Plotting charts with seaborn library

In [46]:
```
# Loading the library
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
```

In [48]:
```
# Load the built-in tips dataset
tips = sns.load_dataset('tips')
tips.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
```

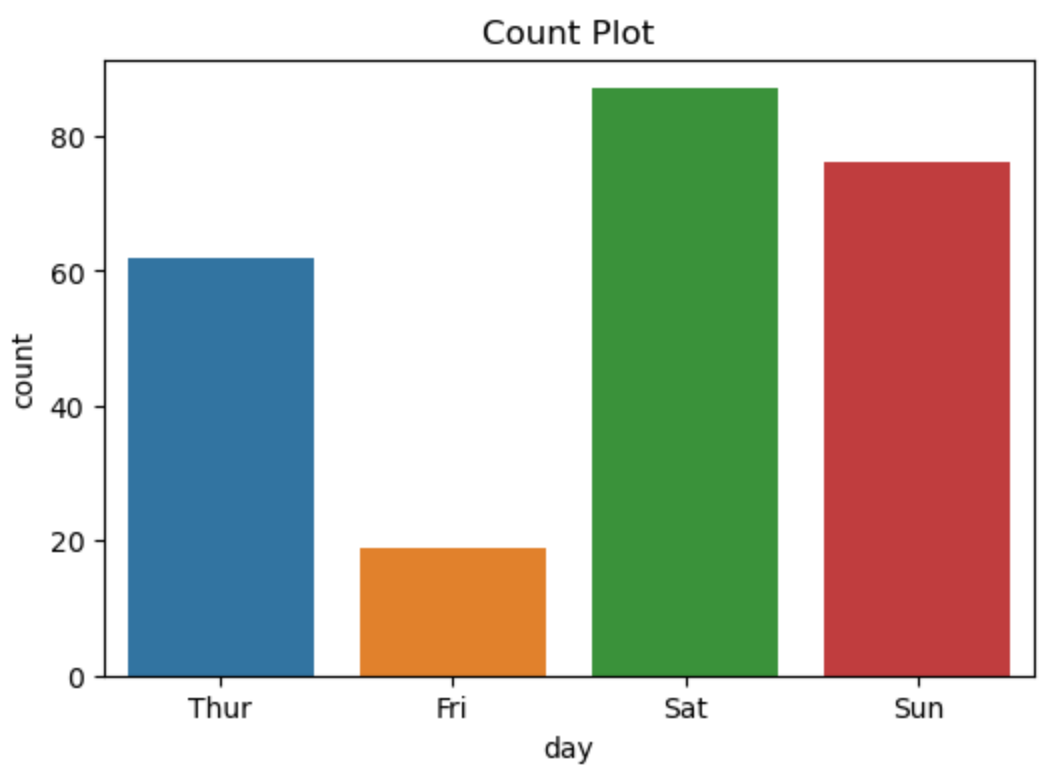
In [50]:
```
tips.head()
```

Out[50]:

	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

1. Count Plot

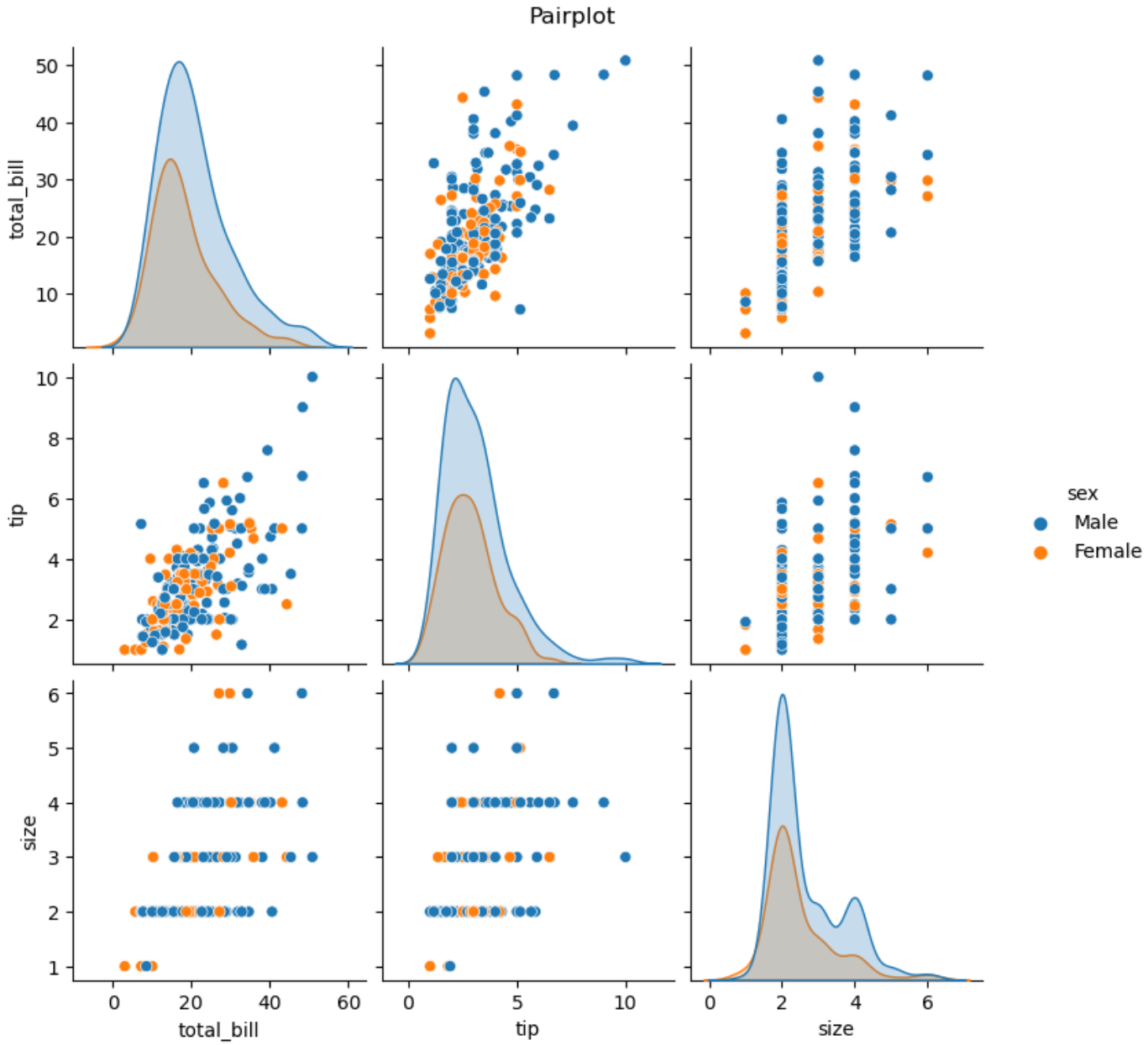
In [53]:
```
plt.figure(figsize=(6,4))
sns.countplot(x='day', data=tips)
plt.title('Count Plot')
plt.show()
```



