Data Elaboration

April 2, 2024

[ ]:
```
%pip install pandas
```

```
Requirement already satisfied: pandas in
c:\users\utente\appdata\local\programs\python\python310\lib\site-packages
(2.1.1)
Requirement already satisfied: numpy>=1.22.4 in
c:\users\utente\appdata\local\programs\python\python310\lib\site-packages (from
pandas) (1.24.3)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\utente\appdata\local\programs\python\python310\lib\site-packages (from
pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\utente\appdata\local\programs\python\python310\lib\site-packages (from
pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\utente\appdata\local\programs\python\python310\lib\site-packages (from
pandas) (2023.3)
Requirement already satisfied: six>=1.5 in
c:\users\utente\appdata\local\programs\python\python310\lib\site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
WARNING: Ignoring invalid distribution -yping-extensions
(c:\users\utente\appdata\local\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -yping-extensions
(c:\users\utente\appdata\local\programs\python\python310\lib\site-packages)

[notice] A new release of pip is available: 23.3.2 -> 24.0
[notice] To update, run: python.exe -m pip install --upgrade pip
```

1 Il fungaiolo và nel bosco...

[ ]:
```
import pandas as pd

df = pd.read_csv("mushroom.csv")
```

1.1 e muore avvelenato...

[ ]:
```
df
```

[ ]:
```
       class  cap-diameter cap-shape cap-surface cap-color  \
0          p         15.26         x           g         o
1          p         16.60         x           g         o
2          p         14.07         x           g         o
3          p         14.17         f           h         e
4          p         14.64         x           h         o
…      …            …        …          …       …
61064      p          1.18         s           s         y
61065      p          1.27         f           s         y
61066      p          1.27         s           s         y
61067      p          1.24         f           s         y
61068      p          1.17         s           s         y

       does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                         f               e          NaN          w
1                         f               e          NaN          w
2                         f               e          NaN          w
3                         f               e          NaN          w
4                         f               e          NaN          w
…                       …               …          …          …
61064                     f               f          f          f
61065                     f               f          f          f
61066                     f               f          f          f
61067                     f               f          f          f
61068                     f               f          f          f

        stem-height  …  stem-root stem-surface stem-color veil-type  \
0             16.95  …          s            y          w         u
1             17.99  …          s            y          w         u
2             17.80  …          s            y          w         u
3             15.77  …          s            y          w         u
4             16.53  …          s            y          w         u
…             …  …       …          …          …       …
61064          3.93  …        NaN          NaN          y       NaN
61065          3.18  …        NaN          NaN          y       NaN
61066          3.86  …        NaN          NaN          y       NaN
61067          3.56  …        NaN          NaN          y       NaN
61068          3.25  …        NaN          NaN          y       NaN

       veil-color has-ring ring-type spore-print-color habitat season
0               w        t         g               NaN       d      w
1               w        t         g               NaN       d      u
2               w        t         g               NaN       d      w
3               w        t         p               NaN       d      w
```

```
4              w        t         p               NaN       d      w
…          …      …        …              …      …    …
61064        NaN        f         f               NaN       d      a
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
61068        NaN        f         f               NaN       d      u

[61069 rows x 21 columns]
```

1.2 o forse no?

per salvare il nostro amico fungaiolo, nel 2024 abbiamo le tecnologie per decretare quale fungo è velenoso o meno…

[ ]:
```
percentage_missing = df.isnull().sum() * 100 / len(df)
print(percentage_missing)
```

```
class                     0.000000
cap-diameter              0.000000
cap-shape                 0.000000
cap-surface              23.121387
cap-color                 0.000000
does-bruise-or-bleed      0.000000
gill-attachment          16.184971
gill-spacing             41.040462
gill-color                0.000000
stem-height               0.000000
stem-width                0.000000
stem-root                84.393064
stem-surface             62.427746
stem-color                0.000000
veil-type                94.797688
veil-color               87.861272
has-ring                  0.000000
ring-type                 4.046243
spore-print-color        89.595376
habitat                   0.000000
season                    0.000000
dtype: float64
```

Quello che ci manca sono i dati, per precisione una in alcune colonne quasi il 90% dei dati, oltre che la "decriptazione di essi" perché chi ha raccolto i dati ha abbreviato gli stessi, e noi andremo a riallargarli in modo che il nostro dataset sia più semplice da analizzare a colpo d'occhio

[ ]:
```
df1 = df
```

è buona pratica far rimanere il dataframe originale

immutato

, in modo da poter poi recuperare i dati in caso di bisogno

[ ]:
```
df1['class'] = df['class'].replace({"p" : "poisonus", "e" : "edible"})
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26         x           g         o
1      poisonus         16.60         x           g         o
2      poisonus         14.07         x           g         o
3      poisonus         14.17         f           h         e
4      poisonus         14.64         x           h         o
…         …            …        …          …        …
61064  poisonus          1.18         s           s         y
61065  poisonus          1.27         f           s         y
61066  poisonus          1.27         s           s         y
61067  poisonus          1.24         f           s         y
61068  poisonus          1.17         s           s         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                        f               e          NaN          w
1                        f               e          NaN          w
2                        f               e          NaN          w
3                        f               e          NaN          w
4                        f               e          NaN          w
…                      …               …          …        …
61064                    f               f          f          f
61065                    f               f          f          f
61066                    f               f          f          f
61067                    f               f          f          f
61068                    f               f          f          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …          s            y          w         u
1            17.99  …          s            y          w         u
2            17.80  …          s            y          w         u
3            15.77  …          s            y          w         u
4            16.53  …          s            y          w         u
…              …  …        …          …        …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
```

```
2              w         t         g                NaN       d       w
3              w         t         p                NaN       d       w
4              w         t         p                NaN       d       w
…            …       …         …                  …       …     …
61064        NaN         f         f                NaN       d       a
61065        NaN         f         f                NaN       d       a
61066        NaN         f         f                NaN       d       u
61067        NaN         f         f                NaN       d       u
61068        NaN         f         f                NaN       d       u

[61069 rows x 21 columns]
```

1.3 Magari non morirà, dopotutto

ok, adesso faremo la stessa cosa con tutte le altre colonne, a partire da cap-shape, ma questa volta useremo un dizionario nel replace, decisamente più comodo

[ ]:
```
legend_mapping = {
    'b': 'bell',
    'c': 'conical',
    'x': 'convex',
    'f': 'flat',
    's': 'sunken',
    'p': 'spherical',
    'o': 'others'
}

df1['cap-shape'] = df1['cap-shape'].replace(legend_mapping)
df1
```

[ ]:
```
           class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex           g         o
1      poisonus         16.60    convex           g         o
2      poisonus         14.07    convex           g         o
3      poisonus         14.17      flat           h         e
4      poisonus         14.64    convex           h         o
…         …           …        …         …        …
61064  poisonus          1.18    sunken           s         y
61065  poisonus          1.27      flat           s         y
61066  poisonus          1.27    sunken           s         y
61067  poisonus          1.24      flat           s         y
61068  poisonus          1.17    sunken           s         y

       does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                         f               e          NaN          w
1                         f               e          NaN          w
2                         f               e          NaN          w
```

```
3                      f              e          NaN          w
4                      f              e          NaN          w
…                    …              …          …          …
61064                  f              f          f          f
61065                  f              f          f          f
61066                  f              f          f          f
61067                  f              f          f          f
61068                  f              f          f          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …          s            y          w         u
1            17.99  …          s            y          w         u
2            17.80  …          s            y          w         u
3            15.77  …          s            y          w         u
4            16.53  …          s            y          w         u
…              …  …        …          …        …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
2              w        t         g               NaN       d      w
3              w        t         p               NaN       d      w
4              w        t         p               NaN       d      w
…            …      …       …               …     …    …
61064        NaN        f         f               NaN       d      a
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
61068        NaN        f         f               NaN       d      u

[61069 rows x 21 columns]
```

1.4 adesso ancora, ancora e ancora...

[ ]:
```
cap_surface_mapping = {
    'i': 'fibrous',
    'g': 'grooves',
    'y': 'scaly',
    's': 'smooth',
    'h': 'shiny',
    'l': 'leathery',
    'k': 'silky'
```

```
}

df1['cap-surface'] = df1['cap-surface'].replace(cap_surface_mapping)
df1
```

[ ]:

```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                        f               e          NaN          w
1                        f               e          NaN          w
2                        f               e          NaN          w
3                        f               e          NaN          w
4                        f               e          NaN          w
…                      …               …          …       …
61064                    f               f          f            f
61065                    f               f          f            f
61066                    f               f          f            f
61067                    f               f          f            f
61068                    f               f          f            f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …          s            y          w         u
1            17.99  …          s            y          w         u
2            17.80  …          s            y          w         u
3            15.77  …          s            y          w         u
4            16.53  …          s            y          w         u
…            …  …       …          …        …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
```

```
2              w         t         g                NaN       d       w
3              w         t         p                NaN       d       w
4              w         t         p                NaN       d       w
…            …       …         …                  …       …     …
61064        NaN         f         f                NaN       d       a
61065        NaN         f         f                NaN       d       a
61066        NaN         f         f                NaN       d       u
61067        NaN         f         f                NaN       d       u
61068        NaN         f         f                NaN       d       u

[61069 rows x 21 columns]
```

Il data scientis che ha fatto questa ricerca, non ha saputo fare di meglio che mettere solo l'iniziale (presumibilmente) del colore, ma purtroppo dovremo accontentarci della migliore interpretazione possibile, adesso andiamo a vedere quante lettere appaiono e le confrontiamo con la lista (scarsa) di colori che l'autore ci ha dato:

brown, buff, gray, green, pink, purple, red, white

[ ]:
```
unique_cap_colors = df1['cap-color'].unique()
print(unique_cap_colors)
```

```
['o' 'e' 'n' 'g' 'r' 'w' 'y' 'p' 'u' 'b' 'l' 'k']
```

Con questo `unique()` selezionamo tutti gli elementi che appaiono una volta sola Adesso inizieremo ad assegnare ad ogni lettera un plausibile colore:

[ ]:
```
# gray, pink
# ['o' 'e' 'u' 'l' 'k']
colors_legenda = {
    'w' : 'white',
    'r' : 'red',
    'p' : 'purple',
    'g' : 'green',
    'y' : 'yellow',
    'b' : 'buff',
    'n' : 'brown',
}
```

Ma qui è dove i colori e le associazioni ragionevoli si fermano, quindi ho dovuto gettare la spugna ed eliminare l'idea di sostiuire e rendere leggibile per quella specifica colonna (che comunque non andremo ad eliminare); proseguiamo con le altre colonne

[ ]:
```
df1['does-bruise-or-bleed'] = df1['does-bruise-or-bleed'].replace({"f" :␣
↪"False", "t" : "true"})
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
```

```
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …        …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False               e          NaN          w
1                    False               e          NaN          w
2                    False               e          NaN          w
3                    False               e          NaN          w
4                    False               e          NaN          w
…                    …                 …          …        …
61064                False               f            f          f
61065                False               f            f          f
61066                False               f            f          f
61067                False               f            f          f
61068                False               f            f          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …          s            y          w         u
1            17.99  …          s            y          w         u
2            17.80  …          s            y          w         u
3            15.77  …          s            y          w         u
4            16.53  …          s            y          w         u
…            …  …        …          …        …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
2              w        t         g               NaN       d      w
3              w        t         p               NaN       d      w
4              w        t         p               NaN       d      w
…          …       …        …               …     …    …
61064        NaN        f         f               NaN       d      a
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
```

```
61068        NaN        f         f                NaN       d      u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_gill_attachment = {
    'a': 'adnate',
    'x': 'adnexed',
    'd': 'decurrent',
    'e': 'free',
    's': 'sinuate',
    'p': 'pores',
    'f': 'none'
}

df1['gill-attachment'] = df1['gill-attachment'].
 ↪replace(legend_mapping_gill_attachment)
df1
```

[ ]:
```
           class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …        …          …          …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

       does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                     False            free          NaN          w
1                     False            free          NaN          w
2                     False            free          NaN          w
3                     False            free          NaN          w
4                     False            free          NaN          w
…                     …              …           …          …
61064                 False            none            f          f
61065                 False            none            f          f
61066                 False            none            f          f
61067                 False            none            f          f
61068                 False            none            f          f

       stem-height  …  stem-surface stem-color veil-type veil-color  \
0            16.95  …             y          w         u          w
1            17.99  …             y          w         u          w
```

```
2            17.80  …             y          w         u          w
3            15.77  …             y          w         u          w
4            16.53  …             y          w         u          w
…              …  …           …          …         …         …
61064         3.93  …           NaN          y       NaN        NaN
61065         3.18  …           NaN          y       NaN        NaN
61066         3.86  …           NaN          y       NaN        NaN
61067         3.56  …           NaN          y       NaN        NaN
61068         3.25  …           NaN          y       NaN        NaN

      has-ring ring-type spore-print-color habitat season  \
0            t         g                NaN       d      w
1            t         g                NaN       d      u
2            t         g                NaN       d      w
3            t         p                NaN       d      w
4            t         p                NaN       d      w
…          …         …                …       …    …
61064        f         f                NaN       d      a
61065        f         f                NaN       d      a
61066        f         f                NaN       d      u
61067        f         f                NaN       d      u
61068        f         f                NaN       d      u

      gill-attachment-column
0                       free
1                       free
2                       free
3                       free
4                       free
…                       …
61064                   none
61065                   none
61066                   none
61067                   none
61068                   none

[61069 rows x 22 columns]
```

[ ]:
```
legend_mapping_gill_spacing = {
    'c': 'close',
    'd': 'distant',
    'f': 'none'
}

df1['gill-spacing'] = df1['gill-spacing'].replace(legend_mapping_gill_spacing)
df1
```

[ ]:

	class	cap-diameter	cap-shape	cap-surface	cap-color	\
0	poisonus	15.26	convex	grooves	o	
1	poisonus	16.60	convex	grooves	o	
2	poisonus	14.07	convex	grooves	o	
3	poisonus	14.17	flat	shiny	e	
4	poisonus	14.64	convex	shiny	o	
…	…	…	…	…	…	
61064	poisonus	1.18	sunken	sunken	y	
61065	poisonus	1.27	flat	sunken	y	
61066	poisonus	1.27	sunken	sunken	y	
61067	poisonus	1.24	flat	sunken	y	
61068	poisonus	1.17	sunken	sunken	y	

	does-bruise-or-bleed	gill-attachment	gill-spacing	gill-color	\
0	False	free	NaN	w	
1	False	free	NaN	w	
2	False	free	NaN	w	
3	False	free	NaN	w	
4	False	free	NaN	w	
…	…	…	…	…	
61064	False	none	none	f	
61065	False	none	none	f	
61066	False	none	none	f	
61067	False	none	none	f	
61068	False	none	none	f	

	stem-height	…	stem-root	stem-surface	stem-color	veil-type	\
0	16.95	…	s	y	w	u	
1	17.99	…	s	y	w	u	
2	17.80	…	s	y	w	u	
3	15.77	…	s	y	w	u	
4	16.53	…	s	y	w	u	
…	…	…	…	…	…	…	
61064	3.93	…	NaN	NaN	y	NaN	
61065	3.18	…	NaN	NaN	y	NaN	
61066	3.86	…	NaN	NaN	y	NaN	
61067	3.56	…	NaN	NaN	y	NaN	
61068	3.25	…	NaN	NaN	y	NaN	

	veil-color	has-ring	ring-type	spore-print-color	habitat	season
0	w	t	g	NaN	d	w
1	w	t	g	NaN	d	u
2	w	t	g	NaN	d	w
3	w	t	p	NaN	d	w
4	w	t	p	NaN	d	w
…	…	…	…	…	…	…
61064	NaN	f	f	NaN	d	a

```
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
61068        NaN        f         f               NaN       d      u

[61069 rows x 21 columns]
```

Per il gill-color è lo stesso discorso del cap-color

teniamo i dati numerci intoccati

[ ]:
```
legend_mapping_stem_root = {
    'b': 'bulbous',
    's': 'swollen',
    'c': 'club',
    'u': 'cup',
    'e': 'equal',
    'z': 'rhizomorphs',
    'r': 'rooted'
}

df1['stem-root'] = df1['stem-root'].replace(legend_mapping_stem_root)
df1
```

[ ]:
```
           class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                    …              …           …        …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
```

```
61068               False            none         none          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen            y          w         u
1            17.99  …    swollen            y          w         u
2            17.80  …    swollen            y          w         u
3            15.77  …    swollen            y          w         u
4            16.53  …    swollen            y          w         u
…              …  …        …           …         …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
2              w        t         g               NaN       d      w
3              w        t         p               NaN       d      w
4              w        t         p               NaN       d      w
…          …      …       …               …     …    …
61064        NaN        f         f               NaN       d      a
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
61068        NaN        f         f               NaN       d      u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_stem_surface = {
    'i': 'fibrous',
    'g': 'grooves',
    'y': 'scaly',
    's': 'smooth',
    'h': 'shiny',
    'l': 'leathery',
    'k': 'silky',
    'f': 'none'
}

df1['stem-surface'] = df1['stem-surface'].replace(legend_mapping_stem_surface)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
```

```
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …        …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                    …              …           …        …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen        scaly          w         u
1            17.99  …    swollen        scaly          w         u
2            17.80  …    swollen        scaly          w         u
3            15.77  …    swollen        scaly          w         u
4            16.53  …    swollen        scaly          w         u
…            …  …        …           …        …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
2              w        t         g               NaN       d      w
3              w        t         p               NaN       d      w
4              w        t         p               NaN       d      w
…          …       …        …               …     …    …
61064        NaN        f         f               NaN       d      a
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
```

```
61068        NaN        f         f               NaN       d      u

[61069 rows x 21 columns]
```

```
[ ]: df1['stem-color'] = df1['stem-color'].replace({"f" : "None"})
     df1
```

```
[ ]:          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                    …              …           …        …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen        scaly          w         u
1            17.99  …    swollen        scaly          w         u
2            17.80  …    swollen        scaly          w         u
3            15.77  …    swollen        scaly          w         u
4            16.53  …    swollen        scaly          w         u
…            …  …       …          …         …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
```

```
0              w        t         g                NaN       d       w
1              w        t         g                NaN       d       u
2              w        t         g                NaN       d       w
3              w        t         p                NaN       d       w
4              w        t         p                NaN       d       w
…            …      …       …                …      …     …
61064        NaN        f         f                NaN       d       a
61065        NaN        f         f                NaN       d       a
61066        NaN        f         f                NaN       d       u
61067        NaN        f         f                NaN       d       u
61068        NaN        f         f                NaN       d       u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_veil_color = {
    'f': 'none'
}

df1['veil-color'] = df1['veil-color'].replace(legend_mapping_veil_color)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                    …              …           …        …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f
```

```
       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen        scaly          w         u
1            17.99  …    swollen        scaly          w         u
2            17.80  …    swollen        scaly          w         u
3            15.77  …    swollen        scaly          w         u
4            16.53  …    swollen        scaly          w         u
…              …  …          …            …          …         …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w        t         g               NaN       d      w
1              w        t         g               NaN       d      u
2              w        t         g               NaN       d      w
3              w        t         p               NaN       d      w
4              w        t         p               NaN       d      w
…            …      …       …               …     …    …
61064        NaN        f         f               NaN       d      a
61065        NaN        f         f               NaN       d      a
61066        NaN        f         f               NaN       d      u
61067        NaN        f         f               NaN       d      u
61068        NaN        f         f               NaN       d      u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_has_ring = {
    't': 'ring',
    'f': 'none'
}
df1['has-ring'] = df1['has-ring'].replace(legend_mapping_has_ring)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y
```

```
      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                      …               …            …          …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen        scaly          w         u
1            17.99  …    swollen        scaly          w         u
2            17.80  …    swollen        scaly          w         u
3            15.77  …    swollen        scaly          w         u
4            16.53  …    swollen        scaly          w         u
…              …  …          …            …          …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w     ring         g               NaN       d      w
1              w     ring         g               NaN       d      u
2              w     ring         g               NaN       d      w
3              w     ring         p               NaN       d      w
4              w     ring         p               NaN       d      w
…            …      …       …               …      …    …
61064        NaN     none         f               NaN       d      a
61065        NaN     none         f               NaN       d      a
61066        NaN     none         f               NaN       d      u
61067        NaN     none         f               NaN       d      u
61068        NaN     none         f               NaN       d      u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_ring_type = {
    'c': 'cobwebby',
    'e': 'evanescent',
    'r': 'flaring',
    'g': 'grooved',
```

```
    'l': 'large',
    'p': 'pendant',
    's': 'sheathing'
}
df1['ring-type'] = df1['ring-type'].replace(legend_mapping_ring_type)
df1
```

[ ]:

	class	cap-diameter	cap-shape	cap-surface	cap-color	\
0	poisonus	15.26	convex	grooves	o	
1	poisonus	16.60	convex	grooves	o	
2	poisonus	14.07	convex	grooves	o	
3	poisonus	14.17	flat	shiny	e	
4	poisonus	14.64	convex	shiny	o	
…	…	…	…	…	…	
61064	poisonus	1.18	sunken	sunken	y	
61065	poisonus	1.27	flat	sunken	y	
61066	poisonus	1.27	sunken	sunken	y	
61067	poisonus	1.24	flat	sunken	y	
61068	poisonus	1.17	sunken	sunken	y	

	does-bruise-or-bleed	gill-attachment	gill-spacing	gill-color	\
0	False	free	NaN	w	
1	False	free	NaN	w	
2	False	free	NaN	w	
3	False	free	NaN	w	
4	False	free	NaN	w	
…	…	…	…	…	
61064	False	none	none	f	
61065	False	none	none	f	
61066	False	none	none	f	
61067	False	none	none	f	
61068	False	none	none	f	

	stem-height	…	stem-root	stem-surface	stem-color	veil-type	\
0	16.95	…	swollen	scaly	w	u	
1	17.99	…	swollen	scaly	w	u	
2	17.80	…	swollen	scaly	w	u	
3	15.77	…	swollen	scaly	w	u	
4	16.53	…	swollen	scaly	w	u	
…	…	…	…	…	…	…	
61064	3.93	…	NaN	NaN	y	NaN	
61065	3.18	…	NaN	NaN	y	NaN	
61066	3.86	…	NaN	NaN	y	NaN	
61067	3.56	…	NaN	NaN	y	NaN	
61068	3.25	…	NaN	NaN	y	NaN	

	veil-color	has-ring	ring-type	spore-print-color	habitat	season

```
0              w     ring   grooved               NaN       d      w
1              w     ring   grooved               NaN       d      u
2              w     ring   grooved               NaN       d      w
3              w     ring   pendant               NaN       d      w
4              w     ring   pendant               NaN       d      w
…            …      …       …                    …      …    …
61064        NaN     none         f               NaN       d      a
61065        NaN     none         f               NaN       d      a
61066        NaN     none         f               NaN       d      u
61067        NaN     none         f               NaN       d      u
61068        NaN     none         f               NaN       d      u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_habitat = {
    'g': 'grasses',
    'l': 'leaves',
    'm': 'meadows',
    'p': 'paths',
    'h': 'heaths',
    'u': 'urban',
    'w': 'waste',
    'd': 'woods'
}
df1['habitat'] = df1['habitat'].replace(legend_mapping_habitat)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                    …              …           …        …
```

```
61064               False           none         none          f
61065               False           none         none          f
61066               False           none         none          f
61067               False           none         none          f
61068               False           none         none          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen        scaly          w         u
1            17.99  …    swollen        scaly          w         u
2            17.80  …    swollen        scaly          w         u
3            15.77  …    swollen        scaly          w         u
4            16.53  …    swollen        scaly          w         u
…              …  …        …           …         …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat season
0              w     ring   grooved               NaN   woods      w
1              w     ring   grooved               NaN   woods      u
2              w     ring   grooved               NaN   woods      w
3              w     ring   pendant               NaN   woods      w
4              w     ring   pendant               NaN   woods      w
…          …      …       …                …      …    …
61064        NaN     none         f               NaN   woods      a
61065        NaN     none         f               NaN   woods      a
61066        NaN     none         f               NaN   woods      u
61067        NaN     none         f               NaN   woods      u
61068        NaN     none         f               NaN   woods      u

[61069 rows x 21 columns]
```

[ ]:
```
legend_mapping_season = {
    's': 'spring',
    'u': 'summer',
    'a': 'autumn',
    'w': 'winter'
}
df1['season'] = df1['season'].replace(legend_mapping_season)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
```

```
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …           …       …          …       …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free          NaN          w
1                    False            free          NaN          w
2                    False            free          NaN          w
3                    False            free          NaN          w
4                    False            free          NaN          w
…                    …              …           …        …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f

       stem-height  …  stem-root stem-surface stem-color veil-type  \
0            16.95  …    swollen        scaly          w         u
1            17.99  …    swollen        scaly          w         u
2            17.80  …    swollen        scaly          w         u
3            15.77  …    swollen        scaly          w         u
4            16.53  …    swollen        scaly          w         u
…            …  …        …           …        …       …
61064         3.93  …        NaN          NaN          y       NaN
61065         3.18  …        NaN          NaN          y       NaN
61066         3.86  …        NaN          NaN          y       NaN
61067         3.56  …        NaN          NaN          y       NaN
61068         3.25  …        NaN          NaN          y       NaN

      veil-color has-ring ring-type spore-print-color habitat  season
0              w     ring   grooved               NaN   woods  winter
1              w     ring   grooved               NaN   woods  summer
2              w     ring   grooved               NaN   woods  winter
3              w     ring   pendant               NaN   woods  winter
4              w     ring   pendant               NaN   woods  winter
…          …       …        …                …     …      …
61064        NaN     none         f               NaN   woods  autumn
61065        NaN     none         f               NaN   woods  autumn
61066        NaN     none         f               NaN   woods  summer
61067        NaN     none         f               NaN   woods  summer
61068        NaN     none         f               NaN   woods  summer
```

```
[61069 rows x 21 columns]
```

1.5 adesso controlliamo quanti dati mancano

sì, di nuovo

[ ]:
```
percentage_missing = df.isnull().sum() * 100 / len(df)
print(percentage_missing)
```

```
class                    0.000000
cap-diameter             0.000000
cap-shape                0.000000
cap-surface             23.121387
cap-color                0.000000
does-bruise-or-bleed     0.000000
gill-attachment         16.184971
gill-spacing            41.040462
gill-color               0.000000
stem-height              0.000000
stem-width               0.000000
stem-root               84.393064
stem-surface            62.427746
stem-color               0.000000
veil-type               94.797688
veil-color              87.861272
has-ring                 0.000000
ring-type                4.046243
spore-print-color       89.595376
habitat                  0.000000
season                   0.000000
dtype: float64
```

Questo vuol dire che quelle con 0.0 possiamo lasciarle stare, perché sono complete, quelle con più di 80 ci conviene eliminarle in realtà poiché non sarebbe una media oggettiva

[ ]:
```
missing_values = df1['cap-surface'].isna().sum()
if missing_values > 0:
    mode_value = df1['cap-surface'].mode()[0]
    df1['cap-surface'].fillna(mode_value, inplace=True)
    print(f"Replaced {missing_values} missing values with the mode␣
 ↪({mode_value}) in the 'cap-surface' column.")
```

[ ]:
```
missing_values = df1['gill-attachment'].isna().sum()
if missing_values > 0:
    mode_value = df1['gill-attachment'].mode()[0]
    df1['gill-attachment'].fillna(mode_value, inplace=True)
```

```
    print(f"Replaced {missing_values} missing values with the mode␣
 ↪({mode_value}) in the 'gill-attachment' column.")
```

```
Replaced 9884 missing values with the mode (adnate) in the 'gill-attachment'
column.
```

[ ]:
```
missing_values = df1['gill-spacing'].isna().sum()
if missing_values > 0:
    mode_value = df1['gill-spacing'].mode()[0]
    df1['gill-spacing'].fillna(mode_value, inplace=True)
    print(f"Replaced {missing_values} missing values with the mode␣
 ↪({mode_value}) in the 'gill-spacing' column.")
```

```
Replaced 25063 missing values with the mode (close) in the 'gill-spacing'
column.
```

[ ]:
```
df1.drop(columns=['stem-root'], inplace=True)
df1
```

[ ]:
```
           class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …            …        …          …          …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

       does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                     False            free        close          w
1                     False            free        close          w
2                     False            free        close          w
3                     False            free        close          w
4                     False            free        close          w
…                     …              …           …          …
61064                 False            none         none          f
61065                 False            none         none          f
61066                 False            none         none          f
61067                 False            none         none          f
61068                 False            none         none          f

       stem-height  stem-width stem-surface stem-color veil-type veil-color  \
0            16.95       17.09        scaly          w         u          w
1            17.99       18.19        scaly          w         u          w
```

```
2            17.80       17.74        scaly          w         u          w
3            15.77       15.98        scaly          w         u          w
4            16.53       17.20        scaly          w         u          w
…              …           …            …            …         …          …
61064         3.93        6.22          NaN          y       NaN        NaN
61065         3.18        5.43          NaN          y       NaN        NaN
61066         3.86        6.37          NaN          y       NaN        NaN
61067         3.56        5.44          NaN          y       NaN        NaN
61068         3.25        5.45          NaN          y       NaN        NaN

      has-ring ring-type spore-print-color habitat  season
0         ring   grooved               NaN   woods  winter
1         ring   grooved               NaN   woods  summer
2         ring   grooved               NaN   woods  winter
3         ring   pendant               NaN   woods  winter
4         ring   pendant               NaN   woods  winter
…          …       …                   …      …      …
61064     none         f               NaN   woods  autumn
61065     none         f               NaN   woods  autumn
61066     none         f               NaN   woods  summer
61067     none         f               NaN   woods  summer
61068     none         f               NaN   woods  summer

[61069 rows x 20 columns]
```

[ ]:
```
missing_values = df1['stem-surface'].isna().sum()
if missing_values > 0:
    mode_value = df1['stem-surface'].mode()[0]
    df1['stem-surface'].fillna(mode_value, inplace=True)
    print(f"Replaced {missing_values} missing values with the mode␣
 ↪({mode_value}) in the 'stem-surface' column.")
```

```
Replaced 38124 missing values with the mode (smooth) in the 'stem-surface'
column.
```

[ ]:
```
df1.drop(columns=['veil-type'], inplace=True)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …            …          …          …          …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
```

```
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free        close          w
1                    False            free        close          w
2                    False            free        close          w
3                    False            free        close          w
4                    False            free        close          w
…                        …               …            …        …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f

       stem-height  stem-width stem-surface stem-color veil-color has-ring  \
0            16.95       17.09        scaly          w          w     ring
1            17.99       18.19        scaly          w          w     ring
2            17.80       17.74        scaly          w          w     ring
3            15.77       15.98        scaly          w          w     ring
4            16.53       17.20        scaly          w          w     ring
…                …           …            …          …        …      …
61064         3.93        6.22       smooth          y        NaN     none
61065         3.18        5.43       smooth          y        NaN     none
61066         3.86        6.37       smooth          y        NaN     none
61067         3.56        5.44       smooth          y        NaN     none
61068         3.25        5.45       smooth          y        NaN     none

      ring-type spore-print-color habitat  season
0       grooved               NaN   woods  winter
1       grooved               NaN   woods  summer
2       grooved               NaN   woods  winter
3       pendant               NaN   woods  winter
4       pendant               NaN   woods  winter
…         …                   …      …      …
61064         f               NaN   woods  autumn
61065         f               NaN   woods  autumn
61066         f               NaN   woods  summer
61067         f               NaN   woods  summer
61068         f               NaN   woods  summer

[61069 rows x 19 columns]
```

[ ]:
```
df1.drop(columns=['veil-color'], inplace=True)
df1
```

[ ]:

	class	cap-diameter	cap-shape	cap-surface	cap-color	\
0	poisonus	15.26	convex	grooves	o	
1	poisonus	16.60	convex	grooves	o	
2	poisonus	14.07	convex	grooves	o	
3	poisonus	14.17	flat	shiny	e	
4	poisonus	14.64	convex	shiny	o	
…	…	…	…	…	…	
61064	poisonus	1.18	sunken	sunken	y	
61065	poisonus	1.27	flat	sunken	y	
61066	poisonus	1.27	sunken	sunken	y	
61067	poisonus	1.24	flat	sunken	y	
61068	poisonus	1.17	sunken	sunken	y	

	does-bruise-or-bleed	gill-attachment	gill-spacing	gill-color	\
0	False	free	close	w	
1	False	free	close	w	
2	False	free	close	w	
3	False	free	close	w	
4	False	free	close	w	
…	…	…	…	…	
61064	False	none	none	f	
61065	False	none	none	f	
61066	False	none	none	f	
61067	False	none	none	f	
61068	False	none	none	f	

	stem-height	stem-width	stem-surface	stem-color	has-ring	ring-type	\
0	16.95	17.09	scaly	w	ring	grooved	
1	17.99	18.19	scaly	w	ring	grooved	
2	17.80	17.74	scaly	w	ring	grooved	
3	15.77	15.98	scaly	w	ring	pendant	
4	16.53	17.20	scaly	w	ring	pendant	
…	…	…	…	…	…	…	
61064	3.93	6.22	smooth	y	none	f	
61065	3.18	5.43	smooth	y	none	f	
61066	3.86	6.37	smooth	y	none	f	
61067	3.56	5.44	smooth	y	none	f	
61068	3.25	5.45	smooth	y	none	f	

	spore-print-color	habitat	season
0	NaN	woods	winter
1	NaN	woods	summer
2	NaN	woods	winter
3	NaN	woods	winter
4	NaN	woods	winter
…	…	…	…
61064	NaN	woods	autumn

```
61065               NaN   woods  autumn
61066               NaN   woods  summer
61067               NaN   woods  summer
61068               NaN   woods  summer

[61069 rows x 18 columns]
```

[ ]:
```
missing_values = df1['ring-type'].isna().sum()
if missing_values > 0:
    mode_value = df1['ring-type'].mode()[0]
    df1['ring-type'].fillna(mode_value, inplace=True)
    print(f"Replaced {missing_values} missing values with the mode␣
 ↪({mode_value}) in the 'ring-type' column.")
```

```
Replaced 2471 missing values with the mode (f) in the 'ring-type' column.
```

[ ]:
```
df1.drop(columns=['spore-print-color'], inplace=True)
df1
```

[ ]:
```
          class  cap-diameter cap-shape cap-surface cap-color  \
0      poisonus         15.26    convex     grooves         o
1      poisonus         16.60    convex     grooves         o
2      poisonus         14.07    convex     grooves         o
3      poisonus         14.17      flat       shiny         e
4      poisonus         14.64    convex       shiny         o
…         …            …        …          …          …
61064  poisonus          1.18    sunken      sunken         y
61065  poisonus          1.27      flat      sunken         y
61066  poisonus          1.27    sunken      sunken         y
61067  poisonus          1.24      flat      sunken         y
61068  poisonus          1.17    sunken      sunken         y

      does-bruise-or-bleed gill-attachment gill-spacing gill-color  \
0                    False            free        close          w
1                    False            free        close          w
2                    False            free        close          w
3                    False            free        close          w
4                    False            free        close          w
…                    …              …           …          …
61064                False            none         none          f
61065                False            none         none          f
61066                False            none         none          f
61067                False            none         none          f
61068                False            none         none          f

       stem-height  stem-width stem-surface stem-color has-ring ring-type  \
0            16.95       17.09        scaly          w     ring   grooved
1            17.99       18.19        scaly          w     ring   grooved
```

```
2            17.80       17.74        scaly          w     ring   grooved
3            15.77       15.98        scaly          w     ring   pendant
4            16.53       17.20        scaly          w     ring   pendant
…            …           …            …              …     …      …
61064         3.93        6.22       smooth          y     none         f
61065         3.18        5.43       smooth          y     none         f
61066         3.86        6.37       smooth          y     none         f
61067         3.56        5.44       smooth          y     none         f
61068         3.25        5.45       smooth          y     none         f

      habitat  season
0       woods  winter
1       woods  summer
2       woods  winter
3       woods  winter
4       woods  winter
…       …      …
61064   woods  autumn
61065   woods  autumn
61066   woods  summer
61067   woods  summer
61068   woods  summer

[61069 rows x 17 columns]
```

2 Perché tutto questo lavoro?

per provare un punto, ovvero che puoi essere il miglior data scientist al mondo, ma non sarai niente senza una conoscenza della programmazione, perché tutto questo poteva essere fatto in 11 linee di codice

[ ]:
```python
df2 = df
for column in df2.columns:
    missing_percentage = (df2[column].isna().sum() / len(df2)) * 100
    if missing_percentage == 0:
        continue
    elif missing_percentage > 84:
        df2.drop(columns=[column], inplace=True)
        print(f"Dropped the '{column}' column due to {missing_percentage}%␣
 ↪missing values.")
    else:
        mean_value = df2[column].mean()
        df2[column].fillna(mean_value, inplace=True)
        print(f"Replaced {missing_percentage}% missing values with the mean␣
 ↪({mean_value}) in the '{column}' column.")
```

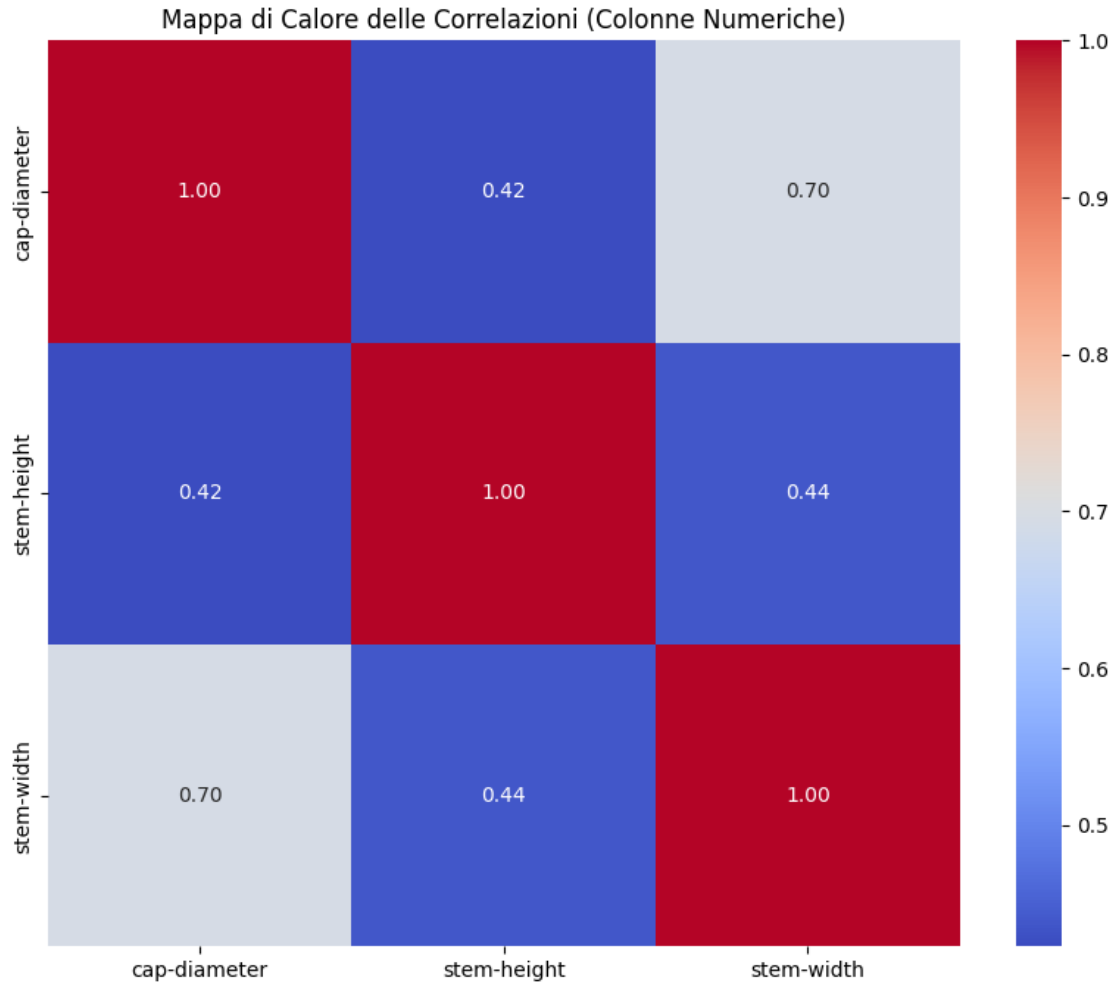
controlliamo se ha funzionato

[ ]:
```
if df1.equals(df2):
    print("df1 is equal to df2")
else:
    print("df1 is not equal to df2")
```

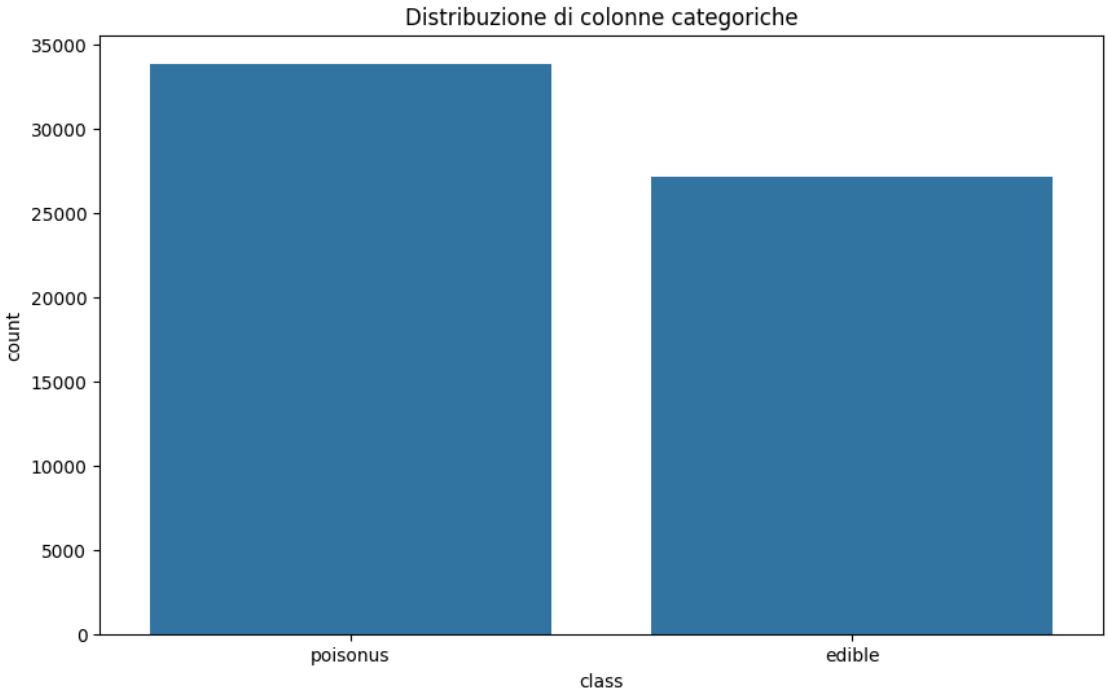
```
df1 is equal to df2
```

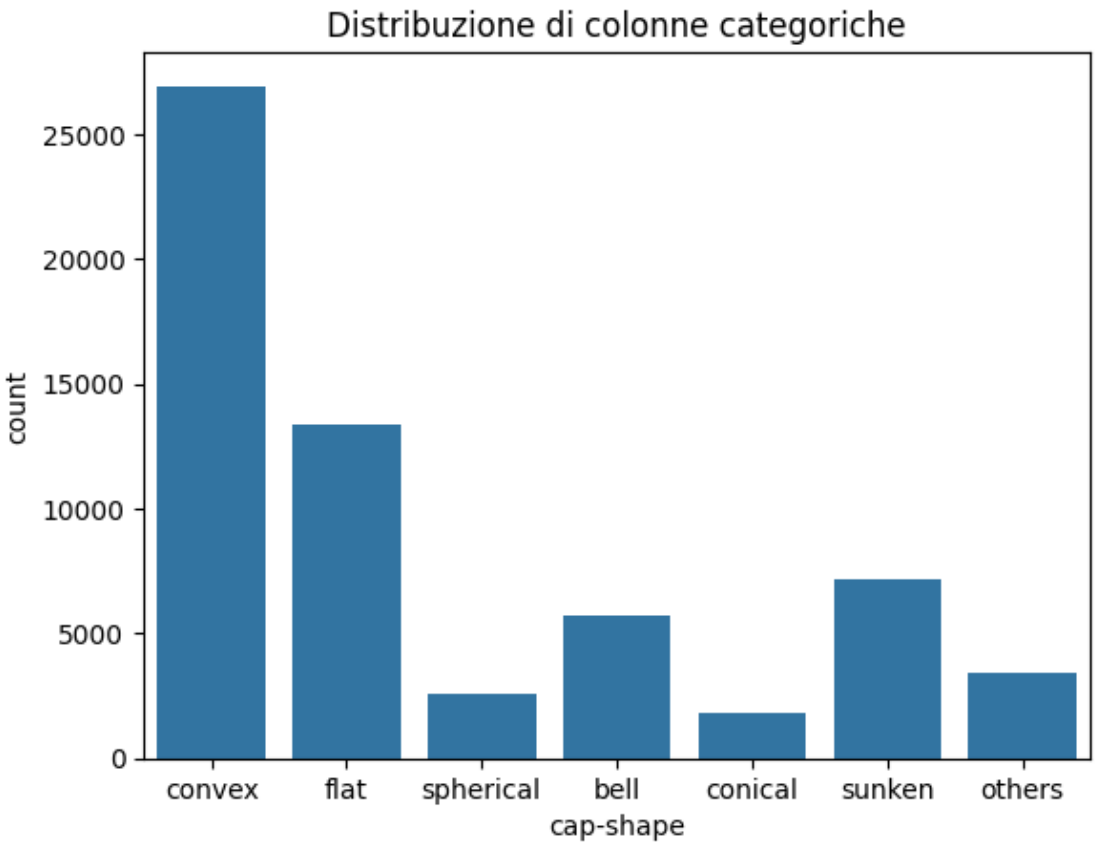
Data tutta la fatica fatta per questa dimostrazione userò df2 perché ci piacciono le cose efficienti

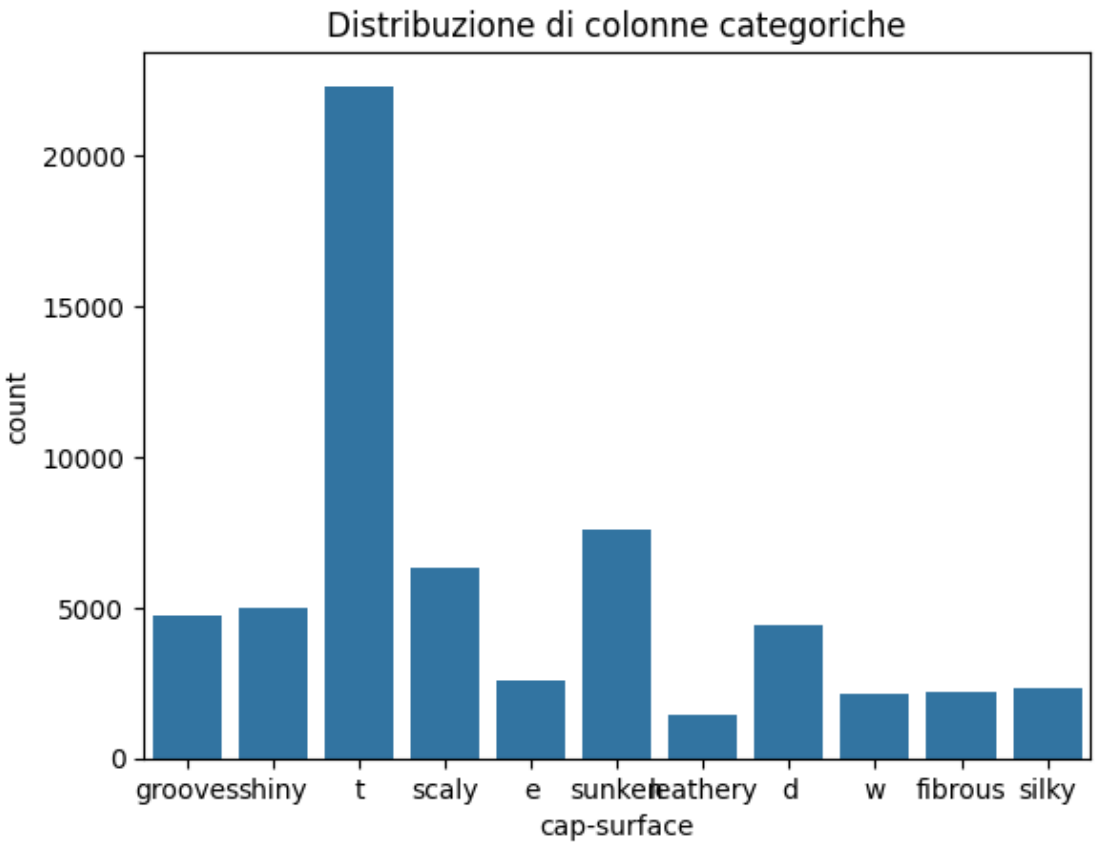
[ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

numeric_columns = df2.select_dtypes(include=['float64', 'int64'])
correlation_matrix = numeric_columns.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Mappa di Calore delle Correlazioni (Colonne Numeriche)')
plt.show()
```

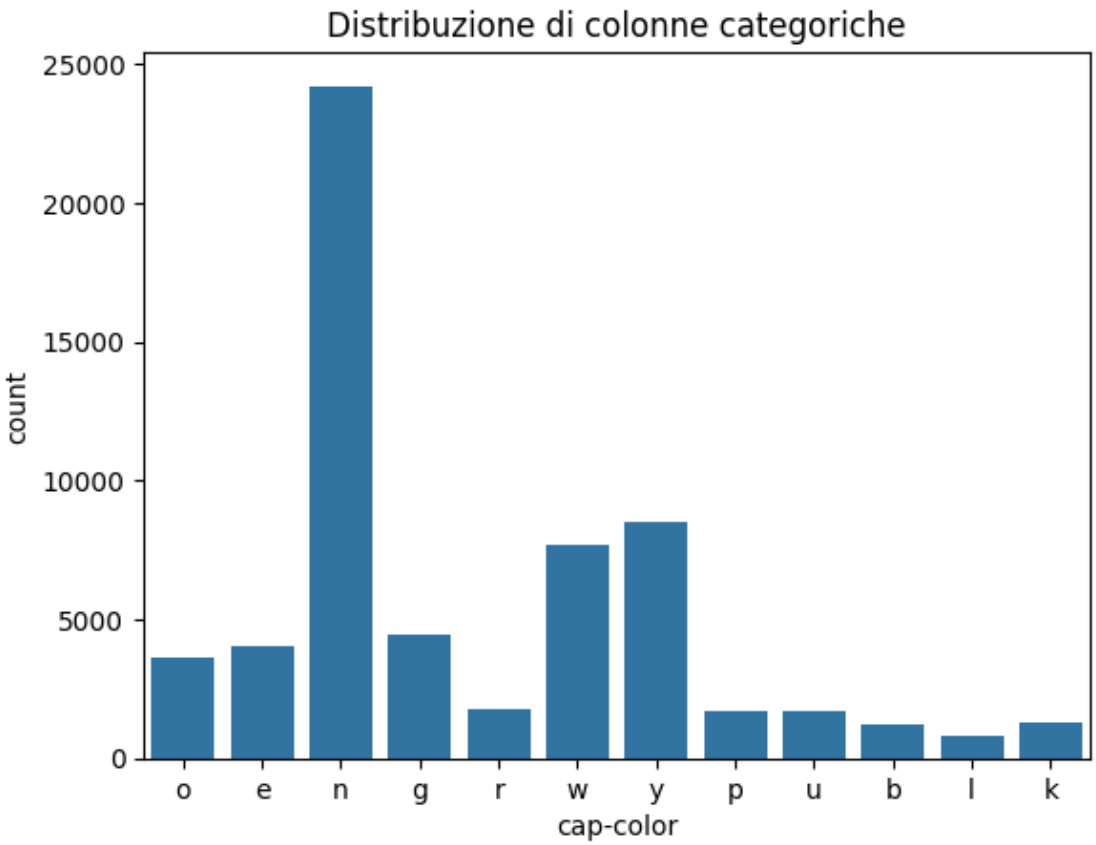


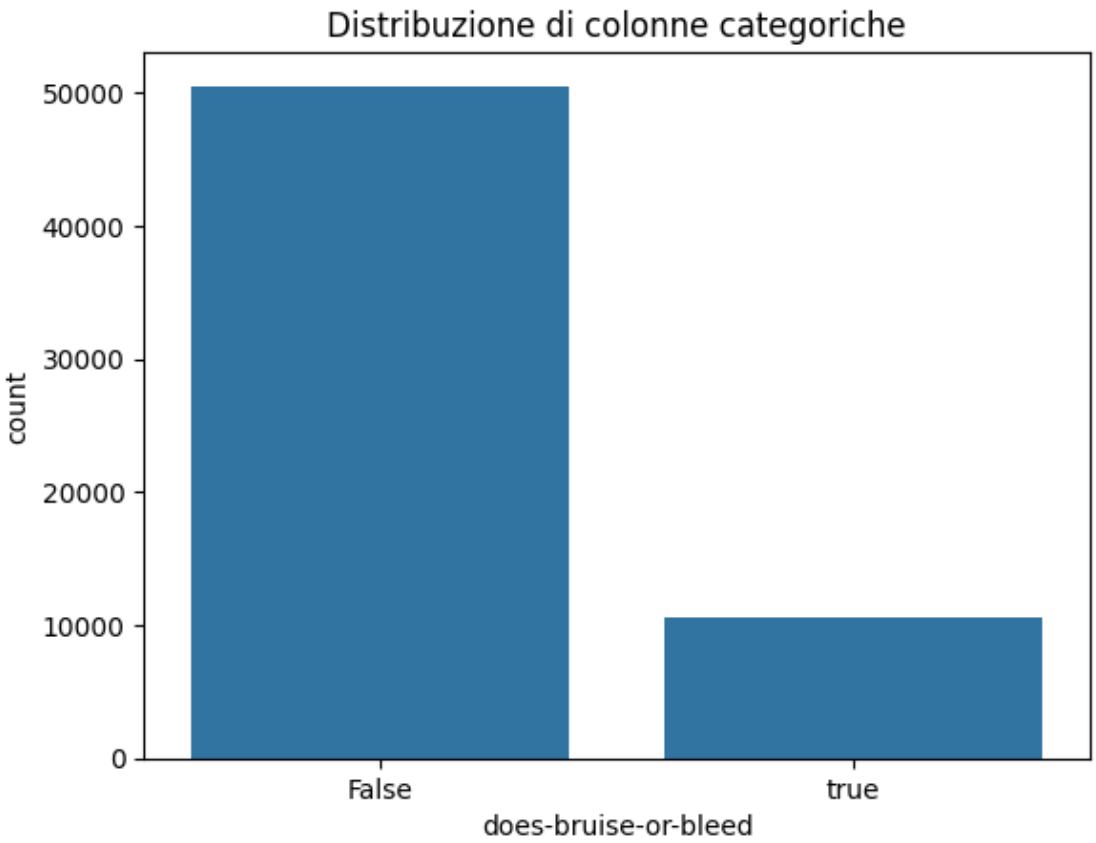
[ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

text_columns = df2.select_dtypes(include=['object'])
plt.figure(figsize=(10, 6))
for col in text_columns.columns:
    sns.countplot(x=col, data=df2)
    plt.title(f'Distribuzione di colonne categoriche')
    plt.show()
```

