Assignment 5 - Renduchintala Navya - 2306AML112

Complete data visualization on Mushroom dataset using Matplotlib and Seaborn libraries

Dataset link:- https://archive.ics.uci.edu/dataset/73/mushroom

This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family (pp. 500-525). Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like ``leaflets three, let it be'' for Poisonous Oak and Ivy.

1. cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
2. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
3. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y
4. bruises?: bruises=t,no=f
5. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s
6. gill-attachment: attached=a,descending=d,free=f,notched=n
7. gill-spacing: close=c,crowded=w,distant=d
8. gill-size: broad=b,narrow=n
9. gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y
   A. stalk-shape: enlarging=e,tapering=t
   B. stalk-root: bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
   C. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
   D. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
   E. stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
   F. stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
   G. veil-type: partial=p,universal=u
   H. veil-color: brown=n,orange=o,white=w,yellow=y
   I. ring-number: none=n,one=o,two=t
   J. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z
   K. spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,white=w,yellow=y
   L. population: abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitary=y
   M. habitat: grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d

In [25]:

```
import pandas as pd
import numpy as np

# Load the dataset from the provided URL
features = ["poisonous","cap-shape","cap-surface","cap-color","bruises","odor","gill-attachment","gill-spacing"

url = 'agaricus-lepiota.data'
df = pd.read_csv(url, names=features)
df
```

Out[25]:

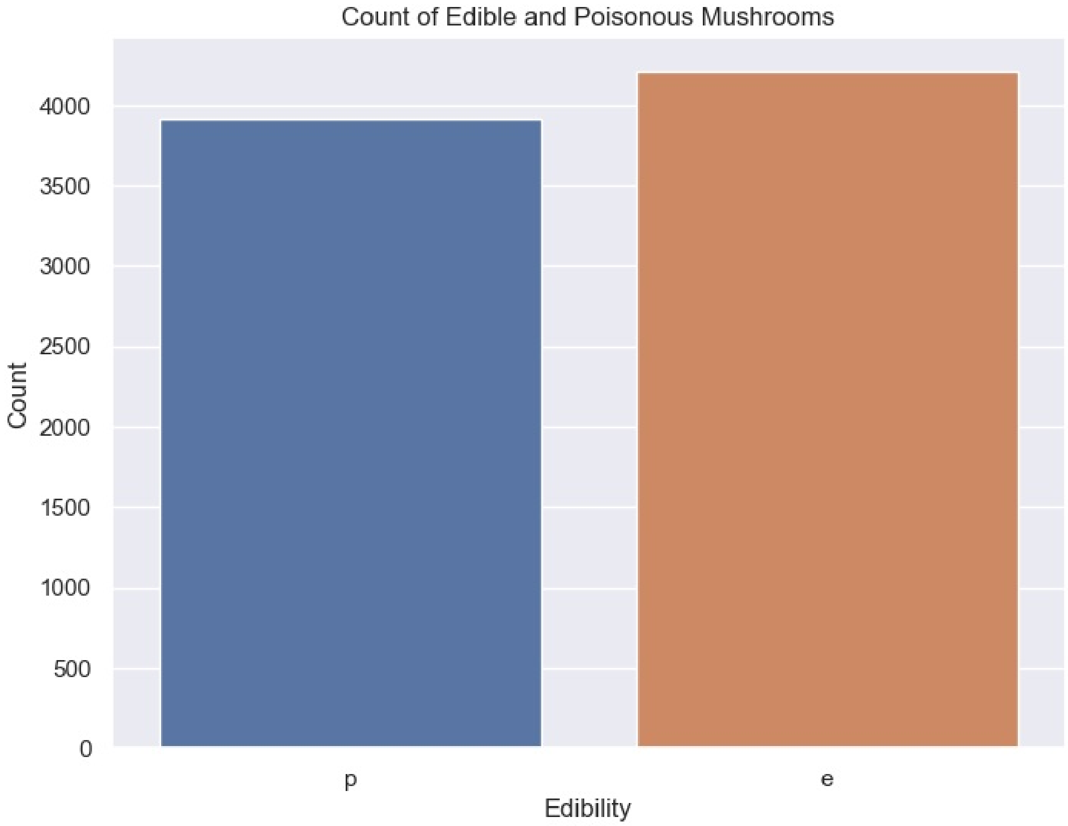
	poisonous	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	o
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	o
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	o
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	o
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	o
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8119	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o
8120	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	n	o
8121	e	f	s	n	f	n	a	c	b	n	...	s	o	o	p	o	o
8122	p	k	y	n	f	y	f	c	n	b	...	k	w	w	p	w	o
8123	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o

