Course	CSD101 Fundamentals of Data Science	Semester	Monsoon Semester 2024
Faculty Name(s)	Dinesh Barot, Jinal Parikh, Kunjal Gajjar, Kuntalkumar Patel, Mitaxi Mehta, Najma Barkat, Shefali Naik, Vinay Vachharajani	Contact	dinesh.barot@ahduni.edu.in, jinal.parikh@ahduni.edu.in, kunjal.gajjar@ahduni.edu.in, kuntal.patel@ahduni.edu.in, mitaxi.mehta@ahduni.edu.in, najma.barkat@ahduni.edu.in, shefali.naik@ahduni.edu.in, vinay.vachharajani@ahduni.edu.in
School	SEAS	Credits	3
GER Category:	Mathematical and Physical Sciences	Teaching Pedagogy Enable:NO	P/NP Course: Can not be taken as P/NP

Schedule

Section 1	08:00 am to 09:30 am	Tue	29-07-24 to 26-11-24
	08:00 am to 09:30 am	Thu	29-07-24 to 26-11-24
Section 2	08:00 am to 09:30 am	Mon	29-07-24 to 26-11-24
	08:00 am to 09:30 am	Fri	29-07-24 to 26-11-24
Section 4	08:00 am to 09:30 am	Tue	29-07-24 to 26-11-24
	08:00 am to 09:30 am	Thu	29-07-24 to 26-11-24
Section 5	08:00 am to 09:30 am	Sat	29-07-24 to 26-11-24
	09:30 am to 11:00 am	Sat	29-07-24 to 26-11-24
Section 8	08:00 am to 09:30 am	Fri	29-07-24 to 26-11-24
	08:00 am to 09:30 am	Mon	29-07-24 to 26-11-24
Section 3	11:00 am to 12:30 pm	Sat	29-07-24 to 26-11-24
	01:00 pm to 02:30 pm	Sat	29-07-24 to 26-11-24
Section 10	04:00 pm to 05:30 pm	Tue	29-07-24 to 26-11-24
	04:00 pm to 05:30 pm	Thu	29-07-24 to 26-11-24
Section 6	04:00 pm to 05:30 pm	Mon	29-07-24 to 26-11-24
	04:00 pm to 05:30 pm	Fri	29-07-24 to 26-11-24
Section 7	04:00 pm to 05:30 pm	Tue	29-07-24 to 26-11-24
	04:00 pm to 05:30 pm	Thu	29-07-24 to 26-11-24

Section 9	04:00 pm to 05:30 pm	Mon	29-07-24 to 26-11-24
	04:00 pm to 05:30 pm	Fri	29-07-24 to 26-11-24

Prerequisite	Not Applicable
Antirequisite	CSD100 Introduction to Data Science OR CSD102 Data Science or CSD102 Advanced Level Data Science
Corequisite	Not Applicable
Course Description	Data science is an interdisciplinary area that involves recording, storing and analyzing data to gain insights and knowledge for decision making. This is an intermediate level course providing foundation in data science and programming for first year undergraduate students. The course covers data science process and its life cycle, data collection using sampling/surveys, ordering/organizing, statistical analysis and visualization of data. Cases, examples and practical applications of data science are discussed using spreadsheet and python programming.
Course Objectives	The main objectives of course are to make students • Familiar with concepts of statistics useful in data science. • Learn tools for data analysis such as MS Excel, Tableau and Piktochart. • Acquaint with basics of GIS and GIS mapping. • Understand methods of data collection, data organization, data analysis and data presentation. • Study basics of Python programming.
Learning Outcomes	Upon the completion of this course, the students will be able to : • Create data file, clean and organize the data, analyze data and visualize the data in Excel and Tableau. • Prepare GIS maps using the tool Quantum GIS. • Prepare infographics using the tool Piktochart. • Do statistical analysis and write small programs using Python.

