assignment-5-1

June 30, 2023

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,
 4. bruises?:                 bruises=t,no=f
 5. odor:                     almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,punge
 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=
10. stalk-shape:              enlarging=e,tapering=t
11. stalk-root:               bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
14. stalk-color-above-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,
15. stalk-color-below-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,
16. veil-type:                partial=p,universal=u
17. veil-color:               brown=n,orange=o,white=w,yellow=y
18. ring-number:              none=n,one=o,two=t
19. ring-type:                cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,shea
20. spore-print-color:        black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,whi
21. population:               abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitar
22. habitat:                  grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d
```

[1]:

```
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

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

[1]:

```
     poisonous 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
…        …       …         …       …     …  …             …
8119         e         k           s         n       f    n               a
8120         e         x           s         n       f    n               a
8121         e         f           s         n       f    n               a
8122         p         k           y         n       f    y               f
8123         e         x           s         n       f    n               a

     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
…           …       …        …  …                      …
8119            c         b          y  …                        s
8120            c         b          y  …                        s
8121            c         b          n  …                        s
8122            c         n          b  …                        k
8123            c         b          y  …                        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
…                     …                    …       …        …
8119                      o                      o         p          o
8120                      o                      o         p          n
8121                      o                      o         p          o
8122                      w                      w         p          w
8123                      o                      o         p          o

     ring-number ring-type spore-print-color population habitat
```

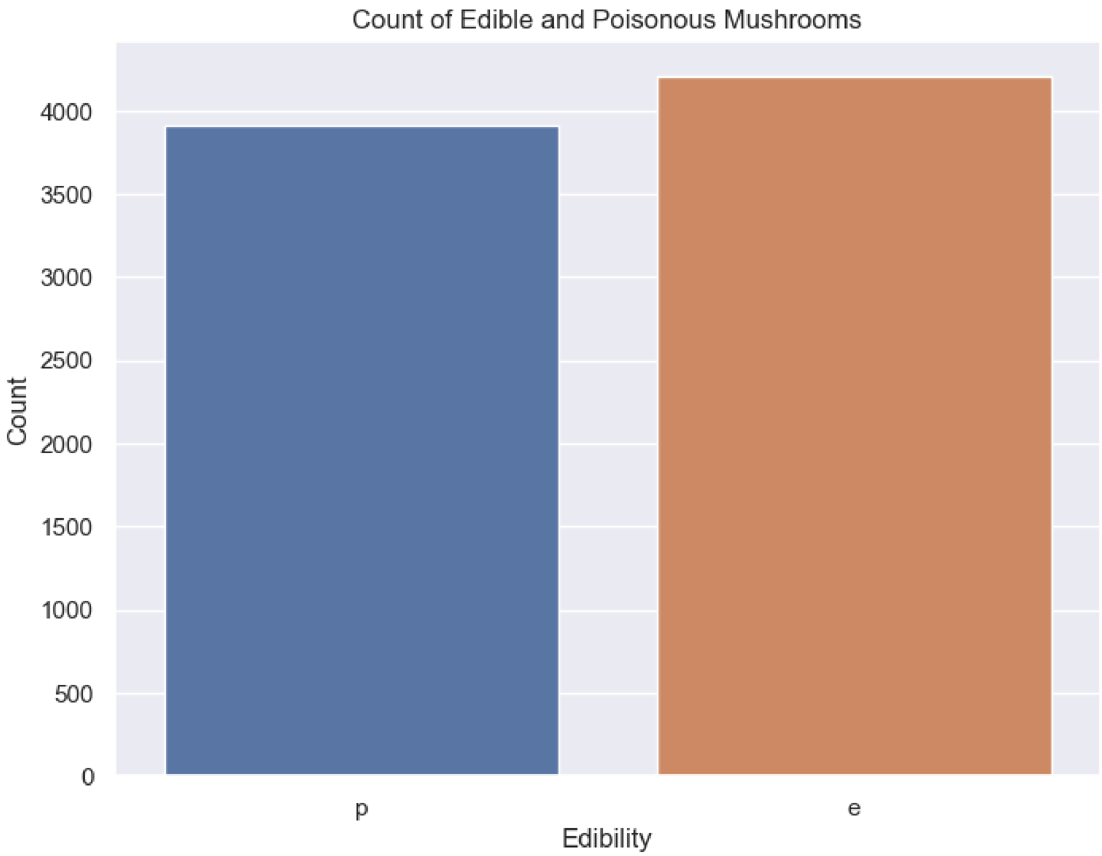
```
0              o         p                 k          s       u
1              o         p                 n          n       g
2              o         p                 n          n       m
3              o         p                 k          s       u
4              o         e                 n          a       g
…          …       …                …       …      …
8119           o         p                 b          c       l
8120           o         p                 b          v       l
8121           o         p                 b          c       l
8122           o         e                 w          v       l
8123           o         p                 o          c       l

[8124 rows x 23 columns]
```

