

```
pip install pandas
```

Requirement already satisfied: pandas in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (3.0.3)

Requirement already satisfied: numpy>=1.26.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from pandas) (2.5.0)

Requirement already satisfied: python-dateutil>=2.8.2 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from pandas) (2.9.0.post0)

Requirement already satisfied: tzdata in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from pandas) (2026.2)

Requirement already satisfied: six>=1.5 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

Note: you may need to restart the kernel to use updated packages.

```
pip install scikit-learn
```

Requirement already satisfied: scikit-learn in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (1.9.0)

Requirement already satisfied: numpy>=1.24.1 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from scikit-learn) (2.5.0)

Requirement already satisfied: scipy>=1.10.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from scikit-learn) (1.18.0)

Requirement already satisfied: joblib>=1.4.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from scikit-learn) (1.5.3)

Requirement already satisfied: narwhals>=2.0.1 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from scikit-learn) (2.22.1)

Requirement already satisfied: threadpoolctl>=3.5.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from scikit-learn) (3.6.0)

Note: you may need to restart the kernel to use updated packages.

```
pip install pandas
```

Requirement already satisfied: pandas in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (3.0.3)

Requirement already satisfied: numpy>=1.26.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from pandas) (2.5.0)

Requirement already satisfied: python-dateutil>=2.8.2 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from pandas) (2.9.0.post0)

Requirement already satisfied: tzdata in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from pandas) (2026.2)

Requirement already satisfied: six>=1.5 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

Note: you may need to restart the kernel to use updated packages.

```
import pandas as pd
df= pd.read_csv('diabetes missing data.csv')
```

```
df.head()
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	\
0	6	148.0	72.0	35.0	NaN	33.6	
1	1	85.0	66.0	29.0	NaN	26.6	
2	8	183.0	64.0	NaN	NaN	23.3	
3	1	89.0	66.0	23.0	94.0	28.1	
4	0	137.0	40.0	35.0	168.0	43.1	

	Diabetes_Pedigree	Age	Class
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

```
df.isnull().sum()
```

```
Pregnant      0
Glucose        5
Diastolic_BP  35
Skin_Fold     227
Serum_Insulin 374
BMI           11
Diabetes_Pedigree  0
Age           0
Class         0
dtype: int64
```

```
df.head()
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	\
0	6	148.0	72.0	35.0	NaN	33.6	
1	1	85.0	66.0	29.0	NaN	26.6	
2	8	183.0	64.0	NaN	NaN	23.3	
3	1	89.0	66.0	23.0	94.0	28.1	
4	0	137.0	40.0	35.0	168.0	43.1	

	Diabetes_Pedigree	Age	Class
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

```
import pandas as pd
df= pd.read_csv('diabetes missing data.csv')
df.isnull().sum()
```

```

Pregnant          0
Glucose           5
Diastolic_BP     35
Skin_Fold        227
Serum_Insulin    374
BMI              11
Diabetes_Pedigree 0
Age              0
Class            0
dtype: int64

len(df)

missing_precentage=df.isnull().mean()*100

columns_with_missing_30 = missing_percentage[missing_percentage > 30]
print('columns with more than 30% missing values:')
print(columns_with_missing_30)

print('columns with more than 30% missing values')
print(columns_with_missing_30)

columns_with_missing_30 = missing_percentage[missing_percentage > 30]
missing_percentage= df.isnull().mean()*100
columns_with_missing_30 = missing_percentage[missing_percentage > 30]
print('columns with more than 30% missing values')
print(columns_with_missing_30)

missing_percentage= df.isnull().mean()*100
threshold = 30

columns_to_drop = missing_percentage[missing_percentage >
threshold].index

df_cleaned = df.drop(columns=columns_to_drop)

print(f"columns dropped: {list(columns_to_drop)}")
print(f"remaining columns:{df_cleaned.columns}")

import pandas as pd
import numpy as np

data= pd.DataFrame({
    'Skin_Fold':[35.0, 29.0, None, 23.0, 35.0]
})
data['Skin_Fold'] = data['Skin_Fold'].fillna(data['Skin_Fold'].mean())
print(data)

```

	Skin_Fold
0	35.0
1	29.0
2	30.5
3	23.0
4	35.0

```
import numpy as np
import pandas as pd
```

```
df=pd.read_csv("diabetes missing data.csv")
df.head()
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	\
0	6	148.0	72.0	35.0	NaN	33.6	
1	1	85.0	66.0	29.0	NaN	26.6	
2	8	183.0	64.0	NaN	NaN	23.3	
3	1	89.0	66.0	23.0	94.0	28.1	
4	0	137.0	40.0	35.0	168.0	43.1	

	Diabetes_Pedigree	Age	Class
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