Pedagogy	Classroom Teaching: Students will be taught practical implementation of statistical methods used for data science. Flipped Classroom: Classroom teaching that will be flipped with practical demonstration. Activity Based Learning: Concept wise interesting activities will be given to individuals or groups which they can complete within few days.
Expectation From Students	• Interactive during the sessions • Curious to learn new concepts • Readiness for hands on using various tools and for programming • Read prescribed books, reference books and reading material • Submit assignments and projects on time

Assessment/Evaluation	• Mid-Semester Examination:
	  ◦ Written - 20%
	• End Semester Examination:
	  ◦ Written - 30%
	• Other Components:
	  ◦ Assignment (Statistics) - 15%
	  ◦ Assignment (Computer Science) - 15%
	  ◦ Infographics Project - 20%

Attendance Policy	As per Ahmedabad University Policy.
Project / Assignment Details	• Assignments • Small group/individual projects/activities

Course Material	Text Book(s) • Business Statistics, J. Joseph Francis, Second Edition, Pearson Education, ISBN: 978-98-5350-219-5, Year: 2024, Reference Book • Statistics, David Freedman, Robert Pisani and Roger Purves, Fourth Edition, W. W. Norton & Company Ltd., • MS Excel 2013 Bible, John Walkenbach, First Edition, Wiley and Sons Inc., • Step by Step Microsoft Excel 2013, Curtis D. Frye, Microsoft Press, Year: 2013, • Learning QGIS, Anita Graser, Third Edition, Packt Publishing, Year: 2016, Coursepacks • Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description, • Course Pack for Computer Science related sessions, Description, • Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description, • Course Pack for Computer Science related sessions, Description, • Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description,
Additional Information	Students are expected to bring laptops during the sessions.

Session Plan

NO.	TOPIC TITLE	TOPIC & SUBTOPIC DETAILS	READINGS,CASES,ETC.	ACTIVITIES	IMPORTANT DATES
1	Introduction to key terminology in data science	Data Science terminology Data Science in different domains Concepts of Big Data and Data Mining Importance of Datasets.	Web Ref : https://machinelearning-blog.com, https://www.edureka.co/blog/what-is-data-science	Discussion of Course Objectives and Expectations,Discussions on basic concepts of Data Science and its applications	
2	Introduction to key terminology in data science	Relationship between Artificial Intelligence, Machine and Deep Learning, Data Analysis Process.	Teaching Notes from the Book: Machine Learning using Python (2019), M Pradhan, UD Kumar Ch. 1 Introduction to Machine Learning Web Ref : https://machinelearning-blog.com	Discussion on importance of data, data life cycle and its applications through Machine and Deep Learning.	
3	Introduciton to geographical infromation systems	Concepts of GIS and Maps, Examples of GIS Applications, Vector and Raster data, Overview of Projection	Teaching Notes from the Book: Practical GIS (2017), Gábor Farkas, ISBN : 9781787123328	Demonstration of vector and raster data from various sources and understanding the formats	
4	Introduciton to geographical infromation systems	Understanding Attribute Table, Geo-referencing	Teaching Notes from the Book: Practical GIS (2017), Gábor Farkas, ISBN : 9781787123328	1. Demonstration of geo referencing on the map of Gujarat 2. Students will perform an exercise for the given map of India	
5	Introduciton to geographical infromation systems	Creating vector point, vector line and polygon data on map	Teaching Notes from the Book: Practical GIS (2017), Gábor Farkas, ISBN : 9781787123328	Demonstration of plotting points, lines and polygons on the map to represent Cities, Highways and Lakes respectively.	