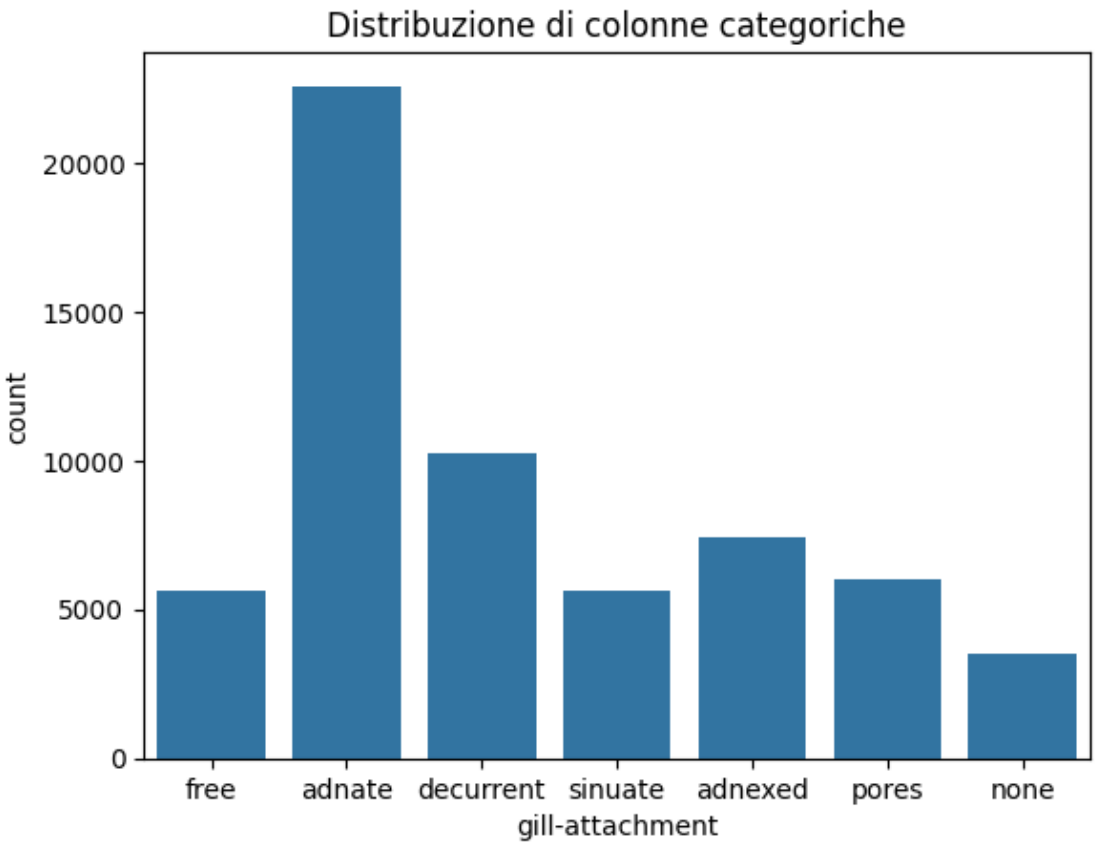


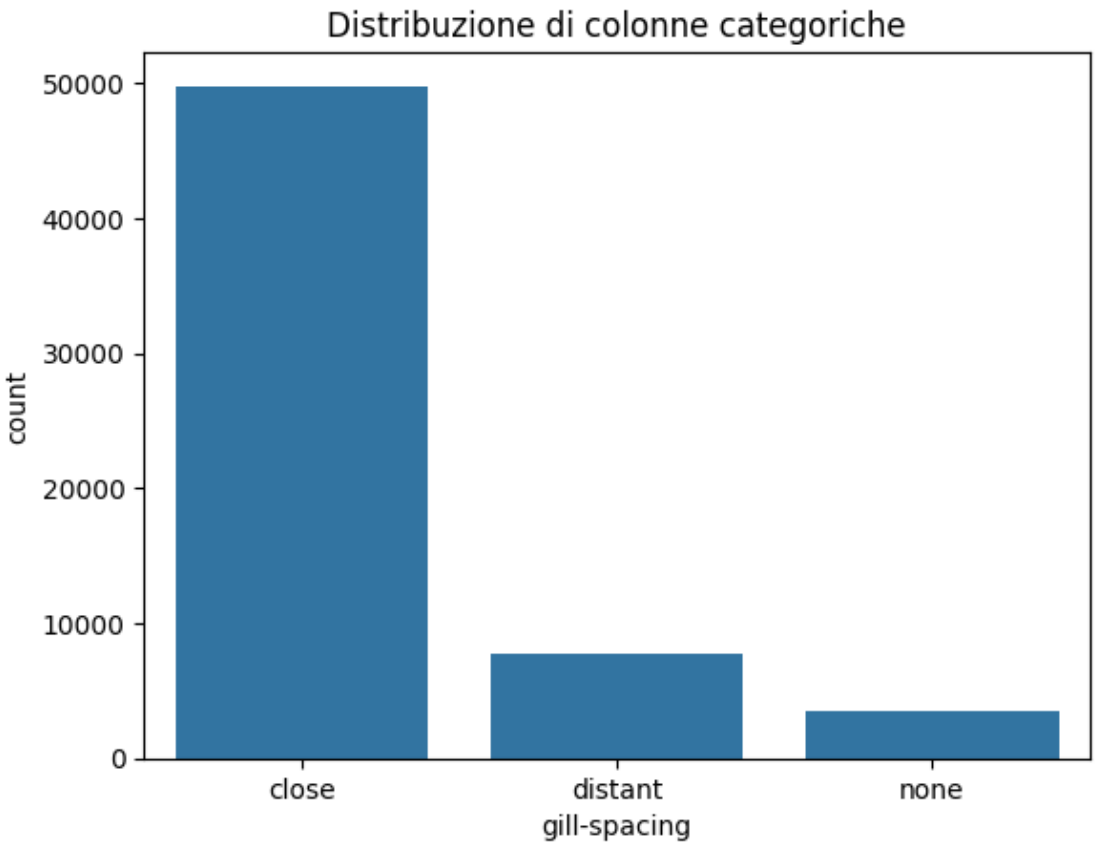

Distribuzione di colonne categoriche
count
25000
20000
15000
10000
5000
0
convex
flat
spherical
bell
conical
sunken
others
cap-shape

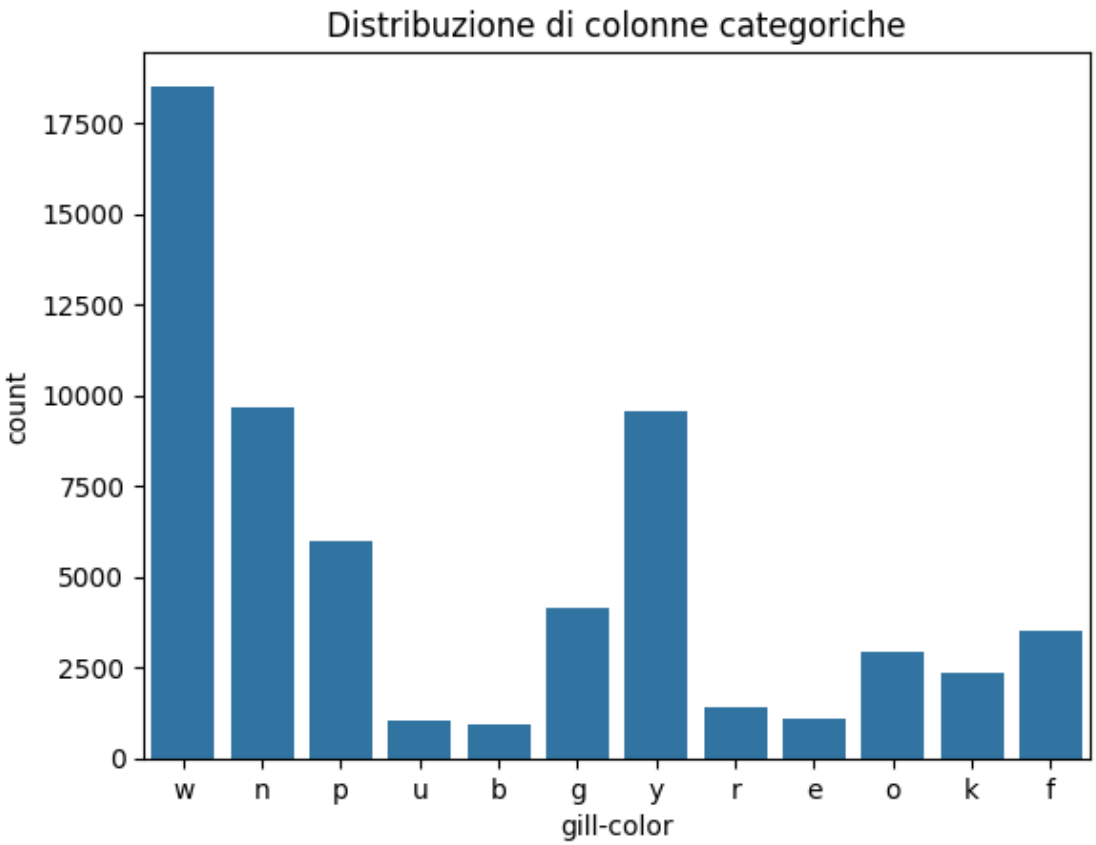

Distribuzione di colonne categoriche
count
20000
15000
10000
5000
0
groovesshiny
t
scaly
e
sunken
leathery
d
w
fibrous
silky
cap-surface

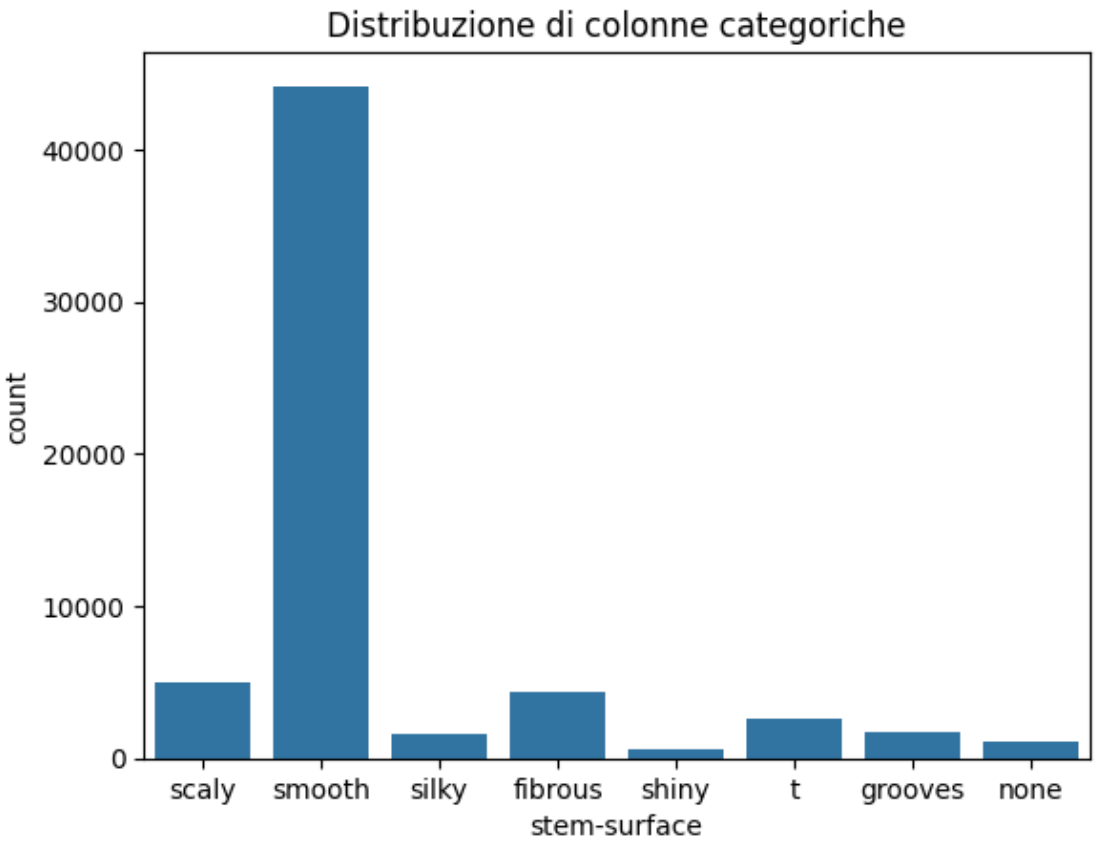

Distribuzione di colonne categoriche
25000
20000
15000
10000
5000
0
count
o
e
n
g
r
w
y
p
u
b
l
k
cap-color

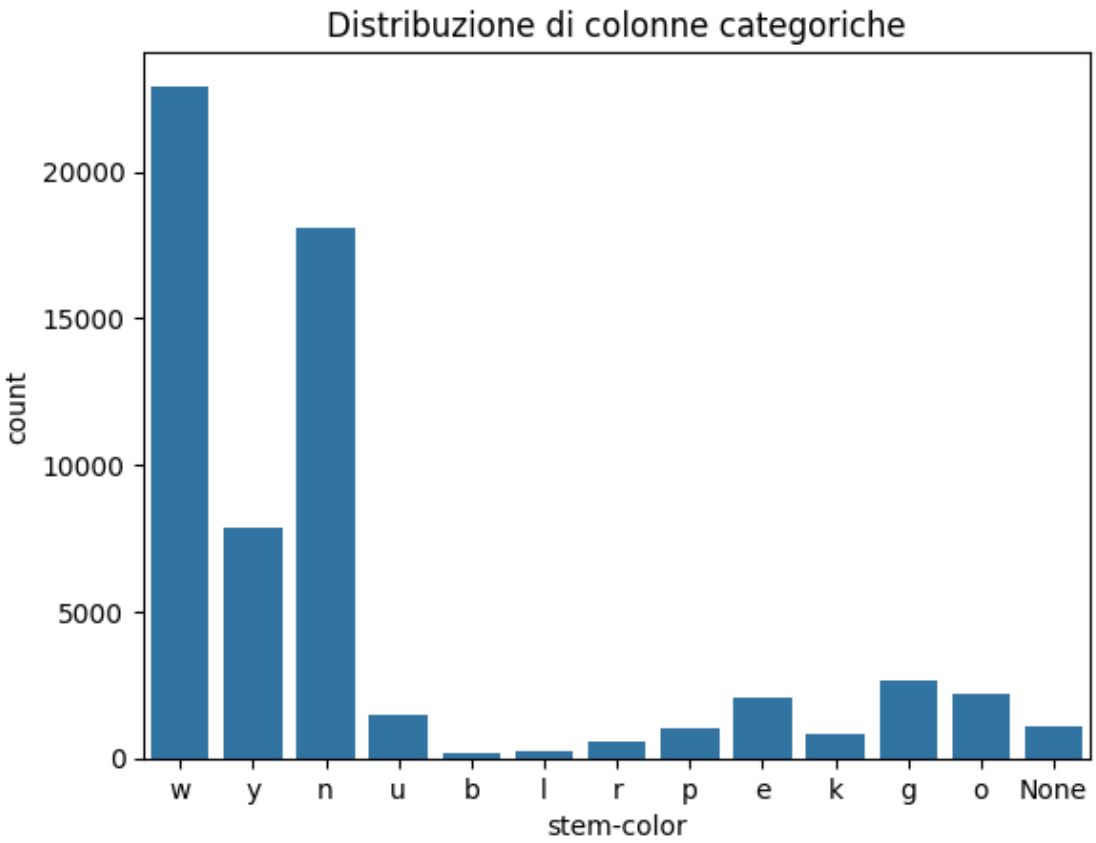

Distribuzione di colonne categoriche
50000
40000
30000
20000
10000
0
count
False
true
does-bruise-or-bleed

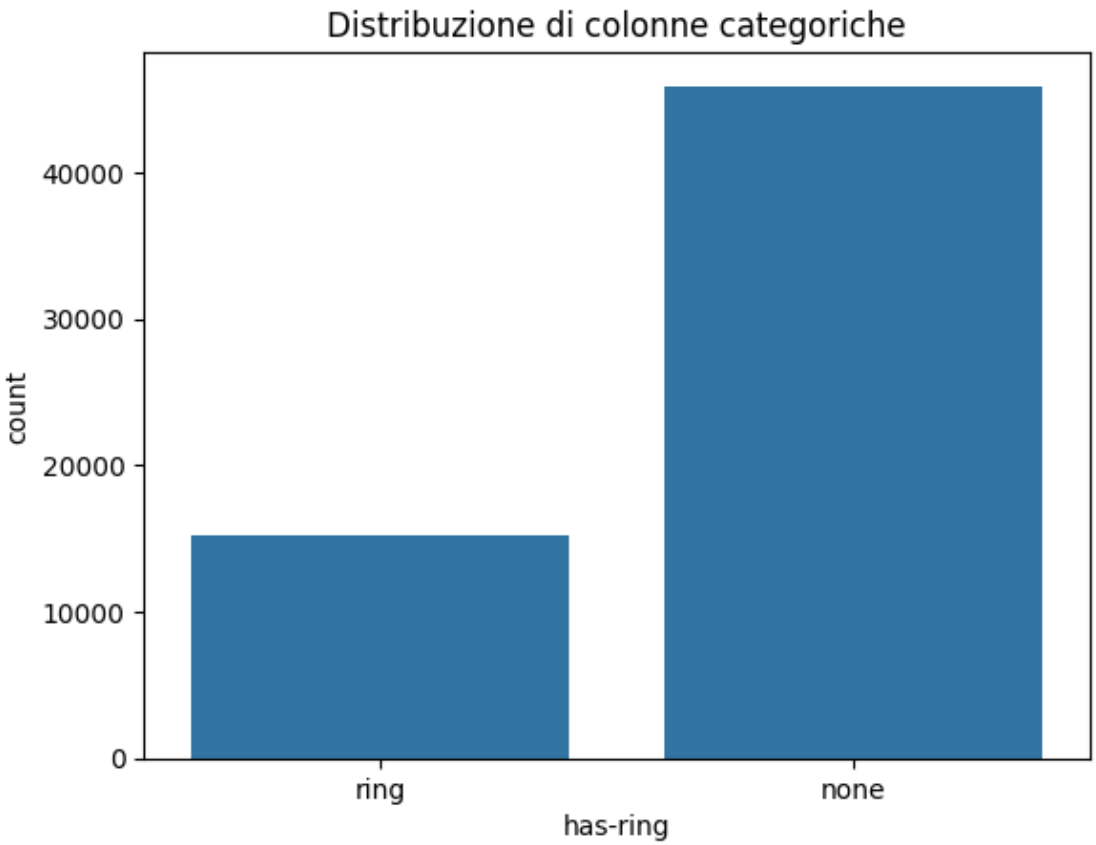

Distribuzione di colonne categoriche
count
20000
15000
10000
5000
0
free
adnate
decurrent
sinuate
adnexed
pores
none
gill-attachment

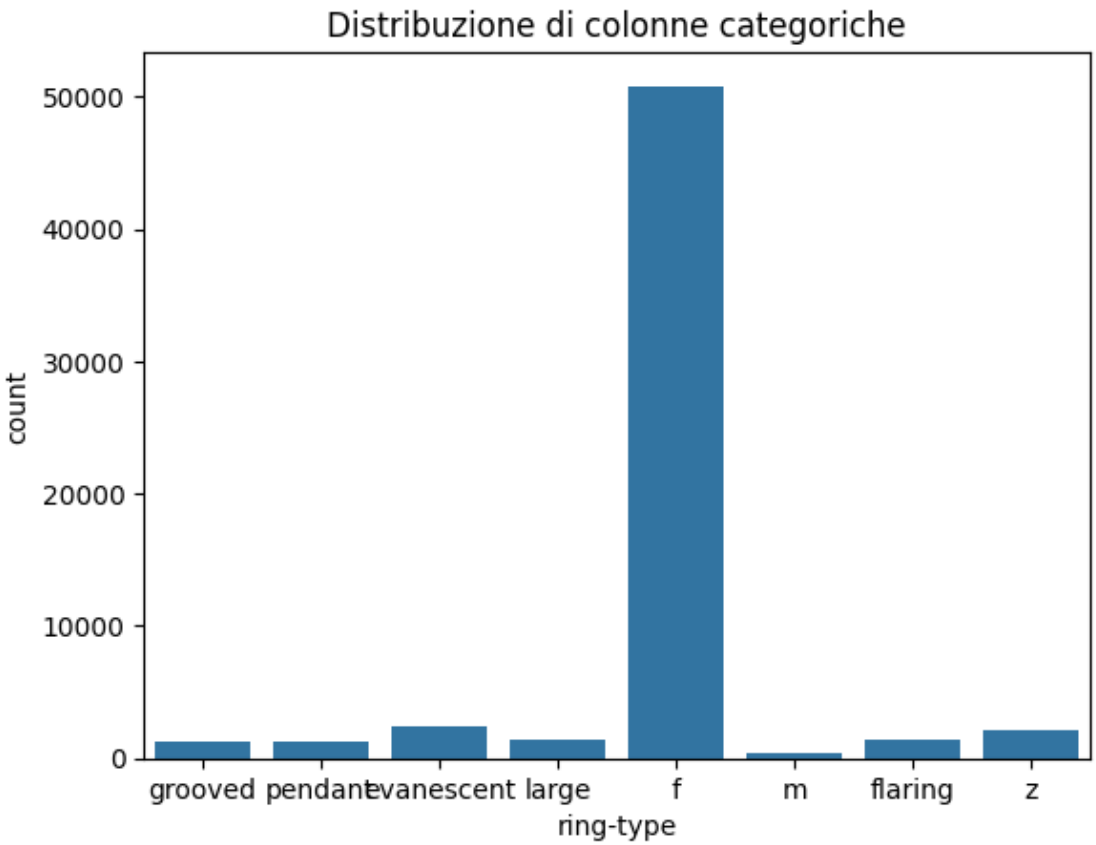

Distribuzione di colonne categoriche
50000
40000
30000
20000
10000
0
count
close
distant
none
gill-spacing

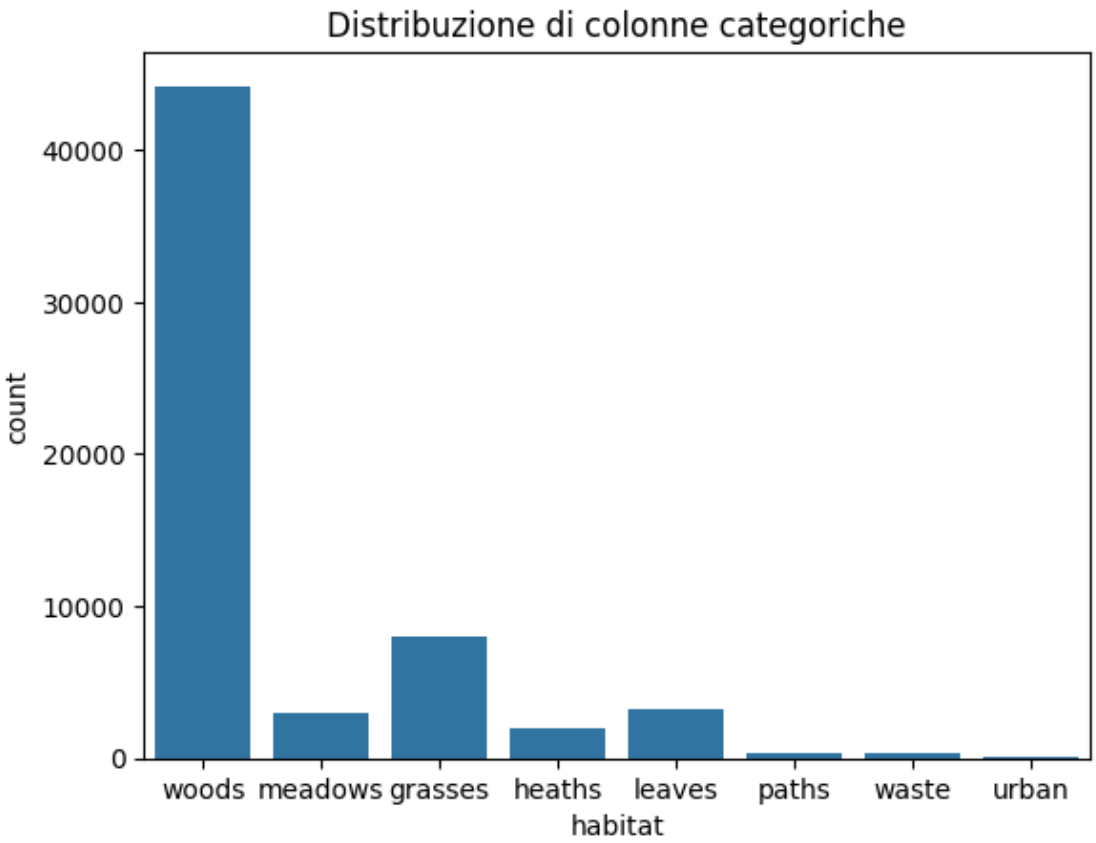

Distribuzione di colonne categoriche
count
17500
15000
12500
10000
7500
5000
2500
0
w
n
p
u
b
g
y
r
e
o
k
f
gill-color

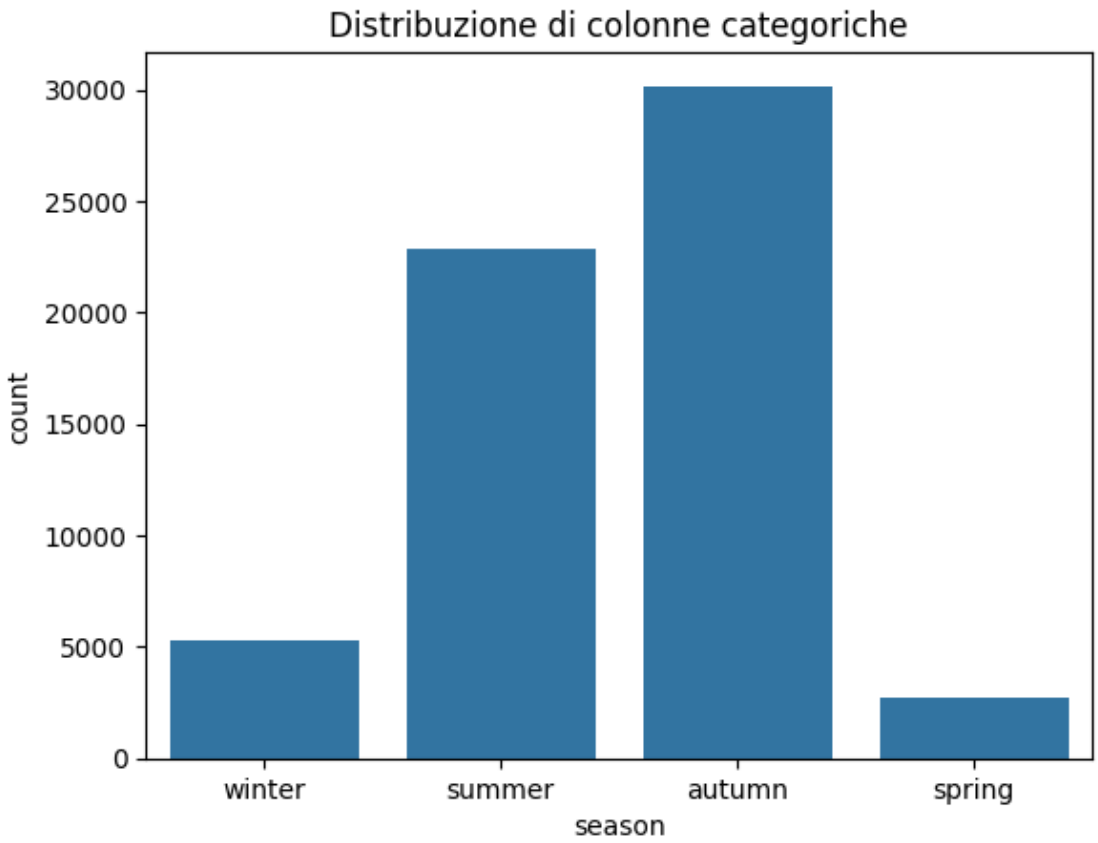

Distribuzione di colonne categoriche
count
40000
30000
20000
10000
0
scaly
smooth
silky
fibrous
shiny
t
grooves
none
stem-surface


Distribuzione di colonne categoriche
count
20000
15000
10000
5000
0
w
y
n
u
b
l
r
p
e
k
g
o
None
stem-color


Distribuzione di colonne categoriche
count
40000
30000
20000
10000
0
ring
none
has-ring


Distribuzione di colonne categoriche
count
50000
40000
30000
20000
10000
0
grooved
pendant
evanescent
large
f
m
flaring
z
ring-type


Distribuzione di colonne categoriche
count
40000
30000
20000
10000
0
woods
meadows
grasses
heaths
leaves
paths
waste
urban
habitat



2.1 l'unico fungo verde non velenoso del bosco?

Gli outliers sono punti dati che si discostano significativamente dal resto del set di dati. In altre parole, sono valori che sono molto diversi dagli altri valori nel dataset. Gli outliers possono influenzare negativamente alcune analisi statistiche, come la media e la deviazione standard, poiché possono distorcere le stime e le conclusioni.

Gli outliers possono essere causati da vari motivi, come errori nei dati, errori di misurazione, anomalie reali nel fenomeno studiato o semplicemente dalla casualità. È importante rilevare e gestire gli outliers in modo appropriato durante l'analisi dei dati per evitare interpretazioni errate o risultati distorti.

Per individuare e rimuovere gli outlier in ogni colonna del DataFrame e quindi eliminare le righe contenenti tali outlier, puoi seguire un approccio basato sulla deviazione standard

[ ]:
```
import pandas as pd

# Assume che il tuo DataFrame sia già definito come df2

# Seleziona solo le colonne numeriche
numeric_columns = df2.select_dtypes(include=['float64', 'int64'])
```

```
# Calcola il primo e terzo quartile per ogni colonna numerica
Q1 = numeric_columns.quantile(0.25)
Q3 = numeric_columns.quantile(0.75)

# Calcola l'IQR per ogni colonna numerica
IQR = Q3 - Q1

# Calcola i limiti per gli outlier utilizzando l'IQR (ad es. considerando outlier valori oltre 1.5 volte l'IQR sopra il terzo quartile e sotto il primo quartile)
outlier_limits = (Q3 + 1.5 * IQR, Q1 - 1.5 * IQR)

# Trova gli outlier per ogni colonna numerica
outliers = ((numeric_columns < outlier_limits[1]) | (numeric_columns > outlier_limits[0])).any(axis=1)

# Elimina le righe contenenti outlier solo dalle colonne numeriche
df2_cleaned = df2[~outliers]

df2_cleaned
```

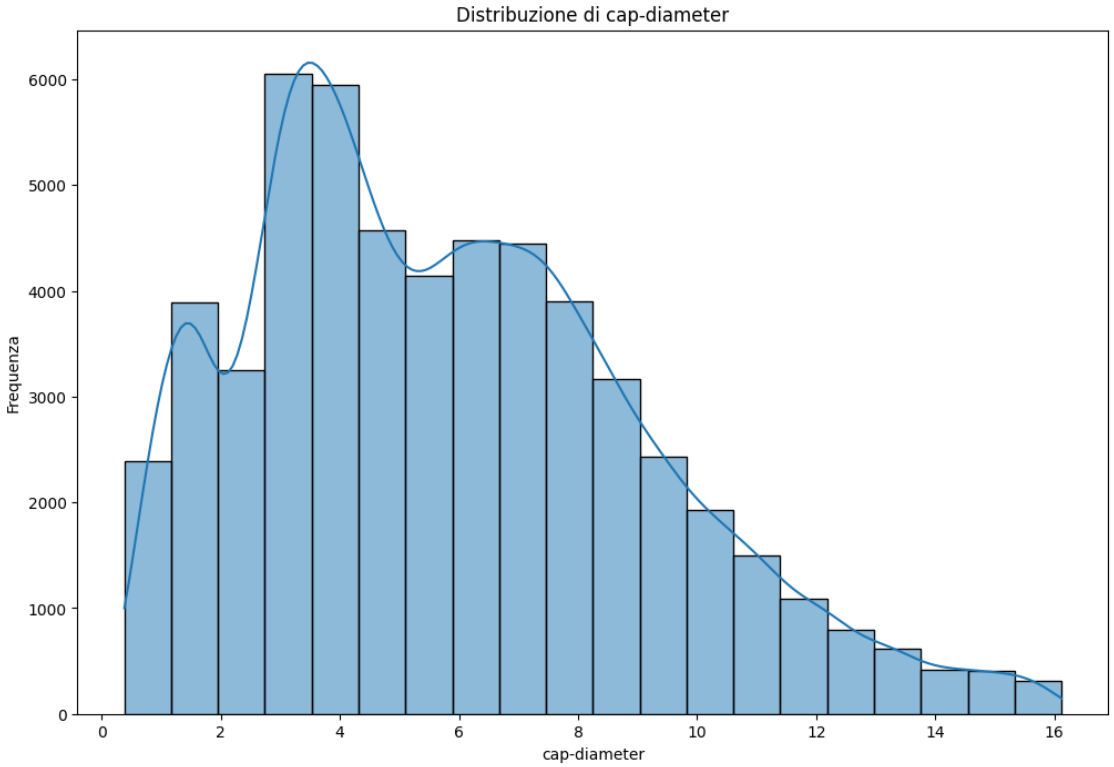
[ ]:

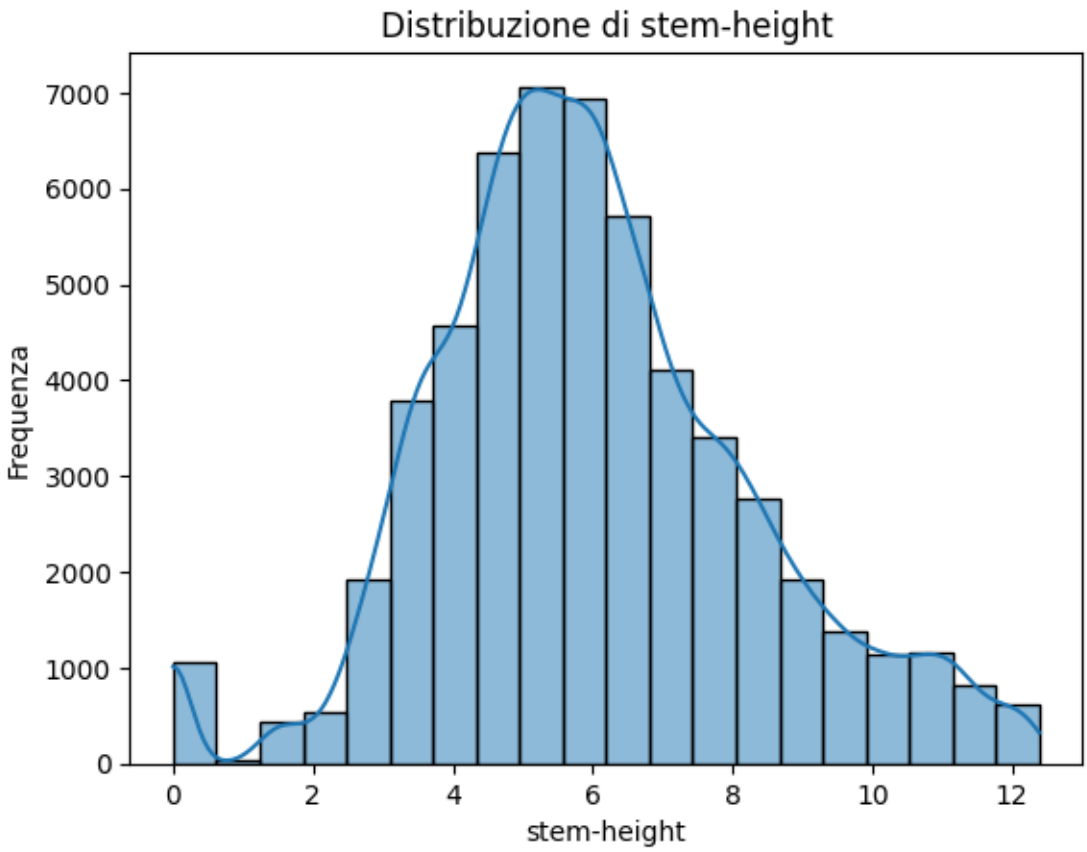
	class	cap-diameter	cap-shape	cap-surface	cap-color	\
353	poisonus	6.87	convex	grooves	n	
354	poisonus	8.59	spherical	grooves	n	
355	poisonus	5.95	spherical	grooves	n	
356	poisonus	6.51	convex	grooves	n	
357	poisonus	7.66	convex	grooves	n	
…	…	…	…	…	…	
61064	poisonus	1.18	sunken	sunken	y	
61065	poisonus	1.27	flat	sunken	y	
61066	poisonus	1.27	sunken	sunken	y	
61067	poisonus	1.24	flat	sunken	y	
61068	poisonus	1.17	sunken	sunken	y	

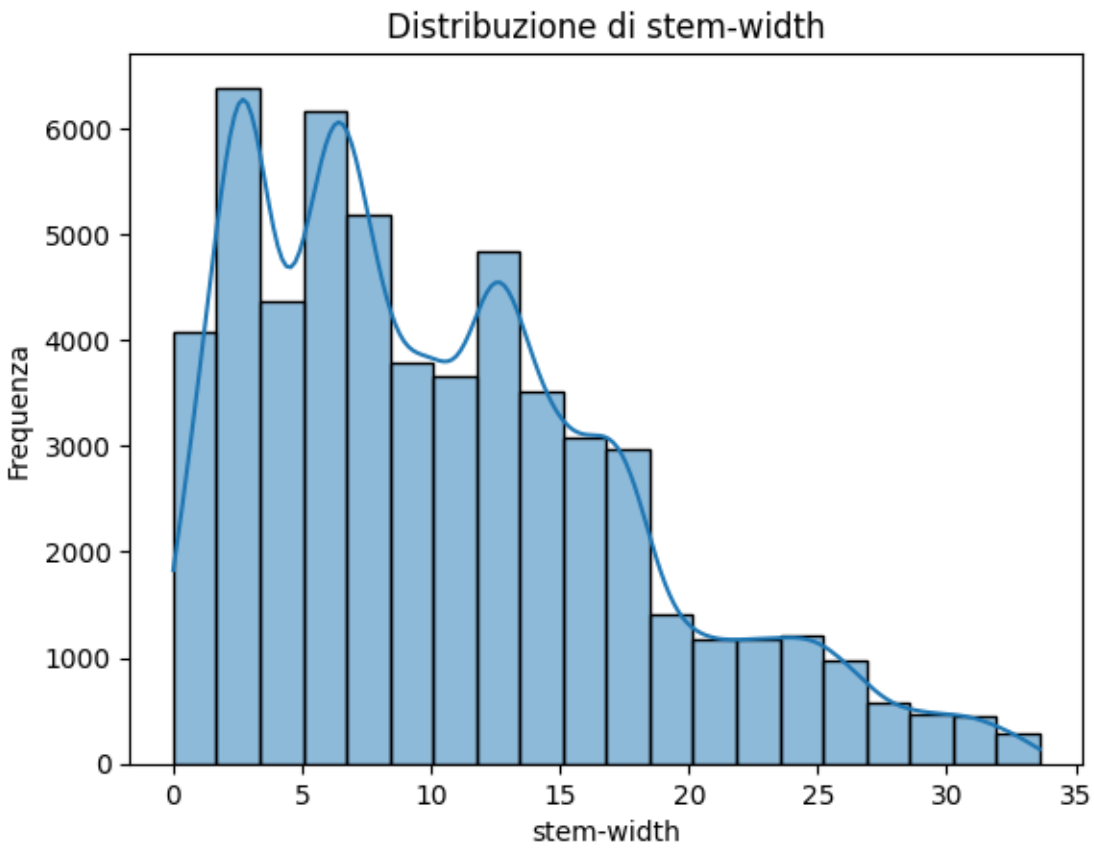
	does-bruise-or-bleed	gill-attachment	gill-spacing	gill-color	\
353	False	free	close	w	
354	False	free	close	w	
355	False	free	close	w	
356	False	free	close	w	
357	False	free	close	w	
…	…	…	…	…	
61064	False	none	none	f	
61065	False	none	none	f	
61066	False	none	none	f	
61067	False	none	none	f	
61068	False	none	none	f	

```
       stem-height  stem-width stem-surface stem-color has-ring ring-type  \
353           6.88       13.64        scaly          w     ring   pendant
354           9.15       17.34        scaly          w     ring   pendant
355           7.54       12.73        scaly          w     ring   pendant
356           6.80       12.92        scaly          w     ring   pendant
357           8.55       14.98        scaly          w     ring   pendant
…              …          …           …        …      …       …
61064         3.93        6.22       smooth          y     none         f
61065         3.18        5.43       smooth          y     none         f
61066         3.86        6.37       smooth          y     none         f
61067         3.56        5.44       smooth          y     none         f
61068         3.25        5.45       smooth          y     none         f

      habitat  season
353     woods  autumn
354     woods  autumn
355     woods  summer
356     woods  autumn
357     woods  autumn
…       …      …
61064   woods  autumn
61065   woods  autumn
61066   woods  summer
61067   woods  summer
61068   woods  summer

[55729 rows x 17 columns]
```

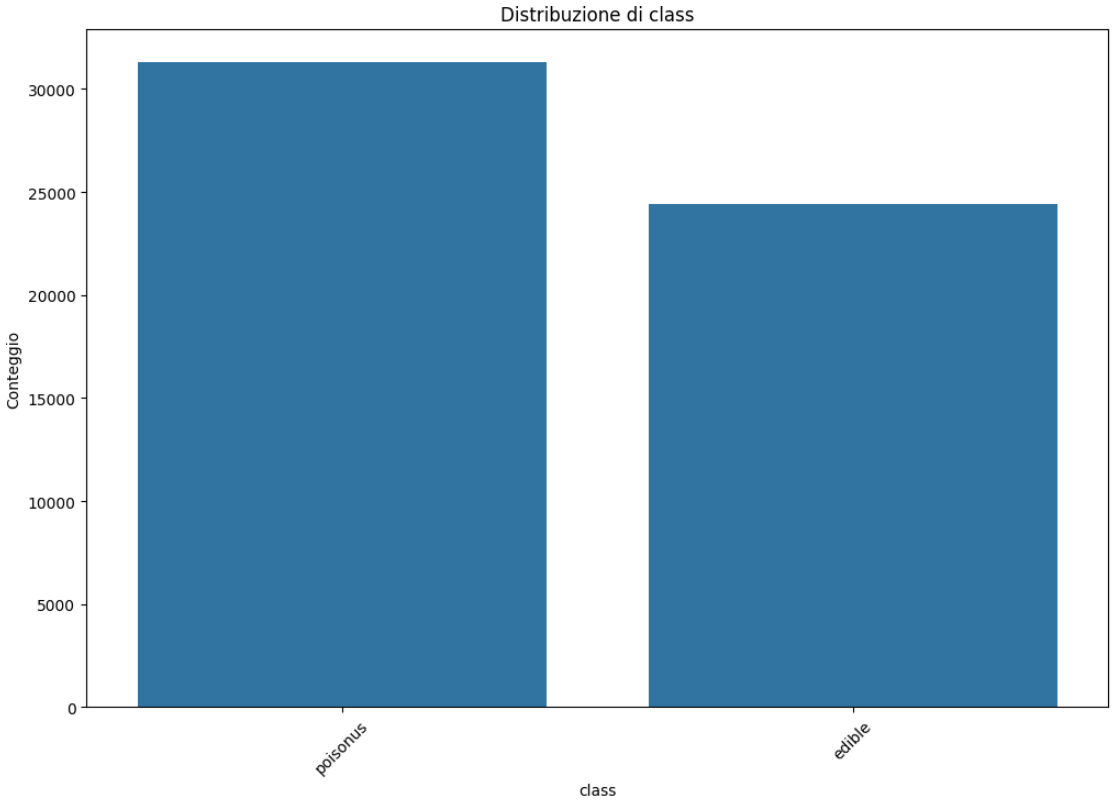
[ ]:
```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Assume che il tuo nuovo DataFrame sia già definito come df2_cleaned

# Seleziona solo le colonne numeriche
numeric_columns = df2_cleaned.select_dtypes(include=['float64', 'int64'])

# Visualizzazione della distribuzione delle colonne numeriche utilizzando un␣
 ↪istogramma per ciascuna colonna
plt.figure(figsize=(12, 8))
for col in numeric_columns.columns:
    sns.histplot(data=df2_cleaned, x=col, kde=True, bins=20)
    plt.title(f'Distribuzione di {col}')
    plt.xlabel(col)
    plt.ylabel('Frequenza')
    plt.show()
```

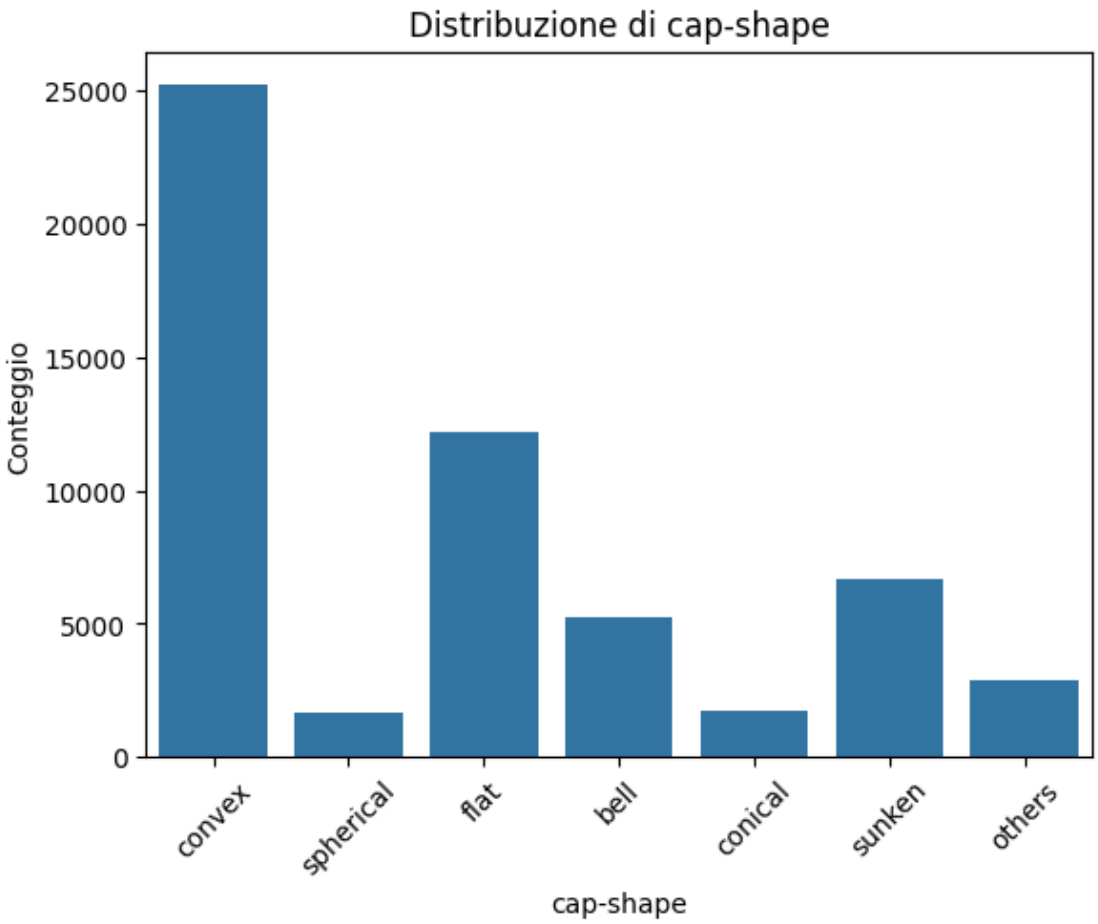

Distribuzione di cap-diameter
Frequenza
cap-diameter
0
1000
2000
3000
4000
5000
6000
0
2
4
6
8
10
12
14
16

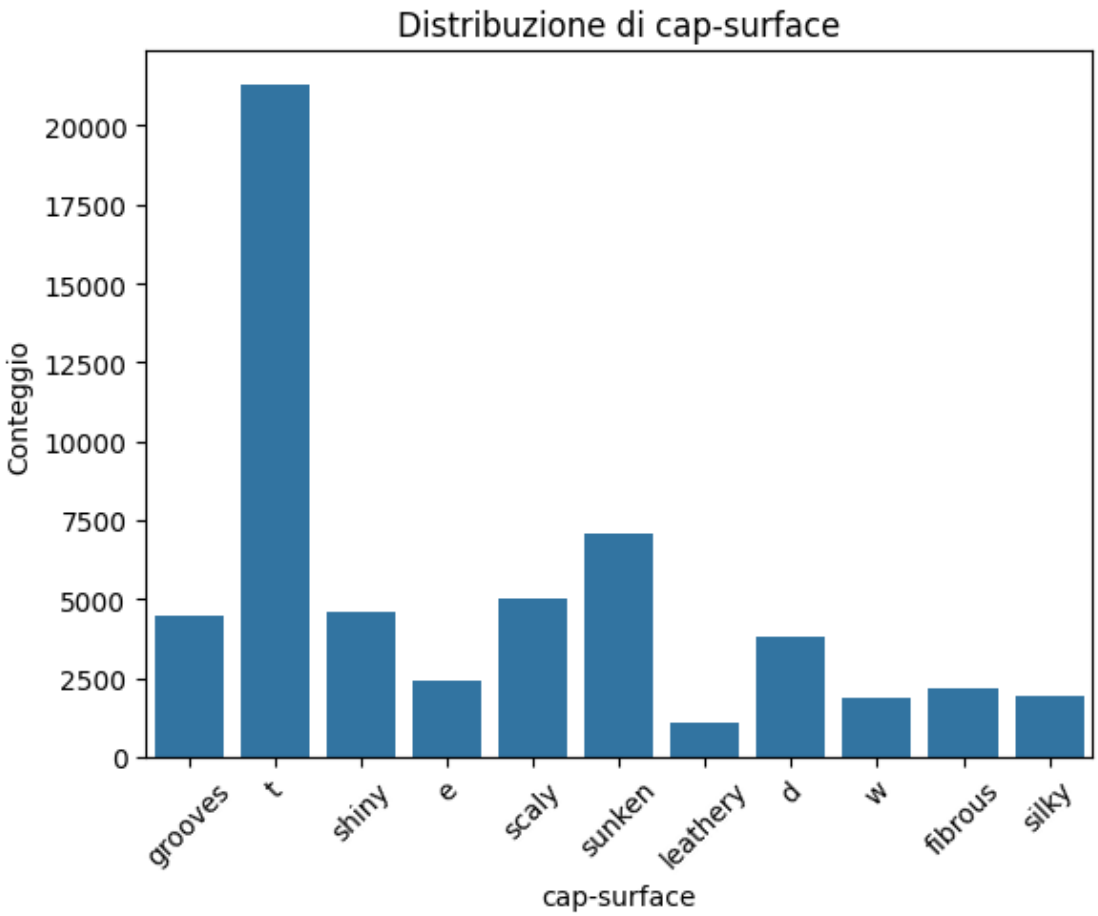

Distribuzione di stem-height
Frequenza
7000
6000
5000
4000
3000
2000
1000
0
0
2
4
6
8
10
12
stem-height

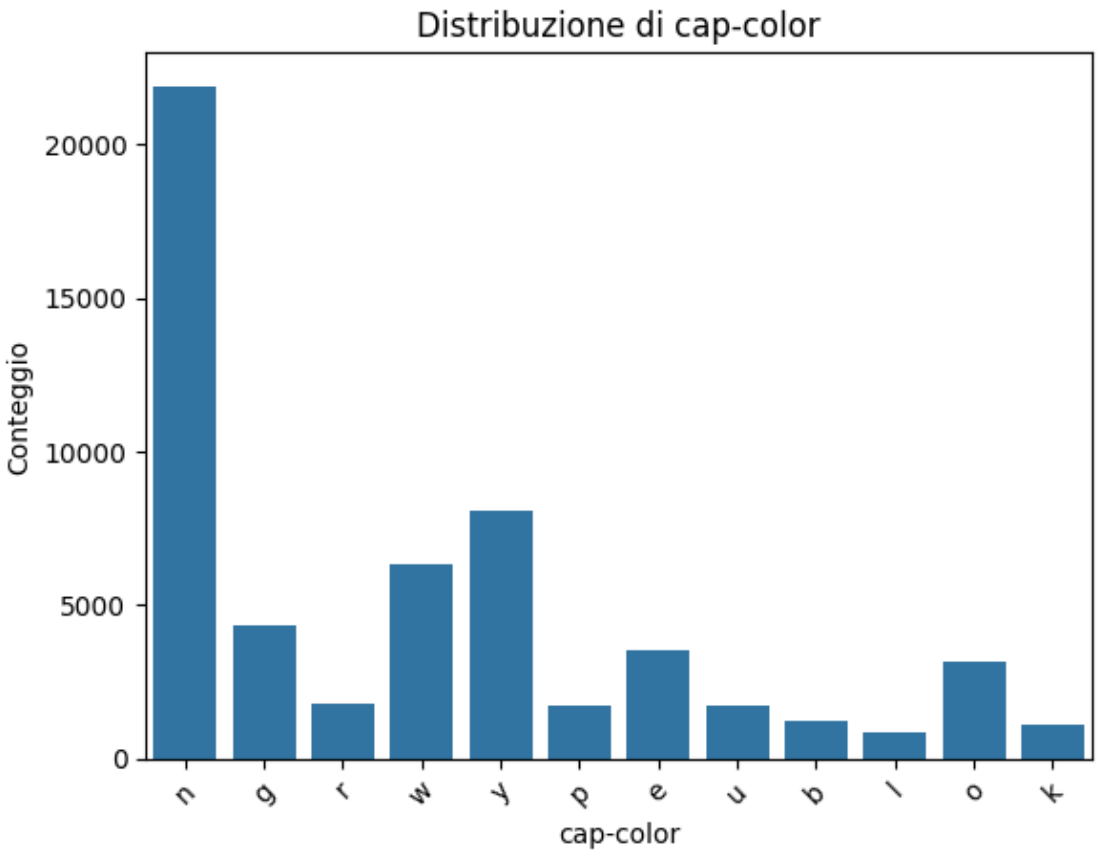


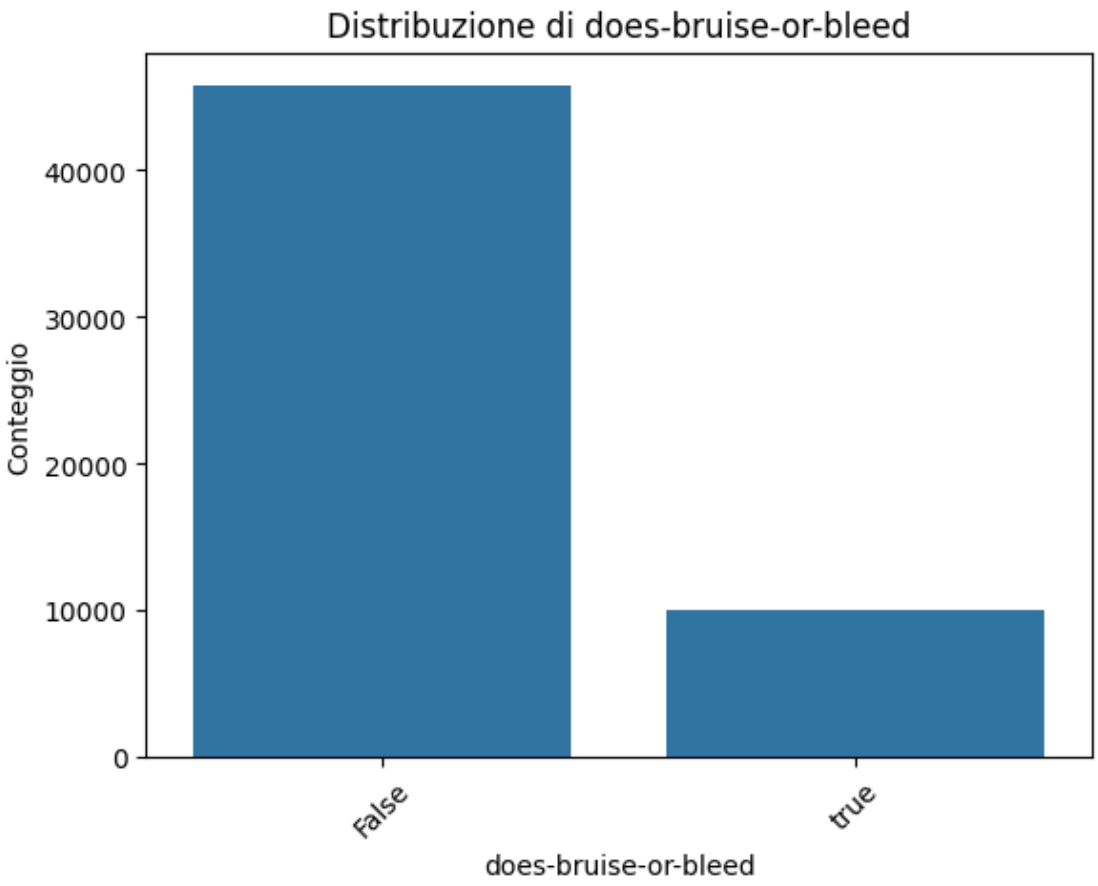
[ ]:
```
# Seleziona solo le colonne categoriche
categorical_columns = df2_cleaned.select_dtypes(include=['object'])

# Visualizzazione delle distribuzioni delle colonne categoriche utilizzando␣
 ↪grafici a barre
plt.figure(figsize=(12, 8))
for col in categorical_columns.columns:
    sns.countplot(data=df2_cleaned, x=col)
    plt.title(f'Distribuzione di {col}')
    plt.xlabel(col)
    plt.ylabel('Conteggio')
    plt.xticks(rotation=45)
    plt.show()
```

