

- הכרות עם שפת התכנות פייתון (PYTHON)
- הכרות עם סביבת העבודה של PYTHON
- הכרות עם התכנות הבסיסי בפייתון

➤ ספריות – הכנה ללמידת מכונה:

matplotlib✓

pandas✓

Numpy✓

seaborn✓

הכל נמצא ברשת, צריך רק לדעת איפה לחפש:

אתרים ללמוד עצמי – pandas

- ❑ [Pandas Tutorial W3Schools](#)
- ❑ [10 minutes to pandas](#)
- ❑ [Python pandas tutorial: The ultimate guide for beginners](#)
- ❑ [Pandas Tutorial - geeksforgeeks.org](#)

- ספריה לייצוג ולניהול נתונים במבנה של טבלאות (data frames)
- מאפשר ניתוח סטטיסטי של המידע, התמודדות עם ערכים חסרים, ערכים כפולים

ספריית pandas

➤ **ייבוא -** `import pandas as pd`

➤ **טעינת קובץ –** הספרייה כוללת פונקציות לטעינה של קבצים במגוון

פורמטים:

CSV: `read_csv (filename)`

Excel : `read_excel (filename)`

HTML: `read_html (filename)`

דוגמא: קריאת טבלה מקובץ CSV

```
df = pd.read_csv('/content/drive/My Drive/Data/datasets_19_420_Iris.csv')
df
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

	A	B	C	D	E	F
1	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
2	1	5.1	3.5	1.4	0.2	Iris-setosa
3	2	4.9	3	1.4	0.2	Iris-setosa
4	3	4.7	3.2	1.3	0.2	Iris-setosa
5	4	4.6	3.1	1.5	0.2	Iris-setosa
6	5	5	3.6	1.4	0.2	Iris-setosa
7	6	5.4	3.9	1.7	0.4	Iris-setosa
8	7	4.6	3.4	1.4	0.3	Iris-setosa
9	8	5	3.4	1.5	0.2	Iris-setosa
10	9	4.4	2.9	1.4	0.2	Iris-setosa
11	10	4.9	3.1	1.5	0.1	Iris-setosa
12	11	5.4	3.7	1.5	0.2	Iris-setosa
13	12	4.8	3.4	1.6	0.2	Iris-setosa
14	13	4.8	3	1.4	0.1	Iris-setosa
15	14	4.3	3	1.1	0.1	Iris-setosa
16	15	5.8	4	1.2	0.2	Iris-setosa
17	16	5.7	4.4	1.5	0.4	Iris-setosa
18	17	5.4	3.9	1.3	0.4	Iris-setosa
19	18	5.1	3.5	1.4	0.3	Iris-setosa
20	19	5.7	3.8	1.7	0.3	Iris-setosa
21	20	5.1	3.8	1.5	0.3	Iris-setosa
22	21	5.4	3.4	1.7	0.2	Iris-setosa
23	22	5.1	3.7	1.5	0.4	Iris-setosa
24	23	4.6	3.6	1	0.2	Iris-setosa
25	24	5.1	3.3	1.7	0.5	Iris-setosa
26	25	4.8	3.4	1.9	0.2	Iris-setosa
27	26	5	3	1.6	0.2	Iris-setosa

- פתחו מחברת חדשה
- הדפיסו hello world
- חברו את המחברת לGoogle drive שלכם
- חפשו בגוגל את בסיס הנתונים iris.csv. הורידו לgoogle drive שלכם
- טענו את טבלת נתוני Iris לסביבת colab באמצעות Pandas
- הציגו את הטבלה

- ❑ הצגת ראש הטבלה
- ❑ בחירת שמות העמודות
- ❑ בחירת עמודות
- ❑ בחירת שורות
- ❑ מיון
- ❑ חישוב מדדים סטטיסטיים (סטטיסטיקה תיאורית)

הצגת ראש הטבלה - df.head()



```
df.head()
```



	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa

הצגת שמות העמודות - df.columns



df.columns

```
[> Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',  
        'variety'],  
        dtype='object')
```

בחירת עמודה - df[col] או df[[col]]

אפשרות ב: df[[col]] :
התוצאה היא טבלה

אפשרות א: df[col] :
התוצאה היא רשימה

```
[ ] df[['sepal.length']]
```



sepal.length

0	5.1
1	4.9
2	4.7
3	4.6
4	5.0
...	...
145	6.7
146	6.3
147	6.5
148	6.2
149	5.9

150 rows × 1 columns



```
df['sepal.length']
```



```
0      5.1
1      4.9
2      4.7
3      4.6
4      5.0
...
145     6.7
146     6.3
147     6.5
148     6.2
149     5.9
```

Name: sepal.length, Length: 150, dtype: float64

בחירת עמודות df[[cols]]



```
df[['sepal.length', 'sepal.width']]
```



	sepal.length	sepal.width
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6
...
145	6.7	3.0
146	6.3	2.5
147	6.5	3.0
148	6.2	3.4
149	5.9	3.0

150 rows × 2 columns

בחירת שורות לפי אינדקס df[from:to]

▶ df[0:10]

	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa
5	5.4	3.9	1.7	0.4	Setosa
6	4.6	3.4	1.4	0.3	Setosa
7	5.0	3.4	1.5	0.2	Setosa
8	4.4	2.9	1.4	0.2	Setosa
9	4.9	3.1	1.5	0.1	Setosa

בחירת שורות באמצעות תנאי