2. pairplot

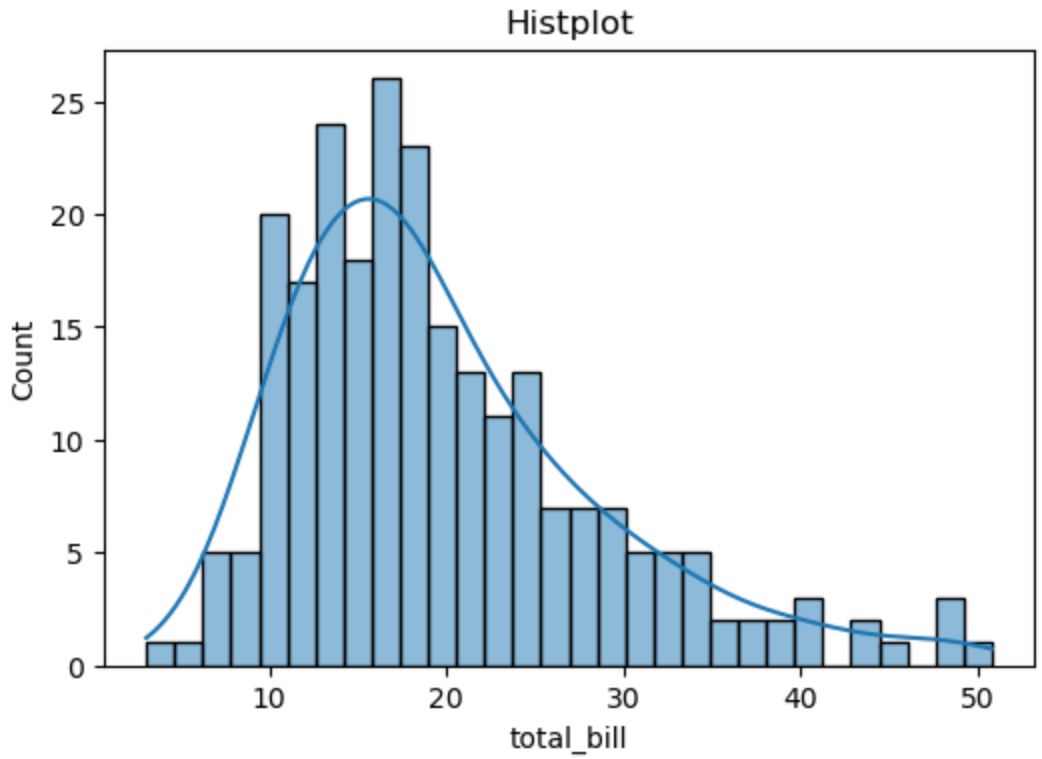
In [56]:
```
sns.pairplot(tips, hue='sex')
plt.suptitle('Pairplot', y=1.02)
plt.show()
```

```
/Users/josephkambham/anaconda3/lib/python3.11/site-packages/seaborn/axisgrid.py:118: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



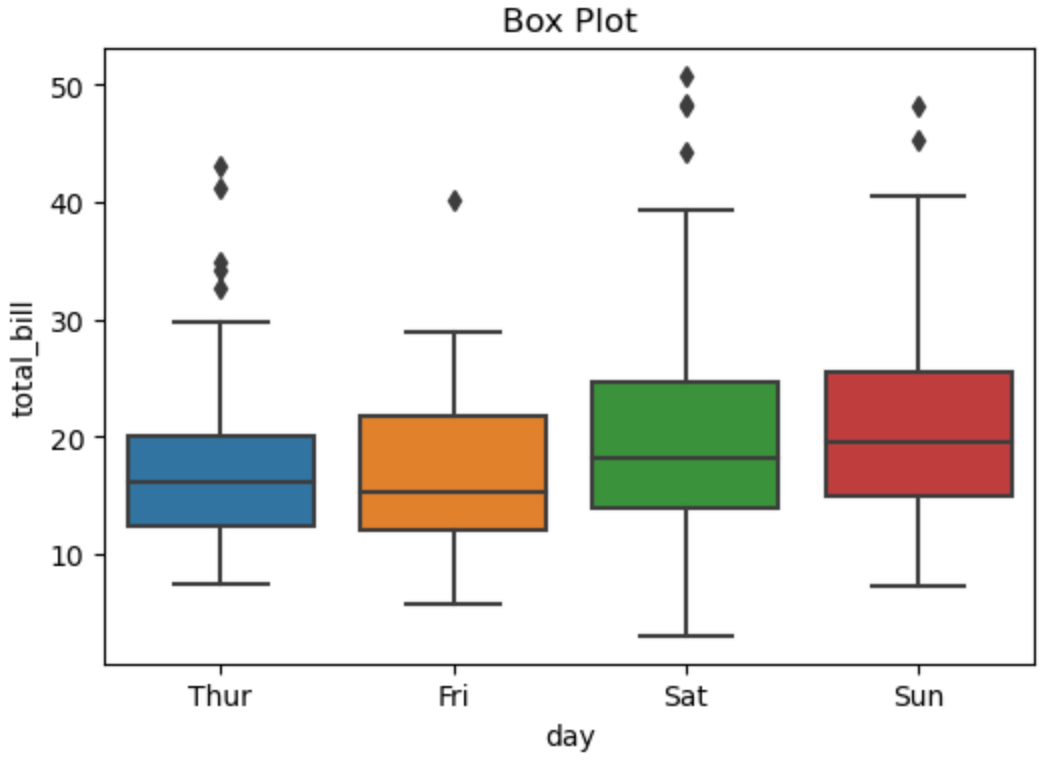
3. Histplot

In [59]:
```
plt.figure(figsize=(6,4))
sns.histplot(tips['total_bill'], kde=True, bins=30)
plt.title('Histplot')
plt.show()
```



4. Box plot

In [62]:
```
plt.figure(figsize=(6,4))
