WORKSHEET - 1

A) PYTHON WORKSHEET

1) C - %
2) B - 0
3) C – 24
4) A – 2
5) D – 6
6) C - The finally block will be executed no matter if the try block raises an error or not.
7) A - It is used to raise an exception.
8) C - In defining a generator.
9) A & C - _abc & abc2
10) A – Yield,
    B – Raise.

B) STAISTIC WORKSHEET

1) A – True
2) A – Central limit theory
3) D – all of the above
4) A
5) C – Poisson random variable
6) B - False
7) B - Hypothesis
8) A -0
9) C - Outliers cannot conform to the regression relationship.
10) Normal distribution – it is a bell curve, which is symmetrical about its y-axis ie mean, the data which is closer to mean are likely to occur frequently than others.
11) Missing data – missing data can be handled by either deleting or imputing it with some values, deleting data is not at all recommended (except entire rows or column is having almost missing values).
    Based on observations, there are several ways to impute missing values like NaN can be replace with MEAN, MODE, MEDIAN or with some arbitrary values, there are several sklearn libraries available like Knn imputer , ittirative imputer, which can be used to handle missing data.
12) A/B testing is an experiment in which we compare two different version of variable to find out which one is performing better.
13) Mean imputer can be used to fill missing values, but it is **NOT A GOOD PRACTICE**, as it can lead to unbiased estimates.
14) Linear regression is a analysis which helps to predict the dependent variable value based on another independent variable value.
15) There are two main branches of statistics:
    15.1) Descriptive statistics – it deals withcollection of data and its representation in the form of table,graph etc.
    15.2) Inferential statistics – it deals with analysis of data ,making estimates and giving conclusion from the data.

WORKSHEET - 1

C) MACHINE LEARNING

1) D – both A & B
2) A – linear regression is sensitive to outliers
3) B – negative
4) B – correlation
5) C – low bias and high variance
6) B – predictive model
7) D – regularization
8) A – cross validation
9) C – sensitivity and specificity
10) B – false
11) B –PCA
12) A,B,C
13) Regularization – it is a technique used to prevent the model from overfitting.
14) Some of the common regularization technique are – lasso and ridge technique, dropout technique etc.
15) Error in linear regression equation represents the average distance that the observed values fall from the regression line.