hp-05

July 3, 2023

[36]:

```
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

[22]:

```
features = ['type','cap-shape','cap-surface','cap-color','bruises?
 ↪','odor','gill-attachment','gill-spacing',
          ␣
 ↪'gill-size','gill-color','stalk-shape','stalk-root','stalk-surface-above-ring','stalk-surfa
          ␣
 ↪'stalk-color-above-ring','stalk-color-below-ring','veil-type','veil-color','ring-number','r
           'spore-print-color','population','habitat']

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

[22]:

```
     type 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]
```

[23]:

```
df.isnull().sum()
```

[23]:

```
type                        0
cap-shape                   0
cap-surface                 0
cap-color                   0
bruises?                    0
odor                        0
gill-attachment             0
gill-spacing                0
gill-size                   0
```

```
gill-color                  0
stalk-shape                 0
stalk-root                  0
stalk-surface-above-ring    0
stalk-surface-below-ring    0
stalk-color-above-ring      0
stalk-color-below-ring      0
veil-type                   0
veil-color                  0
ring-number                 0
ring-type                   0
spore-print-color           0
population                  0
habitat                     0
dtype: int64
```

[26]:
```
df.drop(columns = ['veil-type'], inplace = True)
```

[27]:
```
categorical = df.select_dtypes(include = "object").columns
for i in categorical:
    print(df[categorical][i].value_counts())
    print('________________________________________')
```

```
e    4208
p    3916
Name: type, dtype: int64
________________________________________
x    3656
f    3152
k     828
b     452
s      32
c       4
Name: cap-shape, dtype: int64
________________________________________
y    3244
s    2556
f    2320
g       4
Name: cap-surface, dtype: int64
________________________________________
n    2284
g    1840
e    1500
y    1072
w    1040
b     168
```

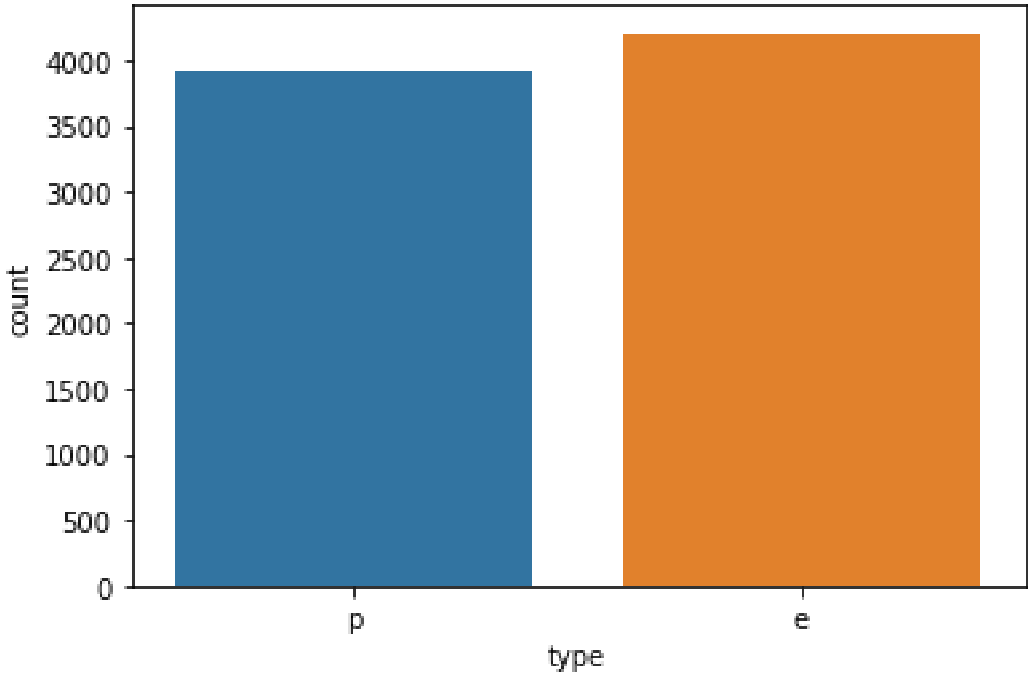
```
p     144
c      44
u      16
r      16
Name: cap-color, dtype: int64
----------------------------------------
f    4748
t    3376
Name: bruises?, dtype: int64
----------------------------------------
