Assignment 5

In [1]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```python
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/agaricus-lep
column_names = ['class', 'cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor', '
                'gill-spacing', 'gill-size', 'gill-color', 'stalk-shape', 'stalk-root'
                'stalk-surface-below-ring', 'stalk-color-above-ring', 'stalk-color-bel
                'veil-color', 'ring-number', 'ring-type', 'spore-print-color', 'popula
data = pd.read_csv(url, names=column_names)
```

In [3]:

```python
print(data.head())
```

```
  class cap-shape cap-surface cap-color bruises odor gill-attachment  \
0     p         x           s         n       t    p               f   
1     e         x           s         y       t    a               f   
2     e         b           s         w       t    l               f   
3     p         x           y         w       t    p               f   
4     e         x           s         g       f    n               f   

  gill-spacing gill-size gill-color  ... stalk-surface-below-ring  \
0            c         n          k  ...                        s   
1            c         b          k  ...                        s   
2            c         b          n  ...                        s   
3            c         n          n  ...                        s   
4            w         b          k  ...                        s   

  stalk-color-above-ring stalk-color-below-ring veil-type veil-color  \
0                      w                      w         p          w   
1                      w                      w         p          w   
2                      w                      w         p          w   
3                      w                      w         p          w   
4                      w                      w         p          w   

  ring-number ring-type spore-print-color population habitat  
0           o         p                 k          s       u  
1           o         p                 n          n       g  
2           o         p                 n          n       m  
3           o         p                 k          s       u  
4           o         e                 n          a       g  

[5 rows x 23 columns]
```

In [4]:

```python
print(data.info())
```

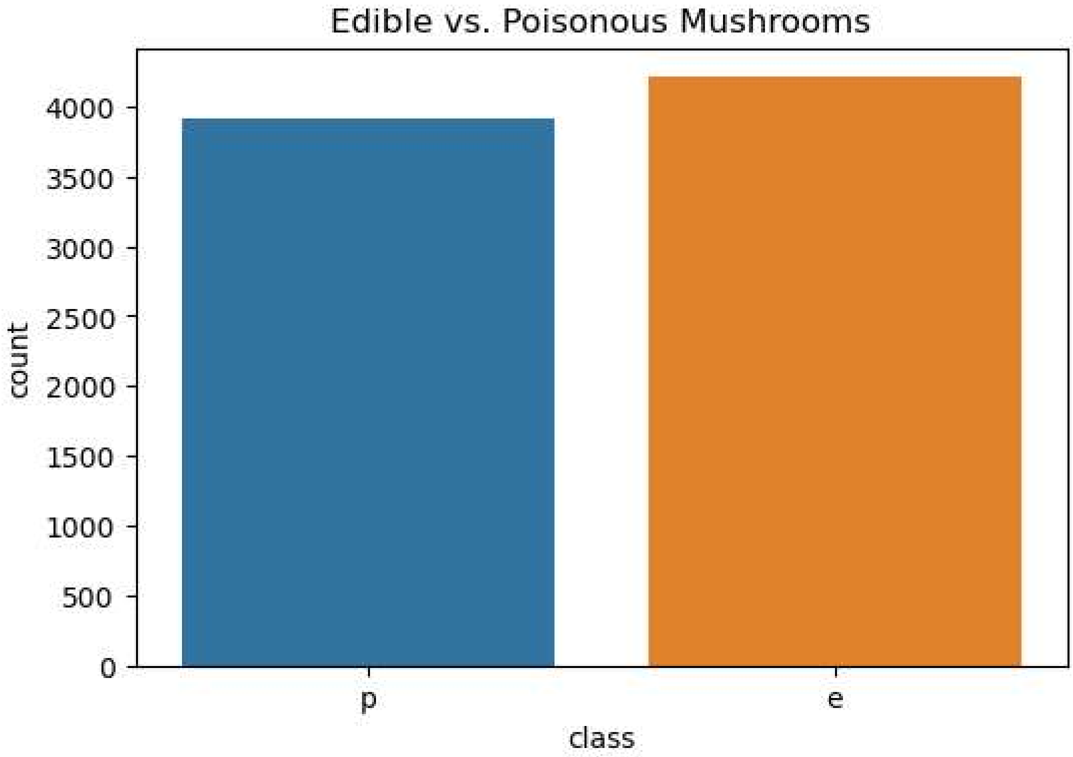
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
 14  stalk-color-above-ring    8124 non-null   object
 15  stalk-color-below-ring    8124 non-null   object
 16  veil-type                 8124 non-null   object
 17  veil-color                8124 non-null   object
 18  ring-number               8124 non-null   object
 19  ring-type                 8124 non-null   object
 20  spore-print-color         8124 non-null   object
 21  population                8124 non-null   object
 22  habitat                   8124 non-null   object
dtypes: object(23)
memory usage: 1.4+ MB
None
```

