```
import seaborn as sns
import pandas as pd
import numpy as np


df=sns.load_dataset('tips')
df
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype   
---  ------      --------------  -----   
 0   total_bill  244 non-null    float64 
 1   tip         244 non-null    float64 
 2   sex         244 non-null    category
 3   smoker      244 non-null    category
 4   day         244 non-null    category
 5   time        244 non-null    category
 6   size        244 non-null    int64   
dtypes: category(4), float64(2), int64(1)
memory usage: 7.4 KB
```

```
from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()
```

```
df['smoker']=lb.fit_transform(df['smoker'])
df['sex']=lb.fit_transform(df['sex'])
df['time']=lb.fit_transform(df['time'])
df['day']=lb.fit_transform(df['day'])
```

```
df.dtypes
```

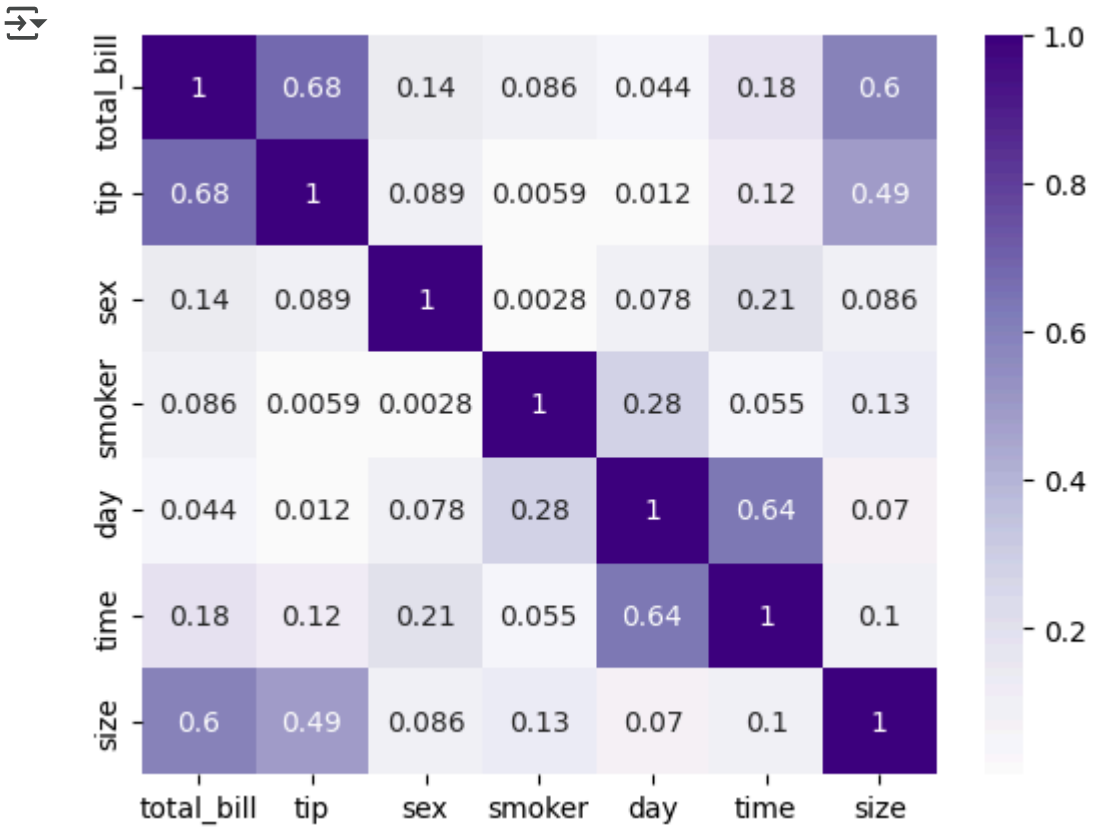
	0
total_bill	float64
tip	float64
sex	int64
smoker	int64
day	int64
time	int64
size	int64

dtype: object

```
df.corr()
```

	total_bill	tip	sex	smoker	day	time	size
total_bill	1.000000	0.675734	0.144877	0.085721	-0.043550	-0.183118	0.598315
tip	0.675734	1.000000	0.088862	0.005929	-0.011548	-0.121629	0.489299
sex	0.144877	0.088862	1.000000	0.002816	-0.078292	-0.205231	0.086195
smoker	0.085721	0.005929	0.002816	1.000000	-0.282721	-0.054921	-0.133178
day	-0.043550	-0.011548	-0.078292	-0.282721	1.000000	0.638019	0.069510
time	-0.183118	-0.121629	-0.205231	-0.054921	0.638019	1.000000	-0.103411
size	0.598315	0.489299	0.086195	-0.133178	0.069510	-0.103411	1.000000

```
import matplotlib.pyplot as plt
sns.heatmap(np.abs(df.corr()),cmap='Purples',annot=True)
plt.show()
```


total_bill
tip
sex
smoker
day
time
size
1 0.68 0.14 0.086 0.044 0.18 0.6
0.68 1 0.089 0.0059 0.012 0.12 0.49
0.14 0.089 1 0.0028 0.078 0.21 0.086
0.086 0.0059 0.0028 1 0.28 0.055 0.13
0.044 0.012 0.078 0.28 1 0.64 0.07
0.18 0.12 0.21 0.055 0.64 1 0.1
0.6 0.49 0.086 0.13 0.07 0.1 1
1.0
0.8
0.6
0.4
0.2

```
df.describe()
```

	total_bill	tip	sex	smoker	day	time	size
count	244.000000	244.000000	244.000000	244.000000	244.000000	244.000000	244.000000

```
from sklearn.preprocessing import StandardScaler
std_scaler=StandardScaler()
```

min	3.070000	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000

```
scale_array=std_scaler.fit_transform(df)
```

```
scale_array.shape
```

```
(244, 7)
```

max	50.810000	10.000000	1.000000	1.000000	3.000000	1.000000	6.000000

Start coding or generate with AI.