S11 T01 Agrupa els diferents vols

May 26, 2022

1 Nivell 1

1.1 Exercici 1

Agrupa els diferents vols utilitzant l'algorisme de K-means.

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

[2]:
```
fly = pd.read_csv('C:/Users/Guillermo/Desktop/Curs Data Scientist/Sprint 2/S02␣
↪T05 Exploració de les dades/DelayedFlights.csv')
```

[3]:
```
fly = fly.drop('Unnamed: 0',axis = 1)
```

[4]:
```
print(fly['Cancelled'].value_counts()/len(fly))
```

```
0    0.999673
1    0.000327
Name: Cancelled, dtype: float64
```

[5]:
```
fly['Diverted'].value_counts()/len(fly)
```

[5]:
```
0    0.995996
1    0.004004
Name: Diverted, dtype: float64
```

Se va a estratificar el muestreo para que esten todos los casos de vuelos cancelados y desviados

[6]:
```
NoCanDiv = fly[(fly['Cancelled']== 0) & (fly['Diverted']==0)]
Can = fly[fly['Cancelled']!= 0 ]
Div = fly[fly['Diverted']!=0]
```

[7]:
```
print(NoCanDiv.shape)
print(Can.shape)
print(Div.shape)
print(len(fly))
```

```
(1928371, 29)
(633, 29)
(7754, 29)
1936758
```

[8]:
```
pd.crosstab(fly['Cancelled'], fly['Diverted'])
```

[8]:
```
Diverted         0     1
Cancelled
0          1928371  7754
1              633     0
```

Al dividir el dataframe en diferents subdataframes per asegurar que el mostreig es estratificat es mostra en la taula anterior que no es solapen, com és evident, la variable cancel·lada i desviada. Per tant, al dividir-ho en dos dataframes i després tornar-ho a montar no hi ha risc de que es solapin les variables

[9]:
```
#Mirem per a 10000 mostres quantes han de pertanyer als vols cancel·lats o␣
↪desviats
print('Can: ', round(10000*len(Can)/len(fly),0))
print('Div: ', round(10000*len(Div)/len(fly),0))
```

```
Can:  3.0
Div:  40.0
```

[10]:
```
SampCan = Can.sample(n=40, random_state = 42)
SampDiv = Div.sample(n=3, random_state = 42)
SampNoCanDiv = NoCanDiv.sample(n = (10000-43), random_state=42)
print(SampCan.shape)
print(SampDiv.shape)
print(SampNoCanDiv.shape)
```

```
(40, 29)
(3, 29)
(9957, 29)
```

[11]:
```
fly2 = pd.concat([SampCan, SampDiv, SampNoCanDiv], ignore_index=True)
print(fly2.shape)
```

```
(10000, 29)
```

[12]:
```
fly2.head()
fly2.tail()
```

[12]:
```
      Year  Month  DayofMonth  DayOfWeek  DepTime  CRSDepTime  ArrTime  \
9995  2008      1           4          5   1604.0         936   1912.0
9996  2008      3          30          7   1923.0        1855   2207.0
9997  2008      7          30          3    715.0         700    955.0
```

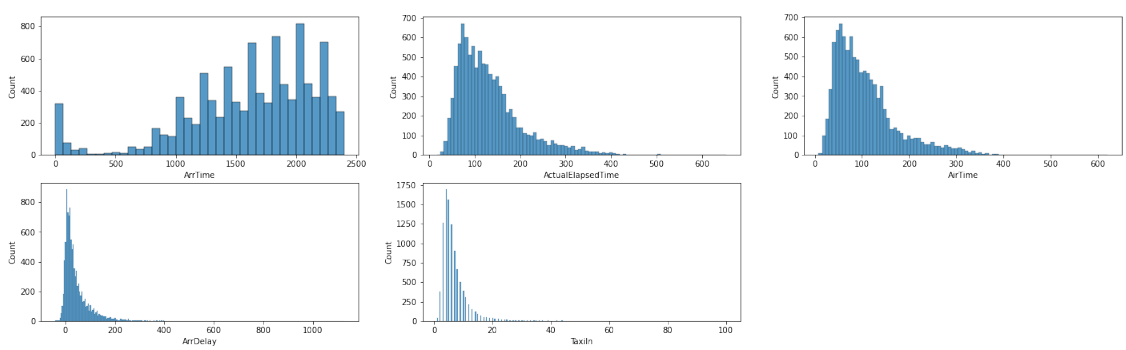
```
9998  2008     12          18          4   2125.0         2108   2145.0
9999  2008     12           1          1   2029.0         1910   2251.0

      CRSArrTime UniqueCarrier  FlightNum  … TaxiIn  TaxiOut  Cancelled  \
9995        1243            OO       6064  …    8.0     11.0          0
9996        2205            AA       1831  …    3.0     18.0          0
9997         937            OH       5565  …   11.0     15.0          0
9998        2128            US        656  …    6.0     17.0          0
9999        2148            CO       1593  …   10.0     25.0          0

      CancellationCode  Diverted  CarrierDelay WeatherDelay NASDelay  \
9995                 N         0           0.0          0.0    389.0
9996                 N         0           NaN          NaN      NaN
9997                 N         0          15.0          0.0      3.0
9998                 N         0          17.0          0.0      0.0
9999                 N         0           0.0          0.0      0.0

      SecurityDelay  LateAircraftDelay
9995            0.0                0.0
9996            NaN                NaN
9997            0.0                0.0
9998            0.0                0.0
9999            0.0               63.0

[5 rows x 29 columns]
```

[13]:
```
fly2.isna().sum()
```

[13]:
```
Year                  0
Month                 0
DayofMonth            0
DayOfWeek             0
DepTime               0
CRSDepTime            0
ArrTime              43
CRSArrTime            0
UniqueCarrier         0
FlightNum             0
TailNum               0
ActualElapsedTime    43
CRSElapsedTime        0
AirTime              43
ArrDelay             43
DepDelay              0
Origin                0
Dest                  0
Distance              0
```

```
TaxiIn                   43
TaxiOut                  26
Cancelled                 0
CancellationCode          0
Diverted                  0
CarrierDelay           3493
WeatherDelay           3493
NASDelay               3493
SecurityDelay          3493
LateAircraftDelay      3493
dtype: int64
```

[14]:
```
#Comprovem la distribució de les variables que tenen NANs i no han sigut␣
 ↪substituits en exercicis anteriors
plt.figure(figsize = (10,7))
plt.subplot(2,3,1)
sns.histplot(fly2['ArrTime'])
plt.subplot(2,3,2)
sns.histplot(fly2['ActualElapsedTime'])
plt.subplot(2,3,3)
sns.histplot(fly2['AirTime'])
plt.subplot(2,3,4)
sns.histplot(fly2['ArrDelay'])
plt.subplot(2,3,5)
sns.histplot(fly2['TaxiIn'])
plt.subplots_adjust(right = 2)
plt.show()
```



Oservem que cap de les variables segueix una distribució normal. Per tant, els NaNs seràn també substituits per la mitjana per aquestes variables com en els exercicis anteriors.

Per tant, al igual que en exercicis anteriors dividirem el dataframe en variables cualitatives i cuantitatives

[15]:
```
fly2.columns
```

[15]: Index(['Year', 'Month', 'DayofMonth', 'DayOfWeek', 'DepTime', 'CRSDepTime',
       'ArrTime', 'CRSArrTime', 'UniqueCarrier', 'FlightNum', 'TailNum',
       'ActualElapsedTime', 'CRSElapsedTime', 'AirTime', 'ArrDelay',
       'DepDelay', 'Origin', 'Dest', 'Distance', 'TaxiIn', 'TaxiOut',
       'Cancelled', 'CancellationCode', 'Diverted', 'CarrierDelay',
       'WeatherDelay', 'NASDelay', 'SecurityDelay', 'LateAircraftDelay'],
      dtype='object')

[16]:
```
fly2.CancellationCode.value_counts()
```

[16]:
```
N    9960
B      21
A      14
C       5
Name: CancellationCode, dtype: int64
```

[17]:
```
#Eliminem l'any perquè no aporta informació
fly2 = fly2.drop('Year', axis=1)
```

[18]:
```
num = ['Month', 'DayofMonth', 'DayOfWeek', 'DepTime', 'CRSDepTime', 'ArrTime',
       'CRSArrTime', 'FlightNum', 'ActualElapsedTime', 'CRSElapsedTime',␣
 ↪'AirTime', 'ArrDelay',
       'DepDelay','Distance', 'TaxiIn', 'TaxiOut', 'CarrierDelay',
       'WeatherDelay', 'NASDelay', 'SecurityDelay', 'LateAircraftDelay']
cat = ['UniqueCarrier', 'TailNum','Origin', 'Dest', 'CancellationCode']
catNoDum = ['Cancelled', 'Diverted']#Excluim del dummies les variables que␣
 ↪només son 0 o 1
```