In [5]:

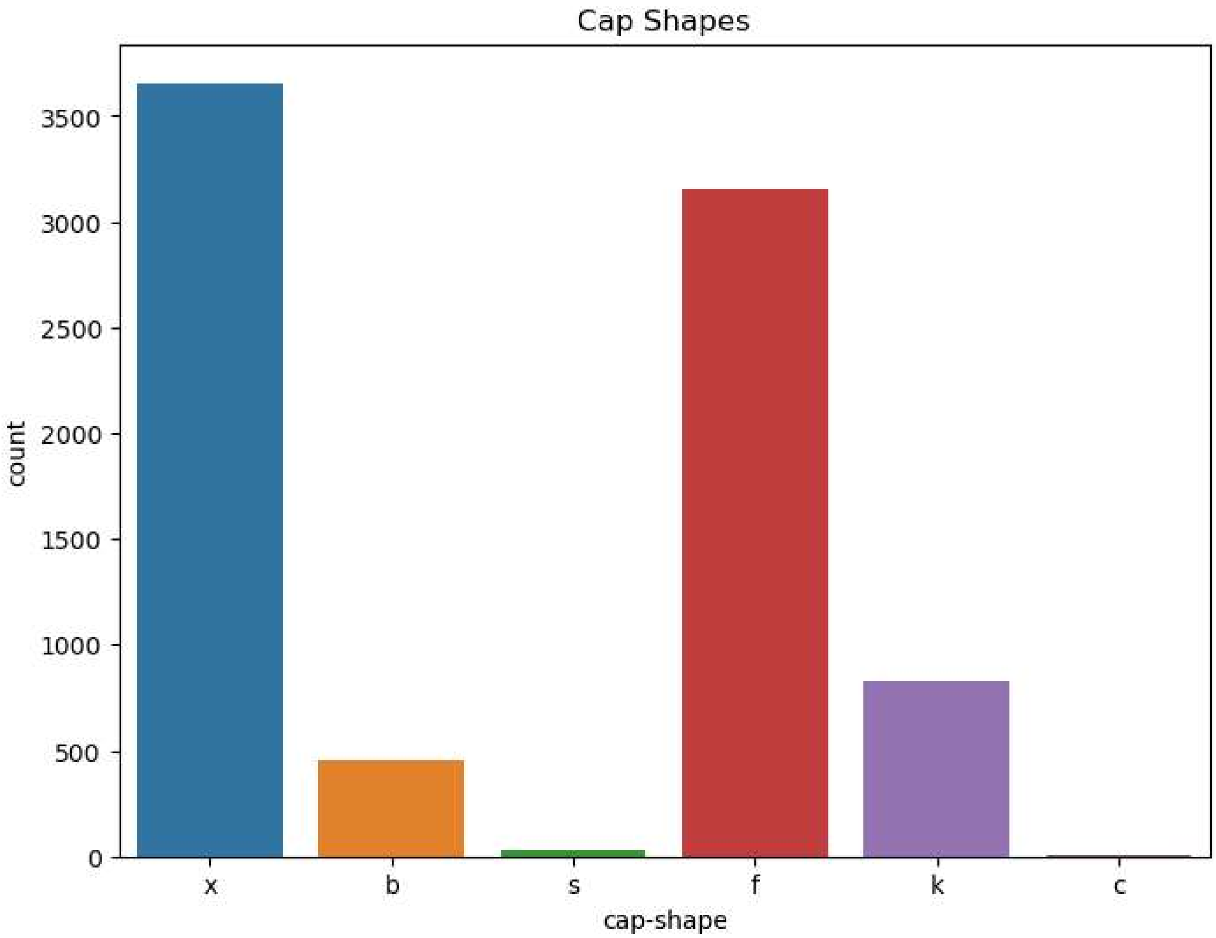
```
print(data.describe())
```

```
       class cap-shape cap-surface cap-color bruises  odor gill-attachment  \
count   8124      8124        8124      8124    8124  8124            8124   
unique     2         6           4        10       2     9               2   
top        e         x           y         n       f     n               f   
freq    4208      3656        3244      2284    4748  3528            7914   

       gill-spacing gill-size gill-color  ... stalk-surface-below-ring  \
count          8124      8124       8124  ...                     8124   
unique            2         2         12  ...                        4   
top               c         b          b  ...                        s   
freq           6812      5612       1728  ...                     4936   

       stalk-color-above-ring stalk-color-below-ring veil-type veil-color  \
count                    8124                   8124      8124       8124   
unique                      9                      9         1          4   
top                         w                      w         p          w   
freq                     4464                   4384      8124       7924   

       ring-number ring-type spore-print-color population habitat  
count         8124      8124              8124       8124    8124  
unique           3         5                 9          6       7  
top              o         p                 w          v       d  
freq          7488      3968              2388       4040    3148  

[4 rows x 23 columns]
```

