

У цьому домашньому завданні ми будемо працювати з набором даних Tips. Нижче інформація про нього.

Про датасет:

Tips - це датасет, який містить інформацію про рахунки та чайові в ресторані.

Колонки датасету:

- total_bill - загальна сума рахунку (долари)
- tip - розмір чайових (долари)
- sex - стать відвідувача (Male/Female)
- smoker - чи курить (Yes/No)
- day - день тижня (Thur/Fri/Sat/Sun)
- time - час відвідування (Lunch/Dinner)
- size - розмір групи (кількість людей)

Джерело: Оригінальні дані зібрані в одному з американських ресторанів у 1990-х роках. Скачати і переглянути набір даних можна тут: <https://www.kaggle.com/datasets/sakhisatre/tips-dataset>

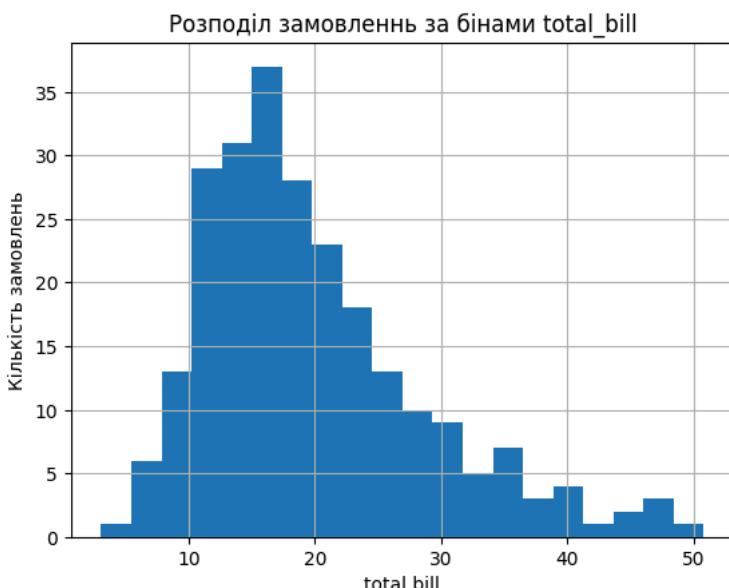
А також ми вже вивантажили файл з даними і завантажили вам у Матеріали до цього модуля.

1. Гістограма розподілу рахунків Завантажте датасет Tips (Чайові) з файла у матеріалах `tips.csv`. Створіть гістограму для колонки `total_bill` з 20 bins. Опціонально можна додати підписи осей ОХ, ОУ, а також назву графіку.

```
import pandas as pd
import matplotlib.pyplot as plt

df=pd.read_csv('tip_dataset.csv')
df.columns
Index(['total_bill', 'tip', 'sex', 'smoker', 'day', 'time', 'size'], dtype='object')

df['total_bill'].hist(bins=20);
plt.title('Розподіл замовлень за бінами total_bill');
plt.xlabel('otal bill');
plt.ylabel('Кількість замовлень');
```

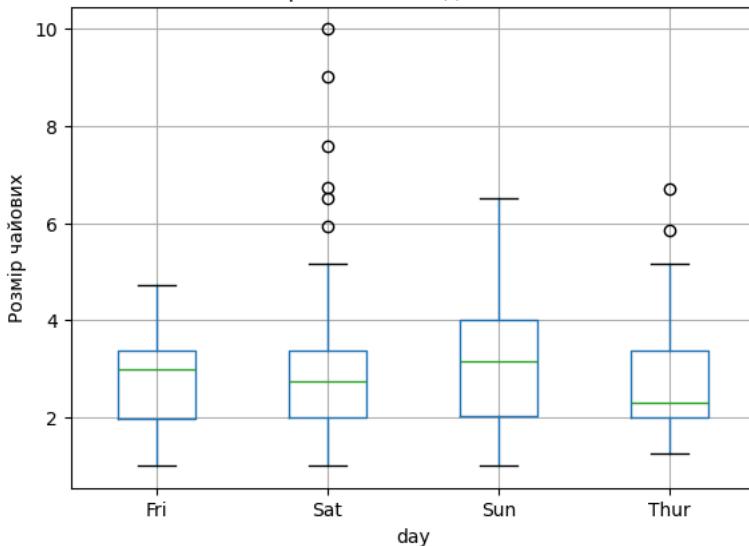


2. Box plot чайових за днями тижня

Створіть box plot, який показує розподіл чайових (`tip`) за днями тижня (`day`). В який день навище медіанне (середнє) значення чайових? Зелена горизонтальна лінія всередині кожного box-plot.

```
df.boxplot(column = 'tip', by= 'day');
plt.title('Розмір чайових за днями тижня');
plt.ylabel('Розмір чайових');
```

