pca crabs

August 15, 2023

1 Import Libraries

```
[]: import pandas as pd
     pd.set_option('display.precision',3)
     import io
     from google.colab import files
[]: uploaded = files.upload()
    <IPython.core.display.HTML object>
    Saving 3 - crabs.csv to 3 - crabs.csv
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs data
[]:
                          FL
                                 RW
                                       CL
                                             CW
                                                   BD
                 index
         sp sex
                         8.1
                                6.7
                                     16.1
                                           19.0
                                                  7.0
          В
              М
                     1
     1
          В
                     2
                         8.8
                                7.7
                                           20.8
                                                  7.4
              М
                                     18.1
     2
          В
                     3
                         9.2
                                7.8
                                     19.0
                                           22.4
                                                  7.7
          В
                     4
                         9.6
     3
              М
                                7.9
                                     20.1
                                           23.1
                                                  8.2
          В
                     5
                         9.8
                                8.0 20.3
                                           23.0
                                                  8.2
              М
                        21.4
                              18.0 41.2
     195
         0
              F
                                          46.2
                                                 18.7
                    46
     196 0
              F
                    47
                        21.7
                              17.1 41.7 47.2
                                                 19.6
              F
                              17.2
                                    42.6 47.4
     197 0
                    48
                        21.9
                                                 19.5
                        22.5
                              17.2
                                    43.0 48.7
     198
         0
              F
                    49
                                                 19.8
                        23.1
                              20.2 46.2 52.5
     199
                    50
                                                 21.1
     [200 rows x 8 columns]
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.head()
                                         CW
[]:
                                   CL
                                              BD
       sp sex
               index
                       FL
                            RW
                                       19.0
       В
            Μ
                      8.1
                           6.7
                                 16.1
                                             7.0
                   1
       В
                   2
                      8.8
                           7.7
                                 18.1
                                       20.8
                                             7.4
     1
            Μ
       В
                      9.2
                                 19.0
                                       22.4
     2
            Μ
                   3
                           7.8
                                             7.7
     3 B
            Μ
                      9.6
                          7.9
                                20.1
                                       23.1
                                             8.2
```

```
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.tail()
[]:
                 index
                          FL
                                RW
                                      CL
                                            CW
                                                   BD
         sp sex
     195
         0
              F
                    46
                        21.4
                              18.0
                                    41.2
                                          46.2
                                                 18.7
              F
                        21.7
                              17.1
                                    41.7
                                          47.2
     196 0
                    47
                                                 19.6
     197 0
              F
                        21.9
                              17.2
                                   42.6 47.4 19.5
                    48
                        22.5
                             17.2 43.0 48.7
     198
              F
                    49
                                                19.8
     199
                    50
                        23.1
                             20.2 46.2 52.5
                                                21.1
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.shape
[]: (200, 8)
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.info()
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 200 entries, 0 to 199
    Data columns (total 8 columns):
         Column Non-Null Count Dtype
     0
         sp
                 200 non-null
                                  object
     1
         sex
                 200 non-null
                                  object
     2
         index
                 200 non-null
                                  int64
     3
         FL
                 200 non-null
                                  float64
     4
         R.W
                 200 non-null
                                  float64
     5
         CL
                                 float64
                 200 non-null
     6
         CW
                 200 non-null
                                  float64
     7
         BD
                 200 non-null
                                 float64
    dtypes: float64(5), int64(1), object(2)
    memory usage: 12.6+ KB
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.info
[]: <bound method DataFrame.info of
                                                                       CL
                                                                             CW
                                                                                   BD
                                         sp sex
                                                  index
                                                           FL
                                                                 RW
     0
          B
              М
                     1
                         8.1
                               6.7 16.1 19.0
                                                  7.0
     1
          В
                     2
                         8.8
                               7.7 18.1 20.8
                                                  7.4
              М
     2
          В
                     3
                         9.2
                               7.8 19.0 22.4
                                                  7.7
              Μ
     3
          В
              М
                     4
                         9.6
                               7.9 20.1
                                          23.1
                                                  8.2
                     5
     4
          В
              М
                         9.8
                               8.0 20.3 23.0
                                                  8.2
                              18.0 41.2 46.2
     195
         0
              F
                    46 21.4
     196
        0
              F
                    47
                        21.7
                              17.1 41.7 47.2
                                                19.6
```

```
197 0
         F
               48
                   21.9
                         17.2
                               42.6 47.4
                                           19.5
198
                   22.5
                         17.2
                              43.0 48.7
   0
         F
               49
                                           19.8
199
         F
               50
                   23.1
                         20.2
                               46.2
                                     52.5
                                           21.1
```