sns.boxplot(x='day', y='total_bill', data=tips)
plt.title('Box Plot')
plt.show()
```



5. Heatmap

In [65]:
```
plt.figure(figsize=(6,4))
correlation_matrix = tips.corr()
```

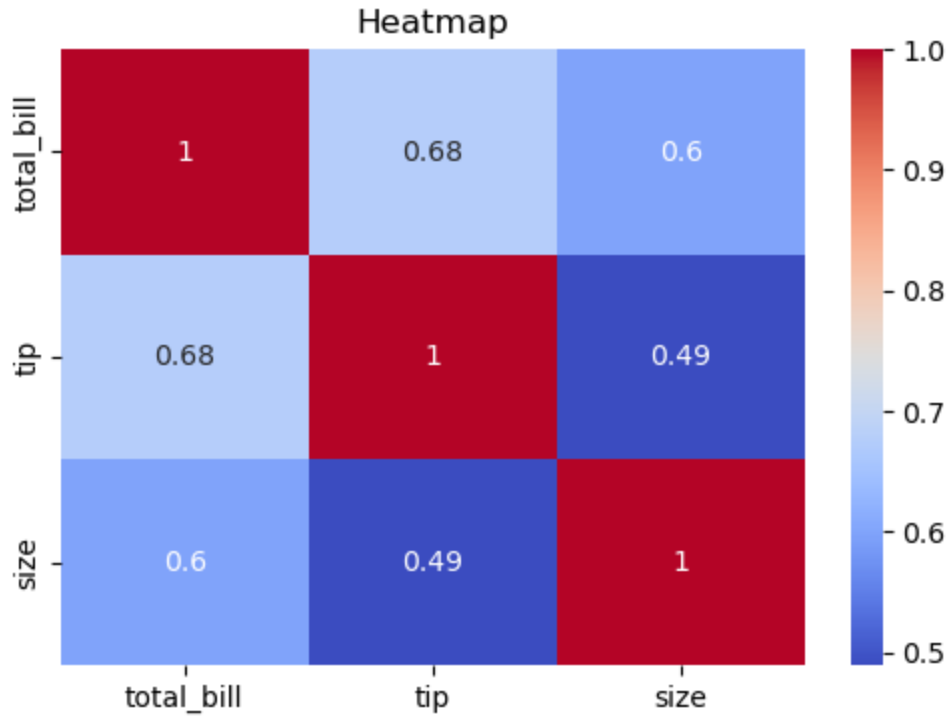
```
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Heatmap')
plt.show()
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[65], line 2
      1 plt.figure(figsize=(6,4))
----> 2 correlation_matrix = tips.corr()
      3 sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
      4 plt.title('Heatmap')

