| 1. | Using various techniques and processes the raw data is transformed into a format which could be further visualized and be made human readable. | Data analytics |
|---|---|---|
| 2. | The process of describing historical trends in data and aims to answer what happened which involves measuring traditional indicators such as return of investment , indicators used will be different for each industry | Descriptive analytics |
| 3. | A request for information with certain characteristics from a database | Data query |
| 4. | Collections of tables, charts, maps, and summary statistics that are updated as new data become available. | Data dashboards |
| 5. | The use of analytical techniques for better understanding patterns and relationships that exist in large data sets | Data mining |
| 6. | Examples of techniques defined in #5 | Cluster analysis, sentiment analysis |
| 7. | Consists of techniques that use models constructed from past data to predict the future or ascertain the impact of one variable on another | Predictive analysis |
| 8. | Examples of techniques defined in #7 | Linear regression, time series ana. |
| 9. | involves the use of probability and statistics to construct a computer model to study the impact of uncertainty on a decision. | Simulation |
| 10. | Prescriptive models that rely on a rule or set of rules are often referred to as | Rule-based models |
| 11. | Prescriptive analytics provide a forecast or prediction, but do not provide a ___. A forecast or prediction, when combined with a rule, becomes a ___. | Decision Prescriptive model |
| 12. | Use historical investment return data to determine the mix of investments that yield the highest expected return while controlling or limiting exposure to risk. | Portfolio models |
| 13. | Provide the cost-minimizing plant and distribution center locations subject to meeting the customer service requirements | Supply network design models |
| 14. | Use historical data to yield revenue-maximizing discount levels and the timing of discount offers when goods have not sold as planned. | Price-markdown models |
| 15. | Models that give the best decision subject to constraints of the situation | Optimization models |
| 16. | Combines the use of probability and statistics to model uncertainty with optimization techniques to find good decisions in highly complex and highly uncertain settings | Simulation optimization |
| 17. | Used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events. | Decision analysis |
| 18. | This theory performs the assignment of values to outcomes based on the decision maker's attitude toward risk, loss, and other factors | Utility theory |
| 19. | Every day, we create ___ bytes of data | 2.5 quintillion |
| 20. | It is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it | Big data |
| 21. | Types of charts:<br>• Visual representation of text data. This chart needs a large data and the discrimination's degree.<br>• Represents the transformation of data across a continuous time interval<br>• It gives the exact timing and the activity's progress in contrast to the requirements.<br>• It is used to fill the area between the axis and the polyline with color to represent better trend data | Word cloud chart<br>Line chart<br>Gantt chart<br>Area chart |

| | | |
|---|---|---|
| | • Describes which data variable has higher values and which has the lower values. Moreover, it uses to compare various quantized charts | Radar chart |
| | • Represents the proportion of various groups. It is mostly used for only one data series. | Pie chart |
| | • Presents the variable distribution in points to reveal the correlation among the rectangular coordinate system variables. | Scatter plot |
| | • It is used to represent the proportion of individual stages and reflects the individual module's size. It also supports ranking comparison. | Funnel chart |
| | • It is a scatter plot's variation. Here, the bubble area can be represented as the 3rd value | Bubble chart |
| 22. | AI and ML applications of data analytics | |
| | • It is a system that can improve its structure depending on the information that flows within the network. It considers being highly dependable in forecasting applications and business classification. | Artificial neural networks |
| | • A tree-shaped model that describes a regression or classification model. It distributes the data sets into smaller subsets and develops in a similar tree simultaneously | Decision trees |
| | • It is domain-independent DA techniques that use to examine extensive search space efficiently. | Evolutionary programming |
| | • Depends on the probability that helps in handling the difficulties in data mining techniques. | Fuzzy logic |
| 23. | The 5 Vs of data: | Volume, variety, velocity, veracity, value |
| 24. | Hurdles and risks of big data | Unstructured data, lack of privacy and security, inconsistent/poor quality |
| 25. | An area of science that manages, manipulates, extracts, and interprets knowledge from tremendous amount of data, especially on big data | Data science |
| 26. | A particular unit or thing is called an _____ or object | Entity |
| 27. | It is a measurable or observable property of an entity | attribute |
| 28. | A measurement of an attribute | Data |
| 29. | Data _____ an entity. | Defines |
| 30. | Identify the operations allowed and types used in the following properties/operations of data | |
| |     Distinctiveness | = != ->Categorical |
| |     Order | < <= > >= ->Categorical |
| |     Addition | + - Numerical |
| |     Multiplication | * / Numerical |
| 31. | What property/ies do each of the NOIR classification exhibit? | N – distinctiveness<br>O – distinctiveness and order<br>I – Distinct, order, additive<br>R – All properties |
| 32. | A variable that takes a value among a set of mutually exclusive codes that have no logical order is known as a: | Nominal variable |
| 33. | The nominal scale is used to label data categorization using a consistent naming conventional ___.<br>A nominal data may be numerical in form, but the numerical values have no ___ | Naming convention<br>Mathematical interpretation |
| 34. | A nominal variable with exactly two mutually exclusive categories that have no logical order is known as a: | Binary variable |

