

Imputing Missing Prices - Complete Solution

Question and Screenshot:

Learn / Courses / Exploratory Data Analysis In Python

← Course Outline →

Daily XP 977

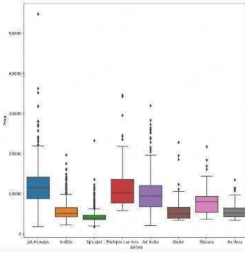
Exercise

Imputing missing plane prices

Now there's just one column with missing values left!

You've removed the "Additional_Info" column from planes—the last step is to impute the missing data in the "Price" column of the dataset.

As a reminder, you generated this boxplot, which suggested that imputing the median price based on the "Airline" is a solid approach!



Instructions 3/3 30 XP

- Conditionally impute missing values for "Price" by mapping values in the "Airline" column based on prices_dict.
- Check for remaining missing values.

Take Hint (-9 XP)

script.py

```
1 # Calculate median plane ticket prices by Airline
2 airline_prices = planes.groupby("Airline")["Price"].median()
3
4 print(airline_prices)
5
6 # Convert to a dictionary
7 prices_dict = airline_prices.to_dict()
8
9 # Map the dictionary to missing values of Price by Airline
10 planes["Price"] = ____["____"].____(____["____"]).____(____)
11
12 # Check for missing values
13 print(____)
```

Run Code Submit Answer

IPython Shell

```
<script.py> output:
Airline
Air Asia      5192.0
Air India     9443.0
GoAir         5083.5
IndiGo        5854.0
Jet Airways   11587.0
Multiple carriers 10197.0
SpiceJet      3873.0
Vistara       8028.0
Name: Price, dtype: float64

In [1]:
```

Question Explanation:

This task involves calculating the median ticket prices for each 'Airline', converting the results to a dictionary, and mapping the dictionary to the 'Price' column to impute missing values. Finally, it verifies if any missing values remain.

Code Solution:

```
# Calculate median plane ticket prices by Airline
airline_prices = planes.groupby("Airline")["Price"].median()

print(airline_prices)

# Convert to a dictionary
```

```
prices_dict = airline_prices.to_dict()

# Map the dictionary to the missing values
planes["Price"] = planes["Price"].fillna(planes["Airline"].map(prices_dict))

# Check for missing values
print(planes.isna().sum())
```

Solution Explanation:

1. The `groupby()` method groups the data by 'Airline', and the `median()` function calculates the median ticket price for each airline.
2. The resulting Series is converted into a dictionary (`prices_dict`) for easier mapping.
3. The `fillna()` method imputes missing values in the 'Price' column by mapping the `prices_dict` to the 'Airline' column.
4. The `isna().sum()` function checks if any missing values remain in the DataFrame, confirming successful imputation.