Boxplot grouped by day
Розмір чайових за днями тижня

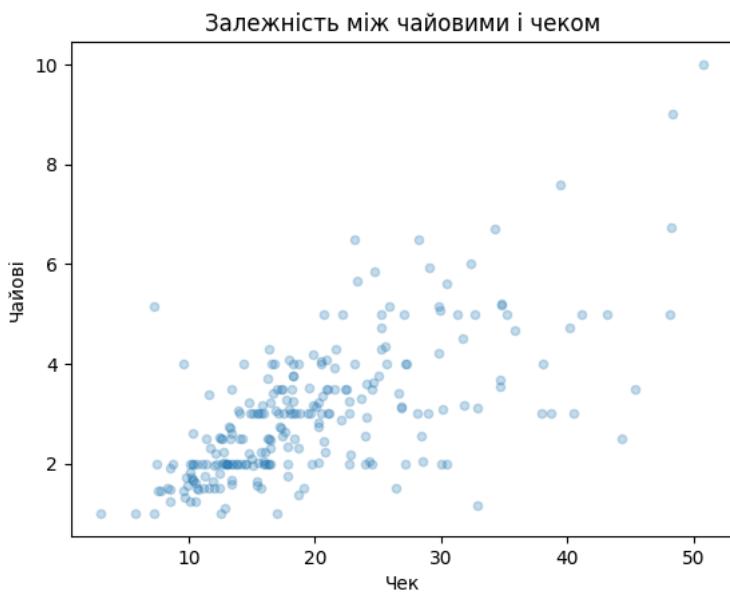


```
print ('Найбільше чайових залишають в неділю.')
```

3. Scatter plot: рахунок vs чайові

Створіть діаграму розсіювання для `total_bill` та `tip`.

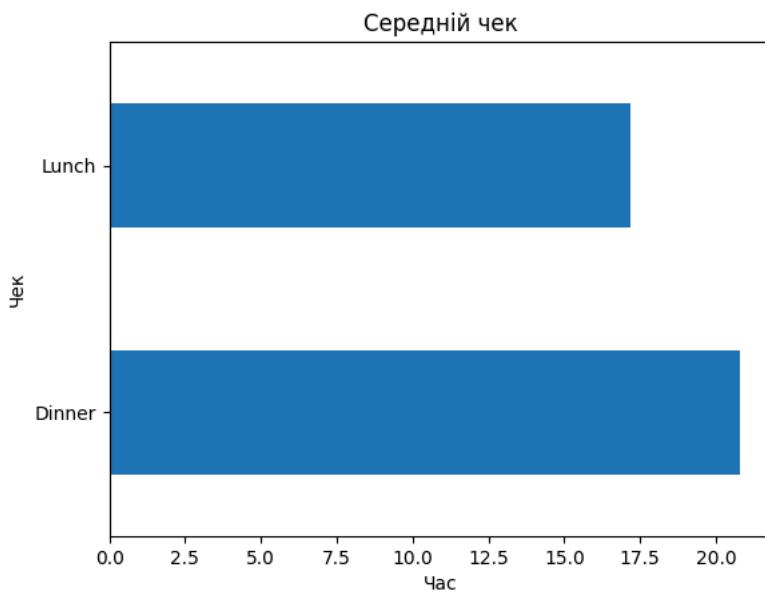
```
df.plot.scatter(x= 'total_bill', y='tip', alpha=0.25);
plt.title('Залежність між чайовими і чеком');
plt.xlabel('Чек');
plt.ylabel('Чайові');
```



4. Bar chart: середній рахунок за часом

Створіть стовпчасту діаграму середнього рахунку за часом дня (`time`).

```
df.groupby(['time'])['total_bill'].mean().plot(kind='barh');
plt.title('Середній чек');
plt.xlabel('Час');
plt.ylabel('Чек');
```



5. Інтерактивна діаграма розсіювання

Створіть з допомогою бібліотек Plotly інтерактивний scatter plot для `total_bill` vs `tip` з кольором за статтю (`sex`).

```
import plotly.express as px
import plotly.graph_objects as go
```

```
df['percent_tip'] = df['tip']/df['total_bill']
df['percent_tip']= (df['percent_tip']*100).round(2)
df.head()
```