```
df[df['variety']=='Virginica']
```



	sepal.length	sepal.width	petal.length	petal.width	variety
100	6.3	3.3	6.0	2.5	Virginica
101	5.8	2.7	5.1	1.9	Virginica
102	7.1	3.0	5.9	2.1	Virginica
103	6.3	2.9	5.6	1.8	Virginica
104	6.5	3.0	5.8	2.2	Virginica
105	7.6	3.0	6.6	2.1	Virginica
106	4.9	2.5	4.5	1.7	Virginica
107	7.3	2.9	6.3	1.8	Virginica
108	6.7	2.5	5.8	1.8	Virginica
109	7.2	3.6	6.1	2.5	Virginica

בחירת שורות באמצעות תנאי (2)

```
df[df['petal.length'] > 6]
```



	sepal.length	sepal.width	petal.length	petal.width	variety
105	7.6	3.0	6.6	2.1	Virginica
107	7.3	2.9	6.3	1.8	Virginica
109	7.2	3.6	6.1	2.5	Virginica
117	7.7	3.8	6.7	2.2	Virginica
118	7.7	2.6	6.9	2.3	Virginica
122	7.7	2.8	6.7	2.0	Virginica
130	7.4	2.8	6.1	1.9	Virginica
131	7.9	3.8	6.4	2.0	Virginica
135	7.7	3.0	6.1	2.3	Virginica

בחירת שורות באמצעות תנאי מורכב

```
df[(df['variety'] == 'Virginica') & (df['petal.length'] > 6)]
```



	sepal.length	sepal.width	petal.length	petal.width	variety
105	7.6	3.0	6.6	2.1	Virginica
107	7.3	2.9	6.3	1.8	Virginica
109	7.2	3.6	6.1	2.5	Virginica
117	7.7	3.8	6.7	2.2	Virginica
118	7.7	2.6	6.9	2.3	Virginica
122	7.7	2.8	6.7	2.0	Virginica
130	7.4	2.8	6.1	1.9	Virginica
131	7.9	3.8	6.4	2.0	Virginica
135	7.7	3.0	6.1	2.3	Virginica

```
df.sort_values(by='petal.length')
```

	sepal.length	sepal.width	petal.length	petal.width	variety
22	4.6	3.6	1.0	0.2	Setosa
13	4.3	3.0	1.1	0.1	Setosa
14	5.8	4.0	1.2	0.2	Setosa
35	5.0	3.2	1.2	0.2	Setosa
36	5.5	3.5	1.3	0.2	Setosa
...
131	7.9	3.8	6.4	2.0	Virginica
105	7.6	3.0	6.6	2.1	Virginica
117	7.7	3.8	6.7	2.2	Virginica
122	7.7	2.8	6.7	2.0	Virginica
118	7.7	2.6	6.9	2.3	Virginica

150 rows × 5 columns



```
df['petal.length.in'] = df['petal.length'] / 2.54
df
```



	sepal.length	sepal.width	petal.length	petal.width	variety	petal.length.in
0	5.1	3.5	1.4	0.2	Setosa	0.551181
1	4.9	3.0	1.4	0.2	Setosa	0.551181
2	4.7	3.2	1.3	0.2	Setosa	0.511811
3	4.6	3.1	1.5	0.2	Setosa	0.590551
4	5.0	3.6	1.4	0.2	Setosa	0.551181
...
145	6.7	3.0	5.2	2.3	Virginica	2.047244
146	6.3	2.5	5.0	1.9	Virginica	1.968504
147	6.5	3.0	5.2	2.0	Virginica	2.047244
148	6.2	3.4	5.4	2.3	Virginica	2.125984
149	5.9	3.0	5.1	1.8	Virginica	2.007874

150 rows × 6 columns

אינפורמציה חשובה של הטבלה

✓
1s



```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   sepal_length    150 non-null   float64
1   sepal_width     150 non-null   float64
2   petal_length    150 non-null   float64
3   petal_width     150 non-null   float64
4   species         150 non-null   object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```



```
df.describe()
```



	sepal.length	sepal.width	petal.length	petal.width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

בצעו והציגו את החיתוכים הבאים:

❑ כל הפרחים מסוג virginica

❑ כל הפרחים שאינם מסוג virginica

❑ 20 הפרחים הראשונים

❑ 20 הפרחים הראשונים לאחר ערבול הטבלה

❑ כל הפרחים עם sepal length מעל 6.0

הערה: יש להתאים את שמות העמודות ושמות הפרחים לטבלה

שברשותכם

הוספת עמודות תוך ביצוע פעולות חשבוניות

✓
0s



```
df3['E']=df3['A']+df3['B']+1  
df3
```


	A	B	C	D	E
a	18	8	-6	-13	27
b	0	18	-2	2	19
c	-10	-10	3	15	-19
d	19	3	-18	1	23
e	-19	3	9	17	-15



מחיקת שורות / מחיקת עמודות

✓
0s [28] df3 = df3.drop('E',axis=1)
df3

מחיקת עמודה E

✓
0s  # YOUR CODE HERE
df3 = df3.drop(['b'], axis=0)
df3

מחיקת שורה a

```
df4 = df4.dropna(axis=1)
```

מחיקת עמודות עם ערכים ריקים

```
df4 = df4.dropna(axis=0)
```

מחיקת שורות עם ערכים ריקים