**Name of the Teacher: Ms. P. S. Patil**

**Class: BE** **Subject: Data Analytics**
**AY: 2020-21** **SEM: II**

**UNIT-1**

| | |
|---|---|
| 1) | What is Big Data? |
| | a) Huge amount of data<br>b) Small amount of data<br>c) Huge File<br>d) Big Storage |
| **Ans:** | **a** |
| **Explanation:** | It is Huge amount of data |
| 2) | According to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop? |
| | a) Big data management and data mining<br>b) Data warehousing and business intelligence<br>c) Management of Hadoop clusters<br>d) Collecting and storing unstructured data |
| **Ans:** | **a** |
| **Explanation:** | Big data management and data mining |
| 3) | What are the main components of Big Data? |
| | a)MapReduce<br>b)HDFS<br>c)YARN<br>d)All of these |
| **Ans:** | **d** |
| **Explanation:** | All of these |
| 4) | The sources of Big Data are |
| | a)Stock Exchange<br>b)Transport Data<br>c) Banking Data<br>d) All of the Above |
| **Ans:** | **d** |
| **Explanation:** | |
| 5) | Big Data Characteristics are: |
| | a) Structured data<br>b) Semi-structured data<br>c) Quasi-structured data<br>d) All of the above |
| **Ans:** | **d** |
| **Explanation:** | |
| 6) | Bl tends to provide reports, dashboards, and queries on business |

| | |
|---|---|
| | questions for the current period or in the past. |
| | a) True<br>b) False |
| **Ans:** | **a** |
| **Explanation:** | |
| 7) | Big data can come in multiple forms, including structured and non-structured data |
| | a) True<br>b) False |
| **Ans:** | **a** |
| **Explanation:** | |
| 8) | BI problems tend to require highly structured data organized |
| | a) Rows<br>b) Columns<br>c) Accurate Reporting<br>d) All of the Above |
| **Ans:** | **d** |
| **Explanation:** | |
| 9) | EDW achieves the objective of reporting and sometimes the creation of dashboards, perform analysis on unstructured data |
| | a) High-value data is hard to reach and leverage<br>b) Data moves in batches from EDW to local analytical tools<br>c) Data Science projects will remain isolated<br>d) All of the Above |
| **Ans:** | **d** |
| **Explanation:** | |
| 10) | Drivers of Big Data |
| | a) Medical information<br>b) Photos and video footage uploaded to the World Wide Web<br>c) data extracts<br>d) Both a and b |
| **Ans:** | d |
| **Explanation:** | |
| 11) | According to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop? |
| | a) Big data management and data mining<br>b) Data warehousing and business intelligence<br>c) Management of Hadoop clusters<br>d) Collecting and storing unstructured data |
| **Ans:** | **a** |
| **Explanation:** | |
| 12) | Select from option which is not the phase of data analytics |

| | |
|---|---|
| | |
| | a) model planning<br>b) testing<br>c) discovery<br>d) operationalize |
| **Ans:** | **b** |
| **Explanation:** | |
| 13) | Which phase of data analytics require more time to complete |
| | a) Data preparation<br>b) model building<br>c) communicate results<br>d) Discovery |
| **Ans:** | **a** |
| **Explanation:** | |
| 14) | What is analytic sandbox? |
| | a) Tool<br>b) Separate repository<br>c) data cleaning<br>d) Data conditioning |
| **Ans:** | **b** |
| **Explanation:** | |
| 15) | The person which provides analytic techniques and modeling is called as. |
| | a) Data Engineer<br>b) Data scientist<br>c) Business user<br>d) Project manager |
| **Ans:** | **b** |
| **Explanation:** | |
| 16) | What is task of Project manager? |
| | a) analytic modelling<br>b) Provide requirement<br>c) ensure meeting objectives<br>d) creates DB environment |
| **Ans:** | **c** |

