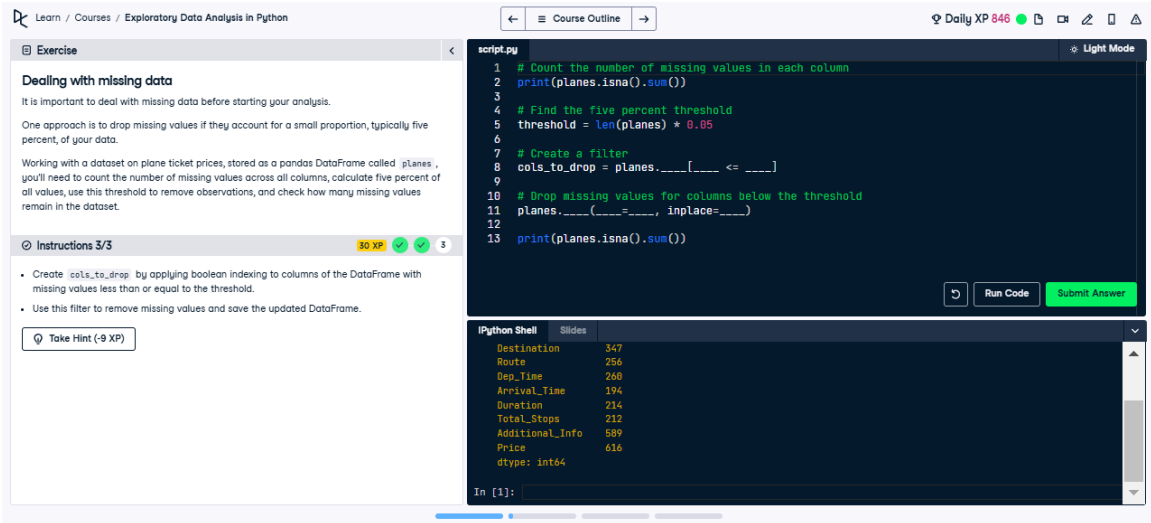
Dealing with Missing Data - Dropping Columns Below Threshold

Question and Screenshot:



Question Explanation:

This task involves identifying columns in the 'planes' DataFrame where the count of missing values is less than or equal to the calculated five percent threshold. The identified columns are then used to drop missing values specifically from these columns.

Code Solution:

```
import pandas as pd

# Count the number of missing values in each column
print(planes.isna().sum())

# Find the five percent threshold
threshold = len(planes) * 0.05

# Create a filter for columns with missing values less than or equal to the threshold
cols_to_drop = planes.columns[planes.isna().sum() <= threshold]

# Drop missing values for columns below the threshold
planes.dropna(subset=cols_to_drop, inplace=True)
```

```
# Print the remaining missing values in the DataFrame
print(planes.isna().sum())
```

Solution Explanation:

1. The `isna().sum()` function identifies and counts missing values in each column.
2. The threshold is calculated as five percent of the total number of rows.
3. Boolean indexing is used to create `cols_to_drop`, which includes column names with missing values less than or equal to the threshold.
4. The `dropna()` method removes rows with missing values only in the filtered columns, and the changes are applied in place.
5. The missing values in the updated DataFrame are printed for verification.