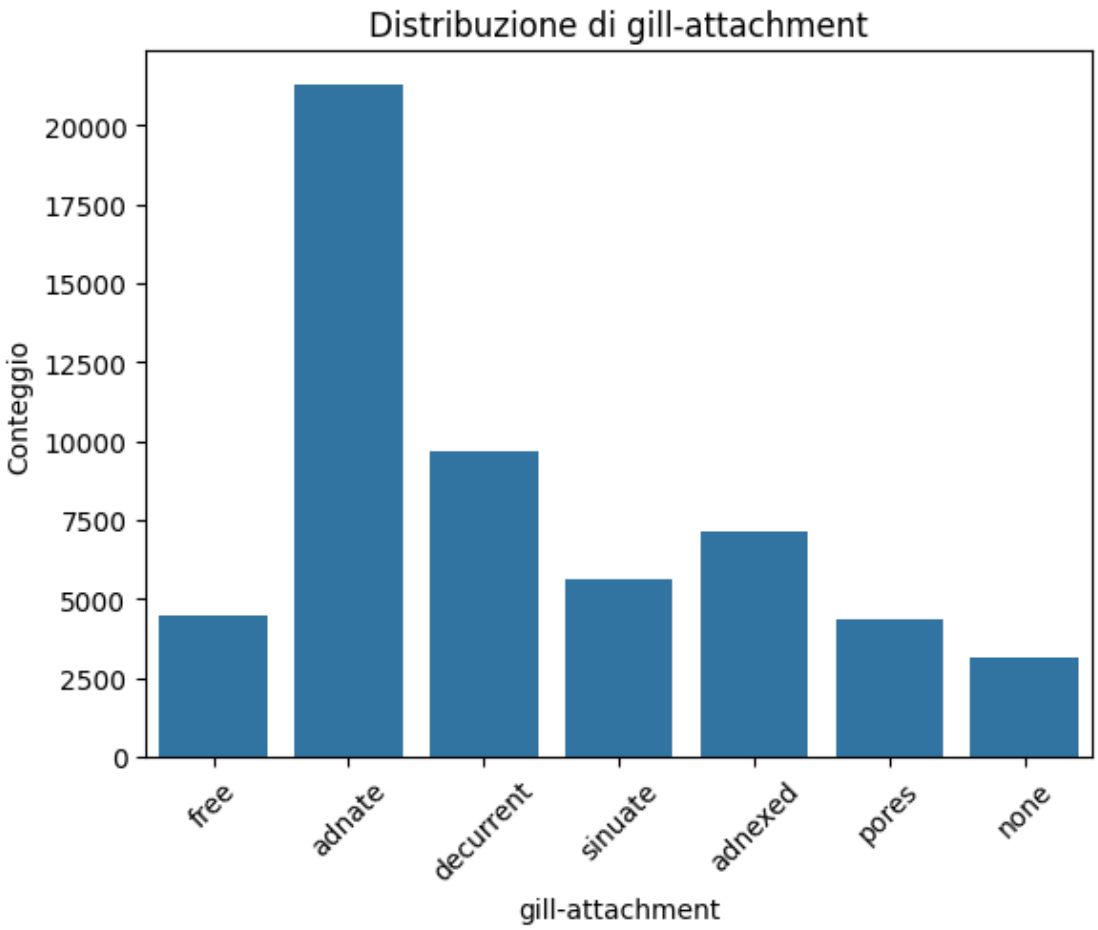

Distribuzione di class
Conteggio
30000
25000
20000
15000
10000
5000
0
poisonus
edible
class

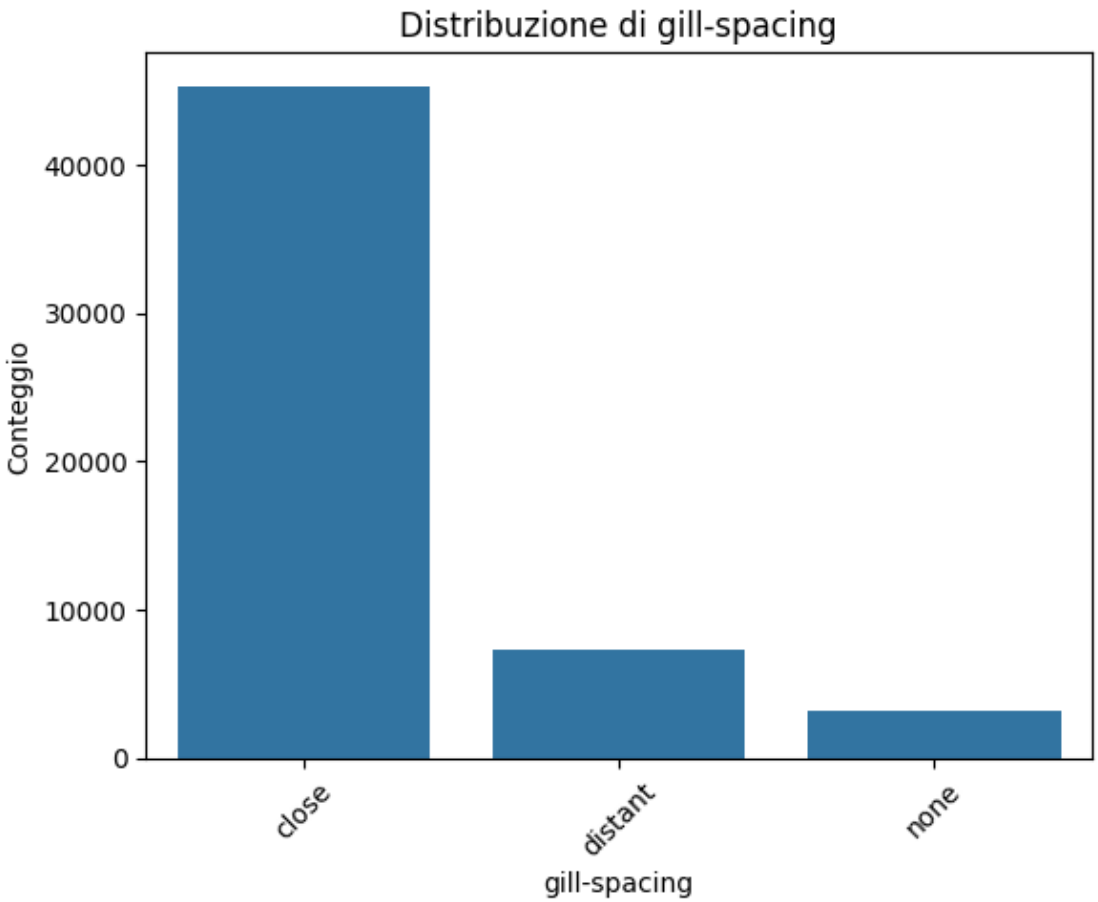

Distribuzione di cap-shape
Conteggio
25000
20000
15000
10000
5000
0
convex
spherical
flat
bell
conical
sunken
others
cap-shape

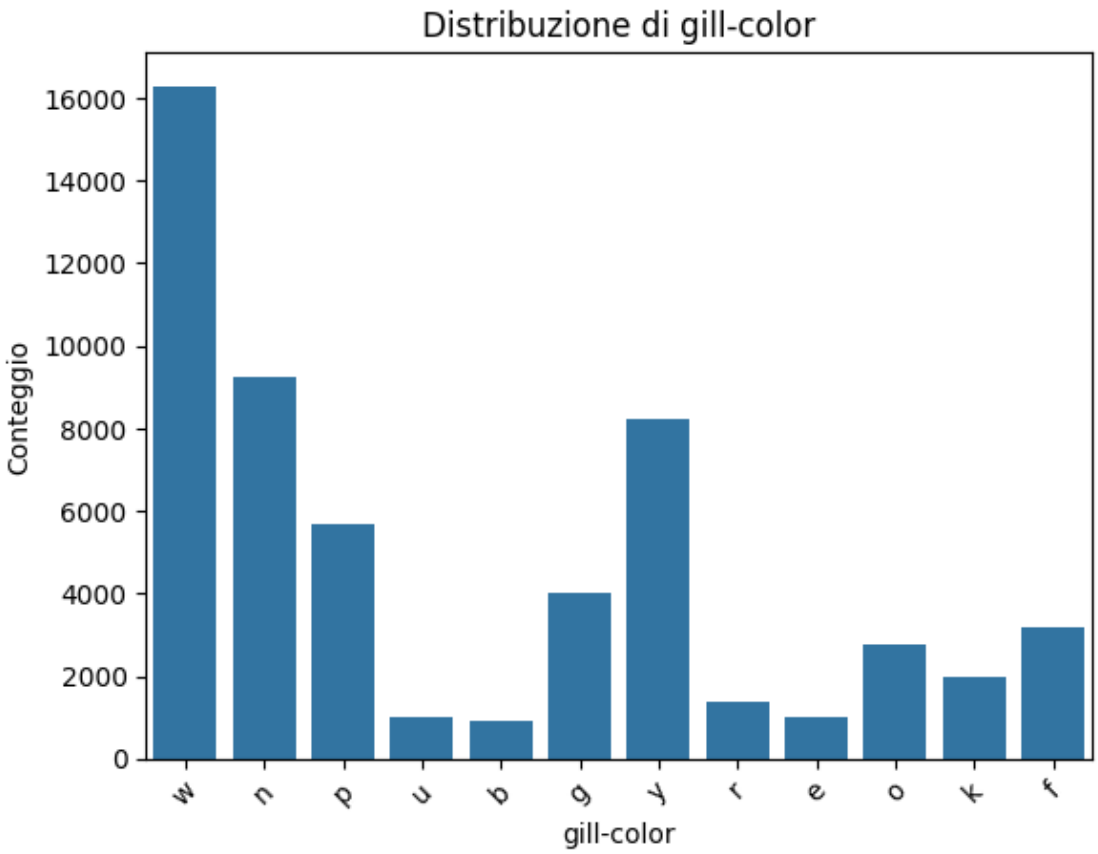

Distribuzione di cap-surface
Conteggio
20000
17500
15000
12500
10000
7500
5000
2500
0
grooves
t
shiny
e
scaly
sunken
leathery
d
w
fibrous
silky
cap-surface

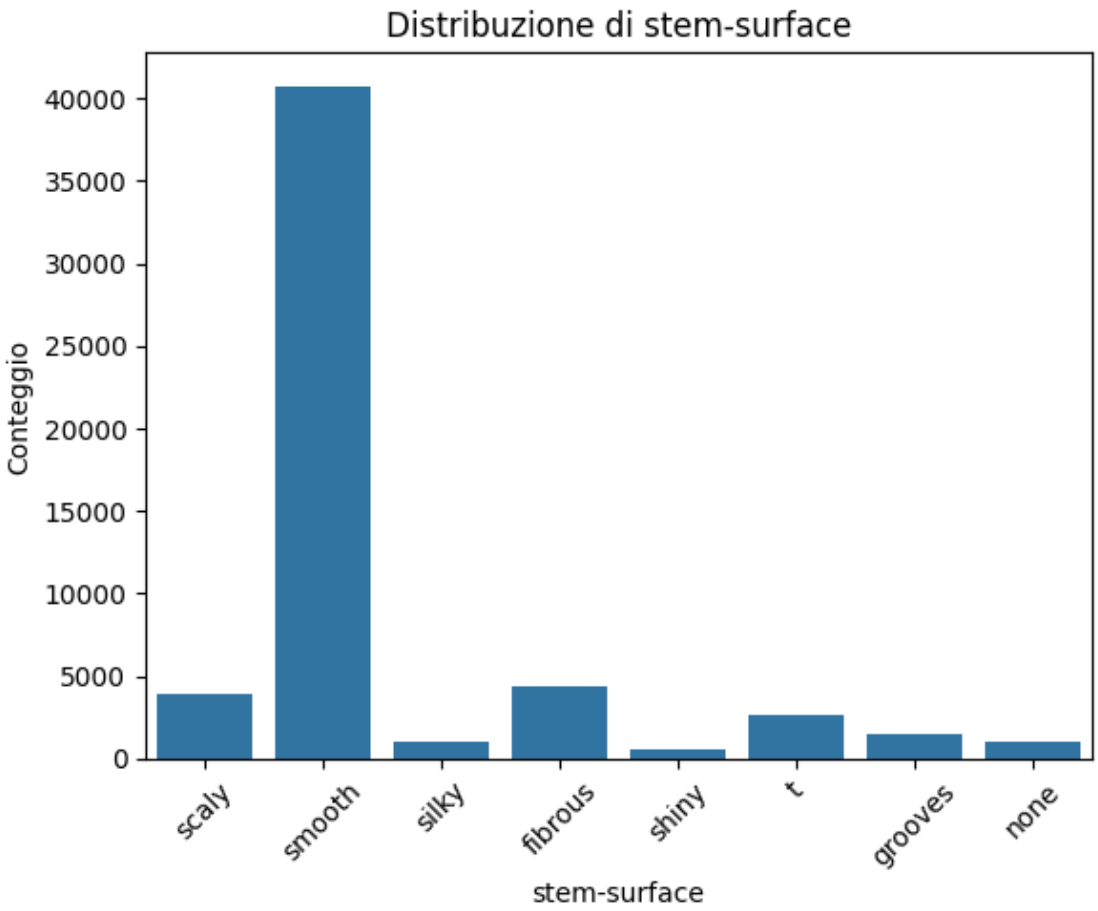

Distribuzione di cap-color
Conteggio
20000
15000
10000
5000
0
n
g
r
w
y
p
e
u
b
l
o
k
cap-color

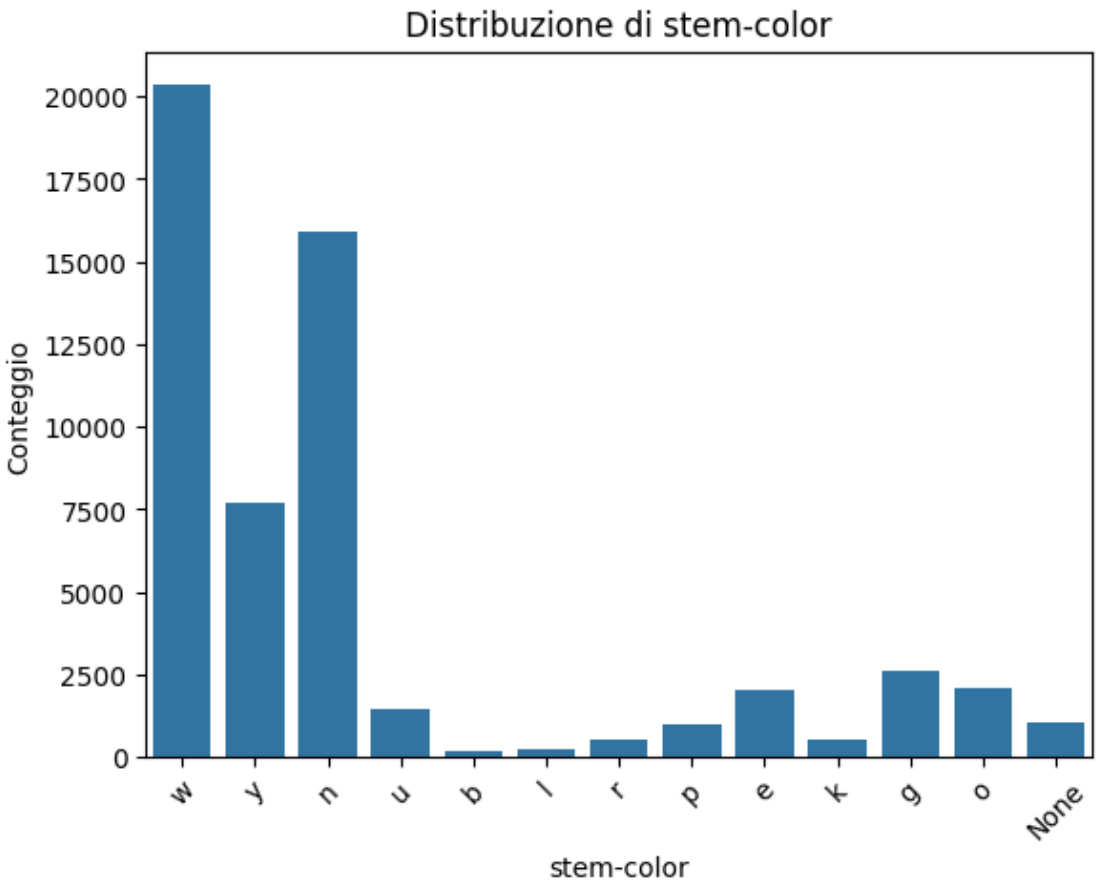

Distribuzione di does-bruise-or-bleed
Conteggio
40000
30000
20000
10000
0
False
true
does-bruise-or-bleed

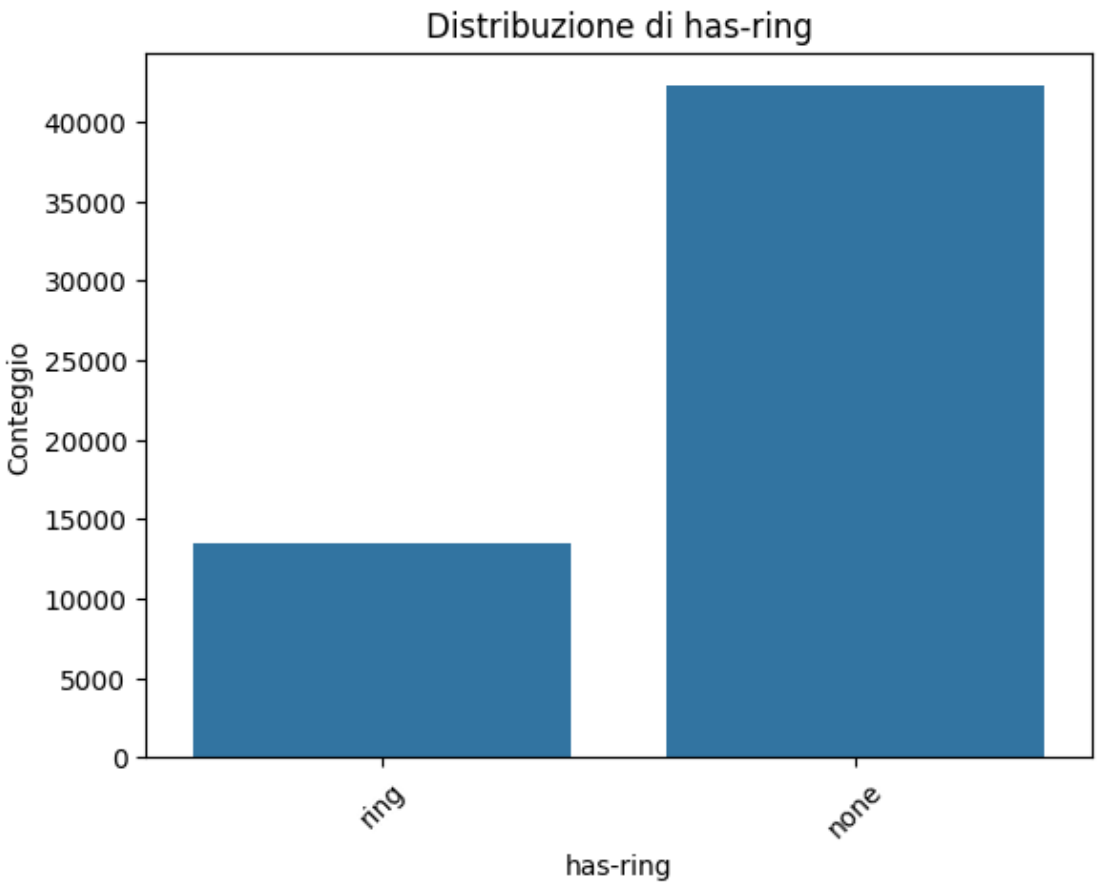

Distribuzione di gill-attachment
Conteggio
0
2500
5000
7500
10000
12500
15000
17500
20000
free
adnate
decurrent
sinuate
adnexed
pores
none
gill-attachment

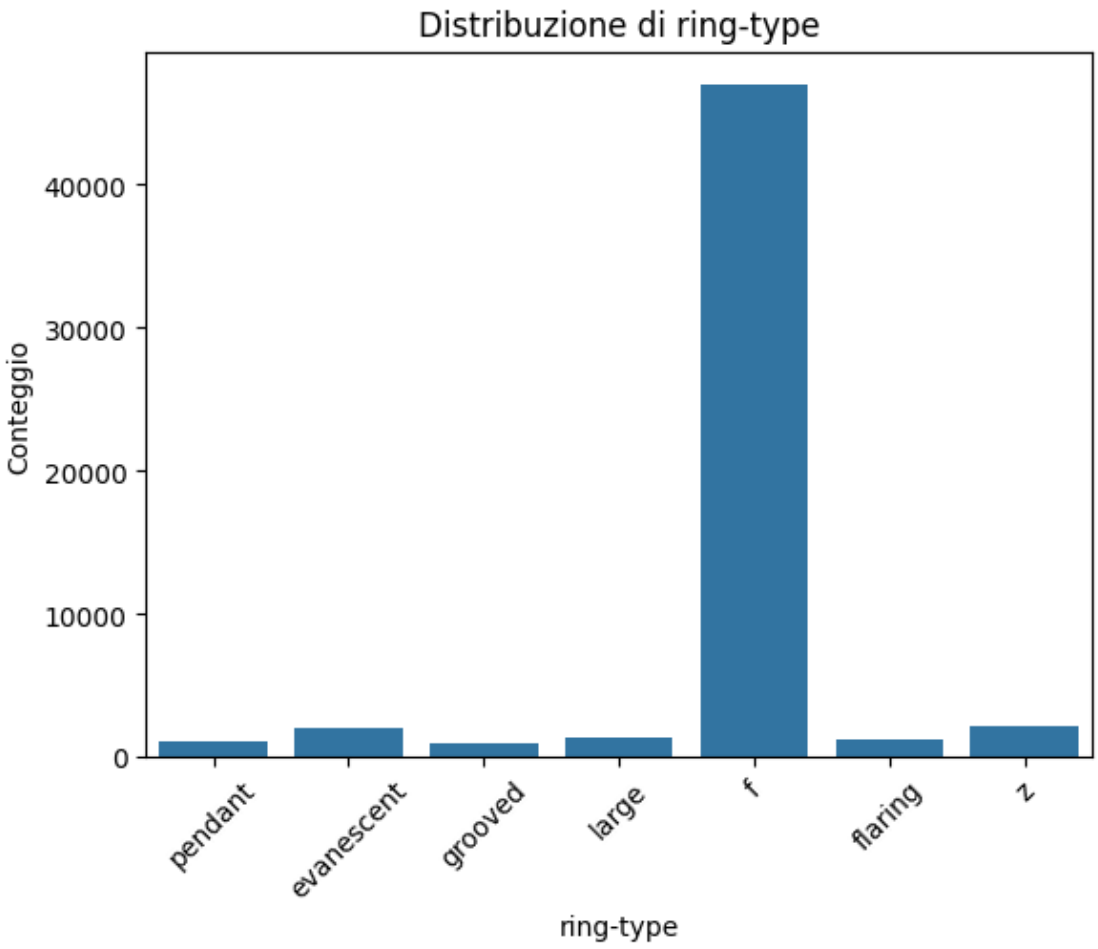

Distribuzione di gill-spacing
Conteggio
40000
30000
20000
10000
0
close
distant
none
gill-spacing

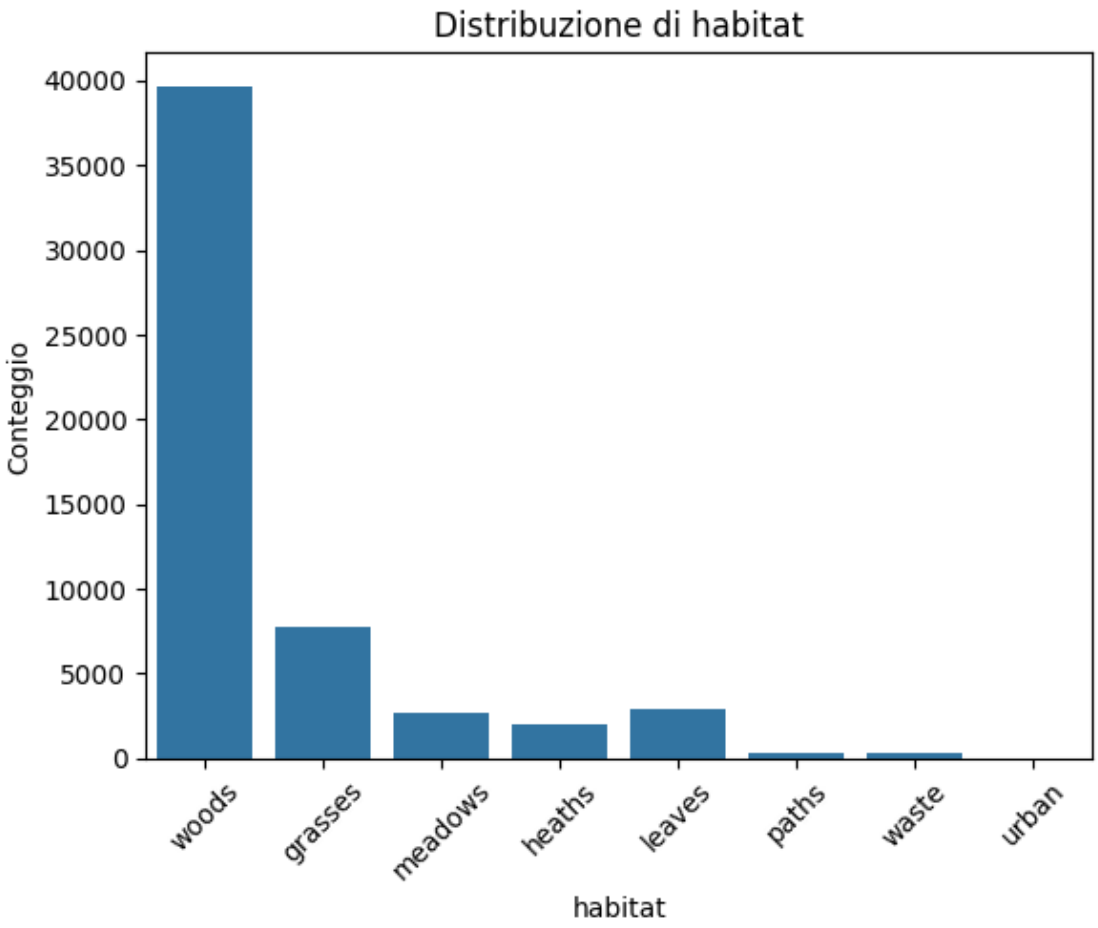

Distribuzione di gill-color
Conteggio
16000
14000
12000
10000
8000
6000
4000
2000
0
w
n
p
u
b
g
y
r
e
o
k
f
gill-color

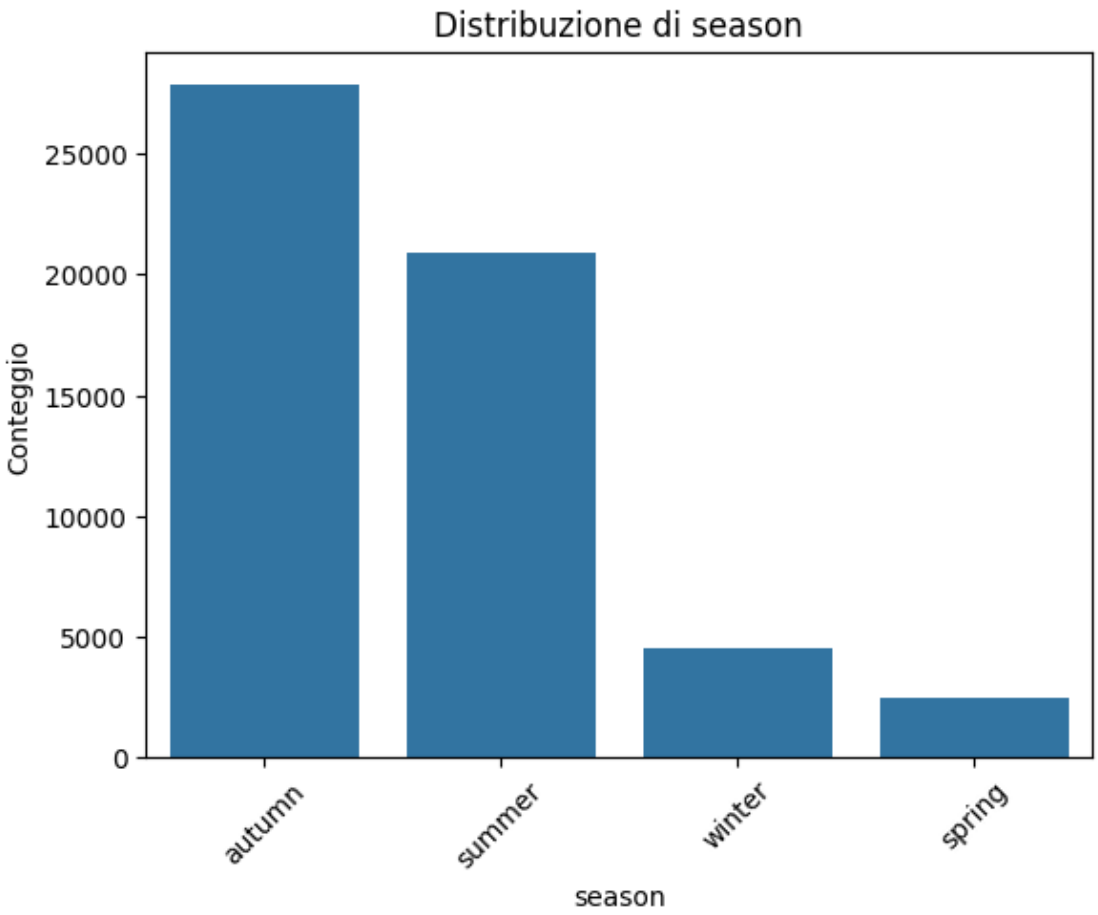

Distribuzione di stem-surface
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
scaly
smooth
silky
fibrous
shiny
t
grooves
none
stem-surface


Distribuzione di stem-color
Conteggio
20000
17500
15000
12500
10000
7500
5000
2500
0
w
y
n
u
b
l
r
p
e
k
g
o
None
stem-color


Distribuzione di has-ring
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
ring
none
has-ring


Distribuzione di ring-type
Conteggio
40000
30000
20000
10000
0
pendant
evanescent
grooved
large
f
flaring
z
ring-type


Distribuzione di habitat
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
woods
grasses
meadows
heaths
leaves
paths
waste
urban
habitat



2.2 I computer non sanno leggere però...

dobbiamo preparare i dati per darli ad un intelligenza artificiale

[ ]:
```
from sklearn.preprocessing import StandardScaler

# Seleziona solo le colonne numeriche nel DataFrame
numeric_columns = df2_cleaned.select_dtypes(include=['float64', 'int64'])

# Inizializza lo StandardScaler
scaler = StandardScaler()

# Applica lo scaling alle colonne numeriche
scaled_numeric_columns = scaler.fit_transform(numeric_columns)

# Crea un nuovo DataFrame con le colonne numeriche scalate
df2_scaled = pd.DataFrame(scaled_numeric_columns, columns=numeric_columns.
 ↪columns)
```

```
# Visualizza il nuovo DataFrame scalato
df2_scaled
```

[ ]:

	cap-diameter	stem-height	stem-width
0	0.305815	0.387899	0.423194
1	0.828934	1.374421	0.927539
2	0.026007	0.674729	0.299153
3	0.196325	0.353132	0.325052
4	0.546085	1.113667	0.605849
…	…	…	…
55724	-1.424737	-0.894145	-0.588220
55725	-1.397365	-1.220089	-0.695905
55726	-1.397365	-0.924567	-0.567774
55727	-1.406489	-1.054944	-0.694542
55728	-1.427779	-1.189667	-0.693179

[55729 rows x 3 columns]

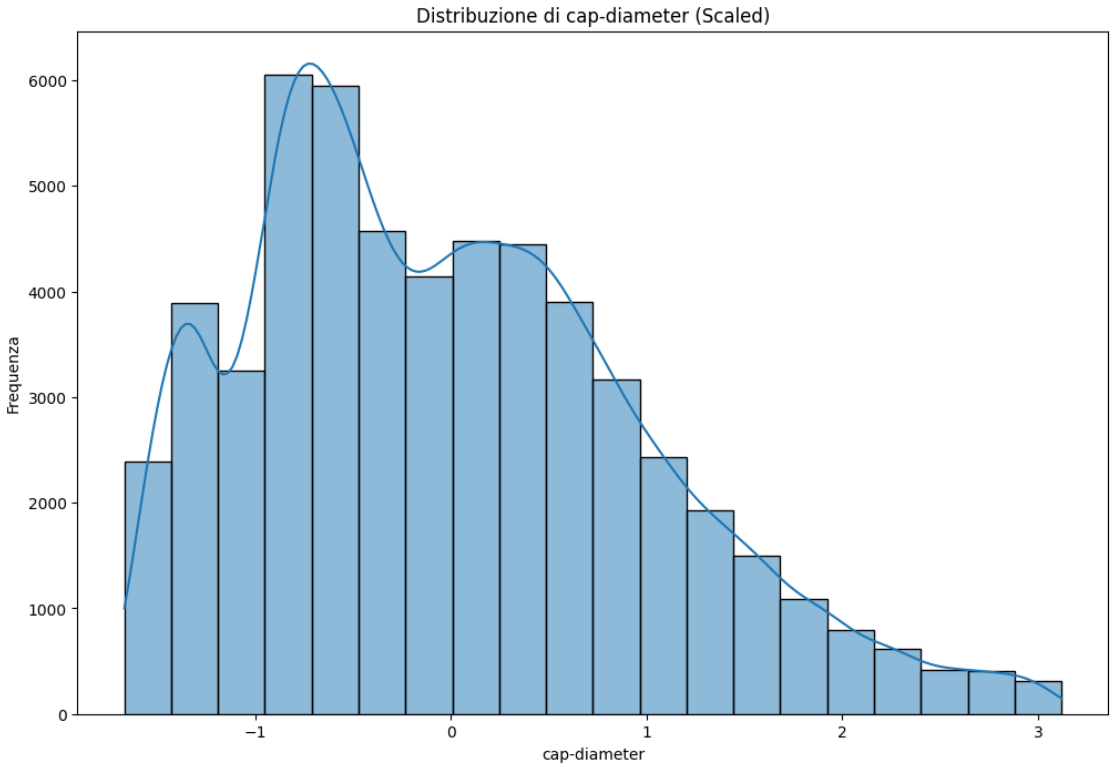
[ ]:
```
from sklearn.preprocessing import OneHotEncoder

# Seleziona solo le colonne categoriche nel DataFrame
categorical_columns = df2_cleaned.select_dtypes(include=['object'])

# Inizializza OneHotEncoder
encoder = OneHotEncoder()

# Applica l'encoding alle colonne categoriche e trasforma i dati
encoded_categorical_columns = encoder.fit_transform(categorical_columns)

# Ottieni i nomi delle categorie dall'encoder
encoded_categories = encoder.categories_

# Crea i nomi delle nuove colonne dopo l'encoding
encoded_column_names = []
for i, col in enumerate(categorical_columns.columns):
    encoded_column_names.extend([f"{col}_{category}" for category in␣
 ↪encoded_categories[i]])

# Crea un nuovo DataFrame con le colonne categoriche codificate e i nomi delle␣
 ↪colonne
df2_encoded = pd.DataFrame(encoded_categorical_columns.toarray(),␣
 ↪columns=encoded_column_names)

# Visualizza il nuovo DataFrame codificato
print(df2_encoded)
```

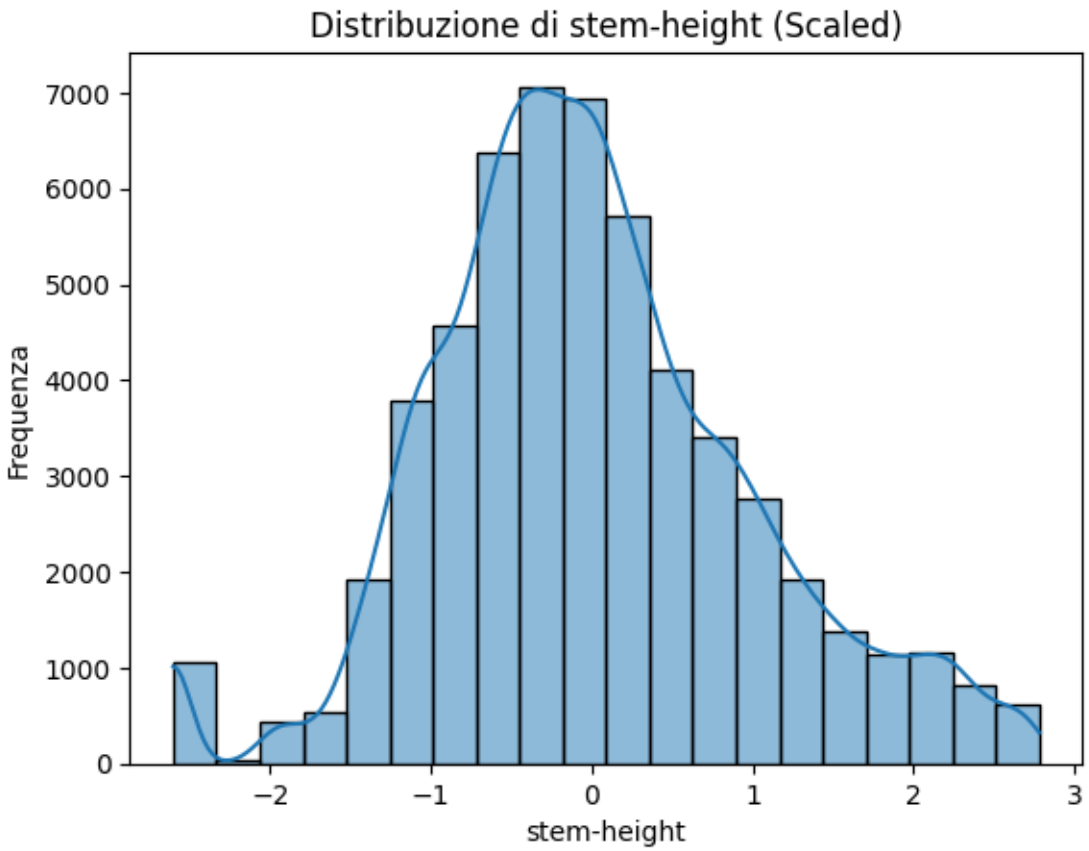
```
       class_edible  class_poisonus  cap-shape_bell  cap-shape_conical  \
0               0.0             1.0             0.0                0.0
1               0.0             1.0             0.0                0.0
2               0.0             1.0             0.0                0.0
3               0.0             1.0             0.0                0.0
4               0.0             1.0             0.0                0.0
…             …               …               …                  …
55724           0.0             1.0             0.0                0.0
55725           0.0             1.0             0.0                0.0
55726           0.0             1.0             0.0                0.0
55727           0.0             1.0             0.0                0.0
55728           0.0             1.0             0.0                0.0

       cap-shape_convex  cap-shape_flat  cap-shape_others  \
0                   1.0             0.0               0.0
1                   0.0             0.0               0.0
2                   0.0             0.0               0.0
3                   1.0             0.0               0.0
4                   1.0             0.0               0.0
…                 …               …                 …
55724               0.0             0.0               0.0
55725               0.0             1.0               0.0
55726               0.0             0.0               0.0
55727               0.0             1.0               0.0
55728               0.0             0.0               0.0

       cap-shape_spherical  cap-shape_sunken  cap-surface_d  …  \
0                      0.0               0.0            0.0  …
1                      1.0               0.0            0.0  …
2                      1.0               0.0            0.0  …
3                      0.0               0.0            0.0  …
4                      0.0               0.0            0.0  …
…                    …                 …              …  …
55724                  0.0               1.0            0.0  …
55725                  0.0               0.0            0.0  …
55726                  0.0               1.0            0.0  …
55727                  0.0               0.0            0.0  …
55728                  0.0               1.0            0.0  …

       habitat_leaves  habitat_meadows  habitat_paths  habitat_urban  \
0                 0.0              0.0            0.0            0.0
1                 0.0              0.0            0.0            0.0
2                 0.0              0.0            0.0            0.0
3                 0.0              0.0            0.0            0.0
4                 0.0              0.0            0.0            0.0
…               …                …              …              …
55724             0.0              0.0            0.0            0.0
55725             0.0              0.0            0.0            0.0
```

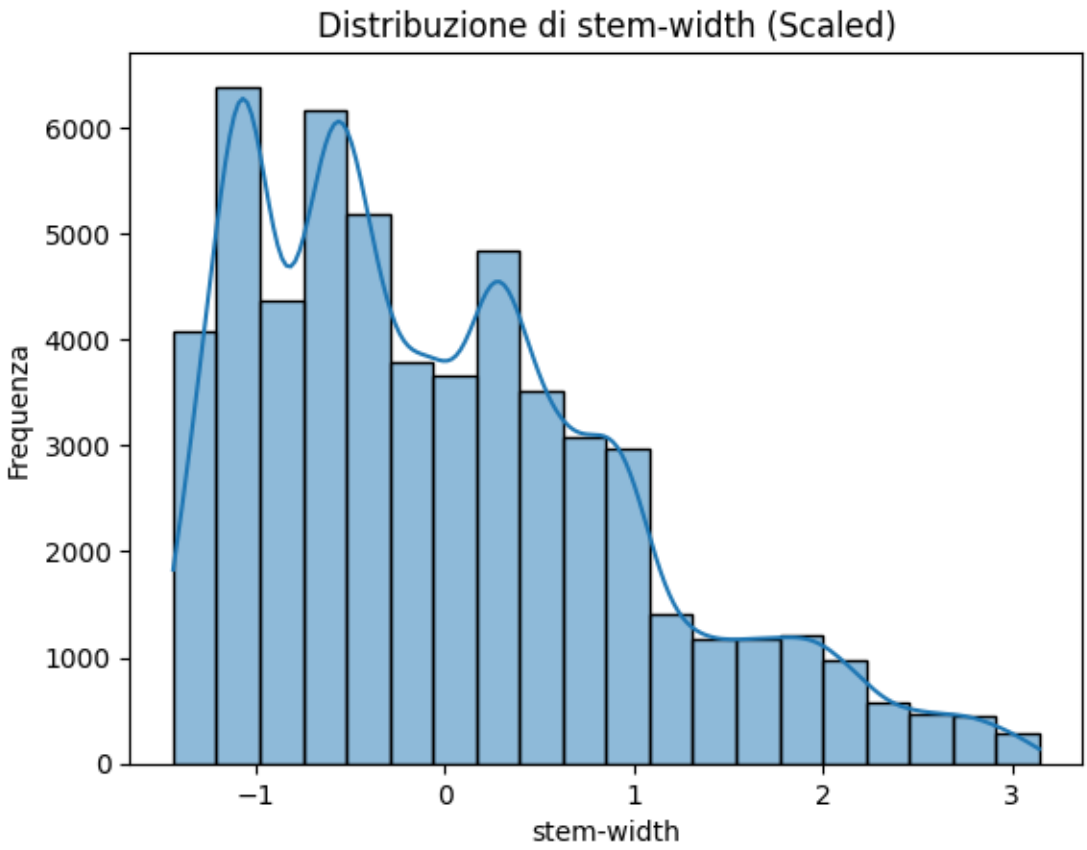
```
55726             0.0              0.0             0.0             0.0
55727             0.0              0.0             0.0             0.0
55728             0.0              0.0             0.0             0.0

       habitat_waste  habitat_woods  season_autumn  season_spring  \
0                0.0            1.0            1.0            0.0
1                0.0            1.0            1.0            0.0
2                0.0            1.0            0.0            0.0
3                0.0            1.0            1.0            0.0
4                0.0            1.0            1.0            0.0
…                …              …              …              …
55724            0.0            1.0            1.0            0.0
55725            0.0            1.0            1.0            0.0
55726            0.0            1.0            0.0            0.0
55727            0.0            1.0            0.0            0.0
55728            0.0            1.0            0.0            0.0

       season_summer  season_winter
0                0.0            0.0
1                0.0            0.0
2                1.0            0.0
3                0.0            0.0
4                0.0            0.0
…                …              …
55724            0.0            0.0
55725            0.0            0.0
55726            1.0            0.0
55727            1.0            0.0
55728            1.0            0.0

[55729 rows x 98 columns]
```

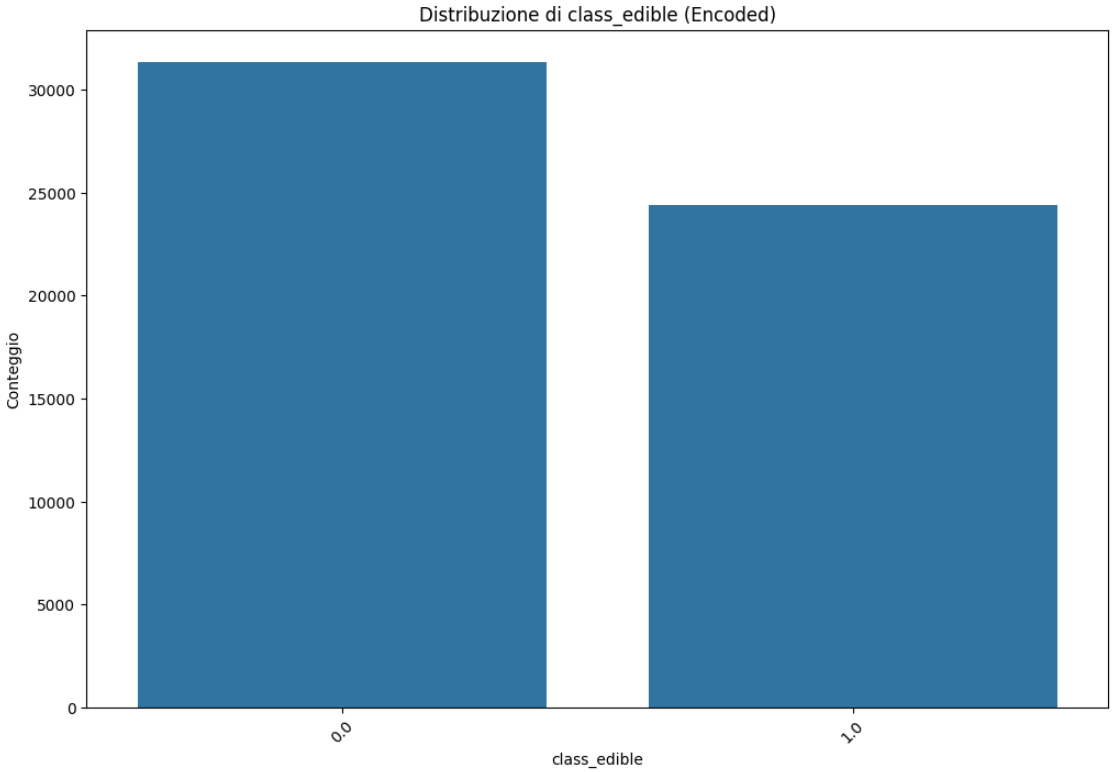
[ ]:
```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler, OneHotEncoder

# Assume che tu abbia già applicato lo scaling e l'encoding al DataFrame
# Quindi, abbiamo df2_scaled per le colonne numeriche scalate e df2_encoded per␣
 ↪le colonne categoriche codificate

# Visualizzazione delle distribuzioni delle colonne numeriche scalate␣
 ↪utilizzando un istogramma per ciascuna colonna
plt.figure(figsize=(12, 8))
for col in df2_scaled.columns:
    sns.histplot(data=df2_scaled, x=col, kde=True, bins=20)
    plt.title(f'Distribuzione di {col} (Scaled)')
```

```
    plt.xlabel(col)
    plt.ylabel('Frequenza')
    plt.show()

# Visualizzazione delle distribuzioni delle colonne categoriche codificate␣
 ↪utilizzando grafici a barre
plt.figure(figsize=(12, 8))
for col in df2_encoded.columns:
    sns.countplot(data=df2_encoded, x=col)
    plt.title(f'Distribuzione di {col} (Encoded)')
    plt.xlabel(col)
    plt.ylabel('Conteggio')
    plt.xticks(rotation=45)
    plt.show()
```

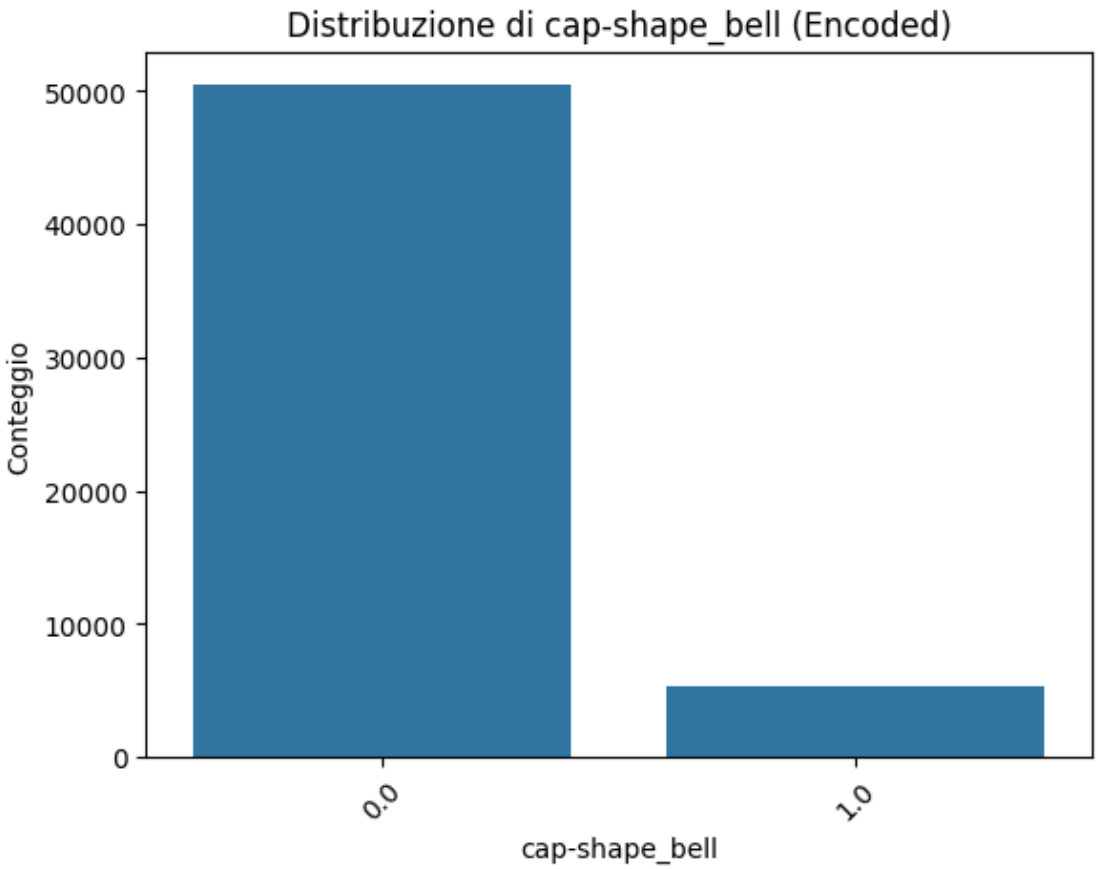


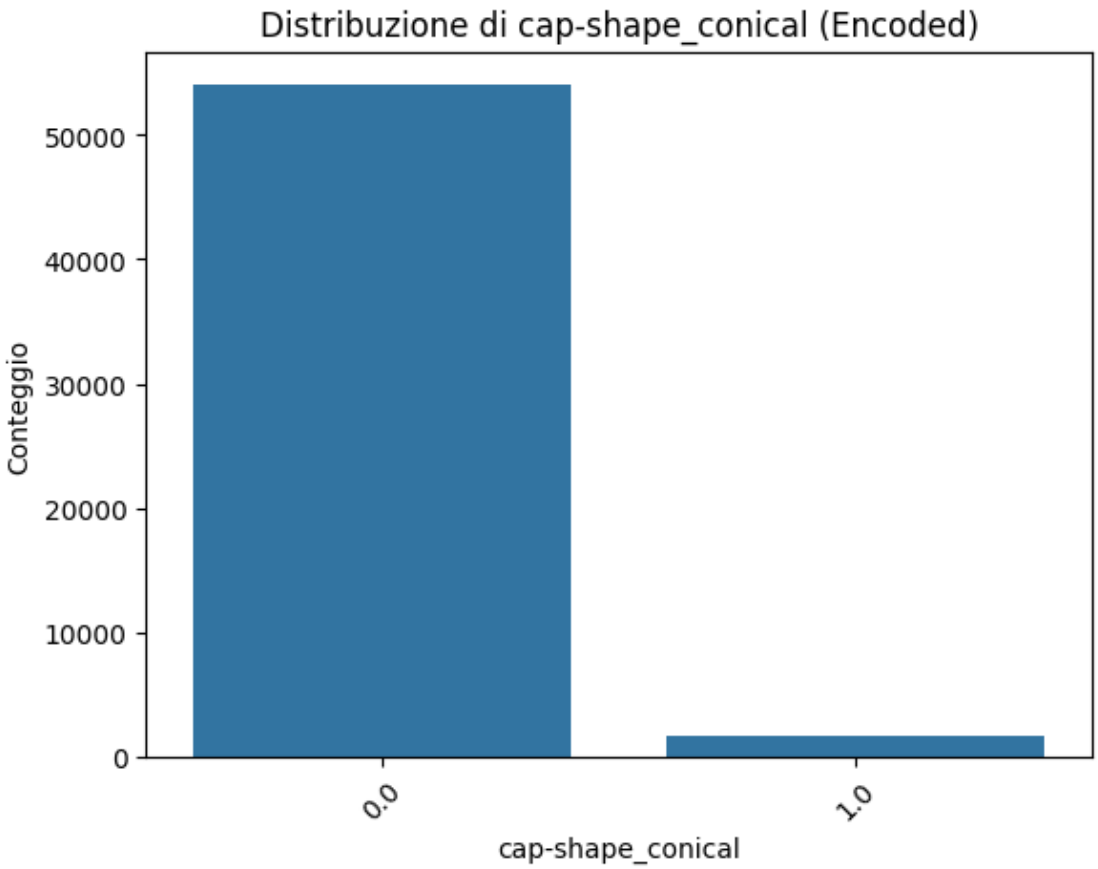

Distribuzione di stem-height (Scaled)
Frequenza
7000
6000
5000
4000
3000
2000
1000
0
−2
−1
0
1
2
3
stem-height

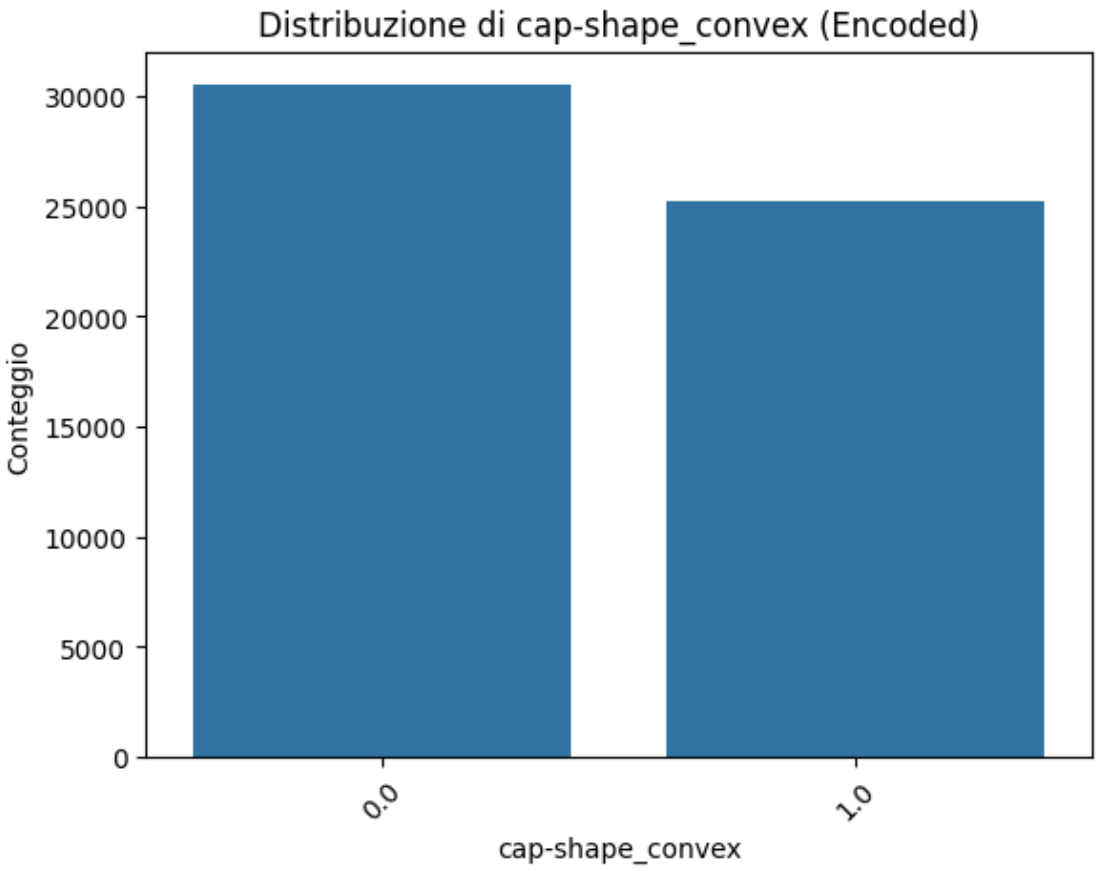

Distribuzione di stem-width (Scaled)
Frequenza
stem-width
0
1000
2000
3000
4000
5000
6000
−1
0
1
2
3