| | |
|---|---|
| **Explanation:** | |
| 17) | Identifying Key Stakeholders this task is performed in which phase? |
| | a) Data preparation<br>b) model building<br>c) Discovery<br>d) communicate results |
| **Ans:** | **c** |
| **Explanation:** | |
| 18) | ETL process is performed in which phase |
| | a) Discovery<br>b) communicate results<br>c) model planning<br>d) Data preparation |
| **Ans:** | **d** |
| **Explanation:** | |
| 19) | How much data Data science teams prefer for analysis? |
| | a) too little<br>b) average<br>c) more<br>d) more than average |
| **Ans:** | **c** |
| **Explanation:** | |
| 20) | select from option tool which is not used in model planning phase |
| | a) Data wrangler<br>b) R<br>c) SQL Analysis service<br>d) SAS/ACESS |
| **Ans:** | **c** |
| **Explanation:** | |

| | |
|---|---|
| 21) | if reports and dashboards will be impacted and need to change this task is performed by. |
| | a) Project sponsor<br>b) BI Analyst<br>c) Data Engineer<br>d) Project manager |
| **Ans:** | **b** |
| **Explanation:** | |
| 22) | What is need of data analytic lifecycle. |
| | a) Data cleaning<br>b) To solve Big data problems<br>c) Data conditioning<br>d) Data Exploration |
| **Ans:** | **b** |
| **Explanation:** | |
| 23) | How many phases are there in data analytic lifecycle? |
| | a) 4<br>b) 5<br>c) 6<br>d) 7 |
| **Ans:** | **c** |
| 24) | The person with technical skills is called as? |
| | a) Business user<br>b) Data Engineer<br>c) Data scientist<br>d) Project sponsor |
| **Ans:** | **b** |
| 25) | What is outcome of Model building phase? |
| | a) Analytic results<br>b) Quality data<br>c) Data<br>d) Potential resources |
| **Ans:** | **a** |

Pravin S.Patil

Subject Teacher

**Name of the Teacher: Ms. P. S. Patil**

**Class: BE** **Subject: Data Analytics**
**AY: 2020-21** **SEM: II**

**UNIT-1I**

| | |
|---|---|
| 1) | 1. A statement made about a population for testing purpose is called? |
| | a) Statistic<br>b) Hypothesis<br>c) Level of Significance<br>d) Test-Statistic |
| **Ans:** | **b** |
| **Explanation:** | |
| 2) | If the assumed hypothesis is tested for rejection considering it to be true is called? |
| | a) Null Hypothesis<br>b) Statistical Hypothesis<br>c) Simple Hypothesis<br>d) Composite Hypothesis |
| **Ans:** | **a** |
| **Explanation:** | |
| 3) | A statement whose validity is tested on the basis of a sample is called? |
| | a) Null Hypothesis<br>b) Statistical Hypothesis<br>c) Simple Hypothesis<br>d) Composite Hypothesis |
| **Ans:** | **b** |
| **Explanation:** | |
| 4) | A hypothesis which defines the population distribution is called? |
| | a) Null Hypothesis<br>b) Statistical Hypothesis<br>c) Simple Hypothesis<br>d) Composite Hypothesis |
| **Ans:** | **c** |
| **Explanation:** | |
| 5) | If the null hypothesis is false then which of the following is accepted? |
| | a) Null Hypothesis<br>b) Positive Hypothesis<br>c) Negative Hypothesis<br>d) Alternative Hypothesis. |
| **Ans:** | **d** |
| **Explanation:** | |

