**1. What can data do in data science?**

* A. Predict future events
* B. Describe the current state
* C. Detect anomalies
* D. All of the above

**2. Why has data science become so popular recently?**

* A. Increased availability of data
* B. Advanced data collection tools
* C. Higher computational power
* D. All of the above

**3. What is the first stage of the data science workflow?**

* A. Experimentation and Prediction
* B. Data Collection and Storage
* C. Data Visualization
* D. Data Preparation

**4. Which of the following is not a task in building a customer service chatbot?**

* A. Predict possible responses
* B. Create bar charts
* C. Classify customer questions
* D. Build SQL databases

**5. Which application is an example of IoT in data science?**

* A. Analyzing social media data
* B. Smart home energy management
* C. Predicting customer churn
* D. Natural language processing

**6. What is essential to machine learning in fraud detection?**

* A. Identifying customer preferences
* B. Using labeled training data
* C. Collecting feedback
* D. Replacing manual intervention

**7. Which data is considered geospatial data?**

* A. Images of a city skyline
* B. Coordinates of hiking trails
* C. Text reviews of a restaurant
* D. Financial transaction records

**8. In the case of a smart watch, what sensor provides motion data?**

* A. GPS
* B. Thermometer
* C. Accelerometer
* D. Barometer

**9. Which characteristic defines qualitative data?**

* A. Can be measured in numbers
* B. Expressed conceptually or descriptively
* C. Requires a database schema
* D. Visualized through time-series graphs

**10. Which is an example of public open data?**

* A. Social media posts
* B. Census data
* C. Customer purchase history
* D. Company web analytics

**11. Which process is part of data preparation?**

* A. Running statistical tests
* B. Cleaning duplicate or missing values
* C. Designing machine learning algorithms
* D. Conducting user surveys

**12. What is one key purpose of exploratory data analysis (EDA)?**

* A. Designing APIs
* B. Building pipelines
* C. Identifying patterns in data
* D. Training machine learning models

**13. In data visualization, why is it important to use labels?**

* A. To make charts more colorful
* B. To enhance interpretability
* C. To increase statistical accuracy
* D. To reduce processing time

**14. Which of the following defines time-series data?**

* A. Data categorized by location
* B. Data tracked across time intervals
* C. Data split into multiple clusters
* D. Data collected via public surveys

**15. What does a confidence interval indicate in predictions?**

* A. The exact value of a prediction
* B. The degree of certainty around predictions
* C. The accuracy of a dataset
* D. The total error in predictions

**16. Which is an example of supervised machine learning?**

* A. Grouping similar customers
* B. Detecting anomalies in user behavior
* C. Predicting customer churn based on labels
* D. Dividing products into clusters

**17. What is the key difference between features and labels?**

* A. Features are predicted outcomes
* B. Labels define outcomes, features help predict them
* C. Features come after labels
* D. Labels are descriptive data only

**18. What is clustering in unsupervised learning used for?**

* A. Dividing data into similar groups
* B. Predicting numerical values
* C. Analyzing outliers in datasets
* D. Building dashboards

**19. What type of data source is an API typically used for?**

* A. Unstructured data only
* B. Real-time or public data
* C. Company confidential data
* D. Open surveys

**20. What is the transform phase in ETL pipelines?**

* A. Creating visualizations
* B. Changing data structure to fit databases
* C. Cleaning messy data
* D. Automating storage

**21. What is a key advantage of a data pipeline?**

* A. Eliminates the need for visualization tools
* B. Ensures faster model predictions
* C. Automates data movement and storage
* D. Avoids data preparation altogether

**22. Why is data preparation critical in data science?**

* A. To make data suitable for analysis
* B. To automate machine learning tasks
* C. To build deep learning models
* D. To integrate social media data

**23. Which is an example of quantitative data?**

* A. Customer satisfaction comments
* B. Number of products sold daily
* C. User gender categories
* D. Social network relationships

**24. What is a feature of a document database?**

* A. Stores tabular data
* B. Handles structured relational data
* C. Stores unstructured data like emails
* D. Organizes data into geospatial clusters

**25. What is an example of deep learning?**

* A. Predicting store revenue trends
* B. Recognizing images in photos
* C. Analyzing survey feedback
* D. Summarizing financial spreadsheets

**26. Which of these roles focuses on data exploration and visualization?**

* A. Data Engineer
* B. Data Analyst
* C. Data Scientist
* D. Machine Learning Scientist

