

Finding the Number of Unique Values in Non-Numeric Columns

Question and Screenshot:

Learn / Courses / Exploratory Data Analysis In Pyt...

← Course Outline →

Daily XP 1036

Exercise

Finding the number of unique values

You would like to practice some of the categorical data manipulation and analysis skills that you've just seen. To help identify which data could be reformatted to extract value, you are going to find out which non-numeric columns in the `planes` dataset have a large number of unique values.

`pandas` has been imported for you as `pd`, and the dataset has been stored as `planes`.

Instructions 100 XP

- Filter `planes` for columns that are of "object" data type.
- Loop through the columns in the dataset.
- Add the column iterator to the print statement, then call the function to return the number of unique values in the column.

Take Hint (-30 XP)

script.py

Light Mode

```
1 # Filter the DataFrame for object columns
2 non_numeric = planes.____("____")
3
4 # Loop through columns
5 for ____ in non_numeric.____:
6
7     # Print the number of unique values
8     print(f"Number of unique values in {____} column: ",
          non_numeric[____].____())
```

↺ Run Code Submit Answer

IPython Shell Slides

In [1]:

Question Explanation:

This task involves filtering the dataset to focus on non-numeric (object-type) columns and iterating through them to find and print the number of unique values in each column. This helps identify potential categorical features for analysis.

Code Solution:

```
# Filter the DataFrame for object columns
non_numeric = planes.select_dtypes(include=["object"])

# Loop through columns
for column in non_numeric.columns:
    # Print the number of unique values
    print(f"Number of unique values in {column} column: ",
          non_numeric[column].nunique())
```

Solution Explanation:

1. The `select_dtypes()` method filters the DataFrame to include only columns with an object data type.
2. A for-loop iterates through the filtered columns to process each column individually.
3. The `nunique()` function calculates the number of unique values in the current column.
4. The results are printed with a formatted string that includes the column name and the number of unique values.