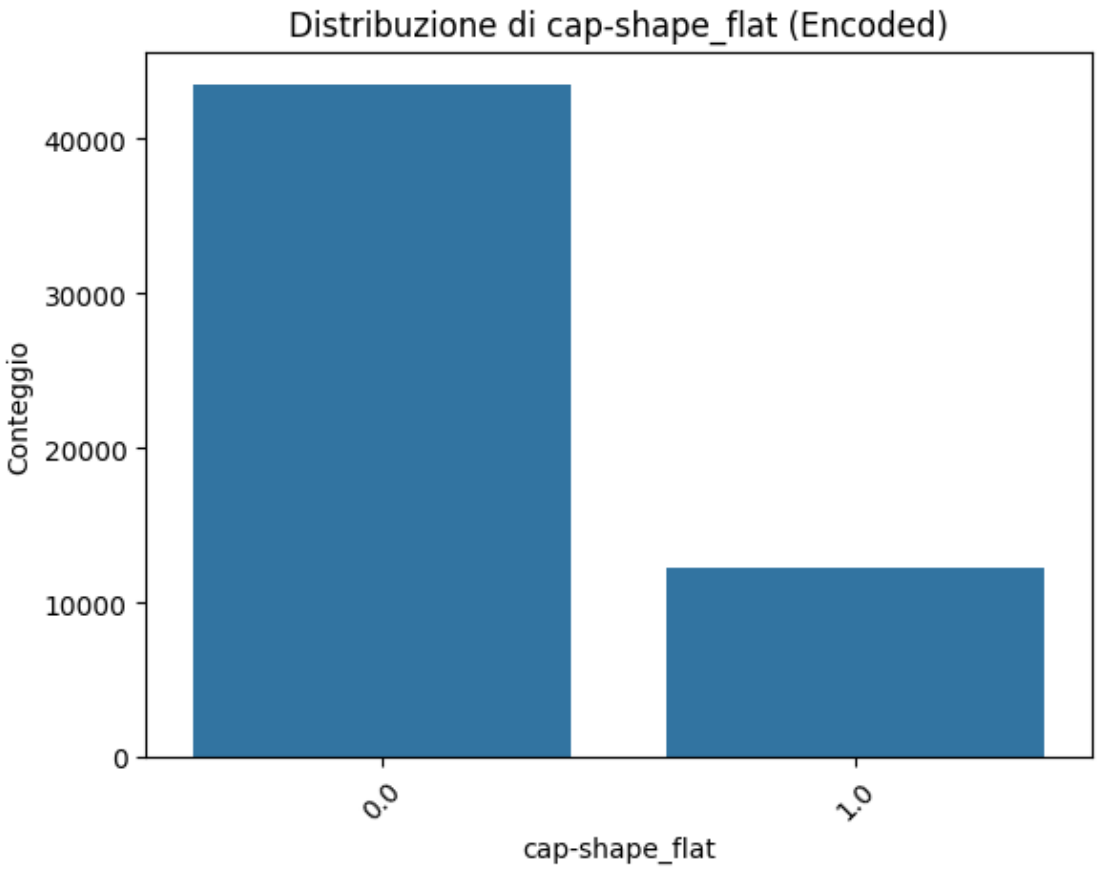

Distribuzione di class_edible (Encoded)
Conteggio
30000
25000
20000
15000
10000
5000
0
0.0
1.0
class_edible

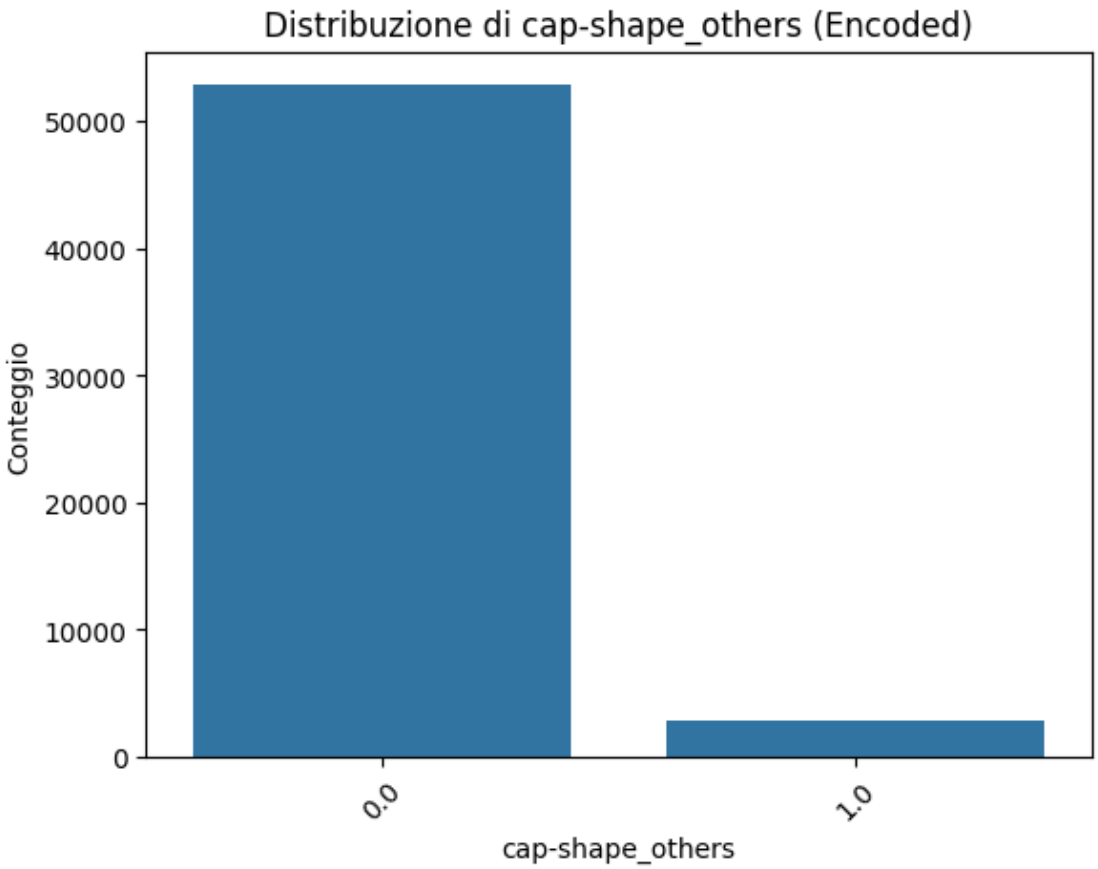

Distribuzione di class_poisonus (Encoded)
Conteggio
30000
25000
20000
15000
10000
5000
0
0.0
1.0
class_poisonus

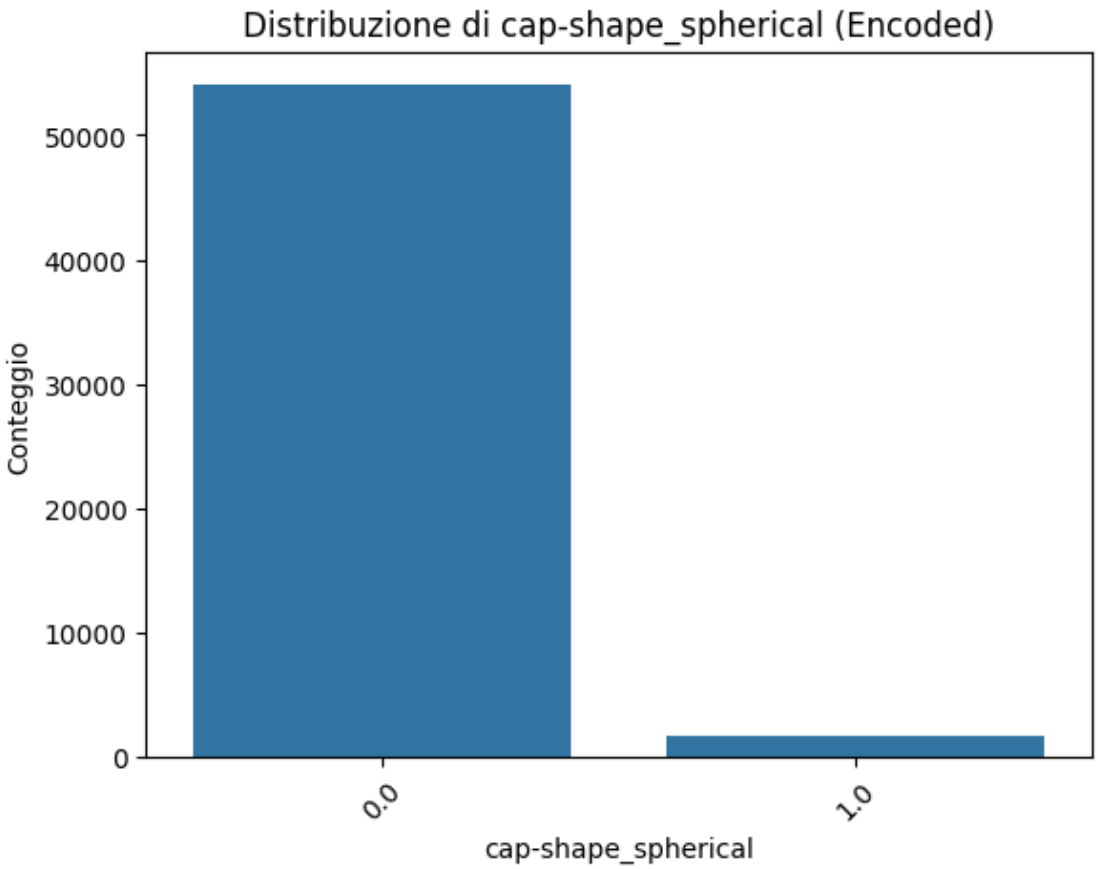

Distribuzione di cap-shape_bell (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-shape_bell

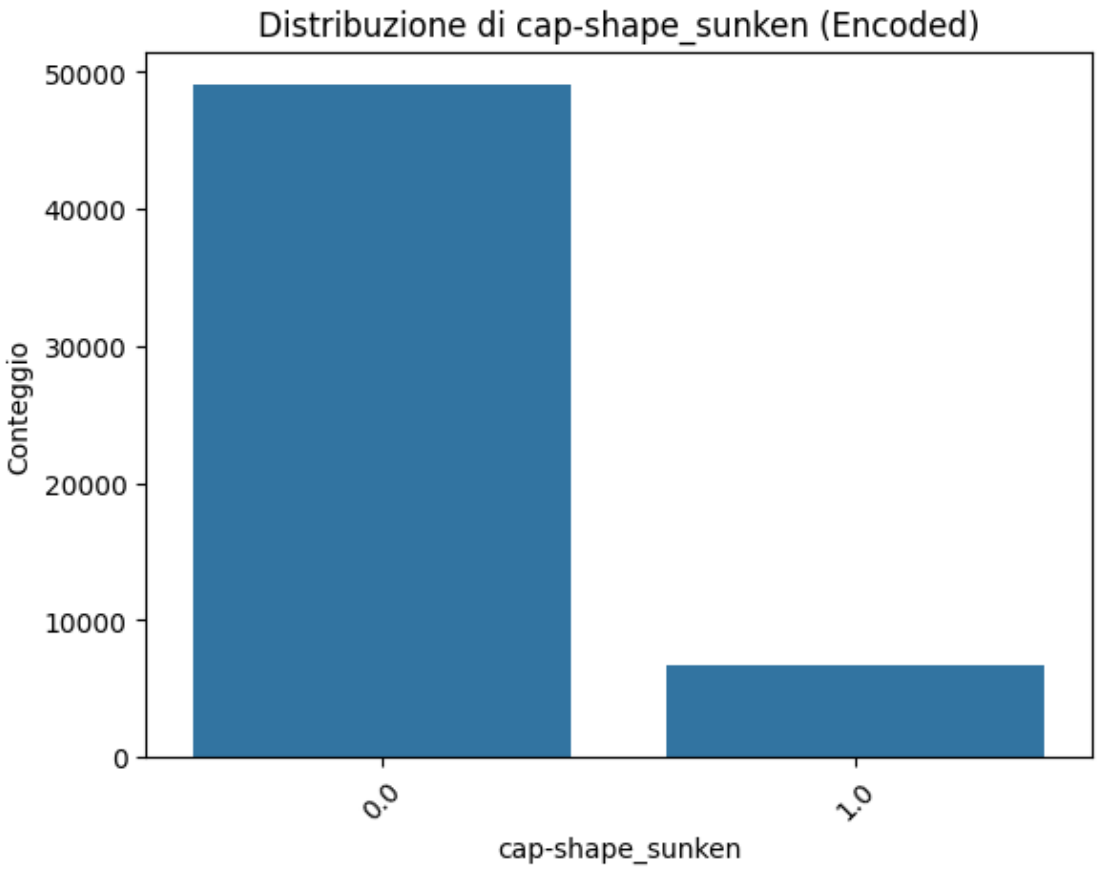

Distribuzione di cap-shape_conical (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-shape_conical

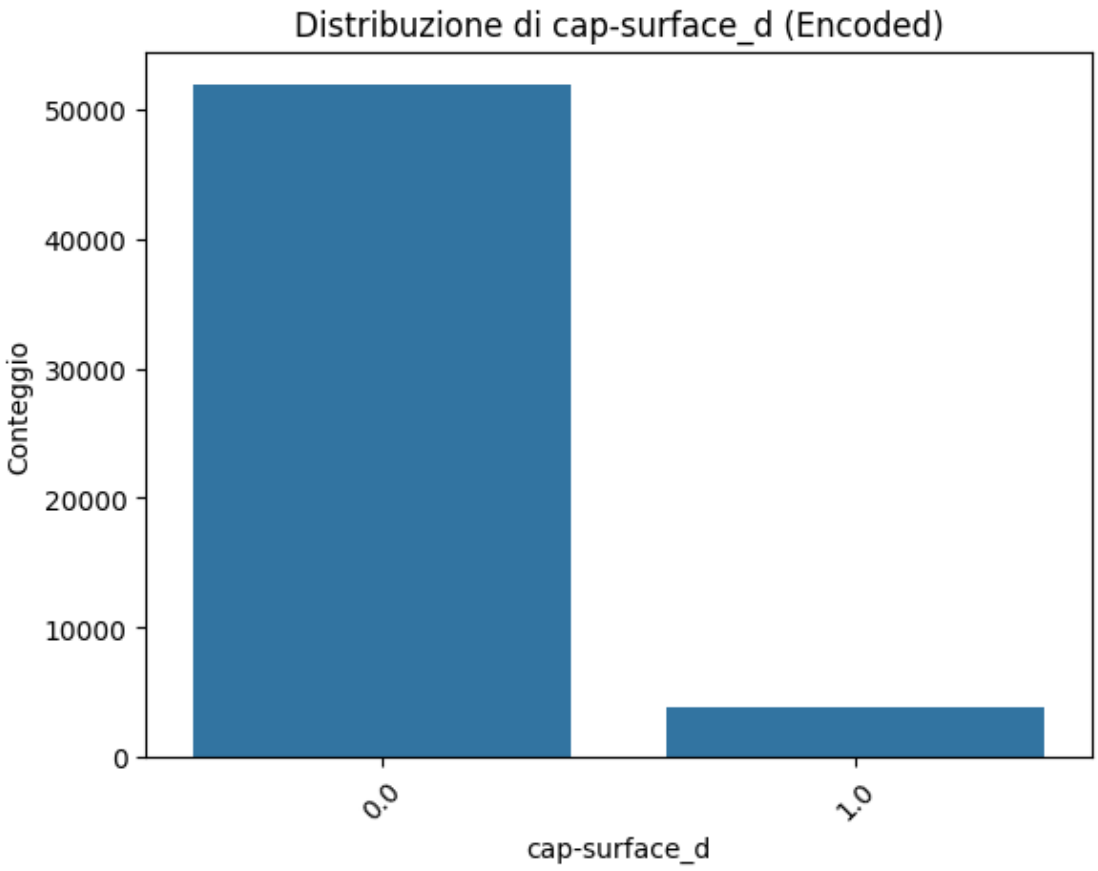

Distribuzione di cap-shape_convex (Encoded)
Conteggio
30000
25000
20000
15000
10000
5000
0
0.0
1.0
cap-shape_convex

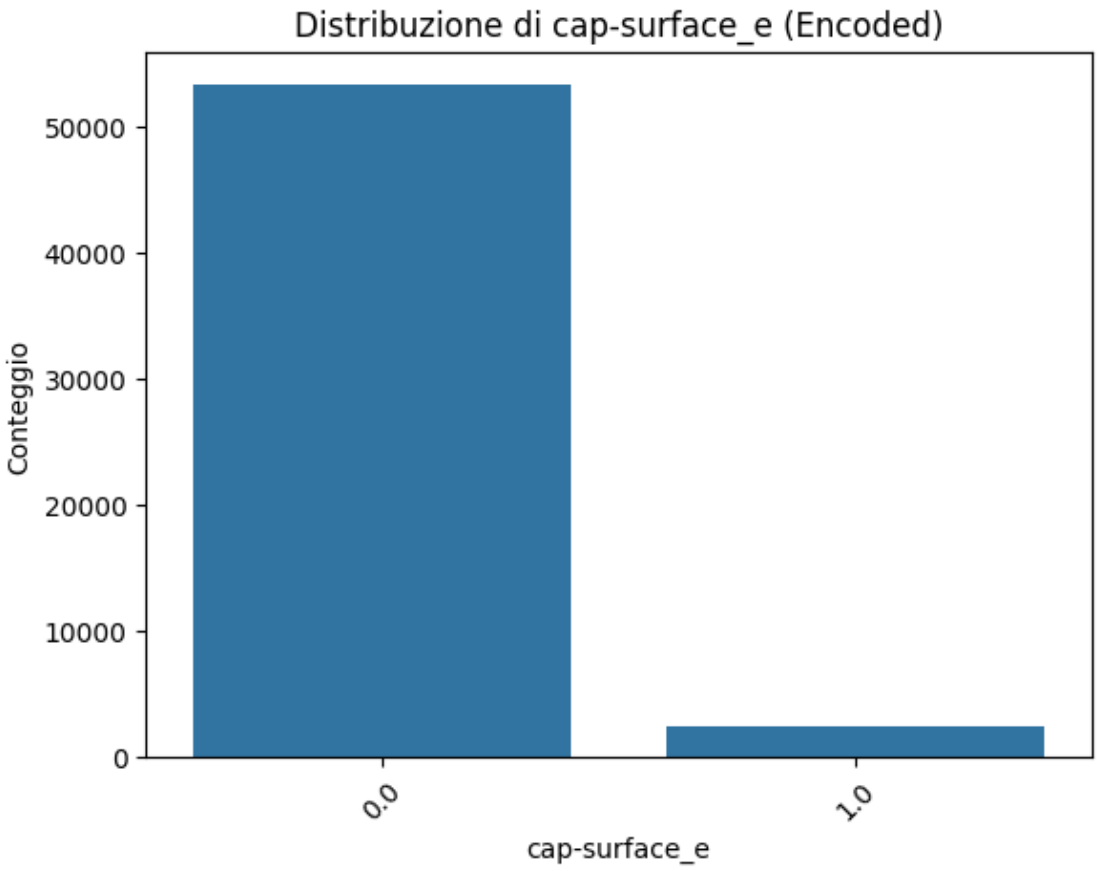

Distribuzione di cap-shape_flat (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
cap-shape_flat

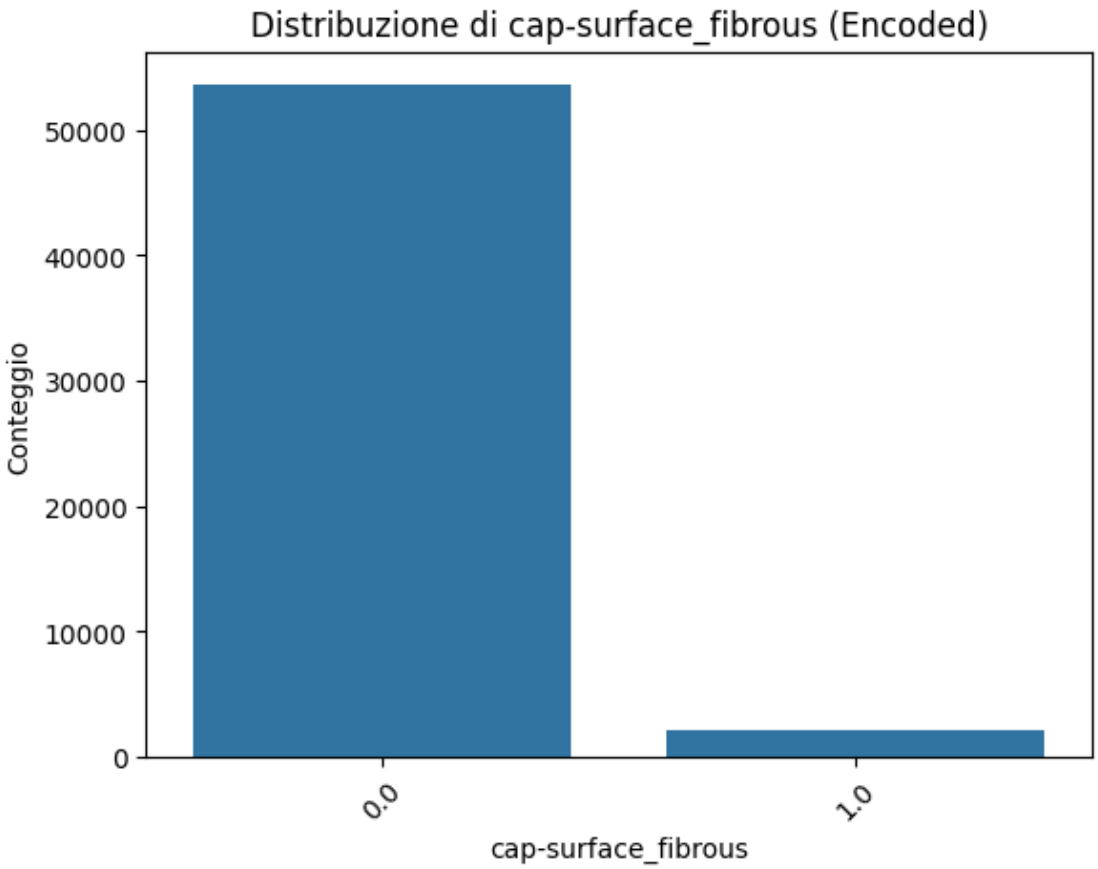

Distribuzione di cap-shape_others (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-shape_others

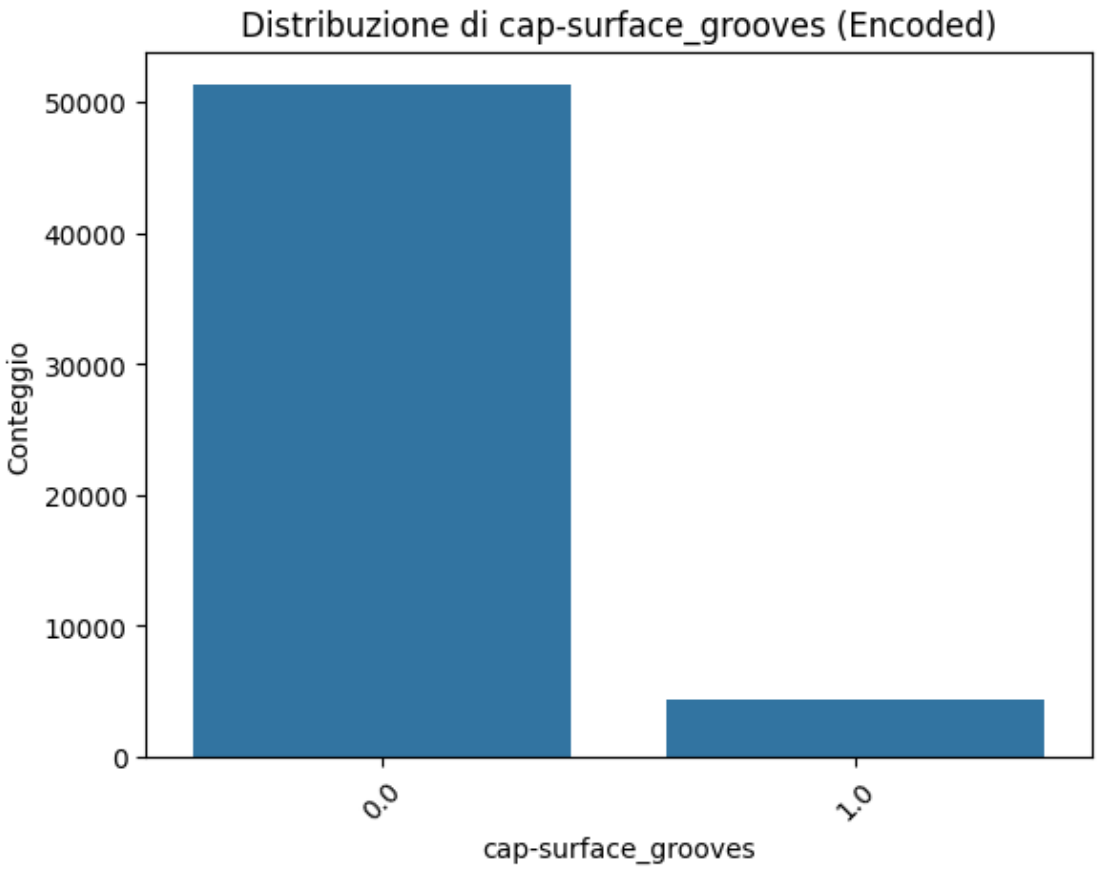

Distribuzione di cap-shape_spherical (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-shape_spherical

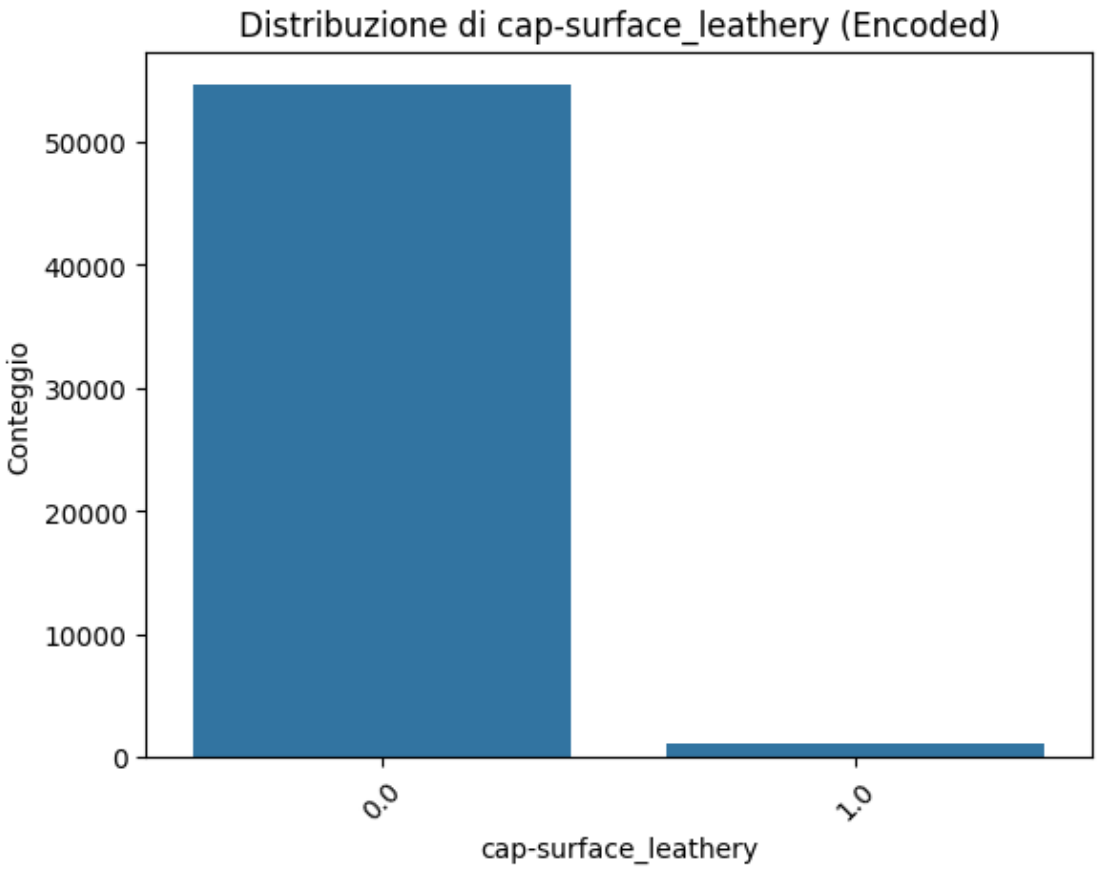

Distribuzione di cap-shape_sunken (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-shape_sunken

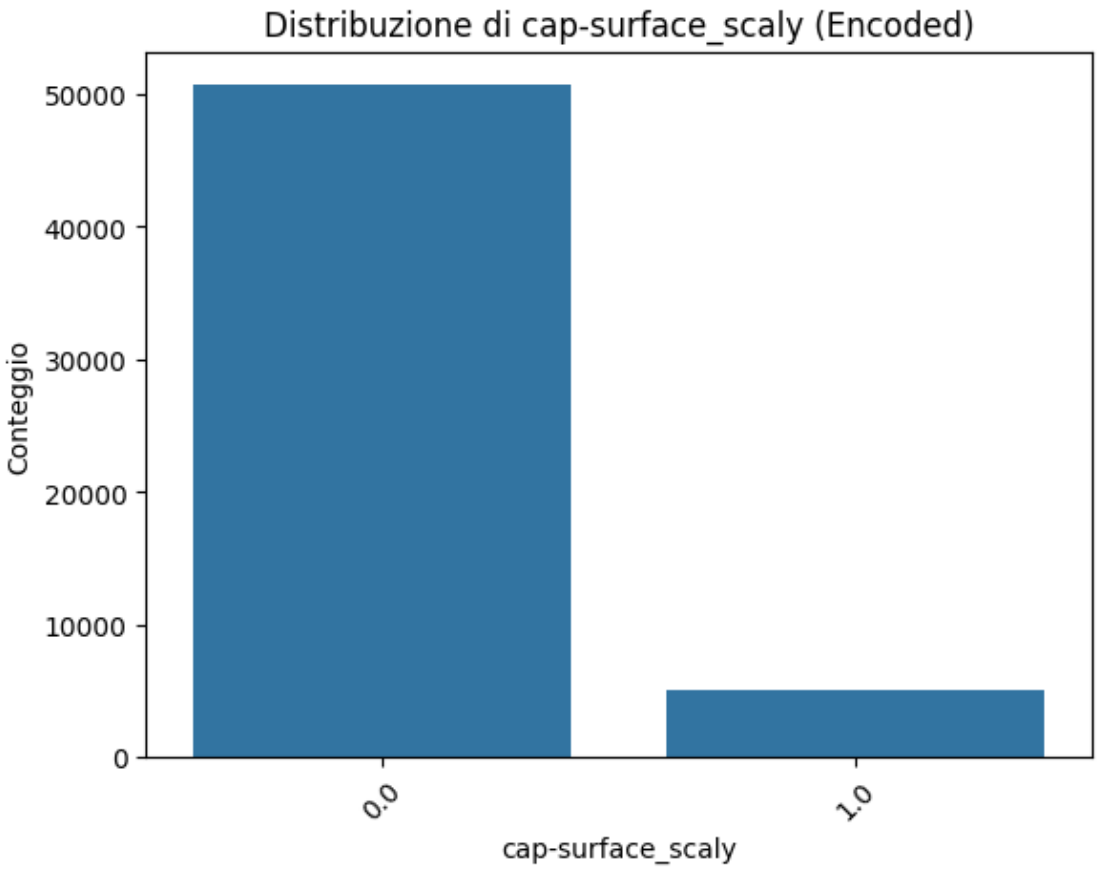

Distribuzione di cap-surface_d (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_d

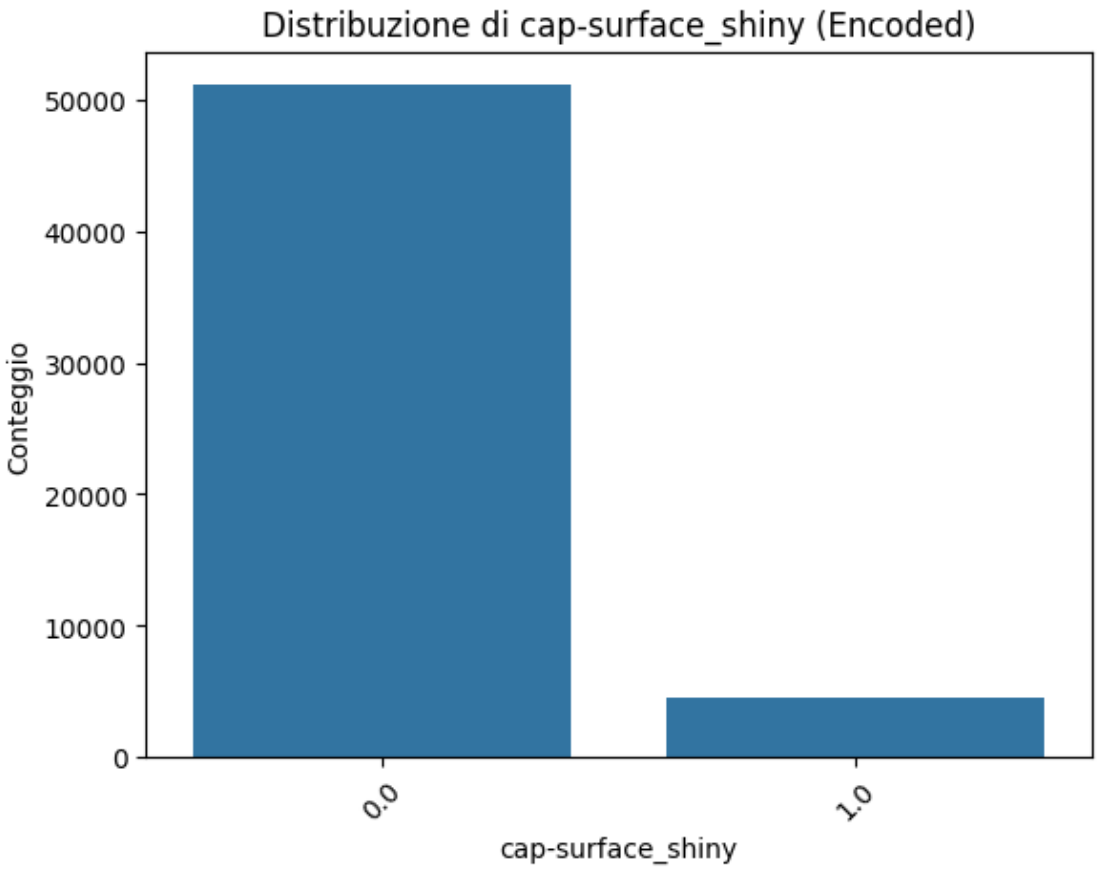

Distribuzione di cap-surface_e (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_e

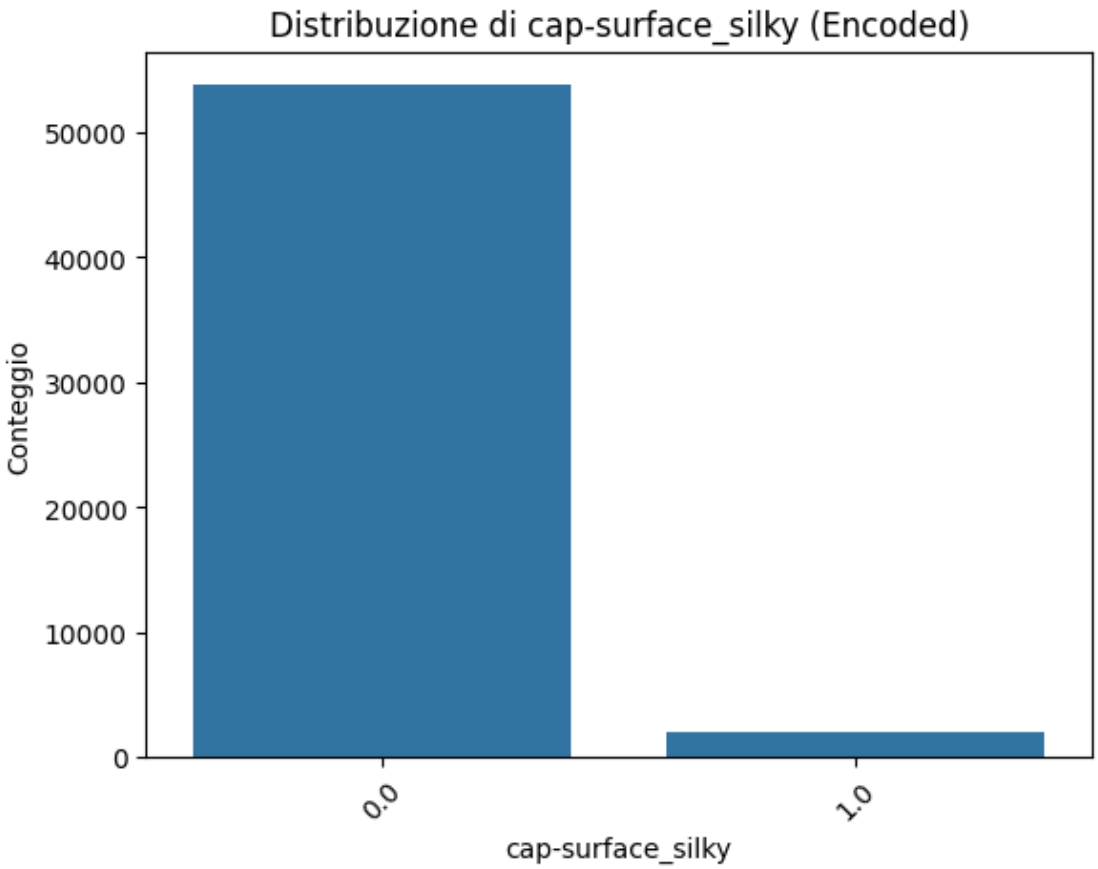

Distribuzione di cap-surface_fibrous (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_fibrous

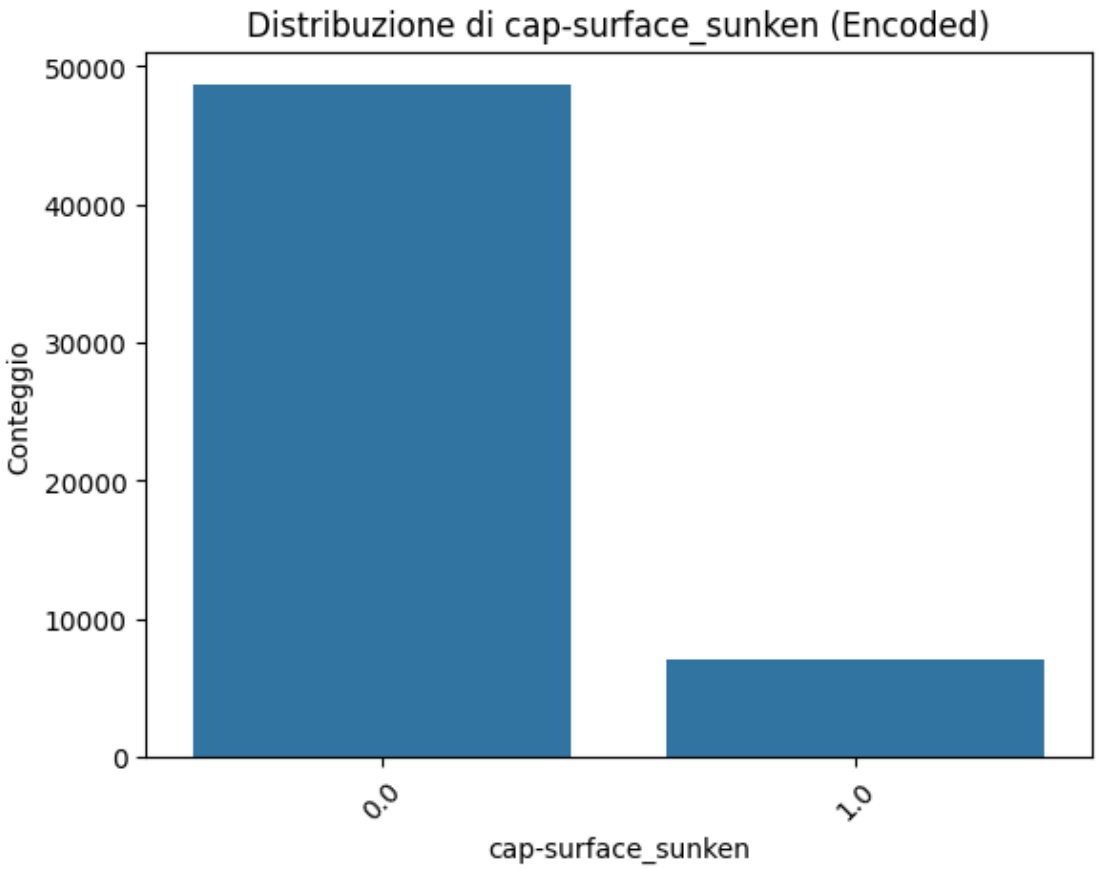

Distribuzione di cap-surface_grooves (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_grooves

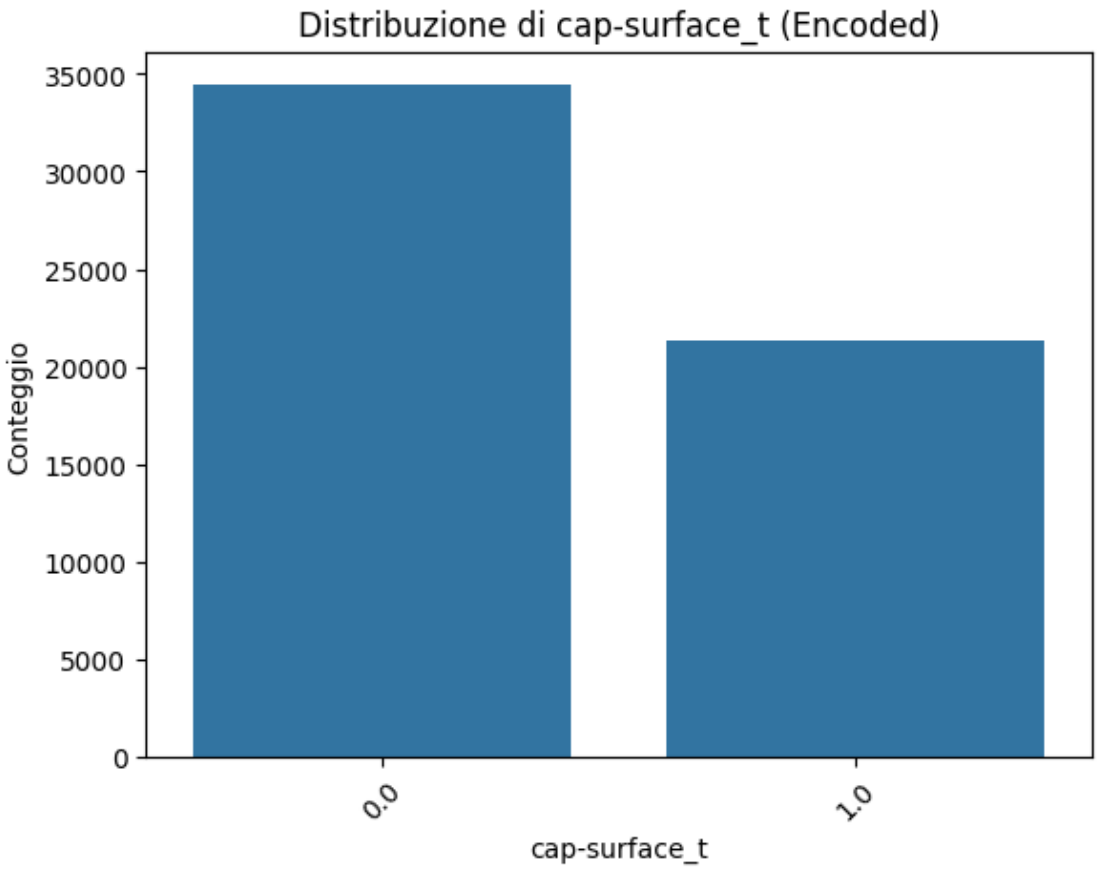

Distribuzione di cap-surface_leathery (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_leathery

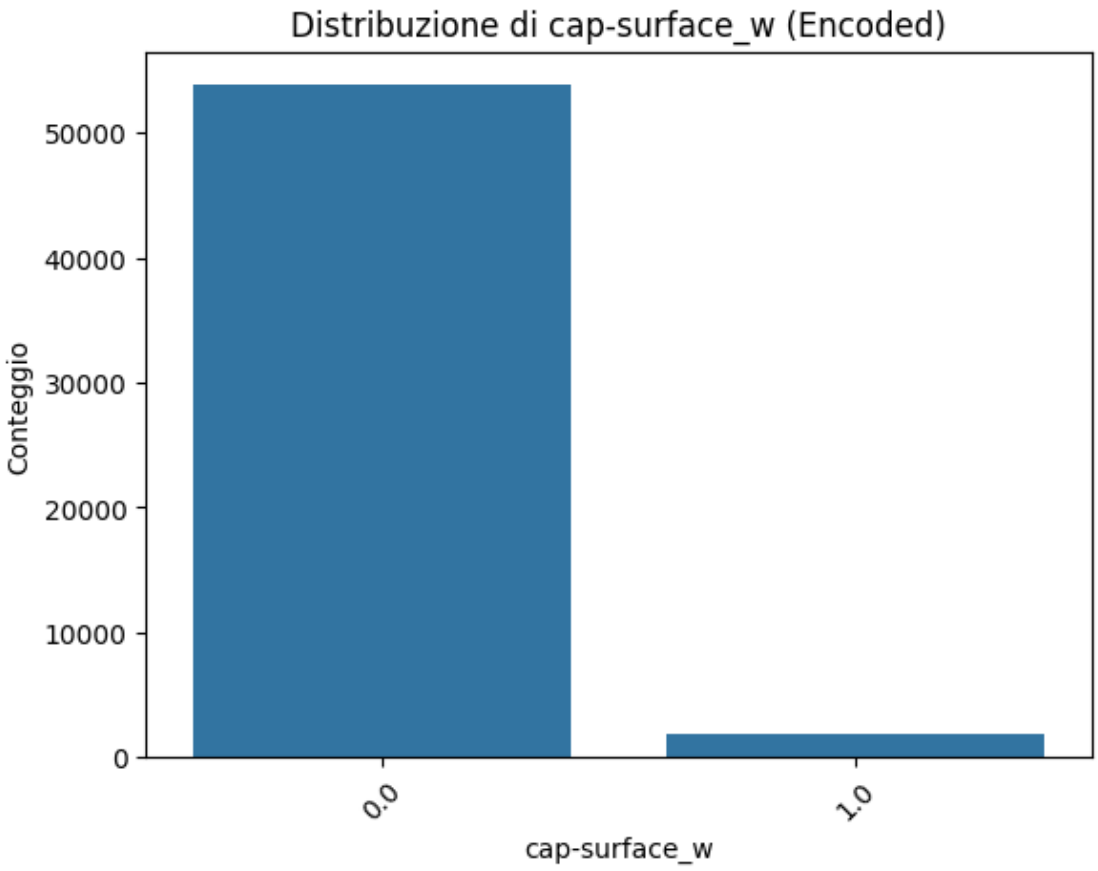

Distribuzione di cap-surface_scaly (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_scaly

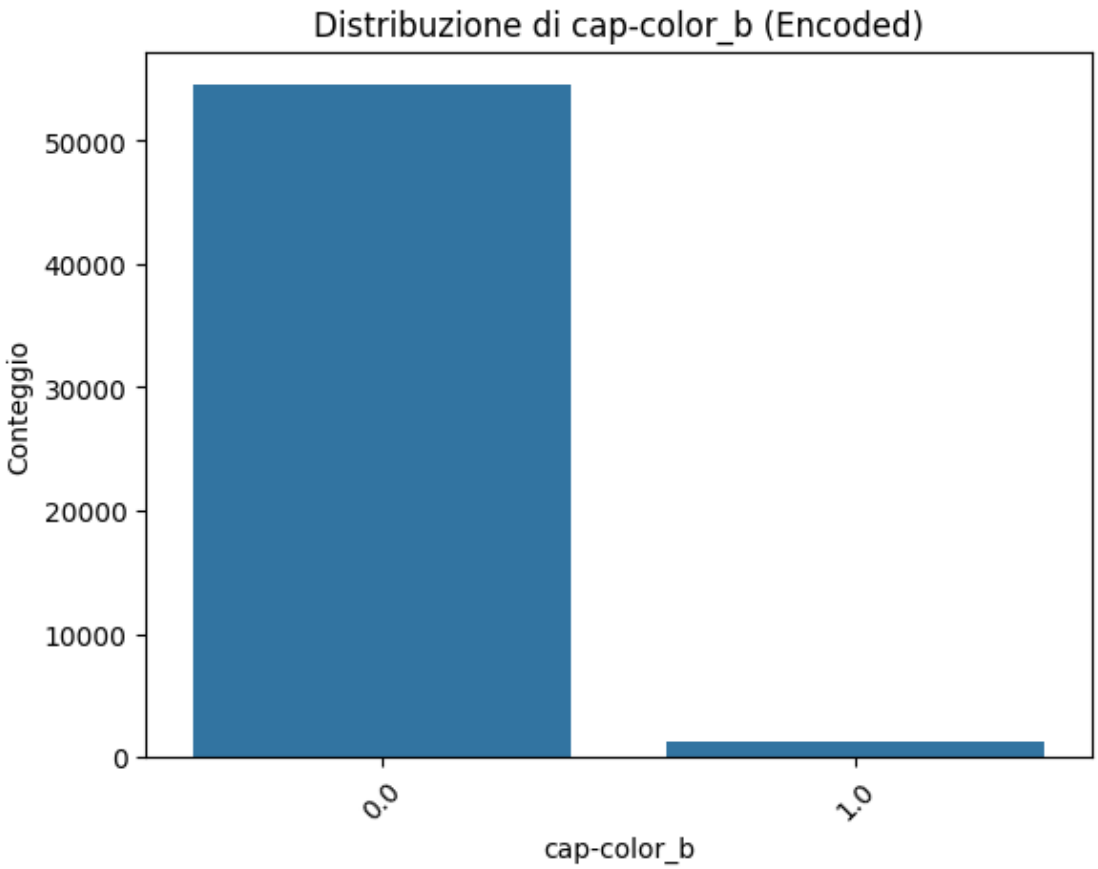

Distribuzione di cap-surface_shiny (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_shiny

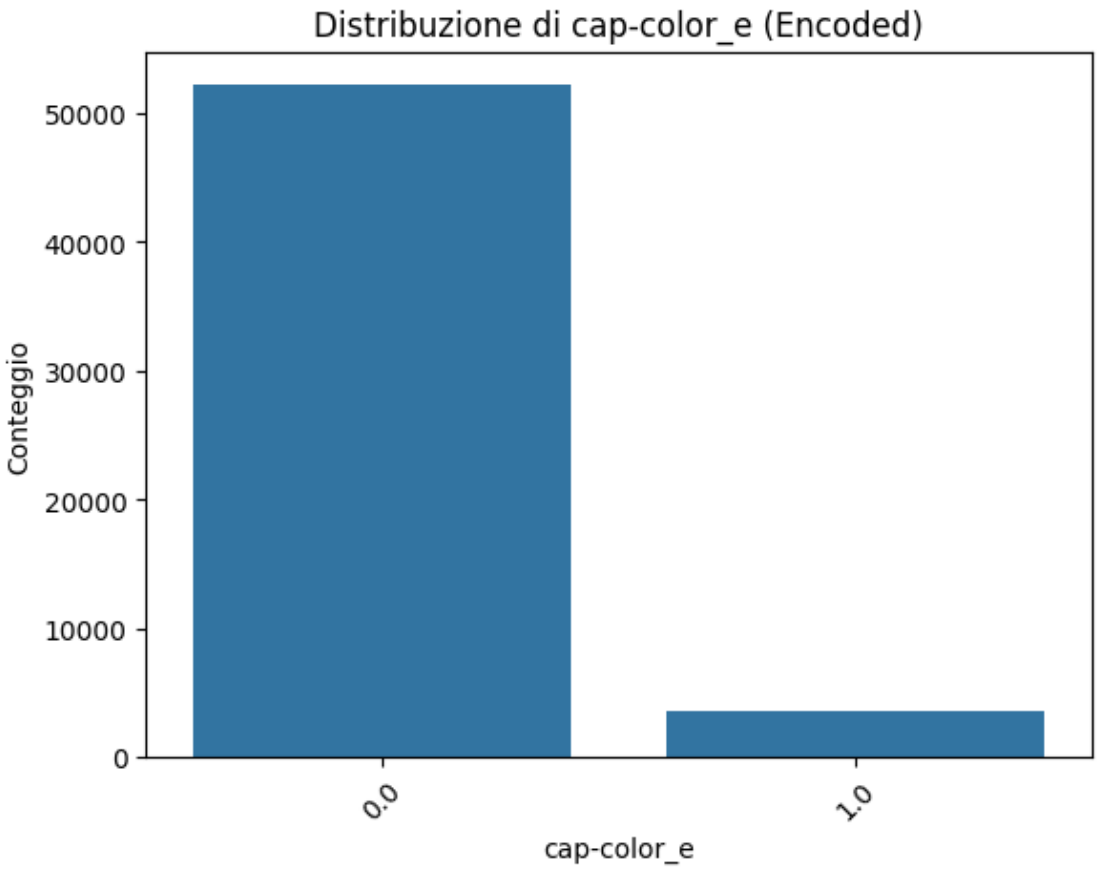

Distribuzione di cap-surface_silky (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_silky

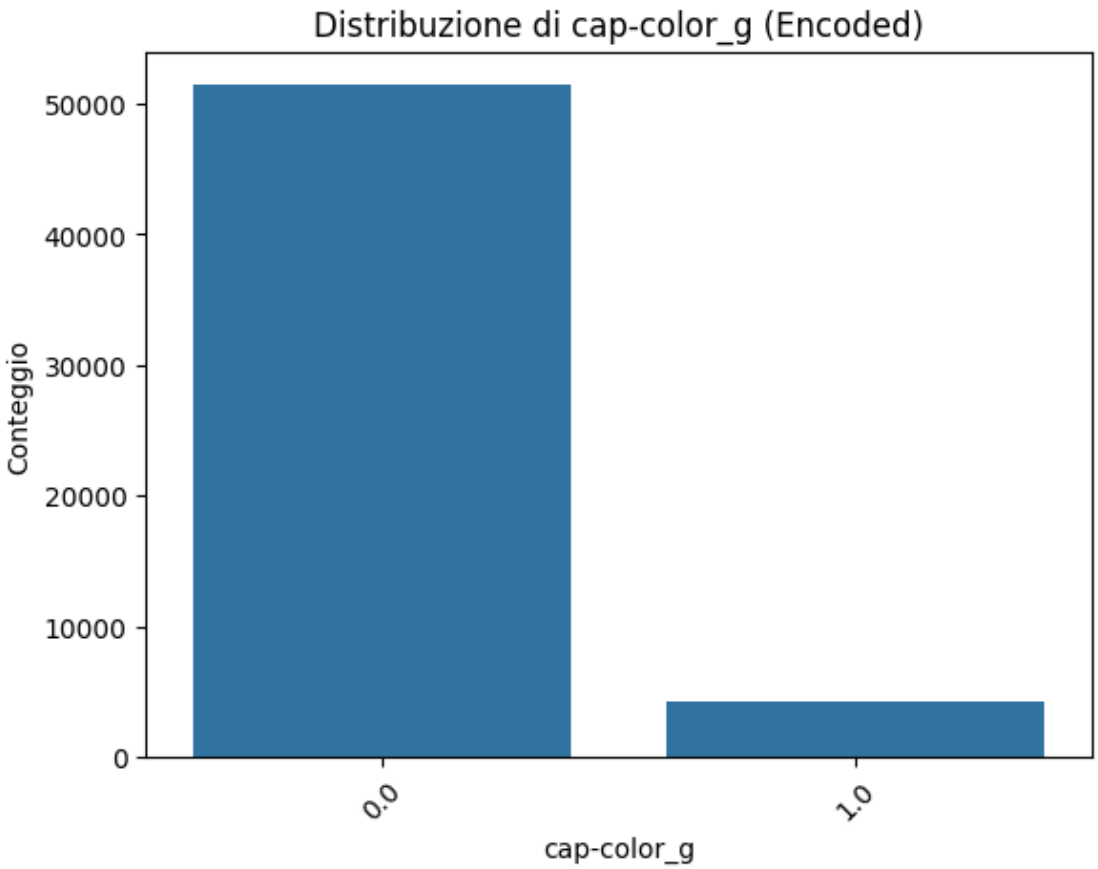

Distribuzione di cap-surface_sunken (Encoded)
50000
40000
30000
20000
10000
0
Conteggio
0.0
1.0
cap-surface_sunken