[19]:
```
flyNum = fly2.loc[:, num]
print(flyNum.shape)
flyCat = fly2.loc[:,cat]
print(flyCat.shape)
flycatNoDum = fly2.loc[:,catNoDum]
print(flycatNoDum.shape)
```

```
(10000, 21)
(10000, 5)
(10000, 2)
```

[20]:
```
from sklearn.impute import SimpleImputer

imp = SimpleImputer(missing_values = np.nan, strategy = 'median')

temp = imp.fit_transform(flyNum)

flyNum = pd.DataFrame(temp, columns = num)
```

```python
print(flyNum.shape)
print(flyNum.isna().sum())
flyNum.head()
```

```
(10000, 21)
Month                0
DayofMonth           0
DayOfWeek            0
DepTime              0
CRSDepTime           0
ArrTime              0
CRSArrTime           0
FlightNum            0
ActualElapsedTime    0
CRSElapsedTime       0
AirTime              0
ArrDelay             0
DepDelay             0
Distance             0
TaxiIn               0
TaxiOut              0
CarrierDelay         0
WeatherDelay         0
NASDelay             0
SecurityDelay        0
LateAircraftDelay    0
dtype: int64
```

[20]:

```
   Month  DayofMonth  DayOfWeek  DepTime  CRSDepTime  ArrTime  CRSArrTime  \
0   12.0        21.0        7.0   2349.0      2005.0   1719.0      2308.0
1   12.0        23.0        2.0   1915.0      1655.0   1719.0      1801.0
2   12.0        18.0        4.0   2043.0      2036.0   1719.0      2051.0
3   11.0        16.0        7.0   2204.0      2130.0   1719.0      2355.0
4   12.0        26.0        5.0   1940.0      1916.0   1719.0      1937.0

   FlightNum  ActualElapsedTime  CRSElapsedTime  …  ArrDelay  DepDelay  \
0      349.0              116.0           243.0  …      25.0     224.0
1     6346.0              116.0            66.0  …      25.0     140.0
2     5655.0              116.0            75.0  …      25.0       7.0
3     1012.0              116.0           145.0  …      25.0      34.0
4     1517.0              116.0            81.0  …      25.0      24.0

   Distance  TaxiIn  TaxiOut  CarrierDelay  WeatherDelay  NASDelay  \
0    1437.0     6.0     14.0           1.0           0.0       2.0
1     221.0     6.0     32.0           1.0           0.0       2.0
2     247.0     6.0     14.0           1.0           0.0       2.0
3     920.0     6.0     21.0           1.0           0.0       2.0
```

```
4     238.0     6.0     14.0           1.0           0.0       2.0

   SecurityDelay  LateAircraftDelay
0            0.0               10.0
1            0.0               10.0
2            0.0               10.0
3            0.0               10.0
4            0.0               10.0

[5 rows x 21 columns]
```

[21]:
```python
import sklearn.preprocessing as sklp

scaler = sklp.RobustScaler()

temp = scaler.fit_transform(flyNum)

minmax = sklp.MinMaxScaler()

temp = minmax.fit_transform(temp)

flyNum = pd.DataFrame(temp, columns = num)

print(flyNum.shape)
flyNum.head()
```

```
(10000, 21)
```

[21]:
```
      Month  DayofMonth  DayOfWeek   DepTime  CRSDepTime   ArrTime  \
0  1.000000    0.666667   1.000000  0.978741    0.849873  0.716132
1  1.000000    0.733333   0.166667  0.797832    0.701442  0.716132
2  1.000000    0.566667   0.500000  0.851188    0.863020  0.716132
3  0.909091    0.500000   1.000000  0.918299    0.902884  0.716132
4  1.000000    0.833333   0.666667  0.808253    0.812129  0.716132

   CRSArrTime  FlightNum  ActualElapsedTime  CRSElapsedTime  …  ArrDelay  \
0    0.978372   0.044473            0.15566        0.342271  …   0.05641
1    0.763359   0.810863            0.15566        0.063091  …   0.05641
2    0.869381   0.722556            0.15566        0.077287  …   0.05641
3    0.998304   0.129201            0.15566        0.187697  …   0.05641
4    0.821035   0.193738            0.15566        0.086751  …   0.05641

   DepDelay  Distance    TaxiIn   TaxiOut  CarrierDelay  WeatherDelay  \
0  0.201107  0.285135  0.050505  0.032609      0.000917           0.0
1  0.123616  0.038532  0.050505  0.081522      0.000917           0.0
2  0.000923  0.043805  0.050505  0.032609      0.000917           0.0
3  0.025830  0.180288  0.050505  0.051630      0.000917           0.0
```

```
4  0.016605  0.041979  0.050505  0.032609      0.000917           0.0

   NASDelay  SecurityDelay  LateAircraftDelay
0  0.005141            0.0           0.023095
1  0.005141            0.0           0.023095
2  0.005141            0.0           0.023095
3  0.005141            0.0           0.023095
4  0.005141            0.0           0.023095

[5 rows x 21 columns]
```

[22]:
```
flyCat.columns
```

[22]:
```
Index(['UniqueCarrier', 'TailNum', 'Origin', 'Dest', 'CancellationCode'],
dtype='object')
```

Per fer el dummies passem a utilitzar els següents prefixes * UniqueCarrier -> UC * TailNum -> TN * Origin -> Or * Dest -> De * CancellationCode -> CC

[23]:
```
flyCat = pd.get_dummies(data=flyCat, prefix = ['UC', 'TN', 'Or', 'De', 'CC'])

print(flyCat.shape)
```

```
(10000, 4599)
```

[24]:
```
print(flyNum.shape)
print( flyCat.shape)
print(flycatNoDum.shape)
```

```
(10000, 21)
(10000, 4599)
(10000, 2)
```

[25]:
```
#Ara toca unificar els 3 dataframes
fly3 = pd.concat([flyNum, flyCat, flycatNoDum], axis = 1)
print(fly3.shape)#Comprovem que el número de columnes es correcte
```

