

1 Check if any of the record of the dataset is duplicated

- ☐ There are 29 duplicated rows
- ☒ There are 9 duplicated rows
- ☐ There are not duplicated rows
- ☐ There are 5 duplicated rows

Submitted Correct!

Report issue

Checking for duplicates in a specific column is an essential step in data preprocessing and analysis. It ensures data accuracy, reliability, and consistency, which are fundamental for drawing meaningful conclusions and making informed decisions based on the data.

2 Your task is to check for duplicates in a specific column of a given DataFrame. Select the correct code

- ☐ df.duplicates('TypeText')
- ☒ df.duplicated('TypeText')
- ☐ df.duplicated('TypeText',axis=1)

Submitted Correct!

Report issue

5 True or False: Ignoring duplicate entries in a dataset is a recommended practice, as they don't significantly impact the accuracy of data analysis and decision-making.

- ☐ True
- ☒ False.

Submitted Correct!

Report issue

3 Drop duplicated

Your task is to drop duplicate rows from a given DataFrame and retain the first occurrence of each duplicated row. Select the correct code.

- ☒ df.drop_duplicates(keep='first',inplace=True)
- ☐ df.dropna()
- ☐ df.drop_duplicates(keep='last',inplace=True)
- ☐ df.drop_duplicates('first')

Submitted Correct!

Report issue

Quiz

4 Scenario question

You work as a data analyst for an e-commerce company that sells a wide range of products. The company has provided you with a dataset of customer reviews, which includes attributes such as the product purchased, review text, rating, and demographic information of customers. Your task is to leverage this data to enhance the product recommendation system.

Scenario: You have conducted an initial analysis of the customer review dataset. During this analysis, you identified that there are several duplicate entries in the dataset, which can lead to biased results and flawed conclusions. What should be your next step to address this issue?

- ☒ Implement data preprocessing techniques to identify and remove duplicate entries effectively.
- ☐ Merge the duplicate entries into single records to increase the dataset's size for more accurate recommendations.
- ☐ Leave the duplicates as is and conduct a separate analysis on them to explore potential patterns.

Submitted Correct!

Report issue