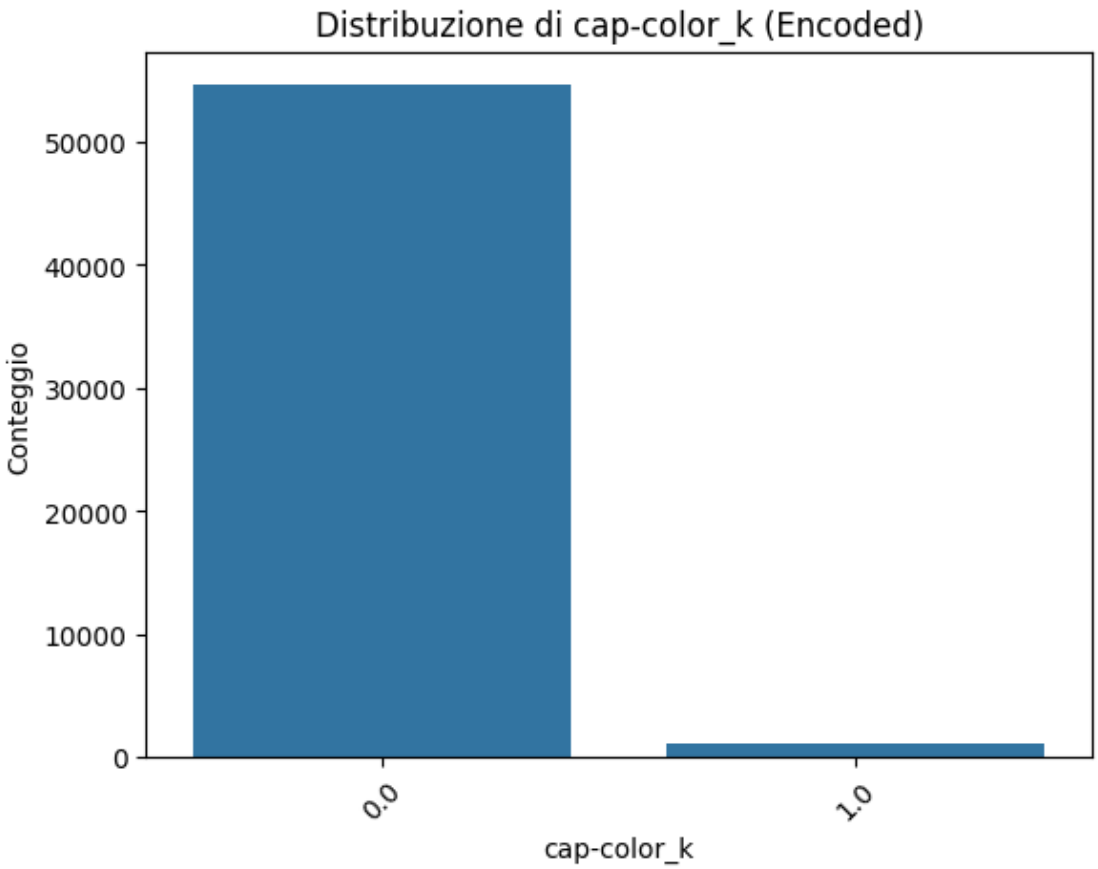

Distribuzione di cap-surface_t (Encoded)
Conteggio
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
cap-surface_t

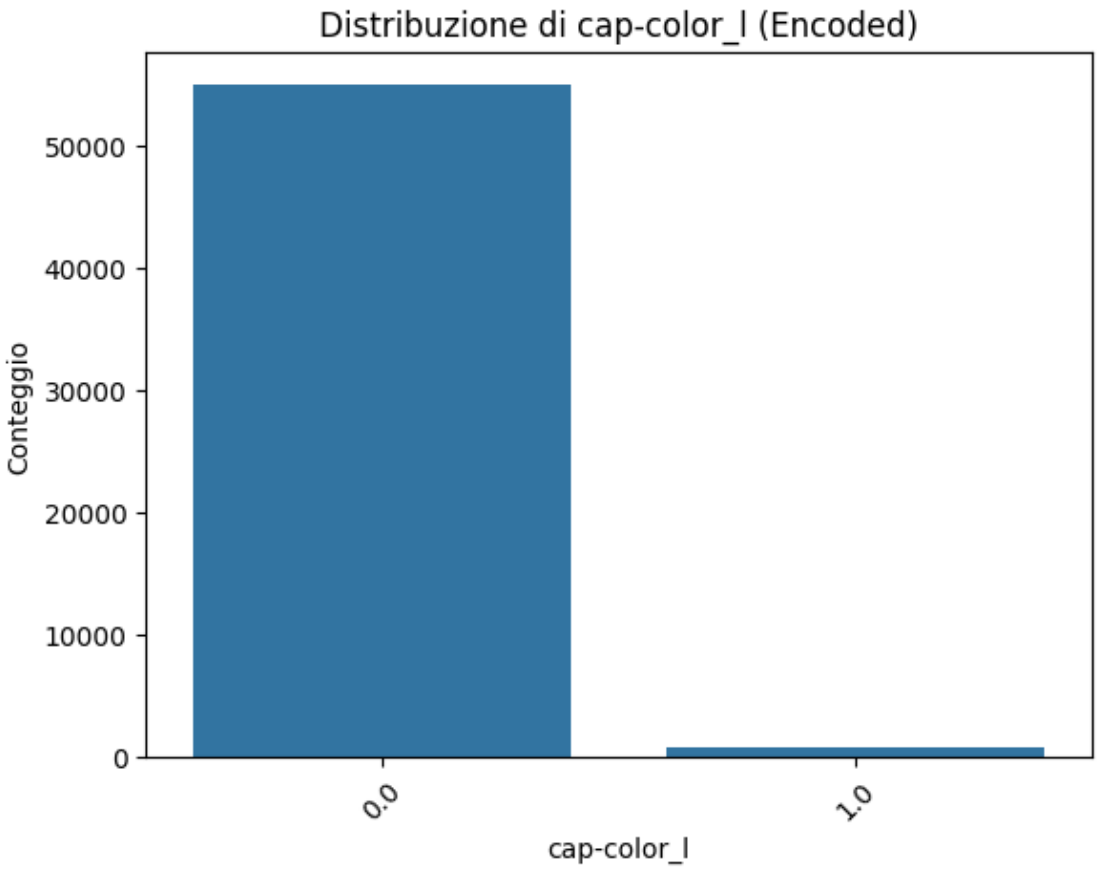

Distribuzione di cap-surface_w (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-surface_w

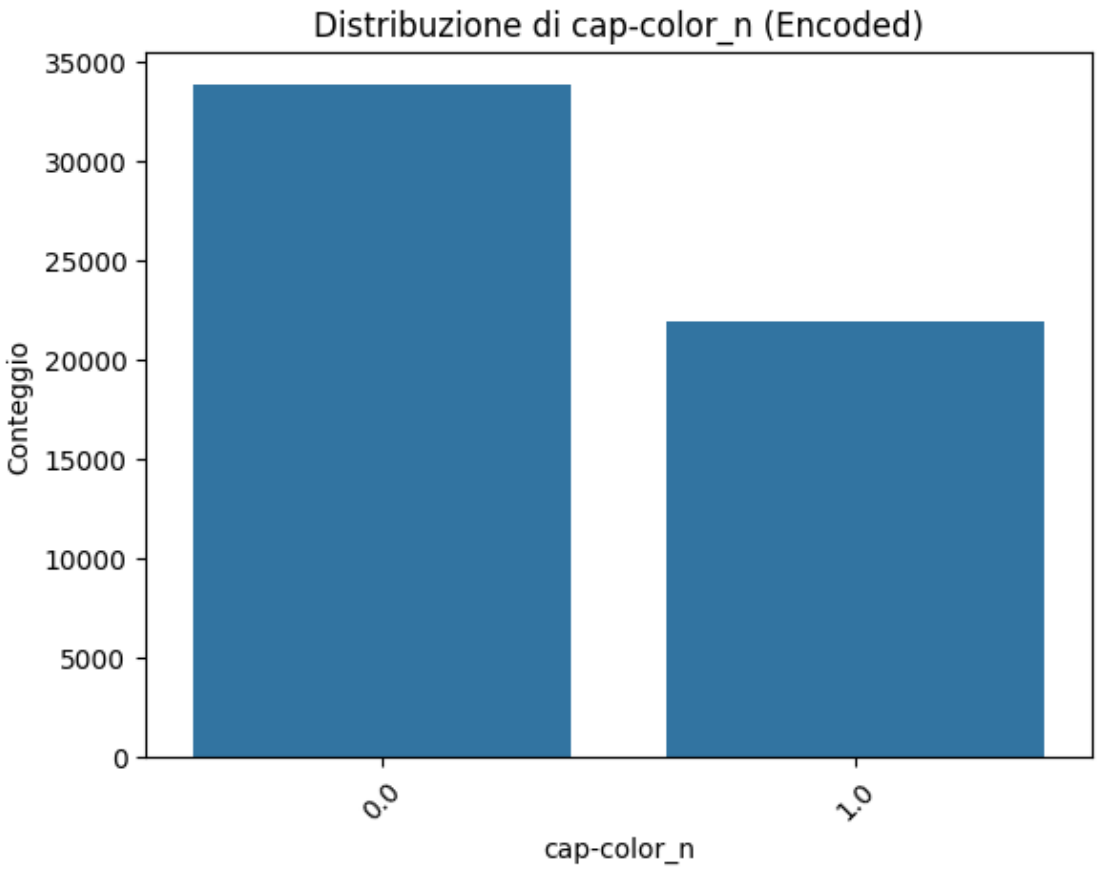

Distribuzione di cap-color_b (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_b

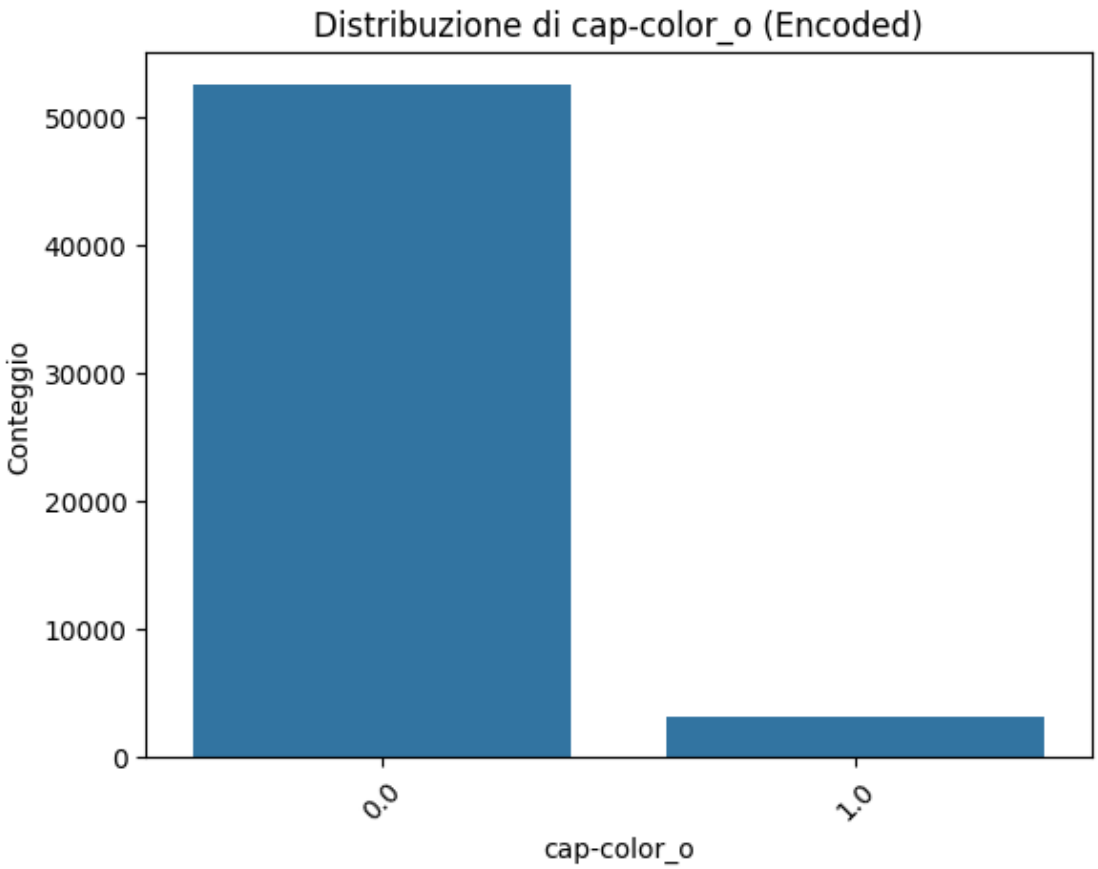

Distribuzione di cap-color_e (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_e

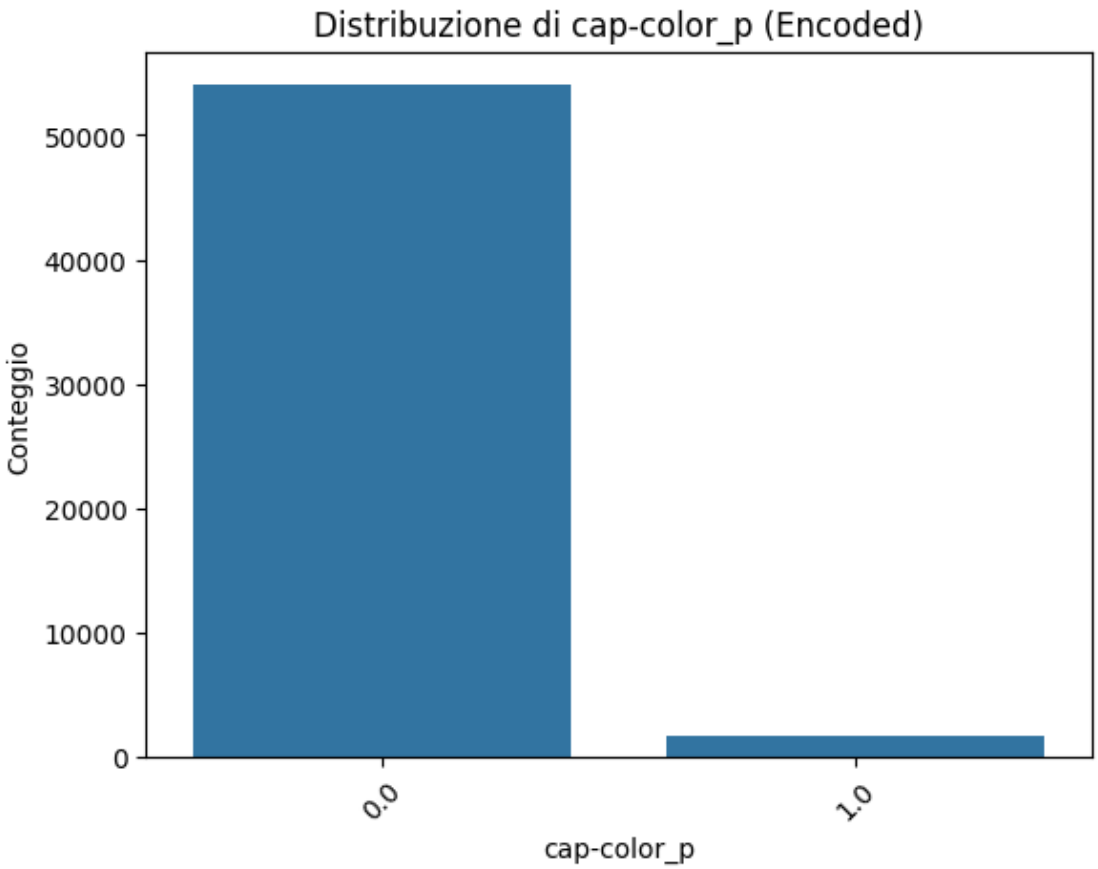

Distribuzione di cap-color_g (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_g

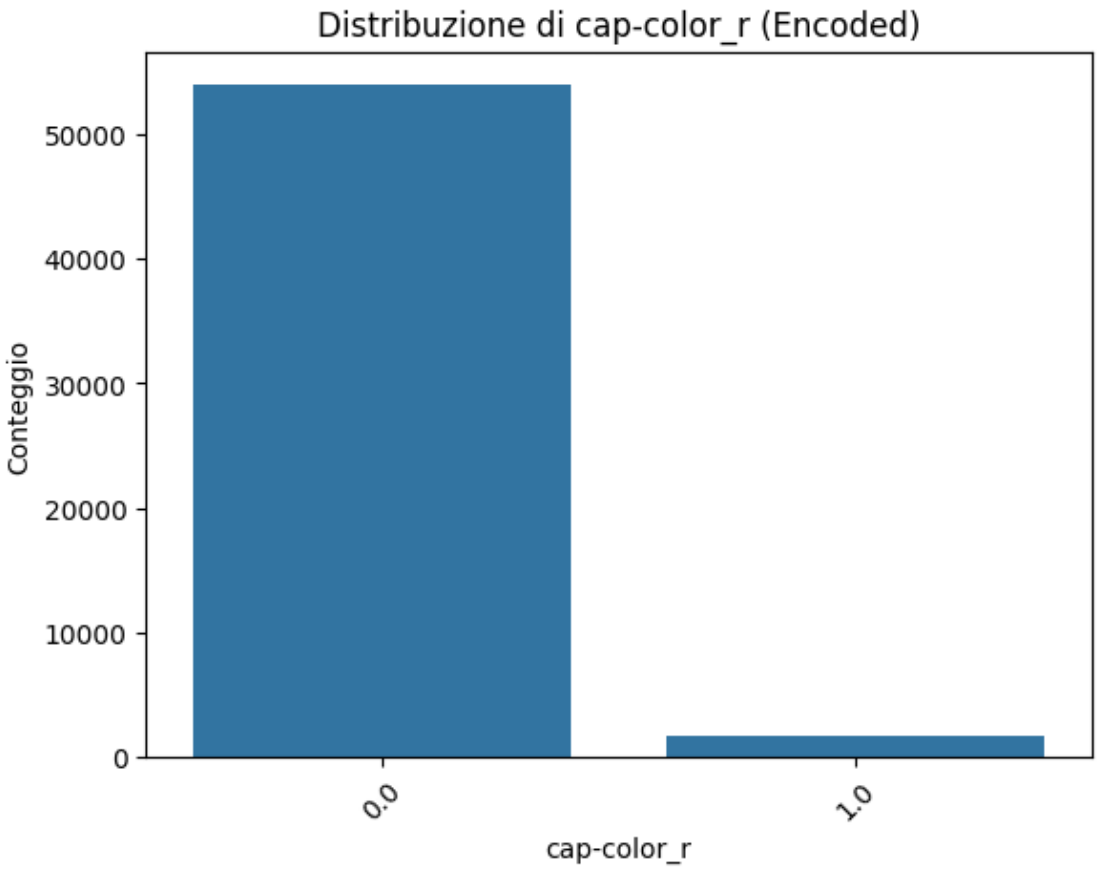

Distribuzione di cap-color_k (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_k

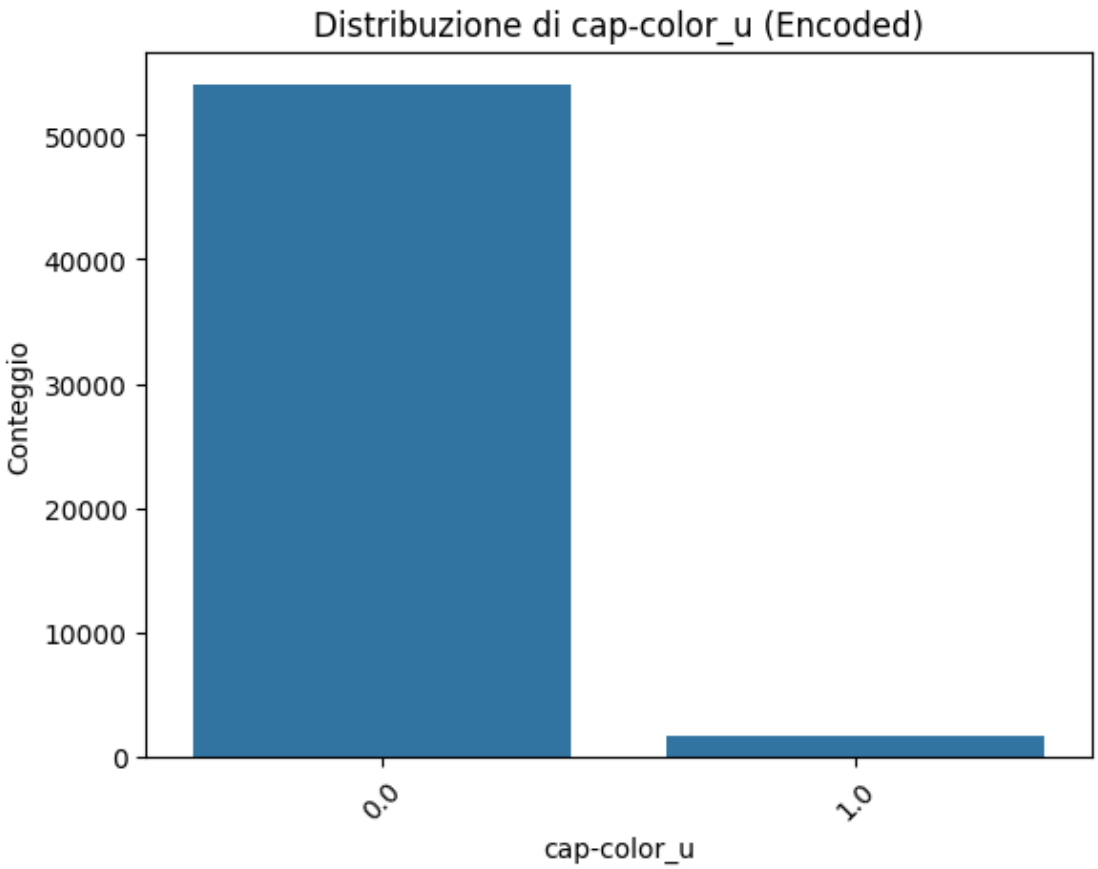

Distribuzione di cap-color_l (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_l

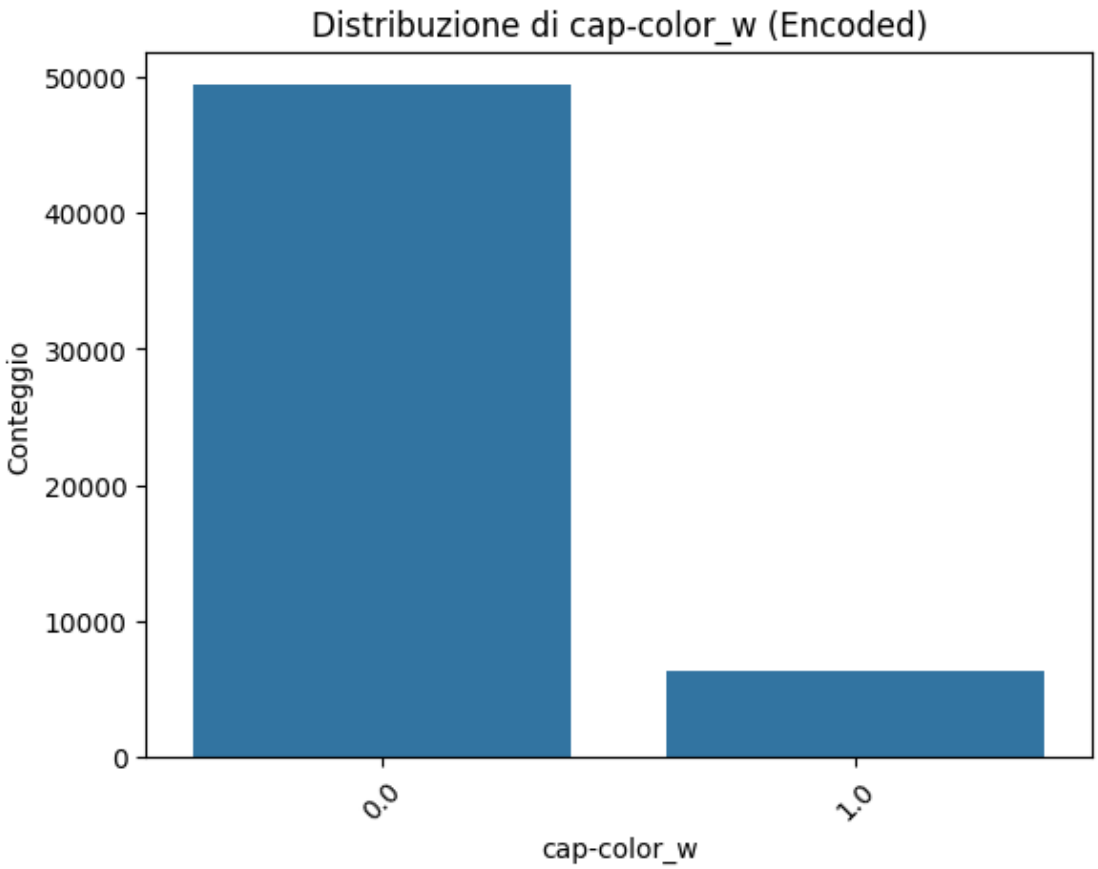

Distribuzione di cap-color_n (Encoded)
Conteggio
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
cap-color_n

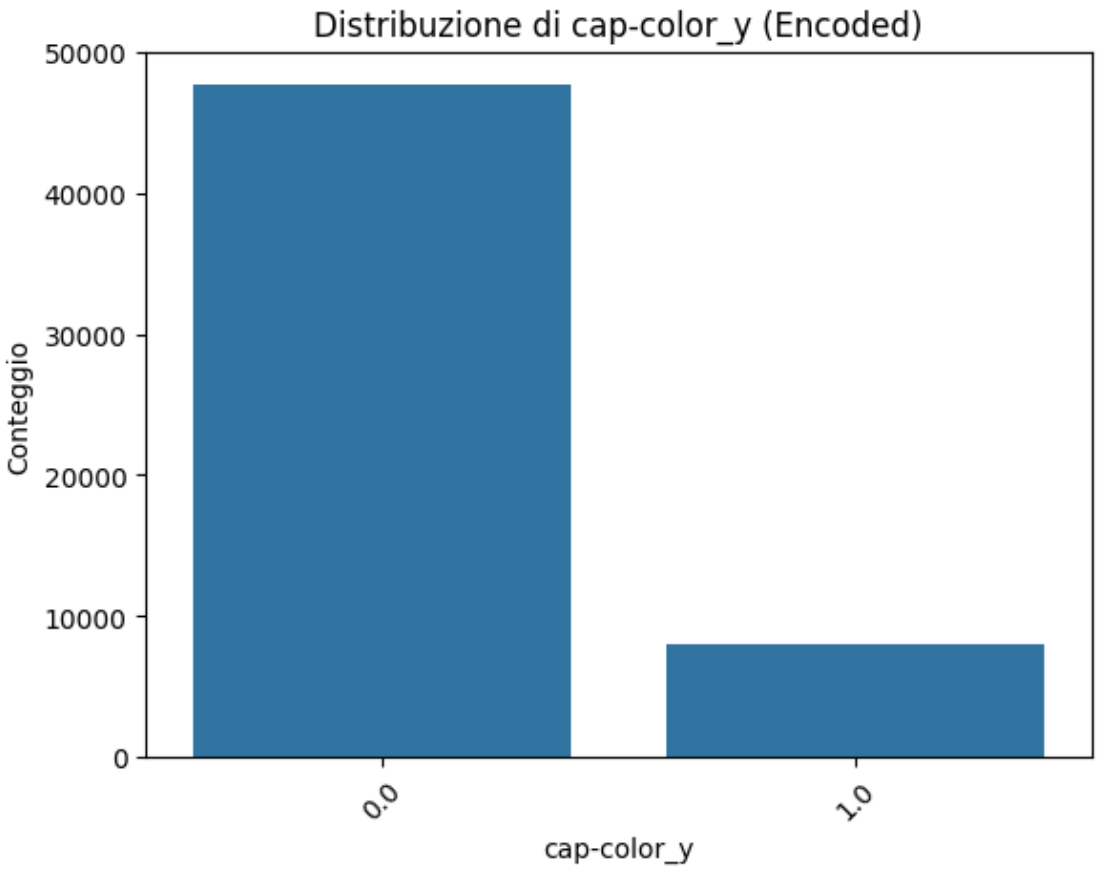

Distribuzione di cap-color_o (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_o

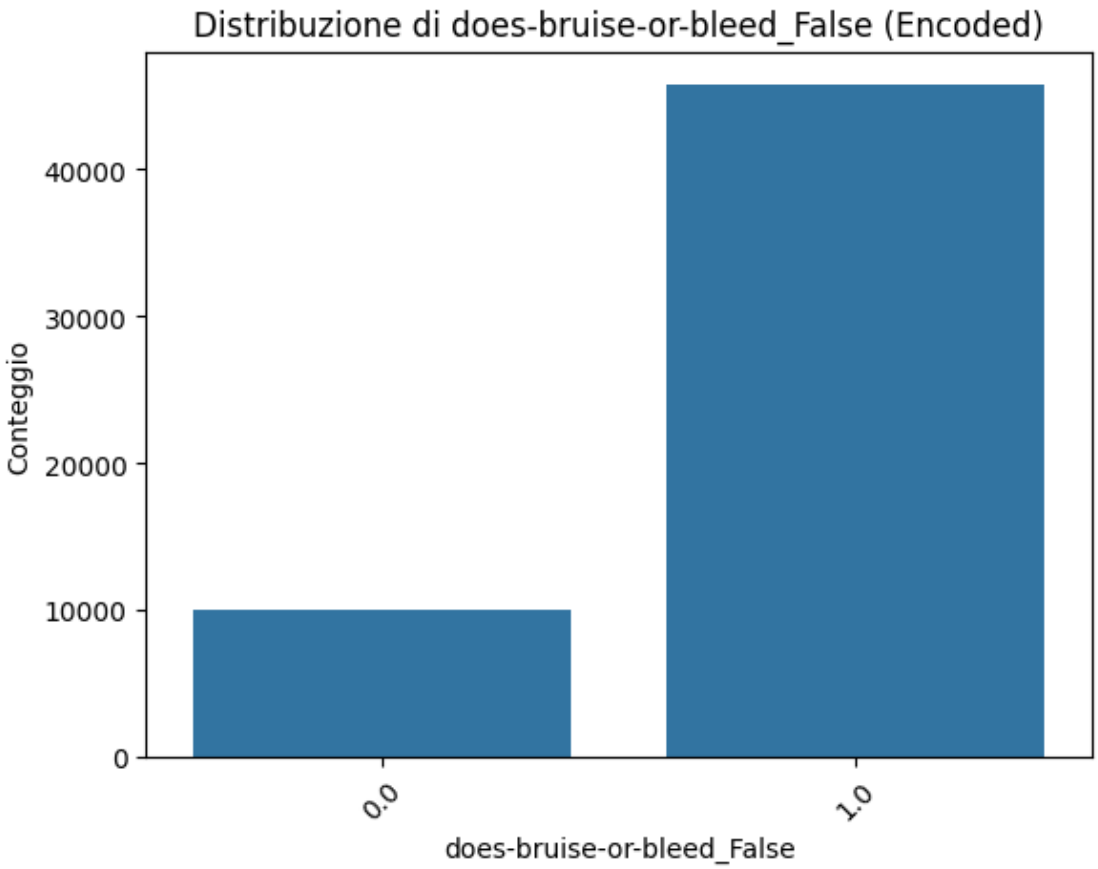

Distribuzione di cap-color_p (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_p

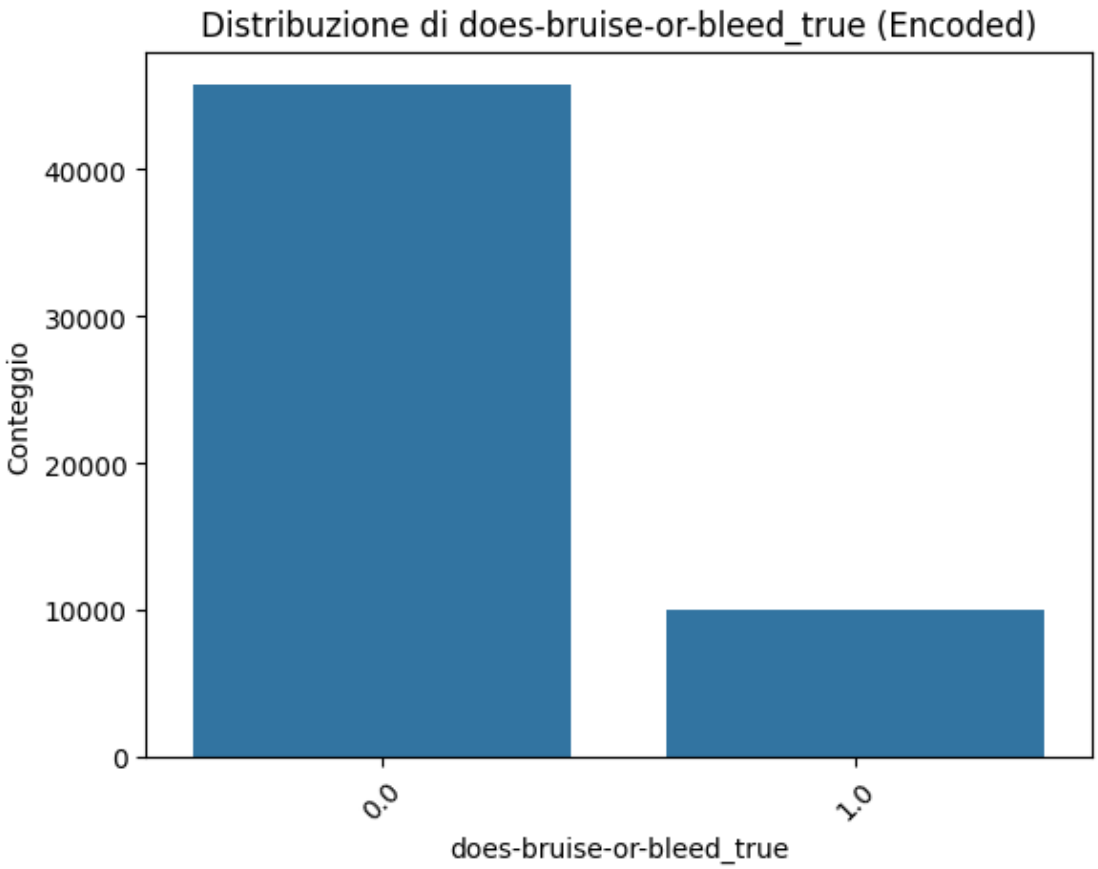

Distribuzione di cap-color_r (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_r

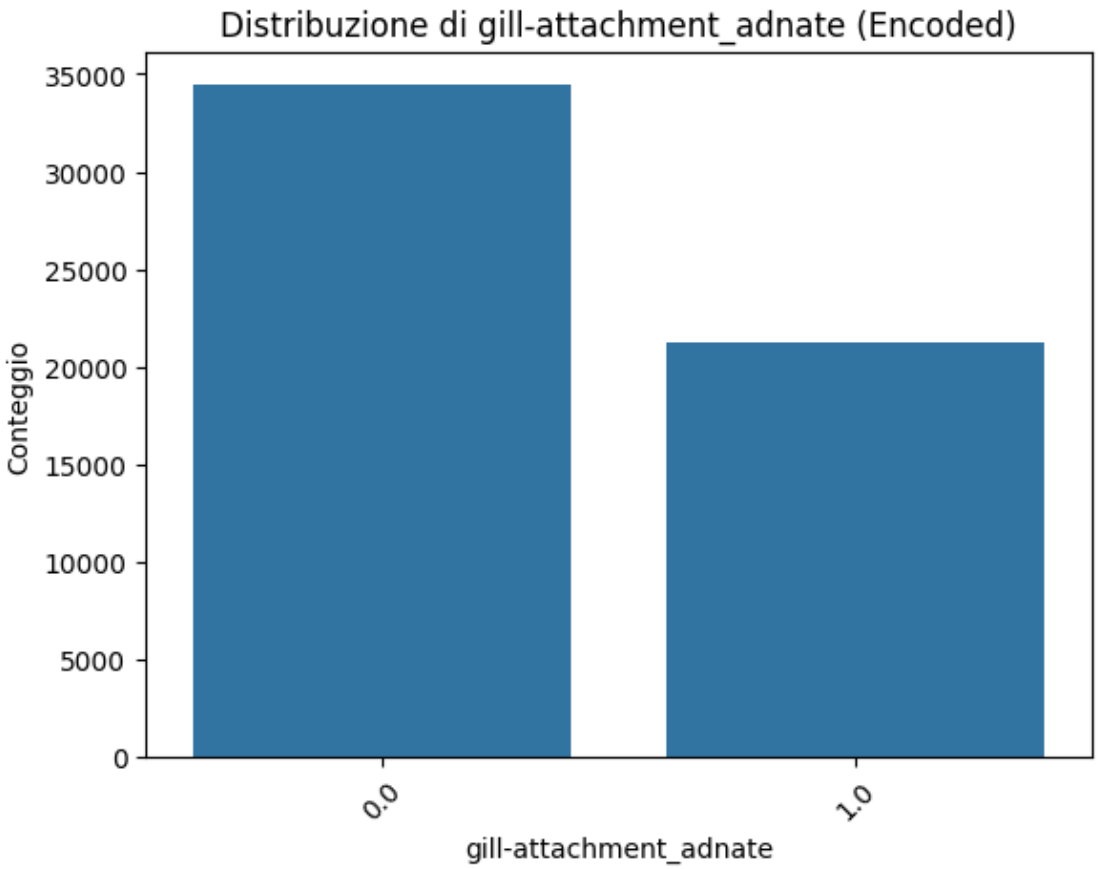

Distribuzione di cap-color_u (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_u

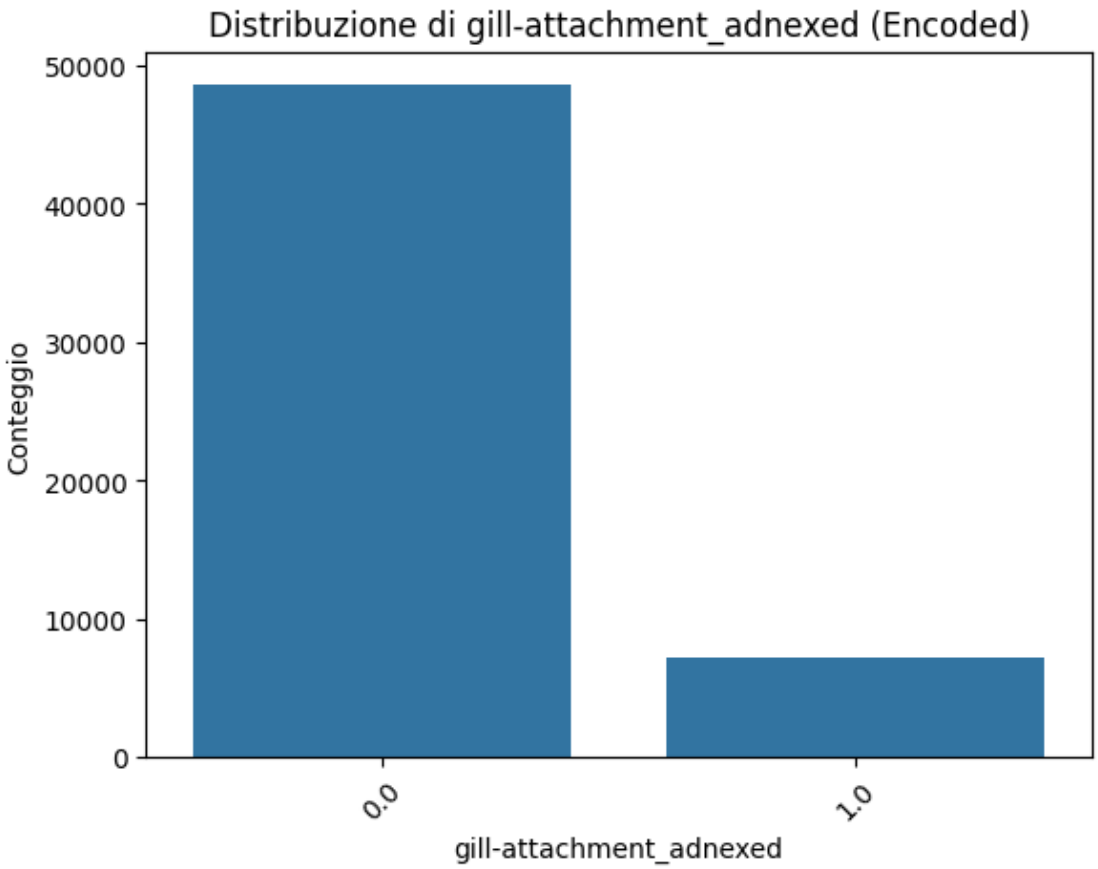

Distribuzione di cap-color_w (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
cap-color_w

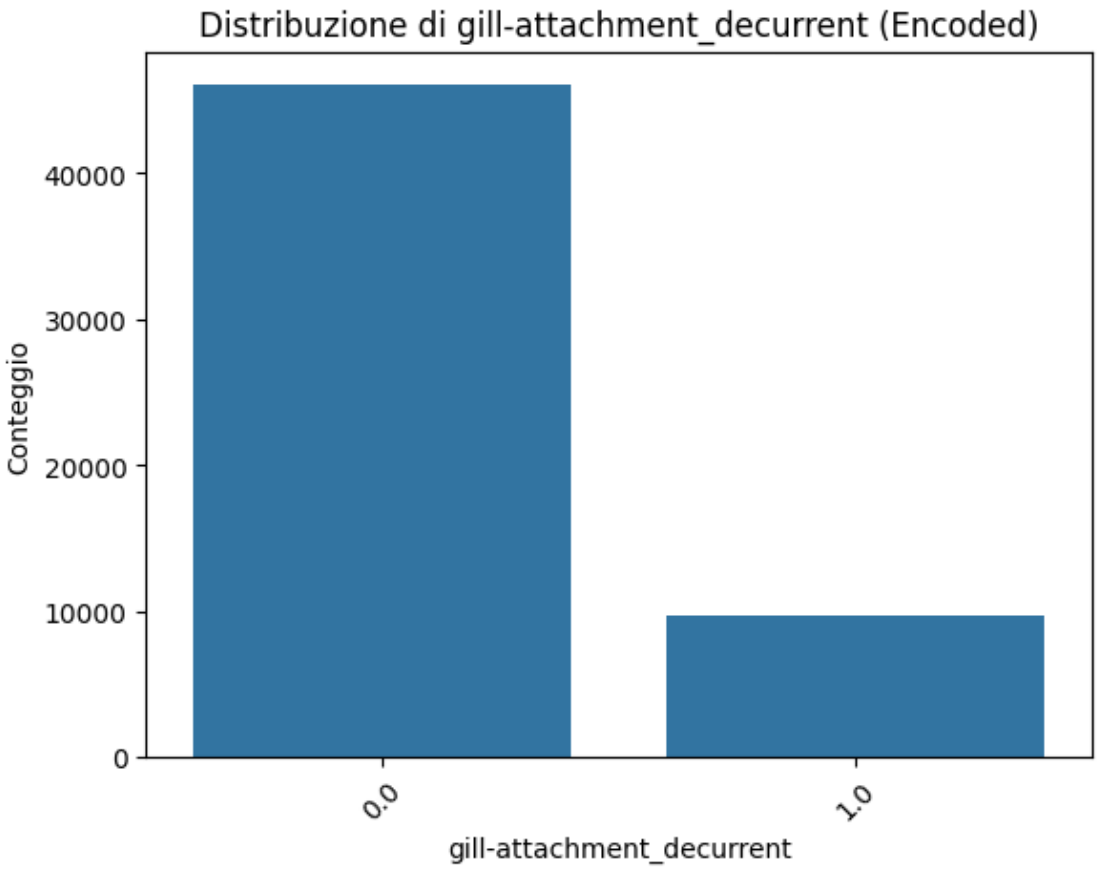

Distribuzione di cap-color_y (Encoded)
50000
40000
30000
20000
10000
0
Conteggio
0.0
1.0
cap-color_y

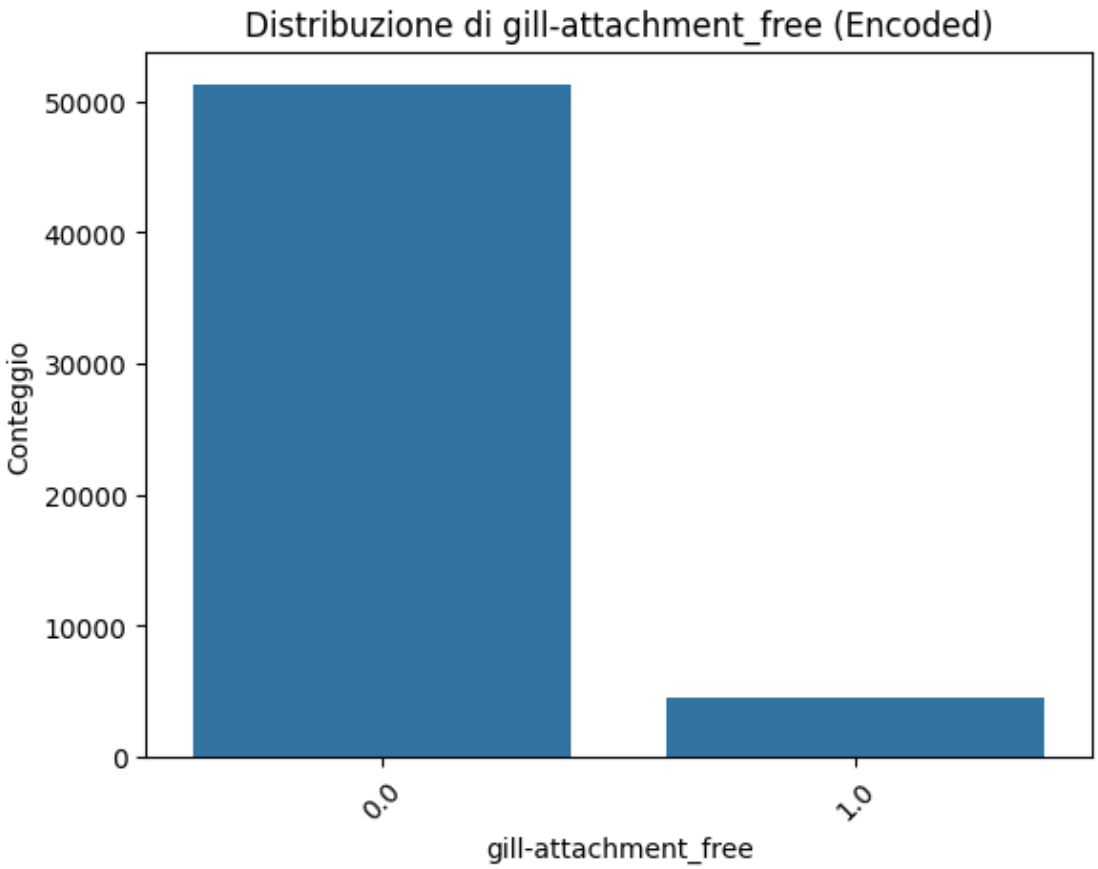

Distribuzione di does-bruise-or-bleed_False (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
does-bruise-or-bleed_False

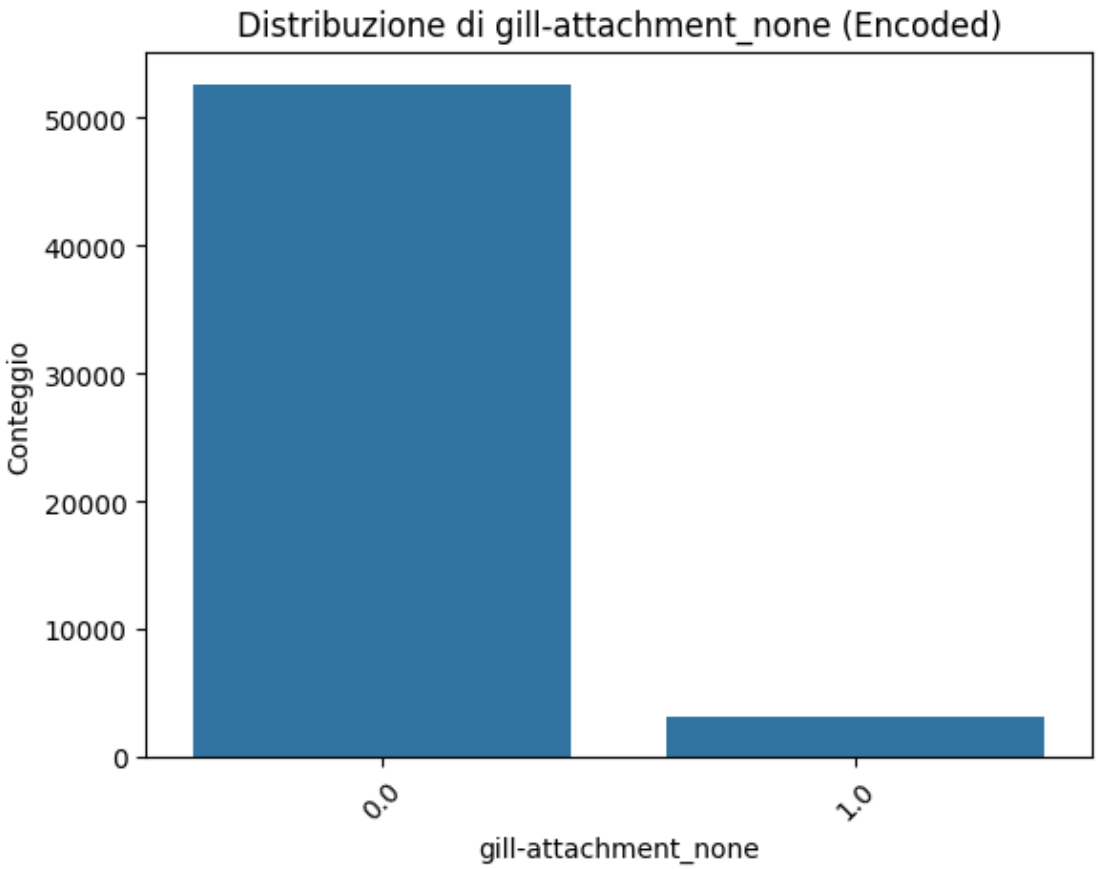

Distribuzione di does-bruise-or-bleed_true (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
does-bruise-or-bleed_true

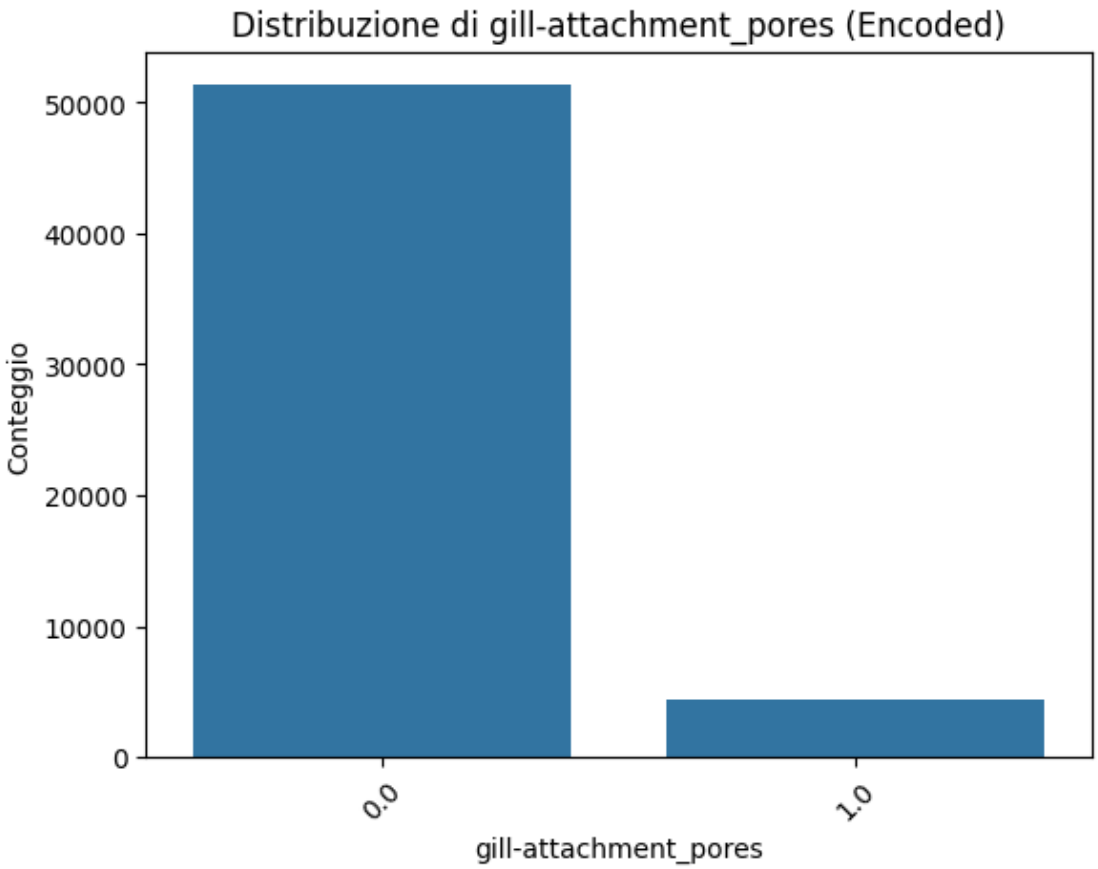

Distribuzione di gill-attachment_adnate (Encoded)
Conteggio
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
gill-attachment_adnate

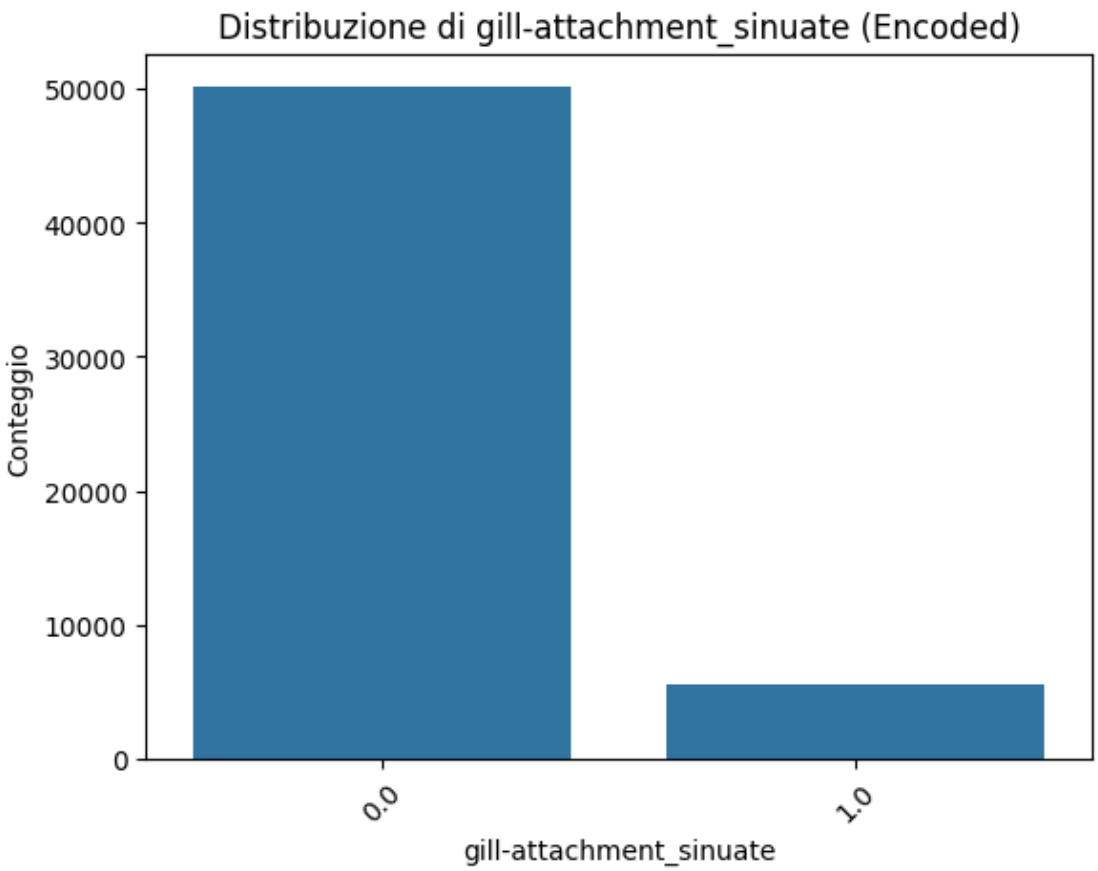

Distribuzione di gill-attachment_adnexed (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-attachment_adnexed


Distribuzione di gill-attachment_decurrent (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
gill-attachment_decurrent


Distribuzione di gill-attachment_free (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-attachment_free


Distribuzione di gill-attachment_none (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-attachment_none