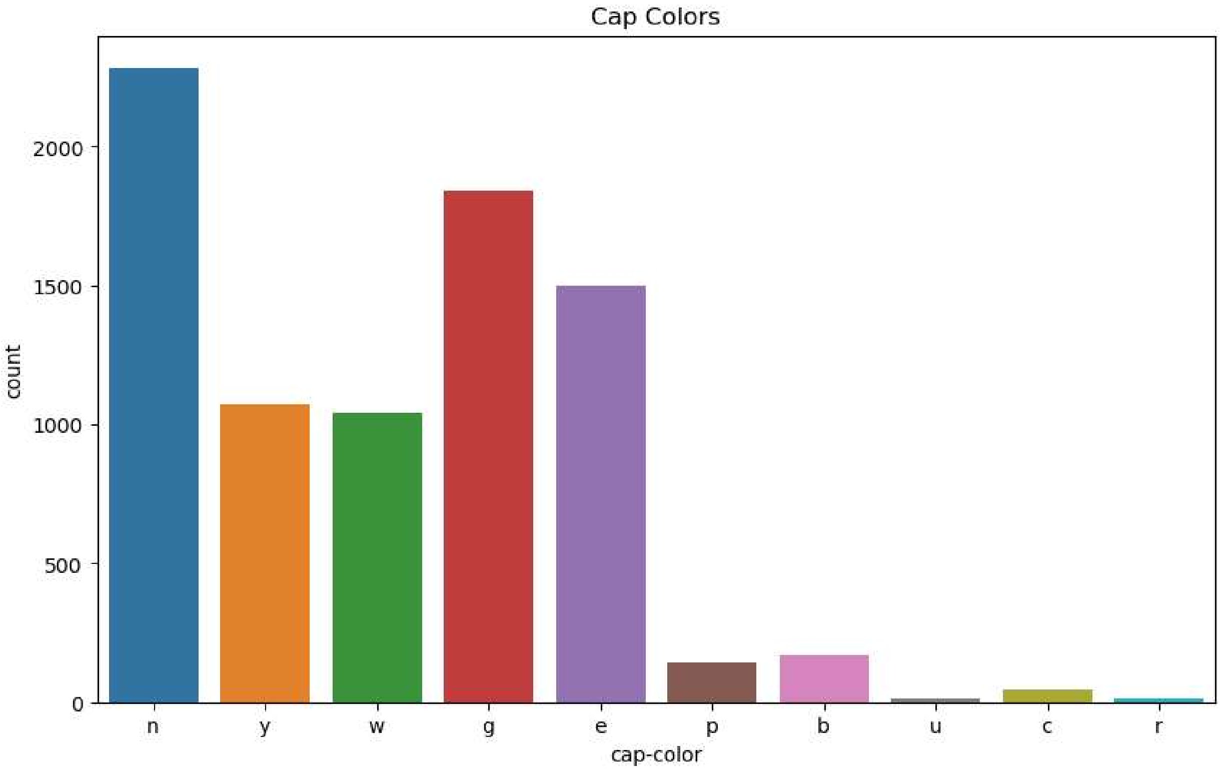
In [6]:
```
plt.figure(figsize=(6, 4))
sns.countplot(x='class', data=data)
plt.title('Edible vs. Poisonous Mushrooms')
plt.show()
```



In [7]:
```
plt.figure(figsize=(8, 6))
sns.countplot(x='cap-shape', data=data)
plt.title('Cap Shapes')
plt.show()
```

