

This document is prepared from the following University Notifications

- https://www.du.ac.in/uploads/new-web/15092023_Indis_sem1.pdf
- https://www.du.ac.in/uploads/new-web/notifications-2021/28032023_nep-Faculty%20of%20Interdisciplinary%20&%20Applied%20Sciences.pdf
- https://www.du.ac.in/uploads/new-web/15092023_Indis_sem3.pdf
- https://www.du.ac.in/uploads/new-web/18092023_Inter_4.pdf

Essential/recommended readings

1. R. L. Boylestad & Louis Nashlesky (2007), Electronic Devices & Circuit Theory, Pearson Education.
2. David A. Bell (2008), Electronic Devices and Circuits, Oxford University Press.
3. B L Theraja and AK Theraja, A Textbook Of Electrical Technology - Vol I.

Suggestive readings

1. Donald A. Neamen, Electronic Circuit Analysis and Design, Tata McGraw Hill (2002)

GENERIC ELECTIVES (GE-2): Data Engineering and Analytics

Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course
		Lecture	Tutorial	Practical/Practice		
Data Engineering and Analytics ELGE-1B	4	3	0	1	None	Basic Knowledge of Python Programming Language

Learning Objectives

The Learning Objectives of this course are as follows:

The objective of this course is to introduce students to data analysis and impart them skills to solve data analytics problem. Data Engineering is basically designing and building pipelines that transform and transport data into a highly usable format before it reaches the Data Scientists or other end users. These pipelines must take data from many disparate sources and collect them into a single warehouse that represents the data uniformly as a single source of truth.

Learning outcomes

The Learning Outcomes of this course are as follows:

- CO1 Use data analysis tools in the pandas library.
- CO2 Develop understanding of basic data analysis techniques.
- CO3 Collect, explore, clean, munge and manipulate data.
- CO4 Solve real world data analysis problems.
- CO5 Build data science applications using Python based toolkits.

This document is prepared from the following University Notifications

- https://www.du.ac.in/uploads/new-web/15092023_Indis_sem1.pdf
- https://www.du.ac.in/uploads/new-web/notifications-2021/28032023_nep_Faculty%20of%20Interdisciplinary%20&%20Applied%20Sciences.pdf
- https://www.du.ac.in/uploads/new-web/15092023_Indis_sem3.pdf
- https://www.du.ac.in/uploads/new-web/18092023_Inter_4.pdf

SYLLABUS OF GE-2

UNIT – I Mathematical Foundation for Data Engineering (12 Hours)

Linear Algebra: Vectors, Matrices; Statistics: Describing a Single Set of Data, Correlation, Simpson's Paradox, Correlation and Causation; Probability: Dependence and Independence, Conditional Probability, Bayes's Theorem, Random Variables, Continuous Distributions, The Normal Distribution, The Central Limit Theorem ; Hypothesis and Inference: Statistical Hypothesis Testing, Confidence Intervals, P-hacking, Bayesian Inference

UNIT – II Introduction to Data Engineering and Data Science (12 Hours)

Relationship between Data Engineering and Data Science, Types of Data, Data file formats. Overview of Data Repositories; Data Warehouses, Data Marts, and Data Lakes. Introduction to ETL, ELT, and Data Pipelines. Data Integration Platforms, Traits of Big data, Analysis vs Reporting, Exploratory Data Analysis and Data Science Process. Motivation for using Python for Data Analysis. Introduction to Cloud Computing in Data Science

Essential Python Libraries: NumPy, pandas, matplotlib, SciPy, scikit-learn, stats models

UNIT – III Understanding Pandas and Data Wrangling (12 Hours)

Getting Started with Pandas: Arrays and vectorized computation, Introduction to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. Data Loading, Cleaning, Preparation and Transformation.

Data Wrangling: Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting.

UNIT – IV Data Aggregation and Analysis (9 Hours)

Data Aggregation and Group operations: Group by Mechanics, Data aggregation, General split-apply-combine, Pivot tables and cross tabulation

Time Series Data Analysis: Date and Time Data Types and Tools, Time series Basics, date Ranges, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion, Moving Window Functions.

Practical component (if any) - Data Engineering and Analytics Lab (Python) (30 Hours)

Learning outcomes

CO1 Implement various data analysis tools in the pandas library.
CO2 Implement various basic data analysis techniques, clean and filter and manipulate data.
CO3 Solve real world data analysis problems.

1. Create a Data Frame and perform Matrix-like Operations on a Data Frame
2. Implement basic array statistical methods (sum, mean, std, var, min, max, argmin, argmax, cumsum and cumprod) and perform sorting operation with sort method.
3. Create a data frame with a following structure using pandas

EMP ID	EMP NAME	SALARY	START DATE
1	Satish	50000	01-11-2017
2	Reeya	75000	12-05-2016

This document is prepared from the following University Notifications

- https://www.du.ac.in/uploads/new-web/15092023_Indis_sem1.pdf
- https://www.du.ac.in/uploads/new-web/notifications-2021/28032023_nep_Faculty%20of%20Interdisciplinary%20&%20Applied%20Sciences.pdf
- https://www.du.ac.in/uploads/new-web/15092023_Indis_sem3.pdf
- https://www.du.ac.in/uploads/new-web/18092023_Inter_4.pdf

3	Jay	100000	22-09-2015
4	Roy	45000	08-01-2017
5	Serah	55000	06-02-2018

4. Load Pima Indians Diabetes dataset (Source: <https://archive.ics.uci.edu/ml/datasets/diabetes>). Implement the following
 - i. Data Cleaning and Filtering methods (Use NA handling methods, fillna function arguments).
 - ii. Implement descriptive and summary statistics.
 - iii. Plot histogram, bar plot, distplot for features/attributes of the dataset
5. Load Boston Housing Price dataset and perform
 - i. Data cleaning and filtering method on the dataset.
 - ii. Implement descriptive and summary statistics
 - iii. Plot 'distplot' for target variable and 'heatmap' for the correlation in dataset.
6. For above data set, perform grouping the data using index in pivot table, aggregate on specific features with values.
7. For Superstore sales data, perform Time Series Data Analysis.
8. Creating cloud account Amazon/Azure/Google/IBM to store images /files / programs..

Note: Students shall sincerely work towards completing all the above listed practicals for this course. In any circumstance, the completed number of practicals shall not be less than seven.

Essential/recommended readings

1. The Data Engineering Cookbook - Mastering The Plumbing Of Data Science by Andreas Kretz.
2. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python by Peter Bruce, Andrew Bruce, Peter Gedeck, Shroff/O'Reilly. ISBN: 8194435006-978
3. Data Engineering A Complete Guide - 2020 Edition by Gerardus Blokdyk, 5starcooks. ISBN: 1867316718-978
4. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling by Ralph Kimball, Margy Ross, Wiley. ISBN: 978-1118530801

Suggestive readings -

1. Python Data Science Handbook by Jake VanderPlas, Shroff/O'Reilly. ISBN: 978-9352134915
2. Data Science from Scratch: First Principles with Python by Joel Grus, Shroff/O'Reilly. ISBN: 9352138326-978

Note: Examination scheme and mode shall be as prescribed by the Examination Branch, University of Delhi, from time to time.