In [11]:

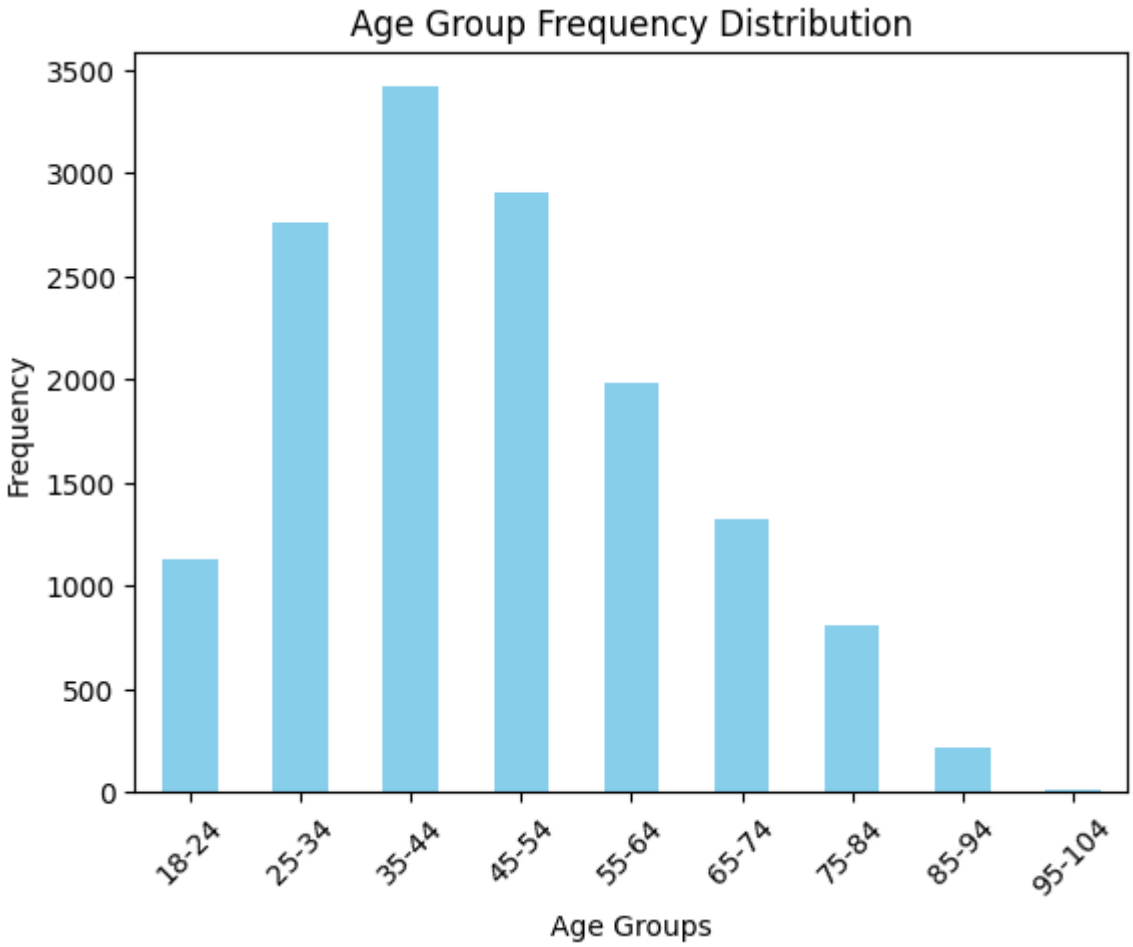
```
import pandas as pd
import matplotlib.pyplot as plt

# Let's have a look at the data frame again
print(df_cleaned)
```

```
      age_group  sex                wine_frequency             wine_amount  \
0         35-44    2  1 or 2 times in the last year    One glass/ container
1         85-94    2           2 to 3 times a month    One glass/ container
2         25-34    2  1 or 2 times in the last year    One glass/ container
3         65-74    2                    Once a week    One glass/ container
4         55-64    2  3 to 6 times in the last year    One glass/ container
...         ...  ...                            ...                     ...
