Assessment #3.

MCQ.

1. a) Python is a high level programming language.
   b) Python support multiple programming paradigms.

2. b) [3, 4, 5]

3. c) Converting dct to DF.

4. a) Specifies diamensions along which an operation is performed.

5. a) Built on top of Numpy.
   b) Provides data structure as Series & DF.

6. d) All of these.

7. c) plt.SaveFig ('output.png').

8. a) df.dropna ()

9. b) data type of array.

10. a) Slicing creates view of the original array.

11. a)
b)
d)

12. c) Michael.

13. a) Generate random int.

14. a)
d)

15. a)

16. c)

17. a)

18. a)

19. a)

20. b)

21. b) b)

Date

Page

22. b)

23. b)

24. b)
c)

25. d)

26. a)

27. a)

28. c)

29. a)

30. b)
d)

Scenario based questions-

1. i) If the missing value present in the data set is less then we will simple delete the rows

ii) If Categorical data is present then we will add up the missing values with, the mode of the Column. (mode:- most frequency data).

iii) Filling up the missing value with the mean

2. For churn prediction we can use linear Regression and Logistic Regression both
For the evaluation of the data we can use Confusion matrix, by examining the no. of observations that are correctly & Incorrectly classified.

3. KNN (K Nearst Number) will help in performing the Customer Segmentation.
With this we can Identify that what are the most sold product so we can create enough stock of it, remove the ~~up~~ unpurchased stock and replace the old stock with the most trending ones.

4. We will replace the Categorical variables with the numerical values like [0, 1, 2] by using 'map' function in a sequence. & when the each variable will become Int. then we will use it in modelling process.

Date

Page

Concept Based Questions —

1. Multicollinearity occurs when two or more Independent variables have Correlation with one another, in a model, which makes it difficult to determine the Individual effect of each Independent Variable on the dependent one.

It affects by reducing the precision of the estimated coefficients, which weakens the Statistical power of your regression model.

To handle multicollinearity first method is to plot the ~~coo~~ correlation matrix of all Independent variables
~~Second to use~~

2. 1) Find point which are far away from the hyperplane.

Presence of outlier can change the magnitude of regression Coefficients & even the direction of Coefficients.

For detecting outliers we can use graphical (such as scatter plot) and analytical method.

We can ~~handle~~ handle outliers by simply removing them.

3. Overfitting is a problem in which our model every single data present in the dataset on we it fits too closely to the training data, and because of this model cannot generalize.

Lasso - It is a regularization tech. used in the feature selection using a Sinkage method.
It act by determing the coefficient in the linear model and Shrunk towards the central point.

Ridge Regression - Ridge Rigression put similar constraint on the coefficient by introducing a penalty factor.

4. It could enhanced by applying correct regression model that help is predicting the a correct values.

Polynomical regression is a form of regression analysis in which the relationship b/w the Independent variables & dependent variables are modeled. in the nth degree polynomical.

We can choose the degree based on the relationship b/w the target & predictor.