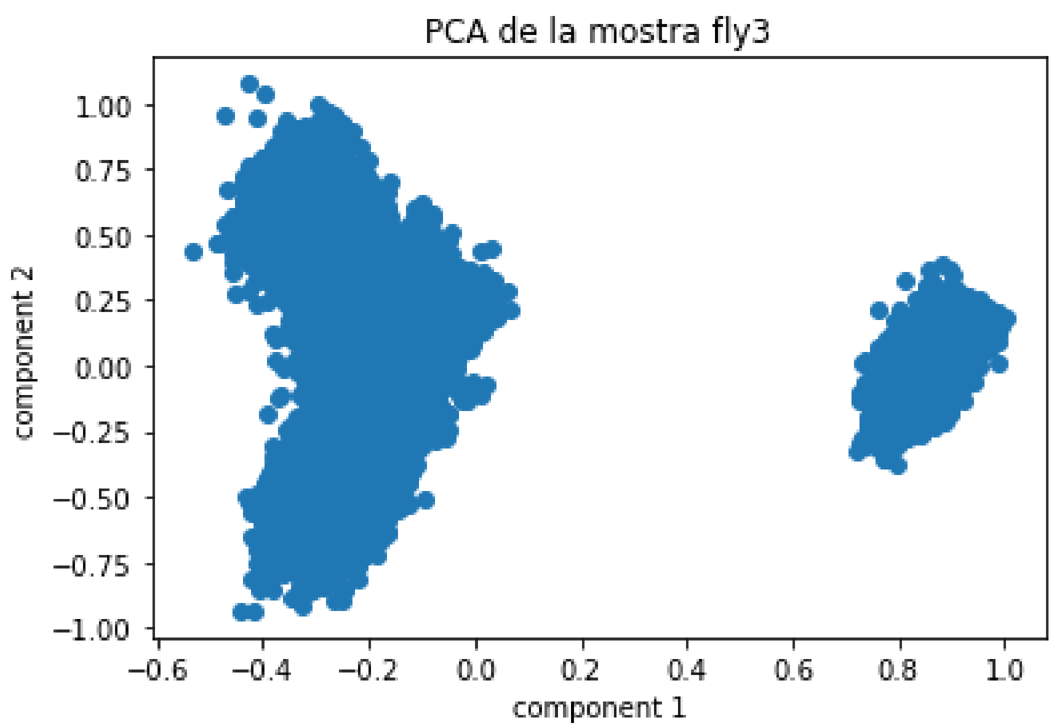
```
(10000, 4622)
```

[26]:
```
fly3.head()
```

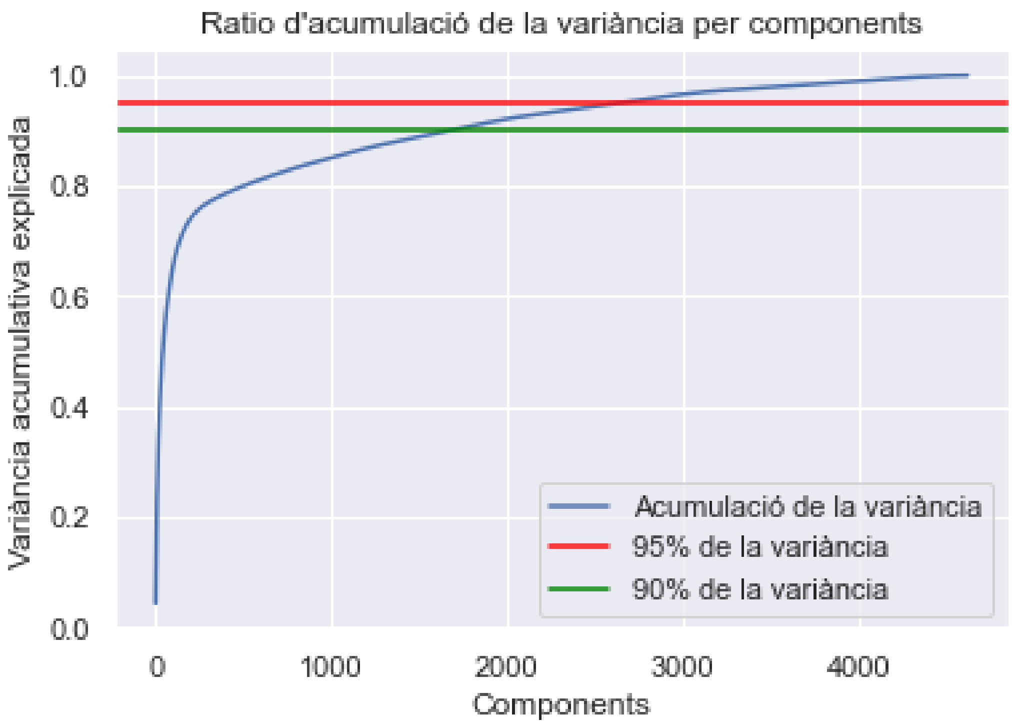
[26]:
```
      Month  DayofMonth  DayOfWeek   DepTime  CRSDepTime   ArrTime  \
0  1.000000    0.666667   1.000000  0.978741    0.849873  0.716132
1  1.000000    0.733333   0.166667  0.797832    0.701442  0.716132
2  1.000000    0.566667   0.500000  0.851188    0.863020  0.716132
3  0.909091    0.500000   1.000000  0.918299    0.902884  0.716132
4  1.000000    0.833333   0.666667  0.808253    0.812129  0.716132
```

```
   CRSArrTime  FlightNum  ActualElapsedTime  CRSElapsedTime  ...  De_WRG  \
0    0.978372   0.044473            0.15566        0.342271  ...       0
1    0.763359   0.810863            0.15566        0.063091  ...       0
2    0.869381   0.722556            0.15566        0.077287  ...       0
3    0.998304   0.129201            0.15566        0.187697  ...       0
4    0.821035   0.193738            0.15566        0.086751  ...       0

   De_XNA  De_YAK  De_YUM  CC_A  CC_B  CC_C  CC_N  Cancelled  Diverted
0       0       0       0     1     0     0     0          1         0
1       0       0       0     0     1     0     0          1         0
2       0       0       0     0     1     0     0          1         0
3       0       0       0     1     0     0     0          1         0
4       0       0       0     0     1     0     0          1         0

[5 rows x 4622 columns]
```

[27]:
```python
from sklearn.decomposition import PCA
pca = PCA()
projected = pca.fit_transform(fly3)

plt.scatter(projected[:,0], projected[:,1])
plt.xlabel('component 1')
plt.ylabel('component 2')
plt.title('PCA de la mostra fly3')
plt.show()
```



[28]:
```
sns.set()
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('Components')
plt.ylabel('Variància acumulativa explicada')
plt.axhline(0.95, color = 'red')
plt.axhline(0.9, color = 'green')
plt.title('Ratio d\'acumulació de la variància per components')
plt.legend(['Acumulació de la variància', '95% de la variància', '90% de la␣
↪variància'])
plt.show()
```



En l'exploració de les dades a través de PCA es mostren dos grups ben diferenciats. A més, tal com mostra la ratio d'acumulació de la variància per components (gràfica anterior) es poden reduir molt el número de components. Per tant, amb l'objectiu de reduir temps de computació es treballarà amb el 90% de les dades, fet que permet reduir el número de variables per sota de la meitat i només desaprofitar un 10% de la informació.

[29]:
```
np.cumsum(pca.explained_variance_ratio_)
```

[29]:
```
array([0.04259262, 0.07116561, 0.09758819, …, 1.        , 1.        ,
       1.        ])
```

1.1.1 k-means

[30]:
```
#Primer reduïm el número de components
pca2 = PCA(0.9)

pca2.fit(fly3)

print(pca2.n_components_)

flyKmeans = pca2.transform(fly3)
```

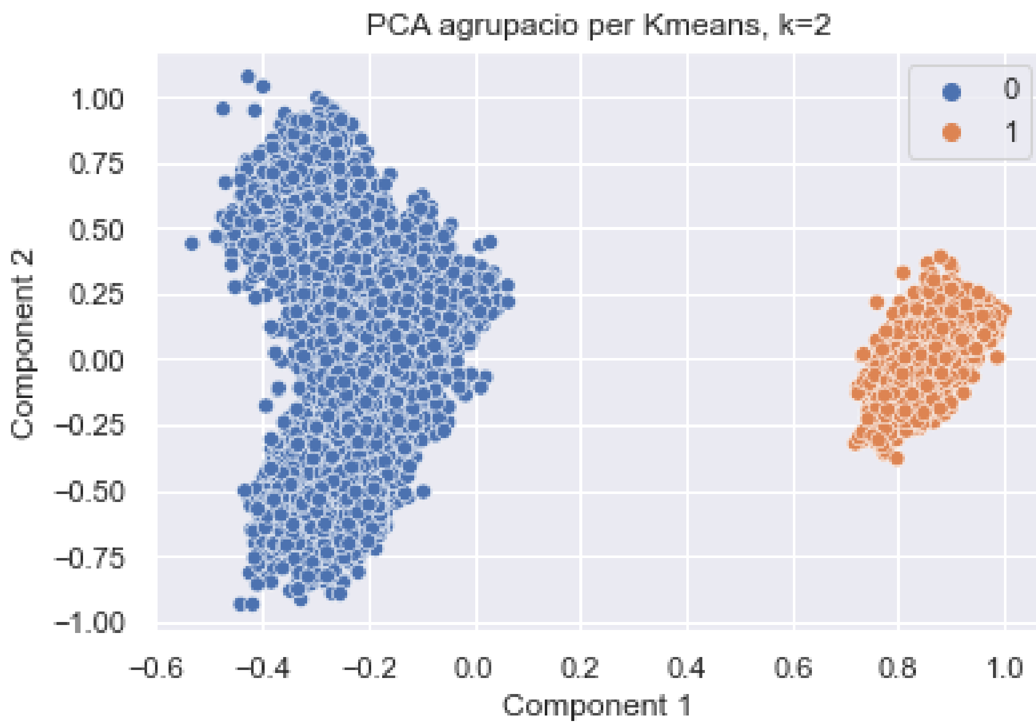
```
1666
```

[31]:
```
#Passem a fer el k-means

from kneed import KneeLocator
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
```

[32]:
```
kmeans = KMeans(init = 'random', n_clusters = 2,
               n_init = 10, max_iter=300, random_state=42)

kmeans.fit(flyKmeans)
```

