

Data Science



JOIN OUR NEW BATCH FOR

ONE PAPER MCQs



Class Time

8:00 PM - 10:00 PM



Duration

3 Months



Schedule

Mon-Fri Classes



Monthly Fee

4000/=

+ Interview Preparation

FOR PPSC, FPSC, SPSC, BPSC, KPSC, OTS, NTS and Others

For Registration Contact:



0336-7483185

▶ **YouTube** For More Updates Follow Our YouTube Channel "**GK with Rana Muhammad Aslam**" Current affairs for CSS, PMS, FPSC, PPSC, KPSC, BPSC, SPSC, AJK Board, NTS, OTS, STS, PTS & Commission Exams **To Attempt Quiz of Current Affairs, Detailed Current Affairs & Download PDF** | Visit our website www.pacegkacademy.com Join us on WhatsApp channel **03367483185** / Follow us on Facebook [@pacegkacademy](https://www.facebook.com/pacegkacademy)

Data Science

Q.1: In EDA, what is the term for a graphical representation that displays the distribution of two continuous variables?

- A. Scatter Matrix
- B. Scatter Plot
- C. Box Plot
- D. Violin Plot

Q.2: What is the primary goal of a bar plot (bar chart) in EDA?

- A. To calculate summary statistics
- B. To display relationships between variables
- C. To compare multiple categories
- D. To visualize the distribution of data

Q.3: Which EDA technique is used to visualize the distribution of two continuous variables using a matrix of scatter plots?

- A. Histogram Matrix
- B. Scatter Matrix
- C. Box Plot Matrix
- D. Correlation Matrix

Q.4: In EDA, what is the term for a graphical representation that displays the distribution of a continuous variable along with a box?

- A. Heat Map
- B. Scatter Plot
- C. Box Plot
- D. Violin Plot

Q.5: What is the primary purpose of a swarm plot in EDA?

- A. To display relationships between variables
- B. To visualize the distribution of data
- C. To compare multiple categories
- D. To identify missing values

Q.6: In EDA, what is the primary purpose of a line plot (time series plot)?

- A. To display relationships between variables
- B. To visualize the distribution of data
- C. To identify missing values
- D. To calculate summary statistics

Q.7: What does the term "mode" refer to in the context of data distribution analysis in EDA?

- A. The presence of outliers
- B. The most frequently occurring value
- C. The average value
- D. The spread of data

Q.8: Which EDA technique is used to visualize the distribution of a single categorical variable as a bar chart?

- A. Scatter Plot
- B. Box Plot
- C. Histogram
- D. Bar Chart

Q.9: In EDA, what is the primary purpose of a histogram?

- A. To calculate summary statistics
- B. To display relationships between variables
- C. **To visualize the distribution of data**
- D. To identify missing values

Q.10: In EDA, what is the term for a graphical representation that displays the distribution of a continuous variable along with outliers?

- A. Heat Map
- B. Scatter Plot
- C. Box Plot
- D. **Violin Plot**

Q.11: What is the primary purpose of a density plot in EDA?

- A. **To visualize the distribution of data**
- B. To display relationships between variables
- C. To identify missing values
- D. To calculate summary statistics

Q.12: In EDA, what is the primary purpose of a pie chart?

- A. To identify missing values
- B. To visualize the distribution of data
- C. To display relationships between variables
- D. **To compare multiple categories**

Q.13: Which EDA technique is commonly used to visualize the relationship between two continuous variables through a 2D plot?

- A. Bar Chart
- B. **Scatter Plot**
- C. Box Plot
- D. Histogram

Q.14: What does the term "scatter matrix" refer to in the context of EDA?

- A. A matrix of bar charts
- B. **A collection of scatter plots**
- C. A matrix of correlation coefficients
- D. A matrix of histogram plots

Q.15: Which EDA technique is used to visualize the relationship between two categorical variables in a contingency table format?

- A. Bar Chart
- B. Box Plot
- C. Scatter Plot
- D. **Crosstabulation**

Q.16: What is the primary purpose of a violin plot in EDA?

- A. To identify missing values
- B. **To visualize the distribution of data**
- C. To display relationships between variables
- D. To compare multiple categories

Q.17: What is the primary goal of a pair plot (scatter plot matrix) in EDA?

- A. To visualize the distribution of data
- B. **To display relationships between variables**
- C. To identify missing values
- D. To calculate summary statistics

Q.18: Which EDA technique is used to visualize the distribution of a single categorical variable?

- A. **Bar Chart**
- B. Scatter Plot
- C. Box Plot
- D. Histogram

Q.19: In EDA, what is the primary purpose of a heat map?