[200 rows x 8 columns]>

```
[]: crabs_data = pd.read_csv("3 - crabs.csv")
crabs_data.T
```

```
[]:
              0
                     1
                                   3
                                         4
                                                5
                                                       6
                                                              7
                                                                    8
                                                                                      190
                В
                       В
                              В
                                     В
                                            В
                                                  В
                                                         В
                                                                В
                                                                       В
                                                                             В
                                                                                        0
     sp
                                                                                 ...
                М
                       М
                              Μ
                                     М
                                           Μ
                                                  Μ
                                                         М
                                                                                        F
     sex
                                                                Μ
                                                                       М
                                                                             Μ
     index
                1
                       2
                              3
                                     4
                                            5
                                                  6
                                                         7
                                                                8
                                                                       9
                                                                            10
                                                                                       41
     FL
              8.1
                     8.8
                            9.2
                                  9.6
                                         9.8
                                               10.8
                                                      11.1
                                                            11.6
                                                                   11.8
                                                                          11.8
                                                                                    20.3
              6.7
                            7.8
                                  7.9
                                                              9.1
                                                                          10.5
     RW
                     7.7
                                         8.0
                                                9.0
                                                       9.9
                                                                    9.6
                                                                                    16.0
     CL
             16.1
                    18.1
                           19.0
                                 20.1
                                        20.3
                                               23.0
                                                      23.8
                                                            24.5
                                                                   24.2
                                                                          25.2
                                                                                    39.4
     CW
             19.0
                                        23.0
                    20.8
                          22.4
                                 23.1
                                               26.5
                                                      27.1
                                                             28.4
                                                                   27.8
                                                                          29.3
                                                                                    44.1
              7.0
                                                       9.8
                                                             10.4
     BD
                     7.4
                            7.7
                                  8.2
                                         8.2
                                                9.8
                                                                     9.7
                                                                          10.3 ...
                                                                                    18.0
              191
                     192
                            193
                                   194
                                         195
                                                196
                                                       197
                                                              198
                                                                     199
                       0
                              0
                0
                                     0
                                            0
                                                  0
                                                         0
                                                                0
                                                                       0
     sp
                       F
                F
                              F
                                     F
                                            F
                                                  F
                                                         F
                                                                F
                                                                       F
     sex
                             44
                                    45
     index
               42
                      43
                                           46
                                                 47
                                                        48
                                                               49
                                                                      50
     FL
             20.5
                    20.6
                          20.9
                                 21.3
                                        21.4
                                               21.7
                                                      21.9
                                                            22.5
                                                                   23.1
             17.5
                    17.5
                          16.5
                                 18.4
                                        18.0
                                               17.1
                                                      17.2
                                                            17.2
                                                                   20.2
     RW
     CL
             40.0
                    41.5
                          39.9
                                 43.8
                                        41.2 41.7
                                                      42.6
                                                            43.0
                                                                   46.2
     CW
             45.5
                    46.2
                          44.7
                                 48.4
                                        46.2 47.2
                                                     47.4
                                                            48.7
                                                                   52.5
                          17.5
     BD
             19.2
                    19.2
                                 20.0
                                        18.7
                                               19.6 19.5
                                                            19.8
                                                                   21.1
```

[8 rows x 200 columns]

```
[]:
         species
                            index
                                  Frontal Lobe Length Rear Width Carepace Length \
                      sex
     0
            BLUE
                     MALE
                                1
                                                     8.1
                                                                  6.7
                                                                                   16.1
            BLUE
                     MALE
                                2
                                                                  7.7
     1
                                                     8.8
                                                                                   18.1
     2
            BLUE
                     MALE
                                                     9.2
                                                                  7.8
                                                                                   19.0
                                3
                                                                  7.9
     3
            BLUE
                     MALE
                                4
                                                     9.6
                                                                                   20.1
                                5
                                                     9.8
     4
            BLUE
                     MALE
                                                                  8.0
                                                                                   20.3
     195
          ORANGE
                   FEMALE
                               46
                                                    21.4
                                                                 18.0
                                                                                   41.2
```