[32]:
```
KMeans(init='random', n_clusters=2, random_state=42)
```

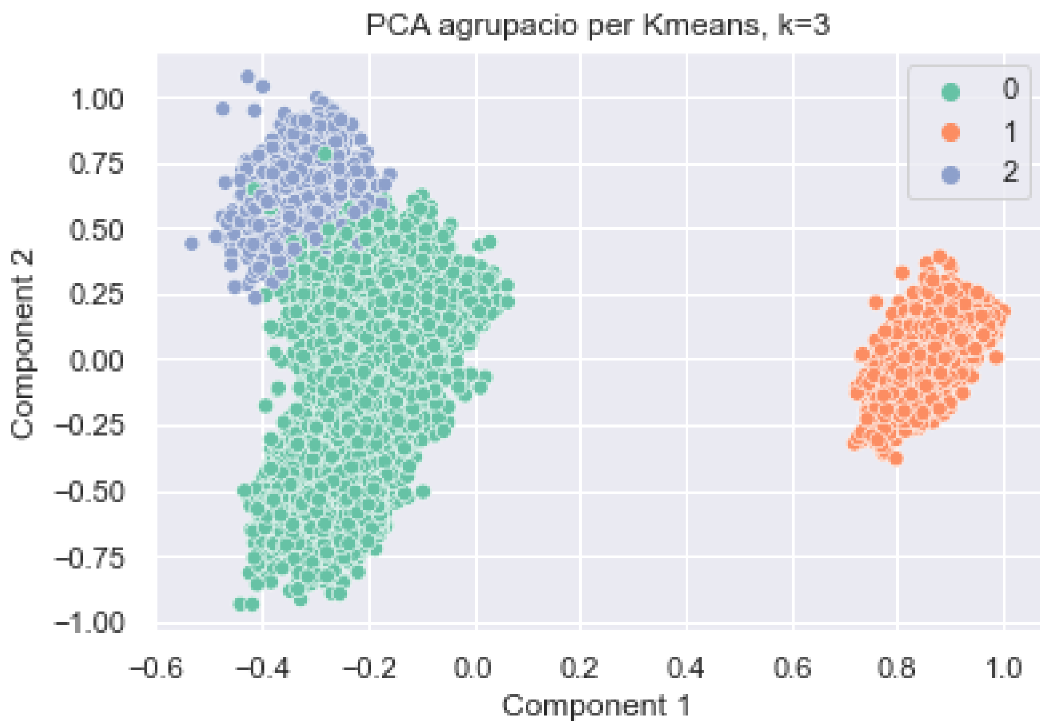
[33]:
```
sns.scatterplot(x=flyKmeans[:,0], y=flyKmeans[:,1], hue = kmeans.labels_)
plt.xlabel('Component 1')
plt.ylabel('Component 2')
plt.title('PCA agrupacio per Kmeans, k=2')
plt.show()
```



[34]:
```
#Ara ho fem sobre 3
kmeans3 = KMeans(init = 'random', n_clusters = 3,
                 n_init = 10, max_iter=300, random_state=42)

kmeans3.fit(flyKmeans)
```

[34]: KMeans(init='random', n_clusters=3, random_state=42)

[35]:
```
sns.scatterplot(x=flyKmeans[:,0], y=flyKmeans[:,1], hue = kmeans3.labels_,␣
 ↪palette = 'Set2')
plt.xlabel('Component 1')
plt.ylabel('Component 2')
plt.title('PCA agrupacio per Kmeans, k=3')
plt.show()
```



[36]:
```
kmeans_kwargs = {
    'init':'random',
    'n_init':10,
    'max_iter':300,
    'random_state':42,
}

sse = []

for k in range(1,30):
    kmeans = KMeans(n_clusters=k, **kmeans_kwargs)
    kmeans.fit(flyKmeans)
    sse.append(kmeans.inertia_)
```

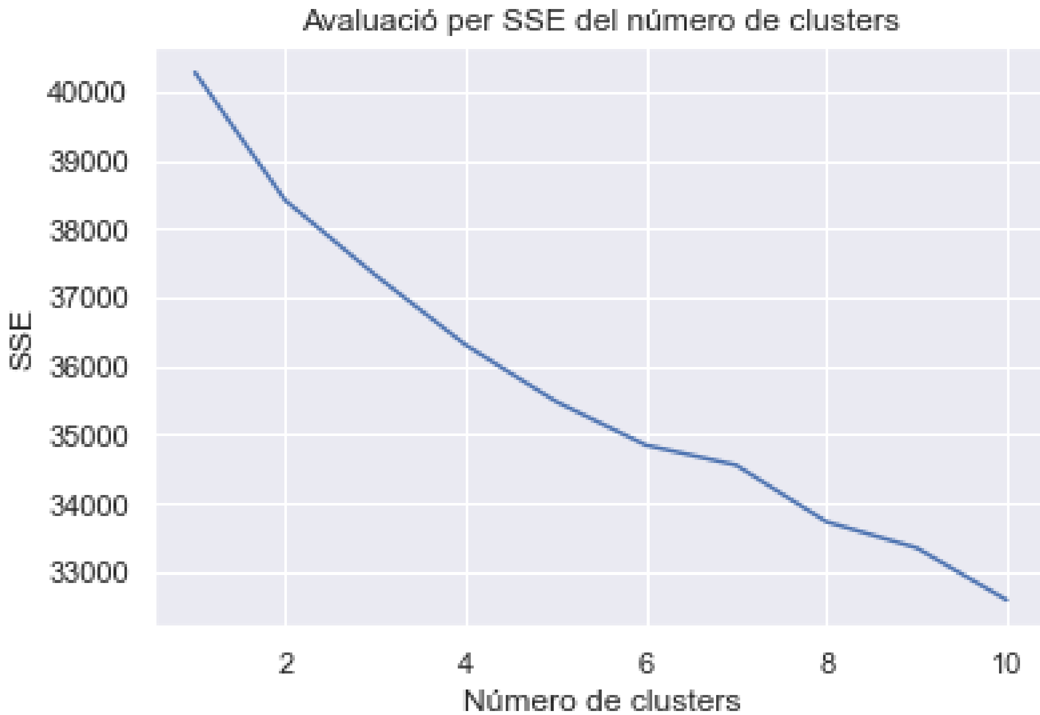
[37]:
```
sse
```

[37]:
```
[40276.6867431205,
 38413.987047026705,
 37324.45988266639,
 36307.02890261333,
 35486.15604027451,
 34847.49220913241,
 34558.398701984624,
```

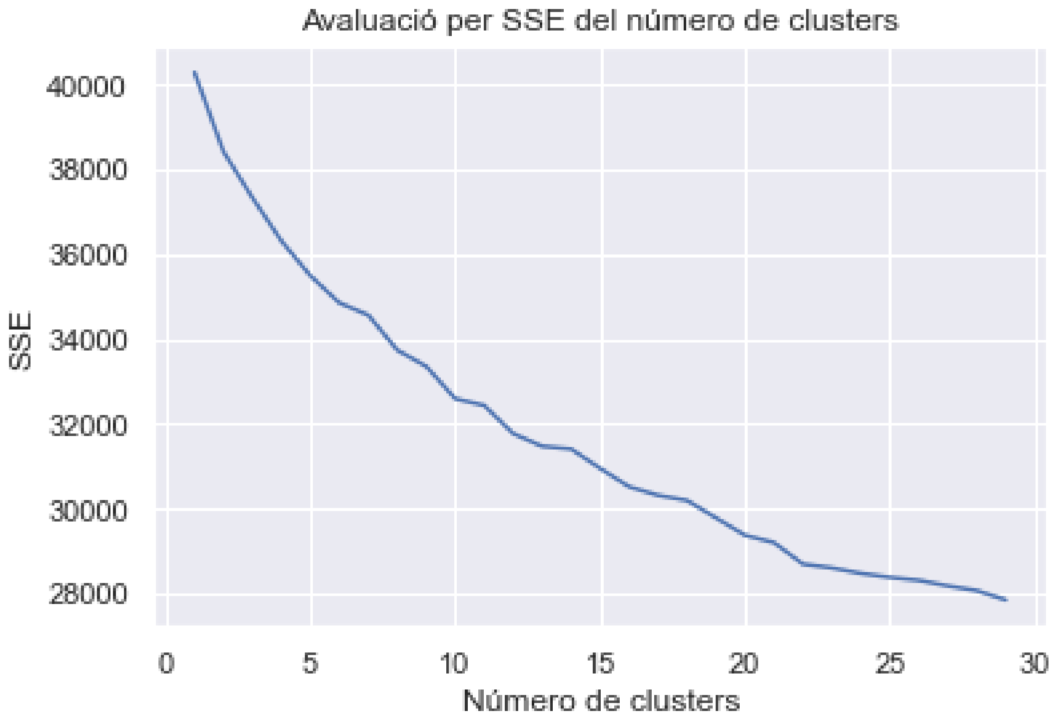
```
33730.87036252899,
33351.50461018891,
32590.54985925202,
32432.269312836845,
31767.55083462334,
31467.77730360981,
31404.746507126965,
30947.562520408715,
30513.599222289562,
30312.022529551425,
30191.83335241132,
29779.01798989931,
29369.878705078558,
29197.839254933646,
28689.833071983725,
28599.433693485073,
28470.636189496647,
28376.691239826727,
28307.42319217917,
28175.58423237423,
28064.487223400607,
27844.08218523256]
```

