Name:

Enrolment No:

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2019

Course: Business Analytics **Semester: II**
Programme: MBA Business Analytics **CC:DSBA 7005**
Time: 03 hrs. **Max. Marks: 100**

SET I

Instructions:

All Questions are COMPULSORY. Marks for each Question is indicated along with the question.

There is an MS Excel data file (Examination.xlsx) on your desktop that contains 3 worksheets corresponding to the first three questions of this question paper. Use the data of each of the worksheets to answer the respective questions and write the solutions in the answer book provided.

For Question No. 4 create a new worksheet with the data provided and answer the questions.

Do not create additional worksheets to answer the questions.

Once you complete the question paper save your work with your SAPID (5000abcd.xlsx) as the file name.

		Marks	CO
Q. 1	Suppose that the average waiting time for a patient at a physician's office is just over 29 minutes. To address the issue of long patient wait times, some physician's offices are using wait-tracking systems to notify patients of expected wait times. Patients can adjust their arrival times based on this information and spend less time in waiting rooms. The data in **SHEET1: PatientData** shows wait times (in minutes) for a sample of patients at offices that do not have a wait tracking system and wait times for a sample of patients at offices with such systems. a. What are the mean and median patient wait times for offices with a wait-tracking system? What are the mean and median patient wait times for offices without a wait-tracking system? b. What are the variance and standard deviation of patient wait times for offices with a wait-tracking system? What are the variance and standard deviation of patient wait times for visits to offices without a wait tracking system? c. Create a box plot on your answer sheet for patient wait times for offices without a wait-tracking system and for patient wait times for offices with a wait-tracking system. d. Do offices with a wait-tracking system have shorter patient wait times than offices without a wait-tracking system? Explain.	**(5 + 5 + 5 + 5 = 20 Marks)**	**CO 1/2**

Q. 2	In **SHEET 2: GDPYears** contains gross domestic product (GDP) values for 30 countries from 2010 to 2015 in equivalent U.S. dollars ($). How could you improve the readability of this table? Create a table that provides all these data for a user. Format the table to make it as easy to read as possible.	**20 Marks**	**CO 2**
Q. 3	Pure Water Ltd. provides maintenance service for water purifier systems throughout Dehradun. Customers contact Pure Water with requests for maintenance service on their water purifier systems. To estimate the service time and the service cost, Pure Water's managers want to predict the repair time necessary for each maintenance request. Hence, repair time in hours is the dependent variable. Repair time is believed to be related to three factors: the number of months since the last maintenance service, the type of repair problem (mechanical or electrical), and the repairperson who performs the repair (Prashant or Radesh). Data for a sample of ten service calls are reported in **SHEET 3: Repair**. a. Develop the simple linear regression equation to predict repair time given the number of months since the last maintenance service, and use the results to test the hypothesis that no relationship exists between repair time and the number of months since the last maintenance service at the 0.05 level of significance. What is the interpretation of this relationship? What does the coefficient of determination tell you about this model? b. Using the simple linear regression model developed in *part a*, calculate the predicted repair time and residual for each of the ten repairs in the data. Sort the data by residual (so the data are in ascending order by value of the residual). Do you see any pattern in the residuals for the two types of repair? Do you see any pattern in the residuals for the two repairpersons? Do these results suggest any potential modifications to your simple linear regression model? Now create a scatter chart with months since last service on the *x*-axis and repair time in hours on the *y*-axis for which the points representing electrical and mechanical repairs are shown in different shapes and/or colors. Create a similar scatter chart of months since last service and repair time in hours for which the points representing Radesh and Prashant repairs are shown in different shapes and/or colors. Do these charts and the results of your residual analysis suggest the same potential modifications to your simple linear regression model? c. Create a new dummy variable that is equal to zero if the type of repair is mechanical and one if the type of repair is electrical. Develop the multiple regression equation to predict repair time, given the number of months since the last maintenance service and the type of repair. What are the interpretations of the estimated regression parameters? What does the coefficient of determination tell you about this model? d. Create a new dummy variable that is equal to zero if the repairperson is Radesh and one if the repairperson is Prashant. Develop the multiple regression equation to predict repair time, given the number of months since the last maintenance service and the repairperson. What are the interpretations of the estimated regression parameters? What does the coefficient of determination tell you about this model? e. Develop the multiple regression equation to predict repair time, given the number of months since the last maintenance service, the type of repair, and the repairperson. What are the interpretations of the estimated regression parameters? What does the coefficient of determination tell you about this model? f. Which of these models would you use? Why?	**(5 + 5 + 8 + 8 + 8 + 6 = 40 Marks)**	**CO 3**

