# Overview

## Dataset statistics

Number of observations  Missing cells  Missing cells (%)  Duplicate rows  Duplicate rows (%)  Total size in memory  Average record size in memory  Average types  Numeric  150  0.0%  0.0%  0.0%  0.0%  0.7%  14		
Missing cells  Missing cells (%)  Duplicate rows  1  Duplicate rows (%)  Total size in memory  4  Average record size in memory  Auriable types  Numeric  0  0.0%  0.7%  0.7%  4	Number of variables	5
Missing cells (%)  Duplicate rows  Duplicate rows (%)  Total size in memory  Average record size in memory  Average record size in memory  Auniable types  Numeric  0.0%  0.7%  6.0 KiB  40.9 B	Number of observations	150
Duplicate rows (%)  Duplicate rows (%)  Total size in memory  Average record size in memory  /ariable types  Numeric  1  0.7%  40.9 B  4	Missing cells	0
Duplicate rows (%)  Total size in memory  Average record size in memory  /ariable types  Numeric  0.7%  6.0 KiB  40.9 B	Missing cells (%)	0.0%
Total size in memory 6.0 KiB  Average record size in memory 40.9 B  /ariable types  Numeric 4	Duplicate rows	1
Average record size in memory 40.9 B  /ariable types  Numeric  4	Duplicate rows (%)	0.7%
/ariable types Numeric 4	Total size in memory	6.0 KiB
Numeric 4	Average record size in memory	40.9 B
	Variable types	
Categorical 1	Numeric	4
	Categorical	1

## Alerts

Dataset has 1 (0.7%) duplicate rows	Duplicates
sepal_length is highly correlated with petal_length and 1 other fields (petal_length, petal_width)	High correlation
petal_length is highly correlated with sepal_length and 1 other fields (sepal_length, petal_width)	High correlation
petal_width is highly correlated with sepal_length and 1 other fields (sepal_length, petal_length)	High correlation
sepal_length is highly correlated with petal_length and 1 other fields (petal_length, petal_width)	High correlation
petal_length is highly correlated with sepal_length and 1 other fields (sepal_length, petal_width)	High correlation
petal_width is highly correlated with sepal_length and 1 other fields (sepal_length, petal_length)	High correlation
sepal_length is highly correlated with petal_length and 1 other fields (petal_length, petal_width)	High correlation
petal_length is highly correlated with sepal_length and 1 other fields (sepal_length, petal_width)	High correlation
petal_width is highly correlated with sepal_length and 1 other fields (sepal_length, petal_length)	High correlation
sepal_length is highly correlated with sepal_width and 3 other fields (sepal_width, petal_length, petal_width, species)	High correlation
sepal_width is highly correlated with sepal_length and 3 other fields (sepal_length, petal_length, petal_width, species)	High correlation
petal_length is highly correlated with sepal_length and 3 other fields (sepal_length, sepal_width, petal_width, species)	High correlation
petal_width is highly correlated with sepal_length and 3 other fields (sepal_length, sepal_width, petal_length, species)	High correlation
species is highly correlated with sepal_length and 3 other fields (sepal_length, sepal_width, petal_length, petal_width)	High correlation
species is uniformly distributed	Uniform

## Reproduction

Analysis started	2022-09-25 10:40:18.942690
Analysis finished	2022-09-25 10:40:25.282747
Duration	6.34 seconds
Software version	pandas-profiling v3.1.0 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf- 8,%7B%22title%22%3A%20%22Pandas%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%2

## Variables

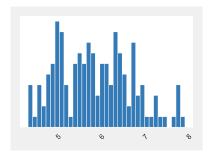
 $\begin{array}{l} sepal\_length \\ Real \ number \ (\mathbb{R}_{\geq 0}) \end{array}$ 

CORRELATION (This variable has a high correlation with 2 fields; petal length, petal width)
HIGH CORRELATION (This variable has a high correlation with 2 fields; petal length, petal width)

HIGH CORRELATION (This variable has a high correlation with 2 fields: petal length, petal width)

HIGH CORRELATION (This variable has a high correlation with 4 fields: sepal width, petal length, petal width, species)

