Lingaya's Vidyapeeth

Deemed-to-be-University u/s 3 of UGC Act 1956, Government of India
NAAC ACCREDITED
Approved by MHRD / AICTE / PCI / BCI / COA / NCTE
Nachauli, Jasana Road, Faridabad- 121002 (Haryana)
Website: www.lingayasvidyapeeth.edu.in | Ph: 0129-2598200-05

COURSE PLAN & COURSE DATA SHEET

PROGRAM: B.Tech(DS)	DEGREE:
COURSE: Data Wrangling	SEMESTER: VI CREDITS: 3
COURSE CODE: CS-314 REGULATION:	COURSE TYPE: CORE
COURSE AREA/DOMAIN:	CONTACT HOURS: 3 hours/Week.
CORRESPONDING LAB COURSE CODE (IF ANY): CS-364	LAB COURSE NAME (IF ANY): Data Wrangling Lab

PROGRAM EDUCATIONAL OBJECTIVES:

SYLLABUS:

UNIT	DETAILS	HOURS
I	**INTRODUCTION TO DATA WRANGLING:** What Is Data Wrangling? - Importance of Data Wrangling -How is Data Wrangling performed? - Tasks of Data Wrangling-Data Wrangling Tools-Introduction to Python-Python Basics-Data Meant to Be Read by Machines-CSV Data-JSON Data-XML Data.	6
II	**WORKING WITH EXCEL FILES AND PDFS:** Installing Python Packages-Parsing Excel Files-Parsing Excel Files -Getting Started with Parsing-PDFs and Problem Solving in Python-Programmatic Approaches to PDF Parsing-Converting PDF to Text-Parsing PDFs Using pdf miner-Acquiring and Storing Data-Databases: A Brief Introduction-Relational Databases: MySQL and PostgreSQL-Non-Relational Databases: NoSQL-When to Use a Simple File-Alternative Data Storage.	8
III	**DATA CLEANUP:** Why Clean Data? Data Clean up Basics-Identifying Values for Data Clean-up-Formatting Data-Finding Outliers and Bad Data-Finding Duplicates-Fuzzy Matching-RegEx Matching-Normalizing and Standardizing the Data-Saving the Data-Determining suitable Data Clean-up Scripting the Clean-up Testing with New Data	7
IV	**DATA EXPLORATION AND ANALYSIS:** Exploring Data-Importing Data-Exploring Table Functions-Joining Numerous Datasets-Identifying Correlations-Identifying Outliers-Creating Groupings-Analysing Data-Separating and Focusing the Data Presenting Data-Visualizing the Data-Charts-Time-Related Data-Maps- Interactives -Words-Images, Video, and Illustrations-Presentation Tools-Publishing the Data-Open Source Platforms.	6
V	**WEB SCRAPING:** What to Scrape and How-Analyzing a Web Page-Network/Timeline-Interacting with JavaScript-In-Depth Analysis of a Page- Getting Pages-Reading a Web Page-Reading a Web Page with LXML-XPath-Advanced Web Scraping -Browser-Based Parsing-Screen Reading with Selenium-Screen Reading with Ghost.PySpidering the Web-Building a Spider with Scrapy-Crawling Whole Websites with Scrapy.	7
	TOTAL HOURS	34

Teacher Centric Approach **TC1: Chalk and Talk, TC2: PPT, TC3: Video Lectures TC4: Blended learning**
Learner Centric Approach: **LC1: Assignment. LC2: Mini project. LC3: Quiz/Class test. LC 4: Seminar on recent trends.** **LC5: Group Task. LC6: Others**

DETAILED SESSION PLAN

Lecture session/ Number	Topics to be covered	CO addressed	Teacher Centric Approach	Learner Centric Approach	References	Relevance with POs and PSOs
1	What Is Data Wrangling? - Importance of Data Wrangling	CO1	TC1	LC1	T1,R2,W1	
2	How is Data Wrangling performed?	CO1	TC1	LC1, LC3	T1,R2,W1	
3	Tasks of Data Wrangling-Data Wrangling Tools	CO1	TC1	LC1, LC3	T1,R1,W2	
4	Introduction to Python-Python	CO1	TC1	LC1	T1,R1,W2	
5	Data Meant to Be Read by Machines	CO1	TC1	LC1	T1,R2,W3	
6	CSV Data-JSON Data-XML Data.	CO1	TC1	LC1,LC3	T1,R3,W3	
7	Installing Python Packages	CO1	TC3	LC4	-	
8	Parsing Excel Files-Parsing Excel Files	CO1	TC1	LC1, LC3	T1,R1	

