1. Data Science Basics

Q1: What is data science mainly concerned with?
A) Only big data
B) Only machine learning
C) Both big and small data
D) Only structured data

Q2: Which of the following is NOT a key component of data science?
A) Machine learning
B) Business analytics
C) Data entry
D) Visualization

Q3: Which of the following best describes the data science pipeline?
A) Data collection → Machine learning → Deployment
B) Data collection → Processing → Exploration → Analysis → Decision making
C) Data visualization → Model training → Result evaluation
D) Database creation → Querying → Reporting

Q4: Which of the following best defines data science?
A) The study of databases and their management
B) The application of computational and statistical techniques to solve real-world problems
C) The process of collecting, storing, and retrieving data
D) The use of artificial intelligence to replace human decision-making

2. Data Representation

Q1: What is an "attribute" in a dataset?
A) A single instance of data
B) A collection of data points
C) A column in a table
D) A relationship between datasets

Q2: In a dataset, a row represents a(n) _____, while a column represents a(n) _____.
A) Attribute, Object
B) Object, Attribute
C) Feature, Dataset
D) Sample, Dataset

Q3: Which of the following is an example of ordered data?
A) Student records
B) Social media posts

C) Time-series data
D) Product inventory

Q6: Which of the following is an example of geometric/structured data?
A) Bank transactions
B) Genetic sequences
C) Social network graphs
D) Customer reviews

3. Data Science Tools

Q7: Which tool is primarily used for data visualization?
A) Pandas
B) Matplotlib
C) NumPy
D) Scikit-Learn

Q8: Which of the following is NOT a database management system?
A) MySQL
B) MongoDB
C) Jupyter Notebook
D) SQLite

4. Python for Data Science

Q9: Which Python library is used for data manipulation and analysis?
A) TensorFlow
B) Pandas
C) Matplotlib
D) Scikit-Learn

Q10: What will be the output of the following Python code?

```python
import numpy as np
arr = np.array([1, 2, 3, 4])
arr += 2
print(arr)
```

A) [1, 2, 3, 4]
B) [3, 4, 5, 6]
C) [2, 3, 4, 5]
D) Error

Q8: Which Python library is used for machine learning?
A) NumPy
B) Scikit-Learn
C) Matplotlib
D) Tableau

Q12: How can you access a specific row in a Pandas DataFrame using an index label?
A) `df.iloc[]`
B) `df.loc[]`
C) `df.index[]`
D) `df.find[]`

Q13: What will the following Pandas code output?

```
import pandas as pd
df = pd.DataFrame({'A': [10, 20, 30], 'B': [40, 50, 60]})
print(df.iloc[1])
```

A) The first row
B) The second row
C) The last row
D) An error

Q14: What does the `drop()` function do in Pandas?
A) Drops missing values from a DataFrame
B) Deletes rows or columns from a DataFrame
C) Sorts values in a DataFrame
D) Adds a new column to a DataFrame

Q3: What is the main advantage of using NumPy over Python lists?
A) NumPy arrays are faster and more memory-efficient
B) NumPy does not require Python installation
C) NumPy only works with small datasets
D) NumPy does not support mathematical operations

Q4: How do you create a NumPy array from a Python list?
A) `array = numpy.array(list)`
B) `array = list.to_numpy()`
C) `array = numpy.create(list)`
D) `array = numpy.list_to_array()`

Q6: How can you create a 2D NumPy array?
A) `np.array([1, 2, 3], [4, 5, 6])`
B) `np.array([[1, 2, 3], [4, 5, 6]])`
C) `np.array({1, 2, 3, 4, 5, 6})`
D) `np.array(1, 2, 3, 4, 5, 6)`

Q9: What will be the output of the following code?

```
import numpy as np
arr = np.array([[1, 2, 3], [4, 5, 6]])
print(arr.shape)
```

A) `(2, 3)`
B) `(3, 2)`
C) `(6,)`
D) `(2, 2)`

Q11: Which function creates an array of zeros?
A) `np.empty()`
B) `np.zeros()`
C) `np.ones()`
D) `np.fill_zeros()`

Q12: What does `np.linspace(1, 10, 5)` return?
A) `[1, 10, 5]`
B) `[1, 3.25, 5.5, 7.75, 10]`
C) `[1, 2, 3, 4, 5]`
D) `[1, 4, 7, 10]`

Q15: How can you select the first row of a 2D NumPy array `arr`?
A) `arr[0, :]`
B) `arr[:, 0]`
C) `arr[1, :]`
D) `arr[0]`

1. Data Collection

Q1: What is the first step in data science?
A) Data visualization
B) Data collection
C) Data modeling
D) Machine learning

Q2: Which of the following is NOT a common method for collecting data?
A) Downloading a data file manually
B) Querying a database
C) Using machine learning models
D) Scraping data from a webpage

Q3: Which Python library is commonly used for issuing HTTP requests?
A) BeautifulSoup
B) Matplotlib
C) Requests
D) NumPy

2. HTTP Requests & APIs

Q4: Which HTTP method is commonly used to retrieve data?
A) GET
B) PUT
C) DELETE
D) UPDATE

Q5: RESTful APIs are based on which of the following principles?
A) Stateful connections
B) Uses standard HTTP methods
C) Requires a continuous connection
D) Only supports GET requests

3. Data Formats

Q6: Which of the following is a common data format used in data science?
A) PNG
B) CSV
C) MP3
D) DOCX

Q7: In JSON format, data is stored as _____.
A) Tables and columns
B) Key-value pairs
C) Image pixels
D) Binary code

Q8: Which Python library is commonly used for parsing JSON data?
A) BeautifulSoup
B) Matplotlib
C) JSON
D) Pandas

4. Web Scraping & XML

Q9: Which Python library is commonly used for web scraping?
A) NumPy
B) BeautifulSoup
C) TensorFlow
D) Scikit-Learn

Q10: XML files use which type of structure to store data?
A) Key-value pairs
B) Tables and columns
C) Hierarchical tags
D) Binary encoding

5. Data Visualization Basics

Q11: Why is data visualization important?
A) To increase the file size of datasets
B) To analyze and interpret data patterns
C) To replace machine learning models
D) To store large amounts of data

Q12: Which Python library is mainly used for creating plots and charts?
A) NumPy
B) Matplotlib
C) Pandas
D) Requests

6. Types of Visualization

Q13: Which visualization technique is used to display multivariate data?
A) Scatter plot
B) Bar chart
C) Spidergram (Star Plot)
D) Histogram

Q14: What is a Chernoff Face used for?
A) Representing data using human face features
B) Displaying time-series data
C) Showing text-based data
D) Analyzing database queries

Q15: A heatmap is commonly used to visualize:
A) Images
B) Matrix-like data
C) Sound waves
D) Text files

Q16: Which type of chart is best suited for showing the distribution of a numerical variable?
A) Bar chart
B) Histogram
C) Line graph
D) Pie chart

7. Python Visualization Libraries

Q17: Which Python library is built on top of Matplotlib and simplifies statistical data visualization?
A) Seaborn
B) Pandas
C) Requests
D) NumPy

Q18: Which function in Seaborn is used to create a scatter plot with a regression line?
A) `sns.barplot()`
B) `sns.histplot()`
C) `sns.lmplot()`
D) `sns.heatmap()`

Q19: Which function in Seaborn allows coloring of data points based on a categorical variable?

A) `hue`
B) `color_palette`
C) `bins`
D) `facetgrid`

Q20: What type of visualization does the Seaborn `swarmplot()` function create?

A) Heatmap
B) Scatter plot with categorical separation
C) Line graph
D) Histogram