- A. To calculate summary statistics
- B. To visualize the distribution of data
- C. **To display relationships between variables**
- D. To identify missing values

Q.20: What does the term "kurtosis" refer to in the context of data distribution analysis in EDA?

- A. The presence of outliers
- B. The symmetry of a data distribution
- C. The spread of a data distribution
- D. **The shape of a data distribution**

Q.21: In EDA, what does the term "box" represent in a box plot?

- A. **Median**
- B. Mode
- C. Range
- D. Standard Deviation

Q.22: Which EDA technique is used to visualize the distribution of a single continuous variable?

- A. Bar Chart
- B. Box Plot
- C. **Histogram**
- D. Scatter Plot

Q.23: What is the primary purpose of a bar chart in EDA?

- A. To identify missing values
- B. To visualize the distribution of data
- C. To display relationships between variables
- D. **To compare multiple categories**

Q.24: What is the primary goal of a density plot in EDA?

- A. To calculate summary statistics
- B. **To visualize the distribution of data**
- C. To display relationships between variables
- D. To identify missing values

Q.25: Which EDA technique is used to display the distribution of a single variable, especially in cases of categorical data?

- A. Bar Chart
- B. Box Plot
- C. Scatter Plot
- D. **Histogram**

Q.26: What does the term "skewness" refer to in the context of data distribution analysis in EDA?

- A. **The symmetry of a data distribution**
- B. The kurtosis of a data distribution
- C. The spread of a data distribution
- D. The presence of outliers

Q.27: Which EDA technique is used to visualize the relationship between a categorical variable and a continuous variable?

- A. Bar Chart
- B. **Box Plot**
- C. Scatter Plot
- D. Histogram

Q.28: In EDA, what is the purpose of a Q-Q plot (Quantile-Quantile plot)?

- A. To display time series data
- B. To visualize the distribution of data
- C. To compare two different datasets
- D. **To check for normality in data**

Q.29: What type of data visualization is commonly used in EDA to represent the distribution of a categorical variable?

- A. **Bar Chart**
- B. Scatter Plot
- C. Box Plot
- D. Histogram

Q.30: What does the term "outlier" refer to in the context of EDA?

- A. A data point that is missing
- B. A point that falls within the data range
- C. A data point that is part of the main cluster
- D. **A data point that falls far from the rest**

Q.31: Which EDA technique is used to visualize the relationship between two continuous variables with a line connecting data points?

- A. Box Plot
- B. **Scatter Plot**
- C. Histogram
- D. Bar Chart

Q.32: What is the purpose of a correlation matrix in EDA?

- A. To calculate summary statistics
- B. To visualize the distribution of data
- C. **To display relationships between variables**
- D. To identify missing values

Q.33: Which EDA technique is used to identify and visualize outliers in a dataset?

- A. Histogram
- B. Bar Chart
- C. **Box Plot**
- D. Scatter Plot

Q.34: In EDA, which measure of central tendency is typically represented by the height of a box in a box plot?

- A. Variance
- B. Mean
- C. **median**
- D. Mode

Q.35: What is the primary purpose of a scatter plot in EDA?

- A. To represent time series data
- B. To show the distribution of a single variable
- C. **To visualize relationships between variables**
- D. To display categorical data

Q.36: In EDA, what is the term for a graphical representation that displays the distribution of a continuous variable?

- A. Scatter Plot
- B. Pie Chart
- C. **Histogram**
- D. Bar Chart

Q.37: Which of the following is NOT a common EDA technique for visualizing data distributions?

- A. Histogram
- B. Box Plot
- C. Scatter Plot
- D. **Bar Chart**

Q.38: What is the primary purpose of Exploratory Data Analysis (EDA) in Data Science?

- A. To simplify complex data
- B. To make data more complicated
- C. To visualize data
- D. **To discover patterns and insights in data**

Q.39: Which of the following function is used for searching text strings by means of regular expression?

- A. gepexpr
- B. grep
- C. **grep**
- D. all of the mentioned

Q.40: Which of the following package is used for tidy data?

- A. souryr
- B. **tidyr**
- C. NumPy
- D. all of the mentioned

Q.41: Which of the following function is used for loading flat files?

- A. **read.table**
- B. read.data
- C. read.sheet
- D. none of the mentioned

Q.42: Which of the following package is used for reading HTML and XML data?

- A. httpx
- B. **httr**
- C. http
- D. all of the mentioned

Q.43: Which of the following request can be issued from http package?

- A. DELETE
- B. GET
- C. PUT
- D. **All of the mentioned**

Q.44: In data preprocessing, what is the term for the process of handling missing data by filling in the missing values with estimates?

- A. Data Encoding
- B. **Data Imputation**
- C. Data Normalization
- D. Data Transformation

Q.45: Which data collection method involves selecting a subset of individuals or elements from a larger population for study?