| 35. | If two choices of a binary variable have equal importance, then it is called: | Symmetric binary variable |
|---|---|---|
| 36. | Summary statistics applicable to nominal data are mode, ___, etc. | Contingency correlation |
| 37. | Summary measures ___ and ___ can be used on ordinal data. | Mode and median |
| 38. | Ordinal data can be ranked (numerically, alphabetically, etc.) Hence, we can find any of the __ of ordinal data. | Percentile measures |
| 39. | ____ can be used as a measure of the strength of association between two sets of ordinal data | Spearman's R |
| 40. | Interval-scale variables are _____ measurements of a roughly linear scale. | Continuous |
| 41. | _____ data has a zero point on origin but does not imply a true absence of the measured characteristics. | Interval |
| 42. | Interval data with a clear definition of "zero" are called ____ data. | Ratio |
| 43. | (T or F) Ratio data may be in linear or non-linear scale. | T |
| 44. | (T or F) Both interval and ratio data can be stored in same data type | T |
| 45. | (T or F) Interval data can be transformed to nominal or ordinal scale, but with loss of information. | T |
| 46. | (T or F) Interval data can be graphed using histogram, frequency polygon, etc. | T |
| 47. | (T or F) Linear (e.g. cx + d ) or Affine transformations are not permissible in interval data | F |
| 48. | (T or F) Negation (changing the sign) and multiplication by a constant are not permitted in interval data. | F |
| 49. | Usually ____ operators can be used on ordinal data | Relational |
| 50. | (T or F) Nominal data cannot be visualized using line charts, bar charts or pie charts. | F |
| 51. | A nominal variable with exactly two ____ categories that have no logical order is known as binary variable | Mutually exclusive |
| 52. | (T or F) In nominal data, Labels cannot be combined for another nominal variable. | F |
| 53. | The _____ method involves human or mechanical observation of what people actually do or what events take place during a buying or consumption situation. | Observation |
| 54. | Best method for gathering quickly needed information; this is where responses are collected from the respondents by the researcher on telephone. | Telephonic interview |
| 55. | It is the most versatile of the all methods. They are used when props are required along with the verbal response non-verbal responses can also be observed | Personal interview |
| 56. | In this method of data gathering, Questionnaires are sent to the respondents; they fill it up and send it back. | Mail survey |
| 57. | A process of recognizing and noting people, objects, occurrences rather than asking for information | Electronic interview |
| 58. | In __, There is no relying on willingness or ability of respondent | Electronic interview |
| 59. | A _____ is one where the treatments are assigned completely at random so that each experimental unit has the same chance of receiving any one treatment. | Completely randomized design (CRD) |
| 60. | In this design, Blocks are formed in such a manner that each block contains as many plots as a number of treatments so that one plot from each is selected at random for each treatment | Randomized block design (RBD) |
| 61. | A ____ is one of the experimental designs which has a balanced two way classification scheme say for example - 4 X 4 arrangement. | Latin Square |
| 62. | The ____ in Latin Square will not get disturbed if any row gets changed with the other, thus giving it its main strength. | Balance arrangement |
| 63. | This design allows the experimenter to test two or more variables simultaneously. It also measures interaction effects of the variables and analyzes the impacts of each of the variables. | Factorial designs |
| 64. | In a true experiment, randomization is essential so the experimenter can infer cause and effect without any _. | Bias |

| 65. | These services are provided by certain organizations which collect and tabulate the marketing information on a regular basis for a number of clients who are the subscribers to these services. | Syndicate services |
|---|---|---|
| 66. | Because preventing data quality problems is typically not an option, data mining focuses on<br>    1. Detection and correction (called _____) of data quality problems<br>    2. Use of _____ that can tolerate poor data quality | Data cleaning<br>Algorithms |
| 67. | Both measurement errors and data collection errors can be either ___ or ___. | Systematic, random |
| 68. | Refers to any problem resulting from the measurement process. A common problem is that the value recorded differs from the true value to some extent. | Measurement error |
| 69. | Refers to errors such as omitting data objects or attribute values, or inappropriately including a data object. | Data collection error |
| 70. | random component of a measurement error; may involve the distortion of a value or addition of false objects | noise |
| 71. | Data errors may be the result of a more deterministic phenomenon, such as a streak in the same place on a set of photographs. Such deterministic distortions of the data are often referred to as ___. | artifact |
| 72. | The closeness of repeated measurements (of the same quantity) to one another. | Precision |
| 73. | Precision is often measured by the _____ of a set of values. | Standard deviation |
| 74. | Bias is measured by taking the difference between the ___ of the set of values and the _____ of the quantity being measured. | Mean, known value |
| 75. | Bias can only be determined for objects whose measured quantity is known by means _____ to the current situation. | External |
| 76. | The closeness of measurements to the true value of the quantity being measured. | Accuracy |
| 77. | These are data objects that have characteristics that are different from most of the other data objects in the data set, or values of an attribute that are unusual with respect to the typical values for that attribute. | Outliers |
| 78. | Strategies for dealing with missing data: | Eliminate data objects or attributes<br>Estimate missing values (in time series)<br>Ignore missing value during analysis |
| 79. | The term ___ is often used to refer to the process of dealing with issues regarding duplicate data. | deduplication |
| 80. | Some data starts to age as soon as it has been collected. This is an issue of: | Timeliness |
| 81. | The available data must contain the information necessary for the application. This ensures the data's ___. | Relevance |
| 82. | Ideally, data sets are accompanied by documentation that describes different aspects of the data. The quality of this documentation can either aid or hinder the subsequent analysis. This ensures the examiner's ___. | Knowledge about the data |
| 83. | These refer to the steps that should be applied to make the data more suitable for data mining. | Data preprocessing |
| 84. | Strategies and techniques of data preprocessing:<br>  • Quantitative attributes are typically grouped by taking a sum or an average.<br>  • An approach for selecting a subset of the data objects to be analyzed.<br>  • These approaches start with a small sample, and then increase the sample size until a sample of sufficient size has been obtained.<br>  • Transform a continuous attribute into a categorical attribute.<br>  • Transform both continuous and discrete attributes into one or more binary attributes.<br>  • Transformation that is applied to all the values of a variable. | |
| 85. | The goal of data preprocessing is to improve the data mining analysis with respect to ___, ___, and ___ | Time, cost, quality |