

Correcting Flight Duration - Removing 'h' and Plotting Histogram

Question and Screenshot:

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

← Course Outline →

Daily XP 1376

Exercise

Flight duration

You would like to analyze the duration of flights, but unfortunately, the "Duration" column in the `planes` DataFrame currently contains string values.

You'll need to clean the column and convert it to the correct data type for analysis. `seaborn` has been imported as `sns`.

Instructions 4/4 25 XP ✓ ✓ ✓ 4

- Plot a histogram of "Duration" values.

Take Hint (-7 XP)

script.py

Light Mode

```
0
7 # Convert to float data type
8 planes["Duration"] = planes["Duration"].astype(float)
9
10 # Plot a histogram
11 ----
12 plt.show()
```

↶ Run Code Submit Answer

IPython Shell Slides

<script.py> output:

```
0          19.0h
1  5.416666666666667h
2          4.75h
3  2.4166666666666665h
4          15.5h
Name: Duration, dtype: object
```

In [1]:

Question Explanation:

The task is to visualize the distribution of flight durations using a histogram. Before plotting, the 'Duration' column must be cleaned to remove the 'h' character and converted to a float for analysis.

Code Solution:

```
# Remove the string character "h"
planes["Duration"] = planes["Duration"].str.replace("h", "", regex=False)

# Convert to float data type
planes["Duration"] = planes["Duration"].astype(float)

# Plot a histogram
sns.histplot(data=planes, x="Duration", kde=True)
plt.show()
```

Solution Explanation:

1. The `str.replace()` method removes the 'h' character from the 'Duration' column values, making them suitable for conversion.
2. The `astype(float)` method converts the cleaned values into a float data type, enabling numerical operations.
3. The `sns.histplot()` function plots a histogram, showing the distribution of durations, with `kde=True` adding a smooth density curve.