	100	OD A MAI			4.7			01 7		17 1		44 7	
	196 197	ORANGI ORANGI			47			21.7 21.9		17.1		41.7 42.6	
					48					17.2			
	198	ORANGI			49			22.5		17.2		43.0	
	199	ORANGI	E FEMA	LE	50			23.1		20.2		46.2	
		Carons	ace Wid	+h Ba	oduil or	agth							
	0	Carepa	19		dyrei	7.0							
	1		20			7.4							
	2		22			7.7							
	3		23			8.2							
	4		23	.0		8.2							
			•••		•••								
	195		46	.2	1	18.7							
	196		47	.2	1	19.6							
	197		47	.4	1	19.5							
	198		48	.7	1	19.8							
	199		52	.5	2	21.1							
	Γ200	rows 2	x 8 col	umnsl									
[]:	crab	s_data	.head()										
[]:	_	ecies		index	Fror	ntal Lo	_				epace Leng		
											10	4	
	0	BLUE	MALE	1			8.1	_	6.7	•	16	. 1	
	0 1	BLUE BLUE	MALE MALE	1 2			8.3 8.8		6.7 7.7		16 18		
								3		•		. 1	
	1	BLUE	MALE	2			8.8	3 2	7.7		18	.1	
	1 2	BLUE BLUE BLUE	MALE MALE	2			8.8 9.2	3 2 3	7.7 7.8	}	18 19	.1 .0 .1	
	1 2 3	BLUE BLUE BLUE	MALE MALE MALE	2 3 4			8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4	BLUE BLUE BLUE BLUE	MALE MALE MALE	2 3 4 5	/lengt	. h	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE	2 3 4 5	_	5h .0	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE	2 3 4 5 Body	7.		8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE Width 19.0 20.8	2 3 4 5 Body	7 . 7 .	. 0	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C 0 1 2	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE Width 19.0 20.8 22.4	2 3 4 5 Body	7. 7. 7.	. 0 . 4 . 7	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C C 0 1 2 3	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1	2 3 4 5 Body	7. 7. 7. 8.	.0 .4 .7	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C 0 1 2	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE Width 19.0 20.8 22.4	2 3 4 5 Body	7. 7. 7. 8.	. 0 . 4 . 7	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C C 0 1 2 3 4	BLUE BLUE BLUE BLUE arepace	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0	2 3 4 5 Body	7. 7. 7. 8.	.0 .4 .7	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C C 0 1 2 3 4	BLUE BLUE BLUE BLUE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0	2 3 4 5 Body	7. 7. 7. 8.	.0 .4 .7	8.8 9.2 9.6	3 2 3	7.7 7.8 7.9	}	18 19 20	.1 .0 .1	
	1 2 3 4 C C 0 1 2 3 4 crab	BLUE BLUE BLUE BLUE arepace	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail()	2 3 4 5 Body	7. 7. 7. 8.	.0 .4 .7 .2 .2	8.8 9.2 9.6	3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	7.7 7.8 7.9 8.0		18 19 20	.1 .0 .1 .3	\
[]:[1 2 3 4 C C 0 1 2 3 4 crab	BLUE BLUE BLUE BLUE arepace	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail()	2 3 4 5 Body	7. 7. 8. 8.	.0 .4 .7 .2 .2	8.8 9.2 9.6 9.8	3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	7.7 7.8 7.9 8.0		18 19 20 20	.1 .0 .1 .3	\
[]:[1 2 3 4 C C 0 1 2 3 4 Crab	BLUE BLUE BLUE arepace s_data species	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail()	2 3 4 5 Body	7. 7. 7. 8. 8.	.0 .4 .7 .2 .2	8.8 9.2 9.6 9.8	ength	7.7 7.8 7.9 8.0	didth	18 19 20 20	.1 .0 .1 .3	\
[]:[1 2 3 4 0 1 2 3 4 crab	BLUE BLUE BLUE BLUE arepace s_data species ORANGE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail() E FEMALE	2 3 4 5 Body	7. 7. 7. 8. 8. 8. andex 46 47	.0 .4 .7 .2 .2	8.8 9.2 9.6 9.8	ength 21.4 21.7	7.7 7.8 7.9 8.0	Tidth 18.0	18 19 20 20	.1 .0 .1 .3	\
[]:[1 2 3 4 C C 0 1 2 3 4 Crab 195 196 197	BLUE BLUE BLUE BLUE arepace s_data species ORANGE ORANGE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail() E FEMALE FEMALE	2 3 4 5 Body ex ir LE LE LE	7. 7. 8. 8. 8. andex 46 47 48	.0 .4 .7 .2 .2	8.8 9.2 9.6 9.8	ength 21.4 21.7 21.9	7.7 7.8 7.9 8.0	Tidth 18.0 17.1 17.2	18 19 20 20	Length 41.2 41.7 42.6	\
[]:[1 2 3 4 C C 0 1 2 3 4 Crab 195 196 197 198	BLUE BLUE BLUE BLUE arepace s_data species ORANGE ORANGE ORANGE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail() E FEMALE E FEMALE E FEMALE	2 3 4 5 Body ex ir LE LE LE	7. 7. 8. 8. 8. adex 46 47 48 49	.0 .4 .7 .2 .2	8.8 9.2 9.6 9.8	ength 21.4 21.7 21.9 22.5	7.7 7.8 7.9 8.0	Tidth 18.0 17.1 17.2 17.2	18 19 20 20	Length 41.2 41.7 42.6 43.0	\
[]:[1 2 3 4 C C 0 1 2 3 4 Crab 195 196 197	BLUE BLUE BLUE BLUE arepace s_data species ORANGE ORANGE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail() E FEMALE E FEMALE E FEMALE	2 3 4 5 Body ex ir LE LE LE	7. 7. 8. 8. 8. andex 46 47 48	.0 .4 .7 .2 .2	8.8 9.2 9.6 9.8	ength 21.4 21.7 21.9	7.7 7.8 7.9 8.0	Tidth 18.0 17.1 17.2	18 19 20 20	Length 41.2 41.7 42.6	\
[]:[1 2 3 4 C C 0 1 2 3 4 Crab 195 196 197 198	BLUE BLUE BLUE BLUE arepace arepace ORANGE ORANGE ORANGE ORANGE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail() E FEMALE FEMALE FEMALE	2 3 4 5 Body ex ir LE LE LE LE LE	7. 7. 8. 8. 8. andex 46 47 48 49 50	.0 .4 .7 .2 .2 Fronta	8.8 9.2 9.6 9.8	ength 21.4 21.7 21.9 22.5	7.7 7.8 7.9 8.0	Tidth 18.0 17.1 17.2 17.2	18 19 20 20	Length 41.2 41.7 42.6 43.0	\
[]:[1 2 3 4 C C 0 1 2 3 4 Crab 195 196 197 198	BLUE BLUE BLUE BLUE arepace arepace ORANGE ORANGE ORANGE ORANGE	MALE MALE MALE MALE Width 19.0 20.8 22.4 23.1 23.0 .tail() E FEMALE E FEMALE E FEMALE	2 3 4 5 Body ex ir LE	7. 7. 8. 8. andex 46 47 48 49 50 andyler	.0 .4 .7 .2 .2 Fronta	8.8 9.2 9.6 9.8	ength 21.4 21.7 21.9 22.5	7.7 7.8 7.9 8.0	Tidth 18.0 17.1 17.2 17.2	18 19 20 20	Length 41.2 41.7 42.6 43.0	\