Q. 4	Consider the following time series data: 	Month	1	2	3	4	5	6	7
---	---	---	---	---	---	---	---
Value	24	13	20	12	19	23	15 | a. Construct a time series plot. What type of pattern exists in the data? b. Develop a three-week moving average for this time series. Compute MSE and a forecast for week 8. c. Use $\alpha = 0.2$ to compute the exponential smoothing values for the time series. Compute MSE and a forecast for week 8. d. Compare the three-week moving average forecast with the exponential smoothing forecast using $\alpha = 0.2$. Which appears to provide the better forecast based on MSE? e. Use trial and error to find a value of the exponential smoothing coefficient α that results in a smaller MSE than what you calculated for $\alpha = 0.2$.	**(4 + 4 + 4 + 4 + 4 = 20 Marks)**	**CO 4**

Name:

Enrolment No:

UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2019

Course: Business Analytics **Semester: II**
Programme: MBA Business Analytics **CC: DSBA 7005**
Time: 03 hrs. **Max. Marks: 100**

SET II

Instructions:

All Questions are COMPULSORY. Marks for each Question is indicated along with the question.

There is an MS Excel data file (Examination.xlsx) on your desktop that contains 3 worksheets corresponding to the first three questions of this question paper. Use the data of each of the worksheets to answer the respective questions and write the solutions in the answer book provided.

For Question No. 4 create a new worksheet with the data provided and answer the questions.

Do not create additional worksheets to answer the questions.

Once you complete the question paper save your work with your SAPID (5000abcd.xlsx) as the file name.

		Marks	CO
Q. 1	In a recent report, based on a sample of 50 Internet users showed that the top five most-visited English-language web sites were google.com (GOOG), facebook.com (FB), youtube.com (YT), yahoo.com (YAH), and wikipedia.com (WIKI). The data of the 50 sample is in **SHEET 1: Websites.** Based on the data provided answer the following: a. Are these data categorical or quantitative? b. Provide frequency and percent frequency distributions. c. Based on the sample, which web site is most frequently the most-often-visited web site for Internet users? Which is second and so on?	**(5 + 7 + 8 = 20 Marks)**	**CO ½**
Q. 2	According to the Central Board of Secondary Education (CBSE) teachers generally spend more than 40 hours each week working on instructional duties. The data in **SHEET 2: Teachers** show the number of hours worked per week for a sample of 13 secondary school science teachers and a sample of 11 secondary school English teachers. a. What is the median number of hours worked per week for the sample of 13 secondary school Science teachers and sample of 11 secondary school English teachers? b. Create a box plot for the number of hours worked for secondary school Science teachers and secondary school English teachers. c. Comment on the differences between the box plots for Science and English teachers.	**(5 + 7 + 8 = 20 Marks)**	**CO 2**
Q. 3	A study investigated the relationship between audit delay (the length of time from a company's fiscal year-end to the date of the auditor's report) and variables that describe the client and the auditor. Some of the independent variables that were included in this study are: **Industry** A dummy variable coded 1 if the firm was an industrial company or 0 if the firm was a bank, savings and loan, or insurance company.	**(5 + 5 + 10 + 20 = 40 Marks)**	**CO 3**

	Public A dummy variable coded 1 if the company was traded on an organized exchange or over the counter; otherwise coded 0. **Quality** A measure of overall quality of internal controls, as judged by the auditor, on a 5-point scale ranging from "virtually none" (1) to "excellent" (5). **Finished** A measure ranging from 1 to 4, as judged by the auditor, where 1 indicates "all work performed subsequent to year-end" and 4 indicates "most work performed prior to year-end." The data for a sample of 40 companies is provided in **SHEET3: Audit.** a. Develop the estimated regression equation using all of the independent variables included in the data. b. How much of the variation in the sample values of delay does this estimated regression equation explain? What other independent variables could you include in this regression model to improve the fit? c. Test the relationship between each independent variable and the dependent variable at the 0.05 level of significance, and interpret the relationship between each of the independent variables and the dependent variable. d. On the basis of your observations about the relationships between the dependent variable Delay and the independent variables Quality and Finished, suggest an alternative model for the regression equation developed in part (a) to explain as much of the variability in Delay as possible.		
Q. 4	Consider the following time series data: Month: 1 2 3 4 5 6 7 Value: 24 13 20 12 19 23 15 a. Construct a time series plot. What type of pattern exists in the data? b. Develop a three-week moving average for this time series. Compute MSE and a forecast for week 8. c. Use $\alpha = 0.2$ to compute the exponential smoothing values for the time series. Compute MSE and a forecast for week 8. d. Compare the three-week moving average forecast with the exponential smoothing forecast using $\alpha = 0.2$. Which appears to provide the better forecast based on MSE? e. Use trial and error to find a value of the exponential smoothing coefficient α that results in a smaller MSE than what you calculated for $\alpha = 0.2$.	**(4 + 4 + 4 + 4 + 4 = 20 Marks)**	**CO 4**