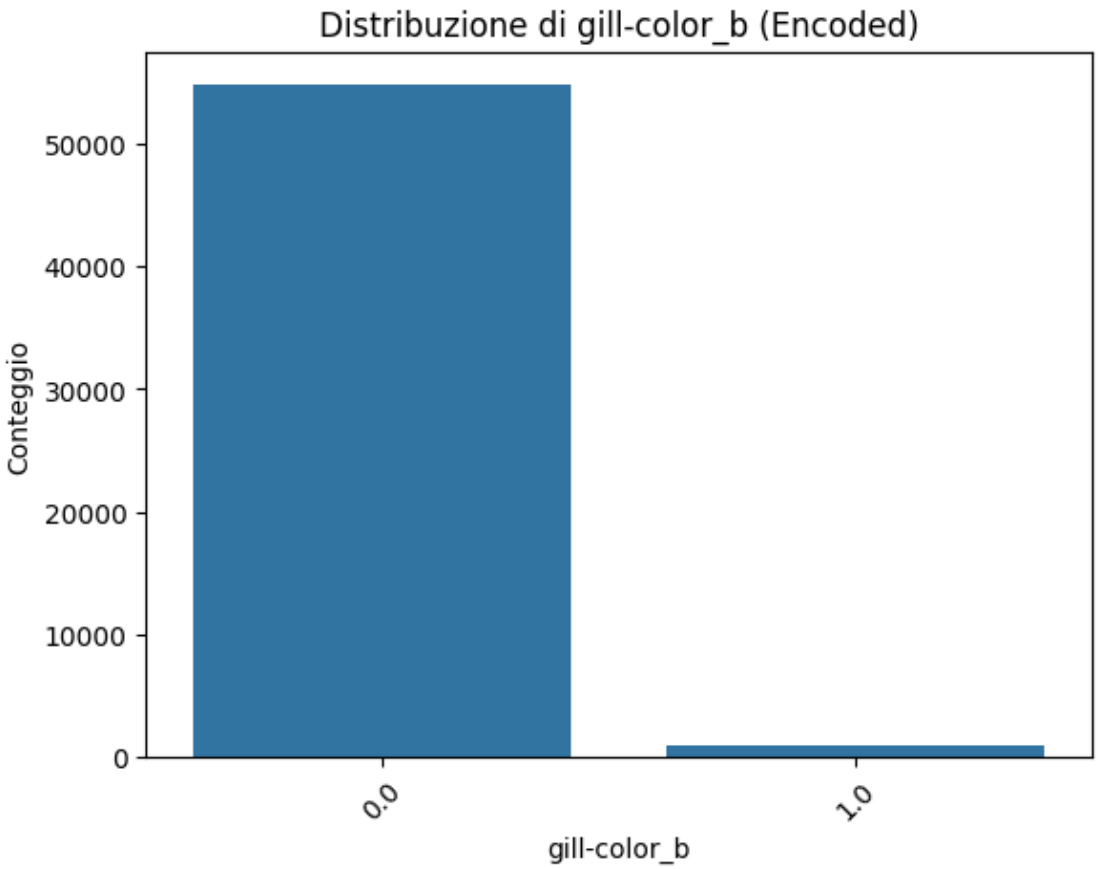

Distribuzione di gill-attachment_pores (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-attachment_pores

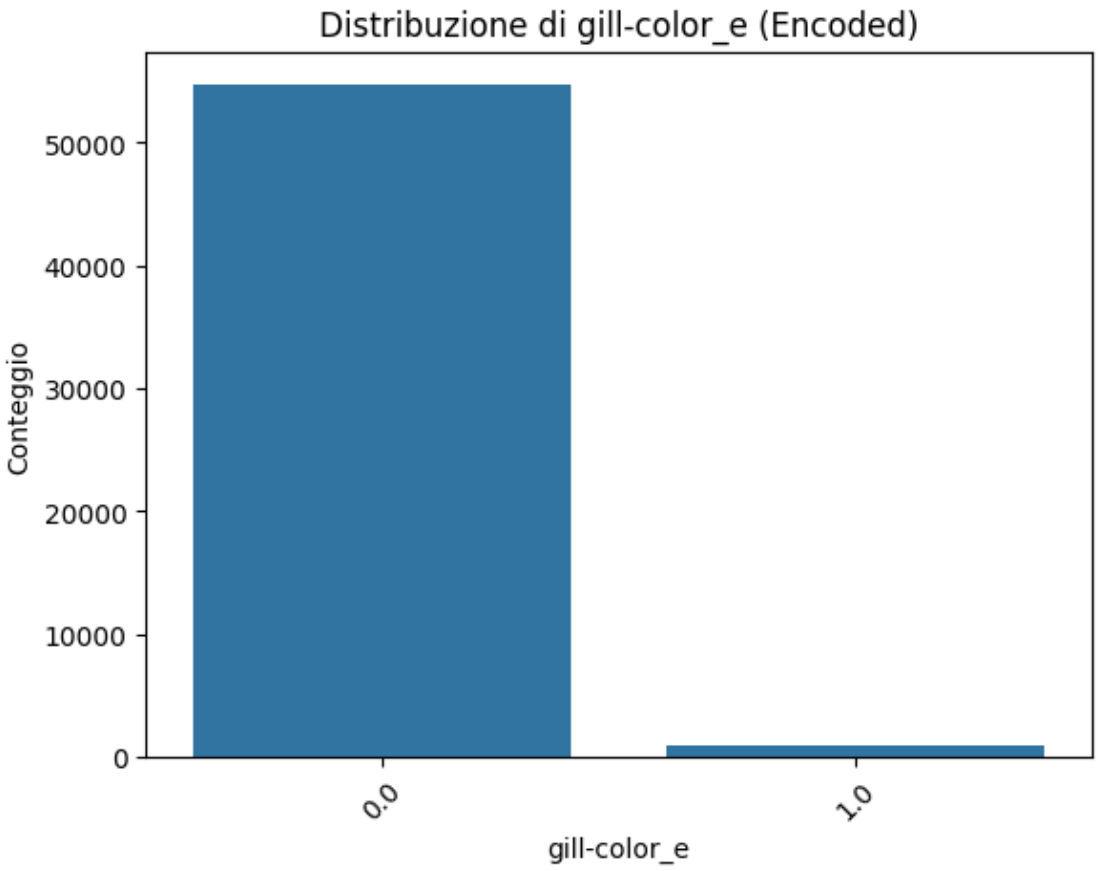

Distribuzione di gill-attachment_sinuate (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-attachment_sinuate

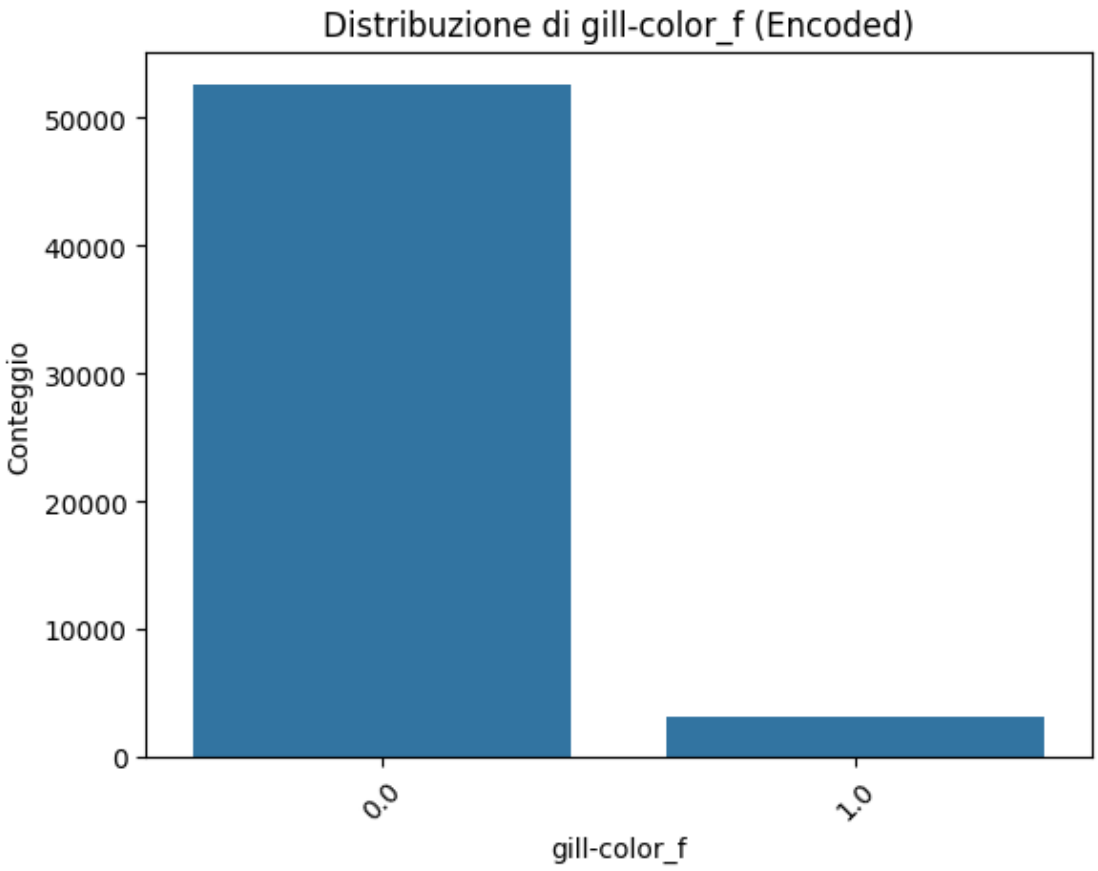

Distribuzione di gill-spacing_close (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
gill-spacing_close

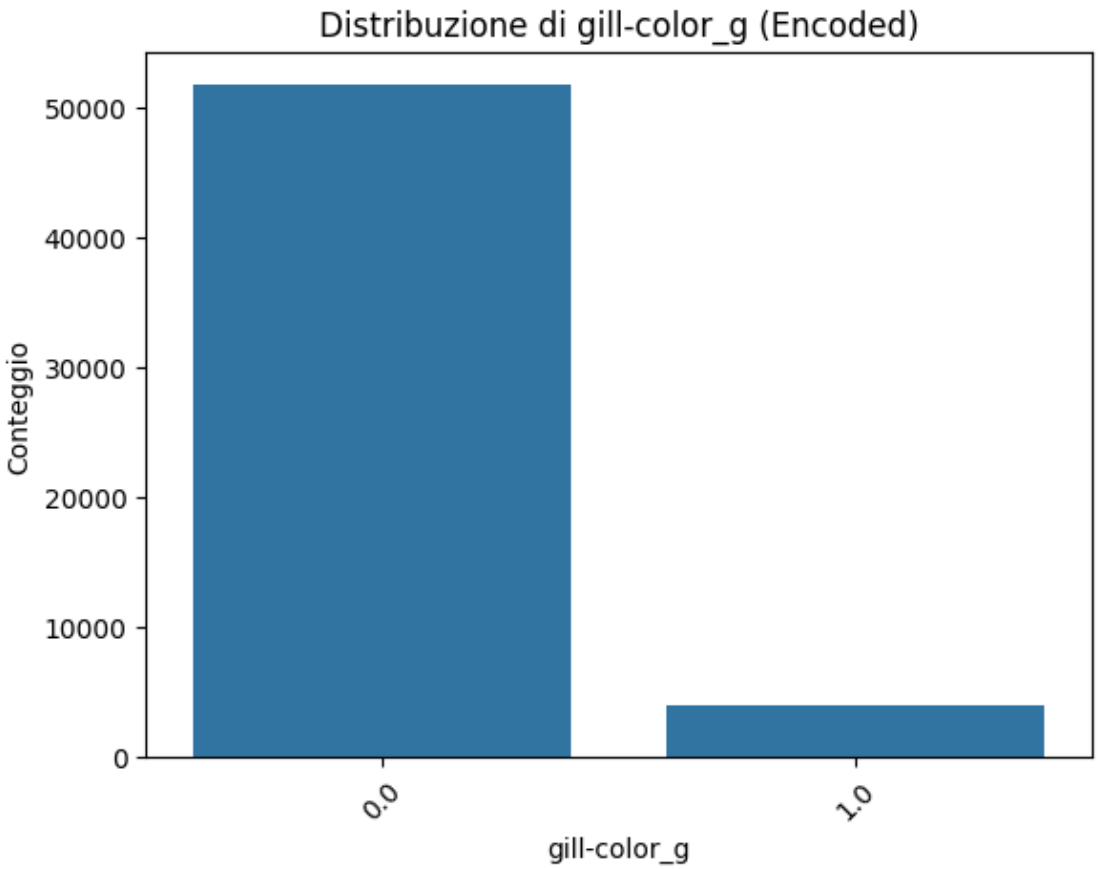

Distribuzione di gill-spacing_distant (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-spacing_distant

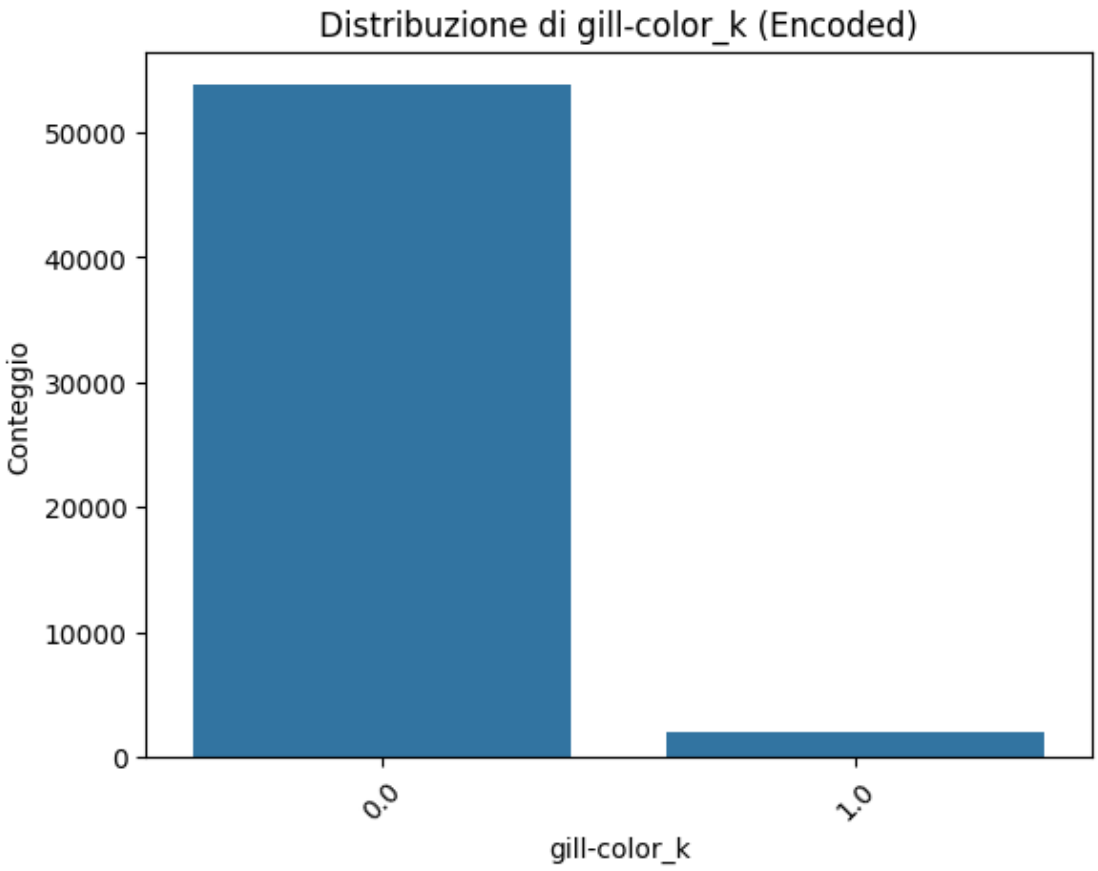

Distribuzione di gill-spacing_none (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-spacing_none

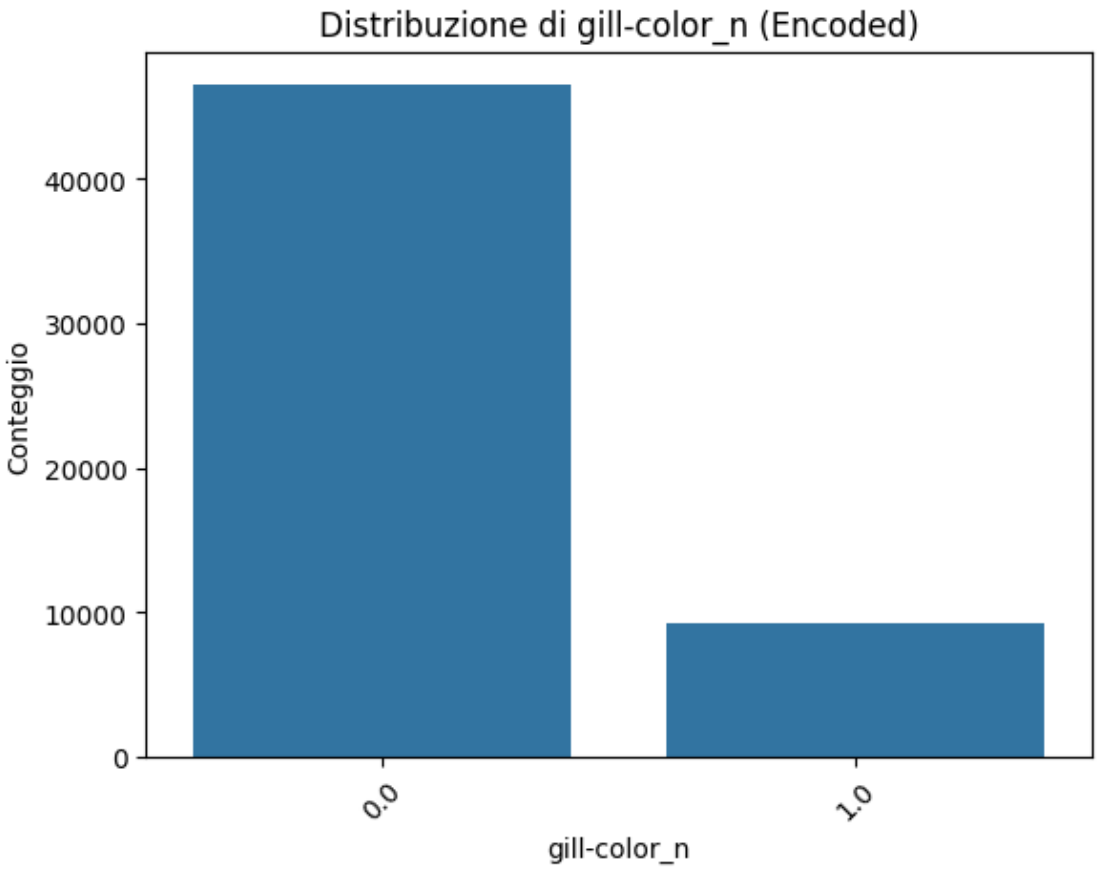

Distribuzione di gill-color_b (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_b

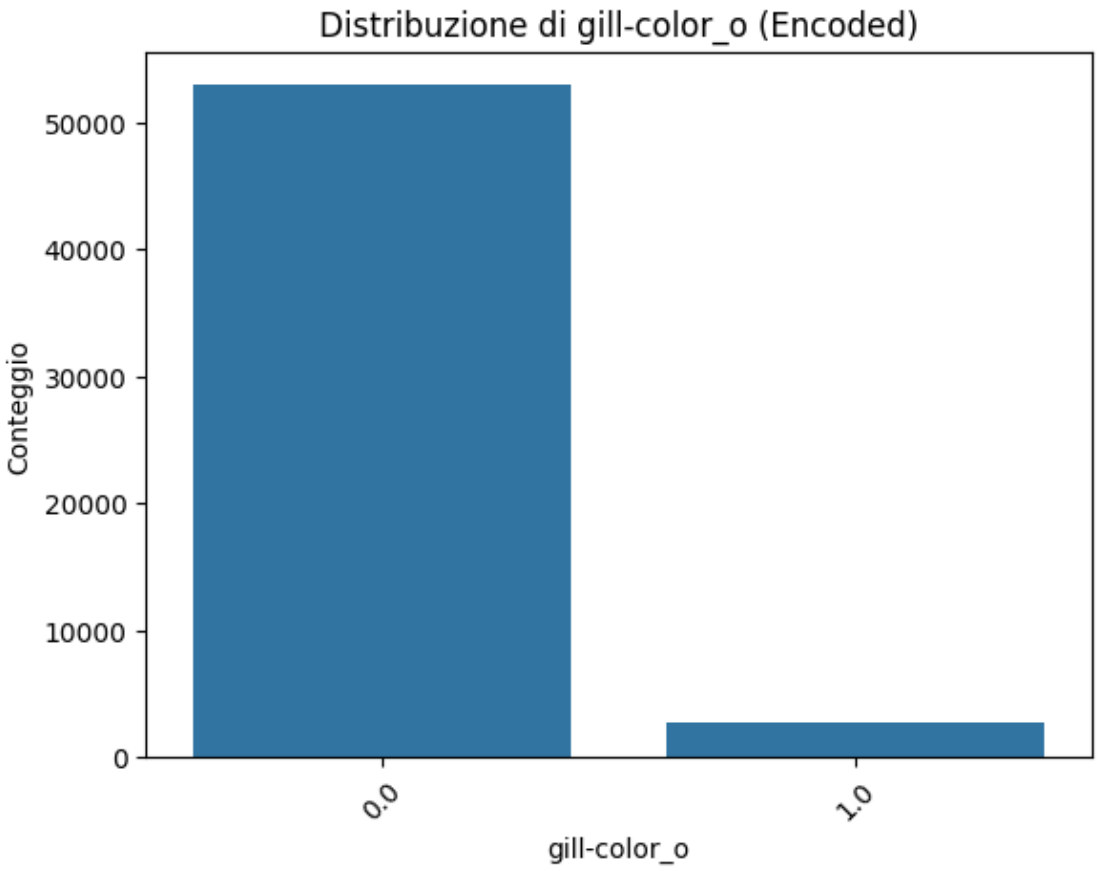

Distribuzione di gill-color_e (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_e

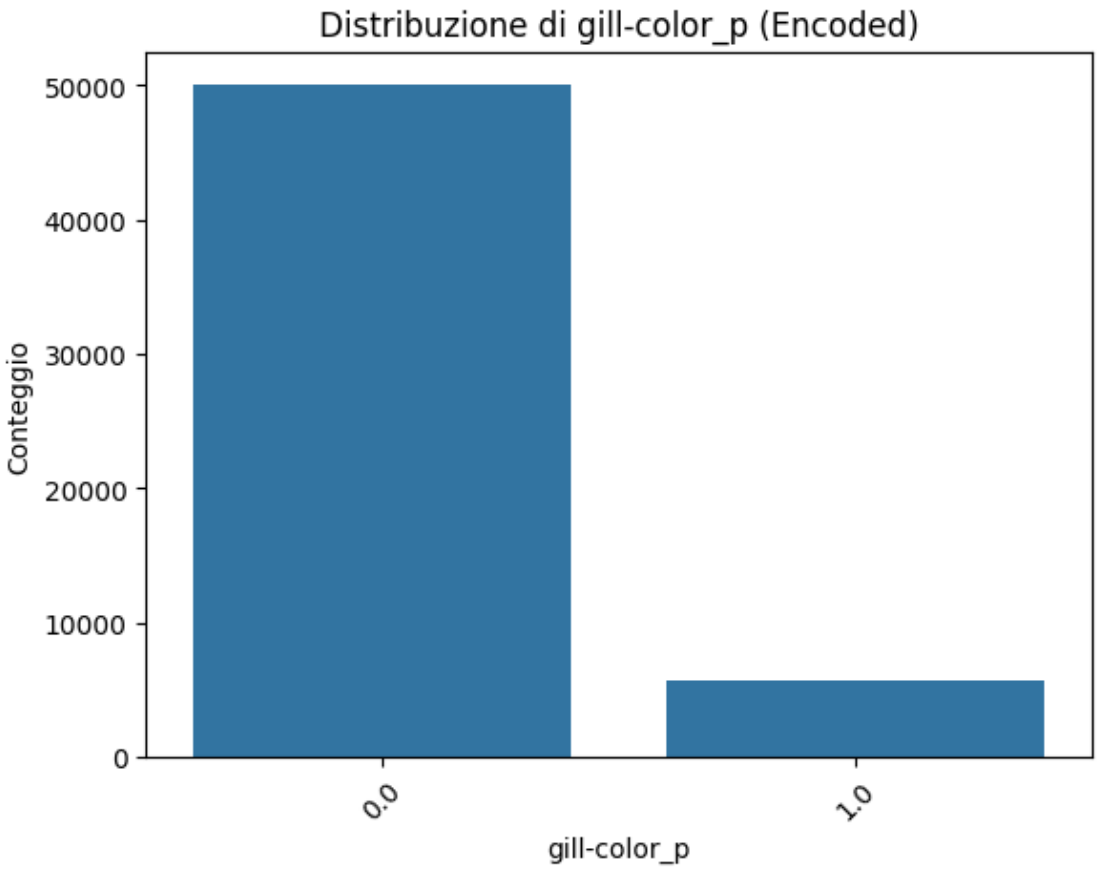

Distribuzione di gill-color_f (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_f

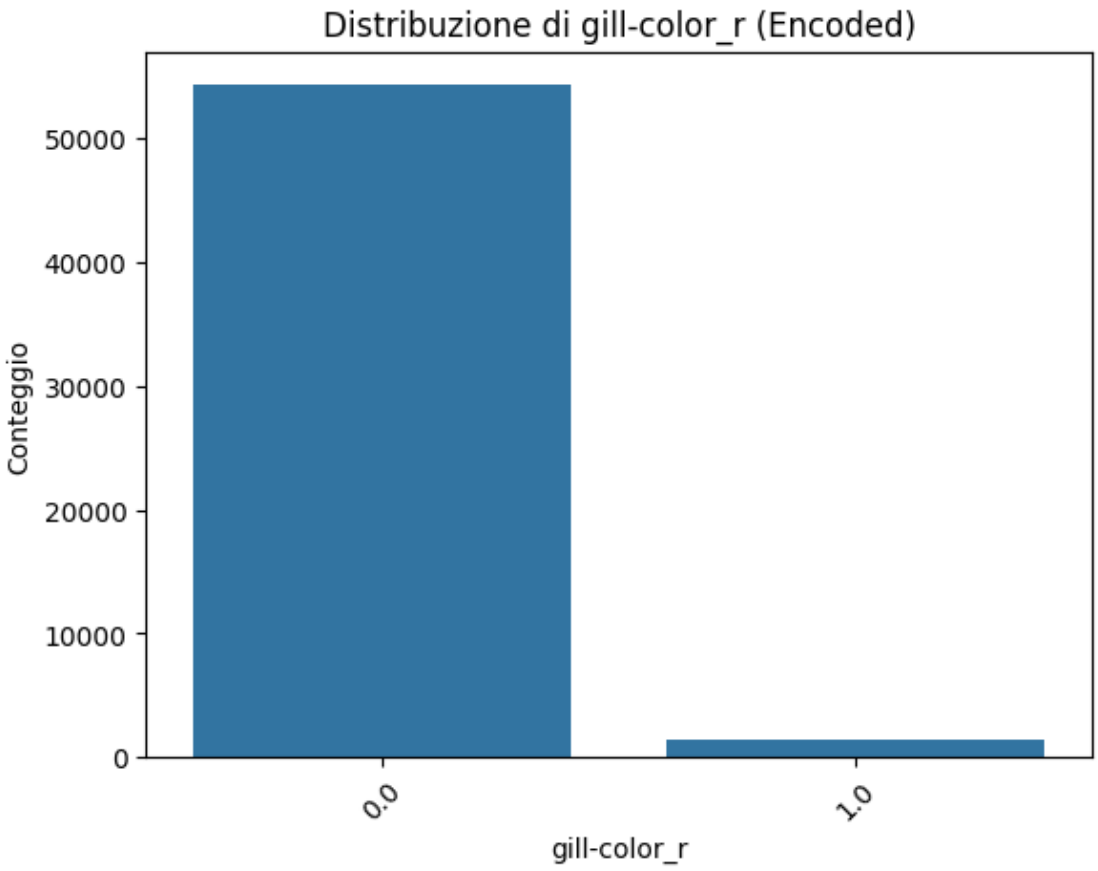

Distribuzione di gill-color_g (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_g

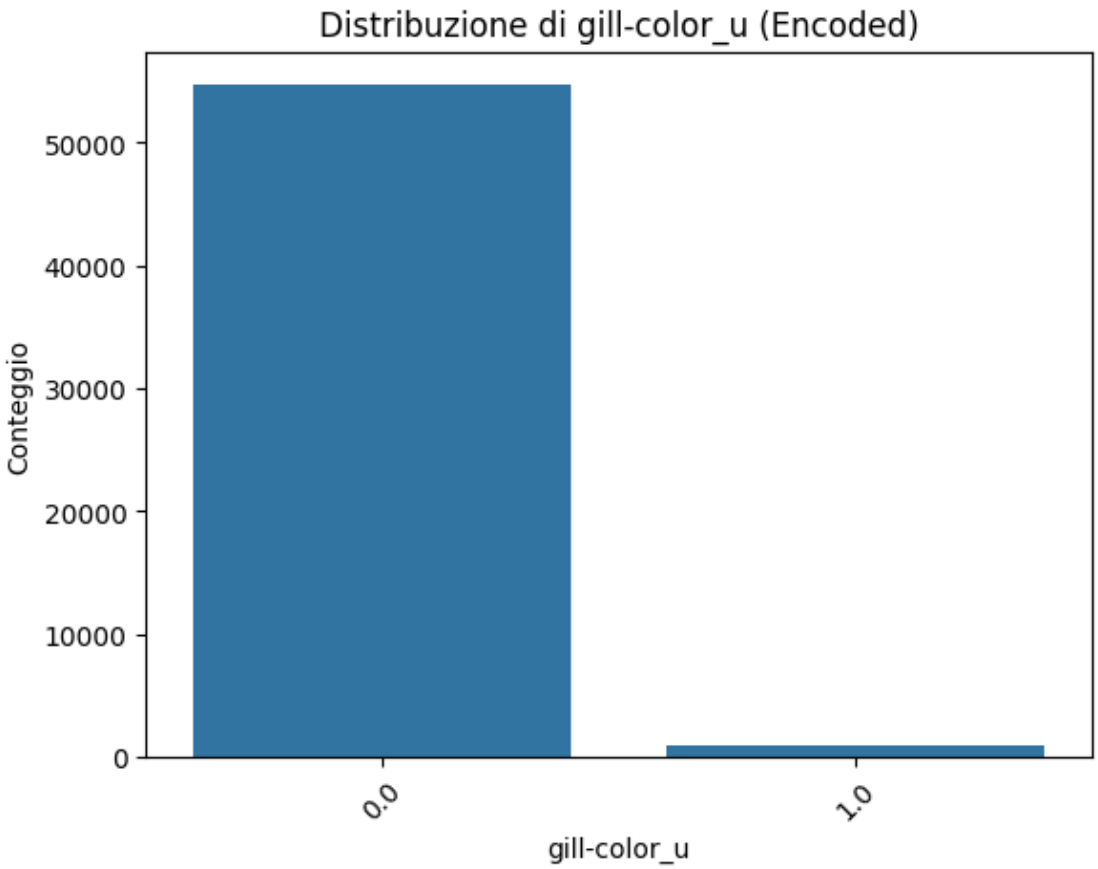

Distribuzione di gill-color_k (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_k

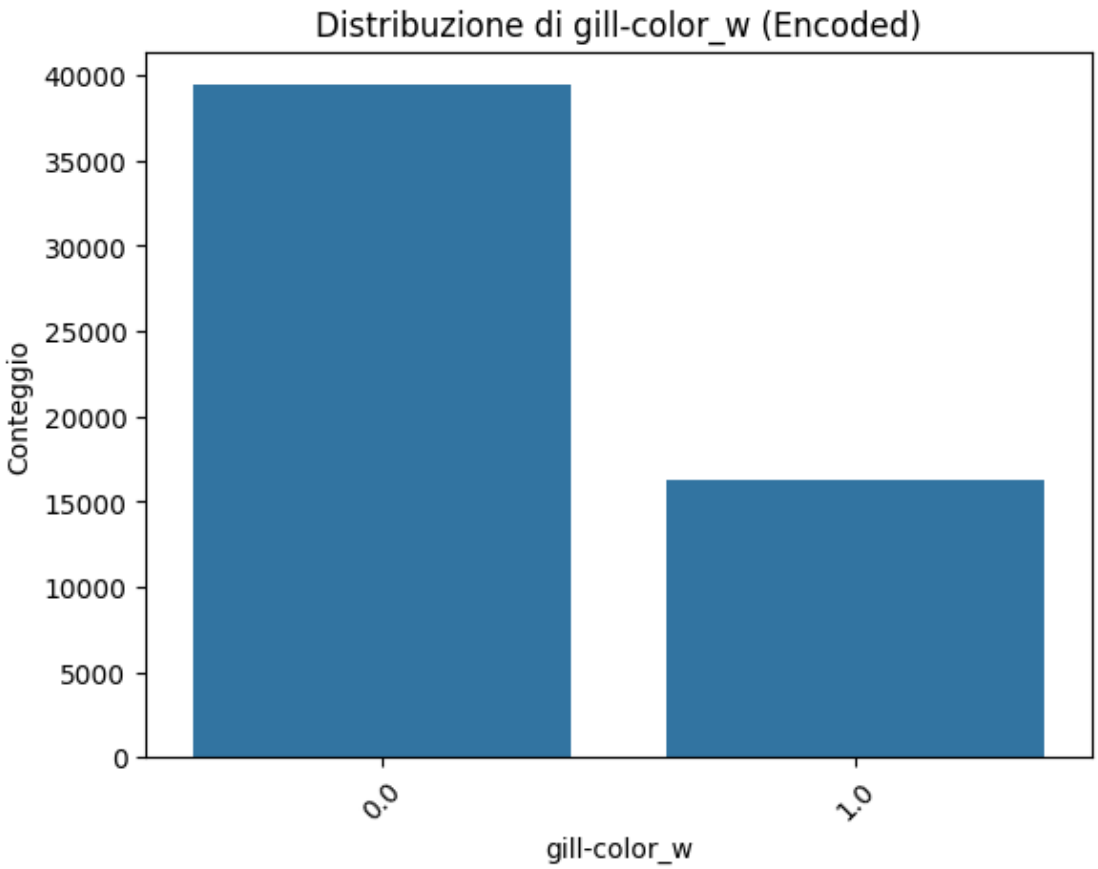

Distribuzione di gill-color_n (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
gill-color_n

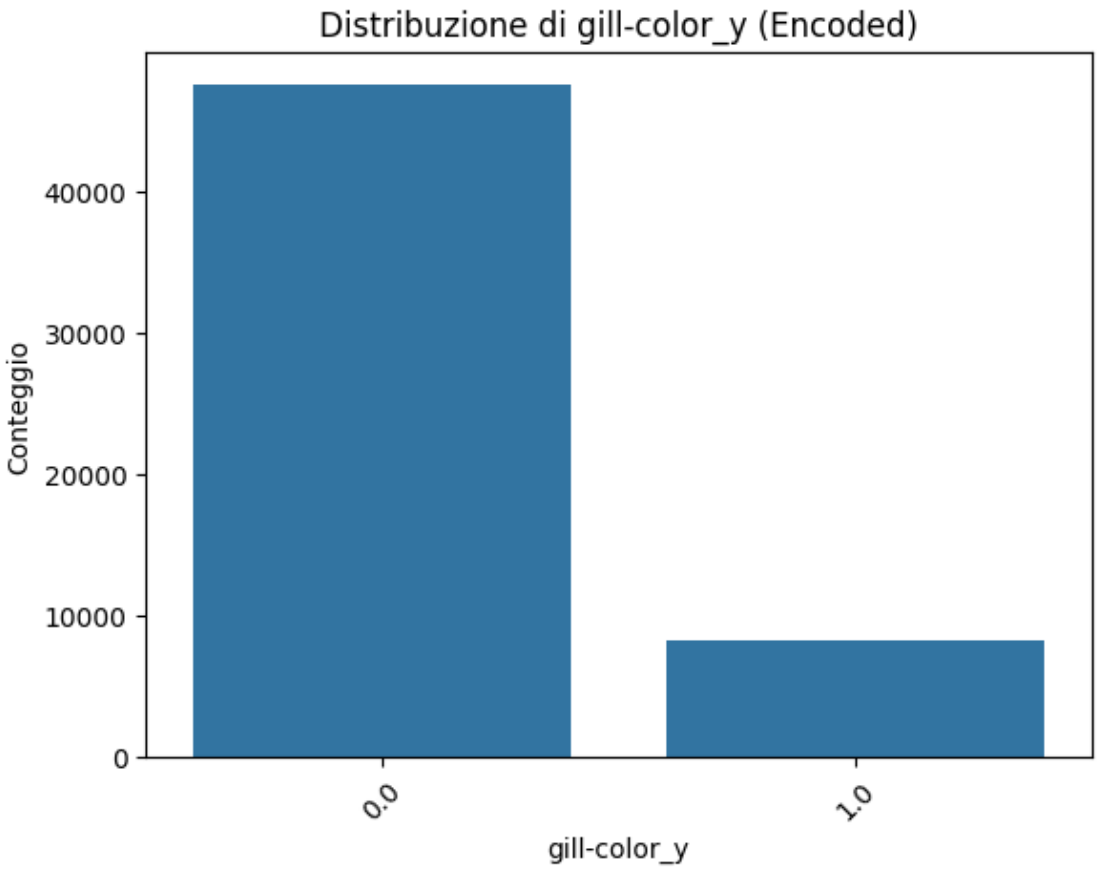

Distribuzione di gill-color_o (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_o

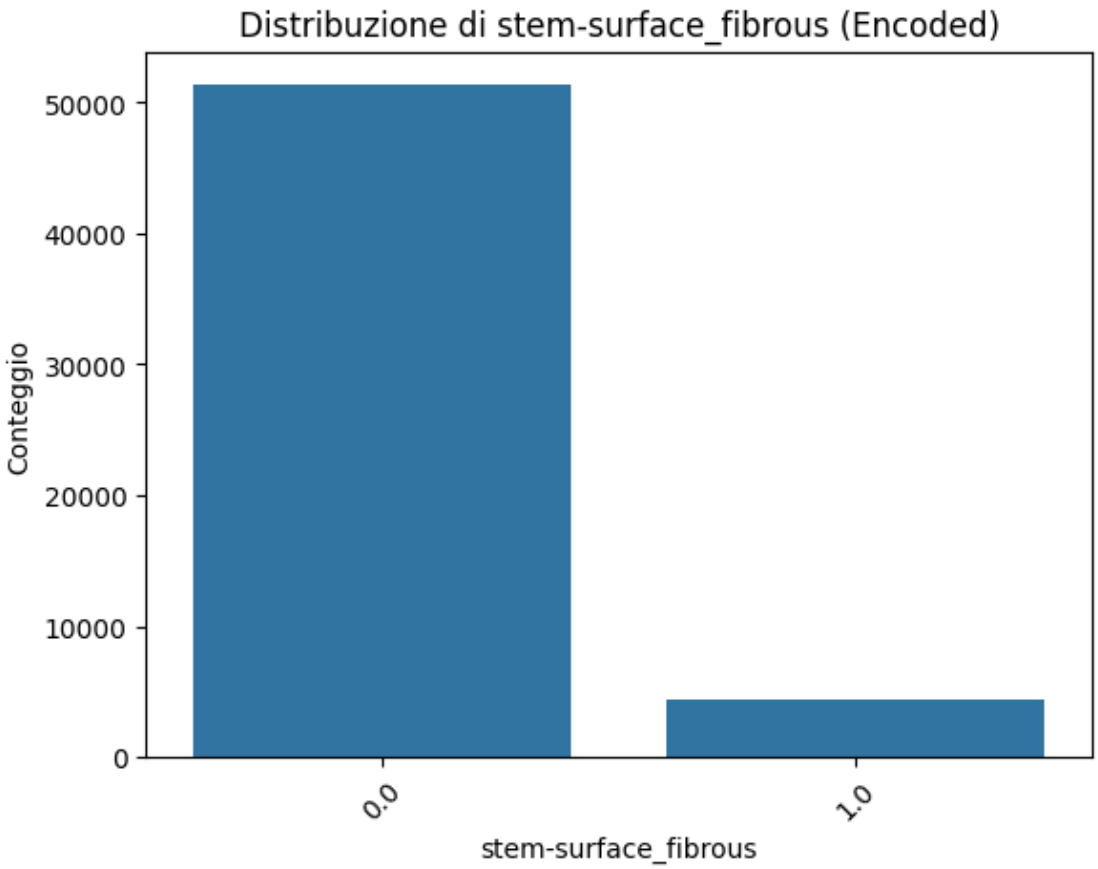

Distribuzione di gill-color_p (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_p

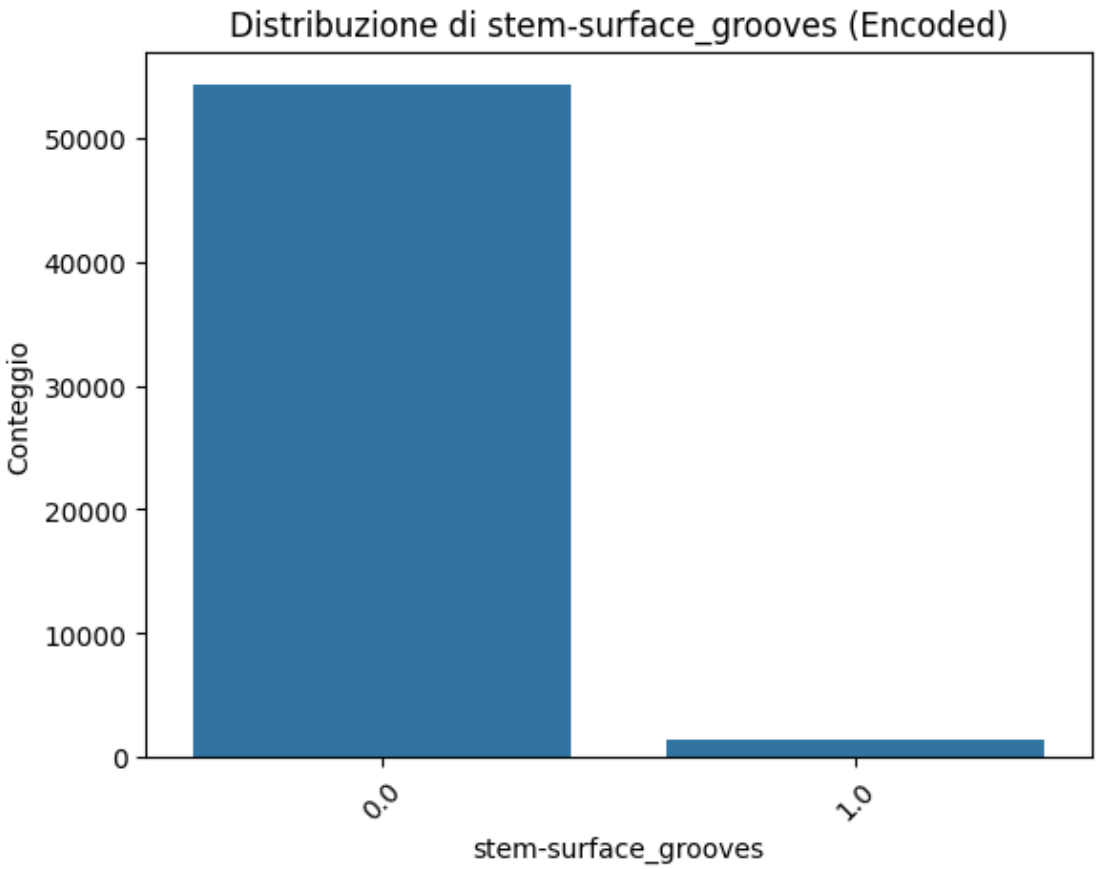

Distribuzione di gill-color_r (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_r

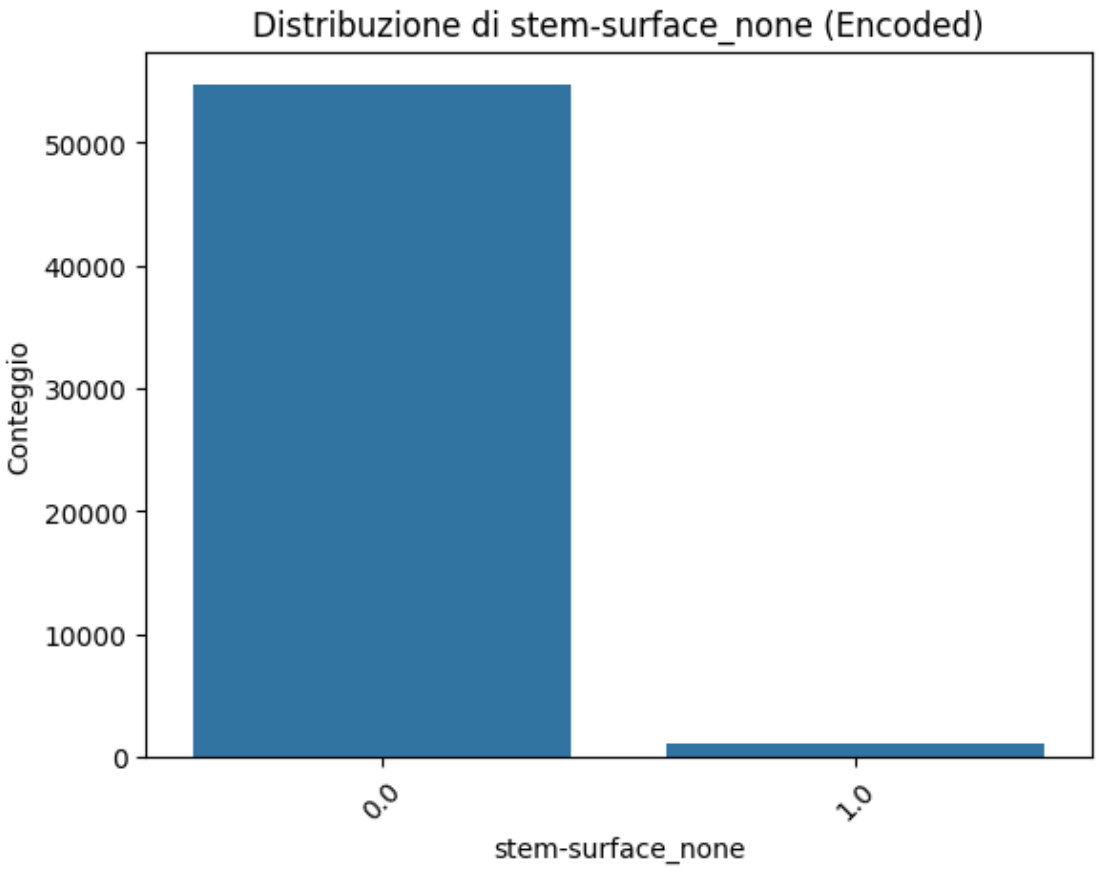

Distribuzione di gill-color_u (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
gill-color_u

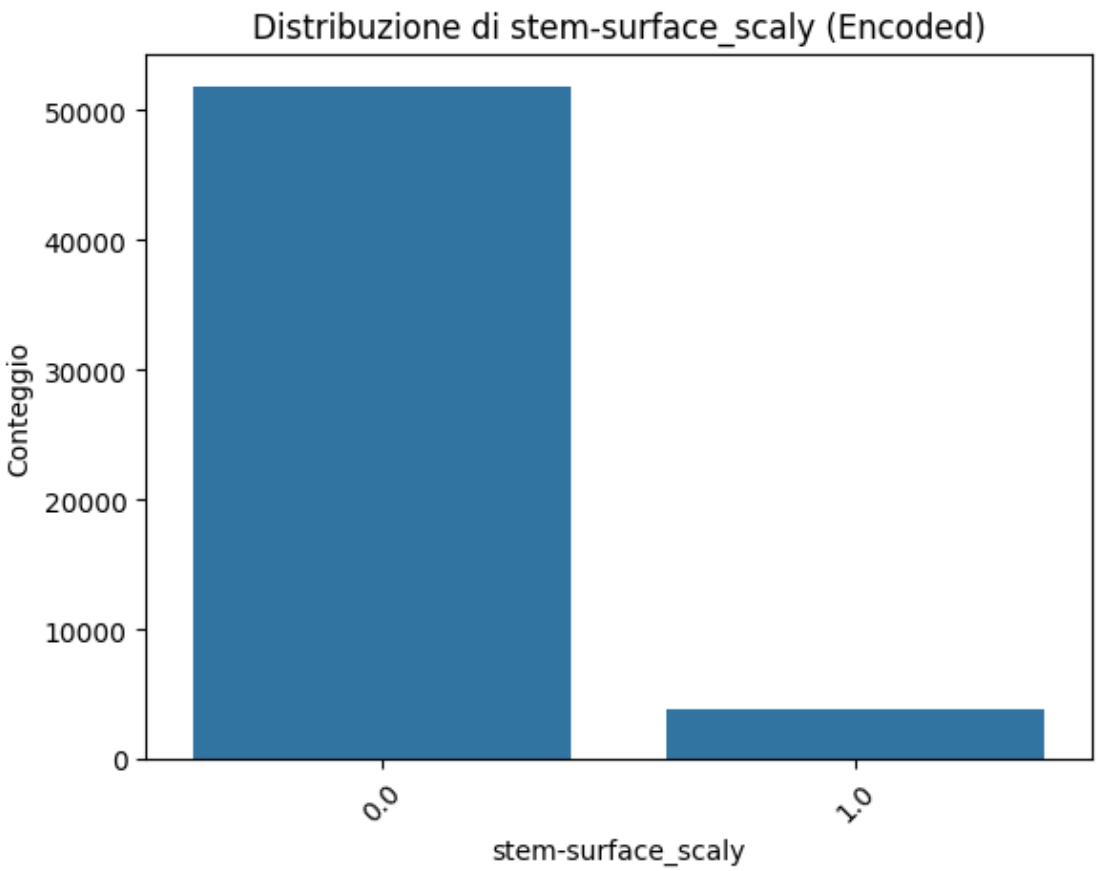

Distribuzione di gill-color_w (Encoded)
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
gill-color_w

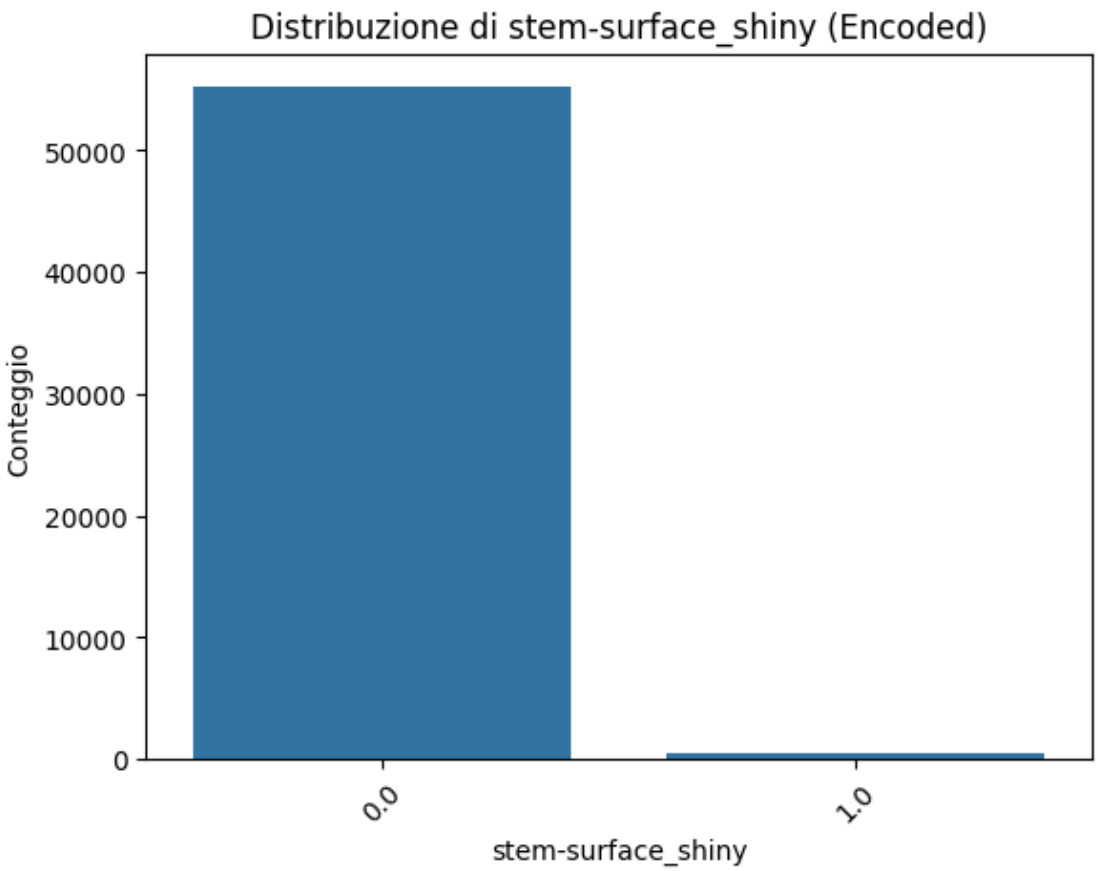

Distribuzione di gill-color_y (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
gill-color_y

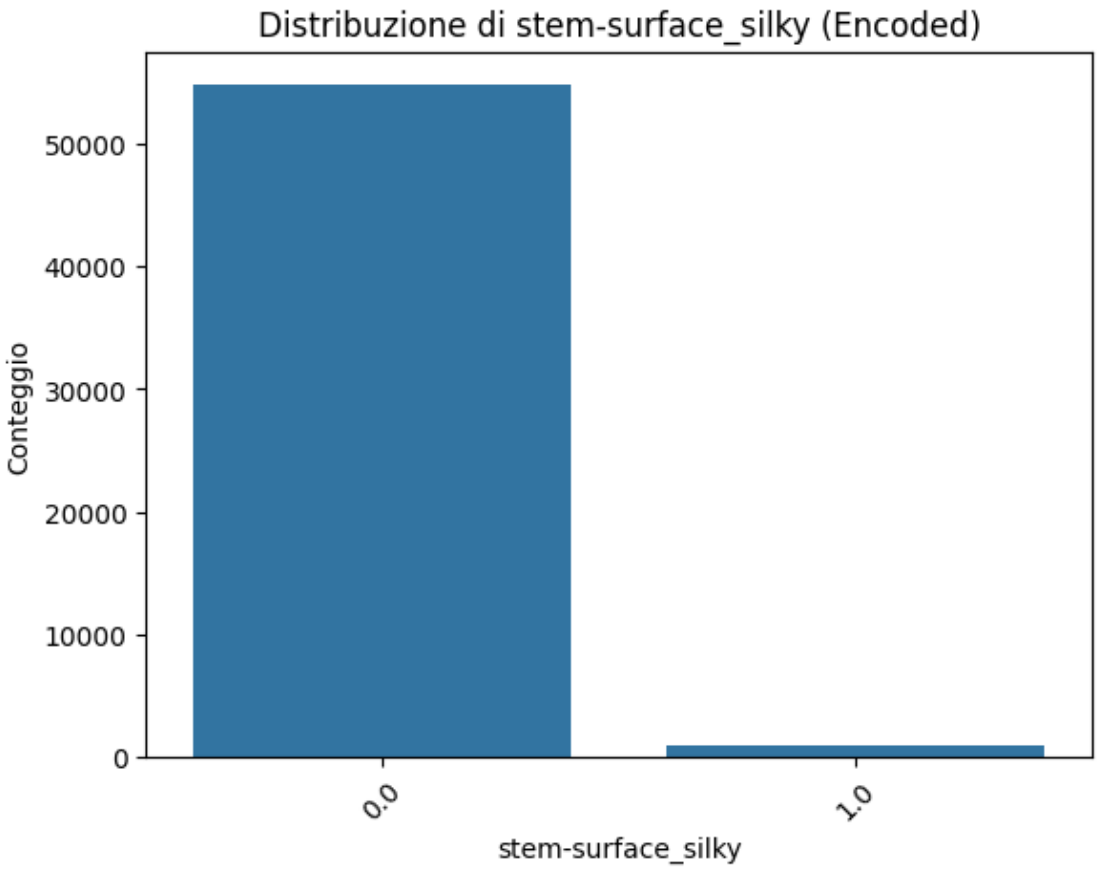

Distribuzione di stem-surface_fibrous (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_fibrous

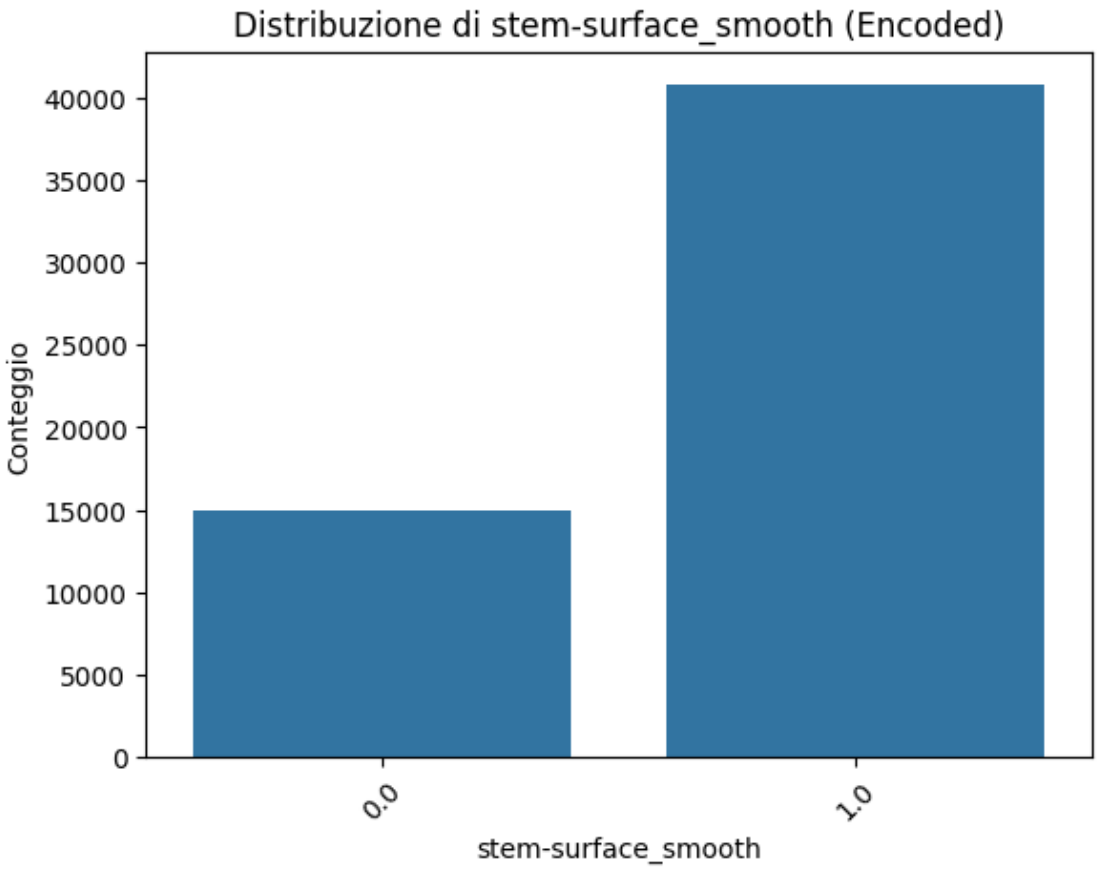

Distribuzione di stem-surface_grooves (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_grooves