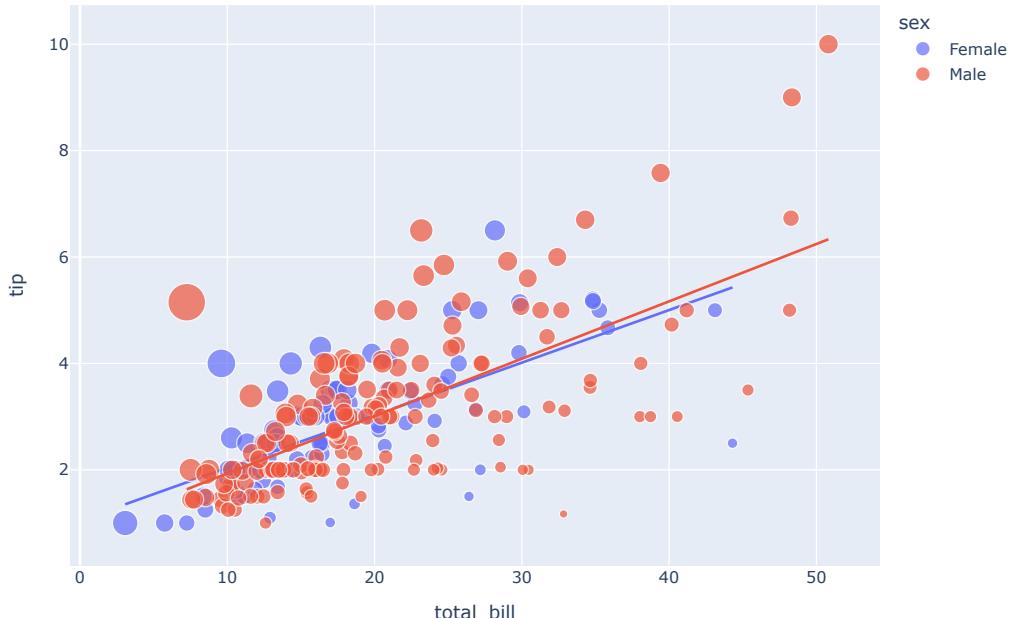
	total_bill	tip	sex	smoker	day	time	size	percent_tip	More
0	16.99	1.01	Female	No	Sun	Dinner	2	5.94	More
1	10.34	1.66	Male	No	Sun	Dinner	3	16.05	More
2	21.01	3.50	Male	No	Sun	Dinner	3	16.66	More
3	23.68	3.31	Male	No	Sun	Dinner	2	13.98	More
4	24.59	3.61	Female	No	Sun	Dinner	4	14.68	More

Далее: [Создать код с переменной df](#) [New interactive sheet](#)

```
from numpy import size

scatter_1 = px.scatter (df, x='total_bill', y= 'tip', color='sex',
                      title='Розмір чека та чаївих',
                      size='percent_tip',
                      hover_data='percent_tip',
                      trendline='ols'
                     )
scatter_1.update_layout(
    width=800,
    height=600
)
scatter_1.show();
```

Розмір чека та чайових



Чтобы изменить содержимое ячейки, дважды нажмите на нее (или выберите "Ввод")

6. Інтерактивний box plot: чайові за днями та статтю

Створіть інтерактивний box plot для чайових за днями тижня з розбивкою за кольором по статі.

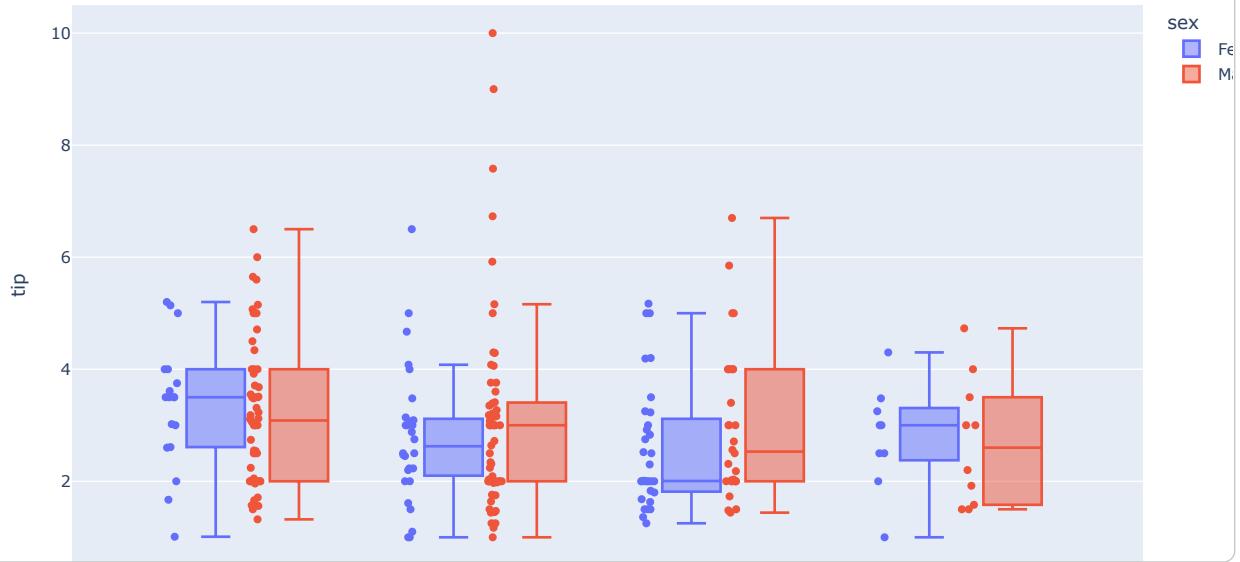
```
medians = df.groupby(['day', 'sex'])[['tip', 'percent_tip']].median()
display(medians);
```

