

## CSV to DataFrame (Final Results)

After completing the data manipulations, the final step is to view the results of the analysis, which includes the total number of passengers, total bumps, and bumps per 10k passengers for each airline.

Instructions:

1. Print the `airline\_totals` DataFrame to see the results of your manipulations.

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 that pandas is imported as pd and the CSV file is in the working directory.

The Python code in the script.py file is as follows:

```
1 # From previous steps
2 airline_bumping = pd.read_csv("airline_bumping.csv")
3 print(airline_bumping.head())
4 airline_totals = airline_bumping.groupby("airline")[["nb_bumped", "total_passengers"]].sum()
5 airline_totals["bumps_per_10k"] = airline_totals["nb_bumped"] / airline_totals
6   ["total_passengers"] * 10000
7
8 # Print airline_totals
9
```

The output of the code is displayed in the Python Shell, showing a DataFrame with columns: airline, nb\_bumped, total\_passengers, and bumps\_per\_10k. The data is as follows:

airline	nb_bumped	total_passengers	bumps_per_10k
ALASKA AIRLINES	1392	36543121	0.381
AMERICAN AIRLINES	11115	197345225	0.563
DELTA AIR LINES	1591	197833215	0.081
EXPRESSJET AIRLINES	3326	27858678	1.194
FRONTIER AIRLINES	1228	22954995	0.535
HAWAIIAN AIRLINES	122	16577572	0.074
JETBLUE AIRWAYS	3615	53245086	0.679
SKYWEST AIRLINES	3894	47892737	0.457
SOUTHWEST AIRLINES	18585	228142636	0.815
SPIRIT AIRLINES	2920	32304571	0.904
UNITED AIRLINES	4941	134468897	0.367
VIRGIN AMERICA	242	12617967	0.201

## Python Code Implementation:

```
# From previous steps
```

```
airline_bumping = pd.read_csv("airline_bumping.csv")
airline_totals = airline_bumping.groupby("airline")[["nb_bumped",
"total_passengers"]].sum()
airline_totals["bumps_per_10k"] = airline_totals["nb_bumped"] /
airline_totals["total_passengers"] * 10000
```

```
# Print airline_totals
print(airline_totals)
```

## Explanation of Code:

1. **\*\*Load and group data\*\***: Previously, the CSV file was read and grouped by `airline` to calculate totals for `nb\_bumped` and `total\_passengers`.
2. **\*\*Calculate bumps per 10k\*\***: A column `bumps\_per\_10k` was added to calculate the number of bumps per 10,000 passengers.
3. **\*\*Print the DataFrame\*\***: Finally, `airline\_totals` is printed to display the results.