| | |
|---|---|
| 6) | The rejection probability of Null Hypothesis when it is true is called as? |
| | a) Level of Confidence<br>b) Level of Significance<br>c) Level of Margin<br>d) Level of Rejection |
| **Ans:** | **b** |
| **Explanation:** | |
| 7) | The point where the Null Hypothesis gets rejected is called as? |
| | a) Significant Value<br>b) Rejection Value<br>c) Acceptance Value<br>d) Critical Value |
| **Ans:** | **d** |
| **Explanation:** | |
| 8) | If the Critical region is evenly distributed then the test is referred as? |
| | a) Two tailed<br>b) One tailed<br>c) Three tailed<br>d) Zero tailed |
| **Ans:** | **a** |
| **Explanation:** | |
| 9) | The type of test is defined by which of the following? |
| | a) Null Hypothesis<br>b) Simple Hypothesis<br>c) Alternative Hypothesis<br>d) Composite Hypothesis |
| **Ans:** | **c** |
| **Explanation:** | |
| 10) | Which of the following is defined as the rule or formula to test a Null Hypothesis? |
| | a) Test statistic<br>b) Population statistic<br>c) Variance statistic<br>d) Null statistic |
| **Ans:** | **a** |
| **Explanation:** | |
| 11) | Type 1 error occurs when? |
| | a) We reject H0 if it is True<br>b) We reject H0 if it is False<br>c) We accept H0 if it is True<br>d) We accept H0 if it is False |
| **Ans:** | **a** |
| **Explanation:** | |
| 12) | The probability of Type 1 error is referred as? |
| | a) 1-α |

| | |
|---|---|
| | b) β<br>c) α<br>d) 1-β |
| **Ans:** | **c** |
| **Explanation:** | |
| 13) | Alternative Hypothesis is also called as? |
| | a) Composite hypothesis<br>b) Research Hypothesis<br>c) Simple Hypothesis<br>d) Null Hypothesis |
| **Ans:** | **b** |
| **Explanation:** | |
| 14) | Which of the following is required by K-means clustering? |
| | a) defined distance metric<br>b) number of clusters<br>c) initial guess as to cluster centroids<br>d) all of the mentioned |
| **Ans:** | **d** |
| **Explanation:** | |
| 15) | Point out the wrong statement. |
| | a) k-means clustering is a method of vector quantization<br>b) k-means clustering aims to partition n observations into k clusters<br>c) k-nearest neighbor is same as k-means<br>d) none of the mentioned |
| **Ans:** | **c** |
| **Explanation:** | |
| 16) | Hierarchical clustering should be primarily used for exploration. |
| | a) True<br>b) False |
| **Ans:** | **a** |
| **Explanation:** | |
| 17) | Which of the following function is used for k-means clustering? |
| | a) k-means<br>b) k-mean<br>c) heatmap<br>d) none of the mentioned |
| **Ans:** | **a** |
| **Explanation:** | |
| 18) | Which of the following clustering requires merging approach? |
| | a) Partitional<br>b) Hierarchical<br>c) Naive Bayes<br>d) None of the mentioned |
| **Ans:** | **b** |
| **Explanation:** | |

| | |
|---|---|
| 19) | K-means is not deterministic and it also consists of number of iterations. |
| | a) True<br>b) False |
| **Ans:** | **a** |
| 20) | Depending on acceptance and rejection of null hypothesis there are 2 types of error produced |
| | a) Type 1<br>b) Type 2<br>c) None of these<br>d) All of these |
| **Ans:** | **d** |
| 21) | The power of a test can be defined as a possibility of … |
| | a) Rejecting null hypothesis<br>b) Accepting null hypothesis<br>c) Increasing null hypothesis<br>d) Decreasing null hypothesis |
| **Ans:** | **a** |
| 22) | For a fixed significance level, a greater sample size is mandatory to discover a |
| | a) Minor difference in mean<br>b) Major difference in mean<br>c) Average difference in mean<br>d) None of the above |
| **Ans:** | **a** |
| 23) | ANNOVA tests if any of the population means vary from other population means |
| | a) True<br>b) False |
| **Ans:** | **a** |
| 24) | Clustering is defined as group of same kind of objects which are gathered by use of |
| | a) Unsupervised method<br>b) Supervised method<br>c) Semi supervised method<br>d) None of these |
| **Ans:** | **a** |
| 25) | Following are the applications of Kmeans |
| | a) Image Processing<br>b) Medical<br>c) Customer Segmentation<br>d) All of the above |

| | |
|---|---|
| **Ans:** | **d** |

Pravin S.Patil

Subject Teacher