- A. Experiments
- B. Observational Studies
- C. Surveys
- D. **Sampling**

Q.46: What is the purpose of feature engineering in data preprocessing?

- A. To visualize data
- B. To make data more complex
- C. **To extract meaningful features**
- D. To perform data cleaning

Q.47: In data preprocessing, what is the term for the process of converting text data into numerical values?

- A. Data Transformation
- B. **Data Encoding**
- C. Data Tokenization
- D. Data Parsing

Q.48: Which of the following is NOT a common data preprocessing technique?

- A. **Data Visualization**
- B. Data Cleaning
- C. Data Encoding
- D. Data Imputation

Q.49: What is the term for the process of organizing and storing data in a structured format that enables efficient retrieval?

- A. **Data Warehousing**
- B. Data Integration
- C. Data Transformation
- D. Data Aggregation

Q.50: What is the primary goal of data preprocessing in the context of machine learning?

- A. To reduce the size of the dataset
- B. To increase data complexity
- C. **To prepare data for analysis**
- D. To create visualizations of the data

Q.51: Which data preprocessing step involves checking for and handling duplicate records in a dataset?

- A. Data Encoding
- B. **Data Deduplication**
- C. Data Aggregation
- D. Data Scaling

Q.52: What is the term for data that is collected from a single source at a specific point in time?

- A. Panel Data
- B. **Cross-Sectional Data**
- C. Time Series Data
- D. Longitudinal Data

Q.53: What is the purpose of data sampling in data collection?

- A. To remove outliers
- B. To analyze the entire dataset
- C. To collect data from various sources
- D. **To select a representative subset**

Q.54: Which data collection method involves manipulating one or more variables to observe their effects on other variables?

- A. Interviews
- B. Observational Studies
- C. **Experiments**
- D. Surveys

Q.55: In data preprocessing, what is the process of converting data values into a standard format to remove noise and variations?

- A. Data Imputation
- B. Data Scaling
- C. Data Transformation
- D. **Data Normalization**

Q.56: What does the acronym "ETL" stand for in the context of data preprocessing and integration?

- A. Estimate, Transform, Load
- B. **Extract, Transform, Load**
- C. Encrypt, Transmit, Log
- D. Evaluate, Test, Learn

Q.57: Which data collection method involves the collection of data at regular intervals over a period of time?

- A. **Time Series Data Collection**
- B. Observational Studies
- C. Experiments
- D. Surveys

Q.58: What is the term for the process of reducing the dimensionality of data while retaining its most important features?

- A. Data Normalization
- B. Data Imputation
- C. Data Scaling
- D. **Data Reduction**

Q.59: What is the term for the process of handling missing data in a dataset by estimating values based on available data?

- A. Data Encoding
- B. **Data Imputation**
- C. Data Normalization
- D. Data Transformation

Q.60: In data preprocessing, what is the term for the process of converting categorical data into binary format?

- A. Data Aggregation
- B. **Data Encoding**
- C. Data Normalization
- D. Data Imputation

Q.61: What is the primary goal of data preprocessing in the context of machine learning?

- A. To create visualizations of the data Discover more Ask question feature Discuss
- B. To make the data more complex
- C. **To prepare the data for analysis**
- D. To reduce the size of the dataset

Q.62: Which data preprocessing step involves checking for and handling duplicate records in a dataset?

- A. Data Encoding
- B. **Data Deduplication**
- C. Data Aggregation
- D. Data Scaling

Q.63: What is the term for the process of selecting a representative subset of a larger dataset for analysis?

- A. Data Cleaning
- B. **Data Sampling**
- C. Data Integration
- D. Data Transformation

Q.64: In data collection, what is the term for data that has not been processed or analyzed and is in its original form?

- A. Structured Data
- B. **Raw Data**
- C. Processed Data
- D. Analytical Data

Q.65: Which of the following is NOT a common data preprocessing technique?

- A. **Data Aggregation**
- B. Data Encoding
- C. Data Normalization
- D. Data Imputation

Q.66: What is the purpose of feature engineering in data preprocessing?

- A. To visualize data
- B. To make data more complex
- C. **To extract meaningful features**
- D. To perform data cleaning

Q.67: Which data collection method involves gathering data by directly interacting with subjects or participants?

- A. Sampling
- B. Surveys
- C. Observations
- D. **Experiments**

Q.68: What is the term for the process of converting data into a structured format suitable for analysis and storage?

- A. Data Cleaning
- B. Data Integration
- C. **Data Transformation**
- D. Data Wrangling

Q.69: Which of the following is a common technique for handling outliers in a dataset?

