Course Assessment Test

Course Title	Intro and Overview of Data Science	**Date**	
Name		**Dept**	

1) *This assessment test is to be given out before course commencement. Answers are to be filled in column entitled "Pre-Course Answer"*
2) *At the end of the course, the same assessment sheet is to be given out where answers are to be filled in column entitled "Post-Course Answer". Instructor will then share the answers and participants need to total the score in both "Pre" and "Post" columns through self-marking.*
3) *Assessment sheets will be collected for filling.*

No	Question	Pre-Course Answer	Post-Course Answer
1	A data analyst would like to compute total, average, min, max of sales and gross / net profits of products as well as compare order quantity / sales / profits across categories such as product, manufacturer, region, country, etc. At what level is this analysis being performed at? a) Descriptive b) Diagnostic c) Predictive d) Prescriptive		A
2	A data analyst would like to determine how to optimize supply chain operations to obtain the ideal manufacturing rate for specific products and shipping to specific countries where high sales / profit is predicted. At what level is this analysis being performed at? a) Descriptive b) Diagnostic c) Predictive d) Prescriptive		D
3	Which of the following statements regarding the machine learning (ML) process is correct? i. ML model is created by providing a data set consisting of independent variables (IV) and a dependent variable (DV) as input to ML algo ii. This process of creating is also known as training / fitting the model iii. The model is a mathematical function that best approximates the relationship between independent variables (IV) and dependent variable (DV) iv. The trained model is used to make prediction of DV values corresponding to new IV values that it has yet to encounter a) Items i), ii) and iii) b) Items i), ii) and iv)		D

	c) Items i), iii) and iv) d) All items are correct		
4	What is the main difference between regression and classification in machine learning? a) Regression is used to perform prediction based on historical data, while classification is used to perform prediction based on patterns from current and historical data b) Classification is used to categorize the accuracy of predictions generated from a regression algorithm. c) Regression is used to predict a continuous target variable, while classification is used to predict a categorical target variable d) Regression is used to predict a categorical target variable, while classification is used to predict a continuous target variable		C
5	Machine Learning (ML) has many common use cases in multiple industry domains. What are the most common use cases for ML in supply chain and manufacturing analytics? a) Fraud Detection and Credit Risk Modeling b) Demand Forecasting and Predictive Maintenance c) Customer Sentiment Analysis and Market Basket Analysis d) Customer Churn Analysis and Customer Segmentation		B
6	In a multiple linear regression equation below, what do we call the constants b_1, b_2, … b_n $Y = b0 + b_1X_1 + b_2X_2 + \ldots. + b_nX_n$ a) Coefficients b) Intercepts c) Dependent variables d) Independent variables		A
7	What is the simplest way to determine which independent variable in a dataset has the biggest influence in minimizing / maximining the dependent variable under study? a) The range correlation between that independent variable and the dependent variable b) The standard deviation of the values within that independent variable column c) The magnitude of its coefficient relative to coefficients of other variables d) The P-value for that independent variable relative to other variables		C
8	What is a popular and widely used metric to assess the prediction accuracy of a regression model? a) Accuracy Ratio b) Predictivity Ratio c) T-Squared		D

	d) R-Squared		
9	What are some of the main reasons why a linear regression model has a low measure of accuracy (using the metric you identified in the previous question) ? i. Important independent variables have been left out in creating the model ii. The standard deviation of the independent variable columns are far apart iii. There is a non-linear functional relationship between the dependent and independent variables iv. All the independent variables have high correlation with the dependent variable a) Reason i) and ii) b) Reason i) and iii) c) Reason ii) and iii) d) Reason iii) and iv)		B
10	What is the purpose of performing dummy variable encoding in a dataset prior to using it to train a regression model? a) To account for missing values in a dataset that may significantly impact the predictive accuracy of the trained regression model b) To provide a buffer column as a countermeasure against potential overfitting of the regression model when it is trained c) To convert categorical variable values to a numeric form necessary for generating a regression model d) To allow variable values that are out of the expected standard deviation range to be normalized to more acceptable range		C
Total			