Distinct	35	
Distinct (%)	23.3%	
Missing	0	
Missing (%)	0.0%	
Infinite	0	
Infinite (%)	0.0%	
Mean	5.843333333	
Minimum		4.3
Maximum		7.9
Zeros		0
Zeros (%)		0.0%
Negative		0
Negative (%)		0.0%
Memory size		1.3 KiB



#### Quantile statistics

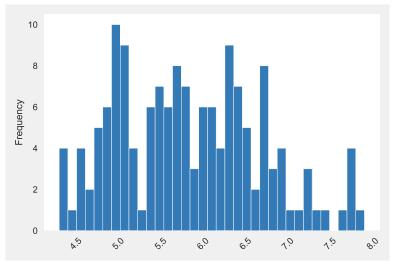
Minimum	4.3
5-th percentile	4.6
Q1	5.1
median	5.8
Q3	6.4
95-th percentile	7.255
Maximum	7.9
Range	3.6
Interquartile range (IQR)	1.3

#### Descriptive statistics

Standard deviation	0.828066128

HIGH

Coefficient of variation (CV)	0.1417112598
Kurtosis	-0.5520640413
Mean	5.843333333
Median Absolute Deviation (MAD)	0.7
Skewness	0.3149109566
Sum	876.5
Variance	0.6856935123
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=35)

Value	Count	Frequency (%)
5	10	6.7%
5.1	9	6.0%
6.3	9	6.0%
5.7	8	5.3%
6.7	8	5.3%
5.8	7	4.7%
5.5	7	4.7%
6.4	7	4.7%
4.9	6	4.0%
5.4	6	4.0%
Other values (25)	73	48.7%

Value	Count	Frequency (%)
4.3	1	0.7%
4.4	3	2.0%
4.5	1	0.7%
4.6	4	2.7%
4.7	2	1.3%
4.8	5	3.3%
4.9	6	4.0%
5	10	6.7%
5.1	9	6.0%
5.2	4	2.7%

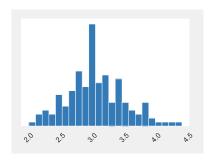
Value	Count	Frequency (%)
7.9	1	0.7%
7.7	4	2.7%
7.6	1	0.7%
7.4	1	0.7%
7.3	1	0.7%
7.2	3	2.0%
7.1	1	0.7%
7	1	0.7%
6.9	4	2.7%
6.8	3	2.0%

## sepal\_width

Real number (R<sub>≥0</sub>)

I	ΙG

ORRELATION (This variable has a high correlation with 4 fields: sepal length, petal length, petal width, species)		
Distinct	23	
Distinct (%)	15.3%	
Missing	0	
Missing (%)	0.0%	
Infinite	0	
Infinite (%)	0.0%	
Mean	3.057333333	
Minimum	2	
Maximum	4.4	
Zeros	0	
Zeros (%)	0.0%	
Negative	0	
Negative (%)	0.0%	



1.3 KiB

## Quantile statistics

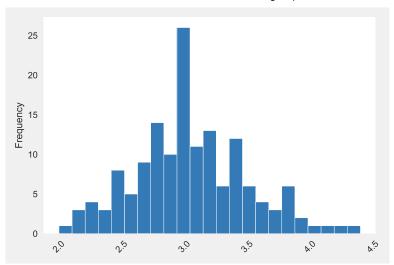
Memory size

Minimum	2
5-th percentile	2.345
Q1	2.8
median	3
Q3	3.3
95-th percentile	3.8
Maximum	4.4
Range	2.4
Interquartile range (IQR)	0.5

## Descriptive statistics

Standard deviation	0.4358662849
Coefficient of variation (CV)	0.1425642014
Kurtosis	0.2282490425
Mean	3.057333333
Median Absolute Deviation (MAD)	0.3
Skewness	0.3189656647

Sum	458.6
Variance	0.1899794183
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=23)

