Test Pandas

For this test create an testResult_topic_yourName.py and for each answer write an example code that is executable and correct!

Pandas Knowledge Test

Question 1:

What is the primary data structure in Pandas for handling tabular data?

- a) `DataArray`
- b) `Series`
- c) `DataFrame`
- d) `Table`

Question 2:

Which of the following methods is used to read a CSV file into a Pandas DataFrame?

- a) `pd.read_file()`
- b) `pd.read_csv()`
- c) `pd.import_csv()`
- d) `pd.load_csv()`

Question 3:

How do you display the first 5 rows of a DataFrame named `df`?

- a) `df.top(5)`
- b) `df.show(5)`
- c) `df.head(5)`
- d) `df.start(5)`

Question 4:

Which function in Pandas is used to get a summary of the basic statistics for numeric columns?

- a) `df.summary()`
- b) `df.info()`
- c) `df.describe()`
- d) `df.statistics()`

Question 5:

If you want to filter a DataFrame `df` to show only rows where the column `age` is greater than 30, what is the correct syntax?

- a) `df.age > 30`
- b) `df[df['age'] > 30]`
- c) `df.where('age', 30)`
- d) `df.filter('age > 30')`

Question 6:

How can you rename the columns of a DataFrame `df` from their current names to new names `['A', 'B', 'C']`?

- a) `df.rename(['A', 'B', 'C'])`
- b) `df.columns(['A', 'B', 'C'])`
- c) `df.rename(columns={'old1':'A', 'old2':'B', 'old3':'C'})`
- d) `df.columns = ['A', 'B', 'C']`

Question 7:

Which method would you use to remove rows with missing data (NaN) in a DataFrame?

- a) `df.remove_na()`
- b) `df.dropna()`

- c) `df.clean()`
- d) `df.clearna()`

Question 8:

How do you calculate the mean of values in the column `salary` of a DataFrame `df`?

- a) `df.mean('salary')`
- b) `df['salary'].avg()`
- c) `df['salary'].mean()`
- d) `df.salary.mean()`

Question 9:

Which of the following methods is used to merge two DataFrames in Pandas?

- a) `pd.combine()`
- b) `pd.append()`
- c) `pd.concat()`
- d) `pd.merge()`

Question 10:

How can you group a DataFrame `df` by the column `department` and compute the sum of the `salary` column for each department?

- a) `df.groupby('department')['salary'].sum()`
- b) `df.group_by('department').sum('salary')`
- c) `df.groupby('department').agg(sum('salary'))`
- d) `df.groupby_sum('department', 'salary')`