	tip	percent_tip	grid
day	sex	bar	bar
Fri	Female	3.000	19.510
	Male	2.600	13.375
Sat	Female	2.625	15.110
	Male	3.000	15.180
Sun	Female	3.500	16.940
	Male	3.085	15.780
Thur	Female	2.005	15.310
	Male	2.530	15.550

Далее: [Создать код с переменной medians](#) [New interactive sheet](#)

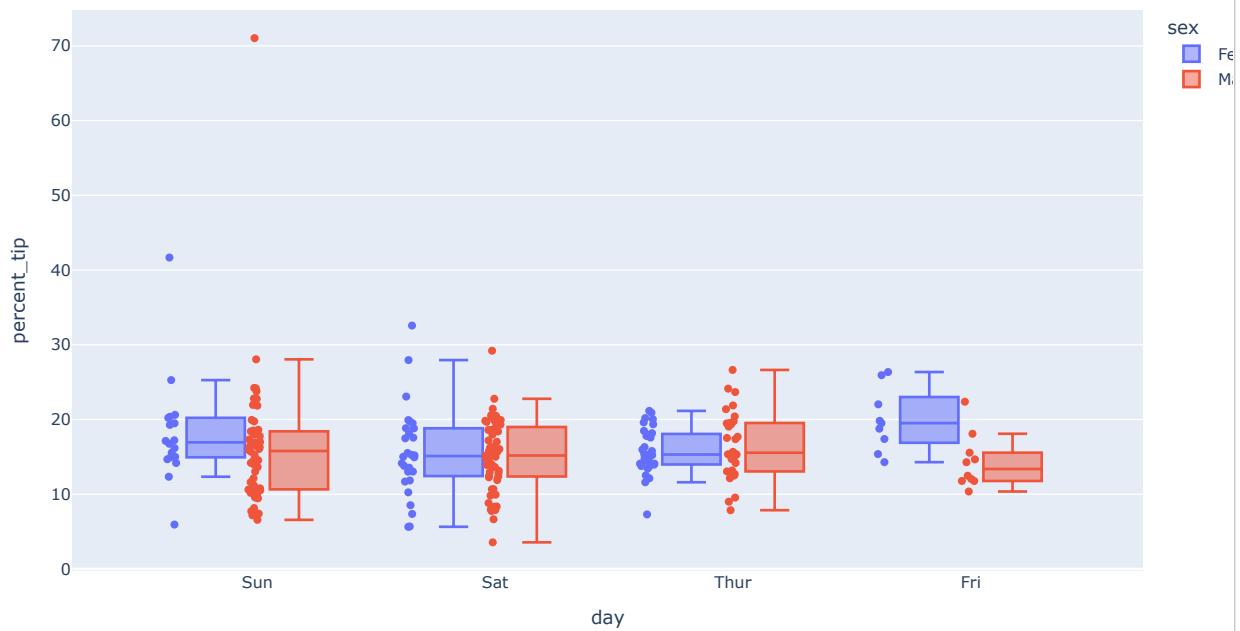
```
boxplot_1=px.box(df, x='day', y='tip',
                  color='sex',
                  points='all',
                  title="Box plot of tips per day of week",);
boxplot_1.update_layout(
    width=1000,
    height=600
)
boxplot_1.show();
```

Box plot of tips per day of week



```
boxplot_2=px.box(df, x='day', y='percent_tip',
                  color='sex',
                  points='all',
                  title="Box plot of %tips per day of week",);
boxplot_2.update_layout(
    width=1000,
    height=600
)
boxplot_2.show();
```

Box plot of %tips per day of week



В абсолютних значеннях чоловіки залишають більші чаїові, а в відносніх (до суми чека) жінки на чай залишають більше.