Value	Count	Frequency (%)
3	26	17.3%
2.8	14	9.3%
3.2	13	8.7%
3.4	12	8.0%
3.1	11	7.3%
2.9	10	6.7%
2.7	9	6.0%
2.5	8	5.3%
3.5	6	4.0%
3.3	6	4.0%
Other values (13)	35	23.3%

Value	Count	Frequency (%)
2	1	0.7%
2.2	3	2.0%
2.3	4	2.7%
2.4	3	2.0%
2.5	8	5.3%
2.6	5	3.3%
2.7	9	6.0%
2.8	14	9.3%
2.9	10	6.7%
3	26	17.3%

Value	Count	Frequenc	су (%)
4.4	1	0	0.7%
4.2	1	0	0.7%
4.1	1	0	0.7%
4	1	0	0.7%
3.9	2	1	1.3%
3.8	6	4.0%	
3.7	3	2	2.0%
3.6	4	2	2.7%
3.5	6	4.0%	
3.4	12	8.0%	,

#### petal\_length

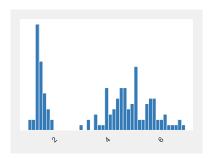
Real number ( $\mathbb{R}_{\geq 0}$ )

HIGH

CORRELATION (This variable has a high correlation with 2 fields: sepal length, netal width)
HIGH CORRELATION (This variable has a high correlation with 2 fields: sepal length, netal width)
HIGH CORRELATION (This variable has a high correlation with 2 fields: sepal length, netal width)

HIGH CORRELATION (This variable has a high correlation with 4 fields: sepal length, sepal width, petal width, species)

Distinct	43
Distinct (%)	28.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.758
Minimum	1
Maximum	6.9
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 KiB



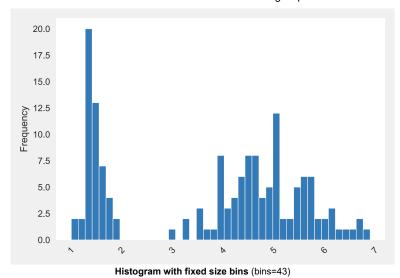
### Quantile statistics

Minimum	1
5-th percentile	1.3
Q1	1.6
median	4.35
Q3	5.1
95-th percentile	6.1
Maximum	6.9
Range	5.9
Interquartile range (IQR)	3.5

#### Descriptive statistics

Standard deviation	1.765298233
Coefficient of variation (CV)	0.4697440748
Kurtosis	-1.402103416
Mean	3.758

Median Absolute Deviation (MAD)	1.25
Skewness	-0.2748841798
Sum	563.7
Variance	3.116277852
Monotonicity	Not monotonic



file:///C:/Users/Lenovo/OneDrive/Documents/GitHub/Analytics Portfolio/Exploratory-Data-Analysis/pandas\_profiling\_iris.html

Value	Count	Frequency (%)
1.4	13	8.7%
1.5	13	8.7%
5.1	8	5.3%
4.5	8	5.3%
1.6	7	4.7%
1.3	7	4.7%
5.6	6	4.0%
4.7	5	3.3%
4.9	5	3.3%
4	5	3.3%
Other values (33)	73	48.7%

Value	Count	Frequency (%)
1	1	0.7%
1.1	1	0.7%
1.2	2	1.3%
1.3	7	4.7%
1.4	13	8.7%
1.5	13	8.7%
1.6	7	4.7%
1.7	4	2.7%
1.9	2	1.3%
3	1	0.7%

Value	Count	Frequency (%)
6.9	1	0.7%
6.7	2	1.3%
6.6	1	0.7%
6.4	1	0.7%
6.3	1	0.7%
6.1	3	2.0%
6	2	1.3%
5.9	2	1.3%
5.8	3	2.0%
5.7	3	2.0%

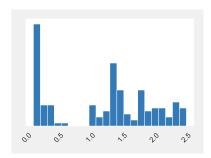
#### petal\_width

Real number (ℝ<sub>≥0</sub>)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: sepal length, petal length)
HIGH CORRELATION (This variable has a high correlation with 2 fields: sepal length, petal length)
HIGH CORRELATION (This variable has a high correlation with 2 fields: sepal length, petal length)
HIGH CORRELATION (This variable has a high correlation with 4 fields: sepal length, sepal width, petal length, species)