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/frame.py:10054, in DataFrame.c
orr(self, method, min_periods, numeric_only)
  10052 cols = data.columns
  10053 idx = cols.copy()
> 10054 mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
  10056 if method == "pearson":
  10057     correl = libalgos.nancorr(mat, minp=min_periods)

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/frame.py:1838, in DataFrame.to
_numpy(self, dtype, copy, na_value)
   1836 if dtype is not None:
   1837     dtype = np.dtype(dtype)
-> 1838 result = self._mgr.as_array(dtype=dtype, copy=copy, na_value=na_value)
   1839 if result.dtype is not dtype:
   1840     result = np.array(result, dtype=dtype, copy=False)

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/internals/managers.py:1732, in
BlockManager.as_array(self, dtype, copy, na_value)
   1730         arr.flags.writeable = False
   1731 else:
-> 1732     arr = self._interleave(dtype=dtype, na_value=na_value)
   1733     # The underlying data was copied within _interleave, so no need
   1734     # to further copy if copy=True or setting na_value
   1736 if na_value is not lib.no_default:

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/internals/managers.py:1788, in
BlockManager._interleave(self, dtype, na_value)
   1782 rl = blk.mgr_locs
   1783 if blk.is_extension:
   1784     # Avoid implicit conversion of extension blocks to object
   1785
   1786     # error: Item "ndarray" of "Union[ndarray, ExtensionArray]" has no
   1787     # attribute "to_numpy"
-> 1788     arr = blk.values.to_numpy(  # type: ignore[union-attr]
   1789         dtype=dtype,
   1790         na_value=na_value,
   1791     )
   1792 else:
   1793     arr = blk.get_values(dtype)

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/arrays/base.py:485, in Extensi
onArray.to_numpy(self, dtype, copy, na_value)
    456 def to_numpy(
    457     self,
    458     dtype: npt.DTypeLike | None = None,
    459     copy: bool = False,
    460     na_value: object = lib.no_default,
    461 ) -> np.ndarray:
    462     """
    463     Convert to a NumPy ndarray.
    464
   (...)
```

```
    483     numpy.ndarray
    484     """
--> 485     result = np.asarray(self, dtype=dtype)
    486     if copy or na_value is not lib.no_default:
    487         result = result.copy()

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/arrays/_mixins.py:86, in ravel_compat.<locals>.method(self, *args, **kwargs)
     83 @wraps(meth)
     84 def method(self, *args, **kwargs):
     85     if self.ndim == 1:
---> 86         return meth(self, *args, **kwargs)
     88     flags = self._ndarray.flags
     89     flat = self.ravel("K")

File ~/anaconda3/lib/python3.11/site-packages/pandas/core/arrays/categorical.py:1344, in Categorical.__array__(self, dtype)
   1342 ret = take_nd(self.categories._values, self._codes)
   1343 if dtype and not is_dtype_equal(dtype, self.categories.dtype):
-> 1344     return np.asarray(ret, dtype)
   1345 # When we're a Categorical[ExtensionArray], like Interval,
   1346 # we need to ensure __array__ gets all the way to an
   1347 # ndarray.
   1348 return np.asarray(ret)

ValueError: could not convert string to float: 'No'
<Figure size 600x400 with 0 Axes>
```

In [67]:
```
# The above error is coming up since there are string values in columns 'sex', 'smoker',
```

In [69]:
```
#Select only the numeric columns
numeric_tips = tips[['total_bill', 'tip', 'size']]