**27. What is the primary role of a data engineer?**

* A. Create visualizations
* B. Train machine learning models
* C. Manage data pipelines and storage
* D. Interpret statistical trends

**28. Which machine learning type predicts labels?**

* A. Supervised
* B. Unsupervised
* C. Clustering
* D. Natural language processing

**29. What is the purpose of A/B testing?**

* A. Creating predictive models
* B. Comparing two scenarios to improve outcomes
* C. Grouping similar users into clusters
* D. Analyzing time-series data

**30. In what scenario might clustering be used?**

* A. Predicting sales trends
* B. Dividing customers into behavior groups
* C. Identifying fraudulent transactions
* D. Summarizing survey data

**1. Which Python library is primarily used for data visualization in this lecture?**

* A. Pandas
* B. Matplotlib
* C. NumPy
* D. Seaborn

**2. What type of plot is typically used to visualize data over a time scale?**

* A. Histogram
* B. Scatter Plot
* C. Line Plot
* D. Bar Chart

**3. What argument is used in Matplotlib to adjust the number of bins in a histogram?**

* A. bins
* B. size
* C. count
* D. range

**4. Which plot is best suited to observe the relationship between two variables?**

* A. Line Plot
* B. Histogram
* C. Scatter Plot
* D. Pie Chart

**5. What scale is suggested for GDP per capita on a scatter plot to improve readability?**

* A. Linear
* B. Exponential
* C. Logarithmic
* D. Polynomial

**6. What should you do to make a scatter plot clearer and easier to interpret?**

* A. Add labels and titles
* B. Increase marker size
* C. Use color coding for categories
* D. All of the above

**7. In Pandas, which function is used to load a CSV file into a DataFrame?**

* A. pd.read\_csv()
* B. pd.load\_file()
* C. pd.import\_csv()
* D. pd.open\_csv()

**8. Which command displays the first 5 rows of a Pandas DataFrame?**

* A. .head()
* B. .tail()
* C. .sample()
* D. .describe()

**9. How do you filter rows in Pandas based on a condition?**

* A. Using an if statement
* B. Using comparison operators inside brackets
* C. Using the .filter() method
* D. Using the .drop() method

**10. Which logical operator is used in Pandas for combining conditions?**

* A. and
* B. or
* C. &
* D. +

**11. What is the use of the .groupby() method in Pandas?**

* A. Visualize grouped data
* B. Group data by a specific column
* C. Count unique values
* D. Sort the dataset

**12. What does the .sort\_values() method do?**

* A. Drops duplicate rows
* B. Sorts a DataFrame based on a column
* C. Groups data into categories
* D. Reverses a DataFrame

**13. Which plot is best to check the distribution of data?**

* A. Line Plot
* B. Scatter Plot
* C. Histogram
* D. Box Plot

**14. What do the xticks and yticks functions control in Matplotlib?**

* A. Chart size
* B. Axis intervals and labels
* C. Data filtering
* D. Color themes

**15. What customization makes a scatter plot more visually informative?**

* A. Adding gridlines
* B. Resizing markers based on data
* C. Changing all points to one color
* D. Removing axis labels

**16. What type of dataset is typically explored with a histogram?**

* A. Time-series data
* B. Data distributions
* C. Categorical data
* D. Geospatial data

**17. What does the apply() function do in Pandas?**

* A. Applies a function to a specific column or row
* B. Filters the dataset
* C. Sorts a DataFrame
* D. Adds a new column

**18. Which Pandas function is used to display DataFrame column information?**

* A. .info()
* B. .describe()
* C. .columns()
* D. .summary()

**19. What plot would you use to visualize sales trends over multiple years?**

* A. Histogram
* B. Line Plot
* C. Bar Chart
* D. Scatter Plot

**20. What function is used in Pandas to iterate through rows of a DataFrame?**

* A. .rows()
* B. .loop()
* C. .iterrows()
* D. .map()

**21. How do you create a new column in a Pandas DataFrame?**

* A. Use square brackets []
* B. Use the .insert() method
* C. Use the .add() method
* D. Use the .new\_column() method

**22. Which plot is best for comparing sales between product categories?**

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

**23. What type of data is best for a scatter plot?**

* A. Categorical data
* B. Numerical data with relationships
* C. Data distributions
* D. Time-series data

**24. What is the purpose of .loc[] in Pandas?**

* A. Load data
* B. Locate and filter rows/columns
* C. Sort values
* D. Visualize data

