

Royal College of Arts, Science and Commerce

T.Y.B.Sc Computer Science

Semester VI 2019-2020

Data Science (Course code: - USCS606)

Sample Questions

- Q1) Data that summarize all observations in a category are called as _____ [1M]
- a) frequency data b) summarized data
c) raw data d) analyzed data
- Q2) Data Science is a field that _____. [1M]
- a) Extracts meaningful information, make decisions and predictions.
b) Extracts raw information
c) Separates structured and unstructured data
d) Contains semi structured data
- Q3) _____ Data is difficult to manipulate and typically needs to be processed in some way before it can be used in standard data analysis software. [1M]
- a) Structured data b) Unstructured data
c) Summarized data d) Frequency data
- Q4) Missing values are coded by _____ symbol and _____ function is used to identify missing values in data set. [1M]
- a) AN , Any() b) NA , is.na()
c) NA, missing() d) AN, isempty()
- Q5) _____ plot shows comparison between discrete categories. [1M]
- a) Scatter b) Bar
c) Histogram d) Line
- Q6) _____ is the process of creating visual representation of data to bring some meaningful insights. [1M]
- a) Data Curation b) Data Visualization
c) Data Cleaning d) Data Transformation

- Q7) _____ is the process of transforming data from one format to another. [1M]
- a) Data Mining
 - b) Data wrangling
 - c) Data visualization
 - d) Data transformation
- Q8) _____ is an unsupervised learning method. [1M]
- a) Linear regression
 - b) Cluster Analysis
 - c) Logistic regression
 - d) SVM
- Q9) For k number of parameters, N number of observations, and a likelihood L, the equation to estimate BIC is _____. [1M]
- a) $-2*\ln(L) + \ln(N)*k$
 - b) $-2*\ln(L) + \ln(k)*N$
 - c) $-2*\ln(L) - \ln(N)*k$
 - d) $2*\ln(L) - \ln(k)*N$
- Q10) Quantitative data is also called as _____ data. [1M]
- a) Ordinal
 - b) Numerical
 - c) Categorical
 - d) Nominal
- Q11) _____ are techniques used to reduce error by avoiding overfitting. [1M]
- a) Smoothing
 - b) Aggregation
 - c) Cross validation
 - d) Regularization
- Q12) The component of TSA that indicates seasonality patterns that are not repeating at regular frequency is _____. [1M]
- a) Trend
 - b) Seasonality
 - c) Noise
 - d) Cycle
- Q13) L1 and L2 regression are _____ techniques. [1M]
- a) Validation
 - b) Regularization
 - c) Approximation
 - d) Ensemble
- Q14) MongoDB is _____. [1M]
- a) Relational database
 - b) Document database
 - c) Cloud database
 - d) SQL database
- Q15) _____ function is used in R for logistic regression. [1M]
- a) rpart()
 - b) glm()
 - c) lm()
 - d) ts()

- Q16) Amazon Web Services falls into _____ cloud-computing category. [1M]
- a) Platform as a Service
 - b) Software as a Service
 - c) Infrastructure as a Service
 - d) Back-end as a Service
- Q17) The penalty term in Ridge regression is _____. [1M]
- a) Absolute value of magnitude of coefficients
 - b) Square of magnitude of coefficients
 - c) Product of magnitude of coefficients
 - d) Sum of product of magnitude of coefficients
- Q18) What is the correct order of life cycle of Data science process? [1M]
- a) Frame the problem, collects raw data , clean data , EDA , generate reports
 - b) Frame the problem, collects raw data , EDA, clean data , communicates the results
 - c) Collects raw data, frame the problem, clean data , generate reports, data analysis
 - d) Collects raw data, clean data , frame the problem, generate reports, data analysis
- Q19) Multivariate data involves _____. [1M]
- a) One variable
 - b) Two variables
 - c) Zero variable
 - d) Three or more variables
- Q20) The best fitting model will have a ____ AIC and a ____ BIC. [1M]
- a) Low, low
 - b) High, high
 - c) Low, high
 - d) High, low