14556     18-24    2  3 to 6 times in the last year    One glass/ container
14557     18-24    1  1 or 2 times in the last year    One glass/ container
14558     45-54    1           2 to 3 times a month    One glass/ container
14559     18-24    1  1 or 2 times in the last year  Two glasses/ containers
14560     18-24    2  1 or 2 times in the last year    One glass/ container

          income_category  noofwines
0      $50,000 to $59,999          1
1      $20,000 to $24,999          1
2      $60,000 to $69,999          1
3      $15,000 to $19,999          1
4      $40,000 to $49,999          1
...                   ...        ...
14556    Less than $5,000          1
14557    Less than $5,000          1
14558  $15,000 to $19,999          1
14559    Less than $5,000          2
14560    Less than $5,000          1

[14561 rows x 6 columns]
```

Generating the univariate graph to illustrate the distributions of week 2

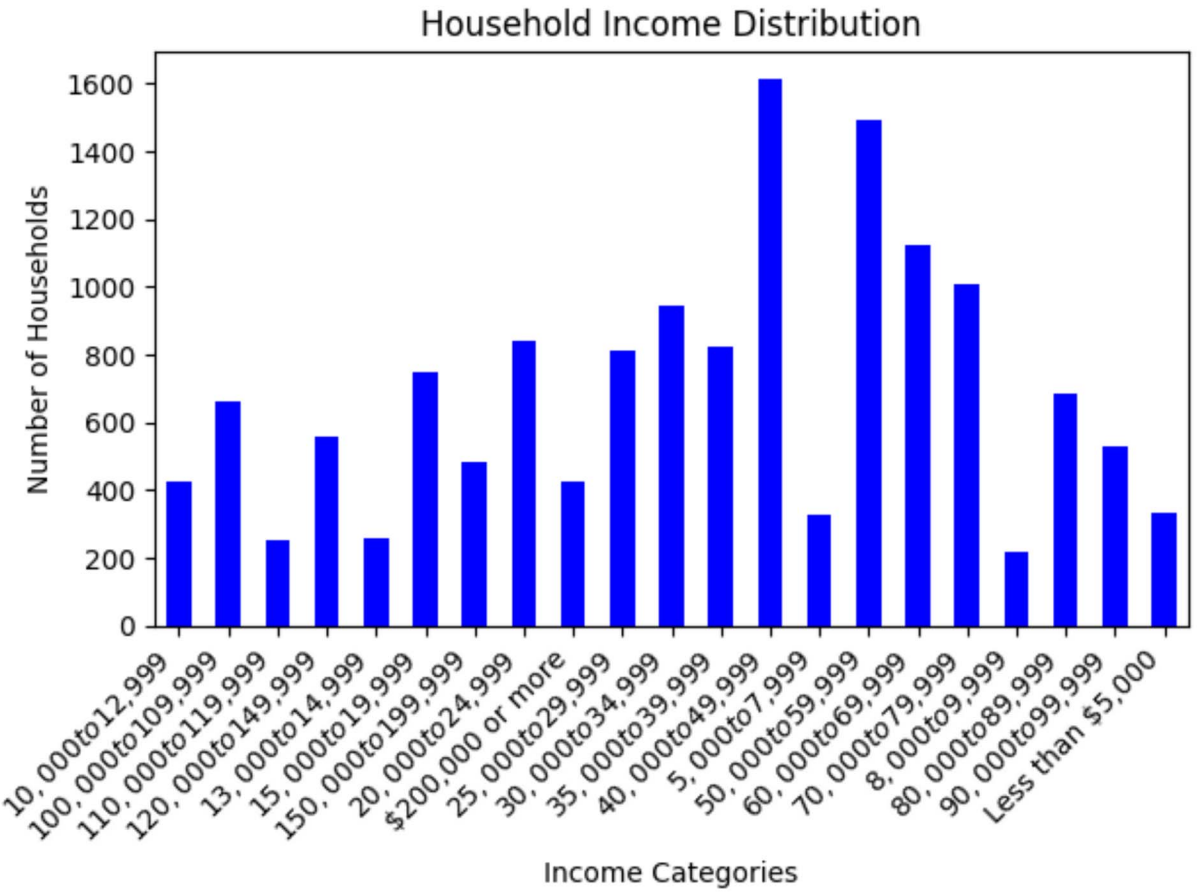
In [2]:

```
# Plotting the Age Group Frequency Distribution
labels = ['18-24', '25-34', '35-44', '45-54', '55-64', '65-74', '75-84', '85-94', '95
