# Creating a Probability Distribution

## Question

# Create probability distribution  
size\_dist = restaurant\_groups['group\_size'].value\_counts() / restaurant\_groups.shape[0]  
# Reset index and rename columns  
size\_dist = size\_dist.reset\_index()  
size\_dist.columns = ['group\_size', 'prob']  
  
# Expected value  
expected\_value = np.sum(size\_dist['group\_size'] \* size\_dist['prob'])  
  
# Subset groups of size 4 or more  
groups\_4\_or\_more = \_\_\_\_  
  
# Sum the probabilities of groups\_4\_or\_more  
prob\_4\_or\_more = \_\_\_\_  
print(prob\_4\_or\_more)

## Explanation of the Question

This question focuses on creating a probability distribution for restaurant group sizes,  
calculating expected values using group size probabilities, and determining the probability  
of groups with sizes 4 or more using Python and pandas.

## Full Answer

# Create probability distribution  
size\_dist = restaurant\_groups['group\_size'].value\_counts() / restaurant\_groups.shape[0]  
  
# Reset index and rename columns  
size\_dist = size\_dist.reset\_index()  
size\_dist.columns = ['group\_size', 'prob']  
  
# Calculate the expected value of group sizes  
expected\_value = np.sum(size\_dist['group\_size'] \* size\_dist['prob'])  
  
# Subset for groups of size 4 or more  
groups\_4\_or\_more = size\_dist[size\_dist['group\_size'] >= 4]  
  
# Sum probabilities for groups of size 4 or more  
prob\_4\_or\_more = groups\_4\_or\_more['prob'].sum()  
  
# Print the probability  
print(prob\_4\_or\_more)

## Explanation of the Answer

The code calculates the probability distribution of group sizes and resets the index  
for easy reference. It computes the expected value by multiplying group sizes with their probabilities.  
Finally, it filters groups with size 4 or more and sums their probabilities.