

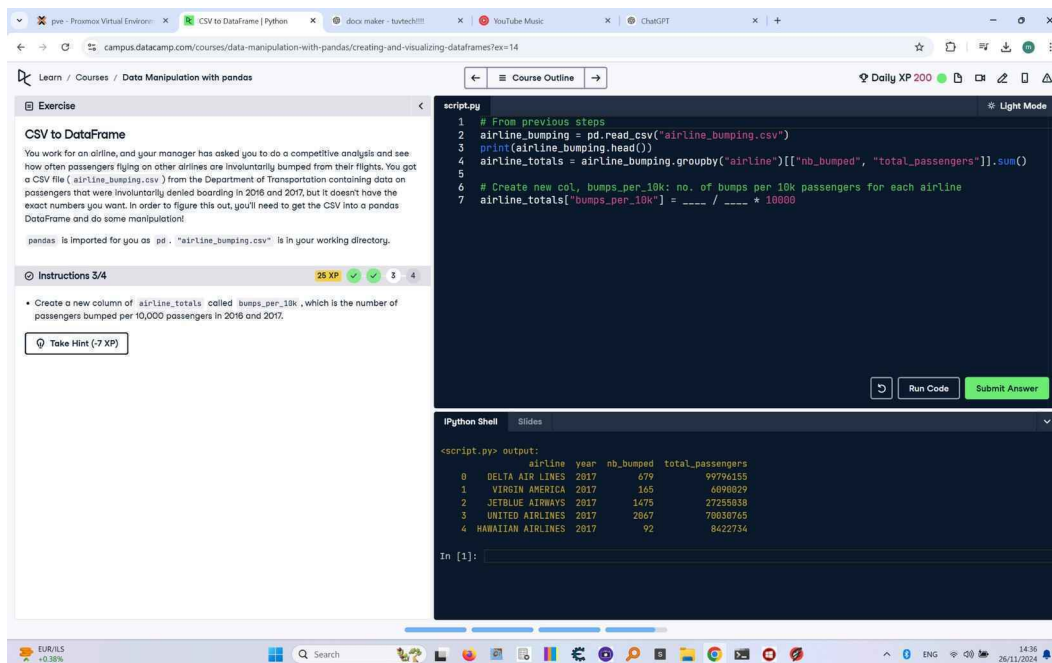
CSV to DataFrame (Add Bumps per 10k Passengers)

In this step, you'll calculate the number of passengers bumped per 10,000 passengers for each airline. This will provide a standardized metric for comparison between airlines.

Instructions:

1. Create a new column in `airline_totals` called `bumps_per_10k`.
2. Calculate the number of passengers bumped per 10,000 passengers for each airline.

Original Uploaded Image:



The screenshot shows a web browser window displaying a DataCamp exercise titled "CSV to DataFrame". The exercise instructions are on the left, and the code editor and terminal are on the right.

Exercise Instructions:

CSV to DataFrame

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!

pandas is imported for you as `pd`. `"airline_bumping.csv"` is in your working directory.

Instructions 3/4 25 XP

- Create a new column of `airline_totals` called `bumps_per_10k`, which is the number of passengers bumped per 10,000 passengers in 2016 and 2017.

[Take Hint \(-7 XP\)](#)

Code Editor:

```
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
6 # Create new col, bumps_per_10k: no. of bumps per 10k passengers for each airline
7 airline_totals["bumps_per_10k"] = ____ / ____ * 10000
```

Terminal Output:

```
<script.py> output:
  airline  year  nb_bumped  total_passengers
0  DELTA AIR LINES  2017      679          99796155
1  VERGIN AMERICA  2017      165           6696829
2  JETBLUE AIRWAYS  2017     1475          27255838
3  UNITED AIRLINES  2017     2867          70830765
4  HAWAIIAN AIRLINES  2017      92          8422734

In [1]:
```

Python Code Implementation:

```
# From previous steps
```

```
airline_bumping = pd.read_csv("airline_bumping.csv")
print(airline_bumping.head())
```

```
# Group by airline and calculate totals
```

```
airline_totals = airline_bumping.groupby("airline")[["nb_bumped",
"total_passengers"]].sum()
```

```
# Create new column, bumps_per_10k: no. of bumps per 10k passengers for
each airline
```

```
airline_totals["bumps_per_10k"] = airline_totals["nb_bumped"] /  
airline_totals["total_passengers"] * 10000
```

```
# Print the updated DataFrame  
print(airline_totals)
```

Explanation of Code:

1. **Group by airline and calculate totals**: Use `groupby("airline")` and `sum()` to aggregate the total `nb_bumped` and `total_passengers` for each airline.
2. **Create a new column**: Add a column `bumps_per_10k` to the `airline_totals` DataFrame. The formula calculates the number of passengers bumped per 10,000 passengers by dividing `nb_bumped` by `total_passengers` and multiplying by 10,000.
3. **Print the updated DataFrame**: Use `print()` to display the updated DataFrame with the new column.