31	Video	1	24
18	Effect of Sample Size on Estimation	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	3.35-3.37	BB	1	25
19	Case Study on Inferential Statistics	T1: Chirag Shah, T3: Julio Cesar Rodriguez	3.37-3.38	Video	1	26
20	Advanced Topics in Inferential Statistics	T1: Chirag Shah, T3: Julio Cesar Rodriguez	3.38-3.39	PPT	1	27
	UNIT IV / ANALYSIS OF VARIANCE				**9**	
21	T-Test for One Sample, Sampling Distribution of T	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	4.1-4.5	BB	1	28
22	T-Test for Two Independent Samples, P-Value	T1: Chirag Shah, T3: Julio Cesar Rodriguez	4.5-4.12	Video	1	29
23	T-Test for Two Related Samples	T1: Chirag Shah, T2: Sinan Ozdemir	4.20-4.24	PPT	2	31
24	F-Test for Variance	T2: Sinan Ozdemir, T3:	4.24-4.30	BB	1	32

		Julio Cesar Rodriguez				
25	ANOVA – One-Way ANOVA	T1: Chirag Shah, T2: Sinan Ozdemir	4.30-4.33	PPT	2	34
26	ANOVA – Two-Factor ANOVA	T3: Julio Cesar Rodriguez	4.34-4.36	Video	1	34
27	Introduction to Chi-Square Tests – Applications	T3: Julio Cesar Rodriguez	4.43-4.49	BB	1	36
	UNIT V / PREDICTIVEANALYTICS				**9**	
28	Linear Least Squares Regression, Implementation	T1: Chirag Shah, T3: Julio Cesar Rodriguez	5.1-5.9	BB	1	37
29	Goodness of Fit, Testing a Linear Model	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	5.12-5.16	PPT	2	39
30	Weighted Resampling, Regression Using Stats Models	T1: Chirag Shah, T2: Sinan Ozdemir	5.19-5.22	Video	1	40
31	Nonlinear Relationships, Logistic Regression	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	5.26-5.28	BB	1	41
32	Time Series Analysis – Moving Averages	T1: Chirag Shah, T3: Julio Cesar Rodriguez	5.32-5.39	PPT	1	42

33	Handling Missing Values, Serial Correlation	T2: Sinan Ozdemir, T3: Julio Cesar Rodriguez	5.40-5.41	Video	1	43
34	Autocorrelation – Concepts and Applications	T1: Chirag Shah, T3: Julio Cesar Rodriguez	5.43-5.44	BB	1	44
35	Introduction to Survival Analysis	T1: Chirag Shah, T3: Julio Cesar Rodriguez	5.44-5.61	PPT	1	45

Course Outcome:

- Understand the real-world data and information
- Apply data science using excel & Python
- Design of mathematical model for problem solving
- Interpret various tools and its advantages.
- Illustrate the different opportunities in industries.
- Apply data modeling for real-world applications.

Course Outcome Vs Program Outcome Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	2	2	1	1	2	1	2	1	3	2	2	1
CO2	3	2	1	2	2	1	1	1	1	2	1	3	3	2	1
CO3	3	3	1	3	2	1	1	1	1	2	1	3	3	3	2
CO4	3	3	2	3	2	1	1	1	1	2	1	3	3	3	2
CO5	3	2	3	3	3	1	1	1	2	3	2	3	3	2	3
AVG	**3.00**	**2.40**	**1.60**	**2.60**	**2.20**	**1.00**	**1.00**	**1.20**	**1.20**	**2.20**	**1.20**	**3.00**	**2.80**	**2.40**	**1.80**

Content beyond Syllabus:

- Advanced Visualization Techniques
- Ethics and Bias in Data Science

Internal Evaluation Components:

Web portal	Assignment	Components	Topic Number with Topic / Unit Details	Relevance to CO
Web portal 1	--	Assessment – I (60)	**Unit I: Introduction to Data Science & Unit II: Descriptive Analytics**	CO 1 & CO2
	1	Assignment – Handwritten (20)	2. Data science process, 5. Exploratory data analysis, 11. Regression line applications	CO 1 & CO2
	2	Assignment – Poster Presentation / PPT (20)	7. Frequency Distributions, Outliers, Interpreting Distributions 8. Graphs, Averages, Describing Variability	CO 1 & CO2
Web portal 2	--	Assessment – II (60)	**Unit III: Inferential Statistics and Unit IV: Analysis of Variance**	CO3 & CO4
	3	Seminar (20)	14. Hypothesis Testing – Z-Test – Procedure	CO3 & CO4
	4	Case Study Report (20)	25. ANOVA – One-Way ANOVA	CO3 & CO4
Web portal 3	--	Model Exam (75)	UNIT I to V	CO1 to CO6
	5	MCQ (15)	UNIT I to V	CO1 to CO6
	-	Course Attendance (10)	--	--

Submission Details:

Phase 1(Before AT 1)		Phase 2 (Before AT 2)		Phase 3 (Model)
Assignment 1	**Assignment 2**	**Assignment 3**	**Assignment 4**	**Assignment 5**

Google Class Code Details: j4jhne7

Class Name: **Fourth Sem**

PLAN OF ASSESSMENT TEST –DISTRIBUTION OF MARKS:

TEST	**CO- MARK WISE DISTRIBUTION**						**BLOOM'S LEVEL MARK WISE DISTRIBUTION**					
AT-1	**CO 1**	**CO 2**	**CO 3**	**CO 4**	**CO 5**	**CO 6**	**BTL 1**	**BTL 2**	**BTL 3**	**BTL 4**	**BTL 5**	**BTL6**
AT-2	**CO 1**	**CO 2**	**CO 3**	**CO 4**	**CO 5**	**CO 6**	**BTL 1**	**BTL 2**	**BTL 3**	**BTL 4**	**BTL 5**	**BTL6**
MODEL	**CO 1**	**CO 2**	**CO 3**	**CO 4**	**CO 5**	**CO 6**	**BTL 1**	**BTL 2**	**BTL 3**	**BTL 4**	**BTL 5**	**BTL6**

Prepared By **Verified By**

Approved By
PRINCIPAL