# PLotting the correlational matrix
plt.figure(figsize=(6,4))
correlation_matrix = numeric_tips.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Heatmap')
plt.show()
```



In [71]:
```
# Since there are columns which having string values the correlation matrix showing erro
# Will convert categorical values using get_dummies method
```

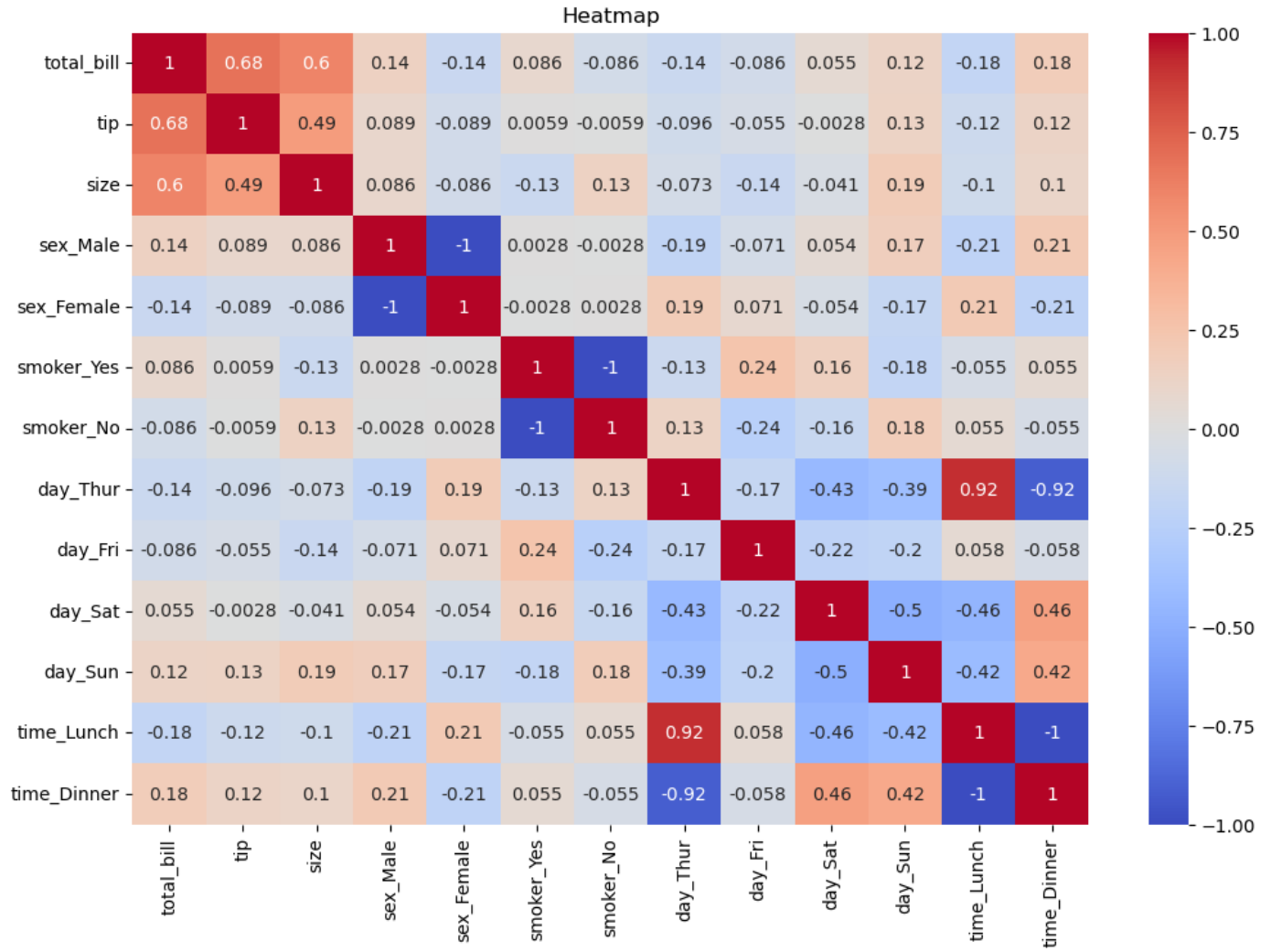
In [73]:
```
# One-hot encode the categorical variables
tips_encoded = pd.get_dummies(tips, columns=['sex', 'smoker', 'day', 'time'])
```

In [75]:
```
tips_encoded.head()
```

Out[75]:

	total_bill	tip	size	sex_Male	sex_Female	smoker_Yes	smoker_No	day_Thur	day_Fri	day_Sat	day_
0	16.99	1.01	2	False	True	False	True	False	False	False	
1	10.34	1.66	3	True	False	False	True	False	False	False	
2	21.01	3.50	3	True	False	False	True	False	False	False	
3	23.68	3.31	2	True	False	False	True	False	False	False	
4	24.59	3.61	4	False	True	False	True	False	False	False	

In [44]:
```
plt.figure(figsize=(12,8))
correlation_matrix = tips_encoded.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Heatmap')
plt.show()
```



In [ ]:
```
!jupyter nbconvert --to webpdf --allow-chromium-download Week6_Lab2.ipynb
```