V2_removing_data

November 2, 2025

1 Activity: Removing Data

1.1 Introduction

In this activity you will practice using Pandas functionality to check for and remove any unwanted data from a dataset. This activity will cover the following topics: - Removing columns from a DataFrame - Removing rows from a DataFrame - Removing rows based on a condition - Checking for duplicate data

Question 1 Create a `DataFrame` called `df` from the given CSV file `exotic_plants_data.csv`, then drop the column `Type` and assign the result to a new `DataFrame` called `df_no_type`.

[ ]:
```
import pandas as pd


# Your code here

# your code here
```

[ ]:
```
# Question 1 Grading Checks

assert isinstance(df, pd.DataFrame), 'Have you created a DataFrame named df?'
assert isinstance(df_no_type, pd.DataFrame), 'Have you created a DataFrame␣
↪named df_no_type?'
```

Question 2 Check the `df DataFrame` for any duplicate rows and assign the result to a new `DataFrame` called `df_duplicates`.

[ ]:
```
# Your code here

# your code here
```

[ ]:
```
# Question 2 Grading Checks

assert isinstance(df_duplicates, pd.DataFrame), 'Have you created a DataFrame␣
↪named df_duplicates?'
```

Question 3 Check the `df DataFrame` for any duplicate rows based on the `Plant Name` and `Type` columns and assign the result to a new `DataFrame` called `df_plant_type_duplicates`.

[ ]:
```
# Your code here

# your code here
```

[ ]:
```
# Question 3 Grading Checks

assert isinstance(df_plant_type_duplicates, pd.DataFrame), 'Have you created a␣
↪DataFrame named df_duplicates?'
```

Question 4 Create a mask called `clean_mask` that will clean up any duplicates in the `df` DataFrame that have the same `Plant Name` and `Origin` and only keep the most up-to-date duplicate entry.

[ ]:
```
# Your code here

# your code here
```

[ ]:
```
# Question 4 Grading Checks

assert isinstance(clean_mask, pd.Series), 'Have you created a Series named␣
↪clean_mask?'
```