**25. How can you combine multiple conditions in Pandas filters?**

* A. Use if-else statements
* B. Use and/or directly
* C. Use & for and, | for or
* D. Use nested for loops

**26. Which method creates a bar chart directly from a DataFrame?**

* A. .bar\_chart()
* B. .plot(kind='bar')
* C. .draw\_bar()
* D. .visualize\_bar()

**27. Which method is used to convert a column to a datetime type in Pandas?**

* A. .to\_date()
* B. .to\_datetime()
* C. .convert\_date()
* D. .datetime()

**28. What feature makes scatter plots more meaningful?**

* A. Uniform marker size
* B. Randomized colors
* C. Marker size based on data
* D. Fewer data points

**29. What is an essential step before visualizing data?**

* A. Cleaning and processing data
* B. Adding more rows to the dataset
* C. Selecting only numerical columns
* D. Using default Python settings

**30. What should be avoided in visualizations?**

* A. Adding descriptive labels
* B. Using too many bins in a histogram
* C. Selecting appropriate marker sizes
* D. Using consistent color schemes

**1. What file format is used to serialize a Pandas DataFrame while preserving its data types?**

* A. CSV
* B. Pickle
* C. JSON
* D. XML

**2. Which Pandas method displays column names, data types, and missing values?**

* A. .head()
* B. .info()
* C. .describe()
* D. .columns()

**3. How can you sort a DataFrame by multiple columns in Pandas?**

* A. Using .multi\_sort()
* B. Using .sort\_values() with a list of columns
* C. Using .sort()
* D. Using .rank()

**4. What argument in .sort\_values() specifies the sorting order?**

* A. ascending
* B. order
* C. sort\_order
* D. direction

**5. How do you filter rows in Pandas for a specific condition?**

* A. Use .filter()
* B. Use conditional expressions inside brackets
* C. Use .select()
* D. Use .mask()

**6. Which operator is used in Pandas to combine multiple conditions?**

* A. and/or
* B. |/&
* C. +/-
* D. xor

**7. What does df['new\_column'] = ... do in Pandas?**

* A. Adds a new column to the DataFrame
* B. Deletes an existing column
* C. Replaces all rows with new data
* D. Creates a DataFrame copy

**8. Which method calculates cumulative sums in a column?**

* A. .sum()
* B. .cumsum()
* C. .accumulate()
* D. .cumulate()

**9. What does the .drop\_duplicates() method do in Pandas?**

* A. Deletes all rows
* B. Removes duplicate rows
* C. Sorts the DataFrame
* D. Splits data into subsets

**10. How can you normalize value counts in Pandas?**

* A. Use normalize=True in .value\_counts()
* B. Divide by the maximum value
* C. Use .normalize() method
* D. Use .scale() method

**11. What does .groupby() in Pandas achieve?**

* A. Splits the data into groups based on a column
* B. Visualizes grouped data
* C. Combines DataFrames
* D. Filters rows

**12. How can you calculate the mean for grouped data?**

* A. .groupby().mean()
* B. .grouped.mean()
* C. .aggregate('mean')
* D. .transform('mean')

**13. Which method resets a DataFrame index?**

* A. .reset\_index()
* B. .reindex()
* C. .drop\_index()
* D. .index\_reset()

**14. Which method helps filter rows based on a list of values?**

* A. .isin()
* B. .contains()
* C. .equals()
* D. .in\_list()

**15. How can you visualize missing values in a DataFrame?**

* A. .plot\_missing()
* B. .isna().sum().plot()
* C. .na\_plot()
* D. .missing\_values()

**16. What does .agg() in Pandas allow you to do?**

* A. Perform multiple aggregations on grouped data
* B. Add columns
* C. Filter rows
* D. Drop duplicates

**17. Which function creates a hierarchical index in a DataFrame?**

* A. .set\_index()
* B. .hierarchy()
* C. .multi\_index()
* D. .group\_index()

**18. How can you subset a DataFrame with hierarchical indexing?**

* A. Using .loc[]
* B. Using .filter()
* C. Using .drop()
* D. Using .subset()

**19. What argument adjusts the transparency of overlapping plots?**

* A. alpha
* B. opacity
* C. transparency
* D. blend

**20. How can you calculate the proportion of total weekly sales for each store type?**

* A. Divide each store's total sales by the overall total sales
* B. Use .groupby()
* C. Use .proportion()
* D. Divide total sales by the maximum value