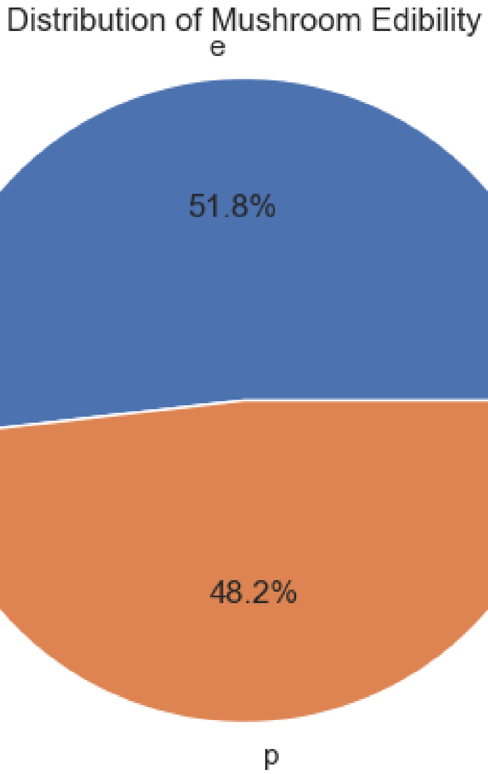
[3]:

```
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()
```



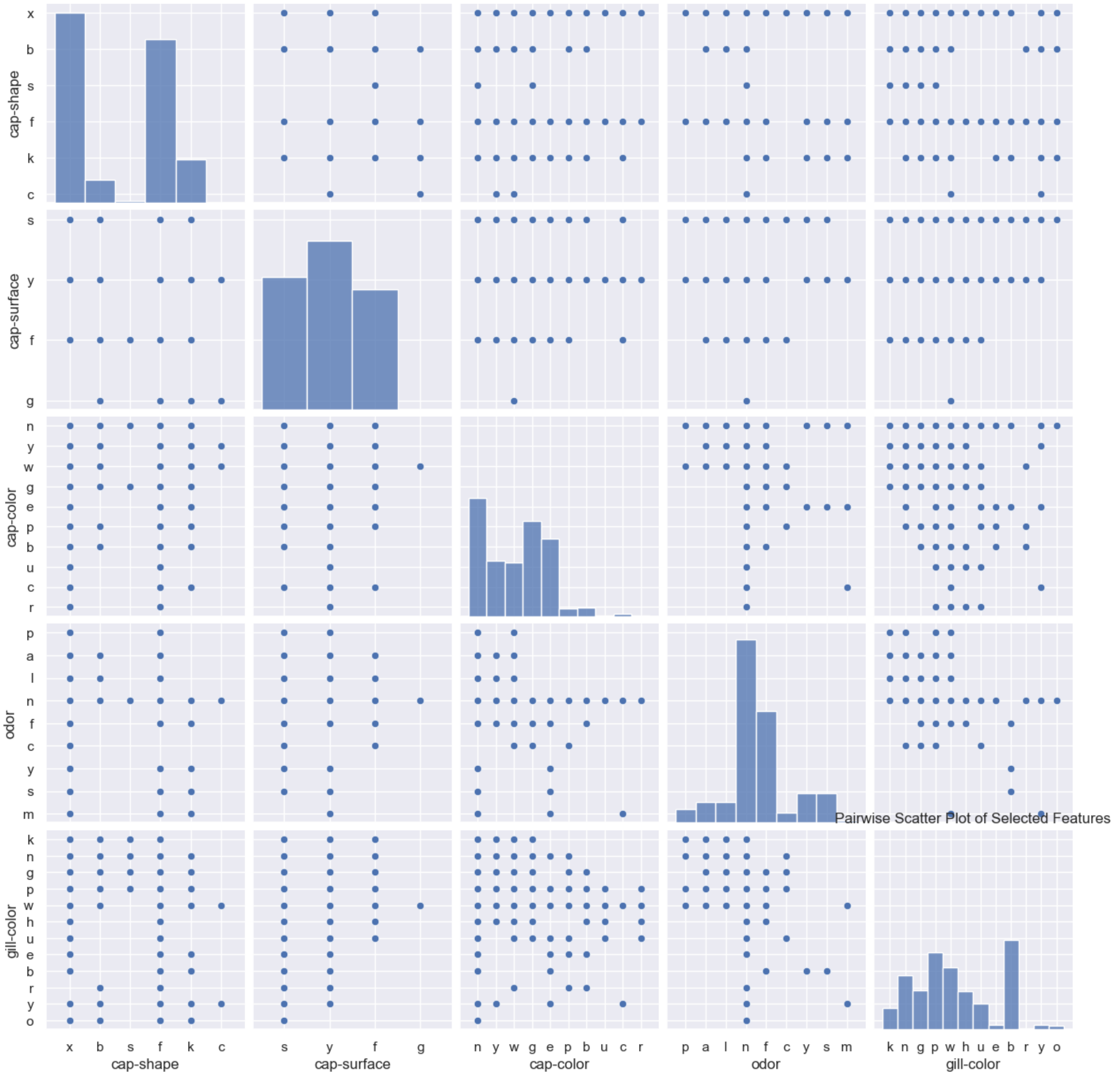
[11]:
```
edibility_counts = df['poisonous'].value_counts()
plt.pie(edibility_counts, labels=edibility_counts.index, autopct='%1.1f%%')
plt.title('Distribution of Mushroom Edibility')
plt.axis('equal')
plt.show()
```



[12]:

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

[15]:
```
selected_features = ['cap-shape', 'cap-surface', 'cap-color', 'odor',␣
 ↪'gill-color']
sns.pairplot(data=df, vars=selected_features)
plt.title('Pairwise Scatter Plot of Selected Features')
plt.show()
```



Pairwise Scatter Plot of Selected Features