

```
import pandas as pd

from google.colab import files

uploaded = files.upload()
```


Choose Files

data.csv **data.csv**(text/csv) - 11037 bytes, last modified: 7/16/2022 - 100% done Saving data.csv to data.csv

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

df.head()
```

	name	emoji	Calories (kcal)	Carbohydrates (g)	Fat (g)	Total Sugar (g)	Protein (g)	Total Fat	Saturated Mon
0	grapes	🍇	0.69		0.1810	0.1548	0.0072	0.0016	0.00054
1	melon	🍈	0.28		0.0658	0.0569	0.0111	0.0010	NaN
2	watermelon	🍉	0.30		0.0755	0.0620	0.0061	0.0015	0.00016
3	tangerine	🍊	0.53		0.1334	0.1058	0.0081	0.0031	0.00039
4	lemon	🍋	0.29		0.0932	NaN	0.0110	0.0030	0.00039





```
df.head(12)
```








	emoji	Calories Sugar (kcal)	Carbohydrates Fat (g)	Fat (g)	Total Protein (g)	Total Saturated (g)	Mo	name
0	grapes	🍇	0.69	0.1810	0.1548	0.0072	0.0016	0.00054
1	melon	🍈	0.28	0.0658	0.0569	0.0111	0.0010	NaN
2	watermelon	🍉	0.30	0.0755	0.0620	0.0061	0.0015	0.00016

Remove Duplicates

3	pineapple		0.53	0.1334	0.1058	0.0081	0.0031	0.00039
4	lemon		0.29	0.0932	NaN	0.0110	0.0030	0.00039

```
df.sort_values("name", inplace=True)
df
```

	name	emoji	Calories (kcal)	Carbohydrates (g)	Total Sugar (g)	Protein (g)	Total Fat (g)	Saturated Fat (g)	Mono
5	banana		0.89	0.2284	0.1223	0.0109	0.0033	0.00112	
11	cherries		0.63	0.1601	0.1282	0.0106	0.0020	0.00038	
0	grapes		0.69	0.1810	0.1548	0.0072	0.0016	0.00054	
8	green apple		0.58	0.1361	0.0959	0.0044	0.0019	NaN	
4	lemon		0.29	0.0932	NaN	0.0110	0.0030	0.00039	
...	
688	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
689	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
690	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
691	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
692	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

693 rows × 13 columns



banana 🍌

pineapple 🍍

```
df.drop_duplicates(subset="name", keep=False, inplace=True) df
```

	name	emoji	Calories (kcal)	Carbohydrates (g)	Total Sugar (g)	Protein (g)	Total Fat (g)	Saturated Fat (g)	Mo
--	------	-------	--------------------	----------------------	-----------------------	----------------	---------------------	-------------------------	----

9	pear	🍐	0.57	0.1523	0.0975	0.0036	0.0014	0.00022	
6	orange	🍊	0.53	0.1334	0.1058	0.0081	0.0031	0.00039	

```
df.dropna()
```

			Calories (kcal)	Carbohydrates (g)	Total Sugar (g)	Protein (g)	Total Fat (g)	Saturated Fat (g)	Mo
5	banana	🍌	0.89	0.2284	0.1223	0.0109	0.0033	0.00112	
11	cherries	🍒	0.63	0.1601	0.1282	0.0106	0.0020	0.00038	
0	grapes	🍇	0.69	0.1810	0.1548	0.0072	0.0016	0.00054	
10	peach	🍑	0.39	0.0954	0.0839	0.0091	0.0025	0.00019	
9	pear	🍐	0.57	0.1523	0.0975	0.0036	0.0014	0.00022	
3	tangerine	🍊	0.53	0.1334	0.1058	0.0081	0.0031	0.00039	
2	watermelon	🍉	0.30	0.0755	0.0620	0.0061	0.0015	0.00016	



```
df.duplicated()
```

5	banana	🍌	0.89	0.2284	0.1223	0.0109	0.0033	0.00112	
11	cherries	🍒	0.63	0.1601	0.1282	0.0106	0.0020	0.00038	
0	grapes	🍇	0.69	0.1810	0.1548	0.0072	0.0016	0.00054	
8	green apple	🍏	0.58	0.1361	0.0959	0.0044	0.0019	NaN	
4	lemon	🍋	0.29	0.0932	NaN	0.0110	0.0030	0.00039	
1	melon	🍈	0.28	0.0658	0.0569	0.0111	0.0010	NaN	
10	peach	🍑	0.39	0.0954	0.0839	0.0091	0.0025	0.00019	
	red apple	🍏							

tangerine 🍊 name
emoji watermelon 🍉

