**DAV QUESTION BANK**

**Class: TE/SEM VI Subject: DAV (CSC601)**

**MODULE 1 Introduction to Data analytics and life cycle**

1. Draw a diagram of Data analytics life cycle. Explain any one phase in detail. Name the

tools used for data preparation phase. 5m

2. Explain different roles of people in data science project?

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. Explain the phases of Data analytics of life cycle in detail.10m

**MODULE 2 Regression Models**

1. What is the difference between Linear Regression and Logistic Regression? 3m

2. Explain following performance metrics terms: -

a. R.M.S.E

b. S.E

c. R2Statistics

d. A.I.C

e. MAE

3. Explain Logit/log-odds function in detail?

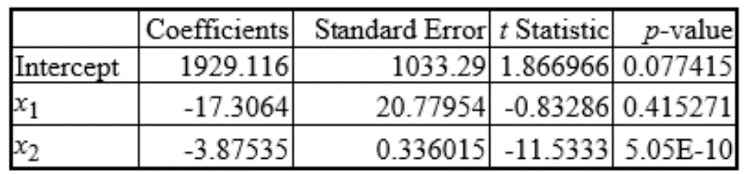
4. Consider the following data 5m

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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.

5. 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. 2m

6. Multiple regression analysis produced the following table.



Calculate Predicted value of y for x1=50 and x2=80. 5m

# 7. What is the difference between R-square and Adjusted R-square? 3m

# 8. What is the difference between Correlation and Regression? 3m

# 9. Explain the cross-validation methods in detail. 5m

# 10. What are the principles and features of model selection? 5m

# 11. Explain step-wise regression and how prediction is done using regression. 10m

# 12. How generalised linear model is used in constructing the model? 5m

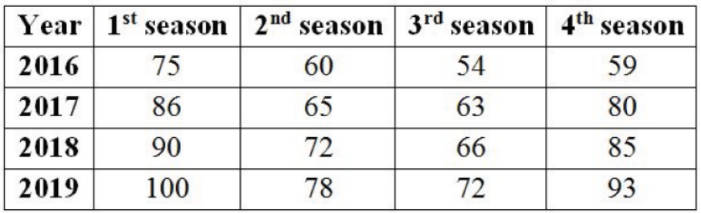
# 13. What terms are to be considered to assess the regression models? Explain them in detail. 5m

# Note: Refer the solved problems on regressions in Techneo publications.

**MODULE 3 Time Series**

1. Explain Time series analysis. What are the components of time series? 3m

2. Calculate the seasonal index for first season for following data (Use method of simple average) 5m



3. Explain Moving Average Model and Autoregressive Model and their orders with example.10m

4. How to build and evaluate ARIMA Model? 5m

5. What do the terms p, d, and q in the ARIMA (p, d, q) model denote? How to choose value of p?3m

6. Explain time series for auto correlation. 5m

7. Explain Box-Jenkins Intervention Analysis. 10m

8. Discuss how to plot the Autocorrelation and Partial Autocorrelation functions. 5m

9. Explain time-series forecasting. 3m

10. What are the advantages’ of ARIMA model. 2m

**Note:** For solved problems refer

https://www.brainkart.com/article/Exercise-9-1--Time-Series-Analysis-and-Measurements-of-Trends\_39020/

**MODULE 4 Text Analytics**

1. Define Text analytics. 2m

2. Write a short note on text mining and its architecture.3m

3. Explain applications and use cases of text mining. 4m

4. Which are 3 different methods to perform sentiment analysis? 2m

5. Explain the methos and features of summarizing text.

6. Explain T.F.I.D.F with example.5m

7. Explain the steps of text analytics. 5m (Kindly refer the notes)

8. Name the seven practices of text analytics.5m

9. Write a short note on POS tagging, lemmatization and tokenization. 5m

10. List advantages and disadvantages of Text Analytics.2m

11. Assume a suitable example for text analytics with the use of seven practices of text analytics.3m

(Refer Techneo)

12. Explain the importance and types of raw data.3m

13. Assume any seven social media comments for xyz Operating system and examine it with any one type of sentiment analysis. 4m

(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!) 5m

(Refer <https://www.analyticsvidhya.com/blog/2021/12/different-methods-for-calculating-sentiment-score-of-text/>)

14. How the data and image are represented in data analysis? 3m

15. Explain about the need of document classification.4m

16. What are the insights of data analytics? 2m

**MODULE 5 Data analytics and visualization with R**

1. List the features of R programming. 2m

2. Write code snippets to create data frame in R programming.3m

3. By using a simple regression method, write steps to import data set and export data using R

programming. 5m

4. Apply data exploration and data visualization on rainfall dataset. 5m

{12,13,23,24,55,32,45,23,15,44,15,23}

5. Which functions are used to handle dirty data or data cleaning in R Programming?3m

6. Explain following related to R Programming. 5m

1. Data types 2. import, export 3.data structures

7. Name any four built in R functions that include descriptive statistics. 2m

8. Explain vector and data frame in R also write the code to create these data structures. 3m

9. Explain any two components of the graphics grammar (ggplot) in R and provide R code example

demonstrating how to apply them. 3m

10. Explain List, array in R also write the code to create these data structures.5m

11. How the factor data type can be used? Give some examples. 3m

12. Give the use of the four window panes of RStudio.5m

13.Write a code to create histogram, bar chart, pie chart, boxplot, scatter plot and line plot in R

(Assume any suitable data) 5m

14. Explain the visualizing of single variables and multiple variables.3m

15. Explain about exploratory data analysis.5m

16. Differentiate data exploration and data presentation.3m

**MODULE 6 Data analytics and Visualization with Python**

1. Write any 4 available functions to perform data analysis in Python. 2m

2. Write a python code to plot A.C.F. 2m

3. Write a short note on Panda, Numpy, SciPy libraries to perform data analysis in Python. 5m

4. Explain use of the seaborn library in data analytics. 2m

5. By using the Matplotlib library, write a python program to plot time series data. 5m

6. By using the panda library, write a python program using series and data frame data structure. 5m

7. Write a code in python to create histogram, bar chart, pie chart, box plot, scatter plot, line plot

horizontal bar plot, violin plot using matplotlib (assume suitable data)10m

8. Explain multiple plots in detail. 4m

9. Explain the following plots in detail. 5m

a. Regression plot b. Regplot

10. What are the functions available in Panda’s library of python? 5m

11. Explain different categories of plot in seaborn. How do you use seaborn library for regression

task? 3m

12. Explain pandas in detail. Explain the utility of ‘group by’ function in pandas with example. 10m

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

EXERCISE ON PYTHON LIBRARIES

