# Creating a Probability Distribution - Step 3/4

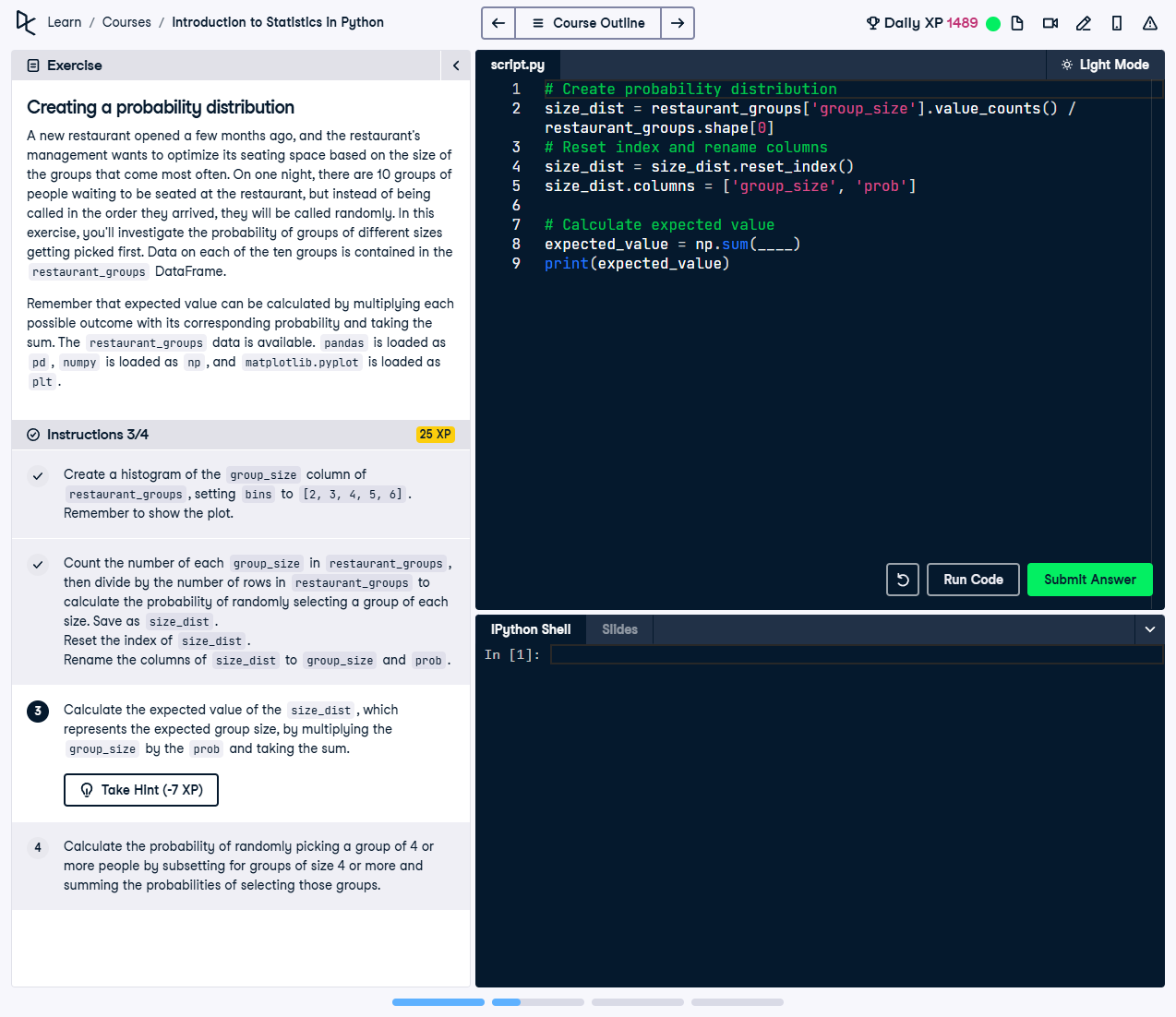


Figure: Screenshot showing the calculation of expected value for group sizes in the restaurant data.

## Question

In step 3/4, you are tasked to calculate the expected value of the `size\_dist` dataset. This expected value represents the average group size based on the given probabilities of each group size.

\*\*Instructions for Step 3/4:\*\*  
1. Divide the frequency of each group size by the total number of rows in the dataset to get the probability distribution.  
2. Reset the index of the resulting probability table and rename its columns for clarity.  
3. Use the `np.sum()` function to calculate the weighted average of group sizes based on their probabilities and print the result.

## Correct Code Solution

# Check if you have calculated the probability distribution  
size\_dist = restaurant\_groups['group\_size'].value\_counts() / restaurant\_groups.shape[0]  
  
# Ensure you reset the index and rename the columns  
size\_dist = size\_dist.reset\_index()  
size\_dist.columns = ['group\_size', 'prob']  
  
# Use this to calculate the expected value  
expected\_value = np.sum(size\_dist['group\_size'] \* size\_dist['prob'])  
print(expected\_value)

### Explanation

1. \*\*Probability Distribution:\*\* The `value\_counts()` method is used to count each group size's occurrence, and dividing by the total number of rows provides the probability of each size.  
2. \*\*Formatting:\*\* Resetting the index and renaming columns ensures the resulting DataFrame is structured for clarity and usability in further calculations.  
3. \*\*Expected Value:\*\* Using the weighted average formula (sum of each group size multiplied by its probability), `np.sum()` computes the average group size considering the given probabilities.  
4. \*\*Verification:\*\* The printed result offers a clear insight into the likely average group size based on historical data.