[38]:

```
plt.plot(range(1,11), sse[0:10])
plt.xlabel('Número de clusters')
plt.ylabel('SSE')
plt.title('Avaluació per SSE del número de clusters')
plt.show()
```



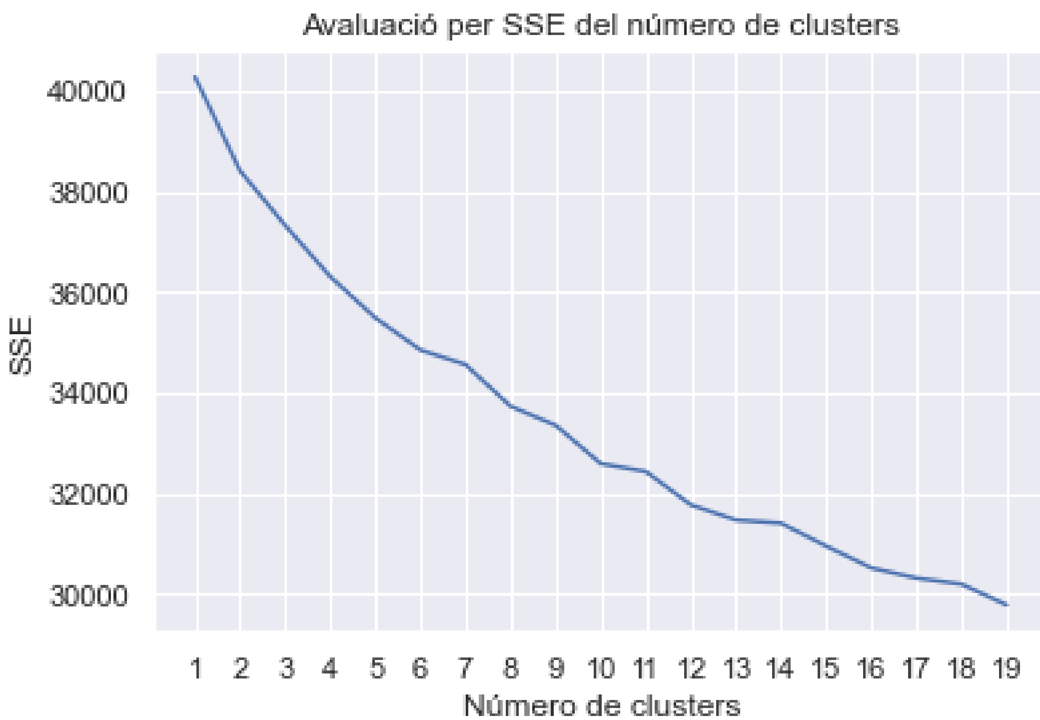
[39]:
```
plt.plot(range(1,30), sse)
plt.xlabel('Número de clusters')
plt.ylabel('SSE')
plt.title('Avaluació per SSE del número de clusters')
plt.show()
```



[40]:
```
from kneed import KneeLocator

kl = KneeLocator(range(1,30), sse, curve='convex', direction='decreasing')

kl.elbow
```

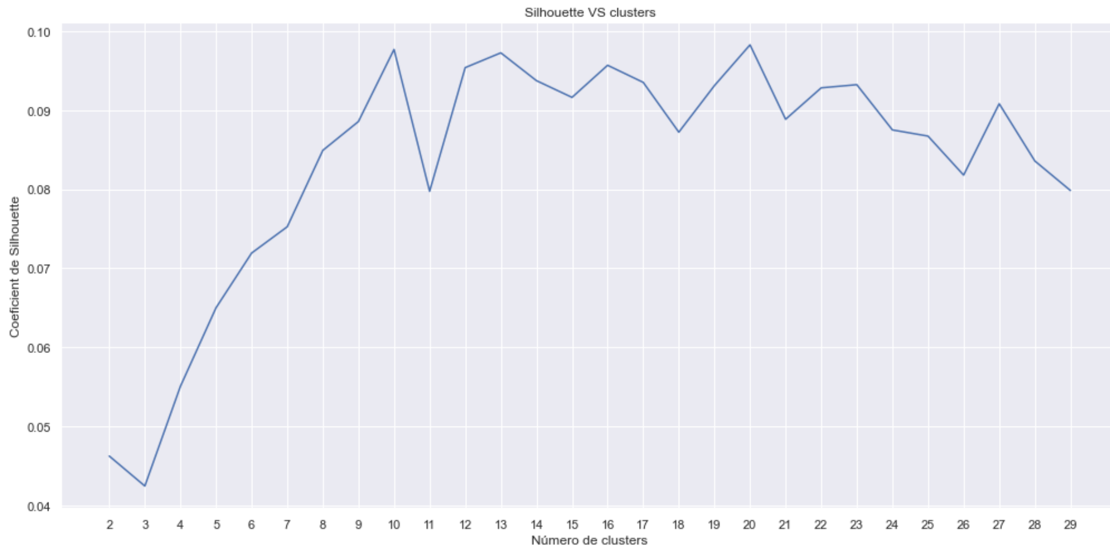
[40]: 12

[41]:
```
plt.plot(range(1,20), sse[0:19])
plt.xlabel('Número de clusters')
plt.ylabel('SSE')
plt.title('Avaluació per SSE del número de clusters')
plt.xticks(range(1,20))
plt.show()
```



[42]:
```
silhouette_coefficientes = []

for k in range(2,30):
    kmeans = KMeans(n_clusters=k, **kmeans_kwargs)
    kmeans.fit(flyKmeans)
    score =  silhouette_score(flyKmeans, kmeans.labels_)
    silhouette_coefficientes.append(score)
```

[43]:
```
plt.figure(figsize=(17,8))
plt.plot(range(2,30), silhouette_coefficientes)
plt.xticks(range(2,30))
plt.xlabel('Número de clusters')
plt.ylabel('Coeficient de Silhouette')
plt.title('Silhouette VS clusters')
plt.show()
```



Tant el mètode del colze com silhouette no han donat bons resultats. Això pot voler dir que és probable que hi hagin més de dos grups, a diferencia del que es veia al explorar les dades en la PCA. Per un altre costat, uns valors tan baixos de silhouette poden indicar que es necessari fer un treball previ més exhaustiu de les dades a través de feature enginering, millorar el preprocessing o utilitzar altres eines per tractar, per exemple, sparse matrix.

Un altre opció de que hagi donat uns resultats tan dolents pot ser degut a com es distribueixen les dades. Ja que K-means necessita que es distribueixin en forma de cercle i pot ser en aquest cas son més el · lipsoidals tal i com es veu reflectit en la PCA. Sent més recomenable Gaussian Mixture Models.

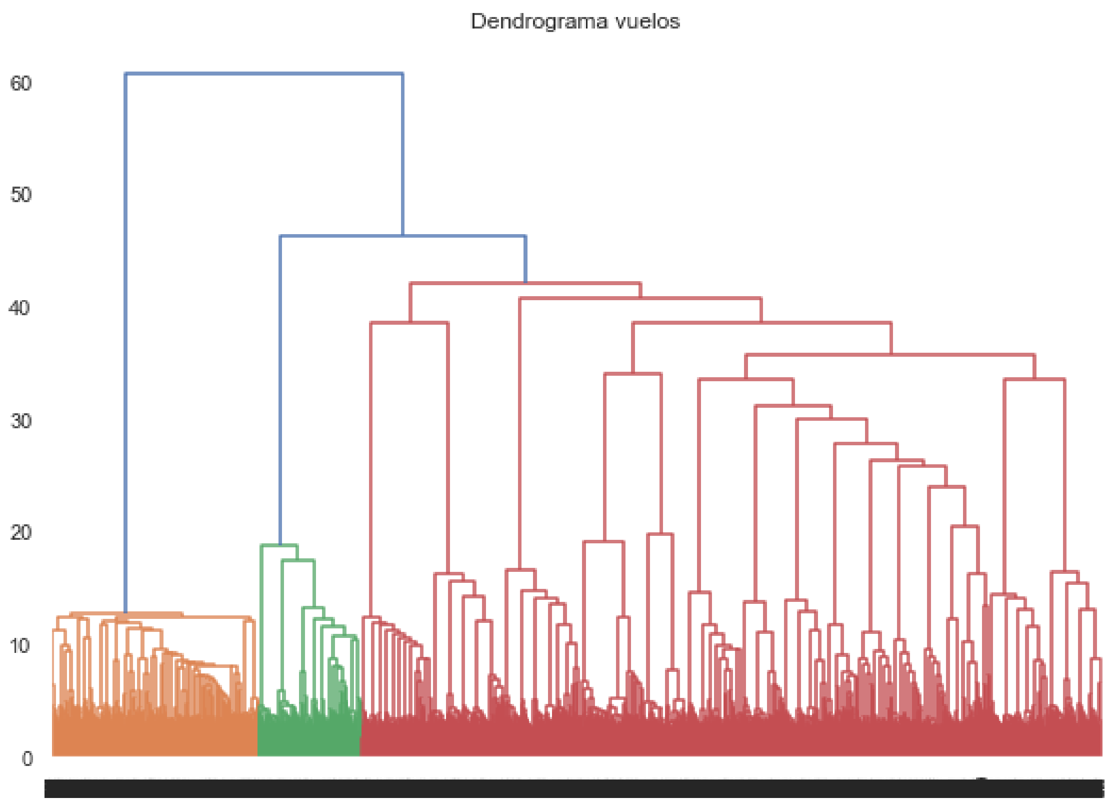
2 Nivell 2

2.1 Exercici 2

Agrupa els diferents vols utilitzant l'algorisme de clustering jeràrquic.