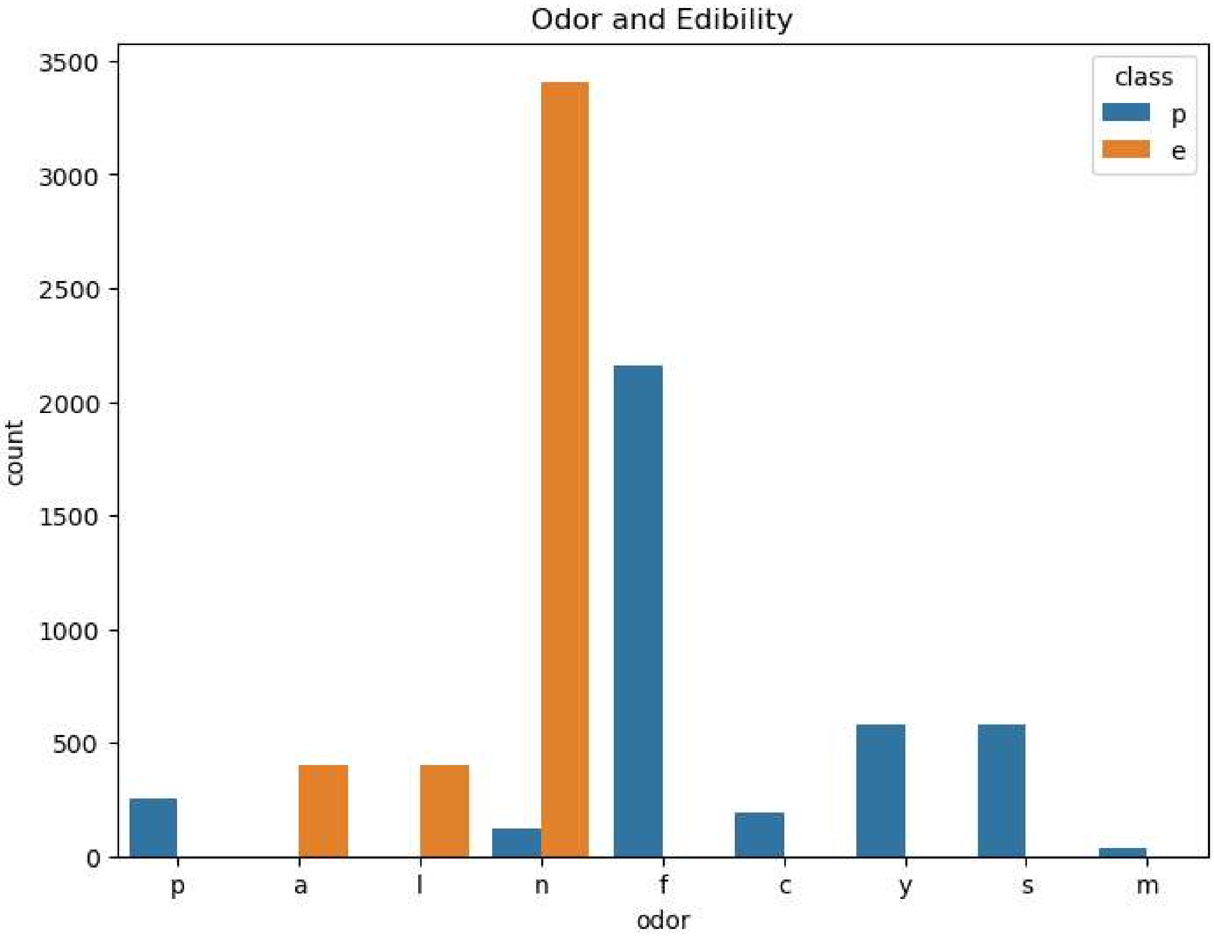


In [8]:
```
plt.figure(figsize=(10, 6))
sns.countplot(x='cap-color', data=data)
plt.title('Cap Colors')
plt.show()
```