**21. Which file type supports reading data using pd.read\_pickle()?**

* A. JSON
* B. Pickle
* C. CSV
* D. Excel

**22. What does .cummax() compute?**

* A. Maximum of the DataFrame
* B. Cumulative maximum of a column
* C. Group-wise maximum
* D. Overall maximum value

**23. How can you filter rows for a specific region and condition in Pandas?**

* A. Use .region()
* B. Use logical expressions inside brackets
* C. Use .mask()
* D. Use .dropna()

**24. How do you calculate the maximum of a column in a DataFrame?**

* A. .max()
* B. .aggregate()
* C. .maximum()
* D. .reduce()

**25. What does .isna() check in a DataFrame?**

* A. Validity of data types
* B. Presence of missing values
* C. Data normalization
* D. Presence of duplicate rows

**26. What does .cumsum() calculate in Pandas?**

* A. Total sum of the column
* B. Cumulative sum of values
* C. Maximum sum of values
* D. Subset of the DataFrame

**27. Which plot type is suitable for visualizing grouped data?**

* A. Bar Chart
* B. Line Plot
* C. Scatter Plot
* D. Histogram

**28. What does the .value\_counts() method provide?**

* A. Unique counts of column values
* B. Mean of numerical data
* C. Summary statistics
* D. Visualization of data

**29. How can you save a Pandas DataFrame to a file format that supports serialization?**

* A. .to\_csv()
* B. .to\_pickle()
* C. .to\_json()
* D. .to\_excel()

**30. Which function aggregates grouped data with multiple statistics?**

* A. .agg()
* B. .groupby().mean()
* C. .aggregate\_stats()
* D. .combine\_stats()

**1. What is the purpose of Exploratory Data Analysis (EDA)?**

* A. Visualizing raw data
* B. Cleaning, reviewing, and deriving insights from data
* C. Applying machine learning algorithms
* D. Removing duplicate rows

**2. Which Pandas method provides a summary of column names, data types, and missing values?**

* A. .head()
* B. .info()
* C. .describe()
* D. .columns()

**3. What does the .value\_counts() method in Pandas do?**

* A. Counts rows with missing values
* B. Counts the occurrences of unique values in a column
* C. Filters rows based on conditions
* D. Displays all unique column values

**4. Which Seaborn plot is ideal for visualizing the distribution of numerical data?**

* A. Line plot
* B. Bar plot
* C. Histogram
* D. Scatter plot

**5. How can categorical data be validated using Pandas?**

* A. Using .isin() to check for specific categories
* B. Using .validate() method
* C. Using .groupby() function
* D. Using .agg()

**6. What does the .select\_dtypes() method do in Pandas?**

* A. Selects columns based on their data type
* B. Filters rows based on conditions
* C. Normalizes numerical data
* D. Visualizes data distributions

**7. What does a boxplot display?**

* A. Mean, median, and standard deviation
* B. Data grouped by quartiles and outliers
* C. Frequency distribution of a variable
* D. Correlation between variables

**8. How can you exclude specific rows using .isin()?**

* A. By using ~ with .isin()
* B. By using .notin()
* C. By using .filter()
* D. By using .groupby()

**9. Which Seaborn plot visualizes aggregated averages and confidence intervals?**

* A. Scatter plot
* B. Histogram
* C. Bar plot
* D. Boxplot

**10. What does the .agg() function allow in Pandas?**

* A. Apply multiple aggregations to columns
* B. Visualize data distributions
* C. Drop missing values
* D. Filter rows based on conditions

**11. How can you rename columns in a grouped aggregation?**

* A. Using the .rename() method
* B. Using named aggregations in .agg()
* C. Using .set\_names()
* D. Using .reset\_index()

**12. How can missing values impact data analysis?**

* A. Introduces bias into distributions
* B. Leads to inaccurate statistical calculations
* C. Reduces representativeness of the data
* D. All of the above

**13. What does the .fillna() method do?**

* A. Drops rows with missing values
* B. Replaces missing values with a specified value
* C. Visualizes missing values
* D. Filters rows with missing data

**14. Which strategy is used to impute missing values for categorical data?**

* A. Mean imputation
* B. Mode imputation
* C. Median imputation
* D. Removing missing rows

**15. What does the .nunique() method calculate?**

* A. The number of missing values in a column
* B. The number of unique values in a column
* C. The number of outliers in a dataset
* D. The total count of values in a column