```
data=df.copy()
```

```
def detect_outliers_iqr(column):
```

```
    q1= column.quotile(0.25)
```

```
    q3= column.quotile(0.27)
```

```
    iqr = q3-q1
```

```
    lower_boun= q1 - 1.5 * iqr
```

```
    upper_bound= q3 + 1.5 * iqr
```

```
    return (column < lower_bound) | (column > upper_bound)
```

```
    for col in data.columns:
```

```
        data[f'{col}_outlier'] = detect_outliers_iqr(data[col])
```

```
        outliers_only = data[data.filter(like =
'_outlier').any(axis=1)]
```

```
        outliers_only
```

```
data=df.copy()
```

```
def detect_outliers_iqr(column):
```

```
    q1= column.quotile(0.25)
```

```

q3= column.quatile(0.27)

iqr = q3-q1
lower_boun= q1 - 1.5 * iqr
upper_bound= q3 + 1.5 * iqr

return (column < lower_bound) | (column > upper_bound)

for col in data.columns:
    data[f'{col}_outlier'] = detect_outliers_iqr(data[col])

    outliers_only = data[data.filter(like =
'_outlier').any(axis=1)]

    outliers_only

df=pd.read_csv("diabetes missing data.csv")
df.head()

```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	\
0	6	148.0	72.0	35.0	NaN	33.6	
1	1	85.0	66.0	29.0	NaN	26.6	
2	8	183.0	64.0	NaN	NaN	23.3	
3	1	89.0	66.0	23.0	94.0	28.1	
4	0	137.0	40.0	35.0	168.0	43.1	

	Diabetes_Pedigree	Age	Class
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

```

data=df.copy()
def detect_outliers_iqr(column):

    Q1= column.quantile(0.25)
    Q3= column.quantile(0.75)

    iqr = Q3-Q1
    lower_bound= Q1 - 1.5 * iqr
    upper_bound= Q3 + 1.5 * iqr

    return (column < lower_bound) | (column > upper_bound)

for col in data.columns:
    data[f'{col}_outlier'] = detect_outliers_iqr(data[col])

outliers_only = data[data.filter(like='_outlier').any(axis=1)]
outliers_only

```

BMI \	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	
4	0	137.0	40.0	35.0	168.0	43.1
8	2	197.0	70.0	45.0	543.0	30.5
12	10	139.0	80.0	NaN	NaN	27.1
13	1	189.0	60.0	23.0	846.0	30.1
18	1	103.0	30.0	38.0	83.0	43.3
..	...	...	...	...	...	...
691	13	158.0	114.0	NaN	NaN	42.3
695	7	142.0	90.0	24.0	480.0	30.4
710	3	158.0	64.0	13.0	387.0	31.2
715	7	187.0	50.0	33.0	392.0	33.9
753	0	181.0	88.0	44.0	510.0	43.3
\	Diabetes_Pedigree	Age	Class	Pregnant_outlier	Glucose_outlier	
4	2.288	33	1	False	False	
8	0.158	53	1	False	False	
12	1.441	57	0	False	False	
13	0.398	59	1	False	False	
18	0.183	33	0	False	False	
..	...	...	...	...	...	...
691	0.257	44	1	False	False	
695	0.128	43	1	False	False	
710	0.295	24	0	False	False	
715	0.826	34	1	False	False	
753	0.222	26	1	False	False	
4	Diastolic_BP_outlier	Skin_Fold_outlier	Serum_Insulin_outlier	\		
	False	False	False	False		

8	False	False	True
12	False	False	False
13	False	False	True
18	True	False	False
..	...	...	...
691	True	False	False
695	False	False	True
710	False	False	True
715	False	False	True
753	False	False	True

	BMI_outlier	Diabetes_Pedigree_outlier	Age_outlier
Class_outlier			
4	False	True	False
False			
8	False	False	False
False			
12	False	True	False
False			
13	False	False	False
False			
18	False	False	False
False			
..	...	...	...
.			
691	False	False	False
False			
695	False	False	False
False			
710	False	False	False
False			
715	False	False	False
False			
753	False	False	False
False			

[84 rows x 18 columns]

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
pip install matplotlib seaborn
```

