Pandas and Data Science Series Quiz

1. Which Python library is most commonly used for data manipulation and analysis?
A. Matplotlib

B. Pandas

C. NumPy

D. Requests

Answer: B

2. What is a "Series" in Pandas?
A. A multi-dimensional table.

B. A one-dimensional labelled array.

C. A type of loop.

D. A database connection.

Answer: B

3. Which command is used to import the Pandas library into your script?
A. import pandas as pd
B. import pd as pandas
C. load pandas
D. include pandas
Answer: A

4. A two-dimensional, size-mutable, and tabular data structure in Pandas is called a:
A. Series

B. Panel

C. DataFrame

D. Matrix

Answer: C

5. Which method is used to view the first 5 rows of a DataFrame?
A. tail()
B. first()
C. show()
D. head()
Answer: D
6. How do you create a Series from a simple Python list named my_list?
A. pd.DataFrame(my_list)

B. pd.Series(my_list)
C. pd.List(my_list)
D. pd.Array(my_list)
Answer: B

7. By default, what is the starting index of a Pandas Series?
A. 1

B. -1

C. 0

D. A

Answer: C

8. Which attribute is used to get the dimensions (number of rows and columns) of a DataFrame?
A. .size
B. .dim
C. .shape
D. .length
Answer: C

9. Which Pandas method is used to read data from a CSV file?
A. read_file()
B. read_csv()
C. open_csv()
D. get_csv()
Answer: B

10. How do you access a specific column named "Age" from a DataFrame called df?
A. df["Age"]
B. df.getColumn("Age")
C. df.Age()
D. df[Age]
Answer: A

11. Which function is used to get a statistical summary (mean, std, min, etc.) of a DataFrame?
A. summary()
B. info()
C. describe()
D. stats()
Answer: C

12. Which method is used to remove rows with missing (null) values?

A. remove_null()
B. dropna()
C. delete_na()
D. clean()
Answer: B

13. To rename a specific column in a DataFrame, which method do you use?
A. change_name()
B. rename()
C. update()
D. set_column()
Answer: B

14. Which attribute returns the data types of each column in a DataFrame?
A. .types
B. .dtype
C. .dtypes
D. .format
Answer: C

15. How do you select a row by its integer location (index position)?
A. .loc[]
B. .iloc[]
C. .index[]
D. .row[]
Answer: B

16. Which method is used to fill missing (NaN) values with a specific value?
A. fillna()
B. replace_na()
C. fix_null()
D. add_data()
Answer: A

17. What does the `info()` method do?
A. Returns the first 10 rows.

B. Provides a summary of the DataFrame, including non-null counts and memory usage.

C. Deletes the DataFrame.

D. Sorts the data.

Answer: B

18. To sort a DataFrame based on the values of a specific column, you use:
A. order_by()
B. sort_values()
C. arrange()
D. rank()
Answer: B
19. Which command is used to find the number of unique values in a column?
A. distinct()
B. nunique()
C. unique_count()
D. count()
Answer: B

20. How can you delete a specific column from a DataFrame?
A. df.remove("column_name")
B. df.drop(columns="column_name")
C. df.clear("column_name")
D. df.delete("column_name")
Answer: B

21. What is the correct way to filter rows where the "Salary" column is greater than 50000?
A. df[df["Salary"] > 50000]
B. df.filter("Salary > 50000")
C. df.where("Salary" > 50000)
D. df["Salary"] > 50000
Answer: A

22. Which Pandas function is used to merge two DataFrames together?
A. combine()
B. join_tables()
C. merge()
D. connect()
Answer: C

23. What does the `isnull()` method return?
A. A single number representing nulls.

B. A DataFrame of Boolean values (True for nulls).

C. The rows that are empty.

D. Nothing.

Answer: B

24. Which method allows you to change the index of a DataFrame to a specific column?
A. set_index()
B. change_index()
C. reset_index()
D. new_index()
Answer: A

25. To export a DataFrame to an Excel file, which method is used?
A. save_excel()
B. to_excel()
C. write_excel()
D. export_excel()
Answer: B

26. Which Python library is most widely used for creating basic 2D graphs and plots?
A. Requests

B. NumPy

C. Matplotlib

D. Flask

Answer: C

27. Which specific module of Matplotlib is usually imported to create plots quickly?
A. matplotlib.graphs
B. matplotlib.pyplot
C. matplotlib.draw
D. matplotlib.show
Answer: B

28. Which type of chart is best for showing the relationship between two numerical variables using dots?
A. Bar Chart

B. Pie Chart

C. Scatter Plot

D. Histogram

Answer: C

29. What is the correct method to add a label to the horizontal axis (X-axis) of a plot?

A. plt.xlabel()
B. plt.xtext()
C. plt.labelX()
D. plt.horizontal_label()

Answer: A

30. Which function is used to actually display the window containing your graph after you have defined it?

A. plt.open()
B. plt.display()
C. plt.show()
D. plt.render()

Answer: C