**16. How can you filter rows containing specific strings in a column?**

* A. Using .str.contains()
* B. Using .str.startswith()
* C. Using .isin()
* D. Using .agg()

**17. What does .groupby() allow in Pandas?**

* A. Combine rows with the same value into groups
* B. Visualize distributions of a variable
* C. Select rows based on data type
* D. Filter rows with missing values

**18. What is the purpose of a histogram?**

* A. Show relationships between variables
* B. Display the distribution of a numerical variable
* C. Visualize grouped data
* D. Highlight outliers in data

**19. How do you calculate interquartile range (IQR)?**

* A. Subtract 25th percentile from 75th percentile
* B. Add mean and standard deviation
* C. Multiply median by the range
* D. Divide the sum of quartiles by 2

**20. What defines outliers using IQR?**

* A. Values below 25th percentile or above 75th percentile
* B. Values outside 1.5 times the IQR from the 25th and 75th percentiles
* C. Values greater than the mean plus standard deviation
* D. Values that are unique

**21. Which Seaborn plot helps identify relationships between two variables?**

* A. Scatter plot
* B. Bar plot
* C. Histogram
* D. Line plot

**22. How can missing values below a threshold be dropped?**

* A. Using .dropna() with a subset of columns
* B. Using .fillna()
* C. Using .drop\_duplicates()
* D. Using .agg()

**23. What does .map() do in Pandas?**

* A. Applies a function to transform values in a column
* B. Groups rows based on column values
* C. Selects non-numeric columns
* D. Counts occurrences of unique values

**24. How can you create new categories for durations in a column?**

* A. Using .groupby()
* B. Using conditions and assigning categories
* C. Using .replace()
* D. Using .cut()

**25. Which of the following is not an EDA step?**

* A. Visualizing distributions
* B. Cleaning missing data
* C. Building machine learning models
* D. Validating data

**26. What does .select\_dtypes('object') filter?**

* A. Numeric columns
* B. Categorical columns
* C. Missing values
* D. Outlier rows

**27. How can the frequency of a categorical column be visualized?**

* A. Scatter plot
* B. Bar chart
* C. Line plot
* D. Boxplot

**28. What method calculates the mean for grouped data?**

* A. .mean()
* B. .agg()
* C. .describe()
* D. .select()

**29. What is the purpose of Seaborn’s boxplot?**

* A. Compare categories based on mean values
* B. Identify data distribution, quartiles, and outliers
* C. Visualize the relationship between two variables
* D. Highlight missing data

**30. Which method allows filtering rows containing multiple strings?**

* A. .filter()
* B. .str.contains() with logical conditions
* C. .isin()
* D. .agg()

**1. What is the primary purpose of converting a column to DateTime format in Pandas?**

* A. To enable statistical analysis
* B. To analyze patterns over time
* C. To clean missing values
* D. To group data by categories

**2. Which method is used to convert a column to DateTime format?**

* A. pd.to\_date()
* B. pd.to\_datetime()
* C. .convert\_to\_date()
* D. .format\_date()

**3. How can you extract the year from a DateTime column in Pandas?**

* A. .year()
* B. .extract\_year()
* C. .dt.year
* D. .date.year()

**4. Which plot is commonly used to analyze trends over time?**

* A. Scatter plot
* B. Histogram
* C. Line plot
* D. Bar chart

**5. What does a correlation coefficient of -1 represent?**

* A. Strong positive relationship
* B. Weak negative relationship
* C. Perfect negative relationship
* D. No relationship

**6. How can you visualize correlations between numeric columns in a dataset?**

* A. Scatter plot
* B. Line plot
* C. Heatmap
* D. Histogram

**7. What does Seaborn’s pairplot show?**

* A. Relationships between selected categorical variables
* B. Pairwise relationships between numerical variables
* C. Time-series trends
* D. Aggregated averages by category

**8. What is the role of the hue argument in Seaborn visualizations?**

* A. Controls the color of the plot
* B. Maps a variable to color for different categories
* C. Adjusts the transparency of the plot
* D. Sets the size of the markers

**9. How do KDE plots improve over histograms?**

* A. They remove outliers
* B. They are smoother and more interpretable
* C. They require less data cleaning
* D. They display all variables at once

**10. What does the cut argument in KDE plots control?**

* A. The number of bins
* B. The x-axis limits for smoothing
* C. The size of the plot
* D. The color intensity

