

Identifying Potential Outliers in Price and Duration

Question and Screenshot:

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

← Course Outline →

Daily XP 1697

Exercise

Identifying outliers

You've proven that you recognize what to do when presented with outliers, but can you identify them using visualizations?

Try to figure out if there are outliers in the "Price" or "Duration" columns of the `planes` DataFrame.

`matplotlib.pyplot` and `seaborn` have been imported for you as `plt` and `sns` respectively.

Instructions 3/3 30 XP

Question

Which column potentially contains outliers?

Possible answers

☒ "Price"

☐ "Duration"

☐ "Price" and "Duration"

☐ Neither

Submit Answer

Take Hint (-9 XP)

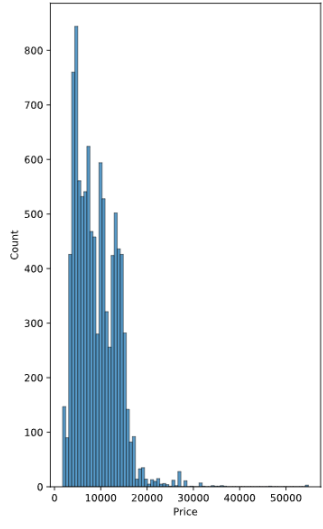
script.py

Light Mode

1

Run Code

Plots



← Previous Plot 1/1 Next Plot →

IPython Shell

Slides

<script.py> output:

count	10446.000
mean	10.724
std	8.472
min	0.083
25%	2.833
50%	8.667
75%	15.500
max	47.667

Name: Duration, dtype: float64

In [1]:

Question Explanation:

The task involves analyzing the 'Price' and 'Duration' columns in the `planes` DataFrame to identify potential outliers. The distribution plot for 'Price' and the descriptive statistics for 'Duration' are provided to aid in this analysis.

Analysis and Conclusion:

1. The histogram of 'Price' indicates a right-skewed distribution, with a small number of flights having significantly higher prices.

2. The descriptive statistics for 'Duration' reveal a maximum value of 47.667 hours, which is unusually high compared to the mean of 10.724 hours and the 75th percentile of 15.560 hours.
3. Both columns contain potential outliers based on the provided visualization and statistics.

Solution and Answer:

The columns 'Price' and 'Duration' both potentially contain outliers based on the evidence provided.

Selecting 'Price' and 'Duration' is the correct choice for this question.