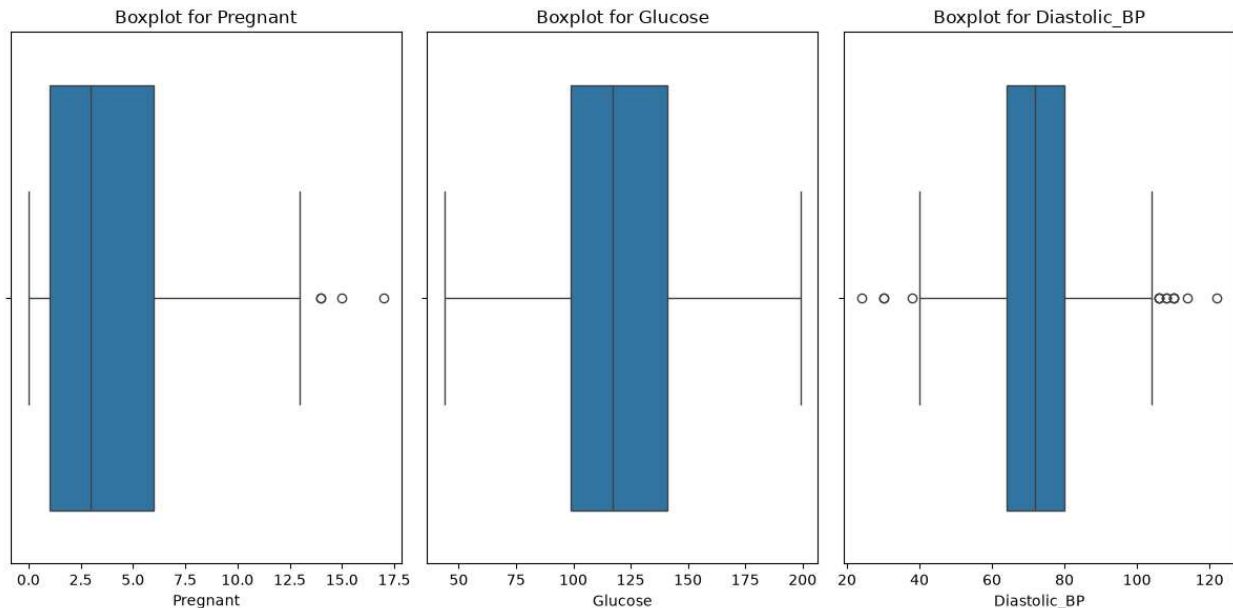
```
Requirement already satisfied: matplotlib in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (3.11.0)
Requirement already satisfied: seaborn in .\AppData\Local\anaconda3\
envs\code1\Lib\site-packages (0.13.2)
Requirement already satisfied: contourpy>=1.0.1 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cycler>=0.10 in .\AppData\Local\
```

```
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (4.63.0)
Requirement already satisfied: kiwisolver>=1.3.1 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (1.5.0)
Requirement already satisfied: numpy>=1.25 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (2.5.0)
Requirement already satisfied: packaging>=20.0 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (26.0)
Requirement already satisfied: pillow>=9 in .\AppData\Local\anaconda3\
envs\code1\Lib\site-packages (from matplotlib) (12.3.0)
Requirement already satisfied: pyparsing>=3 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from matplotlib) (3.3.2)
Requirement already satisfied: python-dateutil>=2.7 in .\AppData\
Local\anaconda3\envs\code1\Lib\site-packages (from matplotlib)
(2.9.0.post0)
Requirement already satisfied: pandas>=1.2 in .\AppData\Local\
anaconda3\envs\code1\Lib\site-packages (from seaborn) (3.0.3)
Requirement already satisfied: tzdata in .\AppData\Local\anaconda3\
envs\code1\Lib\site-packages (from pandas>=1.2->seaborn) (2026.2)
Requirement already satisfied: six>=1.5 in .\AppData\Local\anaconda3\
envs\code1\Lib\site-packages (from python-dateutil>=2.7->matplotlib)
(1.17.0)
```

Note: you may need to restart the kernel to use updated packages.

```
plt.figure(figsize=(12, 6))

for i, col in enumerate(df.columns[:3]):
    plt.subplot(1, 3, i+1)
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot for {col}")
plt.tight_layout()
plt.show()
```



```
pip install "nbconvert[webpdf]"
```

Requirement already satisfied: nbconvert[webpdf] in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (7.17.0)

Requirement already satisfied: beautifulsoup4 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (4.15.0)