**11. How can class imbalance affect EDA?**

* A. Leads to overfitting models
* B. Skews results toward the majority class
* C. Enhances the reliability of predictions
* D. Reduces data quality

**12. Which method shows the relative frequency of categorical data?**

* A. .count()
* B. .value\_counts(normalize=True)
* C. .groupby()
* D. .describe()

**13. What is cross-tabulation used for?**

* A. Aggregating data into bins
* B. Summarizing frequency combinations of categories
* C. Plotting time-series data
* D. Handling missing values

**14. How can missing values be replaced in a DataFrame?**

* A. .replace\_na()
* B. .fillna()
* C. .dropna()
* D. .update\_na()

**15. What technique is used to group numeric data into categories?**

* A. Cross-tabulation
* B. Binning
* C. Normalization
* D. Heatmapping

**16. What does a correlation coefficient close to 0 imply?**

* A. Strong positive relationship
* B. Strong negative relationship
* C. Weak or no linear relationship
* D. Perfect correlation

**17. How can you visualize the distribution of numerical data over time?**

* A. Pairplot
* B. Line plot
* C. Heatmap
* D. Scatter plot

**18. What does the annot=True argument do in Seaborn’s heatmap?**

* A. Adds annotations with correlation values
* B. Adjusts the axis labels
* C. Highlights outliers
* D. Sets the transparency

**19. What is the purpose of Seaborn’s pairplot(vars=[...])?**

* A. Display relationships between all variables
* B. Limit relationships to selected variables
* C. Group data by categories
* D. Aggregate numerical data

**20. How can DateTime attributes like weekday or month be extracted?**

* A. Using .date\_components()
* B. Using .to\_date()
* C. Using .dt accessor
* D. Using .time\_values()

**21. Which method is used to find percentiles in Pandas?**

* A. .percentile()
* B. .quantile()
* C. .describe()
* D. .percentiles()

**22. What does the sns.countplot() function visualize?**

* A. Numeric distributions
* B. Categorical counts
* C. Relationships between variables
* D. Time-series data

**23. How do quartiles split data?**

* A. Into 10 equal parts
* B. Into 25 equal parts
* C. Into 4 equal parts
* D. Into halves

**24. What is a common issue with KDE plots?**

* A. They ignore outliers
* B. They include invalid values due to smoothing
* C. They require categorical data
* D. They rely on bin sizes

**25. What type of correlation is represented by a value of 0.8?**

* A. Strong positive
* B. Weak negative
* C. No correlation
* D. Perfect negative

**26. Which argument is used in pd.to\_datetime() for parsing mixed date formats?**

* A. mixed\_format=True
* B. parse\_date\_format=True
* C. format='mixed'
* D. errors='coerce'

**27. How can you identify the strongest correlations in a heatmap?**

* A. By finding the largest absolute values
* B. By counting the most frequent values
* C. By looking for negative coefficients
* D. By ignoring diagonal values

**28. How can ticket prices be categorized into economy, business, etc.?**

* A. Using .groupby()
* B. Using percentile-based binning
* C. Using .describe()
* D. Using cross-tabulation

**29. What does .str.replace() do in Pandas?**

* A. Replaces values in a column based on conditions
* B. Replaces substrings in text data
* C. Replaces missing values in numeric data
* D. Removes outliers

**30. Why is correlation analysis important in EDA?**

* A. It helps reduce data size
* B. It uncovers relationships between variables
* C. It automates data cleaning
* D. It standardizes numerical data

**1. Which method is used in Pandas to merge two DataFrames?**

* A. .join()
* B. .merge()
* C. .concat()
* D. .append()

**2. What is an inner join?**

* A. Combines all rows from both tables
* B. Returns rows with matching values in both tables
* C. Combines rows with non-matching values only
* D. Appends one table to another

**3. Which argument in .merge() specifies the column(s) to merge on?**

* A. merge\_on
* B. key
* C. on
* D. columns

**4. What happens when there are overlapping column names in a merge?**

* A. The merge fails
* B. A suffix is automatically added to overlapping columns
* C. Only the left table’s column is kept
* D. Only the right table’s column is kept

**5. What does the suffixes argument in .merge() do?**

* A. Sets the suffixes for overlapping columns
* B. Replaces all column names with new ones
* C. Sets prefixes for all columns
* D. Combines column names