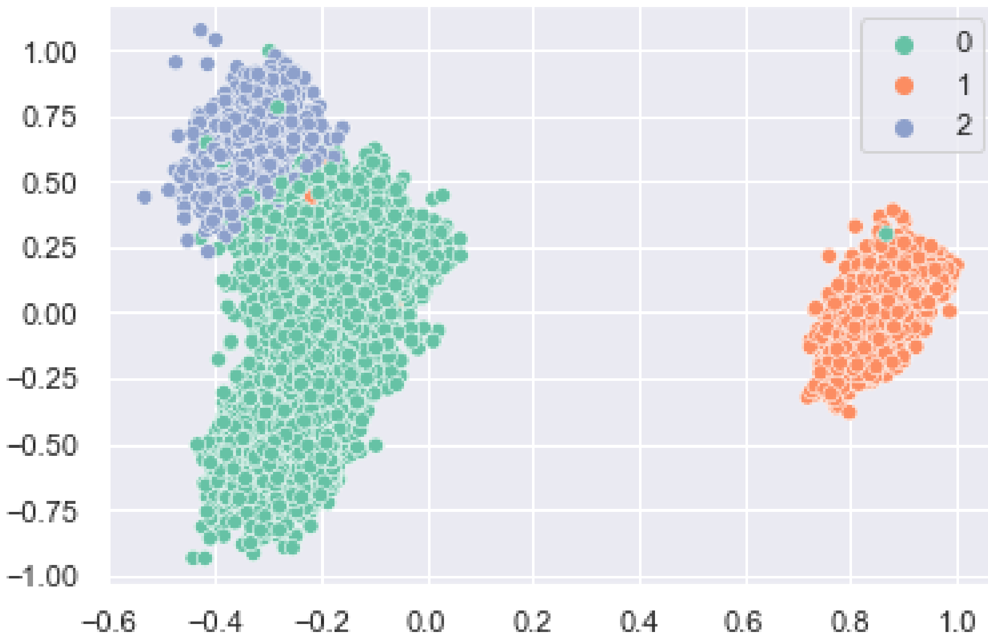
[44]:
```
from sklearn.cluster import AgglomerativeClustering
import scipy.cluster.hierarchy as sch
```

[45]:
```
plt.figure(figsize=(10,7))
plt.title('Dendrograma vuelos')
dendogram = sch.dendrogram(sch.linkage(flyKmeans, method = 'ward'))
plt.show()
```



[46]:
```
model = AgglomerativeClustering(n_clusters=3, affinity = 'euclidean', linkage =␣
↪'ward')

model.fit(flyKmeans)

labels = model.labels_
```

[47]:
```
sns.scatterplot(x=flyKmeans[:,0], y = flyKmeans[:,1], hue=labels,
                palette='Set2')
plt.show()
```



En la gràfica es mostra que els resultats obtinguts amb mètodes jeràrquics i K-means son molt semblants. Tot i així, en el gràfic es veu algun cas entre el grup 0 i 2 que no estan a l'espai que els pertocaria. Això pot ser degut per la resta de components que no apareixen en el gràfic i, per tant, informació que no s'està utilitzant. Per un altre costat, com ja s'ha comentat anteriorment, seria convenient revisar els mètodes empleats en el preprocessament de les dades o aplicar feature engineering per millorar els resultats.

3 Nivell 3

3.1 Exercici 3

Calcula el rendiment del clustering mitjançant un paràmetre com pot ser silhouette.

[48]:
```
score2 =  silhouette_score(flyKmeans, labels)

score2
```

[48]: 0.04123369140080739

Com era d'esperar, degut als resultats anteriors, el valor de silhoutte és molt baix. Això pot ser degut a que la distància entre en els grups es petita i per això el valor és proper a 0. Per tant, com ja s'ha comentat, seria bo tornar a preprocessar les dades per millorar el rendiment.