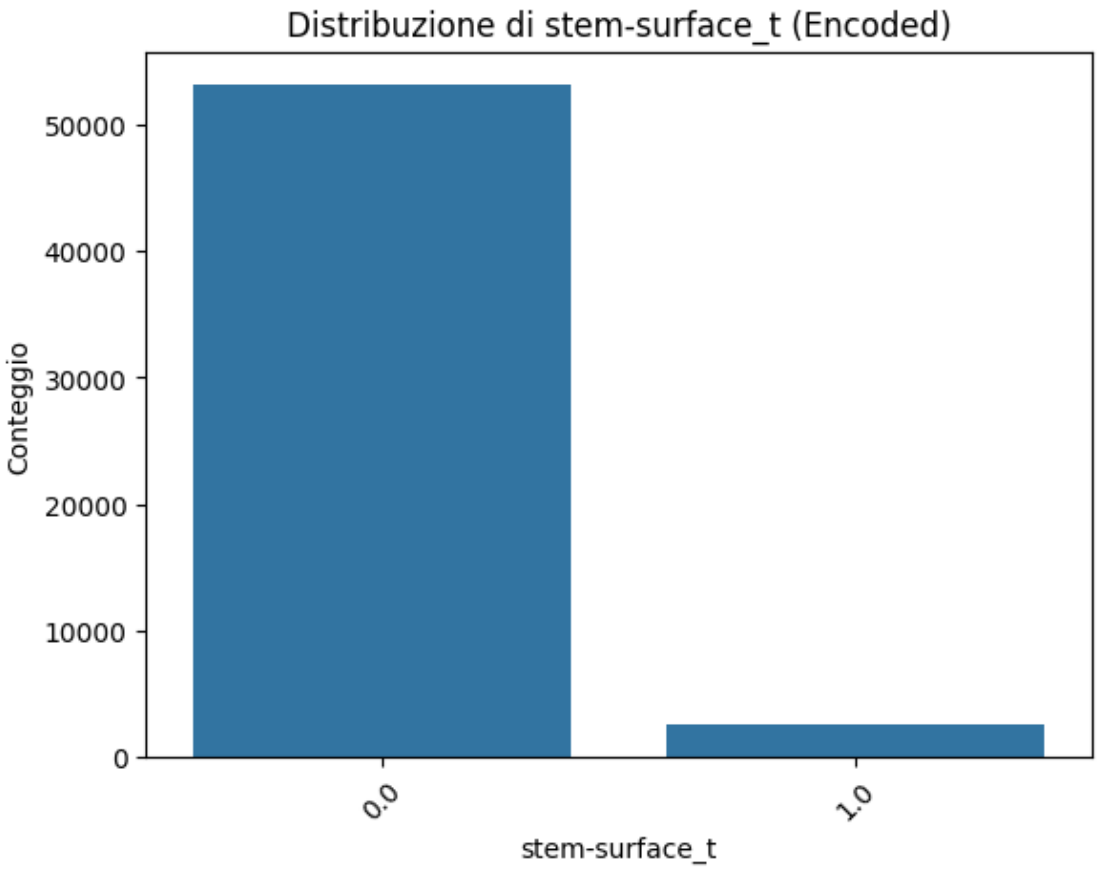

Distribuzione di stem-surface_none (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_none

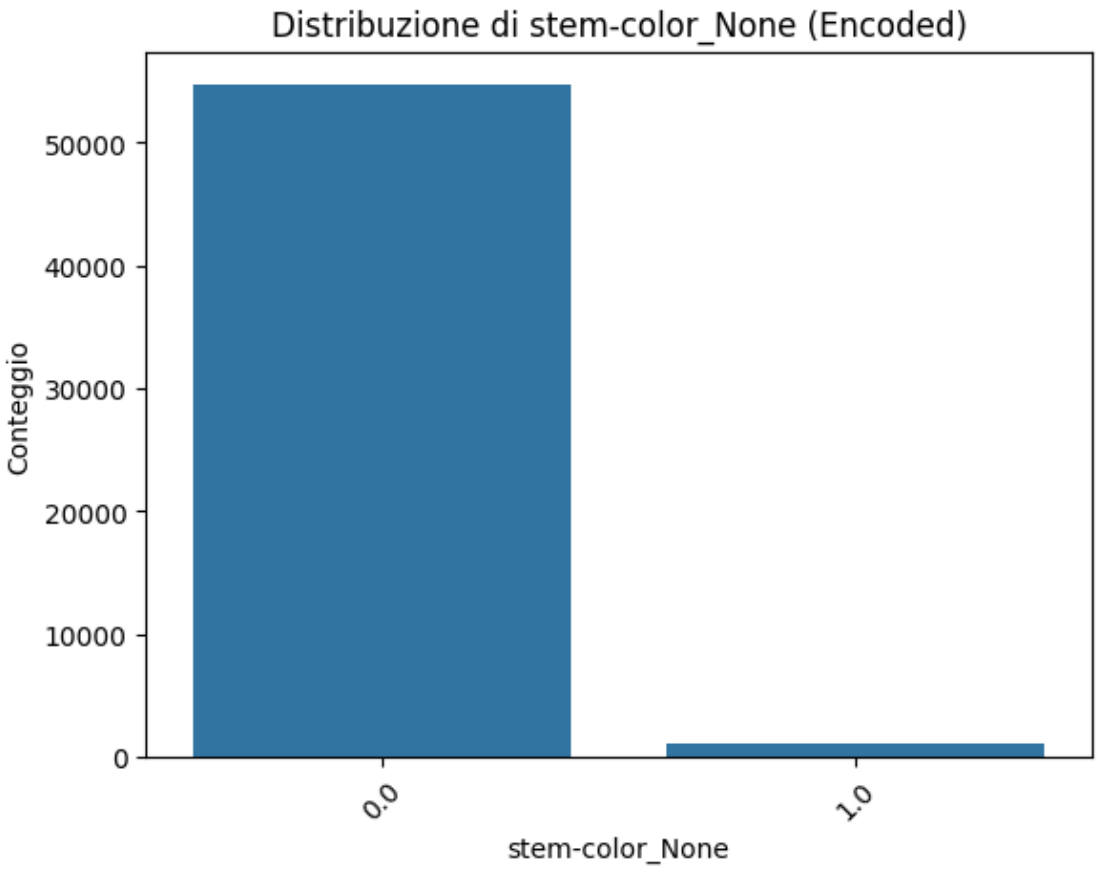

Distribuzione di stem-surface_scaly (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_scaly

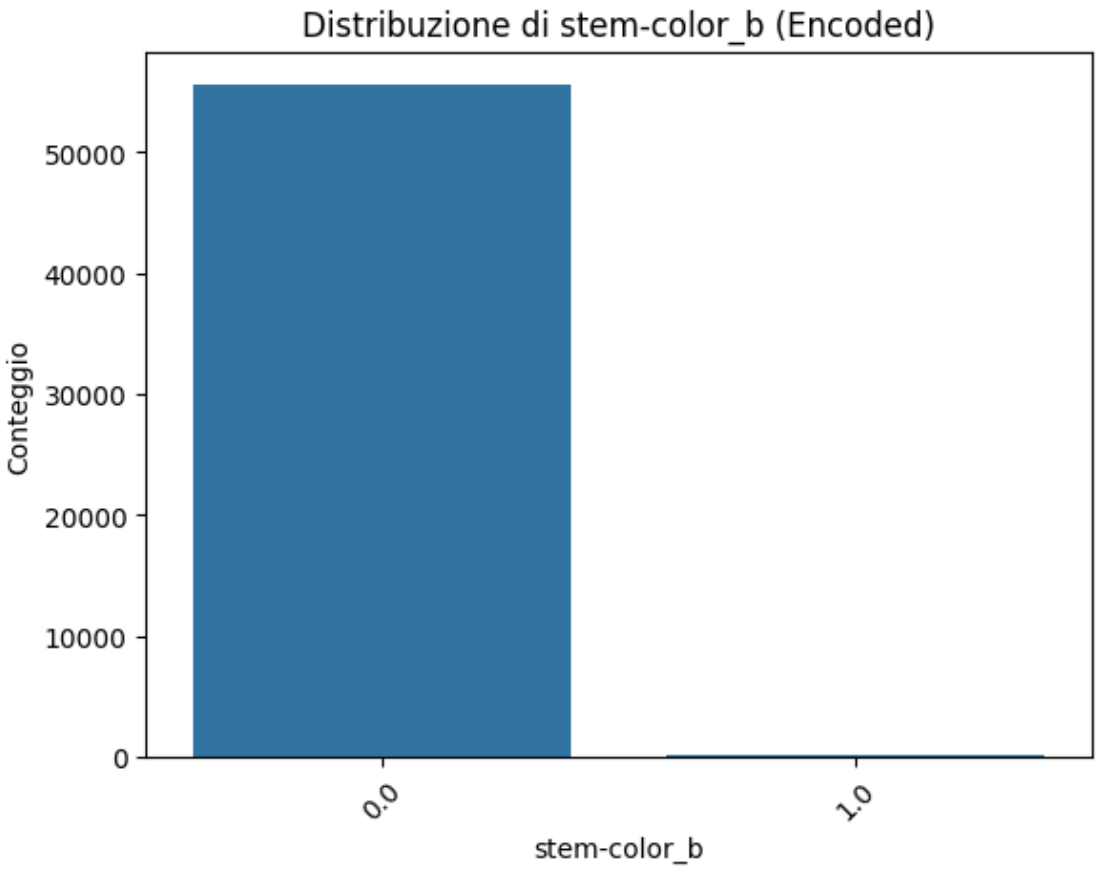

Distribuzione di stem-surface_shiny (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_shiny

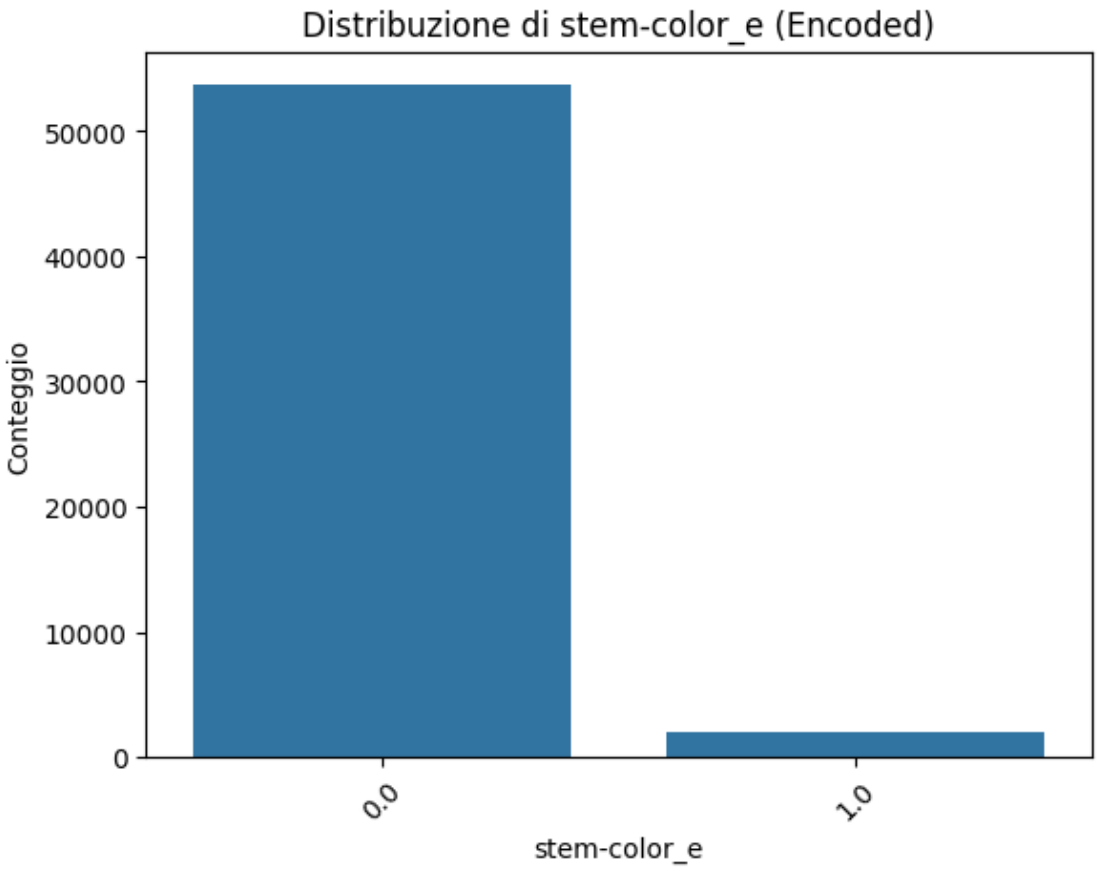

Distribuzione di stem-surface_silky (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_silky

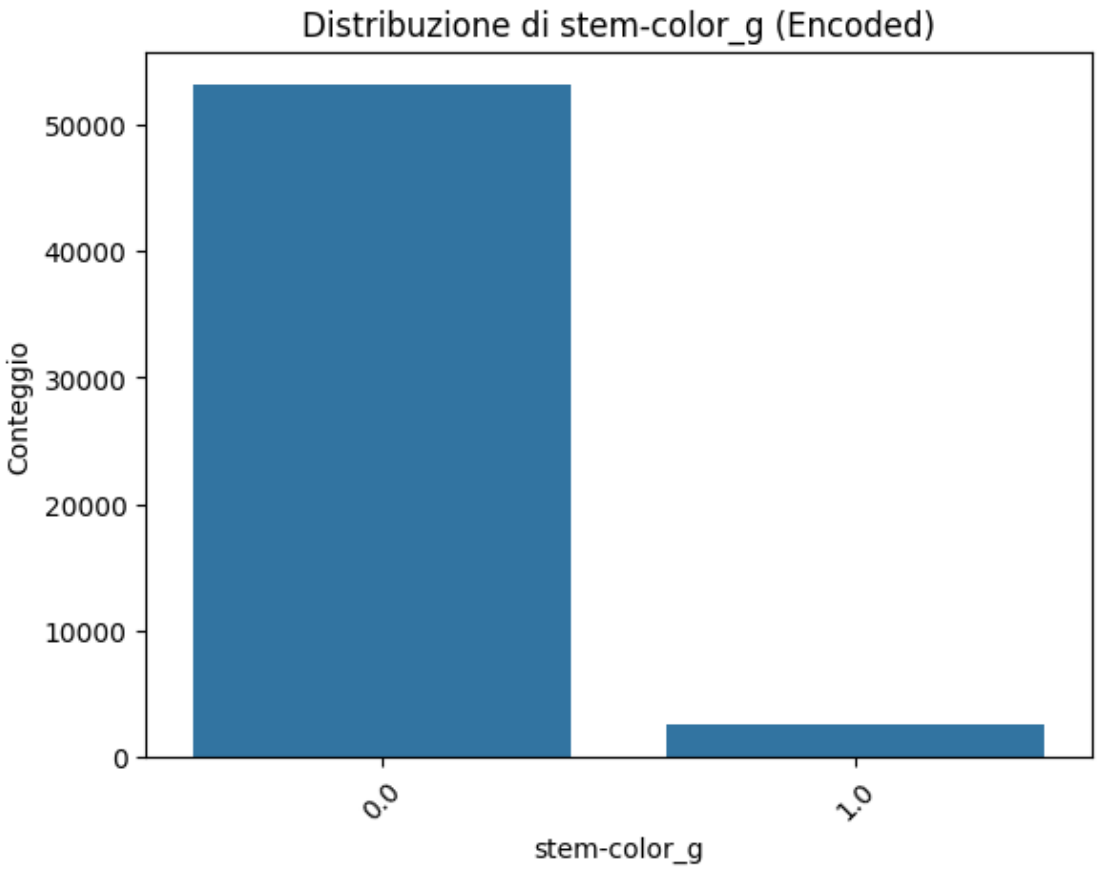

Distribuzione di stem-surface_smooth (Encoded)
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
stem-surface_smooth

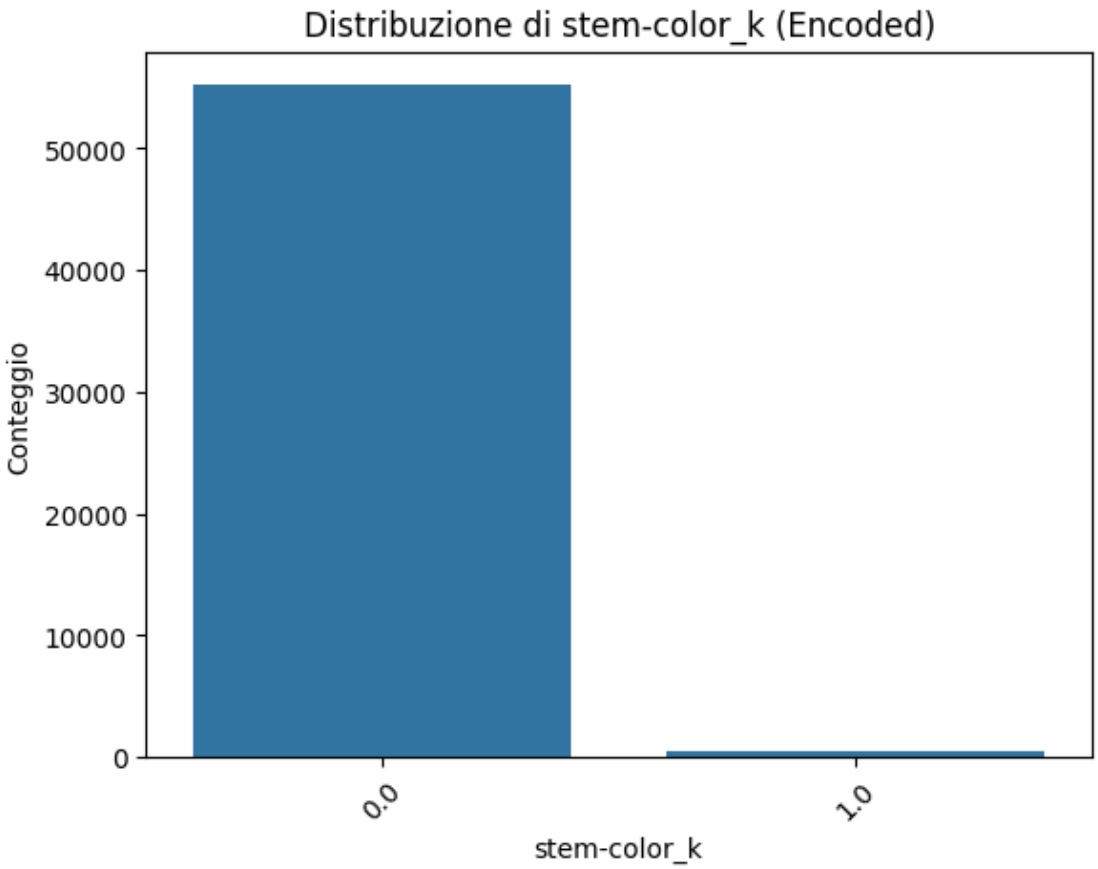

Distribuzione di stem-surface_t (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-surface_t

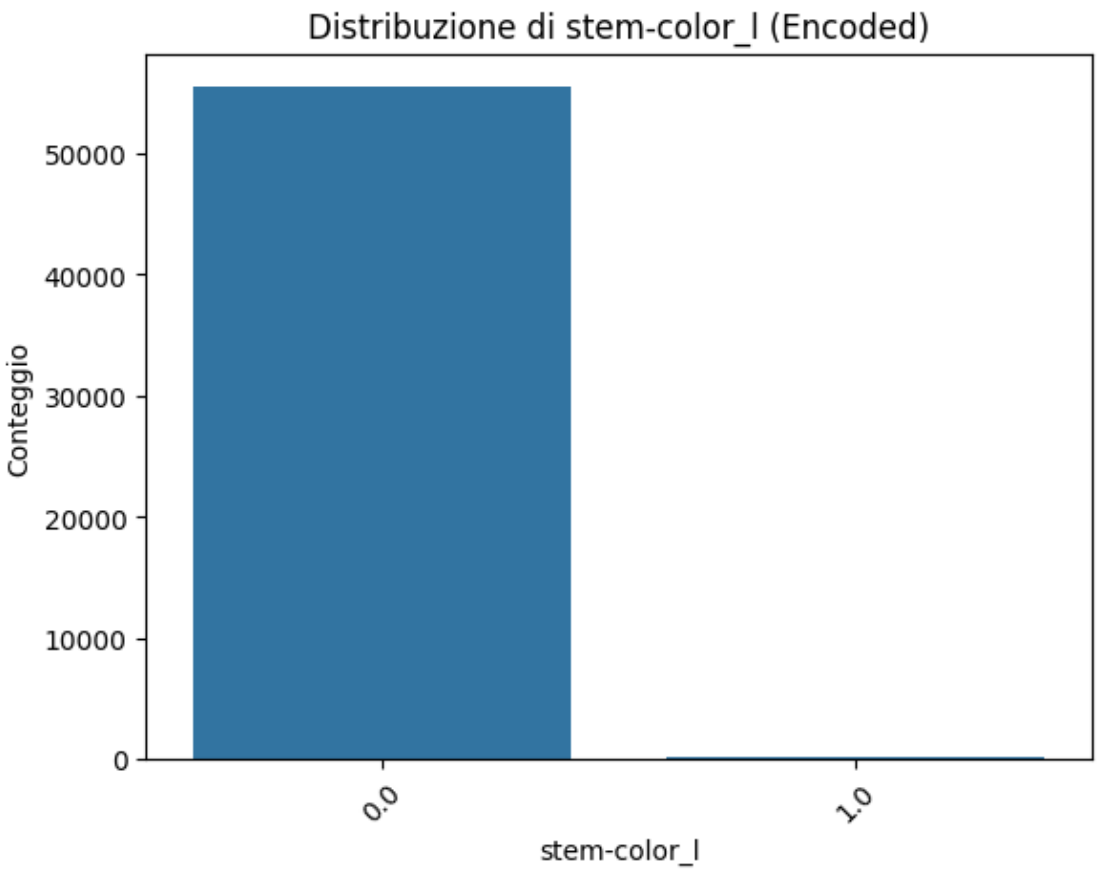

Distribuzione di stem-color_None (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_None

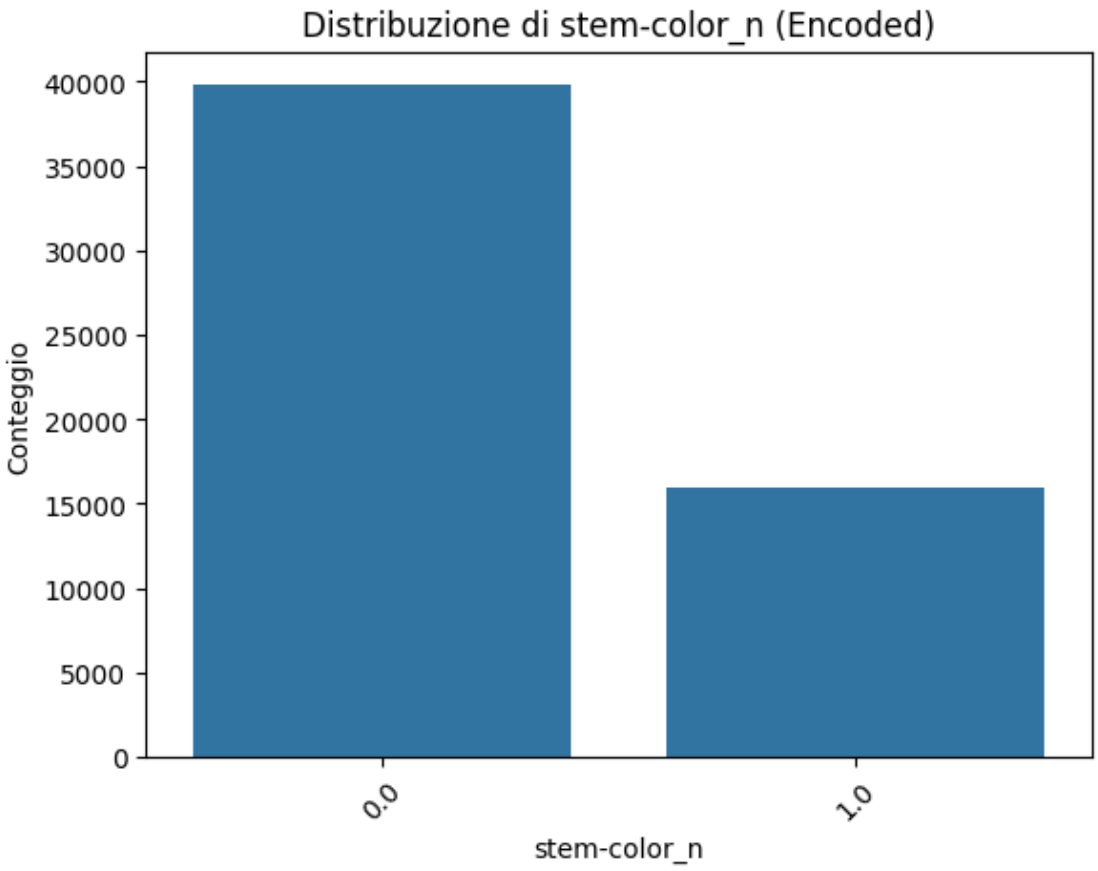

Distribuzione di stem-color_b (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_b

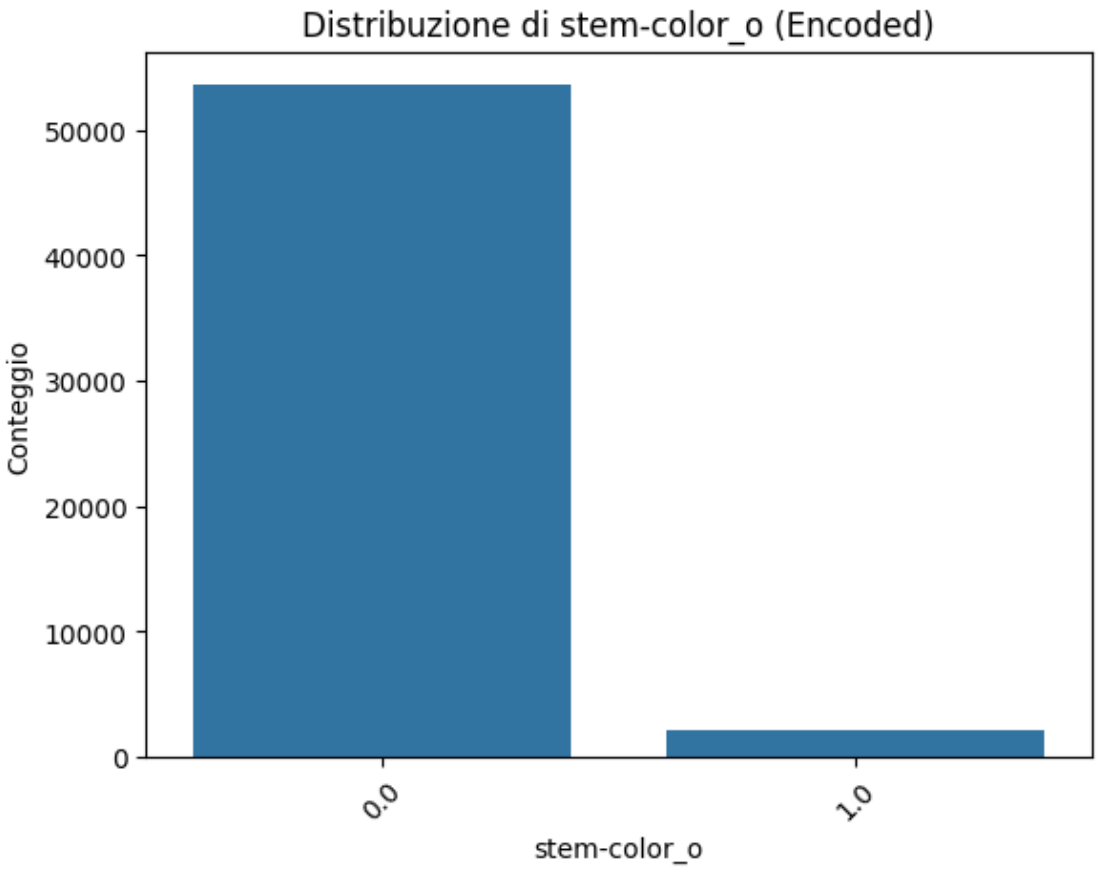

Distribuzione di stem-color_e (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_e

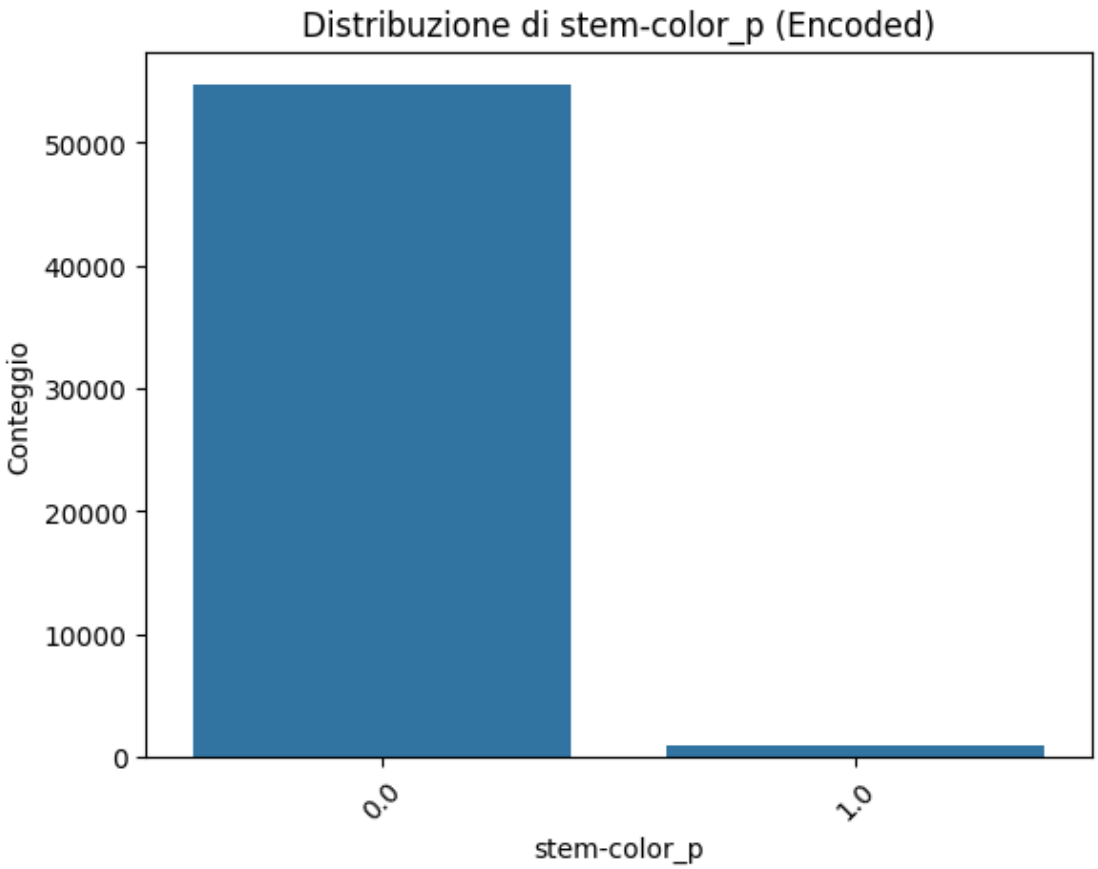

Distribuzione di stem-color_g (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_g

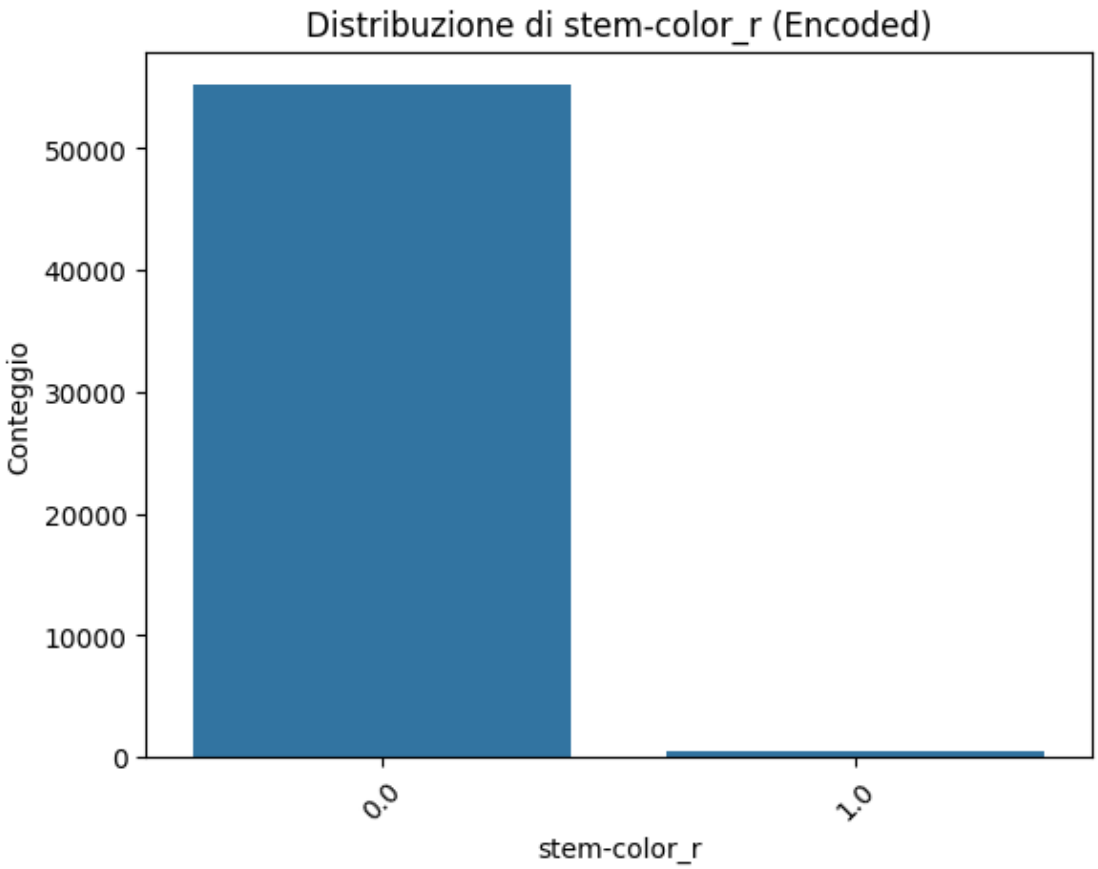

Distribuzione di stem-color_k (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_k

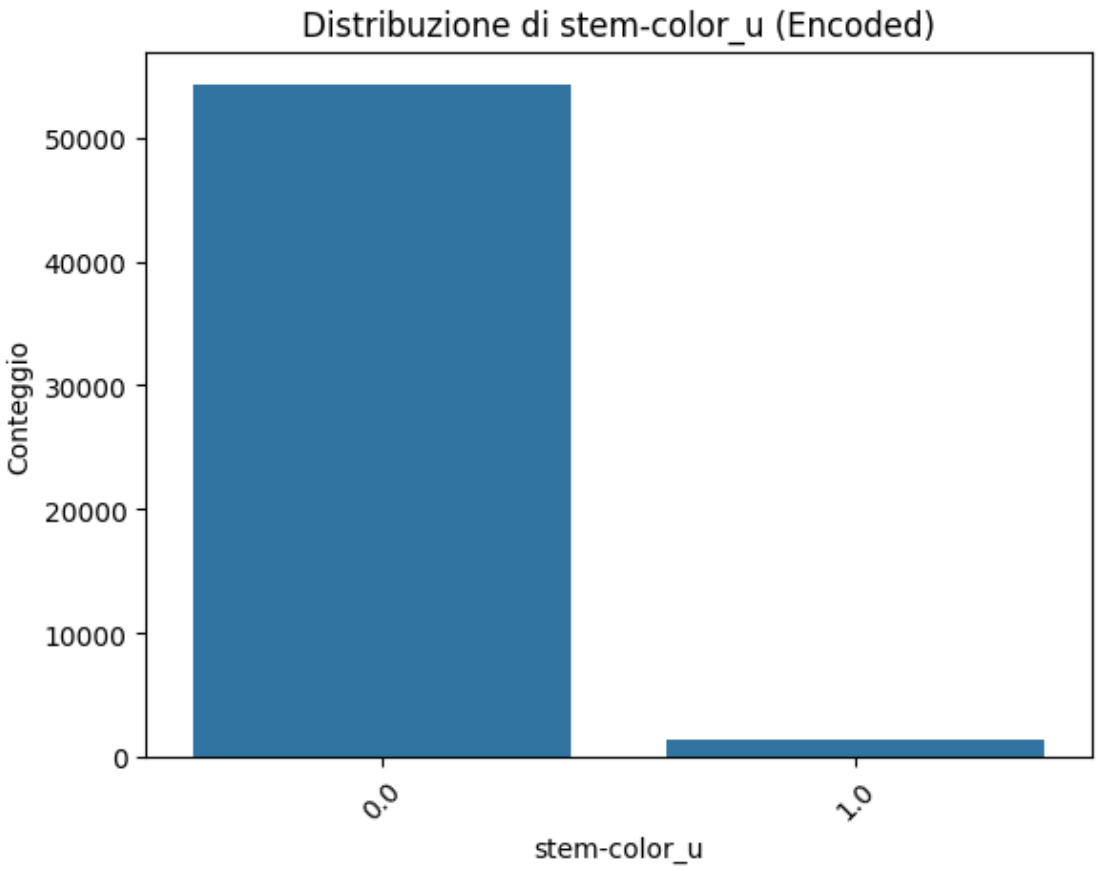

Distribuzione di stem-color_l (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_l

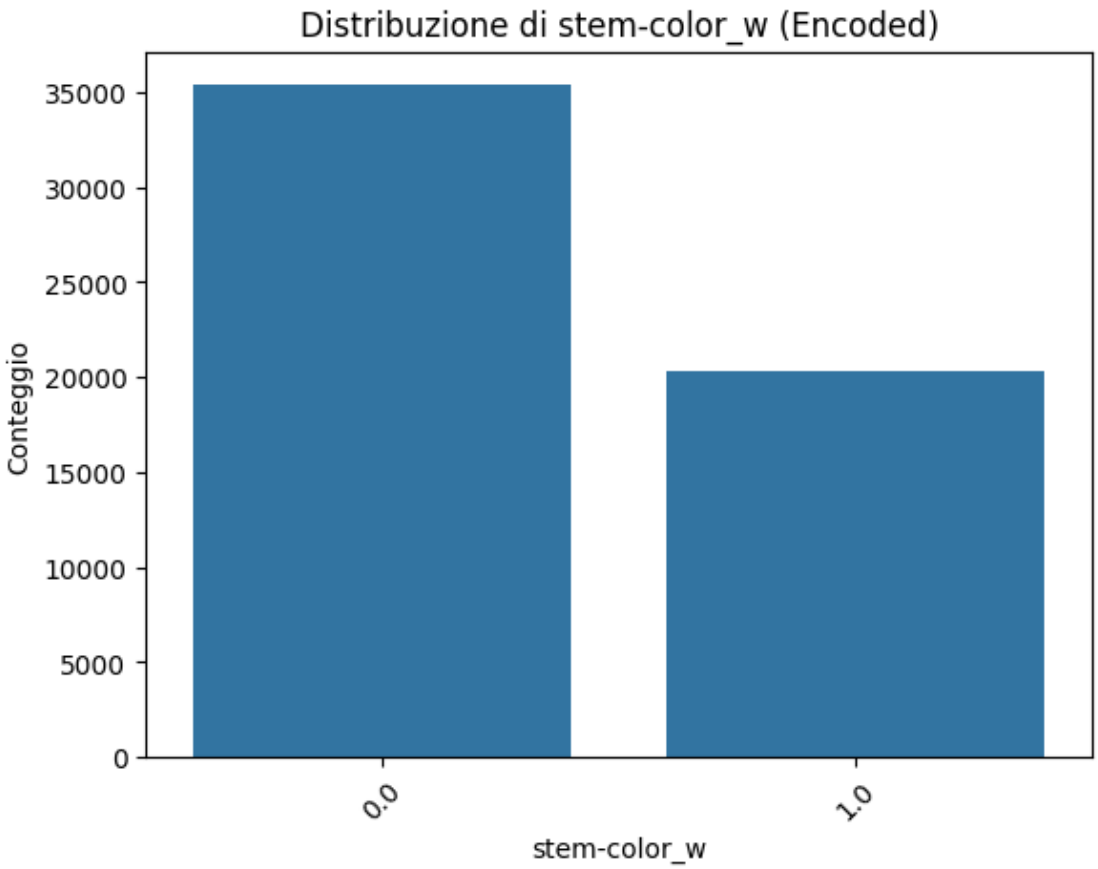

Distribuzione di stem-color_n (Encoded)
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
stem-color_n

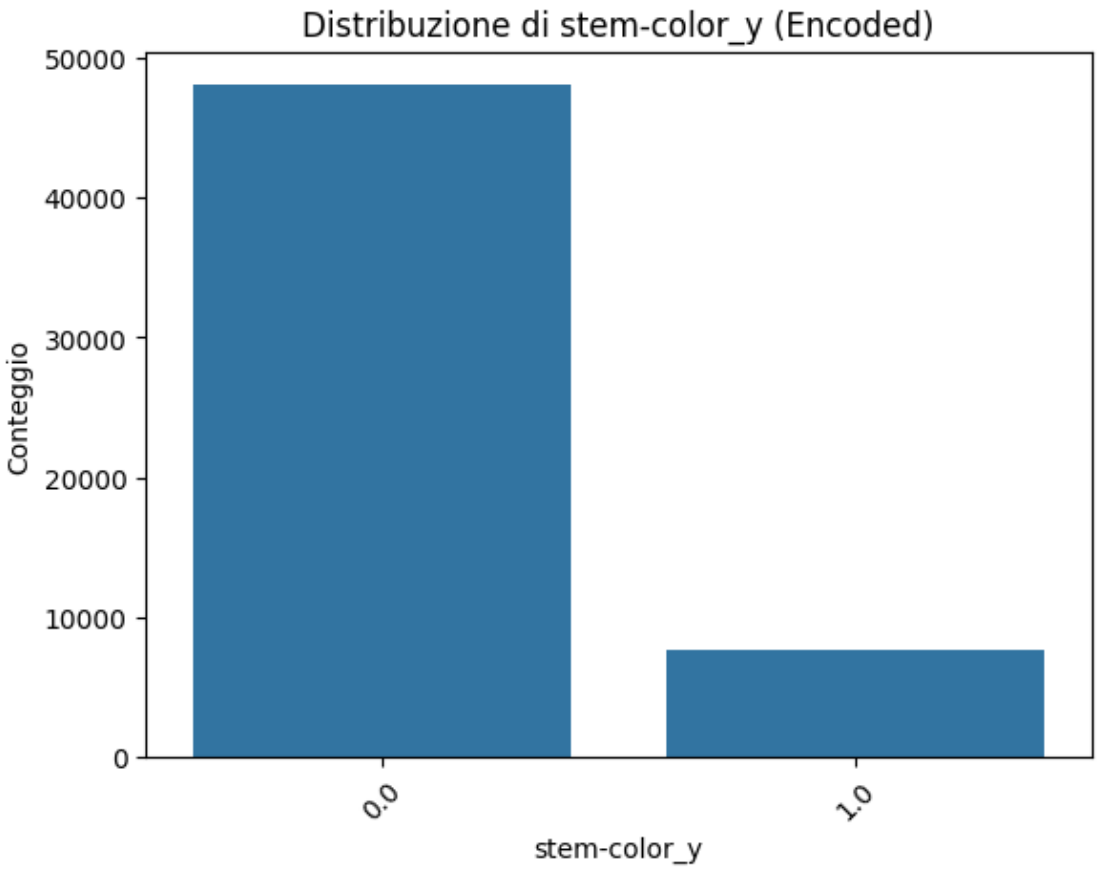

Distribuzione di stem-color_o (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_o

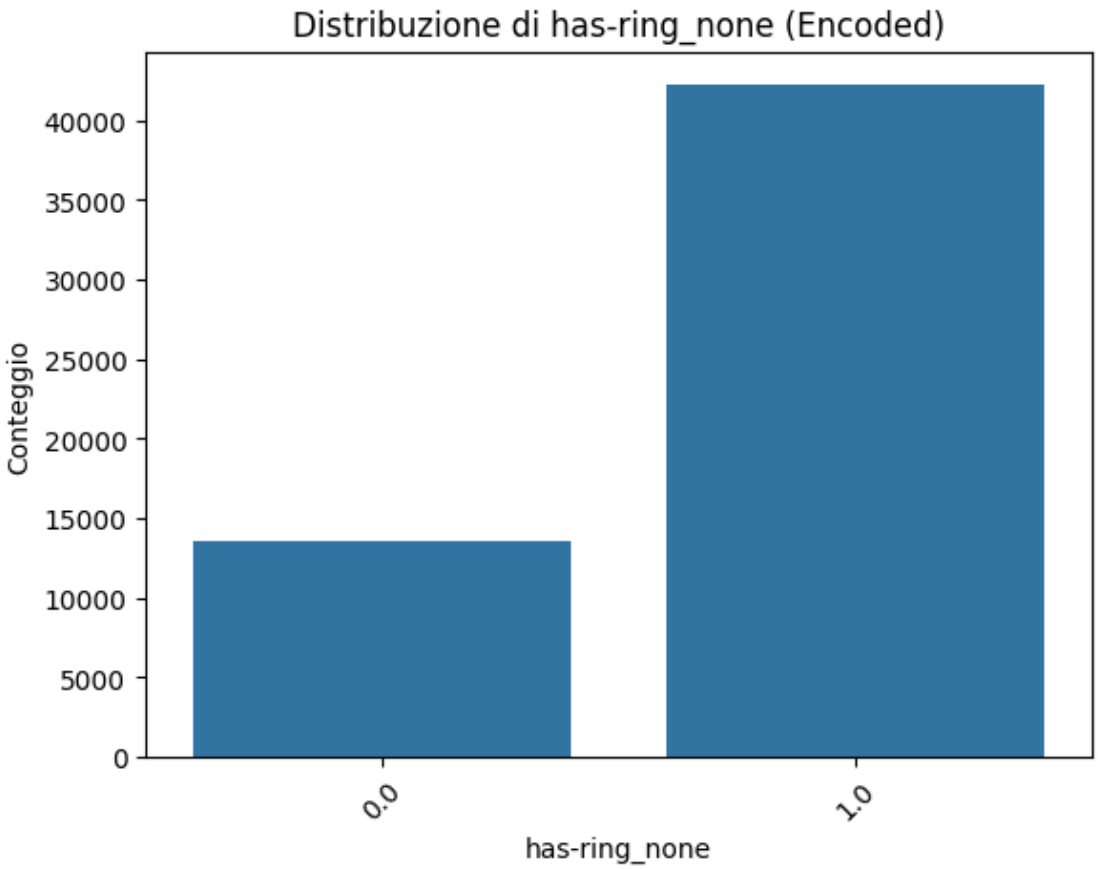

Distribuzione di stem-color_p (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_p

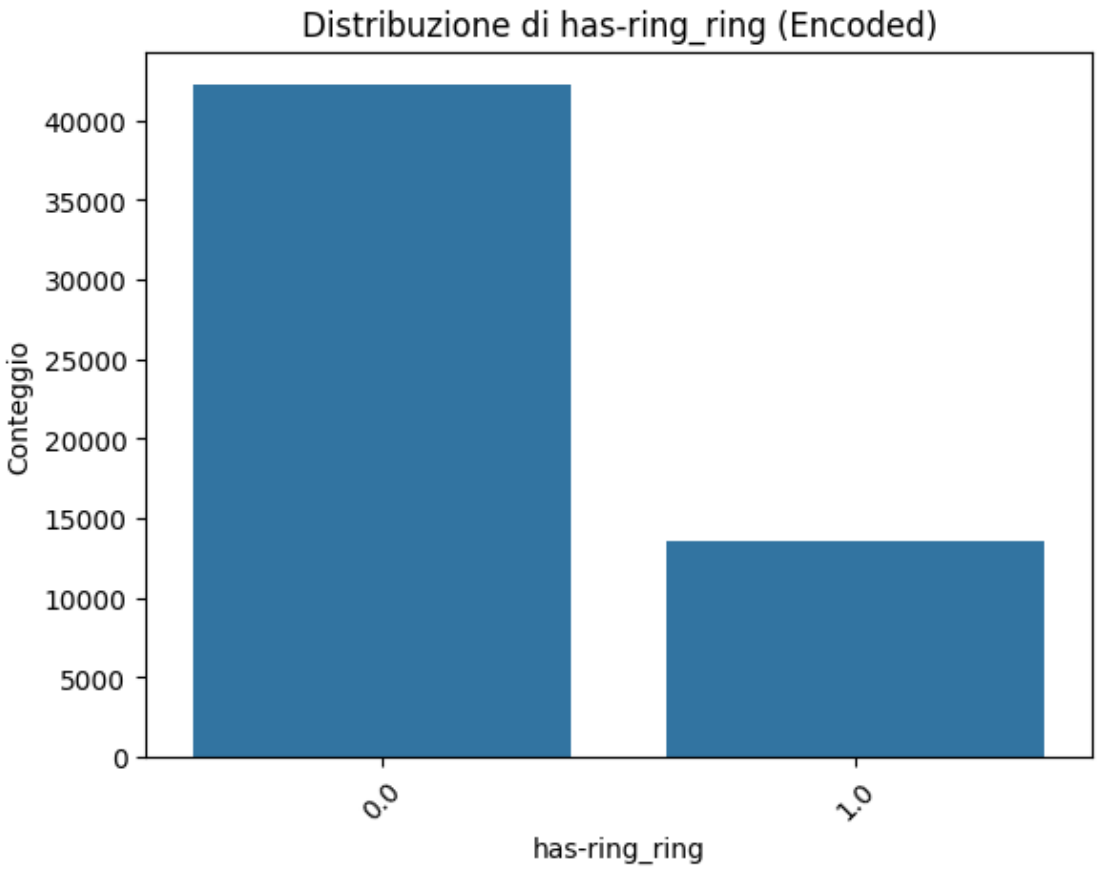

Distribuzione di stem-color_r (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_r

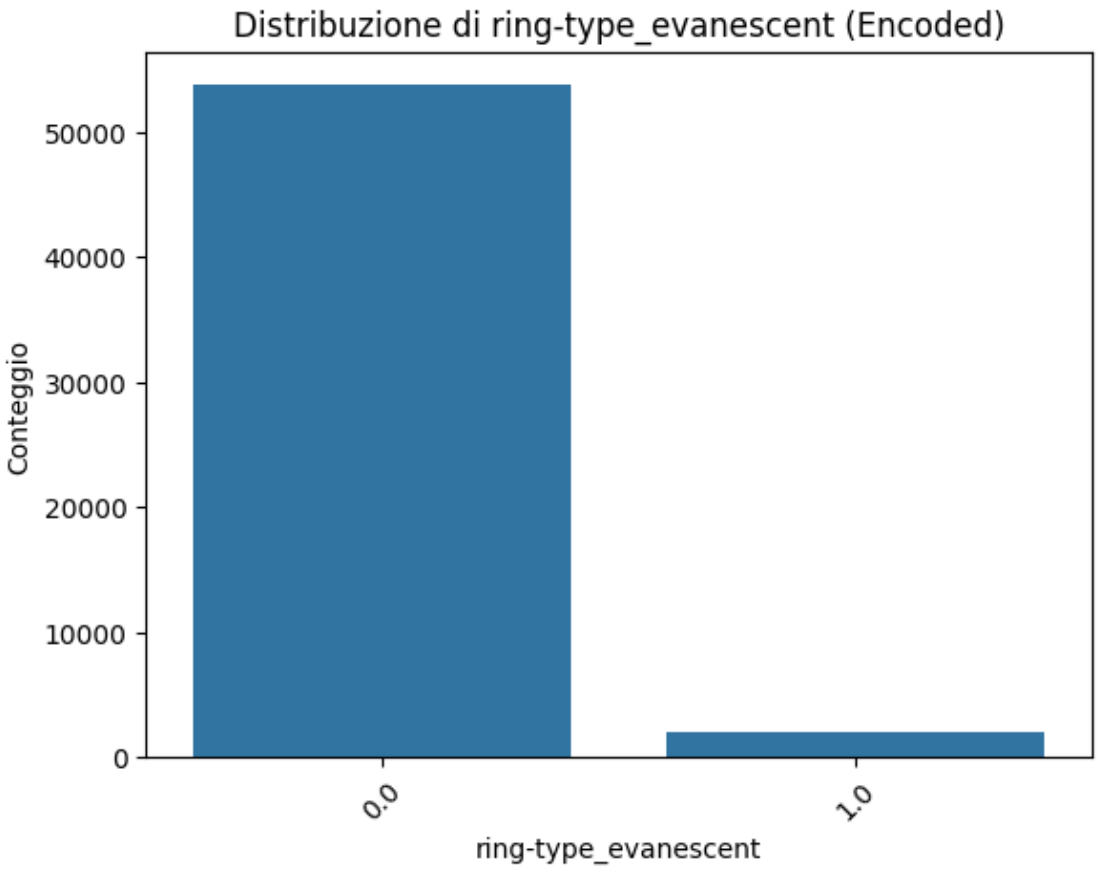

Distribuzione di stem-color_u (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_u

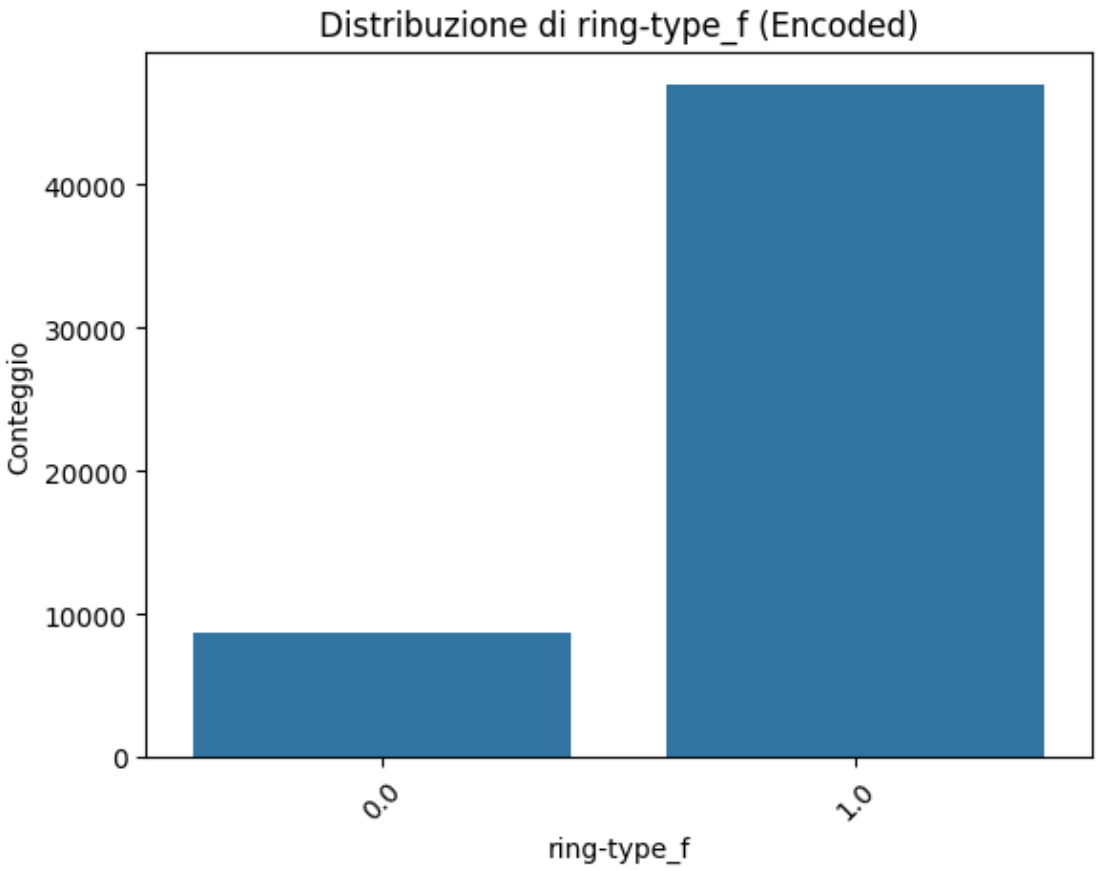

Distribuzione di stem-color_w (Encoded)
Conteggio
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
stem-color_w