- A. **Data Scaling**
- B. Data Imputation
- C. Data Transformation
- D. Data Visualization

Q.70: What type of data collection method involves manipulating one or more variables to observe their effects on other variables?

- A. Interviews
- B. Observations
- C. Surveys
- D. **Experiments**

Q.71: In data preprocessing, what is the purpose of data scaling or normalization?

- A. To create new features
- B. To increase data complexity
- C. **To standardize data**
- D. To remove outliers

Q.72: Which data collection method involves selecting a sample from a larger population and collecting data from the sample only?

- A. **Sampling**
- B. Experiments
- C. Observational Studies
- D. Surveys

Q.73: What is the process of converting text data into a numerical format so that it can be used in machine learning algorithms called?

- A. Data Transformation
- B. Data Encoding
- C. **Data Tokenization**
- D. Data Parsing

Q.74: Which of the following is a common technique for handling imbalanced data in classification problems?

- A. Data Imputation
- B. **Data Augmentation**
- C. Data Scaling
- D. Data Encoding

Q.75: In data preprocessing, what is the term for the identification and removal of duplicate or redundant data?

- A. Data Imputation
- B. **Data Deduplication**
- C. Data Aggregation
- D. Data Normalization

Q.76: Which of the following is NOT a common step in data preprocessing?

- A. Data Integration
- B. Data Cleaning
- C. **Data Visualization**
- D. Data Transformation

Q.77: What is the process of transforming categorical data into numerical values called?

- A. Data Standardization
- B. **Data Encoding**
- C. Data Imputation
- D. Data Transformation

Q.78: Which type of data collection method involves asking questions to gather information from respondents?

- A. Observations
- B. Experiments
- C. **Surveys**
- D. Interviews

Q.79: Which of the following is a common technique used to preprocess text data, involving the removal of common words?

- A. **Data Stopword Removal**
- B. Data Stemming
- C. Data Tokenization
- D. Data Parsing

Q.80: What is the term for the process of reducing the size of a dataset while retaining its essential characteristics?

- A. Data Cleaning
- B. Data Wrangling
- C. **Data Reduction**
- D. Data Sampling

Q.81: Which method of data collection involves systematically observing and recording data without interfering with the subjects?

- A. **Observational Studies**
- B. Surveys
- C. Interviews
- D. Experiments

Q.82: What does the acronym "API" stand for in the context of data collection and retrieval?

- A. Analytical Program Interaction
- B. **Application Programming Interface**
- C. Advanced Protocol Interface
- D. Automated Process Integration

Q.83: Which of the following is a technique used to handle missing data by filling in the missing values with estimated data?

- A. **Data Imputation**
- B. Data Normalization
- C. Data Transformation
- D. Data Encoding

Q.84: What is the process of cleaning and formatting raw data into a usable format called?

- A. Data Integration
- B. Data Visualization
- C. **Data Cleaning**
- D. Data Extraction

Q.85: Which of the following is a type of structured data that is organized in rows and columns, similar to a spreadsheet?

- A. XML
- B. **CSV**
- C. JSON
- D. Unstructured Text Data

Q.86: In data collection, what is the term for the process of recording data in its original form without any changes?

- A. Data Transformation
- B. Data Integration
- C. Data Aggregation
- D. **Raw Data Collection**

Q.87: Which of the following is NOT a common source of data for data collection in Data Science?

- A. Customer Surveys
- B. Social Media
- C. Sensor Data
- D. **Print Media**

Q.88: What is the process of gathering data from various sources and storing it for analysis known as?

- A. **Data Collection**
- B. Data Wrangling
- C. Data Visualization
- D. Data Aggregation

Q.89: Point out the correct statement?

- A. If equations and parameter are not, they may be inferred with data analysis
- B. If equations are known but the parameters are not, they may be inferred with data analysis
- C. If equations are not known but the parameters are, they may be inferred with data Oanalysis
- D. None of the mentioned

Q.90: Which of the following uses relatively small amount of data to estimate about bigger population?

- A. Exploratory
- B. Inferential
- C. Causal
- D. None of the mentioned

Q.91: Which of the following analytical capabilities are provided by information management company?

- A. Content Management
- B. Stream Computing
- C. Information Integration
- D. All of the mentioned

Q.92: Which of the following is the common goal of statistical modelling?

- A. Summarizing
- B. Inference
- C. Subsetting
- D. None of the mentioned

Q.93: Which of the following characteristic of big data is relatively more concerned to data science?

- A. Volume
- B. Velocity
- C. Variety
- D. None of the mentioned

Q.94: Which of the following is the top most important thing in data science?

- A. question
- B. answer
- C. data
- D. None of the mentioned

Q.95: Point out the correct statement?