3.1.1 LLibreries utilitzades

[49]:
```
pip freeze
```

```
alabaster @ file:///home/ktietz/src/ci/alabaster_1611921544520/workNote: you may
need to restart the kernel to use updated packages.
anaconda-client @ file:///C:/ci/anaconda-client_1635342725944/work
anaconda-navigator==2.1.4
anaconda-project @ file:///tmp/build/80754af9/anaconda-
project_1626085644852/work
anyio @ file:///C:/ci/anyio_1620153135622/work/dist

appdirs==1.4.4
argh==0.26.2
argon2-cffi @ file:///C:/ci/argon2-cffi_1613037869401/work
arrow @ file:///C:/ci/arrow_1617738834352/work
asn1crypto @ file:///tmp/build/80754af9/asn1crypto_1596577642040/work
astroid @ file:///C:/ci/astroid_1628063282661/work
astropy @ file:///C:/ci/astropy_1629829318700/work
async-generator @ file:///home/ktietz/src/ci/async_generator_1611927993394/work
atomicwrites==1.4.0
attrs @ file:///tmp/build/80754af9/attrs_1620827162558/work
autopep8 @ file:///tmp/build/80754af9/autopep8_1620866417880/work
Babel @ file:///tmp/build/80754af9/babel_1620871417480/work
backcall @ file:///home/ktietz/src/ci/backcall_1611930011877/work
backports.functools-lru-cache @
file:///tmp/build/80754af9/backports.functools_lru_cache_1618170165463/work
backports.shutil-get-terminal-size @
file:///tmp/build/80754af9/backports.shutil_get_terminal_size_1608222128777/work
backports.tempfile @
file:///home/linux1/recipes/ci/backports.tempfile_1610991236607/work
backports.weakref==1.0.post1
bcrypt @ file:///C:/ci/bcrypt_1607022693089/work
beautifulsoup4 @ file:///tmp/build/80754af9/beautifulsoup4_1631874778482/work
binaryornot @ file:///tmp/build/80754af9/binaryornot_1617751525010/work
bitarray @ file:///C:/ci/bitarray_1629133068652/work
bkcharts==0.2
black==19.10b0
bleach @ file:///tmp/build/80754af9/bleach_1628110601003/work
bokeh @ file:///C:/ci/bokeh_1635306491714/work
boto==2.49.0
Bottleneck @ file:///C:/ci/bottleneck_1607557040328/work
brotlipy==0.7.0
cached-property @ file:///tmp/build/80754af9/cached-property_1600785575025/work
certifi==2022.5.18.1
cffi @ file:///C:/ci/cffi_1625831756778/work
chardet @ file:///C:/ci/chardet_1607706937985/work
charset-normalizer @ file:///tmp/build/80754af9/charset-
```

```
normalizer_1630003229654/work
click==8.0.3
cloudpickle @ file:///tmp/build/80754af9/cloudpickle_1632508026186/work
clyent==1.2.2
colorama @ file:///tmp/build/80754af9/colorama_1607707115595/work
comtypes==1.1.10
conda==4.12.0
conda-build==3.21.6
conda-content-trust @ file:///tmp/build/80754af9/conda-content-
trust_1617045594566/work
conda-pack @ file:///tmp/build/80754af9/conda-pack_1611163042455/work
conda-package-handling @ file:///C:/ci/conda-package-handling_1618262410900/work
conda-repo-cli @ file:///tmp/build/80754af9/conda-repo-cli_1620168426516/work
conda-token @ file:///tmp/build/80754af9/conda-token_1620076980546/work
conda-verify==3.4.2
contextlib2 @ file:///Users/ktietz/demo/mc3/conda-
bld/contextlib2_1630668244042/work
cookiecutter @ file:///tmp/build/80754af9/cookiecutter_1617748928239/work
cryptography @ file:///C:/ci/cryptography_1633520531101/work
cycler==0.10.0
Cython @ file:///C:/ci/cython_1636018292912/work
cytoolz==0.11.0
daal4py==2021.3.0
dask==2021.10.0
debugpy @ file:///C:/ci/debugpy_1629222819322/work
decorator @ file:///tmp/build/80754af9/decorator_1632776554403/work
defusedxml @ file:///tmp/build/80754af9/defusedxml_1615228127516/work
diff-match-patch @ file:///Users/ktietz/demo/mc3/conda-bld/diff-match-
patch_1630511840874/work
distributed @ file:///C:/ci/distributed_1635968318313/work
docutils @ file:///C:/ci/docutils_1620828264669/work
entrypoints==0.3
et-xmlfile==1.1.0
fastcache @ file:///C:/ci/fastcache_1607571310570/work
filelock @ file:///tmp/build/80754af9/filelock_1635402558181/work
flake8 @ file:///tmp/build/80754af9/flake8_1620776156532/work
Flask @ file:///home/ktietz/src/ci/flask_1611932660458/work
fonttools==4.25.0
fsspec @ file:///tmp/build/80754af9/fsspec_1636116461911/work
future @ file:///C:/ci/future_1607568713721/work
gevent @ file:///C:/ci/gevent_1628273776273/work
glob2 @ file:///home/linux1/recipes/ci/glob2_1610991677669/work
greenlet @ file:///C:/ci/greenlet_1628888275363/work
h5py @ file:///C:/ci/h5py_1622088609188/work
HeapDict @ file:///Users/ktietz/demo/mc3/conda-bld/heapdict_1630598515714/work
html5lib @ file:///Users/ktietz/demo/mc3/conda-bld/html5lib_1629144453894/work
idna @ file:///tmp/build/80754af9/idna_1622654382723/work
imagecodecs @ file:///C:/ci/imagecodecs_1635511087451/work
```

```
imageio @ file:///tmp/build/80754af9/imageio_1617700267927/work
imagesize @ file:///Users/ktietz/demo/mc3/conda-bld/imagesize_1628863108022/work
imbalanced-learn @ file:///home/conda/feedstock_root/build_artifacts/imbalanced-
learn_1592145894352/work
importlib-metadata @ file:///C:/ci/importlib-metadata_1631916826748/work
inflection==0.5.1
iniconfig @ file:///home/linux1/recipes/ci/iniconfig_1610983019677/work
intervaltree @ file:///Users/ktietz/demo/mc3/conda-
bld/intervaltree_1630511889664/work
ipykernel @
file:///C:/ci/ipykernel_1633545585502/work/dist/ipykernel-6.4.1-py3-none-any.whl
ipython @ file:///C:/ci/ipython_1635944283918/work
ipython-genutils @
file:///tmp/build/80754af9/ipython_genutils_1606773439826/work
ipywidgets @ file:///tmp/build/80754af9/ipywidgets_1634143127070/work
isort @ file:///tmp/build/80754af9/isort_1628603791788/work
itsdangerous @ file:///tmp/build/80754af9/itsdangerous_1621432558163/work
jdcal @ file:///Users/ktietz/demo/mc3/conda-bld/jdcal_1630584345063/work
jedi @ file:///C:/ci/jedi_1611341083684/work
Jinja2 @ file:///tmp/build/80754af9/jinja2_1612213139570/work
jinja2-time @ file:///tmp/build/80754af9/jinja2-time_1617751524098/work
joblib @ file:///tmp/build/80754af9/joblib_1635411271373/work
json5 @ file:///tmp/build/80754af9/json5_1624432770122/work
jsonschema @ file:///Users/ktietz/demo/mc3/conda-
bld/jsonschema_1630511932244/work
jupyter @ file:///C:/ci/jupyter_1607685287094/work
jupyter-client @ file:///tmp/build/80754af9/jupyter_client_1616770841739/work
jupyter-console @ file:///tmp/build/80754af9/jupyter_console_1616615302928/work
jupyter-core @ file:///C:/ci/jupyter_core_1633420716440/work
jupyter-server @ file:///C:/ci/jupyter_server_1616084298419/work
jupyterlab @ file:///tmp/build/80754af9/jupyterlab_1635799997693/work
jupyterlab-pygments @
file:///tmp/build/80754af9/jupyterlab_pygments_1601490720602/work
jupyterlab-server @
file:///tmp/build/80754af9/jupyterlab_server_1633419203660/work
jupyterlab-widgets @
file:///tmp/build/80754af9/jupyterlab_widgets_1609884341231/work
jupyterthemes==0.20.0
keyring @ file:///C:/ci/keyring_1629321702436/work
kiwisolver @ file:///C:/ci/kiwisolver_1612282555033/work
kneed @
file:///home/conda/feedstock_root/build_artifacts/kneed_1597304056973/work
lazy-object-proxy @ file:///C:/ci/lazy-object-proxy_1616529288960/work
lesscpy==0.15.0
libarchive-c @ file:///tmp/build/80754af9/python-libarchive-c_1617780486945/work
llvmlite==0.37.0
locket==0.2.1
lxml @ file:///C:/ci/lxml_1616443418777/work
```

```
MarkupSafe @ file:///C:/ci/markupsafe_1607027406824/work
matplotlib @ file:///C:/ci/matplotlib-suite_1634667159685/work
matplotlib-inline @ file:///tmp/build/80754af9/matplotlib-
inline_1628242447089/work
mccabe==0.6.1
menuinst @ file:///C:/ci/menuinst_1631733438520/work
mistune @ file:///C:/ci/mistune_1607359457024/work
mkl-fft==1.3.1
mkl-random @ file:///C:/ci/mkl_random_1626186184308/work
mkl-service==2.4.0
mock @ file:///tmp/build/80754af9/mock_1607622725907/work
more-itertools @ file:///tmp/build/80754af9/more-itertools_1635423142362/work
mpmath==1.2.1
msgpack @ file:///C:/ci/msgpack-python_1612287350784/work
multipledispatch @ file:///C:/ci/multipledispatch_1607574329826/work
munkres==1.1.4
mypy-extensions==0.4.3
navigator-updater==0.2.1
nbclassic @ file:///tmp/build/80754af9/nbclassic_1616085367084/work
nbclient @ file:///tmp/build/80754af9/nbclient_1614364831625/work
nbconvert @ file:///C:/ci/nbconvert_1624479160025/work
nbformat @ file:///tmp/build/80754af9/nbformat_1617383369282/work
nest-asyncio @ file:///tmp/build/80754af9/nest-asyncio_1613680548246/work
networkx @ file:///tmp/build/80754af9/networkx_1633639043937/work
nltk==3.6.5