9	Getting Started with Parsing-PDFs and Problem Solving in Python	CO2	TC1	LC1, LC3	T1,R2
10	Programmatic Approaches to PDF Parsing-Converting PDF to Text-Parsing	CO2	TC1	LC1, LC3	T1,R2
11	Acquiring and Storing Data-Databases: A Brief	CO2	TC1	LC1, LC3	T1,R3
12	Relational Databases: MySQL	CO2	TC1	LC1, LC3	T1,R2
13	Non-Relational Databases: NoSQL	CO2	TC1	LC1, LC3	T1,R1
14	When to Use a Simple File-	CO2	TC1	LC1, LC3	T1,R3
15	Why Clean Data? Data Clean up	-	TC1	LC1, LC3	T1,R2
16	Data-Finding Outliers and Bad Data	CO3	TC1	LC1, LC3	T1,R3
17	Duplicates-Fuzzy Matching	CO3	TC1	LC1, LC3	T1,R2
18	RegEx Matching	CO3	TC1	LC1, LC3	T1,R1
19	Normalizing and Standardizing the	CO3	TC1	LC1, LC3	T1,R1
20	Determining suitable Data Clean-	CO3	TC1	LC1, LC3	T1,R2
21	The Clean-up Testing with New	CO3	TC1	LC1, LC3	T1,R2
22	Exploring Data-Importing Data-	CO3	TC1	LC1, LC3	T1,R3
23	Joining Numerous Datasets-Identifying	CO3	TC1	LC1, LC3	T1,R1
24	Creating Groupings-Analysing Data-	-	TC1	LC1, LC3	T1,R3
25	Visualizing the Data-Charts-Time-Related Data-Maps	CO4	TC1	LC1, LC3	T1,R3

26	Interactives - Words-Images,	CO4	TC1	LC1, LC3	T1,R2
27	Presentation Tools-Publishing the	CO4	TC1	LC1, LC3	T1,R1,W3
28	What to Scrape and How-Analyzing a	CO4	TC1	LC1, LC3	T1,R3
29	Interacting with JavaScript-In-Depth Analysis of a Page	CO4	TC1	LC1, LC3	T1,R2
30	Getting Pages-Reading a Web Page-Reading a	-	TC1	LC1, LC3	T1,R3
31	Browser-Based Parsing	CO5	TC1	LC1, LC3	T1,R2
32	Screen Reading with Selenium-Screen Reading	CO5	TC1	LC1, LC3	T1,R1
33	the Web-Building a Spider with Scrapy	CO5	TC1	LC1, LC3	T1,R1
34	Crawling Whole Websites with Scrapy	CO5	TC1	LC1, LC3	T1,R2, W1

TEXT/REFERENCE BOOKS:

T/R	BOOK TITLE/AUTHORS/PUBLICATION
T1	Principles of Data Wrangling: Practical Techniques for Data Preparation
R1	Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython
R2	Data Wrangling with Python: Creating actionable data from raw sources
R3	Data Wrangling with Python: Tips and Tools to Make Your Life Easier
W1	www.simplilearn.com
W2	www.analyticsvidhya.com
W3	www.trifacta.com

COURSE PRE-REQUISITES:

C.CODE	COURSE NAME	DESCRIPTION	SEM
CS-305C	**Python Programming**	**Knowledge of python programming**	**5th**

COURSE OBJECTIVES:

1	Data Wrangling, Analysis and Visualization are critical skills that are necessary in the handling the challenges in a variety of modern-day data-driven businesses.
2	Knowledge in data analysis and visualization provide one with the necessary skills that form the basis for deeper quantitative reasoning needed for data and predictive analytics work.
3	This course enables the skills to turn raw datasets into interactive and dynamic visualizations and dashboards providing the platform to draw insights for decision making.

COURSE OUTCOMES:

S.NO	DESCRIPTION	PO(1..12) MAPPING	PSO(1..3) MAPPING
Cxxx.1	Understanding of Data Wrangling and Tools i.e. Python, CSV, JSON, XML etc..	PO1,PO2,PO3,PO5,PO6	PSO1
Cxxx.2	Understanding of Python Packages, NoSQL, Excel Parsing and PDF parsing.	PO1,PO2,PO3,PO5,PO6	POS1,PSO2
Cxxx.3	Understanding of Data Cleaning Techniques.	PO1,PO2,PO5,PO6,PO8,PO12	PSO1,PSO2,PSO3
Cxxx.4	Understanding of the Data Exploration and Data Analysis.	PO1,PO2,PO4,PO5,PO6,PO9,PO12	PSO1,PSO3
Cxxx.5	Understand of Web Scraping and Analyzing a Web Page-Network/Timeline-Interacting with JavaScript.	PO1,PO2,PO4,PO5,PO6,PO8,PO9,PO12	PSO1,PSO3
COURSE OVERALL PO/PSO MAPPING:			

COURSE OUTCOMES VS POs MAPPING (DETAILED; HIGH: 3; MEDIUM:2; LOW:1**):**

S.NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO 1	3	1	1		1	1							3		
CO.2	2	1	1		3	1							1	2	
CO.3	1	1			1	1		1				1	1	1	1
CO.4	1	1		1	1	1			1			1	1		1
CO.5	1	1		1	1	1		1	1			1	1		1

** For Entire Course, PO & PSO Mapping*

POs & PSO REFERENCE:

PO1	Engineering Knowledge	PO7	Environment & Sustainability	PSO1	To equip the students with theoretical and implementation knowledgebase in all the latest areas of Computer Science & Engineering for a successful career in software industries, pursuing higher studies, or entrepreneurial establishments.
PO2	Problem Analysis	PO8	Ethics	PSO2	To nurture the students with the critical thinking abilities for better decision making by offering them a socially acceptable solutions to real life problems through computing paradigm.