- A. Preprocessed data is original source of data
- B. Raw data is original source of data
- C. Raw data is the data obtained after processing steps
- D. None of the mentioned

Q.96: Which of the following is a good way of performing experiments in data science?

- A. Generalize to the problem
- B. Measure variability
- C. Have Replication
- D. All of the mentioned

Q.97: Which of the following technique comes under practical machine learning?

- A. **Boosting**
- B. Bagging
- C. Forecasting
- D. None of the mentioned

Q.98: Point out the wrong statement?

- A. Complication approached exist for inferring causation
- B. **Randomized studies are not used to identify causation**
- C. Causal relationships may not apply to every individual
- D. All of the mentioned

Q.99: Which of the following relationship are usually identified as average effects?

- A. **Causal**
- B. Descriptive
- C. Predictive
- D. None of the mentioned

Q.100: Which of the following command allows you to update the repository?

- A. pop
- B. **push**
- C. update
- D. one of the mentioned

Q.101: Which of the following is not a step in data analysis?

- A. Clean the data
- B. Obtain the data
- C. EDA
- D. **None of the mentioned**

Q.102: Which of the following is characteristic of Raw Data?

- A. **Original version of data**
- B. Data is ready for analysis
- C. Easy to use for data analysis
- D. None of the mentioned

Q.103: Point out the correct statement?

- A. **CLI can help you to organize files and folders**
- B. You don't need GitHub to use Git
- C. Navigation of directory is possible using CLI
- D. None of the mentioned

Q.104: Which of the following is characteristic of Processed Data?

- A. **All steps should be noted**
- B. Data is not ready for analysis
- C. Hard to use for data analysis
- D. None of the mentioned

Q.105: Which of the following is a revision control system?

- A. NumPy
- B. **Git**
- C. Slidify
- D. None of the mentioned

Q.106: Point out the correct statement____?

- A. **CLI can help you to organize files and folders**
- B. CLI can help you to organize messages
- C. Navigation of directory is possible using CLI
- D. None of the mentioned

Q.107: Which of the following is the correct way of creating GitHub repository in to well labelled commits?

- A. Pop another user's repository
- B. **Fork another user's repository**
- C. Zip another user's repository
- D. None of the mentioned

Q.108: Which of the following command help us to give message description?

- A. git command -d
- B. **git command -m**
- C. git command -message
- D. none of the mentioned

Q.109: Which of the following is one of the key data science skills?

- A. Machine Learning
- B. Statistics
- C. Data Visualization
- D. **All of the mentioned**

Q.110: Which of the following systems record changes to a file over time?

- A. **Version Control**
- B. Record Control
- C. Forecast Control
- D. None of the mentioned

Q.111: Point out the correct statement_____?

- A. Inferential models are useful for discovering new connection
- B. Exploratory analyses are not usually the final way
- C. **Inference involves estimating uncertainty**
- D. All of the mentioned

Q.112: Which of the following allows you to find the relationship you didn't about?

- A. Causal
- B. Inferential
- C. **Exploratory**
- D. None of the mentioned

Q.113: Which of the following analysis helps out to find the effect of variable change?

- A. Exploratory
- B. Inferential
- C. **Causal**
- D. None of the mentioned

Q.114: Which of the following step is performed by data scientist after acquiring the data?

- A. Data Integration
- B. **Data Cleansing**
- C. Data Replication
- D. All of the mentioned

Q.115: Which of the following is not a CLI command?

- A. rm
- B. **delete**
- C. clear
- D. none of the mentioned

Q.116: Point out the wrong statement_____?

- A. The data growth and social media explosion have changed how we look at the data
- B. The big volume indeed represents Big Data
- C. **Big Data is just about lots of data**
- D. All of the mentioned Discover more Answer

Q.117: Which of the following uses data on some object to predict values for other object?

- A. Exploratory
- B. Inferential
- C. **Predictive**
- D. None of the mentioned

Q.118: Point out the wrong statement.

- A. Subsetting can be used to select and exclude variables and observations
- B. Merging concerns combining datasets on the same observations to produce a result with more variables
- C. **Data visualization is the organization of information according to preset specifications**
- D. All of the mentioned

Q.119: Which of the following is a key characteristic of a hacker?

- A. Not Willing to find answers on their own
- B. Afraid to say they don't know the answer
- C. **Willing to find answers on their own**
- D. All of the mentioned

Q.120: Which of the following approach should be used to ask Data Analysis question?

- A. **Find out the question which is to be answered**
- B. Find only one solution for particular problem
- C. Find out answer from dataset without asking question
- D. None of the mentioned

Q.121: Which of the following technique is also referred to as Bagging?