Distinct	22
Distinct (%)	14.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.199333333
Minimum	0.1
Maximum	2.5
Zeros	0
Zeros (%)	0.0%
Zeros (%)	0.0%



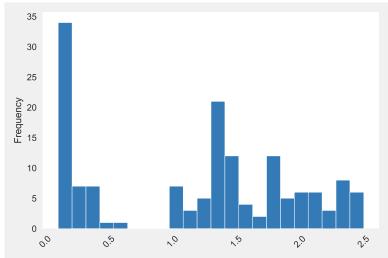
### Quantile statistics

Minimum	0.1
5-th percentile	0.2
Q1	0.3
median	1.3
Q3	1.8
95-th percentile	2.3
Maximum	2.5
Range	2.4
Interquartile range (IQR)	1.5

#### Descriptive statistics

Standard deviation	0.762237669
Coefficient of variation (CV)	0.6355511414
Kurtosis	-1.340603997
Mean	1.199333333

Median Absolute Deviation (MAD)	0.7
Skewness	-0.1029667476
Sum	179.9
Variance	0.581006264
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=22)

Value	Count	Frequency (%)
0.2	29	19.3%
1.3	13	8.7%
1.8	12	8.0%
1.5	12	8.0%
1.4	8	5.3%
2.3	8	5.3%
1	7	4.7%
0.4	7	4.7%
0.3	7	4.7%
2.1	6	4.0%
Other values (12)	41	27.3%

Value	Count	Frequency (%)
0.1	5	3.3%
0.2	29	19.3%
0.3	7	4.7%
0.4	7	4.7%
0.5	1	0.7%
0.6	1	0.7%
1	7	4.7%
1.1	3	2.0%
1.2	5	3.3%
1.3	13	8.7%

Value	Count	Frequency (%)
2.5	3	2.0%
2.4	3	2.0%
2.3	8	5.3%
2.2	3	2.0%
2.1	6	4.0%
2	6	4.0%
1.9	5	3.3%
1.8	12	8.0%
1.7	2	1.3%
1.6	4	2.7%

## species

CORRELATION (This variable has a high correlation with 4 fields: sepal length, sepal width, petal width)
UNIFORM

Distinct

Distinct (%)

Missing

0

Missing (%)

Memory size

1.3 KiB

### Length

Max length	10
Median length	9
Mean length	8.333333333
Min length	6

### Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

### Unique

Unique	0	?
Unique (%)	0.0%	

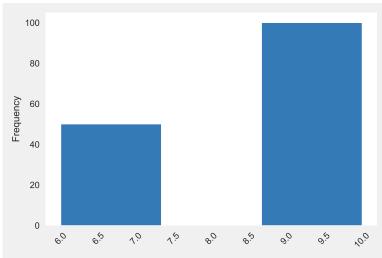
### Sample

1st row	setosa
2nd row	setosa
3rd row	setosa
4th row	setosa
5th row	setosa

### Common Values

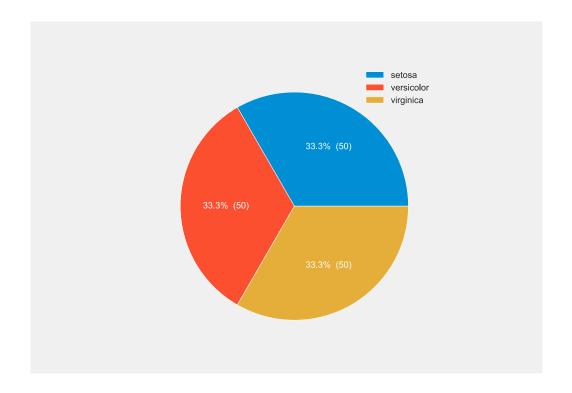
Value	Count	Frequency (%)
setosa	50	33.3%
versicolor	50	33.3%
virginica	50	33.3%