age_group = df_cleaned['age_group'].value_counts().reindex(labels, fill_value=0).sort

age_group.plot(kind='bar', color='skyblue')
plt.title('Age Group Frequency Distribution')
plt.xlabel('Age Groups')
plt.ylabel('Frequency')
plt.xticks(rotation=45)
plt.show()
```


Age Group Frequency Distribution
Frequency
3500
3000
2500
2000
1500
1000
500
0
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-94
95-104
Age Groups

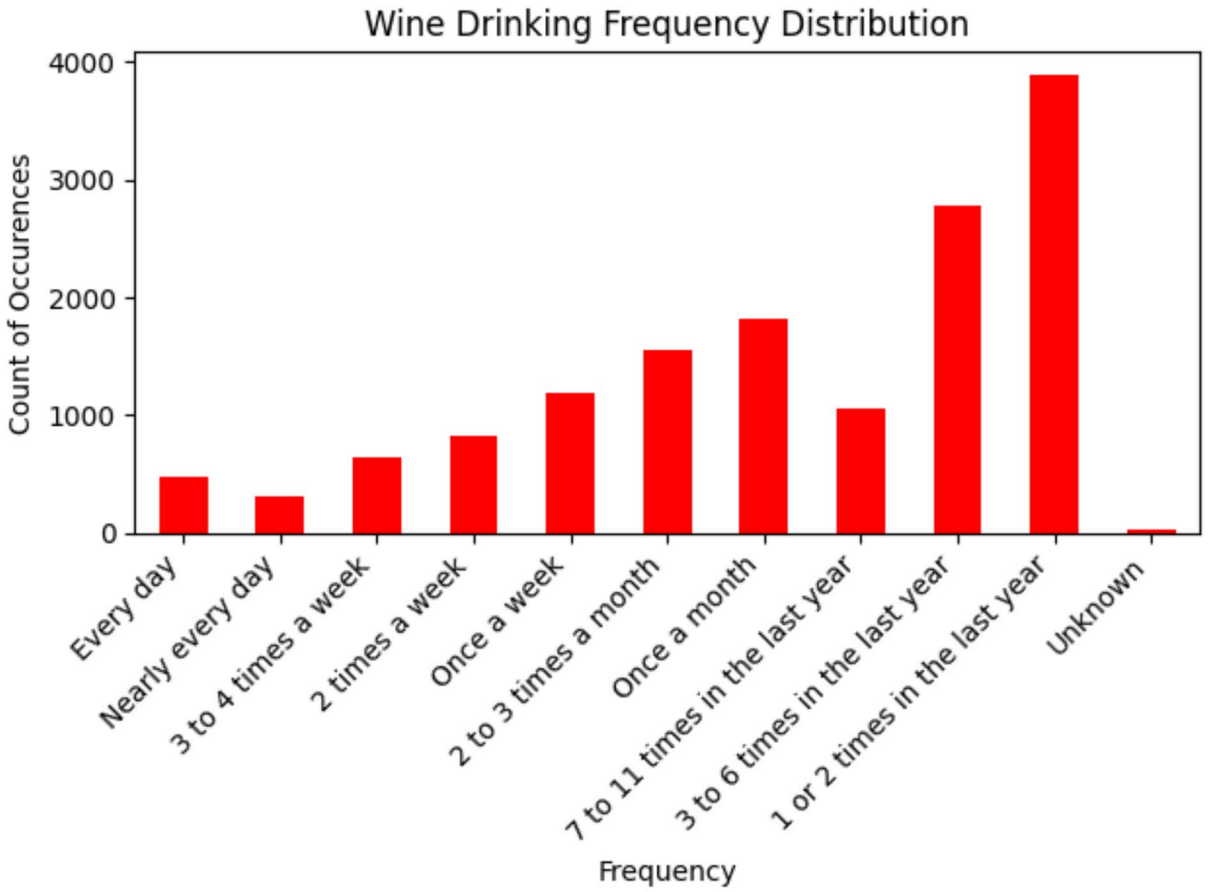
In [3]:

```
# Plotting the Income Category Frequency Distribution
income_labels = ['Less than $5,000',
    '$5,000 to $7,999',
    '$8,000 to $9,999',
    '$10,000 to $12,999',
    '$13,000 to $14,999',
    '$15,000 to $19,999',
    '$20,000 to $24,999',
    '$25,000 to $29,999',
    '$30,000 to $34,999',
    '$35,000 to $39,999',
    '$40,000 to $49,999',
    '$50,000 to $59,999',
    '$60,000 to $69,999',
    '$70,000 to $79,999',
    '$80,000 to $89,999',
    '$90,000 to $99,999',
    '$100,000 to $109,999',
    '$110,000 to $119,999',
    '$120,000 to $149,999',
    '$150,000 to $199,999',
    '$200,000 or more']

income_data = df_cleaned['income_category'].value_counts().reindex(income_labels, fi

