

CSV to DataFrame (Extended)

Continuing with the airline analysis, you'll now calculate the total number of passengers and involuntarily bumped passengers for each airline by grouping the data and summing the relevant columns.

Instructions:

1. For each airline group, select the `nb_bumped` and `total_passengers` columns.
2. Calculate the sum of these columns for both years.
3. Store the result as `airline_totals`.

Original Uploaded Image:

The screenshot shows a web browser window with a DataCamp exercise titled "CSV to DataFrame". The exercise instructions state: "You work for an airline, and your manager has asked you to do a competitive analysis and see how often passengers flying on other airlines are involuntarily bumped from their flights. You got a CSV file (airline_bumping.csv) from the Department of Transportation containing data on passengers that were involuntarily denied boarding in 2016 and 2017, but it doesn't have the exact numbers you want. In order to figure this out, you'll need to get the CSV into a pandas DataFrame and do some manipulation!"

The instructions also mention: "pandas is imported for you as pd. 'airline_bumping.csv' is in your working directory."

The exercise instructions are: "For each airline group, select the nb_bumped and total_passengers columns, and calculate the sum (for both years). Store this as airline_totals."

The Python code in the editor is:

```
1 # From previous step
2 airline_bumping = pd.read_csv("airline_bumping.csv")
3 print(airline_bumping.head())
4
5 # For each airline, select nb_bumped and total_passengers and sum
6 airline_totals = airline_bumping.groupby("airline")[["nb_bumped",
7 "total_passengers"]].sum()
```

The terminal output shows the following table:

	airline	year	nb_bumped	total_passengers
0	DELTA AIR LINES	2017	679	99796155
1	VIRGIN AMERICA	2017	165	6090029
2	JETBLUE AIRWAYS	2017	1475	2725038
3	UNITED AIRLINES	2017	2067	70030905
4	HAWAIIAN AIRLINES	2017	92	8422734

Python Code Implementation:

```
# From previous step
airline_bumping = pd.read_csv("airline_bumping.csv")
print(airline_bumping.head())

# For each airline, select nb_bumped and total_passengers and sum
airline_totals = airline_bumping.groupby("airline")[["nb_bumped",
"total_passengers"]].sum()

# Print the result
```

```
print(airline_totals)
```

Explanation of Code:

1. ****Load CSV file****: Use `pd.read_csv()` to read the `airline_bumping.csv` file into a DataFrame named `airline_bumping`. Print the first few rows to understand the data.
2. ****Group by airline****: Use `groupby("airline")` to group the DataFrame by the `airline` column.
3. ****Select and sum columns****: Use `[["nb_bumped", "total_passengers"]]` to select the relevant columns and call `sum()` to compute their totals for each airline.
4. ****Store the result****: Store the aggregated totals in a new DataFrame called `airline_totals` and print it.