### Length



Histogram of lengths of the category

### Pie chart



Value	Count	Frequency (%)
setosa	50	33.3%
versicolor	50	33.3%
virginica	50	33.3%

### Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

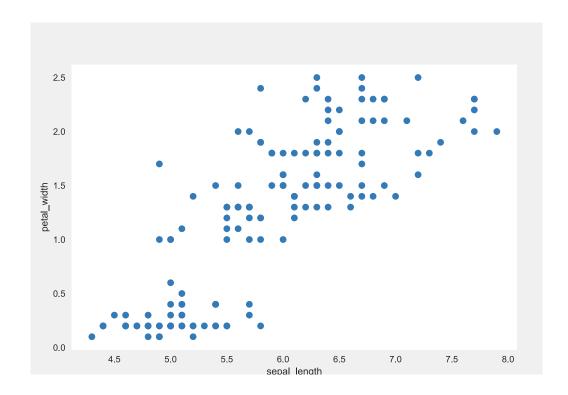
Most occurring blocks

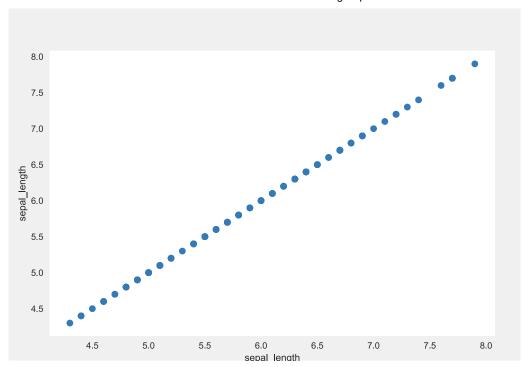
Value Count Frequency (%)

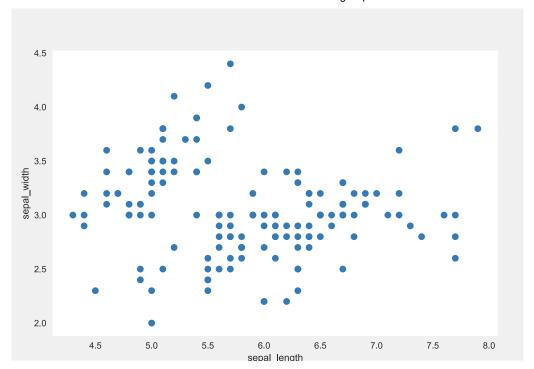
No values found.