8124 rows × 23 columns

In [26]:

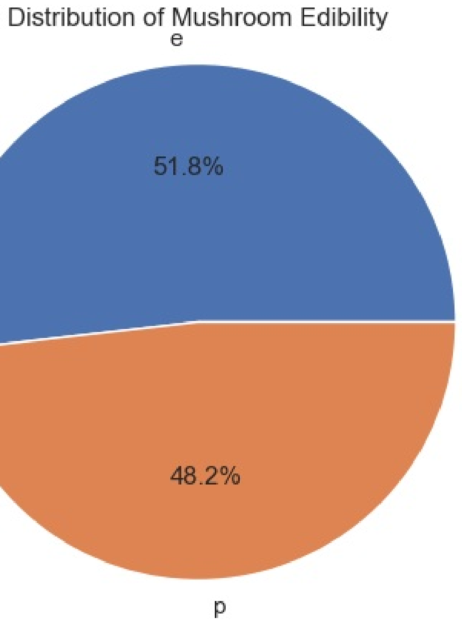
```
import matplotlib.pyplot as plt
```

```
import seaborn as sns

sns.set(style="darkgrid")
plt.figure(figsize=(8, 6))
sns.countplot(x='poisonous', data=df)
plt.xlabel('Edibility')
plt.ylabel('Count')
plt.title('Count of Edible and Poisonous Mushrooms')
plt.show()
```



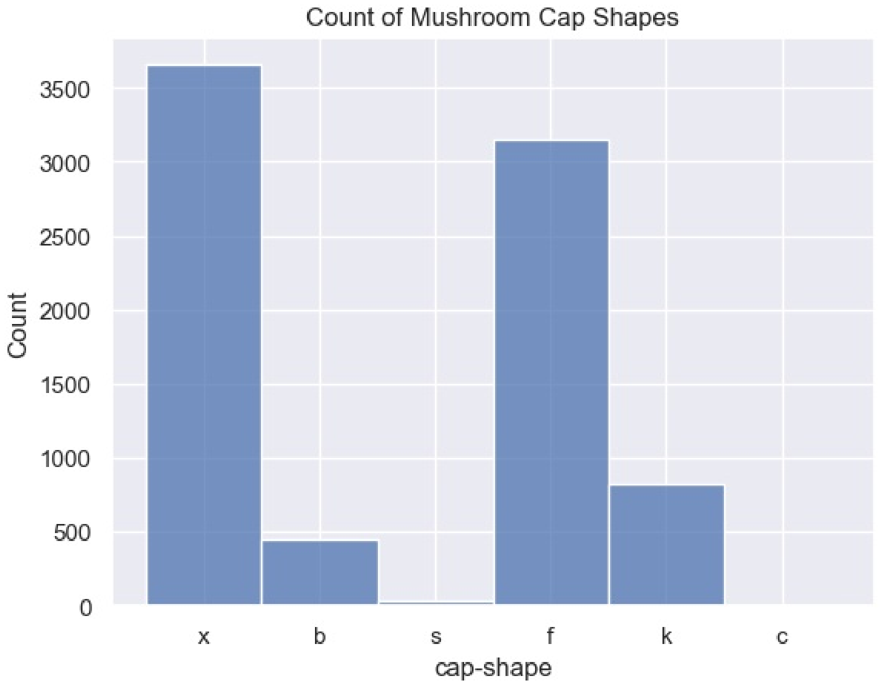
In [27]:

```
edibility_counts = df['poisonous'].value_counts()
plt.pie(edibility_counts, labels=edibility_counts.index, autopct='%1.1f%%')
plt.title('Distribution of Mushroom Edibility')
plt.axis('equal')
plt.show()
```