n    3528
f    2160
y     576
s     576
a     400
l     400
p     256
c     192
m      36
Name: odor, dtype: int64
----------------------------------------
f    7914
a     210
Name: gill-attachment, dtype: int64
----------------------------------------
c    6812
w    1312
Name: gill-spacing, dtype: int64
----------------------------------------
b    5612
n    2512
Name: gill-size, dtype: int64
----------------------------------------
b    1728
p    1492
w    1202
n    1048
g     752
h     732
u     492
k     408
e      96
y      86
o      64
r      24
Name: gill-color, dtype: int64
----------------------------------------
t    4608
```

```
e    3516
Name: stalk-shape, dtype: int64
----------------------------------------
b    3776
?    2480
e    1120
c     556
r     192
Name: stalk-root, dtype: int64
----------------------------------------
s    5176
k    2372
f     552
y      24
Name: stalk-surface-above-ring, dtype: int64
----------------------------------------
s    4936
k    2304
f     600
y     284
Name: stalk-surface-below-ring, dtype: int64
----------------------------------------
w    4464
p    1872
g     576
n     448
b     432
o     192
e      96
c      36
y       8
Name: stalk-color-above-ring, dtype: int64
----------------------------------------
w    4384
p    1872
g     576
n     512
b     432
o     192
e      96
c      36
y      24
Name: stalk-color-below-ring, dtype: int64
----------------------------------------
w    7924
n      96
o      96
y       8
```

```
Name: veil-color, dtype: int64
----------------------------------------
o    7488
t     600
n      36
Name: ring-number, dtype: int64
----------------------------------------
p    3968
e    2776
l    1296
f      48
n      36
Name: ring-type, dtype: int64
----------------------------------------
w    2388
n    1968
k    1872
h    1632
r      72
u      48
o      48
y      48
b      48
Name: spore-print-color, dtype: int64
----------------------------------------
v    4040
y    1712
s    1248
n     400
a     384
c     340
Name: population, dtype: int64
----------------------------------------
d    3148
g    2148
p    1144
l     832
u     368
m     292
w     192
Name: habitat, dtype: int64
----------------------------------------
```