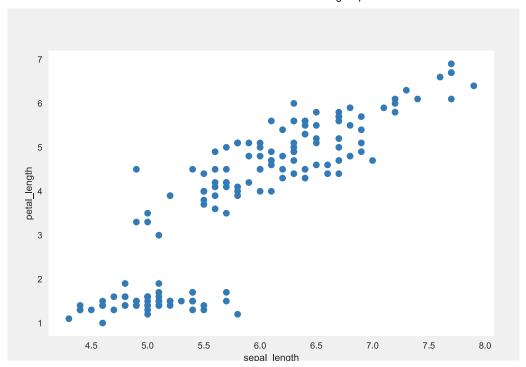
Most frequent character per block

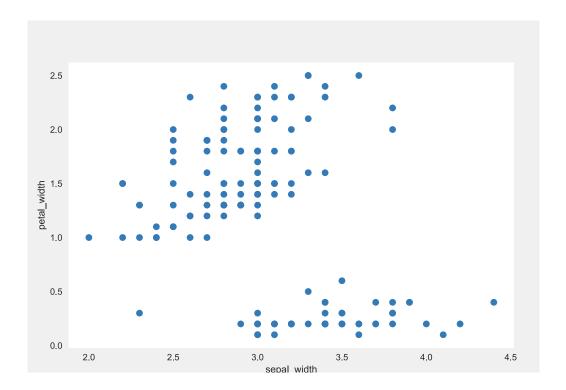
## Interactions

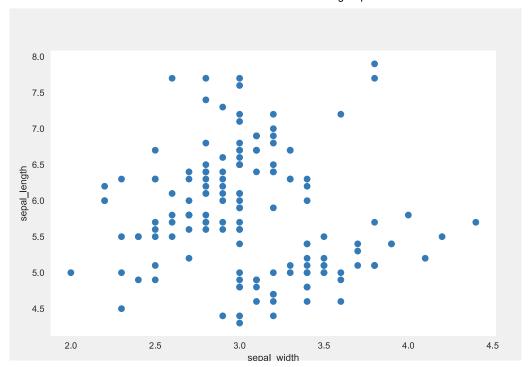


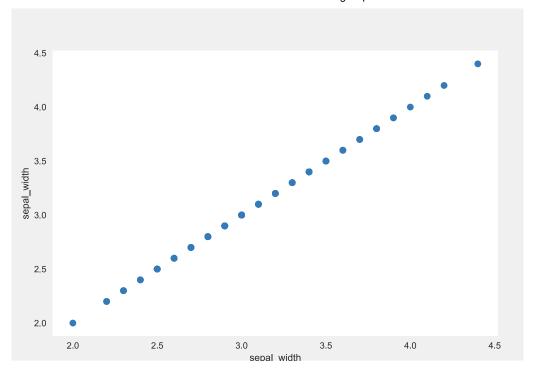


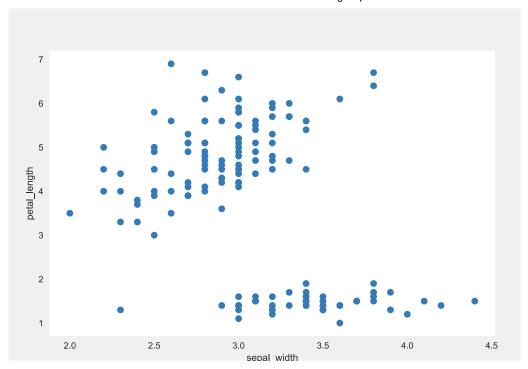


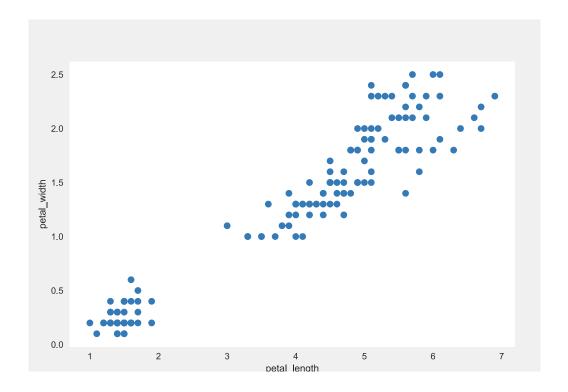


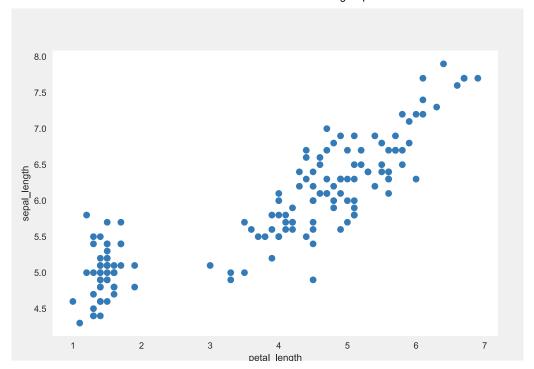


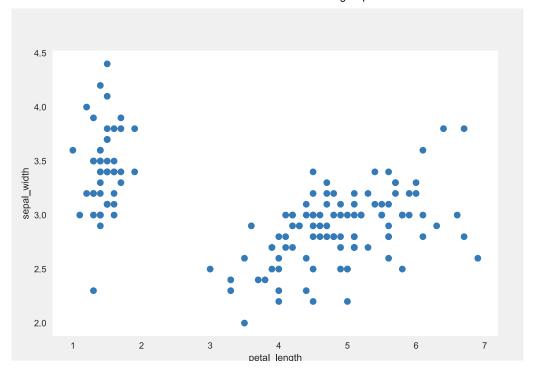


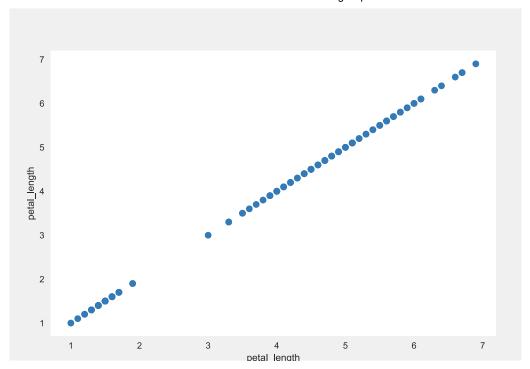


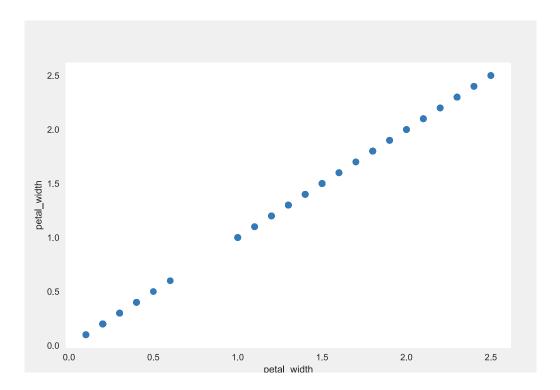


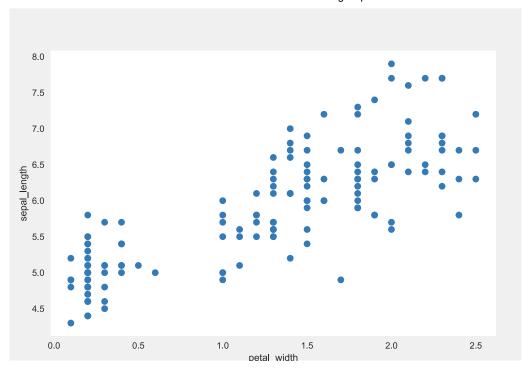


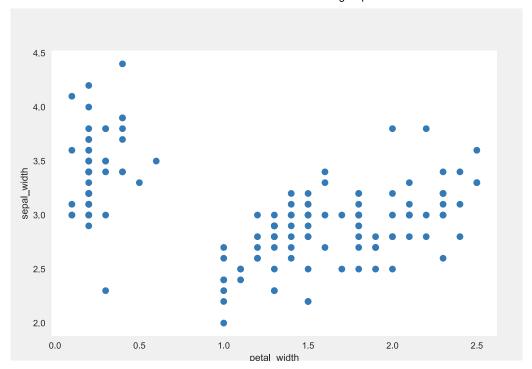


