KU4SECPHY101:Fundamentals of Data analytics

Semester	Course Type	Course Level	Course Code	Credits	Total Hours
IV	SEC	100	KU4SECPHY101	3	60

Learning Approach (Hours/ Week)		Marks Distribution			Duration of ESE (Hours)
Lecture+ Tutorial	Practical	CE	ESE	Total	
2	2	30	45	75	1.5

Course description:

This course provides a comprehensive introduction to statistical concepts and methods. This also introduces data handling modules of Python. Practical applications and problem-solving are emphasized through real-world examples and exercises.

Course Prerequisite: Higher secondary level Mathematics/ statistics

Course Outcomes:

CO No.	Expected Outcome	Learning Domains
1	Understand and apply knowledge about data.	*U,A*
2	Analyse and use different data analysis tools and software	*An , A*
3	Understand different data sampling methods and use graph plotting.	*U, A*
4	Apply different software tools for data analysis.	*A*

**Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C)*

Mapping of Course Outcomes to PSOs

	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6	PSO 7
CO 1	3	3	3	3	1	1	1
CO 2	3	3	2	2	1	0	1
CO 3	3	3	3	3	0	1	0
CO 4	3	3	3	2	2	0	0

**Correlation level 0-None, 1-Slight, 2-Medium, 3-High*

COURSE CONTENTS

Contents for Classroom Transaction:

MODULE	UNIT	DESCRIPTION	HOURS
1	**Sampling and data**		**8**
	1	Definition of statistics, probability and key terms	**2**
	2	Data, sampling, and variation in data and sampling	**2**
	3	Frequency, Frequency tables, and levels of measurement	**2**
	4	Stemplot, Line graphs and Bar graph, histogram, Frequency polygon and time series graph.	**2**
	Sections 1.1,1.2,1.3, 2.1, 2.2 of Book 1		
2	**Data file formats**		**8**
	1	Introducing different data file formats: csv, xls, tab, dat formats.	2
	2	Jupyter Notebooks using Anaconda and Google Colab: introduction	2

	3	Familiarisation with Google Colab 1, Familiarization with Anaconda 2	2
	4	Reading data files in Jupyter Notebooks	2
		Book 1	
3	**Using Pandas for Data Analysis**		**8**
	1	Data Analysis Using Pandas: Series and dataframe, creating data	2
	2	frame from an excel spreadsheet - creating data frame from .csv files. Creating data frame from a python dictionary - creating data frame from python list of tuples - viewing data frame using loc() and iloc().	2
	3	Operations on data frames series object - creating series from a data frame - creating data frame from series - creating series from numpy array.	2
	4	Converting series into numpy array - creating series from a dictionary- accessing elements of a series, Joining data frames - how to join when there is no common column - concatenation of tables - where() method - groupby() method, aggregate functions on data frames.	2
		Chapters 12,13 (SQL & Regular expressions not required**) of Book 1**	
4	**Data Visualization using Seaborn**		**6**
	1	Loading datasets in Seaborn, Distribution plot Count plot, box plot, scatter plot, joint plot. Line Plot, displaying scatter plot with regression	3
	3	Creating subplots 1, Heat map - cat plot 2, Violin plot - pair plot. 2	3
		Chapter 15 of Book 1	
5	**Practical Module**		**30**
	Directions:		
	Practical related with data analysis and manipulations		

Essential Readings:

1. Introductory Statistics, Barbara Illowsky, De Anza College, Susan Dean-De Anza College, OpenStax, 2018.
2. Machine Learning in Data Science using Python, Dr. R. Nageswara Rao, Dreamtech press, 2022

Suggested Readings:

1. Statistics 4th Edition -David Freedman, Robert Pisani, Roger Purves, W. W. Norton& Company, 2007.

Assessment Rubrics:

Evaluation Type			**Marks**	**Evaluation Type**			**Marks**	**Total**
Lecture			**50**	**Practical**			**25**	**75**
a)	ESE		**35**	a)	ESE		**10**	
b)	CCA		**15**	b)	CCA		**15**	
	i	*Test Paper	8		i	Punctuality	5	
	ii	**Book-Article review/ Assignment	2		ii	Skill	5	
	iii	Seminar/ Viva-Voce	5		iii	Record	5	

*Best out of two test papers

** or any other activities like quiz, open book exam, group activity