```
5      False
11     False
0      False
8      False
4      False
1      False
10     False
9      False
6      False
7      False
3      False 2      False dtype: bool
df.name.duplicated().sum()
```

0

```
df.drop_duplicates(subset=['name','emoji'])
```

	name	emoji	Calories (kcal)	Carbohydrates (g)	Total Sugar (g)	Protein (g)	Total Fat (g)	Saturated Mo
5	banana	🍌	0.89	0.2284	0.1223	0.0109	0.0033	0.00112
11	cherries	🍒	0.63	0.1601	0.1282	0.0106	0.0020	0.00038
0	grapes	🍇	0.69	0.1810	0.1548	0.0072	0.0016	0.00054
8	green apple	🍏	0.58	0.1361	0.0959	0.0044	0.0019	NaN
4	lemon	🍋	0.29	0.0932	NaN	0.0110	0.0030	0.00039
1	melon	🍈	0.28	0.0658	0.0569	0.0111	0.0010	NaN

10	peach	🍑	0.39	0.0954	0.0839	0.0091	0.0025	0.00019
9	pear	🍐	0.57	0.1523	0.0975	0.0036	0.0014	0.00022
6	pineapple	🍍	NaN	0.1312	0.0985	0.0054	0.0012	0.00009
7	red apple	🍏	0.63	0.1522	NaN	0.0020	0.0018	NaN
3	tangerine	🍊	0.53	0.1334	0.1058	0.0081	0.0031	0.00039
2	watermelon	🍉	0.30	0.0755	0.0620	0.0061	0.0015	0.00016



```
df['emoji'].isnull()
```

```
5      False
11     False
0      False
8      False
4      False
1      False
10     False
9      False
6      False
7      False
3      False
2      False
Name: emoji, dtype: bool
```

```
df.isnull()
```

	name emoji		Calories (kcal)	Carbohydrates (g)	Fat (g)	Total Sugar (g)	Protein (g)	Total Fat	Saturated Monounsa
5	False	False	False		False	False	False	False	False
11	False	False	False		False	False	False	False	False
0	False	False	False		False	False	False	False	False
8	False	False	False		False	False	False	False	True
4	False	False	False		False	True	False	False	False
1	False	False	False		False	False	False	False	True
10	False	False	False		False	False	False	False	False
9	False	False	False		False	False	False	False	False
6	False	False	True		False	False	False	False	False
7	False	False	False		False	True	False	False	True
3	False	False	False		False	False	False	False	False
2	False	False	False		False	False	False	False	False

Deleting Column with Missing Data

```
new_df=df.dropna(axis=1) new_df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 12 entries, 5 to 2 Data
columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   name                                12 non-null      object
1   emoji                               12 non-null      object
2   Carbohydrates (g)                   12 non-null      float64
3   Protein (g)                         12 non-null      float64
4   Total Fat (g)                       12 non-null      float64
5   Polyunsaturated Fat (g)             12 non-null      float64
6   Total Fiber (g)                     12 non-null      float64
7   Cholesterol (mg)                    12 non-null      float64
8   URL                                12 non-null      object
dtypes: float64(6), object(3) memory usage: 960.0+ bytes

Filling Missing Data
```

```
df.isna().sum().sum()








8
df.isna().sum()/len(df)*100

name                                0.000000 emoji
0.000000 Calories (kcal)
8.333333
Carbohydrates (g)                   0.000000
Total Sugar (g)                      16.666667
Protein (g)                         0.000000
Total Fat (g)                       0.000000
Saturated Fat (g)                    25.000000
Monounsaturated Fat (g)              16.666667
Polyunsaturated Fat (g)              0.000000
Total Fiber (g)                     0.000000
Cholesterol (mg)                     0.000000
URL                                0.000000 dtype:
float64 df.loc[:,df.isnull().any()].columns

Index(['Calories (kcal)', 'Total Sugar (g)', 'Saturated Fat (g)',
      'Monounsaturated Fat (g)'],
      dtype='object')

df.dropna()
```

		Total	Total		
		Calories	Carbohydrates	Protein	Saturated Mo
name	emoji			Sugar	Fat

			(kcal)	(g)	(g)	(g)	(g)	Fat (g)
5	banana		0.89	0.2284	0.1223	0.0109	0.0033	0.00112
11	cherries		0.63	0.1601	0.1282	0.0106	0.0020	0.00038
0	grapes		0.69	0.1810	0.1548	0.0072	0.0016	0.00054
10	peach		0.39	0.0954	0.0839	0.0091	0.0025	0.00019
9	pear		0.57	0.1523	0.0975	0.0036	0.0014	0.00022
3	tangerine		0.53	0.1334	0.1058	0.0081	0.0031	0.00039
2	watermelon		0.30	0.0755	0.0620	0.0061	0.0015	0.00016