- A. Bootstrap subsetting
- B. Bootstrap predicting
- C. **Bootstrap aggregating**
- D. All of the mentioned

Q.122: Point out the correct statement?

- A. **Descriptions can be generalized without statistical modelling**
- B. Descriptive analysis is first kind of data analysis performed
- C. Description and Interpretation are same in descriptive analysis
- D. None of the mentioned

Q.123: Which of the following approach should be used if you can't fix the variable?

- A. non stratify it
- B. randomize it
- C. generalize it
- D. none of the mentioned

Q.124: Which of the following focuses on the discovery of (previously) unknown properties on the data?

- A. Big Data
- B. Data mining
- C. Data wrangling
- D. Machine Learning

Q.125: Which of the following analysis are incredibly hard to infer?

- A. Exploratory
- B. Inferential
- C. Causal
- D. Mechanistic

Q.126: Point out the wrong statement?

- A. Exploratory models are useful for discovering new connection
- B. Exploratory analyses are usually the final way
- C. Exploratory analysis alone should not be used for predicting
- D. All of the mentioned

Q.127: How many principles of analytical graphs exist?

- A. 3
- B. 4
- C. 6
- D. None of the mentioned

Q.128: Point out the correct statement_____?

- A. Correlation does imply causation
- B. Descriptive analysis can be more useful for defining future studies
- C. Inference is commonly the goal of statistical model
- D. None of the mentioned

Q.129: Which of the following web hosting service use Git control system?

- A. Open Hash
- B. GitHub
- C. Git Bash
- D. None of the mentioned

Q.130: Which of the following CLI command can also be used to rename files?

- A. mv
- B. rm
- C. rm -r
- D. none of the mentioned

Q.131: hich of the following command updates tracking for files that are modified?

- A. git add -u
- B. git add.
- C. git add -A
- D. none of the mentioned

Q.132: Which of the following model is usually a gold standard for data analysis?

- A. Descriptive
- B. Inferential
- C. Causal
- D. All of the mentioned

Q.133: Point out the correct statement?

- A. Least square problems falls in to three categories
- B. **Least square is an estimation tool**
- C. Compound least square is one of the category of least square
- D. None of the mentioned

Q.134: Which of the following is performed by Data Scientist?

- A. Create reproducible code
- B. Define the question
- C. Challenge results
- D. **All of the mentioned**

Q.135: Which of the following command is used to give a message description?

- A. git command -d
- B. **git command -m**
- C. git command -message
- D. none of the mentioned

Q.136: Point out the wrong statement?

- A. Compound linear regression is not equipped to handle more than one predictor
- B. **Simple linear regression is equipped to handle more than one predictor**
- C. Linear regression consists of finding the best-fitting straight line through the points
- D. All of the mentioned

Q.137: Which of the following command line environment is used for interacting with Git?

- A. **Git Bash**
- B. GitHub
- C. Git Boot
- D. All of the mentioned

Q.138: Which of the following statement would create branch named as 'example'?

- A. git checkout -c example
- B. **git checkout -b example**
- C. git check -b example
- D. none of the mentioned

Q.139: Point out the wrong statement?

- A. GitHub allows you to share repositories with others
- B. **You need GitHub to use Git**
- C. GitHub allows you to access others repositories
- D. All of the mentioned

Q.140: Which of the following command allows you to change directory to one level above your parent directory?

- A. cd.
- B. cd
- C. **cd..**
- D. none of the mentioned

Q.141: Point out the wrong statement?

- A. There is one and only flag for every command in CLI
- B. Command is the CLI command which does a specific task
- C. Flags are the options given to command for activating particular behaviour
- D. All of the mentioned

Q.142: Which of the following data mining technique is used to uncover patterns in data?

- A. Data booting
- B. Data bagging
- C. Data merging
- D. Data Dredging

Q.143: Which of the following analysis is usually modeled by deterministic set of equations?

- A. Causal
- B. Predictive
- C. Mechanistic
- D. All of the mentioned

Q.144: Which of the following command is used to squash the commits?

- A. squash
- B. rebase
- C. Boot
- D. all of the mentioned

Q.145: Which of the following is NOT typically considered a part of the Data Science process?

- A. Data Visualization
- B. Data Collection
- C. Data Cleaning
- D. Software Development

Q.146: What is the primary objective of exploratory data analysis (EDA) in Data Science?

- A. To find patterns
- B. To make predictions
- C. To transform data
- D. To summarize data

Q.147: What is the primary role of a Data Scientist in a business context?

- A. Developing marketing campaigns
- B. Conducting market research
- C. Leveraging data for insights
- D. Extracting data from databases

Q.148: Which of the following is a common method for dealing with missing data in a dataset?