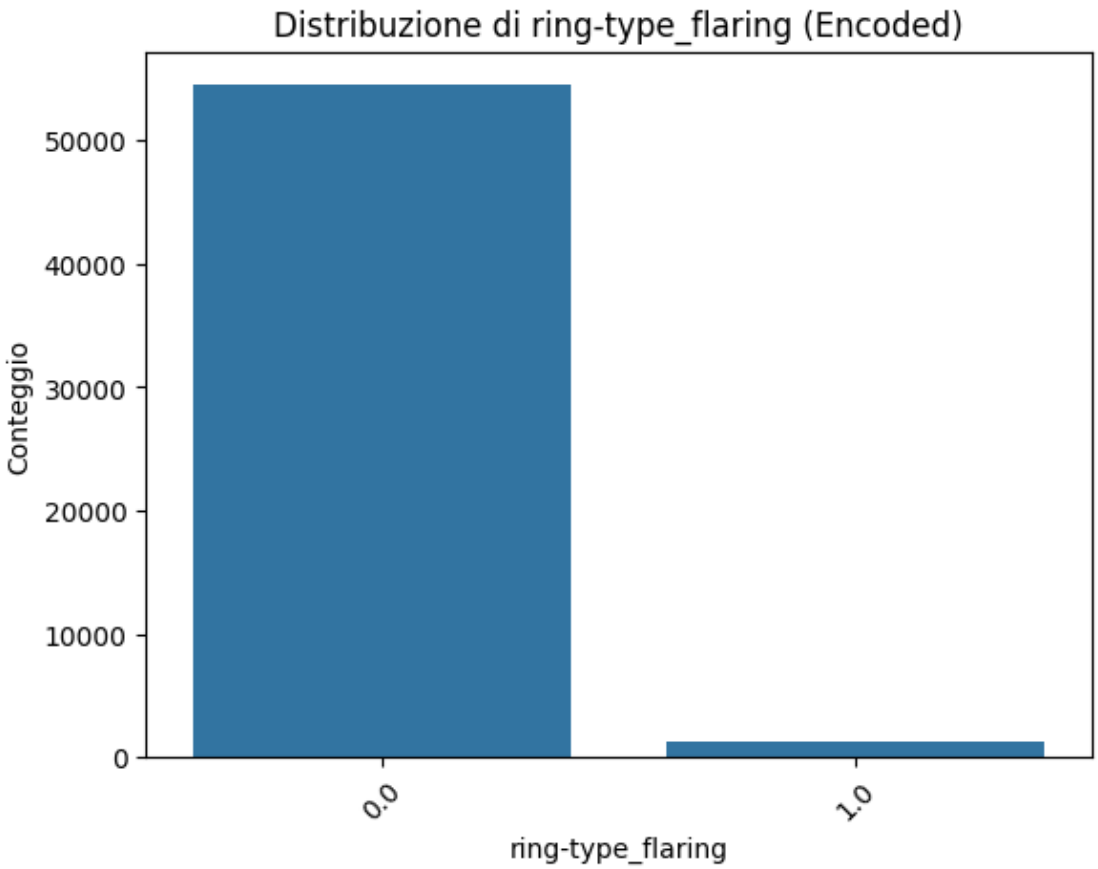

Distribuzione di stem-color_y (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
stem-color_y

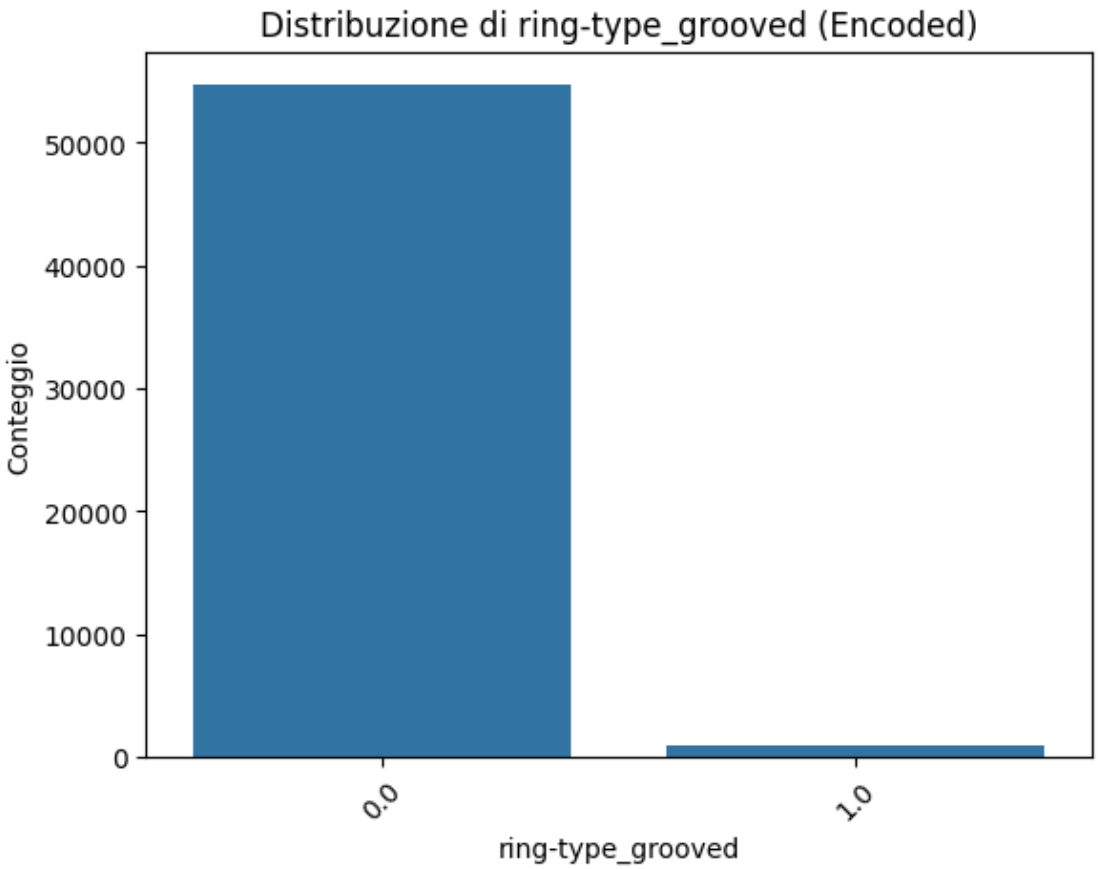

Distribuzione di has-ring_none (Encoded)
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
has-ring_none

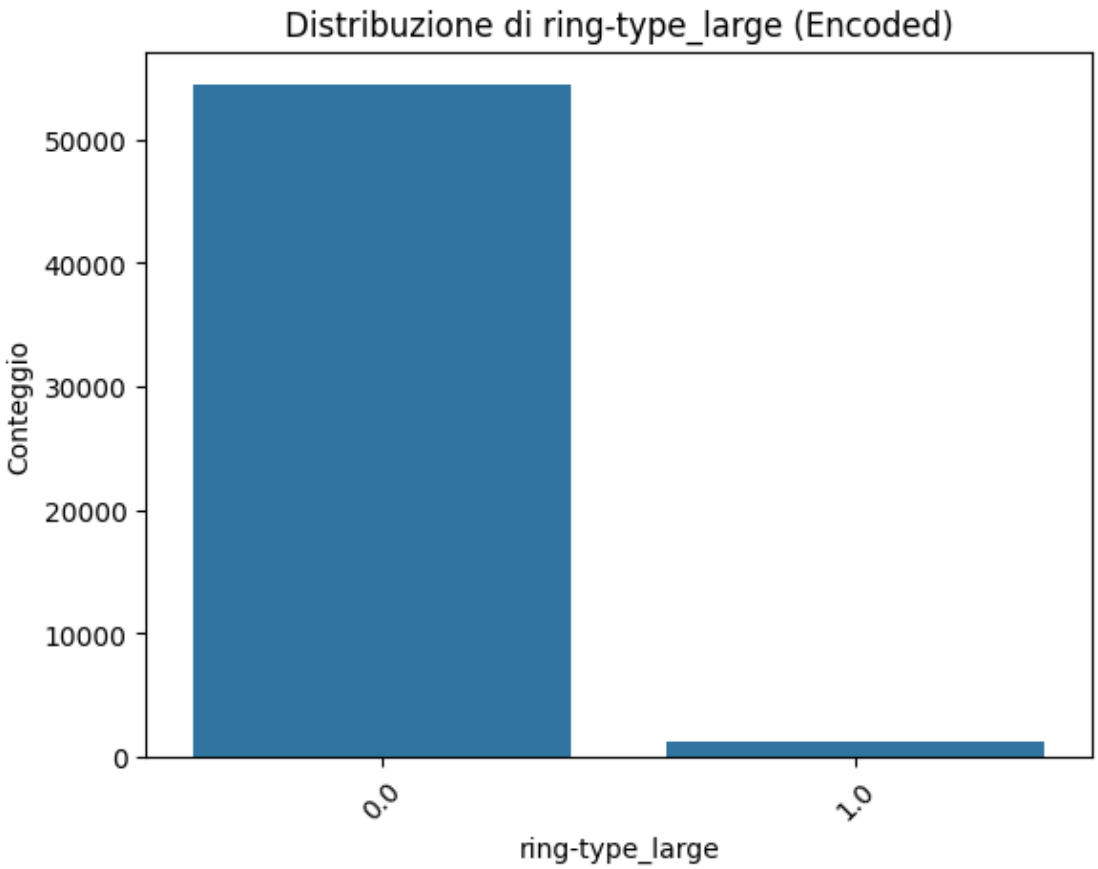

Distribuzione di has-ring_ring (Encoded)
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
has-ring_ring

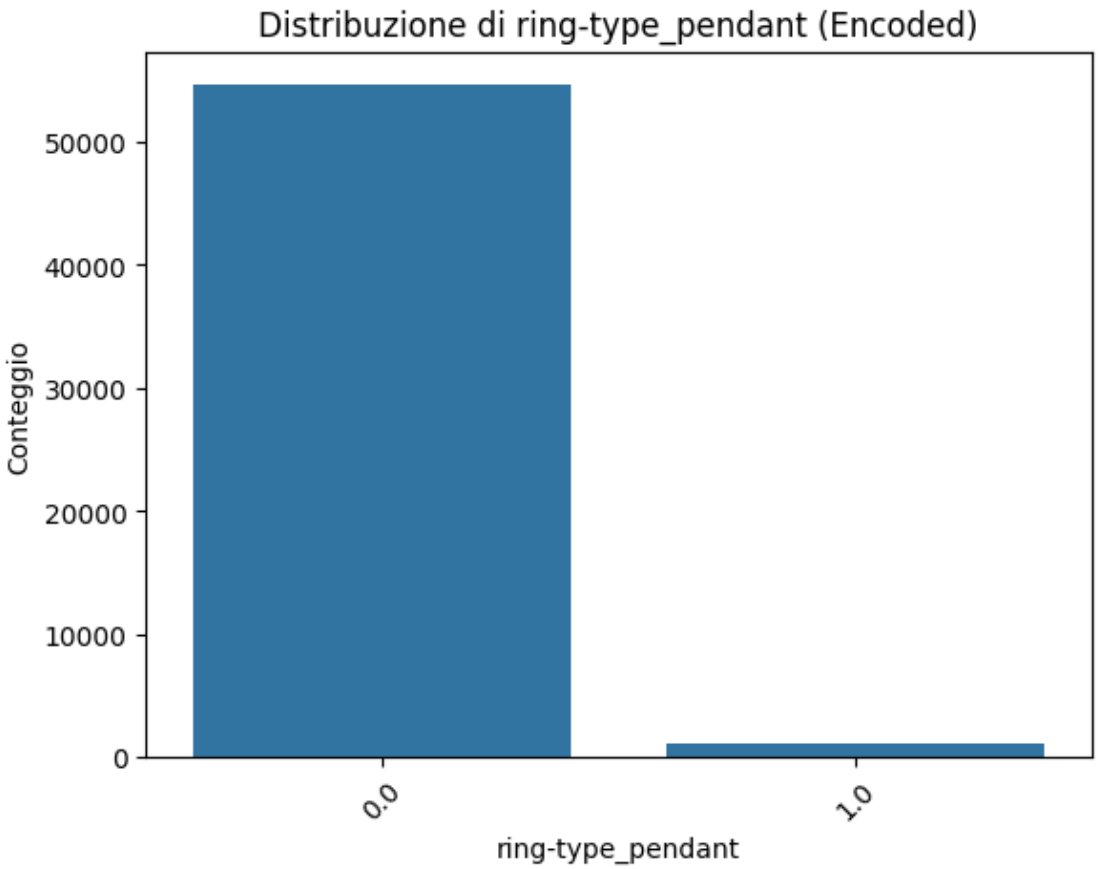

Distribuzione di ring-type_evanescent (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
ring-type_evanescent

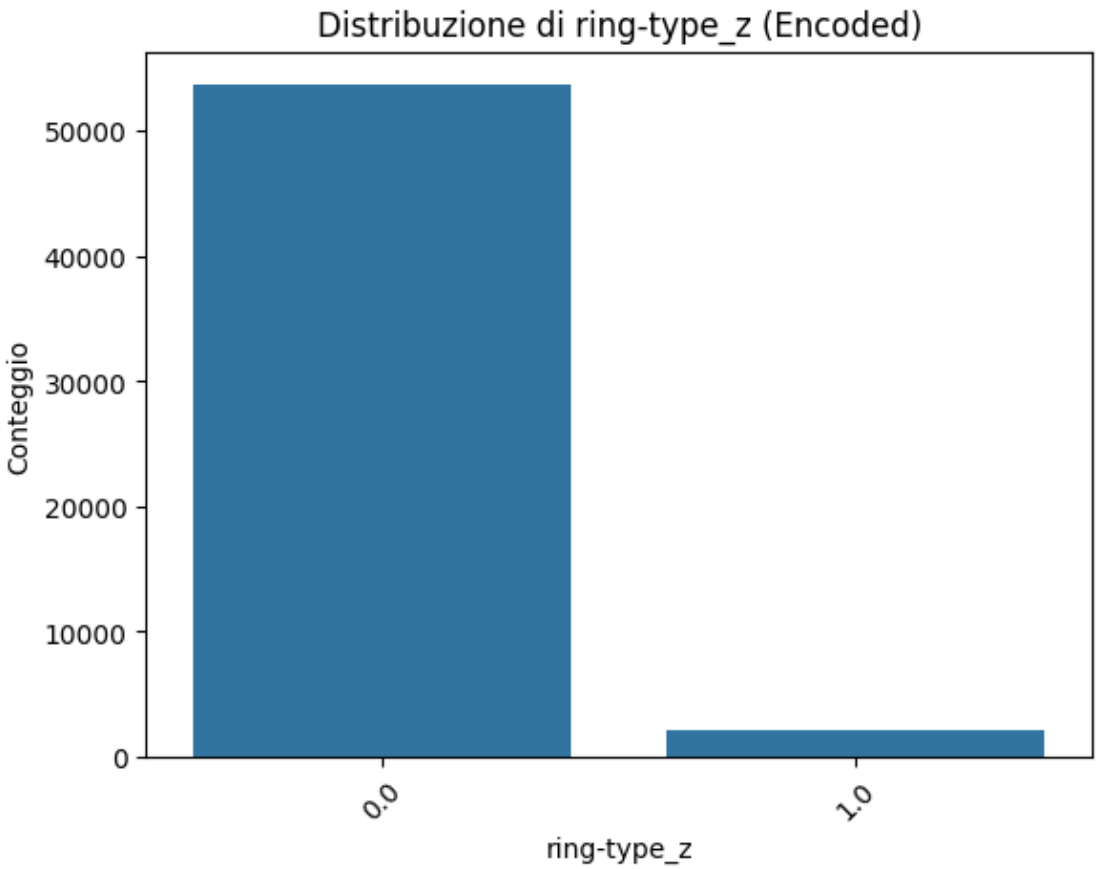

Distribuzione di ring-type_f (Encoded)
Conteggio
40000
30000
20000
10000
0
0.0
1.0
ring-type_f

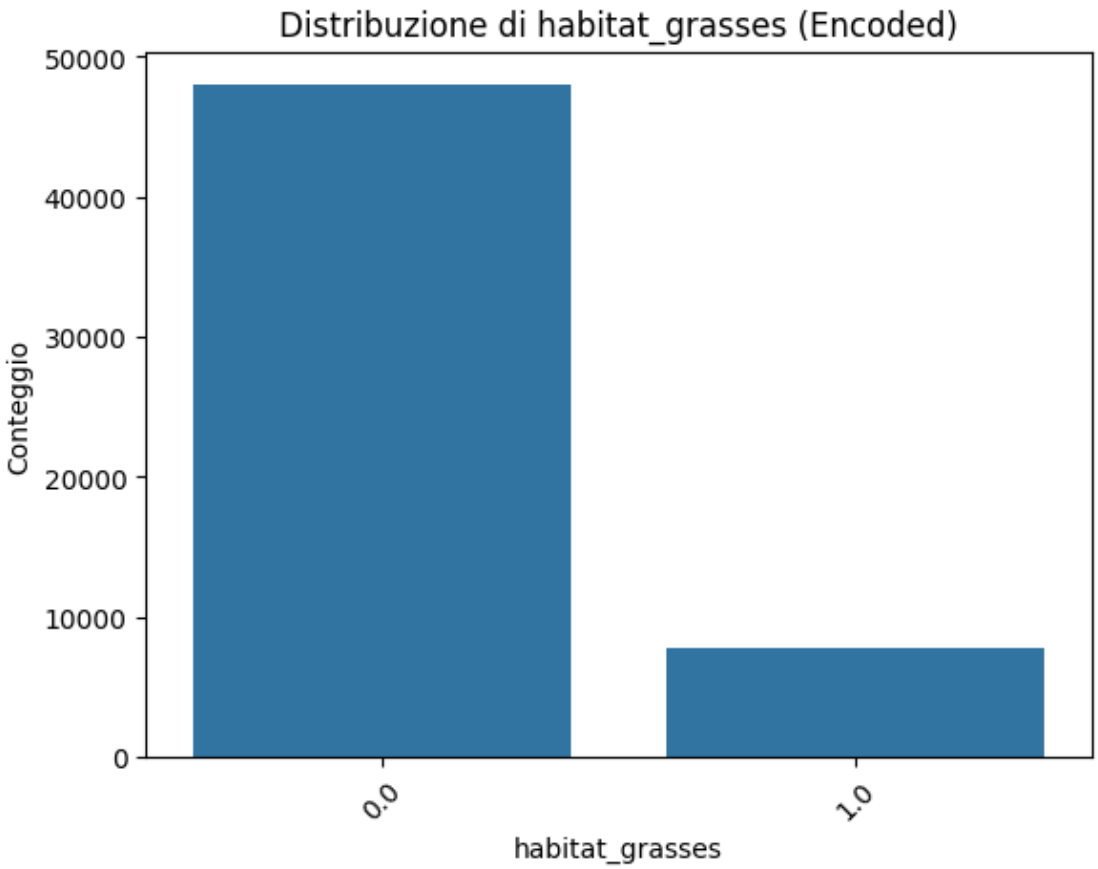

Distribuzione di ring-type_flaring (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
ring-type_flaring

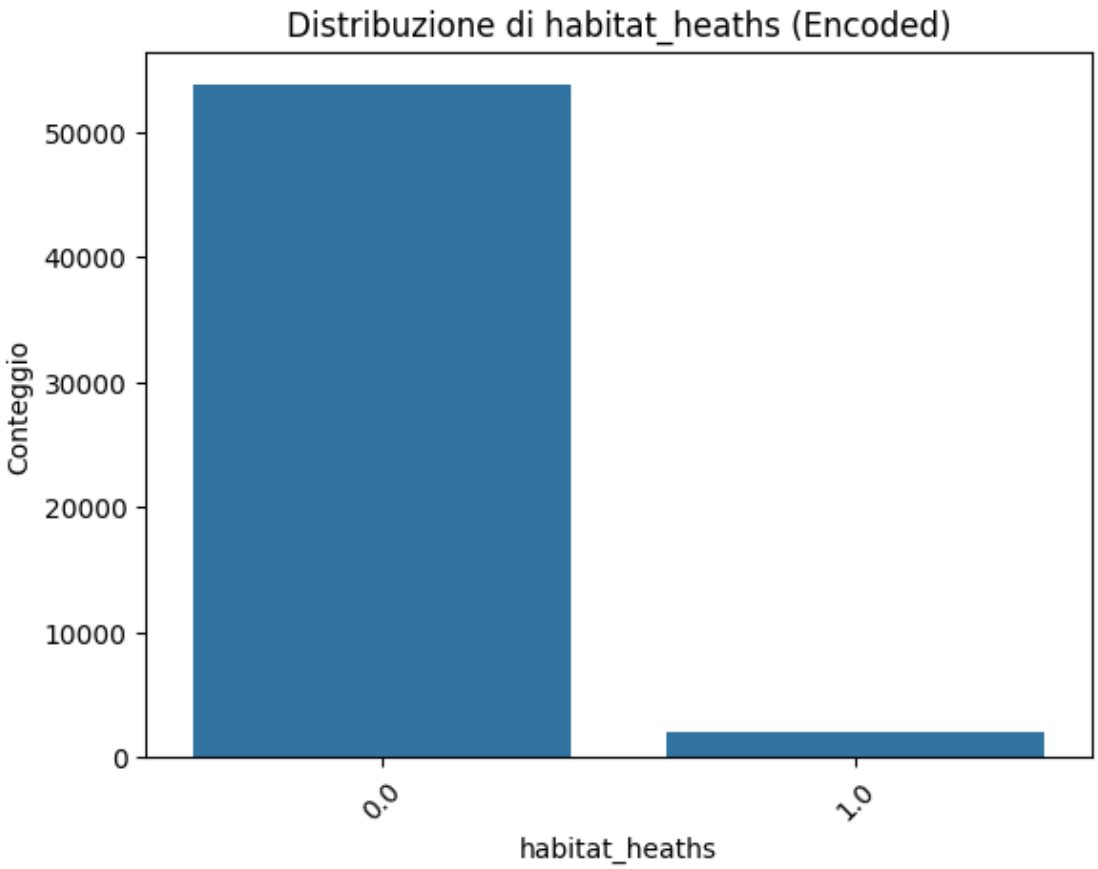

Distribuzione di ring-type_grooved (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
ring-type_grooved

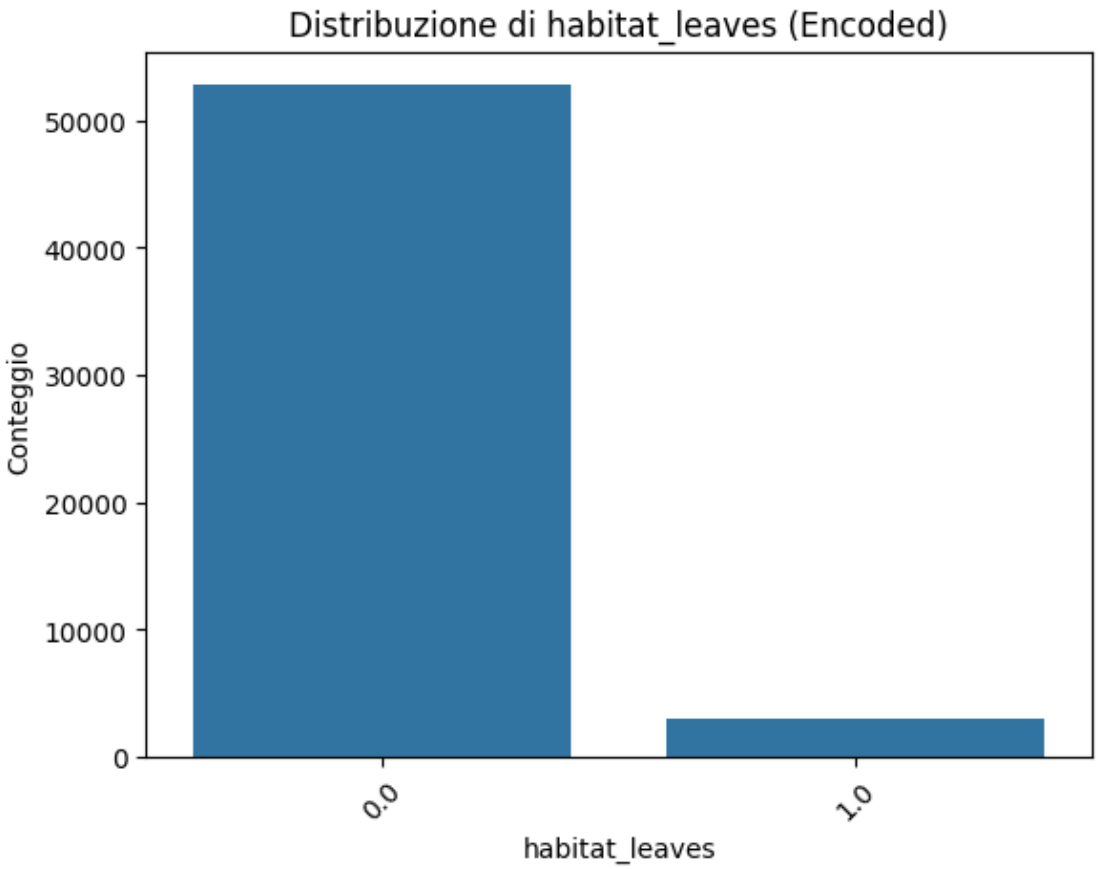

Distribuzione di ring-type_large (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
ring-type_large

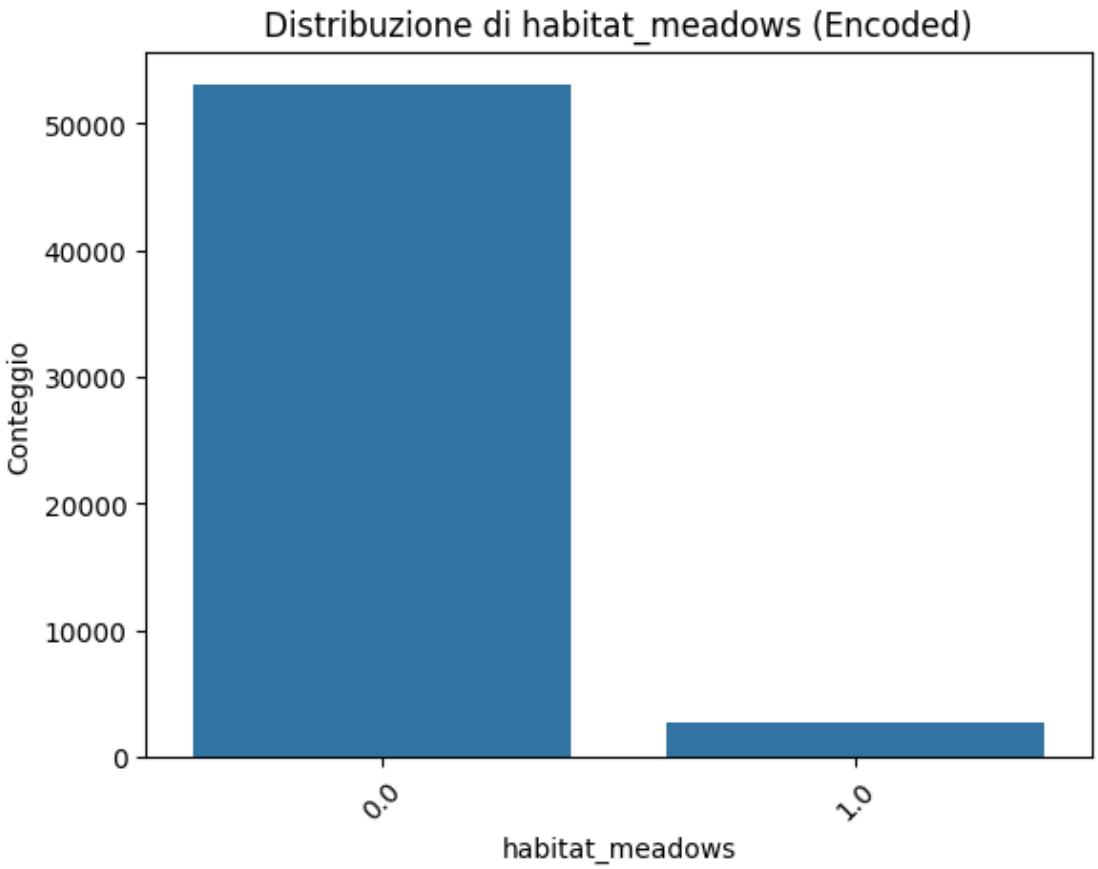

Distribuzione di ring-type_pendant (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
ring-type_pendant

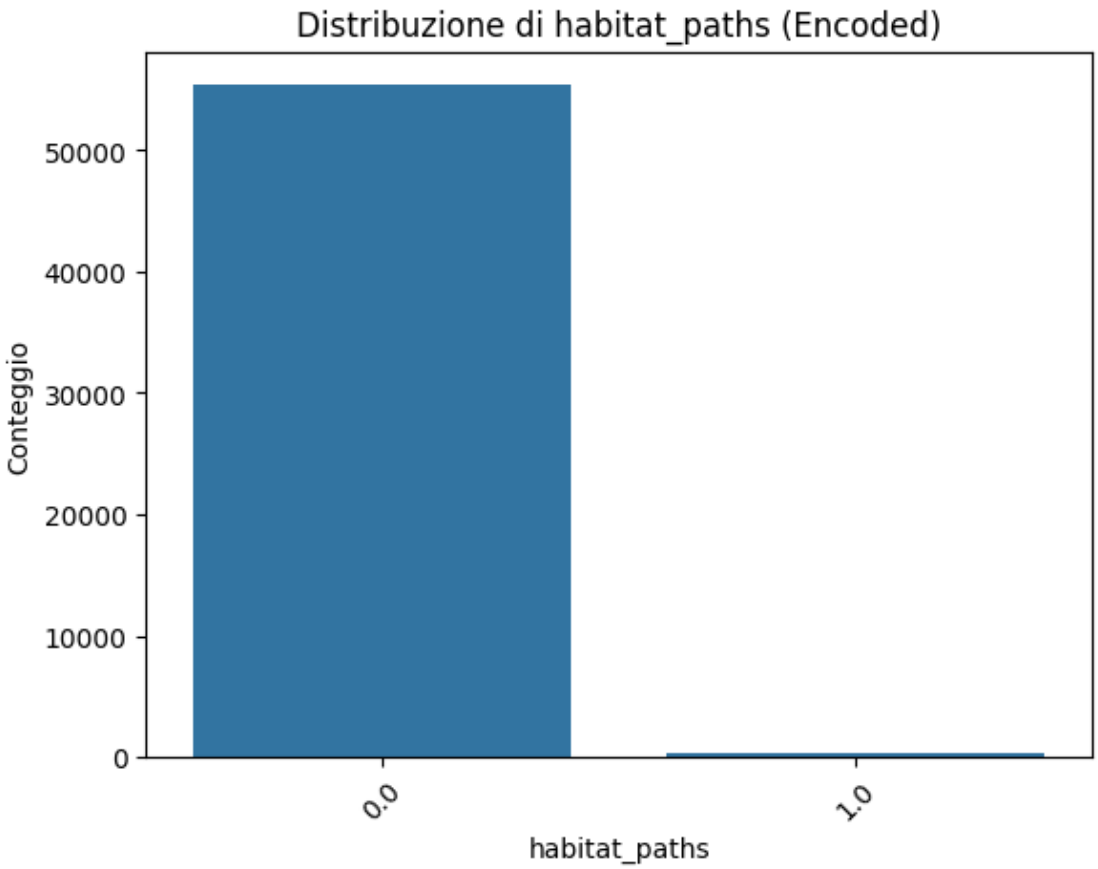

Distribuzione di ring-type_z (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
ring-type_z

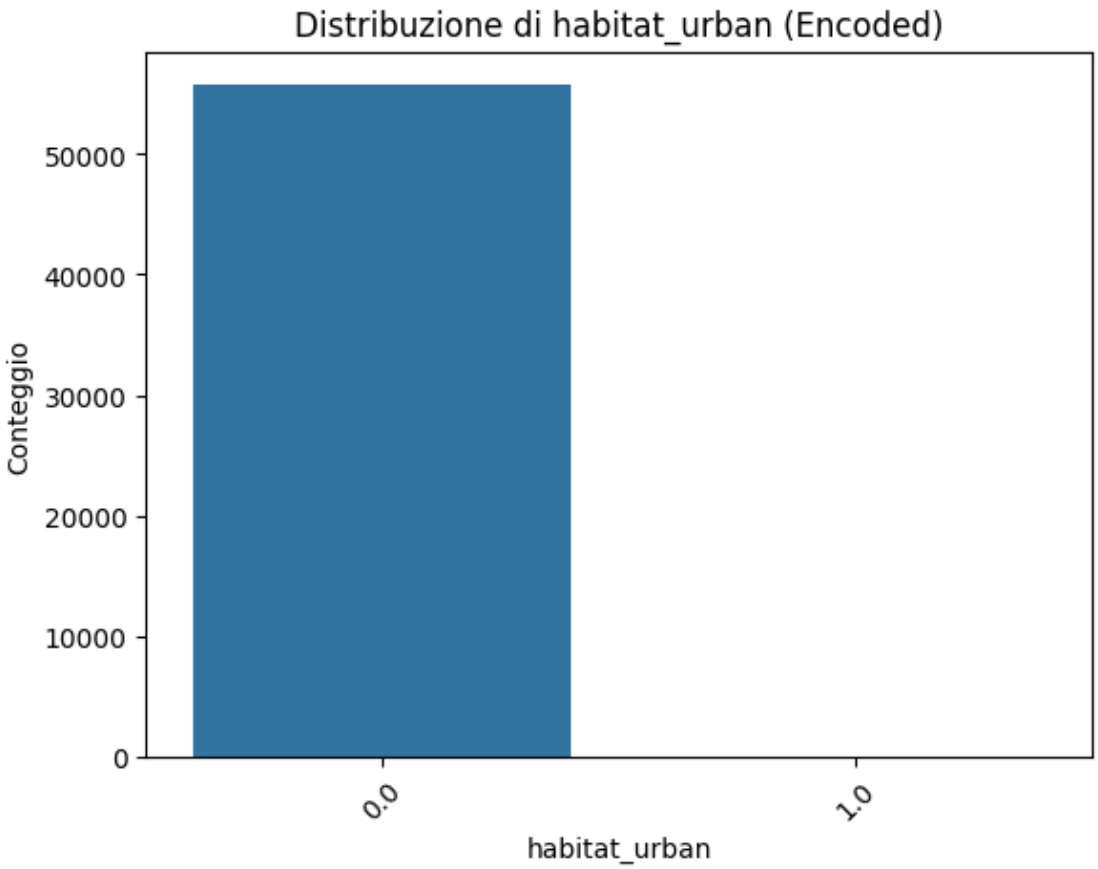

Distribuzione di habitat_grasses (Encoded)
50000
40000
30000
20000
10000
0
Conteggio
0.0
1.0
habitat_grasses

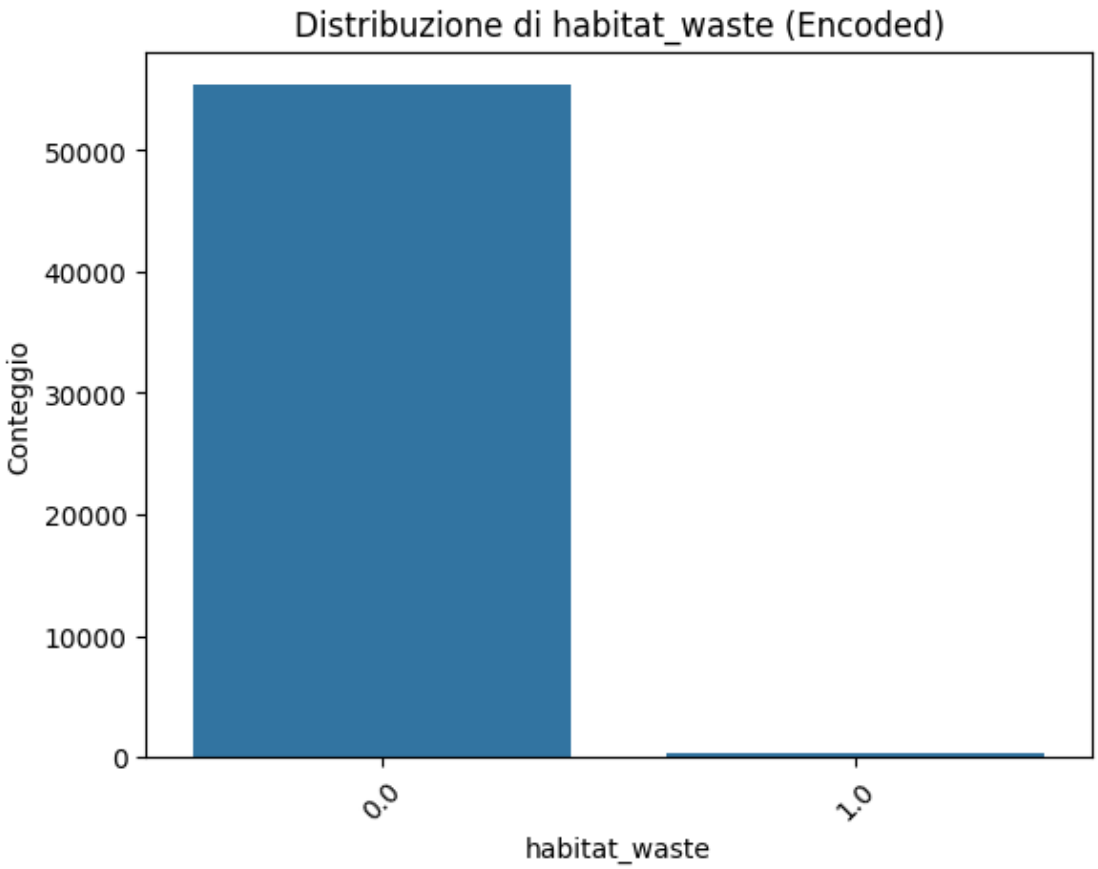

Distribuzione di habitat_heaths (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
habitat_heaths

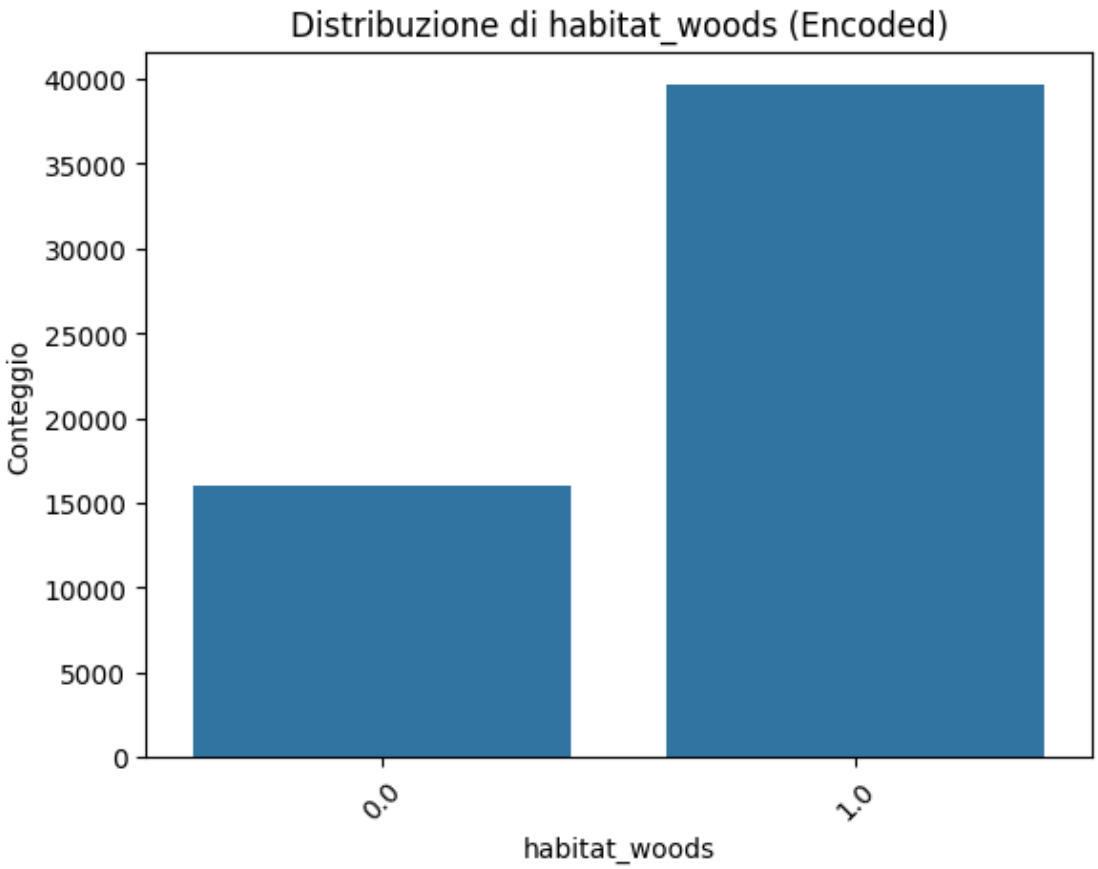

Distribuzione di habitat_leaves (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
habitat_leaves

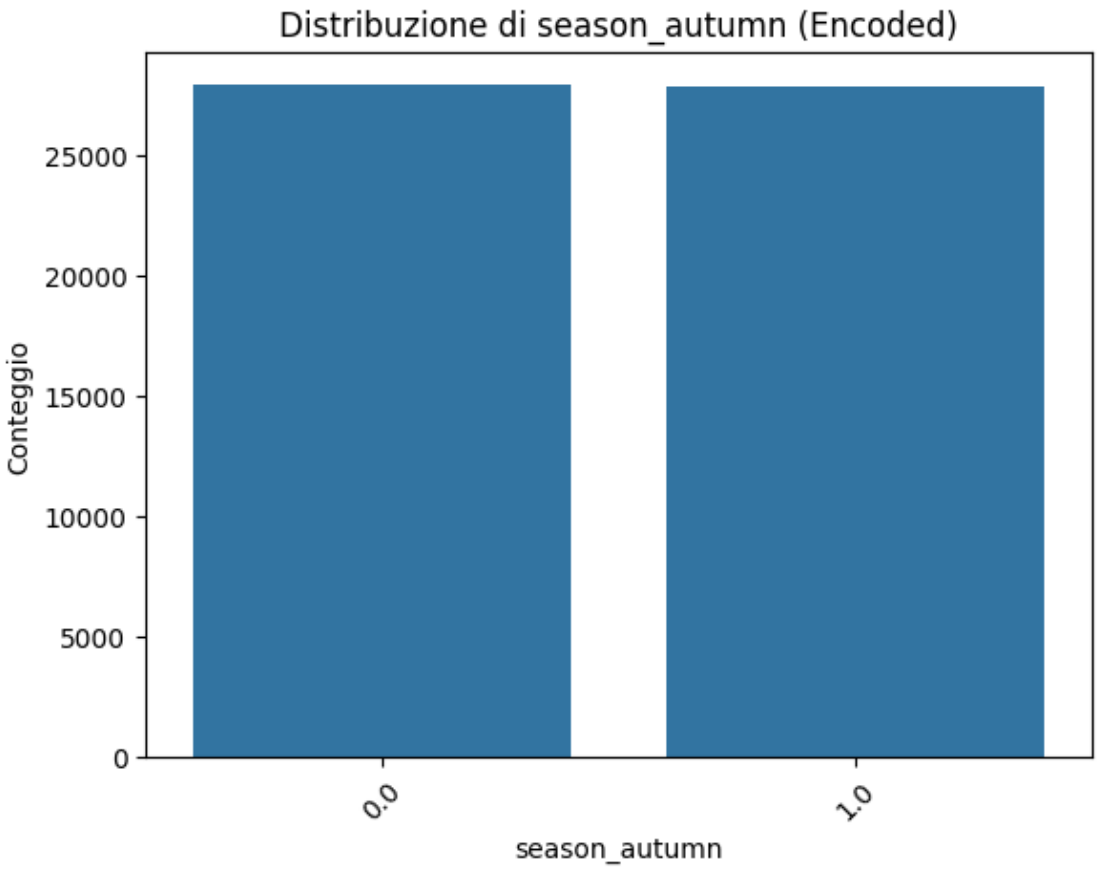

Distribuzione di habitat_meadows (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
habitat_meadows

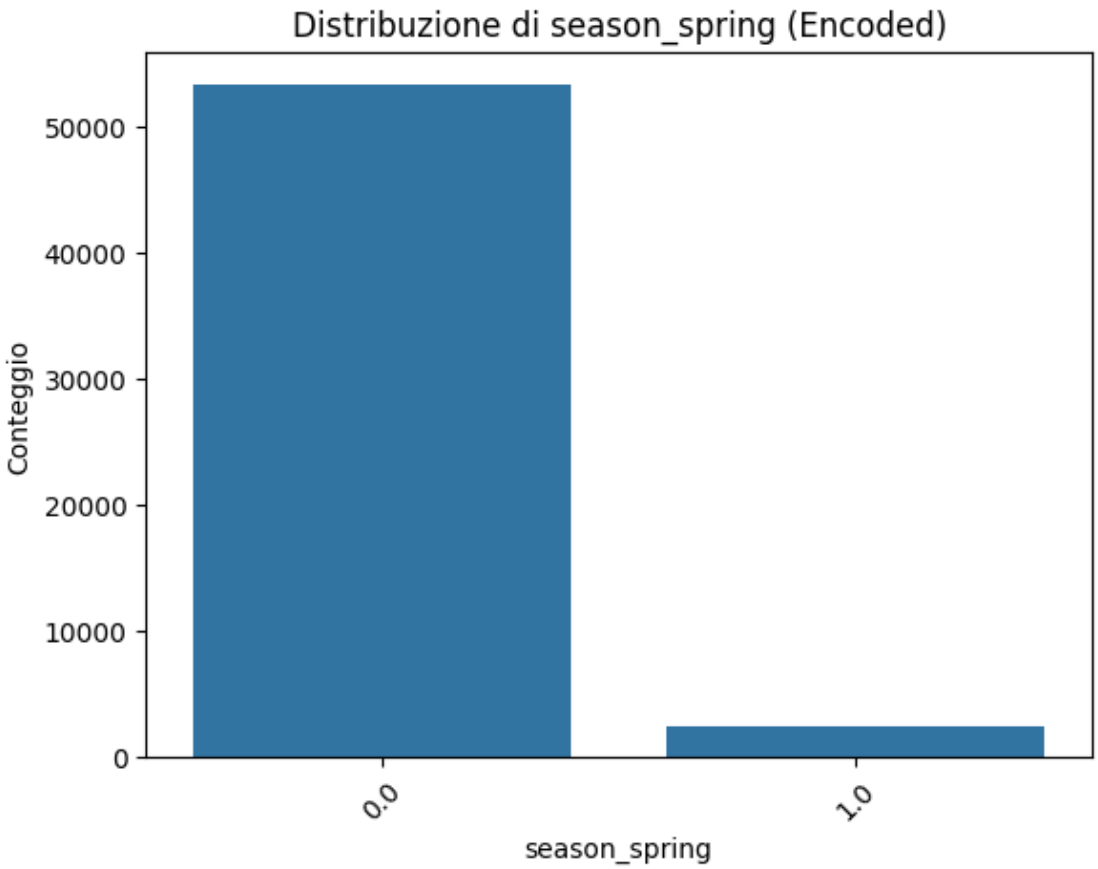

Distribuzione di habitat_paths (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
habitat_paths

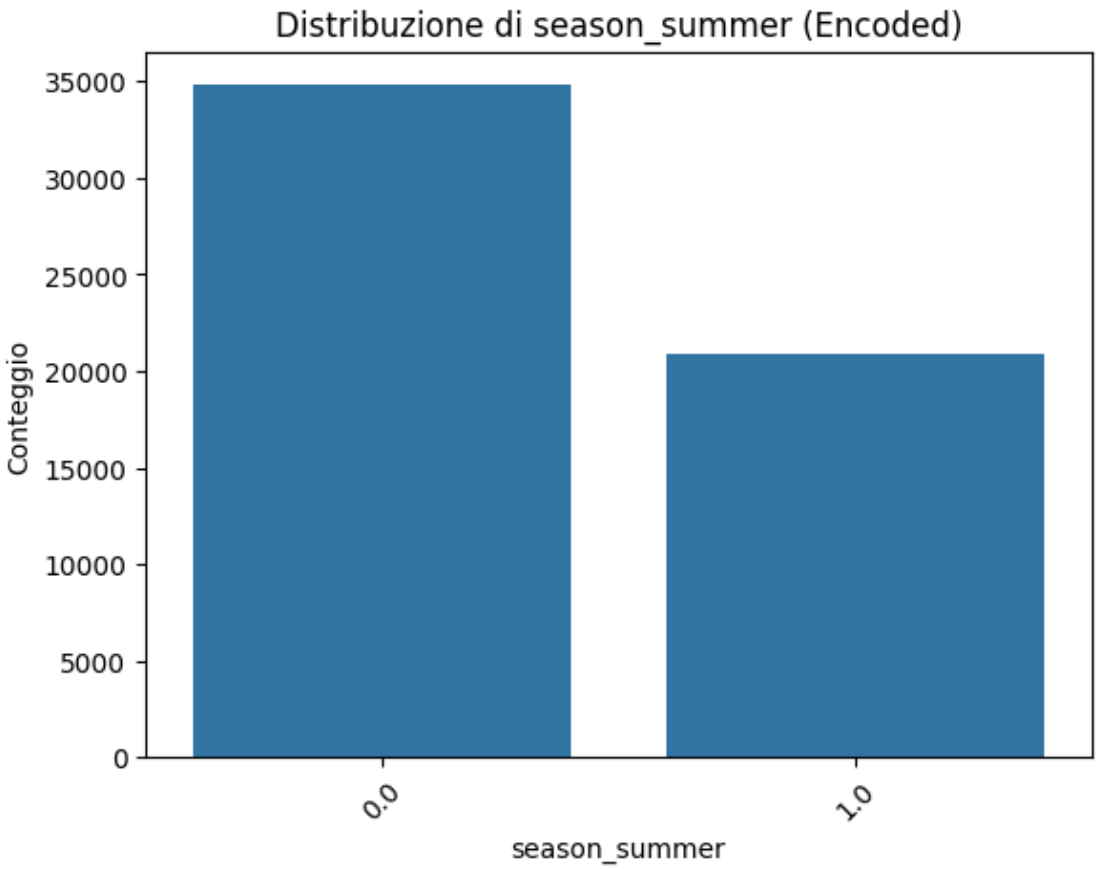

Distribuzione di habitat_urban (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
habitat_urban

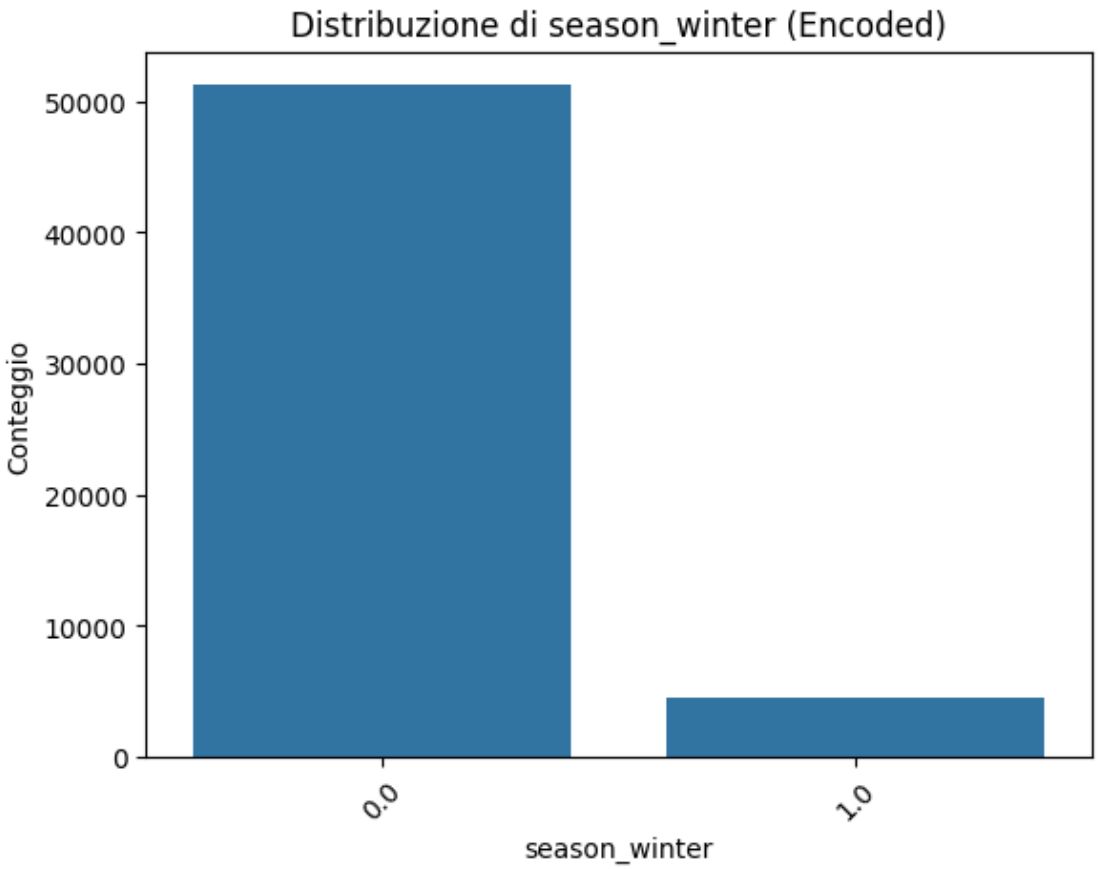

Distribuzione di habitat_waste (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
habitat_waste


Distribuzione di habitat_woods (Encoded)
Conteggio
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
habitat_woods


Distribuzione di season_autumn (Encoded)
Conteggio
25000
20000
15000
10000
5000
0
0.0
1.0
season_autumn


Distribuzione di season_spring (Encoded)
Conteggio
50000
40000
30000
20000
10000
0
0.0
1.0
season_spring


Distribuzione di season_summer (Encoded)
Conteggio
35000
30000
25000
20000
15000
10000
5000
0
0.0
1.0
season_summer



[ ]:
```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Definisci la variabile target
target_variable = df2_cleaned['class']

# Rimuovi la variabile target dal DataFrame codificato
X = df2_encoded
y = target_variable

# Suddivisione dei dati in training e testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

# Inizializza il classificatore Random Forest
rf_classifier = RandomForestClassifier(random_state=42)

# Addestra il modello sul set di dati di addestramento
```

```
rf_classifier.fit(X_train, y_train)

# Effettua le predizioni sul set di dati di test
y_pred = rf_classifier.predict(X_test)

# Valuta le prestazioni del modello
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred)

# Visualizza l'accuratezza e il report di classificazione
print("Accuratezza del modello:", accuracy)
print("\nReport di classificazione:\n", classification_rep)
```

```
Accuratezza del modello: 1.0

Report di classificazione:
               precision    recall  f1-score   support

      edible       1.00      1.00      1.00      4784
    poisonus       1.00      1.00      1.00      6362

    accuracy                           1.00     11146
   macro avg       1.00      1.00      1.00     11146
weighted avg       1.00      1.00      1.00     11146
```

3 e adesso... morirà presto?

[ ]:
```
# Prendi una riga casuale dal dataset di test
sample_row = X_test.sample(n=1, random_state=42)

# Utilizza il modello addestrato per prevedere la variabile target per il dato␣
↪di input
predicted_class = rf_classifier.predict(sample_row)

# Visualizza la riga e la previsione del modello
print("Riga di input:")
print(sample_row)
print("\nPrevisto:", predicted_class)
```

```
Riga di input:
       class_edible  class_poisonus  cap-shape_bell  cap-shape_conical  \
19176           1.0             0.0             0.0                0.0

       cap-shape_convex  cap-shape_flat  cap-shape_others  \
19176               1.0             0.0               0.0
```

```
       cap-shape_spherical  cap-shape_sunken  cap-surface_d  …  \
19176                  0.0               0.0            0.0  …

       habitat_leaves  habitat_meadows  habitat_paths  habitat_urban  \
19176             0.0              1.0            0.0            0.0

       habitat_waste  habitat_woods  season_autumn  season_spring  \
19176            0.0            0.0            1.0            0.0

       season_summer  season_winter
19176            0.0            0.0

[1 rows x 98 columns]

Previsto: ['edible']
```

4 apparentemente no, il modello funziona!