income_data.plot(kind='bar', color='blue')
plt.title('Household Income Distribution')
plt.xlabel('Income Categories')
plt.ylabel('Number of Households')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()  # Adjust layout to make room for x-axis labels
plt.show()
```



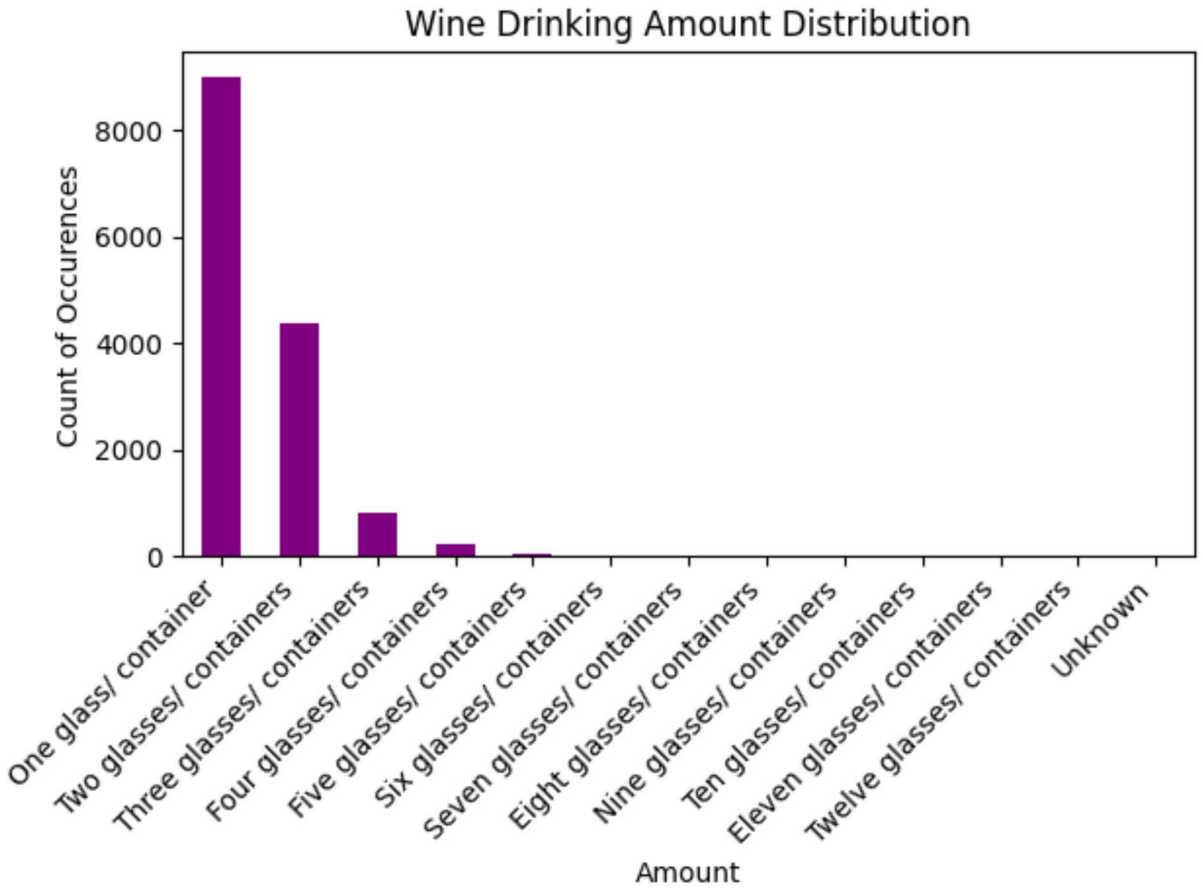
In [4]:

```
# Plotting the Wine Drinking Frequency Distribution
wine_freq_labels = ['Every day', 'Nearly every day',
                            '3 to 4 times a week', '2 times a week',
                            'Once a week', '2 to 3 times a month',
                            'Once a month', '7 to 11 times in the last year',
                            '3 to 6 times in the last year', '1 or 2 times in the last
                            'Unknown']

wine_freq_data = df_cleaned['wine_frequency'].value_counts().reindex(wine_freq_labels

wine_freq_data.plot(kind='bar', color='red')
plt.title('Wine Drinking Frequency Distribution')
plt.xlabel('Frequency')
plt.ylabel('Count of Occurences')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()  # Adjust layout to make room for x-axis labels