Requirement already satisfied: bleach!=5.0.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from bleach[css]!=5.0.0->nbconvert[webpdf]) (6.3.0)

Requirement already satisfied: defusedxml in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (0.7.1)

Requirement already satisfied: jinja2>=3.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (3.1.6)

Requirement already satisfied: jupyter-core>=4.7 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (5.9.1)

Requirement already satisfied: jupyterlab-pygments in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (0.3.0)

Requirement already satisfied: markupsafe>=2.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (3.0.2)

Requirement already satisfied: mistune<4,>=2.0.3 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (3.2.1)

Requirement already satisfied: nbclient>=0.5.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (0.10.4)

Requirement already satisfied: nbformat>=5.7 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (5.10.4)

Requirement already satisfied: packaging in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (26.0)

Requirement already satisfied: pandocfilters>=1.4.1 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (1.5.1)

Requirement already satisfied: pygments>=2.4.1 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (2.20.0)

Requirement already satisfied: traitlets>=5.1 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbconvert[webpdf]) (5.15.0)

Collecting playwright (from nbconvert[webpdf])

 Downloading playwright-1.61.0-py3-none-win_amd64.whl.metadata (3.3 kB)

Requirement already satisfied: webencodings in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from bleach!=5.0.0->bleach[css]!=5.0.0->nbconvert[webpdf]) (0.5.1)

Collecting tinycss2<1.5,>=1.1.0 (from bleach[css]!=5.0.0->nbconvert[webpdf])

 Downloading tinycss2-1.4.0-py3-none-any.whl.metadata (3.0 kB)

Requirement already satisfied: platformdirs>=2.5 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from jupyter-core>=4.7->nbconvert[webpdf]) (4.9.4)

Requirement already satisfied: jupyter-client>=6.1.12 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbclient>=0.5.0->nbconvert[webpdf]) (8.9.1)

Requirement already satisfied: python-dateutil>=2.8.2 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (2.9.0.post0)

Requirement already satisfied: pyzmq>=25.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (27.1.0)

Requirement already satisfied: tornado>=6.4.1 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (6.5.5)

Requirement already satisfied: typing-extensions>=4.13.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (4.15.0)

Requirement already satisfied: fastjsonschema>=2.15 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbformat>=5.7->nbconvert[webpdf]) (2.21.2)

Requirement already satisfied: jsonschema>=2.6 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from nbformat>=5.7->nbconvert[webpdf]) (4.25.1)

Requirement already satisfied: attrs>=22.2.0 in .\AppData\Local\anaconda3\envs\code1\Lib\site-packages (from jsonschema>=2.6-

[illegible]

```

Downloading greenlet-3.5.3-cp313-cp313-win_amd64.whl (239 kB)
Downloading pyee-13.0.1-py3-none-any.whl (15 kB)
Installing collected packages: tinycss2, pyee, greenlet, playwright

```

```
Attempting uninstall: tinycss2
```

```
Found existing installation: tinycss2 1.5.1
```

Uninstalling tinycss2-1.5.1:

[illegible]

[illegible]

[illegible]

Downloading Chrome Headless Shell 149.0.7827.55 (playwright chromium-headless-shell v1228) from
<https://cdn.playwright.dev/builds/cft/149.0.7827.55/win64/chrome-headless-shell-win64.zip>

```
|
| 0% of 113.6 MiB
|â- â- â- â- â- â- â-
| 10% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
| 20% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- | 30% of
113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â-
| 40% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
| 50% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- | 60% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- â- | 70% of 113.6
MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
|
80% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â-
| 90% of 113.6 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- | 100% of 113.6 MiB
```

Chrome Headless Shell 149.0.7827.55 (playwright chromium-headless-shell v1228) downloaded to C:\Users\aalhamdi\AppData\Local\ms-playwright\chromium_headless_shell-1228

Downloading Winldd (playwright winldd v1007) from
<https://cdn.playwright.dev/dbazure/download/playwright/builds/winldd/1007/winldd-win64.zip>

```
|â- â- â- â- â- â-
| 12% of 0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â-
| 25% of 0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
```

```
â- | 38% of
0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
| 50% of 0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- | 63% of 0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- | 76% of 0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
|
89% of 0.1 MiB
|â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â- â-
â- â- â- â- â- â- â- â- â- | 100% of 0.1 MiB
Winldd (playwright winldd v1007) downloaded to C:\Users\aalhamdi\
AppData\Local\ms-playwright\winldd-1007
```