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

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2020

Course: Predictive Modelling **Semester: III**
Program: MBA BA **Time: 03 Hours**
Course code: DSBA 8003 **Max. Marks: 100**
Instructions:

SECTION A

		Marks	CO
	Select the most appropriate answer.	**6 X 5=30**	**CO_1**
1.	**Which of the following methods do we use to best fit the data in Logistic Regression?** (a) Least Square Error (b) Maximum Likelihood (c) Jaccard distance (d) Both A and B		
2.	**What is the minimum no. of variables/ features required to perform clustering?** (a) 0 (b) 1 (c) 2 (d) 3		
3	**Which of the following is true about Residuals?** (a) Lower is better (b) Higher is better (c) A or B depend on the situation (d) None of these		
4	**Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?** (a) Since the there is a relationship means our model is not good (b) Since the there is a relationship means our model is good (c) Can't say (d) None of these		

5	**Which of the following indicates a fairly strong relationship between X and Y?** (a) Correlation coefficient = 0.9 (b) The p-value for the null hypothesis Beta coefficient =0 is 0.0001 (c) The t-statistic for the null hypothesis Beta coefficient=0 is 30 (d) None of these		
6	**In a linear regression problem, we are using "R-squared" to measure goodness-of-fit. We add a feature in linear regression model and retrain the same model.** **Which of the following option is true?** (a) If R Squared increases, this variable is significant. (b) If R Squared decreases, this variable is not significant. (c) Individually R squared cannot tell about variable importance. We can't say anything about it. (d) None of these.		
	SECTION-B		
Q	**Attempt all the questions :**	**70 marks**	

1.	1. A survey was conducted in 295 people asking about arthritic pain on a visual analogue scale (can be treated as continuous) in 295 people. Sex was coded 1 = male 0 = female. Medication was coded 1 = on medication 0 = not on medication. The output from a multiple linear regression computer program is shown in Table 9.8. Table 9.8 Output from survey of 295 people R-squared = 0.0554 Adj R-squared = 0.0489 pain| Coef. Std. Err. t P>|t| [95% Conf. Interval] sex| -5.285991 3.294272 -1.60 0.110 -11.76952 A medication| -9.489177 3.245583 B 0.004 -15.87688 -3.101475 _cons | 94.18235 5.853297 16.09 0.000 82.66236 105.7024 {adjusted r-squared=1-(1-r-squared)(N-1)/(N-k-1) where N is number of subjects and k is the number of predictors.} (i) Deduce the values of A and B. (ii) Is the effect of medication significant? What is the assumption underlying this test? (iii) What is the expected value of the pain score in a woman not on medication? (iv) What is the expected value of the pain score for a man on medication? (v) Is the model a good fit to the data? **OR** A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, affect admission into graduate school. The outcome variable, admit/don't admit, is binary. This data set has a binary response (outcome, dependent) variable called admit, which is equal to 1 if the individual was admitted to graduate school, and 0 otherwise. There are three predictor variables: gre, gpa, and rank. We will treat the variables gre and gpa as continuous. The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. Using the GRE data, please identify how aforesaid variables affect the admission into graduate school. (Data will be shared through email.)	**20**	**CO_2**

2	**Are You Going To Hate Your New Job?** Getting a new job can be an exciting and energizing event in your life. But what if you discover after a short time on the job that you hate your job? Is there any way to determine ahead of time whether you will love or hate your job? Sue Shellenbarger of The Wall *Street Journal* discuss some of the things to look for when interviewing for a position that may provide clues as to whether you will be happy on that job. Among other things, work culture vary from hip, freewheeling start-ups to old school, organizations place pressure on workers to feel tense and to work long hours while others place more emphasis on creativity and the bottom line. Shellenbarger suggest that job interviewees pay close attention to how they are treated in an interview. Are they just another cog in the wheel or are they valued as an individual? Is a work life balance apparent within the company? Ask what a typical workday is like at that firm. Inquire about the values that undergird the management by asking questions such as "What is your proudest accomplishment". Ask about flexible schedules and how job training is managed. For example, does a worker have to go to job training on their own time? A "Work Trends" survey undertaking by the John J. Heldrich Center for Workforce Development at Rutgers University and the Center for survey Research and Analysis at the University for Connecticut posted several questions to employees in a survey to ascertain their job satisfaction. Some other things included in these questions for relationship with your supervisor, overall quality of the work environment, total hours worked each week and opportunities for advancement at the job. Suppose another research are gathered survey data from 19 employees on these questions and also ask the employees to read their job satisfaction on a scale from 0 to 100 (with 100 being perfectly satisfied). Suppose the following data represent the results of this survey. Assume that relationship which supervisor is rated on a scale from 100 to 50 (0 represent poor relationship and 50 represent an excellent relationship), oral quality of the work environment is rated on a scale from 0 to 100 (0 represents poor work environment and 100 represents an excellent work environment), and opportunities for advancement is rated on a scale from 0 to 50 (0 represents no opportunities and 50 represent excellent opportunities).	**25**	**CO_3**

Job Satisfaction	**Relationship with supervisor**	**Overall quality of work environment**	**Total hours worked per week**	**Opportunities for advancement**
55	27	65	50	42
20	12	13	60	28
85	40	79	45	7
65	35	53	65	48

	45	29	43	40	32		
	70	42	62	50	41		
	35	22	18	75	18		
	60	34	75	40	32		
	95	50	84	45	48		
	65	33	68	60	11		
	85	40	72	55	33		
	10	5	10	50	21		
	75	37	64	45	52		
	80	42	82	40	46		
	50	31	46	60	48		
	90	47	95	55	30		
	75	36	82	70	39		
	45	20	42	40	22		
	65	32	73	55	12		

Managerial and Statistical Questions

1. Several variables are presented that maybe related to job satisfaction. Which variables are stronger predictors of job satisfaction? Might other variables not mentioned here be related to job satisfaction?

2. Is it possible to develop a mathematical model to predict job satisfaction using the data given? If so, how strong is the model? With for independent variables, will we need to develop four different simple regression models and compare their results?

3.	Use the air passenger data already in build into R, prepare a time series model & predict the next ten years numbers of passengers (1961-1970). Or Using the K-means clustering algorithm, classify the wines into 3 categories based on the attributes : *Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines,* and *Proline* present in wine data. The data set is a result of a chemical analysis of wines grown in a particular region in Italy but derived from three different cultivars. The data set has 178 observations and no missing values. (data set will be shared through email.)	**25**	**CO_4**