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.