- A. Imputing missing values
- B. Dropping missing values
- C. Normalizing the data
- D. Creating new features

Q.149: Which step in the Data Science process involves visualizing and interpreting the results of data analysis?

- A. Data Cleaning
- B. Data Collection
- C. Data Visualization
- D. Model Building

Q.150: What does the acronym "SQL" stand for in the context of Data Science?

- A. Statistical Query Language
- B. Structured Query Language
- C. Simplified Query Language
- D. Secure Query Language

Q.151: Which of the following is a common algorithm used for classification in supervised learning?

- A. **Decision Tree**
- B. K-Means Clustering
- C. Principal Component Analysis (PCA)
- D. Naive Bayes

Q.152: What is the process of splitting a dataset into a training set and a test set used for machine learning called?

- A. **Data Sampling**
- B. Data Partitioning
- C. Data Shuffling
- D. Data Splitting

Q.153: In Data Science, what is the purpose of data wrangling?

- A. **To transform raw data**
- B. To create complex models
- C. To remove outliers
- D. To visualize data

Q.154: Which type of data is represented by categories or labels and cannot be measured numerically?

- A. **Categorical Data**
- B. Numerical Data
- C. Ordinal Data
- D. Continuous Data

Q.155: What is the primary objective of data exploration in Data Science?

- A. **To find hidden patterns**
- B. To build predictive models
- C. To summarize data
- D. To collect data

Q.156: In Data Science, what is the term for a data point that is missing a value for one or more features?

- A. Anomaly
- B. Outlier
- C. **Null Value**
- D. Feature

Q.157: Which statistical measure represents the spread or dispersion of data values in a dataset?

- A. Mean
- B. Median
- C. **Standard Deviation**
- D. Mode

Q.158: Which of the following best describes the purpose of data sampling in Data Science?

- A. **To select a representative subset**
- B. To analyze the entire dataset
- C. To calculate data statistics
- D. To visualize data

Q.159: What is the term for the process of removing or reducing noise and inconsistencies from data?

- A. Data Transformation
- B. Data Integration
- C. Data Aggregation
- D. **Data Cleansing**

Q.160: In Data Science, what is the term for the process of reducing the dimensionality of a dataset while preserving information?

- A. Data Cleaning
- B. **Dimensionality Reduction**
- C. Feature Engineering
- D. Data Transformation

Q.161: Which of the following is NOT a common data visualization tool or library used in Data Science?

- A. Seaborn
- B. Tableau
- C. Power BI
- D. **Excel**

Q.162: What is the process of converting categorical variables into numerical values for machine learning called?

- A. **Data Encoding**
- B. Feature Extraction
- C. Data Standardization
- D. Label Encoding

Q.163: Which of the following is NOT a key role in a typical Data Science team?

- A. Data Analyst
- B. Data Engineer
- C. Data Scientist
- D. **Database Administrator**

Q.164: What type of data analysis focuses on understanding the relationships and patterns within a dataset?

- A. Predictive Analysis
- B. **Descriptive Analysis**
- C. Inferential Analysis
- D. Diagnostic Analysis

Q.165: Which of the following is a common technique used to handle imbalanced datasets in classification problems?

- A. Downsampling
- B. **Upsampling**
- C. Feature Engineering
- D. Data Wrangling

Q.166: Which step in the Data Science process involves assessing the quality of collected data?

- A. Data Cleaning
- B. Data Collection
- C. **Data Validation**
- D. Data Visualization

Q.167: In data Science, what is the primary purpose of data visualization?

- A. To complicate data
- B. To make data more confusing
- C. To obscure patterns in data
- D. **To help communicate insights from data**

Q.168: What is the term for the process of cleaning and organizing data into a structured format suitable for analysis?

- A. Data Extraction
- B. **Data Transformation**
- C. Data Visualization
- D. Data Aggregation

Q.169: Which library in Python is commonly used for data manipulation and analysis in Data Science?

- A. Matplotlib
- B. **Pandas**
- C. Scikit-Learn
- D. TensorFlow

Q.170: Which of the following is an example of an unsupervised learning algorithm used in clustering data?

- A. Decision Trees
- B. Linear Regression
- C. **K-Means Clustering**
- D. Logistic Regression

Q.171: What is the term for a machine learning algorithm that learns from historical data to make predictions about the future?

- A. Clustering
- B. Regression
- C. Classification
- D. **Supervised Learning**

Q.172: In Data Science, what is the purpose of feature engineering?

- A. To model data features
- B. **To extract features from data**
- C. To visualize data features
- D. To clean data features

Q.173: Which statistical measure represents the middle value of a dataset when it is sorted in ascending order?

- A. Mean
- B. **Median**
- C. Standard Deviation
- D. Mode

Q.174: What is the process of extracting patterns and information from data called?

