Adam Russel Shane P. Oguis

EDA – new dataset for ML algorithm web integration

In [2]:
```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt



# 1. Load the Data - import the necessary libraries (e.g., Pandas)
# to read your data from a file (e.g., CSV) or a database.
# loading the Dataset
df = pd.read_csv('..//Downloads//mushrooms.csv')
```

In [3]:
```
# 2. Basic Data Exploration - Check the first few rows of the dataset
# to get an initial sense of the data's structure.

df.head()
```

Out[3]:

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	h
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	o	p	k	s	
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	o	p	n	n	
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	o	p	n	n	
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	o	p	k	s	
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	o	e	n	a	

In [4]:
```
# 3. Data Summary - Generate descriptive statistics for the data, including mean, median,
# standard deviation, and quartiles, to understand the central tendency and spread of the data.

df.describe()
```

Out[4]:

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	popula
count	8124	8124	8124	8124	8124	8124	8124	8124	8124	8124	...	8124	8124	8124	8124	8124	8124	8124	8124	
unique	2	6	4	10	2	9	2	2	2	12	...	4	9	9	1	4	3	5	9	
top	e	x	y	n	f	n	f	c	b	b	...	s	w	w	p	w	o	p	w	
freq	4208	3656	3244	2284	4748	3528	7914	6812	5612	1728	...	4936	4464	4384	8124	7924	7488	3968	2388	

4 rows × 23 columns

In [5]:
```
#4. Data Information - check the data types of each column,
# the number of non-null values, and memory usage.

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
 14  stalk-color-above-ring    8124 non-null   object
 15  stalk-color-below-ring    8124 non-null   object
 16  veil-type                 8124 non-null   object
 17  veil-color                8124 non-null   object
 18  ring-number               8124 non-null   object
 19  ring-type                 8124 non-null   object
```

In [6]:
```
# 5. Handling Missing Data - identify and handle missing values using techniques
# like imputation or removal.

missing_values = df.isnull().sum()
print("\nMissing Values:")
print(missing_values)
```

```
Missing Values:
class                       0
cap-shape                   0
cap-surface                 0
cap-color                   0
bruises                     0
odor                        0
gill-attachment             0
gill-spacing                0
gill-size                   0
gill-color                  0
stalk-shape                 0
stalk-root                  0
stalk-surface-above-ring    0
stalk-surface-below-ring    0
stalk-color-above-ring      0
stalk-color-below-ring      0
veil-type                   0
veil-color                  0
ring-number                 0
ring-type                   0
spore-print-color           0
population                  0
```