

**PRAGATI ENGINEERING COLLEGE: SURAMPALEM
(AUTONOMOUS)**

III B.Tech II Semester Regular/Supplementary Examinations, April - 2024

DATA WRANGLING IN DATA SCIENCE

(CSE (DS))

Time: 3 hours

Max. Marks: 70 M

**Answer ONE Question from each Unit
All Questions Carry Equal Marks**

Q. No.	Questions	BTL	CO	Marks
UNIT – I				
1.	a) What is Quality Data? What are the advantages of quality data?	K1	CO1	7M
	b) What is CSV? Demonstrate Reading of the CSV data.	K2	CO1	7M
OR				
2.	a) What is Data Wrangling? Explain the benefits of Data Wrangling.	K2	CO1	7M
	b) Demonstrate the reading of JSON data.	K3	CO1	7M
UNIT – II				
3.	a) What is parsing? Demonstrate parsing of PDF file.	K3	CO2	7M
	b) What is NOSQL? How will it differ from SQL?	K2	CO2	7M
OR				
4.	a) List out various alternate data storage approaches.	K2	CO2	7M
	b) Describe when to use a simple file in the context of Data Storage.	K2	CO2	7 M
UNIT – III				
5.	a) What is Data Cleaning? Explain and demonstrate Identifying Values for Data Cleanup.	K2	CO3	7M
	b) What is Regular Expression Matching? How is it useful in Machine Learning.	K2	CO3	7M
OR				
6.	a) Explain and Demonstrate Fuzzy Matching with examples.	K2	CO3	7M
	b) Differentiate Normalization and Standardization? What are different standardization Methods?	K2	CO3	7M
UNIT – IV				
7.	a) List out and explain different libraries used in Data Exploration and Analysis.	K2	CO4	7M
	b) Explain the different ways to handle outliers.	K2	CO4	7M
OR				
8.	a) What is Correlation? How will you find a correlation between attributes or variables?	K3	CO4	7M
	b) What is Map? Explain the different types of Maps used in Data Analysis.	K2	CO4	7M
UNIT – V				
9.	a) What is web Scraping? Demonstrate reading a webpage using Beautiful Soup.	K3	CO5	7M
	b) How will you read the screen using Selenium?	K2	CO5	7M
OR				
10.	a) How will you crawl the whole website using Scrapy?	K3	CO5	7M
	b) What is web Scraping? Explain Browser based parsing.	K2	CO5	7M