- A. Data Visualization
- B. Data Wrangling
- C. Data Engineering
- D. **Data Mining**

Q.175: Which step in the Data Science process involves building and training predictive models?

- A. Data Visualization
- B. Data Collection
- C. Data Cleaning
- D. **Model Building**

Q.176: What is the primary purpose of data visualization in Data Science?

- A. To store data
- B. To make data more complicated
- C. **To simplify complex data**
- D. To increase data complexity

Q.177: Which of the following is a technique used to handle missing data in a dataset?

- A. **Data Imputation**
- B. Data Augmentation
- C. Data Transformation
- D. Data Normalization

Q.178: Which of the following is NOT a common data format used in Data Science projects?

- A. XML
- B. JSON
- C. CSV
- D. **HTML**

Q.179: What is the primary focus of Data Science?

- A. Data Visualization
- B. Data Cleaning
- C. **Data Analysis**
- D. Data Storage

Q.180: In Data Science, what is the term for a dataset that contains both input features and output labels?

- A. **Training Data**
- B. Test Data
- C. Unlabeled Data
- D. Validation Data

Q.181: Which step in the Data Science process involves selecting the appropriate model and algorithm for analysis?

- A. Data Visualization
- B. Data Cleaning
- C. Data Collection
- D. **Model Building**

Q.182: Which of the following is NOT a key skill required for a Data Scientist?

- A. Storytelling
- B. Data Visualization
- C. **Database Administration**
- D. Machine Learning

Q.183: What does the acronym "EDA" stand for in Data Science?

- A. Effective Data Algorithms
- B. **Exploratory Data Analysis**
- C. Extracted Data Aggregation
- D. Efficient Data Assessment

Q.184: Which technology is often used to process and analyze large-scale data sets in Data Science?

- A. SQL
- B. **Hadoop**
- C. Python
- D. HTML

Q.185: Which of the following is NOT a type of machine learning algorithm commonly used in Data Science?

- A. K-Means Clustering
- B. Linear Regression
- C. Decision Trees
- D. **Object-Oriented Programming**

Q.186: What is the term for a data point that falls far from the rest of the data in a dataset?

- A. Median
- B. **Outlier**
- C. Mean
- D. Variance

Q.187: Which step in the Data Science process involves understanding and preparing the data for analysis?

- A. Data Visualization
- B. Data Collection
- C. **Data Cleaning**
- D. Model Building

Q.188: Which programming language is commonly used for Data Science tasks?

- A. **Python**
- B. Java
- C. C++
- D. Java Script

Q.189: What is the primary goal of Data Science?

- A. Data Cleaning
- B. Data Visualization
- C. **Predictive Analytics**
- D. Extracting Data from APIs

Data Science

(Answer Key)

1. A	2. C	3. B	4. D	5. A	6. B	7. B	8. D	9. C	10. D
11. A	12. D	13. B	14. B	15. D	16. B	17. B	18. A	19. C	20. D
21. A	22. C	23. D	24. B	25. D	26. A	27. B	28. D	29. A	30. D
31. B	32. C	33. C	34. C	35. C	36. C	37. D	38. D	39. C	40. B
41. A	42. B	43. D	44. B	45. D	46. C	47. B	48. A	49. A	50. C
51. B	52. B	53. D	54. C	55. D	56. B	57. A	58. D	59. B	60. B
61. C	62. B	63. B	64. B	65. A	66. C	67. D	68. C	69. A	70. D
71. C	72. A	73. C	74. B	75. B	76. C	77. B	78. C	79. A	80. C
81. A	82. B	83. A	84. C	85. B	86. D	87. D	88. A	89. B	90. B
91. D	92. B	93. C	94. A	95. B	96. D	97. A	98. B	99. A	100. B
101. D	102. A	103. A	104. A	105. B	106. A	107. B	108. B	109. D	110. A
111. C	112. C	113. C	114. B	115. B	116. C	117. C	118. C	119. C	120. A
121. C	122. A	123. B	124. B	125. D	126. B	127. C	128. A	129. B	130. A
131. A	132. C	133. B	134. D	135. B	136. B	137. A	138. B	139. B	140. C
141. A	142. D	143. C	144. B	145. D	146. A	147. C	148. A	149. C	150. B
151. A	152. A	153. A	154. A	155. A	156. C	157. C	158. A	159. D	160. B
161. D	162. A	163. D	164. B	165. B	166. C	167. D	168. B	169. B	170. C
171. D	172. B	173. B	174. D	175. D	176. C	177. A	178. D	179. C	180. A
181. D	182. C	183. B	184. B	185. D	186. B	187. C	188. A	189. C	