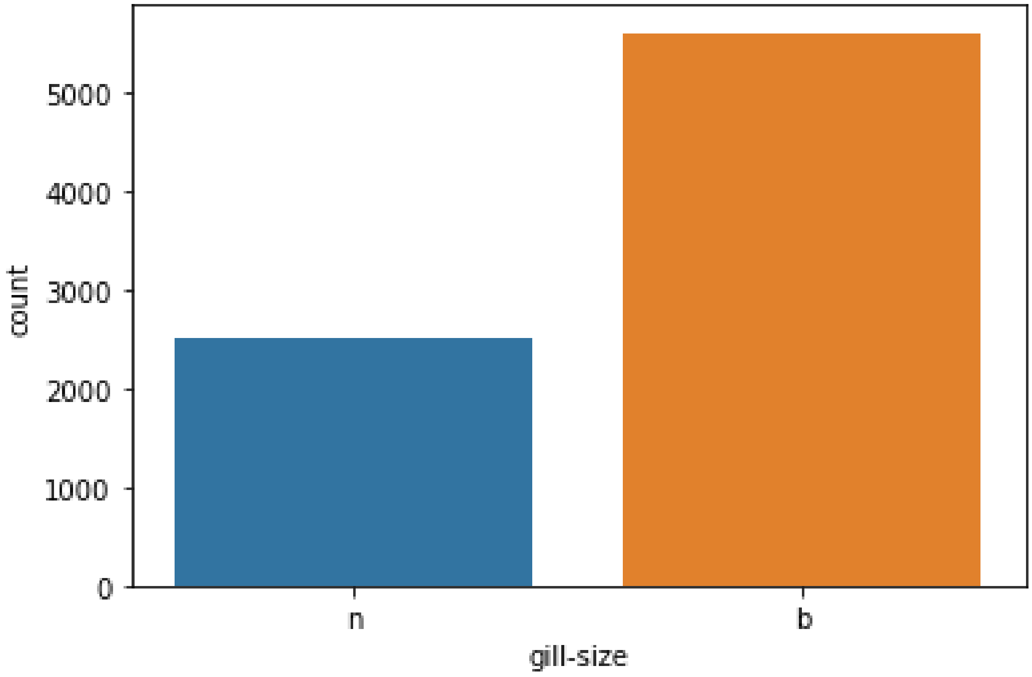
[5]:
```
sns.countplot(df['type'])
```

[5]:
```
<AxesSubplot:xlabel='type', ylabel='count'>
```



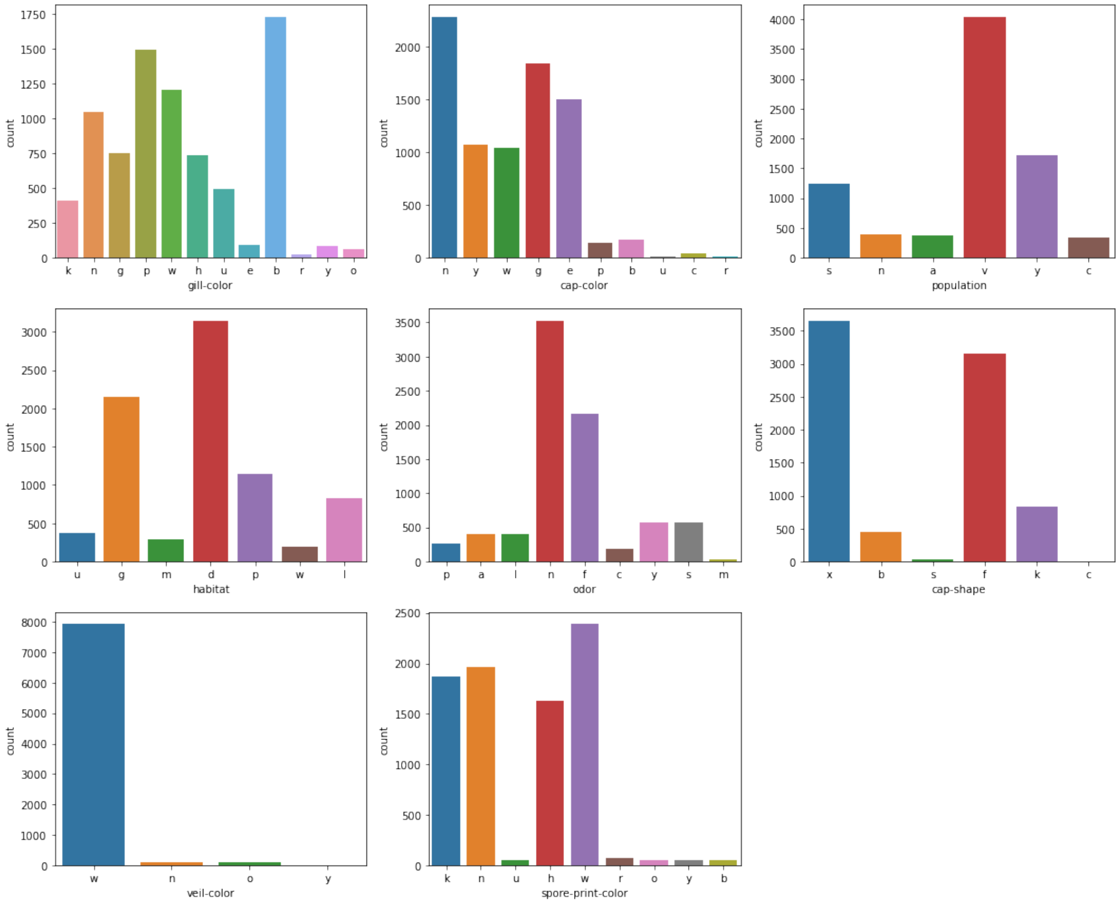
```
[29]: sns.countplot(df['gill-size'])
```

```
[29]: <AxesSubplot:xlabel='gill-size', ylabel='count'>
```



[37]:

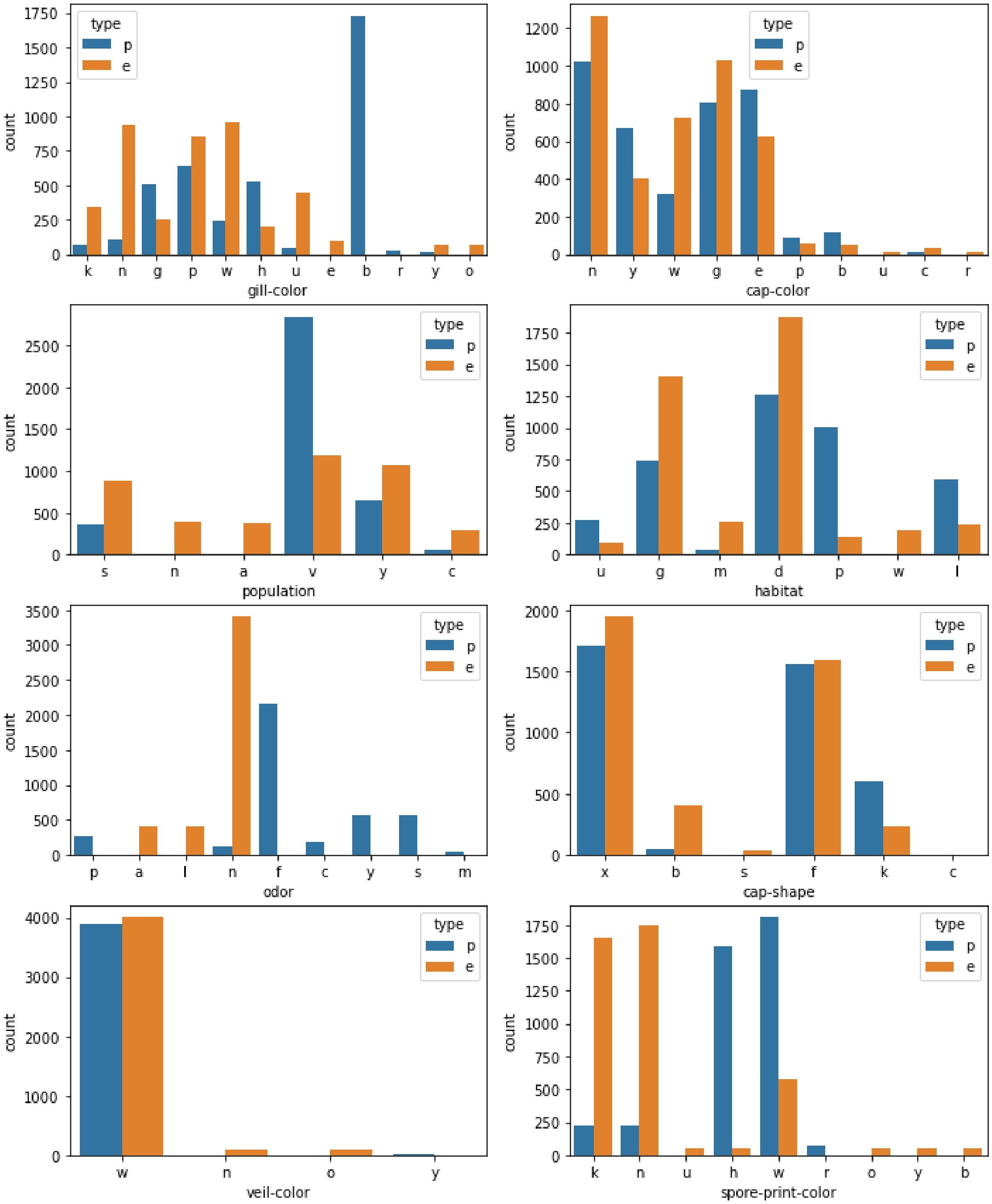
```
plt.figure(figsize=(18,15))

plt.subplot(3,3,1)
sns.countplot(df["gill-color"])

plt.subplot(3,3,2)
sns.countplot(df["cap-color"])

plt.subplot(3,3,3)
sns.countplot(df["population"])

plt.subplot(3,3,4)
sns.countplot(df["habitat"])

plt.subplot(3,3,5)
sns.countplot(df["odor"])

plt.subplot(3,3,6)
sns.countplot(df["cap-shape"])

plt.subplot(3,3,7)
sns.countplot(df["veil-color"])

plt.subplot(3,3,8)
sns.countplot(df["spore-print-color"])


plt.show()
```



[47]:

```
plt.figure(figsize=(18,15))

plt.subplot(4,3,1)
sns.countplot(data = df, x = 'gill-color', hue = 'type')

plt.subplot(4,3,2)
sns.countplot(data = df, x = 'cap-color', hue = 'type')

plt.subplot(4,3,4)
sns.countplot(data = df, x = 'population', hue = 'type')

plt.subplot(4,3,5)
sns.countplot(data = df, x = 'habitat', hue = 'type')

plt.subplot(4,3,7)
sns.countplot(data = df, x = 'odor', hue = 'type')

plt.subplot(4,3,8)
sns.countplot(data = df, x = 'cap-shape', hue = 'type')
```

```
plt.subplot(4,3,10)
sns.countplot(data = df, x = 'veil-color', hue = 'type')

plt.subplot(4,3,11)
sns.countplot(data = df, x = 'spore-print-color', hue = 'type')


plt.show()
```



- Buff gill color is highly poisonous and brown gill color mushrooms are most edible.
- several type mushrooms are poisonous.
- foul odour mushrooms are highly poisonous and no odour are most edible ones.
- white veil-color is least preferred as it contains both edible and poisonous types mushrooms in it.
- Black and brown spore-print-color mushrooms are edible, whereas chocolate and white are poisonous.