Scheme for Honours in Data Science (Computer Engineering) (Autonomous) (DJS23 Scheme)

Sr. No.	Course Code	Course	Teaching Scheme				Semester End Examination (SEE)						Continuous Assessment (CA)			Aggregate (A + B)	Credits Earned	
			Theory (Hrs)	Practical (Hrs)	Tutorial (Hrs)	Credits	Duration (Hrs)	Theory	Oral	Practical	Oral & Practical	SEE (Total)	Theory	Term Work Total	CA Total			
Semester III																		
1	DJS23CCH1301	Statistics for Data Science	3			3	2	60				60	40		40	100	3	3
Semester IV																		
2	DJS23CCH1401	Visualization in Data Science	3			3	2	60				60	40		40	100	3	4
	DJS23CLH1401	Visualization in Data Science Laboratory		2		1			25			25		25	25	50	1	
Semester V																		
3	DJS23CCH1501	Graph Data Science	3			3	2	60				60	40		40	100	3	4
	DJS23CLH1501	Graph Data Science Laboratory		2		1			25			25		25	25	50	1	
Semester VI																		
4	DJS23CCH1601	Essentials of Generative AI and Prompt Engineering	3			3	2	60				60	40		40	100	3	3
Semester VII																		
5	DJS23CCH1701	Cloud Computing for Data Analysis	3			3	2	60				60	40		40	100	3	4
6	DJS23CLH1701	Cloud Computing for Data Analysis Laboratory		2		1			25			25		25	25	50	1	
		Total	15	6		18	10	300	75			375	200	75	275	650	18	18

Program: Computer Engineering Honors in Data Science		S.Y B. Tech.	Semester: III
Course Code: DJS23CCH1301			
Course: Statistics for Data Science			

Pre-requisite: Knowledge of
1. Basics of probability

Objectives:
1. To provide students with a solid understanding of statistical concepts and techniques essential for data science.
2. To develop students' ability to apply statistical methods to real-world data sets and draw meaningful insights.
3. To equip students with the skills to use statistical software tools for data analysis and interpretation.
4. To foster critical thinking and problem-solving skills in the context of statistical analysis for data science.

Outcomes: On completion of the course, the learner will be able to:
1. Describe and summarize data using appropriate statistical measures.
2. Apply correlation analysis and regression to real-world problems.
3. Apply regression techniques to model and forecast time series data.
4. Apply nonparametric test and draw appropriate conclusions for data analysis.

Course name (course code)		
Unit	**Description**	**Duration**
1	**Introduction to Data and Statistics**: Elements, Variables, and Observations, Scales of Measurement, Qualitative and Quantitative Data, Cross-Sectional and Time Series Data, Descriptive Statistics, Statistical Inference, Summarizing Qualitative and Quantitative Data using Tables and Graphs, Cross Tabulations and Scatter Diagram, Measures of Central Tendency, Measures of Dispersion, Skewness, Moments and Kurtosis	08
2	**Correlation Analysis:** Types of correlation, Karl Pearson's Coefficient of Correlation, Coefficient of Determination, Rank Correlation Coefficient, Coefficient of Concurrent Deviation	06
3	**Regression:** **Simple Linear Regression**: Concept of linear relationship between variables Least squares method, Estimating the coefficients (slope and intercept), Assessing the goodness of fit (R-squared), Interpretation of coefficients **Multiple Linear Regression**: Extending simple linear regression to multiple predictors, Estimating the coefficients, Interpreting the coefficients, Assessing the model fit (adjusted R-squared), Multicollinearity and its implications, Heteroscedaticity	08

	Logistic Regression: Modeling binary outcomes, Odds ratios and log odds, Interpreting the coefficients, Assessing model fit (deviance, likelihood ratio test)	
4	**Hypothesis Testing in Regression**: Significance testing for regression coefficients, t-tests and p-values, Confidence intervals for coefficients, F-test for overall model significance	04
5	**Regression for Time Series Data**: Autocorrelation and its impact on regression, Autoregressive models (AR), Moving average models (MA), Autoregressive integrated moving average models (ARIMA)	05
6	**Non-Parametric Statistics**: Nonparametric Methods, Sign Test, Wilcoxon Signed-Rank Test, MannWhitney-Wilcoxon Test, Kruskal Wallis Test, Spearman's Rank Correlation, Runs Test	08
	Total	**39**

Books Recommended:

Textbooks:
1. Dr. S. P. Gupta, "Statistical Methods", 46th Edition, S. Chand and Sons, 2021.
2. James T. McClave, P George Benson, Terry T Sincich, "Statistics for Business and Economics", 14th Edition, Pearson, 2021.

