Name: **Enrolment No:**	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End semester Examination, May 2023

Course: Data Preparation **Semester : II**
Program: MBA Digital Business **Time : 03 hrs.**
Course Code: DSIT 7019 **Max. Marks: 100**

Instructions: Attempt all sections

SECTION A
10Qx2M=20Marks

S. No.		Marks	CO
Q 1	Attempt all Questions in this section		
a.	What is DBMS? i. DBMS is a collection of queries ii. DBMS is a high-level language iii. DBMS is a programming language iv. DBMS stores, modifies and retrieves data	**2**	**CO1**
b.	What is secondary data? i. Data that isn't as good ii. Data that is collected first-hand iii. Data expressed through interpretive analysis. iv. Data that already exists	**2**	**CO1**
c.	Charts that are helpful in making comparisons are: i. Bar charts ii. column charts iii. Pie charts iv. Both Bar & Column Charts	**2**	**CO1**
d.	Which of the following is not a phase in Data Analytics Lifecycle? i. Data discovery ii. Predictive Modelling iii. Model Designing iv. Model building	**2**	**CO1**
e.	Which one of them is a method for Outlier Detection? v. Clustering vi. Boxplots, histograms, and scatterplots vii. Forecasting viii. Classification	**2**	**CO1**
f.	The data that represents the number of tickets sold at a movie theater on any given night is: i. Nominal data ii. Ordinal data	**2**	**CO1**

	iii. Interval data iv. Ratio data		
g.	The statistical data are of two types. These types are : i. technical data and presentation data ii. Primary data and secondary data iii. Primary data and personal data iv. none of the above	**2**	**CO1**
h.	What's the difference between categorical data and continuous data?	**2**	**CO1**
i.	Which of the following is an example of interval level data? i. Temperature in Fahrenheit or Celsius ii. Weight iii. Ethnicity iv. Language ability (e.g., beginner, intermediate, fluent)	**2**	**CO1**
j.	Which of the following are the steps followed in data discovery phase in the data analytics life cycle? i. Problem or Opportunity Identification ii. Model Designing iii. Measuring of Effectiveness	**2**	**CO1**
SECTION B 4Qx5M= 20 Marks			
	Attempt all four Questions in this section		
Q.2.	Explain the difference between the management problem and analysis problem with examples.	**5**	**CO1**
Q.3.	What is data imputation? Explain with examples.	**5**	**CO2**
Q.4.	Explain the process of handling missing data. Why is it important?	**5**	**CO2**
Q.5.	What do you understand by measures of spread/variability? Why is it important?	**5**	**CO1**
SECTION-C 3Qx10M=30 Marks			
	Attempt all three Questions in this section		
Q.6.	Explain the difference between Data Matrix, Graphical and Ordered data using examples.	**10**	**CO2**
Q.7.	Write a program in R programing language to find whether a number is positive or negative.	**10**	**CO2**
Q.8.	What are the different data types in R? How do you check for the date types?	**10**	**CO2**
SECTION-D 2Qx15M= 30 Marks			
	Attempt both the Questions in this section		

<table>
<tr>
<td>Q.9.</td>
<td>Consider the given dataset named as LungCapacity.
Write at least 10 steps for how would you analyse this data in R programing language.

<table>
<tr><th>LungCap(cc)
(double)</th><th>Age(
years)
(double)</th><th>Height(inches)
(double)</th><th>Smoke
(character)</th><th>Gender
(character)</th><th>Caesarean
(character)</th></tr>
<tr><td>6.475</td><td>6</td><td>62.1</td><td>no</td><td>male</td><td>no</td></tr>
<tr><td>10.125</td><td>18</td><td>74.7</td><td>yes</td><td>female</td><td>no</td></tr>
<tr><td>9.550</td><td>16</td><td>69.7</td><td>no</td><td>female</td><td>yes</td></tr>
<tr><td>11.125</td><td>14</td><td>71.0</td><td>no</td><td>male</td><td>no</td></tr>
<tr><td>4.800</td><td>5</td><td>56.9</td><td>no</td><td>male</td><td>no</td></tr>
<tr><td>6.225</td><td>11</td><td>58.7</td><td>no</td><td>female</td><td>no</td></tr>
<tr><td>4.950</td><td>8</td><td>63.3</td><td>no</td><td>male</td><td>yes</td></tr>
</table>
</td>
<td>15</td>
<td>CO3</td>
</tr>
<tr>
<td>Q.10.</td>
<td>Beauty Incorp. was a company carrying on business in cosmetics and perfumes. It was catering to a target market which was using its products. In other words, it only concentrated on what it would make, and did not bother about changes in preferences of its target market.
It was later joined by a new marketing manager who advised the company regarding the changing consumer preferences, and the changes that were necessary to be incorporated in the products. He emphasized upon the income factor. He modernized the products to a great extent, and invested about 30 lakhs on new packing, etc. Even after six months of these changes brought about by him, the company did not seem to have a proportionate increase in sales.
The company management were not very happy with the changes, and thought that although an effort has been made in the right direction, some important factors concerning consumer behavior had been neglected.

As a digital business Analyst, comment on the situation and devise a plan to help the business grow.</td>
<td>15</td>
<td>CO3</td>
</tr>
</table>