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
```

[ ]:

	sp	sex	index	FL	RW	CL	CW	BD
0	B	M	1	8.1	6.7	16.1	19.0	7.0
1	B	M	2	8.8	7.7	18.1	20.8	7.4
2	B	M	3	9.2	7.8	19.0	22.4	7.7
3	B	M	4	9.6	7.9	20.1	23.1	8.2
4	B	M	5	9.8	8.0	20.3	23.0	8.2
..	..	..	…	…	…	…	…	…
195	O	F	46	21.4	18.0	41.2	46.2	18.7
196	O	F	47	21.7	17.1	41.7	47.2	19.6
197	O	F	48	21.9	17.2	42.6	47.4	19.5
198	O	F	49	22.5	17.2	43.0	48.7	19.8
199	O	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.head()
```

[ ]:

	sp	sex	index	FL	RW	CL	CW	BD
0	B	M	1	8.1	6.7	16.1	19.0	7.0
1	B	M	2	8.8	7.7	18.1	20.8	7.4
2	B	M	3	9.2	7.8	19.0	22.4	7.7
3	B	M	4	9.6	7.9	20.1	23.1	8.2

```
4  B   M      5  9.8  8.0  20.3  23.0  8.2
```

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

[ ]:
```
    sp sex  index    FL    RW    CL    CW    BD
195  O   F     46  21.4  18.0  41.2  46.2  18.7
196  O   F     47  21.7  17.1  41.7  47.2  19.6
197  O   F     48  21.9  17.2  42.6  47.4  19.5
198  O   F     49  22.5  17.2  43.0  48.7  19.8
199  O   F     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   RW      200 non-null    float64
 5   CL      200 non-null    float64
 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     sp sex  index    FL    RW    CL    CW    BD
0    B   M      1   8.1   6.7  16.1  19.0   7.0
1    B   M      2   8.8   7.7  18.1  20.8   7.4
2    B   M      3   9.2   7.8  19.0  22.4   7.7
3    B   M      4   9.6   7.9  20.1  23.1   8.2
4    B   M      5   9.8   8.0  20.3  23.0   8.2
..  ..  ..    ...   ...   ...   ...   ...   ...
195  O   F     46  21.4  18.0  41.2  46.2  18.7
196  O   F     47  21.7  17.1  41.7  47.2  19.6
```

```
197  O   F     48  21.9  17.2  42.6  47.4  19.5
198  O   F     49  22.5  17.2  43.0  48.7  19.8
199  O   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     2     3     4     5     6     7     8     9   …   190  \
sp        B     B     B     B     B     B     B     B     B     B   …     O
sex       M     M     M     M     M     M     M     M     M     M   …     F
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
RW      6.7   7.7   7.8   7.9   8.0   9.0   9.9   9.1   9.6  10.5   …  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  20.8  22.4  23.1  23.0  26.5  27.1  28.4  27.8  29.3   …  44.1
BD      7.0   7.4   7.7   8.2   8.2   9.8   9.8  10.4   9.7  10.3   …  18.0

          191   192   193   194   195   196   197   198   199
sp          O     O     O     O     O     O     O     O     O
sex         F     F     F     F     F     F     F     F     F
index      42    43    44    45    46    47    48    49    50
FL       20.5  20.6  20.9  21.3  21.4  21.7  21.9  22.5  23.1
RW       17.5  17.5  16.5  18.4  18.0  17.1  17.2  17.2  20.2
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
BD       19.2  19.2  17.5  20.0  18.7  19.6  19.5  19.8  21.1

[8 rows x 200 columns]
```

[ ]:
```
crabs_data = pd.read_csv("3 - crabs.csv")
crabs_data = crabs_data.rename( columns = {'sp':'species','FL':'Frontal Lobe␣
 ↪Length','RW':'Rear Width','CL':'Carepace Length','CW':'Carepace Width','BD':
 ↪'Bodylength'})
crabs_data['species'] = crabs_data['species'].map({'B':'BLUE','O':'ORANGE'})
crabs_data['sex'] = crabs_data['sex'].map({'M':'MALE','F':'FEMALE'})
crabs_data
```

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

```
196  ORANGE  FEMALE     47                 21.7         17.1              41.7
197  ORANGE  FEMALE     48                 21.9         17.2              42.6
198  ORANGE  FEMALE     49                 22.5         17.2              43.0
199  ORANGE  FEMALE     50                 23.1         20.2              46.2

     Carepace Width  Bodylength
0              19.0         7.0
1              20.8         7.4
2              22.4         7.7
3              23.1         8.2
4              23.0         8.2
..              …           …
195            46.2         18.7
196            47.2         19.6
197            47.4         19.5
198            48.7         19.8
199            52.5         21.1

[200 rows x 8 columns]
```

[ ]:
```
crabs_data.head()
```

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

   Carepace Width  Bodylength
0            19.0         7.0
1            20.8         7.4
2            22.4         7.7
3            23.1         8.2
4            23.0         8.2
```

[ ]:
```
crabs_data.tail()
```

[ ]:
```
     species     sex  index  Frontal Lobe Length  Rear Width  Carepace Length  \
195  ORANGE  FEMALE     46                 21.4         18.0             41.2
196  ORANGE  FEMALE     47                 21.7         17.1             41.7
197  ORANGE  FEMALE     48                 21.9         17.2             42.6
198  ORANGE  FEMALE     49                 22.5         17.2             43.0
199  ORANGE  FEMALE     50                 23.1         20.2             46.2

     Carepace Width  Bodylength
195            46.2        18.7
```

```
196            47.2        19.6
197            47.4        19.5
198            48.7        19.8
199            52.5        21.1
```

[ ]:
```
crabs_data.describe(include='all')
```

[ ]:
```
        species   sex    index  Frontal Lobe Length  Rear Width  \
count      200   200  200.000              200.000     200.000
unique       2     2      NaN                  NaN         NaN
top       BLUE  MALE      NaN                  NaN         NaN
freq       100   100      NaN                  NaN         NaN
mean       NaN   NaN   25.500               15.583      12.738
std        NaN   NaN   14.467                3.495       2.573
min        NaN   NaN    1.000                7.200       6.500
25%        NaN   NaN   13.000               12.900      11.000
50%        NaN   NaN   25.500               15.550      12.800
75%        NaN   NaN   38.000               18.050      14.300
max        NaN   NaN   50.000               23.100      20.200

        Carepace Length  Carepace Width  Bodylength
count           200.000         200.000     200.000
unique              NaN             NaN         NaN
top                 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
max              47.600          54.600      21.600
```

[ ]:
```
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()
```

```
[ ]: BLUEMALE        50
     BLUEFEMALE      50
     ORANGEMALE      50
     ORANGEFEMALE    50
     Name: class, dtype: int64
```