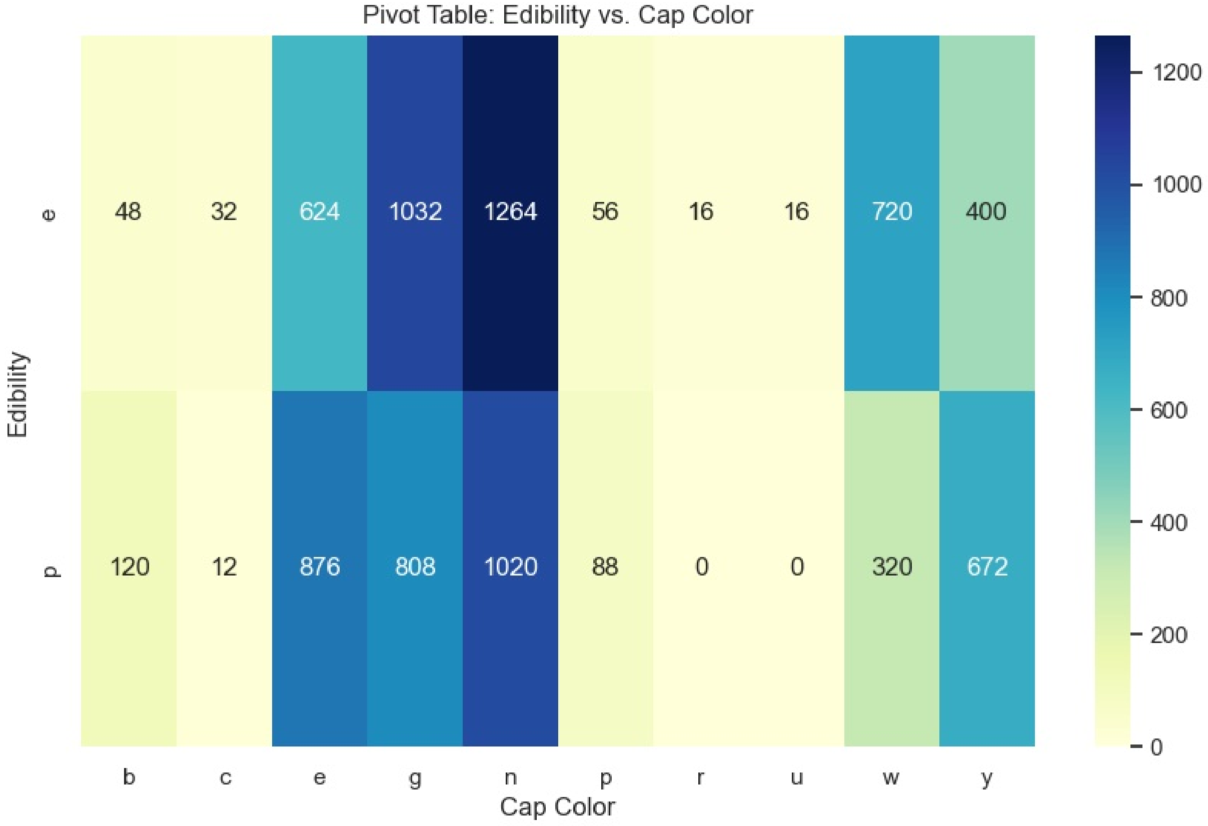
In [28]:

```
sns.histplot(x='cap-shape', data=df)
plt.title('Count of Mushroom Cap Shapes')
plt.show()
```



