

# Dealing with Missing Data - Five Percent Threshold

## Question and Screenshot:

The screenshot shows a web-based coding environment. On the left, there's a sidebar with the exercise title 'Dealing with missing data' and instructions. The main area on the right contains a code editor with a Python script. The script is as follows:

```
1 # Count the number of missing values in each column
2 print(planes.isna().sum())
3
4 # Find the five percent threshold
5 threshold = ____ * ____
```

Below the code editor, there's an 'IPython Shell' window showing the output of the code:

```
Destination    347
Route          256
Dep_Time       260
Arrival_Time   194
Duration       214
Total_Stops    212
Additional_Info 589
Price          616
dtype: int64
```

## Question Explanation:

The task involves calculating the five percent threshold of the number of observations in the 'planes' DataFrame. This threshold is useful for identifying whether columns with missing values exceed a tolerable level and should be removed.

## Code Solution:

```
import pandas as pd
```

```
# Assuming 'planes' DataFrame is already available
# Count the total number of rows in the DataFrame
total_rows = len(planes)
```

```
# Calculate the five percent threshold
threshold = 0.05 * total_rows
```

```
# Print the threshold value
print(f"Five percent threshold: {threshold}")
```

## Solution Explanation:

1. The `len()` function is used to get the total number of rows in the 'planes' DataFrame.
2. The five percent threshold is calculated by multiplying the total rows by

0.05.

3. The threshold value is then printed, indicating the number of rows that correspond to five percent of the dataset.