

DAV QUESTION BANK MODULE WISE

MODULE-I

1. Explain Data Analytics life cycle?
2. Why is Data Analytics life cycle is essential?
3. Explain key output from a successful Analytics project?
4. Explain all phases?
5. Explain the differences between Business Intelligence analyst and Data Scientist.
6. What is an analytic sandbox, and why is it important?
7. What kinds of tools would be used in the following phases, and for which kinds of use scenarios?

MODULE-II

1. What is the difference between Linear Regression and Logistic Regression?
2. Explain the terms: - R.M.S.E., S.E., R^2 Statistics
3. By using formula of simple Linear Regression predict the value of Y
4. By using formula of Multi Linear Regression for 2 input variables predict the value of Y.
5. Explain life cycle of K fold cross validation.
6. Explain Model Selection.
7. Explain two directions model selections
8. Explain Stepwise regression and types of stepwise regression.
9. Explain Prediction using regression with example.
10. Advantages and Disadvantages of Logistic regression.
11. Explain linear classification with logistic regression.

MODULE-III

1. Explain Time series analysis.
2. Discuss time series for Autocorrelation.
3. What is Box-Jenkins intervention analysis methodology?
4. Discuss application of autocorrelation
5. Explain Time series Forecasting
6. Explain ARIMA model in detail.
7. Advantages and Disadvantages of ARIMA MODELS.
8. Explain Building and Evaluating of ARIMA model

MODULE-IV

1. What is text mining? Explain text mining architecture?
2. Explain application of text mining?
3. Explain Advantages and disadvantages of text mining.
4. What are the seven practices of text analytics?
5. Write a short note on POS tagging, lemmatization and tokenization.
6. What is the equation of relative frequency of term t in document d , and write the variants of Term Frequency-Inverse Document Frequency (TFIDF) Weight?
7. How do we determine Sentiment Analysis? / Explain Sentiment Analysis?

1. Explain different roles of people in data science project?
2. List the phases of Life cycle of data analytics.
3. What is an analytic sandbox, and why is it important?
4. Explain the differences between Business Intelligence analyst and Data Scientist.
5. What kinds of tools would be used in the following phases, and for which kinds of use scenarios?
 - a. Phase 2: Data preparation
 - b. Phase 4: Model building
 6. Explain following phases of data analytics in detail:-
 - a. Data Preparation
 - b. Model Building
 7. What is the difference between Linear Regression and Logistic Regression?
 8. Explain following terms:-
R.M.S.E., S.E., R^2 Statistics
 9. Explain Logit/log-odds function in detail?
 10. Write a short note on all 4 components of Time Series.
 11. Explain Autoregressive Model.
 12. Explain Moving Average Model.
 13. Explain A.R.I.M.A model.
 14. By using formula of simple Linear Regression predict the value of Y .
 15. By using formula of Multi Linear Regression for 2 input variables predict the value of Y.

1 Draw a diagram of Data analytics life cycle. Explain any one phase in detail. Name the tools used for data preparation phase

2 A Consider the following data

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

Find the regression equation and coefficient of determination and predict y value for $x=6.2$

B Consider the data collected from 410 customers in a restaurant. It is observed that 40 of 70 customers tipped the server A and 130 of 340 customers tipped the server B. Compute a Logit or log-odds of tipping server A

C Multiple regression analysis produced the following table.

	Coefficients	Standard Error	t Statistic	p-value
Intercept	1929.116	1033.29	1.866966	0.077415
x_1	-17.3064	20.77954	-0.83286	0.415271
x_2	-3.87535	0.336015	-11.5333	5.05E-10

Calculate Predicted value of y for $x_1=50$ and $x_2=80$

3 A What are the components of time series?

B Calculate the seasonal index for first season for following data(Use method of simple average)

Year	1 st season	2 nd season	3 rd season	4 th season
2016	75	60	54	59
2017	86	65	63	80
2018	90	72	66	85
2019	100	78	72	93

1	A	List advantages and disadvantages of Text Analytics
	B	<p>1. Assume a suitable example for text analytics with the use of seven practices of text analytics.</p> <hr/> <p style="text-align: center;">OR</p> <hr/> <p>2. Assume any seven social media comments for xyz Operating system and examine it with any one type of sentiment analysis. (example - "1. Love the user interface. Setup took five minutes and we were ready to go" 2. Took me 2 hours to set up, then I find out I have to update my OS. Love it!")</p>
2	A	List the features of R programming.
	B	<p>1. By using a simple regression method, write steps to import data set and export data using R programming.</p> <hr/> <p style="text-align: center;">OR</p> <hr/> <p>2. Apply data exploration and data visualization on rainfall dataset {12,13,23,24,55,32,45,23,15,44,15,23}</p>
3	A	Explain use of the seaborn library in data analytics.
	B	<p>1. by using the Matplotlib library, write a python program to plot time series data.</p> <hr/> <p style="text-align: center;">OR</p> <hr/> <p>2. By using the panda library, write a python program using series and dataframe data structure.</p>

- a Define Text analytics.
- b Write code snippets to create dataframe in R programming.
- c Write any 4 available functions to perform data analysis in Python.
- d Which are 3 different methods to perform sentiment analysis?
- e Write a python code to plot A.C.F .

- f Which functions are used to handle dirty data in R Programming?

Q 2. Attempt any ONE

- a Explain T.F.I.D.F with example
- b Write a short note on Panda, Numpy, Scipy libraries to perform data analysis in Python

Q 3. Attempt any ONE

- a How to build and evaluate ARIMA Model ?
- b Explain following related to R Programming.
 - 1. Data types 2. import, export 3. data structures

Q.	
1	What do the terms p, d, and q in the ARIMA (p, d, q) model denote? How to choose value of p?
2	a. Name the seven practices of text analytics
	b. Write a short note on POS tagging, lemmatization and tokenization
3	a. Name any four built in R functions that include descriptive statistics.
	b. Explain vector and data frame in R also write the code to create these data structures OR Explain any two components of the graphics grammar (ggplot) in R and provide R code example demonstrating how to apply them.
4	a. Write a code in python to create horizontal bar chart (assume suitable data) OR Write a code in python to create scatter plot (assume suitable data)(also add some color for points)
	b. Write a short note on any one library in python : _____ Numpy or Pandas or Scipy