```
197
                     47.4
                                  19.5
     198
                     48.7
                                  19.8
     199
                     52.5
                                  21.1
[]: crabs_data.describe(include='all')
[]:
             species
                                      Frontal Lobe Length
                                                             Rear Width
                       sex
                               index
                 200
                       200
                            200.000
                                                   200.000
                                                                200.000
     count
     unique
                   2
                         2
                                 NaN
                                                        NaN
                                                                     NaN
     top
                BLUE
                      MALE
                                 NaN
                                                        NaN
                                                                     NaN
                 100
                       100
                                 NaN
                                                        NaN
                                                                    NaN
     freq
     mean
                 NaN
                       NaN
                              25.500
                                                    15.583
                                                                 12.738
     std
                 NaN
                       NaN
                              14.467
                                                     3.495
                                                                  2.573
                 NaN
                       NaN
                                                     7.200
     min
                               1.000
                                                                  6.500
     25%
                       NaN
                              13.000
                 NaN
                                                    12.900
                                                                 11.000
     50%
                 NaN
                       NaN
                              25.500
                                                    15.550
                                                                 12.800
     75%
                 NaN
                       NaN
                              38.000
                                                    18.050
                                                                 14.300
                 NaN
                       NaN
                              50.000
                                                    23.100
                                                                 20.200
     max
             Carepace Length
                                Carepace Width
                                                 Bodylength
                      200.000
                                        200.000
                                                    200.000
     count
     unique
                          NaN
                                                         NaN
                                            NaN
     top
                          {\tt NaN}
                                            NaN
                                                         NaN
     freq
                          NaN
                                            NaN
                                                         NaN
     mean
                       32.105
                                         36.415
                                                      14.030
     std
                        7.119
                                          7.872
                                                       3.425
     min
                       14.700
                                         17.100
                                                       6.100
     25%
                       27.275
                                        31.500
                                                      11.400
     50%
                       32.100
                                        36.800
                                                      13.900
     75%
                       37.225
                                        42.000
                                                      16.600
                       47.600
                                         54.600
                                                     21.600
     max
[]: crabs_data.columns
[]: Index(['species', 'sex', 'index', 'Frontal Lobe Length', 'Rear Width',
             'Carepace Length', 'Carepace Width', 'Bodylength'],
           dtype='object')
     crabs_data.shape
[]: (200, 8)
[]: crabs_data['class'] = crabs_data.species + crabs_data.sex
     crabs_data['class'].value_counts()
```

196

47.2

19.6

[]: BLUEMALE 50
BLUEFEMALE 50
ORANGEMALE 50
ORANGEFEMALE 50

Name: class, dtype: int64