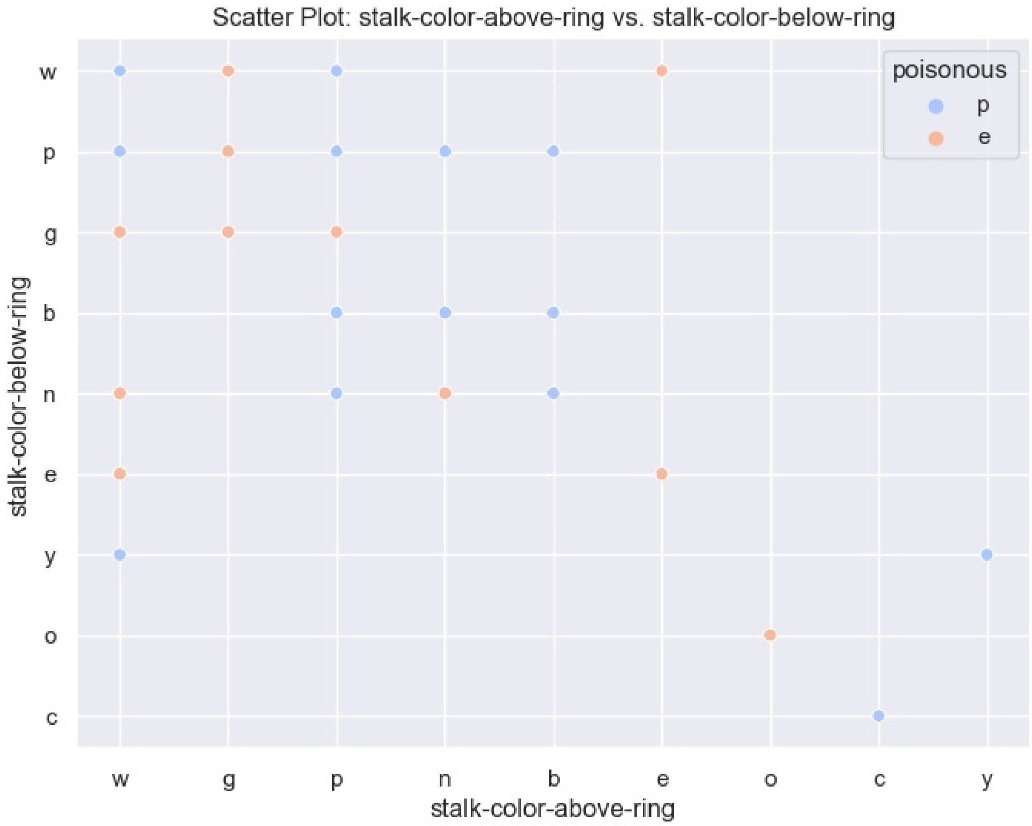
In [50]:

```
# Create a pivot table to aggregate the counts
pivot_table = df.pivot_table(index='poisonous', columns='cap-color', aggfunc='size', fill_value=0)

# Plot the pivot table using a heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(pivot_table, cmap='YlGnBu', annot=True, fmt='d')
plt.title("Pivot Table: Edibility vs. Cap Color")
plt.xlabel("Cap Color")
plt.ylabel("Edibility")
plt.show()
```



In [55]:

```
feature1 = 'stalk-color-above-ring'
feature2 = 'stalk-color-below-ring'

plt.figure(figsize=(8, 6))
sns.scatterplot(data=df, x='stalk-color-above-ring', y='stalk-color-below-ring', hue='poisonous', palette='cool
plt.title("Scatter Plot: {} vs. {}".format(feature1, feature2))
plt.xlabel(feature1)
plt.ylabel(feature2)
plt.show()
```



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js