plt.show()
```



In [5]:

```
# Plotting the Wine Drinking Amount Distribution
wine_amount_labels = ['One glass/ container',
    'Two glasses/ containers',
    'Three glasses/ containers',
    'Four glasses/ containers',
    'Five glasses/ containers',
    'Six glasses/ containers',
    'Seven glasses/ containers',
    'Eight glasses/ containers',
    'Nine glasses/ containers',
    'Ten glasses/ containers',
    'Eleven glasses/ containers',
    'Twelve glasses/ containers',
    'Unknown']

wine_amount_data = df_cleaned['wine_amount'].value_counts().reindex(wine_amount_labe

wine_amount_data.plot(kind='bar', color='purple')
plt.title('Wine Drinking Amount Distribution')
plt.xlabel('Amount')
plt.ylabel('Count of Occurences')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()  # Adjust layout to make room for x-axis labels
plt.show()
```

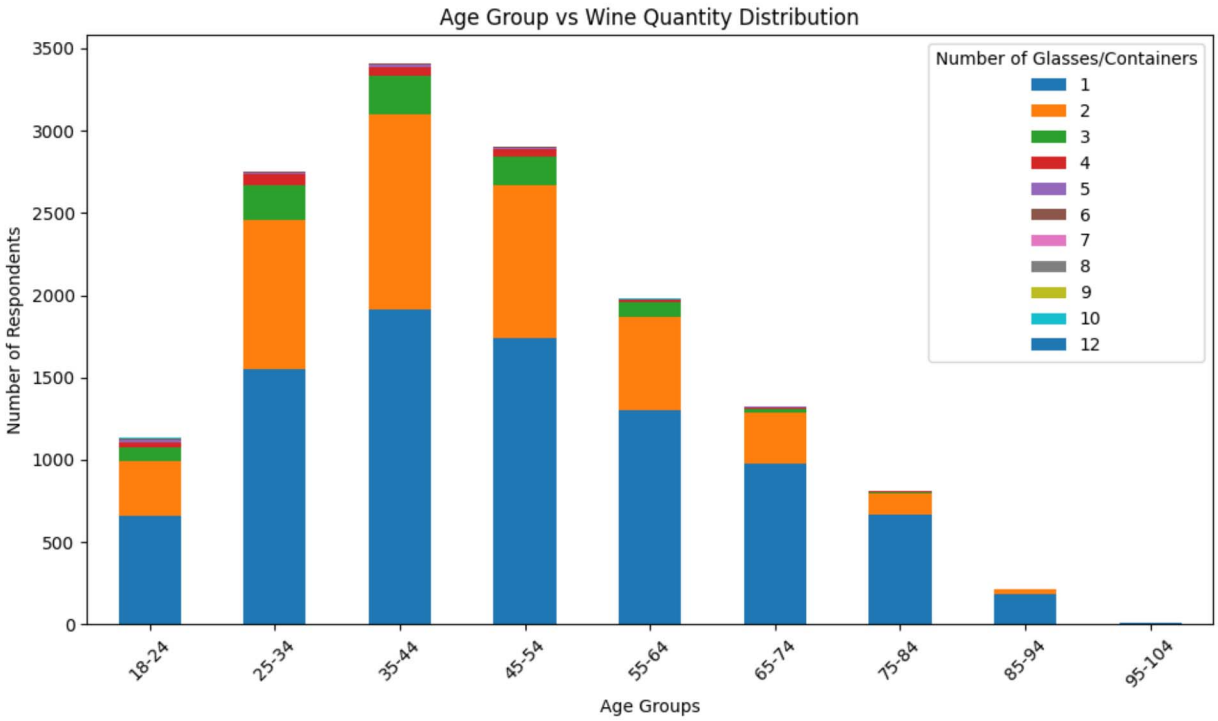


In [14]:

```
# Remove the 99 category (="Unknown") from the data, since it does not benefit the ar
df_cleaned = df_cleaned[~(df_cleaned['noofwines'] == 99)]

# Group by Age Group and Number of Glasses/Containers
age_wine_quantity_distribution = df_cleaned.groupby(['age_group', 'noofwines']).size(

print("Age and Wine Quantity Distribution:")
print(age_wine_quantity_distribution)

# Visualize the distribution
age_wine_quantity_distribution.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.title('Age Group vs Wine Quantity Distribution')
plt.xlabel('Age Groups')
plt.ylabel('Number of Respondents')
plt.xticks(rotation=45)
plt.legend(title='Number of Glasses/Containers')
plt.tight_layout()
plt.show()
```

```
Age and Wine Quantity Distribution:
noofwines    1     2    3   4   5   6   7   8   9   10  12
age_group
18-24       662   328   88  26  17   3   2   1   1   4   0
25-34      1552   906  210  66  11   7   1   0   1   1   0
35-44      1913  1190  231  55  13   4   1   2   0   0   1
45-54      1738   934  167  46  11   5   1   1   0   1   0
55-64      1298   572   87  14   4   1   0   1   0   1   0
65-74       979   306   27   7   2   0   1   0   0   0   0
75-84       669   124   10   5   1   0   0   0   0   0   0
```

```
85-94       187    24    1   0   0   0   0   0   0   0   0
95-104        6     2    0   0   0   0   0   0   0   0   0
```



In [ ]: