Printed I	Page:- Subject Code:- ACSDS0301N Roll. No:
	NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
	(An Autonomous Institute Affiliated to AKTU, Lucknow)
	B.Tech
	SEM: III - CARRY OVER THEORY EXAMINATION - APRIL 2023
Time a. 7	Subject: Foundations of Data Science
Time: 3	Hours Max. Marks: 100 Instructions:
	fy that you have received the question paper with the correct course, code, branch etc.
•	uestion paper comprises of three Sections -A, B, & C. It consists of Multiple Choice
Questions	s (MCQ's) & Subjective type questions.
2. Maxim	um marks for each question are indicated on right -hand side of each question.
	te your answers with neat sketches wherever necessary.
	e suitable data if necessary.
-	ably, write the answers in sequential order.
evaluatea	eet should be left blank. Any written material after a blank sheet will not be
cvaraacca	SECTION A 20
1 Attom	
	pt all parts:-
1-a.	Which of the following focuses on the discovery of (previously) unknown 1 properties on the data [CO1]
	(a) Data mining
	(b) Big Data
	(c) Data wrangling (d) Machine Learning
1 h	(d) Machine Learning
1-b.	Hadoop works in [CO1]
	(a) Centralized processing fashion
	(b) master – slave fashion
	(c) worker/slave fashion
	(d) All of the mentioned
1-c.	Which of the following is characteristic of Processed Data? [CO2]
	(a) Data is not ready for analysis
	(b) All steps should be noted

	(c) Hard to use for data analysis	
	(d) None of the mentioned	
1-d.	What is the use of data cleaning? [CO2]	1
	(a) To remove the noisy data	
	(b) Transformations to correct the wrong data.	
	(c) Correct the inconsistencies in data	
	(d) All of the above	
1-e.	A data warehouse is which of the following? [CO3]	1
	(a) Can be updated by end users.	
	(b) Contains numerous naming conventions and formats.	
	(c) Organized around important subject areas.	
	(d) Contains only current data.	
1-f.	The operation of moving from finer-granularity data to a coarser granularity (by means of aggregation) is called a [CO3]	1
	(a) Rollup	
	(b) Drill down	
	(c) Dicing	
	(d) Pivoting	
1-g.	Which of following is a problem that missing data could cause? [CO4]	1
	(a) Some visualizations will not work with missing data	
	(b) Some types of data cleaning steps won't work with null values	
	(c) Machine learning models will break when they encounter null values	
	(d) All of the above	
1-h.	Where can the data be updated? [CO4]	1
	(a) Informational Environment	
	(b) Datawarehouse environment	
	(c) Operational environment	
	(d) Data mining environment	
1-i.	Which of the following is not a recommended type of graph? [CO5]	1
	(a) Pie chart	
	(b) Histogram	
	(c) Barchart	
	(d) all of the above	