6	Introduciton to geographical infromation systems	Examples of Creating vector point, vector line and polygon data on map	Teaching Notes from the Book: Practical GIS (2017), Gábor Farkas, ISBN : 9781787123328	Students will perform an exercise on the given map of India	
7	Types of data, scales of measurement, and methods for collection	Categorical/Qualitative data, Numerical / Quantitative data, Nominal, Ordinal, Interval and Ratio scales of data			
8	Types of data, scales of measurement, and methods for collection	Collecting raw data, arranging data using arrays, frequency tables, Grouping data			
9	Data cleaning using MS Excel	Whats' and 'Whys' of Data Cleaning Methods		Classroom teaching with examples in Excel	
10	Data cleaning using MS Excel	Data Cleaning using Excel - functions and tactics	https://support.microsoft.com/en-us/office/top-ten-ways-to-clean-your-data-2844b620-677c-47a7-ac3e-c2e157d1db19	Demonstration of the functions and exercise for students on a given dataset	
11	Simple descriptive statistics	Measures of central tendency: Arithmetic mean, Weighted mean, Geometric mean, Median, Mode	*Teaching Notes • Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
12	Simple descriptive statistics	Measures of Dispersion: Range, Standard deviation, Variance, Mean Absolute deviation, Mean Absolute deviation from the median, Quartiles, Deciles, Percentiles, Interquartile range, Quartile deviation, Coefficient of range, Coefficient of quartile deviation, Coefficient of variation, Coefficient of dispersion, Box-plots	• Teaching Notes • Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	

13	Simple descriptive statistics	Measures of Dispersion: Range, Standard deviation, Variance, Mean Absolute deviation, Mean Absolute deviation from the median, Quartiles, Deciles, Percentiles, Interquartile range, Quartile deviation, Coefficient of range, Coefficient of quartile deviation, Coefficient of variation, Coefficient of dispersion, Box-plots	• Teaching Notes • Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
14	Simple descriptive statistics	Measure of Divergence from Normality: Skewness and Kurtosis	• Teaching Notes • Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
15	Data visualisation using MS Excel and other software	Types of graphical displays available to visualize the data - Charts and graphs	https://www.juiceanalytics.com/writing/reading-visualizations-for-beginners, https://queue.acm.org/detail.cfm?id=1805128	Hands-on application to real-time datasets	
16	Data visualisation using MS Excel and other software	Choosing the type of Graphical display and Modifiying the default graphical displays	https://datajournalism.com/read/handbook/one/understanding-data/using-data-visualization-to-find-insights-in-data	Hands-on application to real-time datasets	
17	Data visualisation using MS Excel and other software	Creating a Data Dashboard for story telling	https://www.toptal.com/designers/data-visualization/data-visualization-best-practices	Hands-on application to real-time datasets	
18	Introduction to computer programming	· Introduction to Problem solving and algorithmic thinking \n· Problem solving using Pseudocode, Flowchart and Algorithms	· Teaching Notes	Examples of problem solving,\n Writing Pseudocodes, Drawing Flowcharts	
19	Introduction to computer programming	· Understanding Python Programming Environment\n· Basic syntax and key terminologies\n· Data types	· Teaching Notes\n · docs.python.org/3/	Classroom discussions,\n Activity based on simple python programs	

20	Introduction to computer programming	Python libraries\n· Pandas Series\n· Pandas Data Frames objects	· Python Data Science Handbook by Jake VanderPlas, Ch-3 Data Manipulation with Pandas	Demonstrations - creating series and data frames using Pandas	
21	Introduction to computer programming	Working with datasets\n · Importing data\n · Handling missing data\n · Combine, Concatenate, Append	· Jake VanderPlas, Ch-3 Data Manipulation with Pandas	Classroom teaching, Applying various operations on data sets using Pandas	
22	Introduction to computer programming	Data Visualization using Matplotlib\n · Line plots, Scatter plots, Histograms\n · Customizing plots	· Jake VanderPlas, Ch-4 Visualization with Matplotlib	Classroom teaching, Generating plots/charts using Matplotlib	
23	Infographics project	Developing a story with a strong narrative supported by various graphics. Projected to be selected based on the Foundation Studio.			
24	Infographics project				
25	Infographics project				
26	Infographics project				