10 Quiz Questions

Submitted by Yamini Manral 002703881

Question 1: What is the primary purpose of data cleaning in data science?

A. To increase the size of the dataset

B. To introduce errors into the data

C. To prepare the data for analysis and ensure data quality

D. To remove all missing values

Answer: C

Question 2: Which Python library is commonly used for data cleaning and manipulation?

A. NumPy

B. SciPy

C. TensorFlow

D. PyTorch

Answer: A

Question 3: In pandas, how can you replace all NaN values in a DataFrame with a specific value, such as "Unknown"?

A. df.replace("Unknown", np.nan)

B. df.fillna("Unknown")

C. df.replace(np.nan, "Unknown")

D. df.fillna(np.nan, "Unknown")

Answer: B

Question 4 (Coding): You have a DataFrame df with columns 'A' and 'B'. How can you drop column 'B' from the DataFrame?

A. df.drop('B')

B. df.drop_column('B')

C. df.drop(columns='B')

D. df.remove_column('B')

Answer: C

Explanation: df.drop(columns='B') is the correct way to drop a column in pandas.

Question 5: What is a common technique for handling categorical data during data cleaning?

A. Replacing categorical data with zeros

B. Removing all categorical data from the dataset

C. Converting categorical variables into numerical representations

D. Ignoring categorical data during analysis

Answer: C

Question 6 (Coding): How can you merge two columns 'First_Name' and 'Last_Name' into a new column 'Full_Name' in a DataFrame df?

A. df['Full_Name'] = df['First_Name'] + df['Last_Name']

B. df['Full_Name'] = df['First_Name'] - df['Last_Name']

C. df['Full_Name'] = df['First_Name'].concat(df['Last_Name'])

D. df['Full_Name'] = df['First_Name'].merge(df['Last_Name'])

Answer: A

Explanation: To concatenate two columns, you can use the + operator in pandas.

Question 7: What is the purpose of data validation during data cleaning?

A. To increase the size of the dataset

B. To convert data into a different format

C. To check data against external sources or domain-specific rules

D. To ignore data quality issues

Answer: C

Question 8: Which data cleaning technique is commonly used to handle outliers in numeric columns?

A. Replacing outliers with the column's median

B. Deleting the entire column with outliers

C. Ignoring outliers during analysis

D. Using regular expressions to clean outliers

Answer: A

Question 9: What is the primary purpose of standardizing data during data cleaning?

A. To make all data values the same

B. To ensure data is in a readable format

C. To prepare data for analysis by scaling it to a common range

D. To remove all missing values

Answer: C

Question 10 (Coding): You have a DataFrame df with columns 'A', 'B', and 'C'. How can you reorder the columns so that 'C' appears first?

A. df = df[['C', 'A', 'B']]

B. df.reorder_columns(['C', 'A', 'B'])

C. df.change_column_order(['C', 'A', 'B'])

D. df.set_columns_order(['C', 'A', 'B'])

Answer: A