1-j. Velocity is the speed at which the data is processed [CO5] 1 (a) TRUE (b) FALSE (c) All (d) None 2. Attempt all parts: 2.a. Name few Data Science tools? [CO1] 2 2.b. What is Primary Data? [CO2] 2 2.c. List any five applications of data warehouse? [CO3] 2 2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 3 3. Answer any five of the following: 3. An swer any five of the following: 3a. What are the prerequisites for Data Science? [CO1] 6 3b. Discuss all phases of Data Science [fecycle? [CO1] 6 3c. What kind of data is HTML files, explain briefly? [CO2] 6 3d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 4. Answer any one of the following: 4a. Briefly explain the need of data science with respect to its application areas? [CO1] 4b. Discuss various model planning tools in detail? [CO1] 10 5. Answer any one of the following: 5a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following: 6a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this problem? [CO3]		·	
(b) FALSE (c) All (d) None 2. Attempt all parts:- 2.a. Name few Data Science tools ? [CO1] 2.b. What is Primary Data? [CO2] 2.c. List any five applications of data warehouse? [CO3] 2.d. What are outliers? [CO4] 2.e. What is the lattice package in R used for? [CO5] 30 3. Answer any five of the following:- 3-a. What are the prerequisites for Data Science? [CO1] 3-b. Discuss all phases of Data Science lifecycle? [CO1] 3-c. What kind of data is HTML files, explain briefly? [CO2] 3-d. Differentiate between categorical and graphical data? [CO2] 3-d. Differentiate between data warehouse and database? [CO3] 3-f. Differentiate between data warehouse and database? [CO4] 3-g. Explain what is data visualization and its techniques? [CO4] 3-g. Explain what is data visualization and its techniques? [CO5] 4-Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application areas? [CO1] 4-b. Discuss various model planning tools in detail? [CO1] 5-Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 6-Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	1-j.	Velocity is the speed at which the data is processed [CO5]	1
(c) All (d) None 2. Attempt all parts:- 2.a. Name few Data Science tools? [CO1] 2 2.b. What is Primary Data? [CO2] 2 2.c. List any five applications of data warehouse? [CO3] 2 2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 3 30 3. Answer any five of the following:- 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 4. Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this		(a) TRUE	
2. Attempt all parts:- 2.a. Name few Data Science tools? [CO1] 2 2.b. What is Primary Data? [CO2] 2 2.c. List any five applications of data warehouse? [CO3] 2 2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 3 3. Answer any five of the following:- 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 3.f. Difference between data warehouse with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this		(b) FALSE	
2.At Name few Data Science tools? [CO1] 2 2.b. What is Primary Data? [CO2] 2 2.c. List any five applications of data warehouse? [CO3] 2 2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 5		(c) All	
2.a. Name few Data Science tools? [CO1] 2 2.b. What is Primary Data? [CO2] 2 2.c. List any five applications of data warehouse? [CO3] 2 2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 3 3. Answer any five of the following: 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 4. Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this		(d) None	
2.b. What is Primary Data? [CO2] 2.c. List any five applications of data warehouse? [CO3] 2.d. What are outliers? [CO4] 2.e. What is the lattice package in R used for? [CO5] SECTION B 3. Answer any five of the following: 3-a. What are the prerequisites for Data Science? [CO1] 3-b. Discuss all phases of Data Science lifecycle? [CO1] 3-c. What kind of data is HTML files, explain briefly? [CO2] 3-d. Differentiate between categorical and graphical data? [CO2] 3-e. Differentiate between data warehouse and database? [CO3] 3.f. Difference between univariate and multivariate analysis? [CO4] 3.g. Explain what is data visualization and its techniques? [CO5] 50 4. Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application areas? [CO1] 5. Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 6. Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	2. Atten	npt all parts:-	
2.c. List any five applications of data warehouse? [CO3] 2 2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 3 SECTION B 30 3. Answer any five of the following: 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	2.a.	Name few Data Science tools ? [CO1]	2
2.d. What are outliers? [CO4] 2 2.e. What is the lattice package in R used for? [CO5] 3 3. Answer any five of the following:- 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following:- 10 In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	2.b.	What is Primary Data? [CO2]	2
SECTION B 3. Answer any five of the following:- 3-a. What are the prerequisites for Data Science? [CO1] 3-b. Discuss all phases of Data Science lifecycle? [CO1] 3-c. What kind of data is HTML files, explain briefly? [CO2] 3-d. Differentiate between categorical and graphical data? [CO2] 3.e. Differentiate between data warehouse and database? [CO3] 3.f. Difference between univariate and multivariate analysis? [CO4] 3.g. Explain what is data visualization and its techniques? [CO5] 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 4-b. Discuss various model planning tools in detail? [CO1] 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	2.c.	List any five applications of data warehouse? [CO3]	2
3. Answer any five of the following:- 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	2.d.	What are outliers? [CO4]	2
3. Answer any five of the following: 3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	2.e.	What is the lattice package in R used for? [CO5]	2
3-a. What are the prerequisites for Data Science? [CO1] 6 3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this		SECTION B	30
3-b. Discuss all phases of Data Science lifecycle? [CO1] 6 3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	3. Answ	er any <u>five</u> of the following:-	
3-c. What kind of data is HTML files, explain briefly? [CO2] 6 3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following: 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following: 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following: 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	3-a.	What are the prerequisites for Data Science? [CO1]	6
3-d. Differentiate between categorical and graphical data? [CO2] 6 3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	3-b.	Discuss all phases of Data Science lifecycle? [CO1]	6
3.e. Differentiate between data warehouse and database? [CO3] 6 3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	3-c.	What kind of data is HTML files, explain briefly? [CO2]	6
3.f. Difference between univariate and multivariate analysis? [CO4] 6 3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	3-d.	Differentiate between categorical and graphical data? [CO2]	6
3.g. Explain what is data visualization and its techniques? [CO5] 6 SECTION C 50 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	3.e.	Differentiate between data warehouse and database? [CO3]	6
4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 4-b. Discuss various model planning tools in detail? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	3.f.	Difference between univariate and multivariate analysis? [CO4]	6
 4. Answer any one of the following:- 4-a. Briefly explain the need of data science with respect to its application 10 areas? [CO1] 4-b. Discuss various model planning tools in detail? [CO1] 10 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this 	3.g.	Explain what is data visualization and its techniques? [CO5]	6
 4-a. Briefly explain the need of data science with respect to its application areas? [CO1] 4-b. Discuss various model planning tools in detail? [CO1] 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this 		SECTION C	50
areas? [CO1] 4-b. Discuss various model planning tools in detail? [CO1] 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this	4. Answ	er any <u>one</u> of the following:-	
 5. Answer any one of the following:- 5-a. Describe briefly the steps involved in data manipulation? [CO2] 10 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 10 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this 	4-a.		10
 5-a. Describe briefly the steps involved in data manipulation? [CO2] 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this 	4-b.	Discuss various model planning tools in detail? [CO1]	10
 5-b. Write steps for reading a pdf, csv and json file in python? [CO2] 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this 	5. Answ	er any <u>one</u> of the following:-	
 6. Answer any one of the following:- 6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this 	5-a.	Describe briefly the steps involved in data manipulation? [CO2]	10
6-a. In real world data, tuples with missing values for some attributes are a 10 common occurrence, describe various methods for handling this	5-b.	Write steps for reading a pdf, csv and json file in python? [CO2]	10
common occurrence, describe various methods for handling this	6. Answ	er any <u>one</u> of the following:-	
	6-a.	common occurrence, describe various methods for handling this	10

7. Answe	er any <u>one</u> of the following:-	
7-a.	In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem? [CO4]	10
7-b.	Distinguish between good and bad data ? What are the problems of data analysis? [CO4]	10
8. Answe	er any <u>one</u> of the following:-	
8-a.	What is time series plot. Explain using example? [CO5]	10
8-b.	Explain, you want to a subset of employees who are above 30 years and who make salary less than 10000. How will you execute this in R, Explain with complete steps? [CO5]	10

Explain The Functionality Of OLAP with its types? [CO3]

10

6-b.