PO3	Design & Development	PO9	Individual & Team Work	PSO3	To nurture the students with the comprehensive analytical and design abilities by offering them techno-commercially feasible solutions of real business problems through computing.
PO4	Investigations	PO10	Communication Skills		
PO5	Modern Tools	PO11	Project Mgt. & Finance		
PO6	Engineer & Society	PO12	Life Long Learning		

COs VS POs MAPPING JUSTIFICATION:

S.NO	PO/PSO MAPPED	LEVEL OF MAPPING	JUSTIFICATION
Cxxx.1			
Cxxx.2			
Cxxx.3			
Cxxx.4			
Cxxx.5			
Cxxx*			

GAPS IN THE SYLLABUS - TO MEET INDUSTRY/PROFESSION REQUIREMENTS, POs & PSOs:

SNO	DESCRIPTION	PROPOSED ACTIONS
1		
2		
3		
4		
5		

PROPOSED ACTIONS: TOPICS BEYOND SYLLABUS/ASSIGNMENT/INDUSTRY VISIT/GUEST LECTURER/NPTEL ETC

 # TOPICS BEYOND SYLLABUS/ADVANCED TOPICS/DESIGN:

1	
2	
3	
4	
5	
6	
7	

DELIVERY/INSTRUCTIONAL METHODOLOGIES:

☐ CHALK & TALK	☐ STUD. ASSIGNMENT	☐ WEB RESOURCES	☐ NPTEL/OTHERS
☐ LCD/SMART BOARDS	☐ STUD. SEMINARS	☐ ADD-ON COURSES	☐ WEBNIARS

Lingaya's Vidyapeeth

Deemed-to-be-University u/s 3 of UGC Act 1956, Government of India
NAAC ACCREDITED
Approved by MHRD / AICTE / PCI / BCI / COA / NCTE
Nachauli, Jasana Road, Faridabad- 121002 (Haryana)
Website: www.lingayasvidyapeeth.edu.in | Ph: 0129-2598200-05

ASSESSMENT METHODOLOGIES-DIRECT

☐ ASSIGNMENTS	☐ STUD. SEMINARS	☐ TESTS/MODEL EXAMS	☐ UNIV. EXAMINATION
☐ STUD. LAB PRACTICES	☐ STUD. VIVA	☐ MINI/MAJOR PROJECTS	☐ CERTIFICATIONS
☐ ADD-ON COURSES	☐ OTHERS		

ASSESSMENT METHODOLOGIES-INDIRECT

☐ ASSESSMENT OF COURSE OUTCOMES (BY FEEDBACK, ONCE)	☐ STUDENT FEEDBACK ON FACULTY (TWICE)
☐ ASSESSMENT OF MINI/MAJOR PROJECTS BY EXT. EXPERTS	☐ OTHERS

 # INNOVATIONS IN TEACHING/LEARNING/EVALUATION PROCESSES:

1. **Technology Integration:** Embrace and integrate technology tools in the classroom to enhance the learning experience. This can include interactive whiteboards, educational apps, virtual reality, and online collaboration platforms. Utilizing technology allows for more dynamic and interactive lessons, catering to diverse learning styles.
2. **Personalized Learning Paths:** Implement personalized learning approaches that cater to individual student needs and pace of learning. Adaptive learning platforms and data analytics can help tailor educational content, assignments, and assessments based on the strengths and weaknesses of each student, promoting a more customized learning experience.
3. **Active Learning Strategies:** Move away from traditional lecture-based approaches and incorporate active learning strategies. This involves engaging students in hands-on activities, group discussions, problem-solving exercises, and real-world projects. Active learning fosters critical thinking, collaboration, and practical application of knowledge.
4. **Blended Learning Models:** Adopt blended learning models that combine face-to-face instruction with online resources. This allows for flexibility in learning, enabling students to access materials at their own pace outside the classroom. Flipped classrooms, where students learn new concepts online and engage in discussions and activities during class, are an example of a blended learning approach.
5. **Assessment Innovation:** Rethink assessment methods to go beyond traditional exams and quizzes. Explore alternative forms of assessment, such as project-based assessments, portfolios, presentations, and peer assessments. Additionally, incorporate formative assessments and feedback throughout the learning process to help students track their progress and make improvements.

Prepared by
(Ms. Komal Malsa)

Approved by
(HOD)

 # Additionally, the details to be compiled separately by the Departmental Coordinator for the entire Department.