Name: Enrolment No:	 **UPES** UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, May 2021

Course: Introduction to Business Analytics   Semester: IV
Program: BBA Analytics and Big Data   Time: 03 Hours
Course code: DSBA 2004   Max. Marks: 100
Instructions: Attaché excel file along with word file

SECTION A

	Attempt all Questions	Marks	CO
	Select the most appropriate answer.	6 X 5=30	CO₁
1.	Supervised Learning is (a) Learning with the help of examples (b) Learning without teacher (c) Learning with the help of teacher (d) Learning with computer as teacher		
2.	Write an example where reinforcement learning is applied.		
3.	If the null hypothesis is significant then select the correct one : (a) Accept the null hypothesis Simple random sampling (b) Reject the null hypothesis (c) Both (a) & (b) (d) None of them		
4.	Which one has higher chance of rejecting the null hypothesis : (a) 1 % level of significance (b) 5 % level of significance (c) 10 % level of Significance (d) Same chance in all of the above		
5.	Which of the following analytics use the concept of operations research? (a) Prescriptive (b) Predictive (c) descriptive (d) All of the above		

6.	Each day, the office staff at Oasis Quick Shop prepares a frequency distribution and a histogram of sales transactions by dollar value of the transactions. Friday's histogram follows.		
	On Friday, the approximate number of sales transactions between \$100 and \$150 was _____.		
	(a) 100		
	(b) 200		
	(c) 300		
	(d) 400		
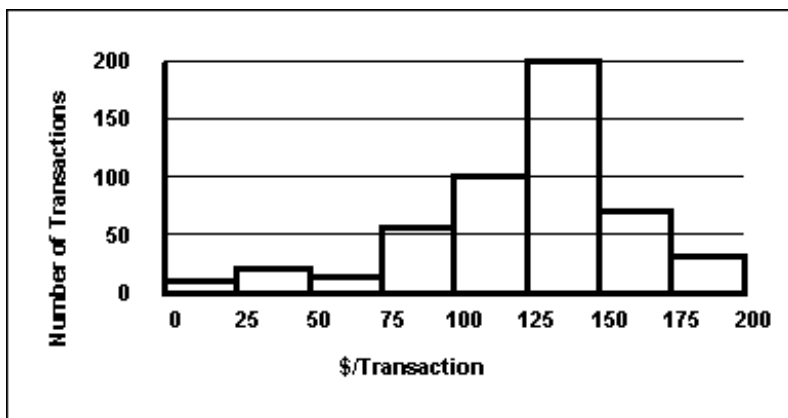
<div align="center">

SECTION B

</div>

Q	**Attempt all the questions**	**10X 5=50**	
1.	Which type of data (cross-sectional or time series) is each variable? (a) Scores of 50 students on a midterm accounting exam last semester. (b) Bob's scores on 10 weekly accounting quizzes last semester. (c) Average score by all takers of the state's CPA exam for each of the last 10 years. (d) Number of years of accounting work experience for each of the 15 partners in a CPA firm. (e) Value of Standard & Poor's 500 stock price index at the close of each trading day last year. **OR** With the help of an example, describe the different stages of life cycle of dada.		CO_2
2.	Using the suitable example, explain the differences between supervised, unsupervised and reinforcement learning.		CO_2
3.	The table given below shows the number of fire insurance claims received by an insurance company in each four month period from 2008 to 2011 :		CO_3

	2008			2009			2010			2011		
Period	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
Claims (y)	7	3	5	9	7	9	12	4	10	13	9	10

(a) Produce a line graph of the quarterly data for fire insurance .Visually establish whether any patterns occur over time.
(b) Calculate the three point and Five point moving averages and compare them
(c) Plot the three point moving and five point moving average with usual claim y
(d) Interpret the graphs plotted in (b)

4.	A tourist car operator finds that during the past few months the car's use has varied so much that the cost of maintaining the car varied considerably. During the past 200days the demand for the car fluctuated as below:	CO3

Trips per week	Frequency
0	16
1	24
2	30
3	60
4	40
5	30

Using the random number simulate the demand for a 10 week period.

