Python Assignment.

1. (C)%
2. (B)0
3. (C)24
4. (A)2
5. (D)6
6. (B) It encloses the lines of code which will be executed if any error occurs while executing the lines of code in
the try block.
7. (A)It is used to raise an exception.
8. (C) in defining a generator.
9. A) _abc and C) abc2
10. A) yield and B) raise.

Coding in Jupitor Notebook

Please see below

Machine Learning Assignment

1. (A)Least Square Error
2. (A)Linear regression is sensitive to outliers
3. (B)Negative
4. (A)Regression
5. (C)Low bias and high variance
6. (B) Predictive Model
7. (D)Regularisation Technique
8. (D) SMOTE
9. (A) TPR and FPR
10. (A) TRUE
11. (A) Construction bag of words from a email
12. (B) Apply PCA to project high dimensional data and (D)Forward selection.

13. Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This

problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

14. Particular algorithms are used for regularization

Ridge Regression (L2 Norm), Lasso (L1 Norm) ,Dropout.

15. Error term, the distance from each point to the line, is minimized. Since the relationship between variables is probably not completely linear and because there are other factors outside the scope of our study (sales on red bull, sales on other caffeine drinks, difficult physics homework sets, etc.)

Please see below:

Statistics Assignment.

1. (A)True
2. (A)Central Limit Theorm
3. (B)Modelling Bound Count Data.
4. (B) Sums of normally distributed random variables are again normally distributed even if the variables are dependent.
5. (C) Poisson.
6. (B) False
7. (B)Hypothesis
8. (A) 0
9. (C) Outliers cannot conform to the regression relationship.

10. A normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3 Normal distributions are symmetrical, but not all symmetrical distributions are normal.In reality, most pricing distributions are not perfectly normal.

11.

Missing data (or missing values) is defined as the data value that is not stored for a variable in the observation of interest. The problem of missing data is relatively common in almost all research and can have a significant effect on the conclusions that can be drawn from the data . Accordingly, some studies have focused on handling the missing data, problems caused by missing data, and the methods to avoid or minimize such in medical research .

However, until recently, most researchers have drawn conclusions based on the assumption of a complete data set. The general topic of missing data has attracted little attention in the field of anesthesiology.

Missing data present various problems. First, the absence of data reduces statistical power, which refers to the probability that the test will reject the null hypothesis when it is false. Second, the lost data can cause bias in the estimation of parameters. Third, it can reduce the representativeness of the samples. Fourth, it may complicate the analysis of the study. Each of these distortions may threaten the validity of the trials and can lead to invalid conclusions.

12.A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.
A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which you can gather both qualitative and quantitative user insights. You can further use this collected data to understand user behavior, engagement rate, pain points, and even satisfaction with website features, including new features, revamped page sections, etc. If you're not A/B testing your website, you're surely losing out on a lot of potential business revenue.

13. The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old

has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14 . Linear regression models the relationships between at least one explanatory variable and an outcome variable. These variables are known as the independent and dependent variables, respectively. When there is one independent variable (IV), the procedure is known as simple linear regression. When there are more IVs, statisticians refer to it as multiple regression.

Fitted line plot for a linear regression model that displays the relationship between height and weight.Learn more about independent and dependent variables.

This flexible analysis allows you to separate the effects of complicated research questions by modeling and controlling all relevant variables. It lets you isolate the role that each variable plays.

This procedure uses sample data to estimate the population parameters. The regression coefficients in your statistical output are the parameter estimates.

15 . There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.