Reference Books:
1. Maurits Kaptein, Edwin van den Heuvel, "Statistics for Data Scientists: An Introduction to Probability, Statistics, and Data Analysis", Springer, 2022.
2. Peter Bruce, Andrew Bruce, "Practical Statistics for Data Scientists", O'Reilly, 2017.
3. C.B. Gupta & Vijay Gupta, "An Introduction to Statistical Methods", 23rd Edition, S. Chand and Sons, 2004.

Prepared by   Checked by   Head of the Department   Principal

Program: Computer Engineering Honors in Data Science	S.Y B. Tech.	Semester: IV
Course: Visualization in Data Science (**DJS23CCH1401**)		
Course: Visualization in Data Science Laboratory (DJS23CLH1401)		

Pre-requisite: Knowledge of
   1. Basics of probability

Objectives:
1. Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories)
2. Learning to analyze and critically evaluate ideas, arguments, and points of view.
3. Learning appropriate methods for collecting, analyzing, and interpreting numerical information
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course
Outcomes: On completion of the course, the learner will be able to:
1. Understand the fundamental design principles and different types of data visualization.
2. Identify both positive and negative impacts of data-informed decision across a variety of domains.
3. Apply the fundamental concepts of data visualization to define a project in your field of study.
4. Practice the core principles using widely available tools (e.g. Tableau).
5. Demonstrate the best practice that presents your story in the process of creating data visualization including connecting to different data sources, assessing to the quality of the data, and converting raw data into data visualizations that provide actionable information.

Course name (course code)		
Unit	**Description**	**Duration**
1	**Introduction to Data Visualization** Importance of Data Visualization - History of Data Visualization - Basic Principles and Concepts - Tools and Technologies Overview Exploring Data with R and Python - Introduction to R for Data Visualization - Introduction to Python (matplotlib, seaborn) - Data Cleaning and Preparation	08
2	**Advanced Visualization Techniques** - Advanced Charts: Histograms, Heatmaps, Box Plots - Multivariate Data Visualization - Geospatial Data Visualization	06
3	**Interactive Dashboards and Storytelling** - Principles of Dashboard Design - Tools for Creating Dashboards (Tableau, Power BI)	08

	- Data Storytelling Techniques Data Sources and Data Wrangling - Finding and Collecting Data - APIs and Web Scraping - Data Wrangling Techniques	
4	**Visual Perception and Design Principles** - Understanding Human Perception in Visualization - Principles of Design in Data Visualization - Color Theory and Its Application in Visualizations	04
5	**Network and Graph Visualization** - Principles of Network Visualization - Tools and Techniques for Graph Visualization **Time Series Data Visualization** - Techniques for Visualizing Time Series Data - Tools and Libraries (e.g., matplotlib, Plotly)	05
6	**Data Ethics and Privacy** - Ethical Considerations in Data Visualization - Privacy and Data Protection - Misleading Visualizations and Avoiding Bias **Evaluating and Improving Visualizations** - Critiquing and Improving Visualizations - Best Practices and Common Mistakes - User Experience in Visualization **Special Topics in Data Visualization** - Visualization for Different Domains (Healthcare, Finance, etc.) - Emerging Trends and Technologies	08
	Total	**39**

Books Recommended:

Textbooks:
1. "The Visual Display of Quantitative Information" by Edward Tufte
2. "Storytelling with Data" by Cole Nussbaumer Knaflic

Reference Books:
1. The Visual Display of Quantitative Information (2nd ed.), Edward Tufte (Read pp. 53-77 & 107-121)
2. The Elements of Graphing Data (1985) by William S. Cleveland (Read pp. 24-55 & 68-88)

3. The Truthful Art: Data, Charts, and Maps for Communication by Alberto Cairo (Read pp. 41-65 & 121- 149)

Online material :
• Designing Great Visualizations:
https://www.tableau.com/sites/default/files/media/designinggreat-visualizations.pdf
One Dataset, Visualized 25 Ways: https://flowingdata.com/2017/01/24/one-datasetvisualized-25-ways/
Tableau tutorial videos Getting started (25 mins):
https://www.tableau.com/learn/tutorials/ondemand/gettingstarted?playlist=554268&signin=4bec0 5bc7a876b95af77 22b08fff9224
& Practice]
 • Tableau how-to videos: https://public.tableau.com/enus/s/resources (31 mins)

Experiment list:

1	Basic visualizations using Excel and Google Charts.	
2	Cleaning and visualizing a dataset.	
3	Advanced visualizations with Tableau and Python.	
4	Creating an interactive dashboard.	
5	Data collection and wrangling in Python.	
6	Designing effective visualizations using color and design principles	
7	Case studies on ethical issues in data visualization.	
8	Evaluating visualizations for ethical concerns.	
9	Network visualization using Gephi.	
10	Time series visualization in Python.	
11	Peer review of visualizations And Improving a given set of visualizations	
12	Mini project :Domain-specific visualization project(Healthcare, Finance, etc.)	

Prepared by   Checked by   Head of the Department   Principal