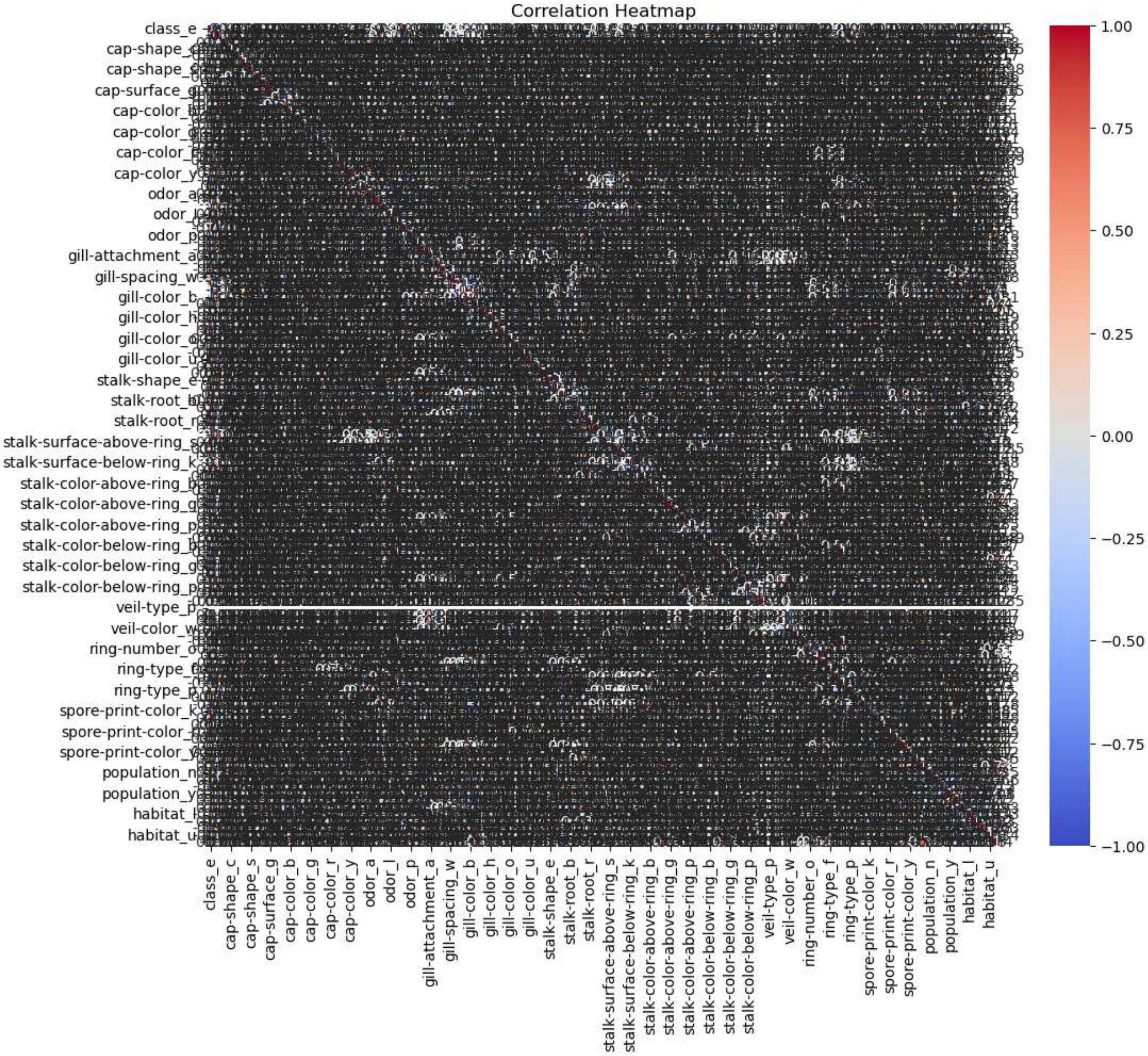
In [9]:

```
plt.figure(figsize=(8, 6))
sns.countplot(x='odor', hue='class', data=data)
plt.title('Odor and Edibility')
plt.show()
```



In [11]:

```
data = data.dropna()
data_encoded = pd.get_dummies(data)
plt.figure(figsize=(12, 10))
sns.heatmap(data_encoded.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```


Correlation Heatmap
class_e
cap-shape_c
cap-shape_s
cap-surface_g
cap-color_b
cap-color_g
cap-color_r
cap-color_y
odor_a
odor_l
odor_p
gill-attachment_a
gill-spacing_w
gill-color_b
gill-color_h
gill-color_o
gill-color_u
stalk-shape_e
stalk-root_b
stalk-root_r
stalk-surface-above-ring_s
stalk-surface-below-ring_k
stalk-color-above-ring_b
stalk-color-above-ring_g
stalk-color-above-ring_p
stalk-color-below-ring_b
stalk-color-below-ring_g
stalk-color-below-ring_p
veil-type_p
veil-color_w
ring-number_o
ring-type_f
ring-type_p
spore-print-color_k
spore-print-color_r
spore-print-color_y
population_n
population_y
habitat_l
habitat_u
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

```
In [ ]:
```