Ahmedabad University

Course	CSD100 Introduction to Data Science	Semester	Monsoon Semester 2024
Faculty Name(s)		Contact	
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			
Prerequisite	Not Applicable		
Antirequisite	CSD101 Fundamentals of Data Science or CSD101 Intermediate Level 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 introductory course provides a foundation in data science for first year undergraduate students. The course covers data science process and its life cycle, data collection using sampling/surveys, ordering/organizing, data processing and visualization of data through charts and maps. Statistical fundamentals needed for analysis and interpretation of data are covered along with cases and examples related to real life applications of data science.		
Course Objectives	- To expose students to different types of data, its collection, cleaning, organizing (including online data repositories). - To expose students to basic statistical analysis that leads to descriptive summaries of datasets using numbers and graphics. - To expose students to the basics of geographical information systems . - To expose students to the basics of computer programming to develop logically thinking and algorithmic skills.		

Learning Outcomes	Upon the completion of this course, the students will be able to : • clean and organize datasets using MS Excel. • apply appropriate statistical analysis to be summarise its key characteristics of produce lucid visualisations. • tell a story from the analysis using infographics. • create thematic spatial maps to represent data using GIS software. • write short computer programs in Python.
Pedagogy	• Classroom Teaching: Students will be taught the practical implementation of statistical methods used for data science. • Flipped Classroom: Classroom teaching that will be flipped with a practical demonstration as and when required. • Activity-Based Learning: Concept wise interesting activities will be given to individuals or groups which they can complete within a few days. • Theoretical concepts to be explained with in-class examples from various disciplines to encourage interdisciplinary learning. • More hands-on time to be spent on software application in the class.
Expectation From Students	• Interactive during the sessions • Curious to learn new concepts • Readiness for hands on using various tools • 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 (Computer Science) - 10%
	  ○ Assignment (Statistics) - 20%
	  ○ Infographics Project - 20%

Attendance Policy	As per Ahmedabad University Policy.
Project / Assignment Details	• Assignments to be given to supplement in-class problem-solving skills. • Small group/individual projects/activities

Course Material	Reference Book
	- MS Excel 2013 Bible, John Walkenbach, First Edition, Wiley and Sons Inc., ISBN: 9788126541720, - Step by Step Microsoft Excel 2013, Curtis D. Frye, Microsoft Press, ISBN: 9780735681019, - Learning QGIS, Anita Graser, Third Edition, Packt Publishing, ISBN: 978-1-78588-033-9, Year: 2016, - Practical GIS: Learn novice to advanced topics such as QGIS, Spatial data analysis, and more, Gabor Farkas, First Edition, Packt Publishing, ISBN: 978-1-78712-332-8, Year: 2017, - MS Excel 2013 Bible, John Walkenbach, First Edition, Wiley and Sons Inc., ISBN: 9788126541720, - Step by Step Microsoft Excel 2013, Curtis D. Frye, Microsoft Press, ISBN: 9780735681019, - Learning QGIS, Anita Graser, Third Edition, Packt Publishing, ISBN: 978-1-78588-033-9, Year: 2016, - Practical GIS: Learn novice to advanced topics such as QGIS, Spatial data analysis, and more, Gabor Farkas, First Edition, Packt Publishing, ISBN: 978-1-78712-332-8, Year: 2017, - MS Excel 2013 Bible, John Walkenbach, First Edition, Wiley and Sons Inc., ISBN: 9788126541720, - Step by Step Microsoft Excel 2013, Curtis D. Frye, Microsoft Press, ISBN: 9780735681019, - Learning QGIS, Anita Graser, Third Edition, Packt Publishing, ISBN: 978-1-78588-033-9, Year: 2016, - Practical GIS: Learn novice to advanced topics such as QGIS, Spatial data analysis, and more, Gabor Farkas, First Edition, Packt Publishing, ISBN: 978-1-78712-332-8, Year: 2017, Other Course Material • QGIS User Guide, Release 2.18, October 2018 , • QGIS User Guide, Release 2.18, October 2018 , • QGIS User Guide, Release 2.18, October 2018 , • QGIS User Guide, Release 2.18, October 2018 , Coursepacks - Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description , - Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description, - Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description, - Statistics for Management, Richard Levin, David Rubin, Masood Siddiqui and Sanjay Rastogi, Eighth, Pearson Education, Description,

| Additional Information | This is a GER course offered to first-year students of the University. |

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	Introduction to geographical information systems	Concepts of GIS and Maps, Examples of GIS Applications, Vector and Raster data	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	Introduction to geographical information systems	Concept of 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	Introduction to geographical information systems	Understanding Attribute Table, Creating vector point data on map	Teaching Notes from the Book: Practical GIS (2017), Gábor Farkas, ISBN : 9781787123328	Demonstration of plotting points on the map to represent cities	
6	Introduction to geographical information systems	Examples of Creating vector point 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	Levin, Chapter-2 Grouping and Displaying Data to Convey Meaning: Tables and Graphs	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
8	Types of data, scales of measurement, and methods for collection	Collecting raw data, arranging data using arrays, frequency tables, Grouping data	Levin, Chapter-2 Grouping and Displaying Data to Convey Meaning: Tables and Graphs	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
9	Data cleaning using MS Excel	What's and Why's of Data Cleaning Methods of Data Cleaning		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	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 central tendency: Arithmetic mean, Weighted mean, Geometric mean, Median, Mode	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	Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	Hands-on application to re• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excelal-time datasets	

14	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	Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
15	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	Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
16	Mid Semester Examination				
17	Simple Descriptive Statistics	Measure of Divergence from Normality: Skewness and Kurtosis	Levin, Chapter-3 Measures of Central Tendency and Dispersion in Frequency Distributions	• Classroom teaching • Classroom discussion • Problem solving • Practical Demonstration using Excel	
18	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	
19	Data visualisation using MS Excel and other software	Choosing the type of Graphical display and Modifying 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	
20	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	

21	Introduction to computer programming	Introduction to Problem solving and algorithmic thinking \ Problem solving using Pseudocode, Flowchart and Algorithms	Teaching Notes	Examples of problem solving, Writing Pseudocodes, Drawing Flowcharts	
22	Introduction to computer programming	· Understanding Python Programming Environment\· Basic syntax and key terminologies\ Data types - Numbers and Strings	· Teaching Notes · docs.python.org/3/	Classroom discussions, Activity based on python numbers and string processing programs	
23	Introduction to computer programming	· Introduction to Python Pandas\ Pandas Series and Data Frames\Importing data\ Handling missing data\Data Visualization using Matplotlib	Python Data Science Handbook by Jake VanderPlas, Ch-3 Data Manipulation with Pandas	Demonstrations · creating series and data frames using Pandas · Generating plots using Matplotlib	
24	Infographics project	Developing a story with a strong narrative supported by various graphics. Project to be selected based on the Foundation Studio			
25	Infographics project	Developing a story with a strong narrative supported by various graphics. Project to be selected based on the Foundation Studio			
26	Infographics project	Developing a story with a strong narrative supported by various graphics. Project to be selected based on the Foundation Studio			
27	Infographics project	Developing a story with a strong narrative supported by various graphics. Project to be selected based on the Foundation Studio			
28	Review & Reflections				
29	Review & Reflections				
30	End Semester Examination				