**6. What is a one-to-many relationship in data merging?**

* A. One row in the left table matches multiple rows in the right table
* B. Every row in both tables matches one other row
* C. Multiple rows in both tables match one row
* D. No rows in either table match

**7. Which merge type returns all rows from the left table and only matching rows from the right table?**

* A. Inner join
* B. Left join
* C. Right join
* D. Outer join

**8. In a right join, what happens to unmatched rows from the left table?**

* A. They are included with null values
* B. They are dropped
* C. They are combined with right table rows
* D. They appear first in the merged DataFrame

**9. What does an outer join return?**

* A. Only matching rows
* B. All rows from the left table
* C. All rows from both tables, matching or not
* D. Rows with null values only

**10. What does merge\_ordered() do?**

* A. Merges two DataFrames with a specific order
* B. Merges and sorts rows by specified columns
* C. Appends one DataFrame to another
* D. Concatenates DataFrames with overlapping rows

**11. What is the result of merging tables using multiple columns?**

* A. Only rows matching all specified columns are merged
* B. Any rows matching any column are merged
* C. All rows from the left table are kept
* D. Duplicates rows are added to the output

**12. How can you control missing values in a merge?**

* A. Using the how argument
* B. Using the null\_policy argument
* C. By forward filling missing data
* D. By dropping null values

**13. What does .concat() do in Pandas?**

* A. Merges tables with overlapping rows
* B. Appends columns or rows to a DataFrame
* C. Combines tables vertically or horizontally
* D. Drops duplicates

**14. Which argument in .concat() sets labels for concatenated tables?**

* A. keys
* B. labels
* C. index
* D. suffixes

**15. What is the default join type for .merge()?**

* A. Inner join
* B. Outer join
* C. Left join
* D. Right join

**16. What is the purpose of the how argument in .merge()?**

* A. Filters rows before merging
* B. Specifies the type of join to use
* C. Defines the output format
* D. Replaces column names

**17. How can you find the most common value in a column after a merge?**

* A. Using .agg()
* B. Using .value\_counts()
* C. Using .groupby()
* D. Using .concat()

**18. What is the purpose of using suffixes in self-joins?**

* A. Prevents column name conflicts
* B. Combines rows with similar values
* C. Creates unique column values
* D. Allows multiple columns to be joined

**19. How can .merge() handle column name mismatches between two DataFrames?**

* A. Using on argument only
* B. Using left\_on and right\_on arguments
* C. Dropping mismatched columns
* D. Renaming columns beforehand

**20. Which argument in .concat() includes only matching columns?**

* A. sort
* B. keys
* C. join="inner"
* D. suffixes

**21. How can missing data in a column be filled after merging?**

* A. Using .merge()
* B. Using .fillna()
* C. Using .replace()
* D. Using .dropna()

**22. What does merging a table to itself achieve?**

* A. Combines rows from different DataFrames
* B. Allows comparisons within the same table
* C. Drops duplicates from the table
* D. Appends rows with missing values

**23. What is a use case for the merge\_ordered() method?**

* A. Merging time-series data
* B. Joining tables with multiple columns
* C. Merging unrelated datasets
* D. Creating histograms

**24. What does a forward fill achieve in ordered merges?**

* A. Fills missing values using the previous value
* B. Deletes rows with null values
* C. Fills missing values with zeros
* D. Adds rows for missing data

**25. How can .merge() be used to identify rows missing in one table?**

* A. Using an outer join
* B. Using a right join
* C. Using a left join with null filtering
* D. Using .concat()

**26. Why would you use pd.crosstab() after merging data?**

* A. To aggregate relationships between categories
* B. To filter null values
* C. To create a heatmap
* D. To normalize columns

**27. How can multiple DataFrames be combined vertically?**

* A. Using .merge()
* B. Using .concat()
* C. Using .join()
* D. Using .groupby()

**28. What does the keys argument do when concatenating DataFrames?**

* A. Sets a multi-index for the combined DataFrame
* B. Sorts the resulting DataFrame
* C. Combines rows with duplicate indices
* D. Removes null values

**29. What does a right join return?**

* A. All rows from the left table
* B. All rows from the right table and matching rows from the left
* C. All rows from both tables
* D. Rows with null values only

**30. Why is joining data important in data analysis?**

* A. To reduce dataset size
* B. To combine and analyze data from multiple sources
* C. To remove duplicates from a dataset
* D. To visualize trends