5.	A firm is engaged in producing two products. A and B. Each unit of product A requires 2 kg of raw material and 4 labour hours for processing, whereas each unit of B requires 3 kg of raw materials and 3 labour hours for the same type. Every week, the firm has an availability of 60 kg of raw material and 96 labour hours. One unit of product A sold yields Rs.40 and one unit of product B sold gives Rs.35 as profit. Formulate this as a Linear Programming Problem to determine as to how many units of each of the products should be produced per week so that the firm can earn maximum.	CO2

SECTION-C

Q	**Attempt the question :**	20 X 1 =20	CO4
1.	An investment analyst studied the relationship between corporate *return on capital* (%) and a number of *performance* and *demographic characteristics* of business entities. The performance measures are *sales*, *margin* % and *debt ratio* (%) and the demographic measures are the *region* (Gauteng, Cape, KZN) and *sector* (agriculture, construction). **The data gathered from annual financial reports are shown in the table below for a sample of 25 business entities.** (a) For each of the categorical variables, *region* and *sector*, set up a binary coding scheme to recode the data.		

(b) Construct a multiple linear regression model to estimate corporate *return on capital* (%) based on the selected performance and demographic measures

(c) What percentage of variation in *return on capital* (%)) can be explained by all the independent variables?

(d) Test whether the overall regression model is statistically significant. Formulate the null and alternative hypotheses, identify *F-crit* and *F-stat* and draw the statistical conclusion. Use $\alpha = 0.05$.

(e) Test whether the *performance measures* (*sales*, *margin %*, *debt ratio* (%)) are each individually statistically significant independent variables to estimate *return on capital %*. For each, formulate the null and alternative hypotheses, identify *t-crit* and *t-stat* and draw the statistical conclusion. Use $\alpha = 0.05$.

(f) Is *region* statistically significant? Test each binary coded variable separately and conclude about the overall significance of *region* as a useful regressor. Formulate the null and alternative hypotheses, identify *t-crit* and *t-stat* and draw the statistical conclusion. Use $\alpha = 0.05$.

(g) Is *sector* statistically significant? Test each binary coded variable separately and conclude about the overall significance of sector as a useful regressor. Formulate the null and alternative hypotheses, identify *t-crit* and *t-stat* and draw the statistical conclusion. Use $\alpha = 0.05$. Test each of the independent variables for statistical significance based on their *p*-values.

(i) Test each of the independent variables for statistical significance using the 95% confidence interval estimate for their regression coefficients, βi.

(j) Prepare a brief report of the overall findings of this study

(k) Estimate, with 95% confidence, the average return on capital % of business entities with a sales volume of 8862; a margin % of 10%; a debt ratio (%) of 22% and operates primarily in the construction sector in the Cape region.

Company	Y ROC(%)	Performance Measures			Demographic Measures	
		Sales	Margin%	Debt ratio(%)	Region	Sector
1	19.7	7178	18.7	28.5	1	1
2	17.2	1437	18.5	24.3	1	1
3	17.1	3948	16.5	65.6	1	1
4	16.6	1672	16.2	26.4	1	1
5	16.6	2317	16.0	20.1	1	2
6	16.5	4123	15.6	46.4	2	1
7	15.9	4418	15.3	60.3	2	2
8	15.4	6804	14.0	17.7	3	1
9	15.3	3592	14.0	17.0	2	1
10	15.1	1570	13.3	14.1	2	2
11	15.0	3802	13.0	26.1	2	2
12	14.7	2594	12.7	19.2	2	2
13	14.7	1414	13.8	0.0	2	1
14	14.6	1991	9.8	21.9	2	2
15	13.8	3379	8.6	10.7	2	1
16	13.7	1910	9.9	21.4	3	2
17	13.4	7548	5.2	29.5	3	2

	18	12.9	1858	7.8	5.1	2	2			
	19	11.9	3085	1.4	14.6	3	2			
	20	11.9	15197	1.9	39.6	3	2			
	21	11.3	2453	10.2	21.7	3	2			
	22	17.7	2485	17.7	38.6	1	1			
	23	15.5	9564	7.5	22.4	2	2			
	24	14.1	3912	16.1	19.3	2	1			
	25	12.5	2006	14.9	17.9	2	2			