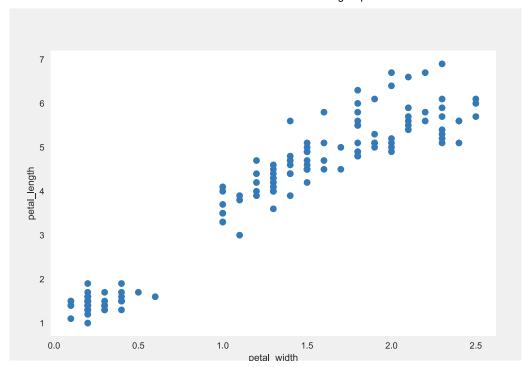




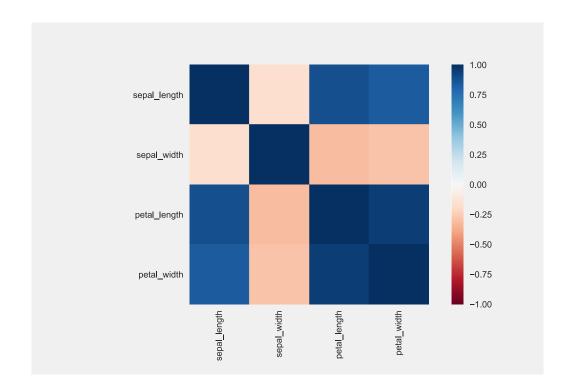


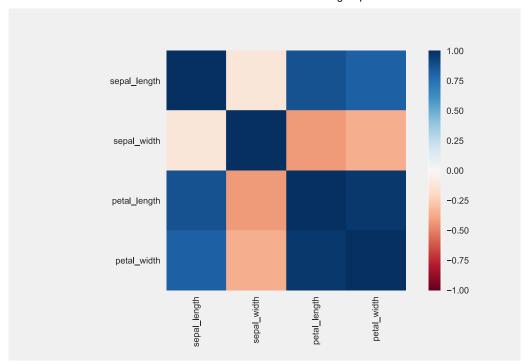


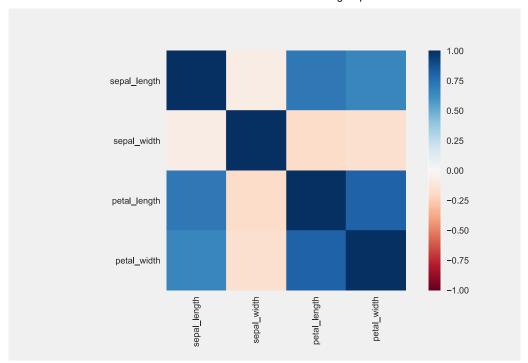


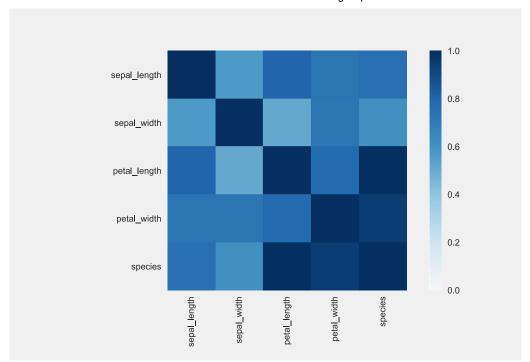


### Correlations

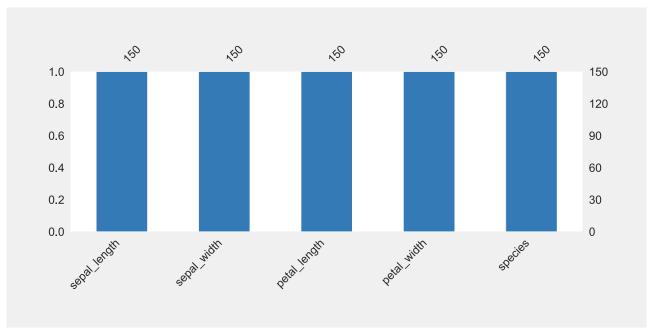




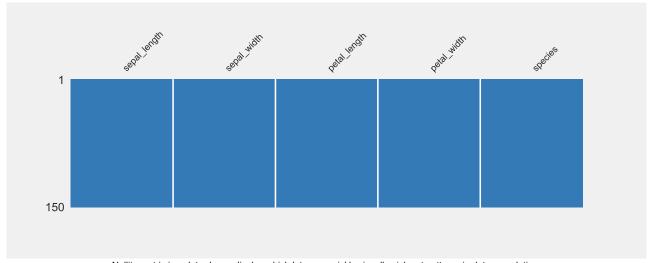




# Missing values



A simple visualization of nullity by column.



Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.

# Sample

### First rows

	sepal_length	sepal_width	petal_length	petal_width	species
C	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
					<b>•</b>

### Last rows

	sepal_length	sepal_width	petal_length	petal_width	species
140	6.7	3.1	5.6	2.4	virginica
141	6.9	3.1	5.1	2.3	virginica
142	5.8	2.7	5.1	1.9	virginica
143	6.8	3.2	5.9	2.3	virginica
144	6.7	3.3	5.7	2.5	virginica
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

# **Duplicate rows**

## Most frequently occurring

	sepal_length	sepal_width	petal_length	petal_width	species	# duplicates
(	0 5.8	2.7	5.1	1.9	virginica	2
4						<b>•</b>

Report generated with pandas-profiling (https://github.com/pandas-profiling/pandas-profiling).