


<b>Name:</b> <b>Enrolment No:</b>	
--------------------------------------	-------------------------------------------------------------------------------------

**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**  
**End Semester Examination, May 2022**

**Course: Data Preparation**  
**Program: BBA Analytics & Big Data**  
**Course Code: DSIT 3004**

**Semester: 4**  
**Time : 03 hrs.**  
**Max. Marks: 100**

**Instructions:** *The student must write his/her name and enrolment no. in the space designated above.*

**SECTION A**  
**10Qx2M=20Marks**

S. No.		Marks	CO
Q1	Data in _____ bytes is known as Big Data. a. Mega b. Giga c. Tera d. Peta	2	CO1
Q2	Choose the languages which are used in data science. a. C++ b. C c. R d. Ruby	2	CO1
Q3	Which of the following can be generally used to clean and prepare big data? a. Pandas b. Data lake c. U-SQL d. Data warehouse	2	CO1
Q4	What is the use of data cleaning? a. To remove the noisy data b. Transformations to correct the wrong data. c. Correct the inconsistencies in data d. All of the above	2	CO1
Q5	Which company was recently implicated in a global data theft crime? a. Amazon b. Google c. Cisco d. Cambridge Analytica	2	CO1
Q6	Which of the following is not a measure of dispersion? a. Range b. Mode c. Mean Deviation d. Standard Deviation	2	CO1

Q7	Microsoft _____ is used to find insights within an organization's data	2	CO1
Q8	Data Analysis is a process of? a. inspecting data b. cleaning data c. transforming data d. All of the above	2	CO1
Q9	The goal of business intelligence is to allow easy interpretation of large volumes of data to identify new opportunities. a. TRUE b. FALSE c. Can be true or false d. Can not say	2	CO1
Q10	Text Analytics, also referred to as Text Mining? a. TRUE b. FALSE c. Can be true or false d. Can not say	2	CO1
<b>SECTION B</b> <b>4Qx5M= 20 Marks</b>			
Q1	List down some of the most popular Data preparation tools.	5	CO1
Q2	Explain 'Data Cleaning'.	5	CO1
Q3	Discuss the benefits of Data preparation.	5	CO2
Q4	In what cases would one need Data Augmentation?	5	CO1
<b>SECTION-C</b> <b>3Qx10M=30 Marks</b>			
Q1	Summarily list down the essential steps in the Data preparation process.	10	CO2
Q2	What are the various personnel roles involved in the Data preparation process? <b>OR</b> What are the benefits of outlier cleaning during Data preparation process.	10	CO2
Q3	What is the significance of 'Data Discovery' process during data preparation?	10	CO2
<b>SECTION-D</b> <b>2Qx15M= 30 Marks</b>			
Q1	How can Artificial Intelligence (AI) play a bigger role in Data Preparation?	15	CO3
Q2	Describe in detail the ETL Cycle. <b>OR</b> What are some of the steps with which Data is prepared/processed for further Analytics?	15	CO3