nose @ file:///tmp/build/80754af9/nose_1606773131901/work
notebook @ file:///C:/ci/notebook_1635393701545/work
numba @ file:///C:/ci/numba_1635186062888/work
numexpr @ file:///C:/ci/numexpr_1618856728739/work
numpy @ file:///C:/ci/numpy_and_numpy_base_1626271900803/work
numpydoc @ file:///tmp/build/80754af9/numpydoc_1605117425582/work
olefile @ file:///Users/ktietz/demo/mc3/conda-bld/olefile_1629805411829/work
openpyxl @ file:///tmp/build/80754af9/openpyxl_1632777717936/work
packaging @ file:///tmp/build/80754af9/packaging_1625611678980/work
pandas @ file:///C:/ci/pandas_1635506685681/work
pandoc==2.0.1
pandocfilters @ file:///C:/ci/pandocfilters_1605114832805/work
paramiko @ file:///tmp/build/80754af9/paramiko_1598886428689/work
parso @ file:///tmp/build/80754af9/parso_1617223946239/work
partd @ file:///tmp/build/80754af9/partd_1618000087440/work
path @ file:///C:/ci/path_1624287837534/work
pathlib2 @ file:///C:/ci/pathlib2_1625585796814/work
pathspec==0.7.0
patsy==0.5.2
pep8==1.7.1
pexpect @ file:///tmp/build/80754af9/pexpect_1605563209008/work
pickleshare @ file:///tmp/build/80754af9/pickleshare_1606932040724/work
Pillow==8.4.0
```

```
pkginfo==1.7.1
plotly @ file:///opt/conda/conda-bld/plotly_1646671701182/work
pluggy @ file:///C:/ci/pluggy_1615976440052/work
plumbum==1.7.2
ply==3.11
poyo @ file:///tmp/build/80754af9/poyo_1617751526755/work
prometheus-client @
file:///tmp/build/80754af9/prometheus_client_1623189609245/work
prompt-toolkit @ file:///tmp/build/80754af9/prompt-toolkit_1633440160888/work
psutil @ file:///C:/ci/psutil_1612298199233/work
ptyprocess @ file:///tmp/build/80754af9/ptyprocess_1609355006118/work/dist/ptypr
ocess-0.7.0-py2.py3-none-any.whl
py @ file:///tmp/build/80754af9/py_1607971587848/work
pycodestyle @ file:///tmp/build/80754af9/pycodestyle_1615748559966/work
pycosat==0.6.3
pycparser @ file:///tmp/build/80754af9/pycparser_1594388511720/work
pycurl==7.44.1
pydocstyle @ file:///tmp/build/80754af9/pydocstyle_1621600989141/work
pydotplus @ file:///tmp/build/80754af9/pydotplus_1594390100614/work
pyee @ file:///home/conda/feedstock_root/build_artifacts/pyee_1601273176314/work
pyerfa @ file:///C:/ci/pyerfa_1621560974055/work
pyflakes @ file:///tmp/build/80754af9/pyflakes_1617200973297/work
Pygments @ file:///tmp/build/80754af9/pygments_1629234116488/work
PyJWT @ file:///C:/ci/pyjwt_1619682721924/work
pylint @ file:///C:/ci/pylint_1627536884966/work
pyls-spyder==0.4.0
PyNaCl @ file:///C:/ci/pynacl_1607612759007/work
pyodbc===4.0.0-unsupported
pyOpenSSL @ file:///tmp/build/80754af9/pyopenssl_1635333100036/work
pyparsing @ file:///tmp/build/80754af9/pyparsing_1635766073266/work
pyppeteer @
file:///home/conda/feedstock_root/build_artifacts/pyppeteer_1600514859019/work
pyreadline==2.1
pyrsistent @ file:///C:/ci/pyrsistent_1636093225342/work
PySocks @ file:///C:/ci/pysocks_1605307512533/work
pytest==6.2.4
python-dateutil @ file:///tmp/build/80754af9/python-dateutil_1626374649649/work
python-lsp-black @ file:///tmp/build/80754af9/python-lsp-
black_1634232156041/work
python-lsp-jsonrpc==1.0.0
python-lsp-server==1.2.4
python-slugify @ file:///tmp/build/80754af9/python-slugify_1620405669636/work
pytz==2021.3
PyWavelets @ file:///C:/ci/pywavelets_1607645631519/work
pywin32==228
pywin32-ctypes @ file:///C:/ci/pywin32-ctypes_1607553594546/work
pywinpty @ file:///C:/ci/pywinpty_1607419945780/work
PyYAML==6.0
```

```
pyzmq @ file:///C:/ci/pyzmq_1628276105793/work
QDarkStyle @ file:///tmp/build/80754af9/qdarkstyle_1617386714626/work
qstylizer @ file:///tmp/build/80754af9/qstylizer_1617713584600/work/dist/qstyliz
er-0.1.10-py2.py3-none-any.whl
QtAwesome @ file:///tmp/build/80754af9/qtawesome_1615991616277/work
qtconsole @ file:///tmp/build/80754af9/qtconsole_1632739723211/work
QtPy @ file:///tmp/build/80754af9/qtpy_1629397026935/work
regex @ file:///C:/ci/regex_1628063427816/work
requests @ file:///tmp/build/80754af9/requests_1629994808627/work
rope @ file:///tmp/build/80754af9/rope_1623703006312/work
Rtree @ file:///C:/ci/rtree_1618421015405/work
ruamel-yaml-conda @ file:///C:/ci/ruamel_yaml_1616016898638/work
scikit-image==0.18.3
scikit-learn @ file:///C:/ci/scikit-learn_1642617276183/work
scikit-learn-intelex==2021.20210714.120553
scipy @ file:///C:/ci/scipy_1630606917240/work
seaborn @ file:///tmp/build/80754af9/seaborn_1629307859561/work
Send2Trash @ file:///tmp/build/80754af9/send2trash_1632406701022/work
simplegeneric==0.8.1
singledispatch @ file:///tmp/build/80754af9/singledispatch_1629321204894/work
sip==4.19.13
six @ file:///tmp/build/80754af9/six_1623709665295/work
sniffio @ file:///C:/ci/sniffio_1614030527509/work
snowballstemmer @ file:///tmp/build/80754af9/snowballstemmer_1611258885636/work
sortedcollections @
file:///tmp/build/80754af9/sortedcollections_1611172717284/work
sortedcontainers @
file:///tmp/build/80754af9/sortedcontainers_1623949099177/work
soupsieve @ file:///tmp/build/80754af9/soupsieve_1616183228191/work
Sphinx==4.2.0
sphinxcontrib-applehelp @ file:///home/ktietz/src/ci/sphinxcontrib-
applehelp_1611920841464/work
sphinxcontrib-devhelp @ file:///home/ktietz/src/ci/sphinxcontrib-
devhelp_1611920923094/work
sphinxcontrib-htmlhelp @ file:///tmp/build/80754af9/sphinxcontrib-
htmlhelp_1623945626792/work
sphinxcontrib-jsmath @ file:///home/ktietz/src/ci/sphinxcontrib-
jsmath_1611920942228/work
sphinxcontrib-qthelp @ file:///home/ktietz/src/ci/sphinxcontrib-
qthelp_1611921055322/work
sphinxcontrib-serializinghtml @ file:///tmp/build/80754af9/sphinxcontrib-
serializinghtml_1624451540180/work
sphinxcontrib-websupport @ file:///tmp/build/80754af9/sphinxcontrib-
websupport_1597081412696/work
spyder @ file:///C:/ci/spyder_1636480369575/work
spyder-kernels @ file:///C:/ci/spyder-kernels_1634237096710/work
SQLAlchemy @ file:///C:/ci/sqlalchemy_1626948551162/work
squarify==0.4.3
```

```
statsmodels==0.12.2
sympy @ file:///C:/ci/sympy_1635219088507/work
tables==3.6.1
TBB==0.2
tblib @ file:///Users/ktietz/demo/mc3/conda-bld/tblib_1629402031467/work
tenacity @ file:///C:/ci/tenacity_1626248381338/work
terminado==0.9.4
testpath @ file:///tmp/build/80754af9/testpath_1624638946665/work
text-unidecode @ file:///Users/ktietz/demo/mc3/conda-bld/text-
unidecode_1629401354553/work
textdistance @ file:///tmp/build/80754af9/textdistance_1612461398012/work
threadpoolctl @ file:///Users/ktietz/demo/mc3/conda-
bld/threadpoolctl_1629802263681/work
three-merge @ file:///tmp/build/80754af9/three-merge_1607553261110/work
tifffile @ file:///tmp/build/80754af9/tifffile_1627275862826/work
tinycss @ file:///tmp/build/80754af9/tinycss_1617713798712/work
toml @ file:///tmp/build/80754af9/toml_1616166611790/work
toolz @ file:///home/linux1/recipes/ci/toolz_1610987900194/work
tornado @ file:///C:/ci/tornado_1606924294691/work
tqdm @ file:///tmp/build/80754af9/tqdm_1635330843403/work
traitlets @ file:///tmp/build/80754af9/traitlets_1632522747050/work
typed-ast @ file:///C:/ci/typed-ast_1624953797214/work
typing-extensions @
file:///tmp/build/80754af9/typing_extensions_1631814937681/work
ujson @ file:///C:/ci/ujson_1611259568517/work
unicodecsv==0.14.1
Unidecode @ file:///tmp/build/80754af9/unidecode_1614712377438/work
urllib3==1.26.7
watchdog @ file:///C:/ci/watchdog_1624955113064/work
wcwidth @ file:///Users/ktietz/demo/mc3/conda-bld/wcwidth_1629357192024/work
webencodings==0.5.1
websockets @ file:///D:/bld/websockets_1610127782618/work
Werkzeug @ file:///tmp/build/80754af9/werkzeug_1635505089296/work
whichcraft @ file:///tmp/build/80754af9/whichcraft_1617751293875/work
widgetsnbextension @ file:///C:/ci/widgetsnbextension_1607531582688/work
win-inet-pton @ file:///C:/ci/win_inet_pton_1605306162074/work
win-unicode-console==0.5
wincertstore==0.2
wrapt @ file:///C:/ci/wrapt_1607574570428/work
xlrd @ file:///tmp/build/80754af9/xlrd_1608072521494/work
XlsxWriter @ file:///tmp/build/80754af9/xlsxwriter_1628603415431/work
xlwings==0.24.9
xlwt==1.3.0
xmltodict @ file:///Users/ktietz/demo/mc3/conda-bld/xmltodict_1629301980723/work
yapf @ file:///tmp/build/80754af9/yapf_1615749224965/work
zict==2.0.0
zipp @ file:///tmp/build/80754af9/zipp_1633618647012/work
zope.event==4.5.0
```

```
zope.interface @ file:///C:/ci/zope.interface_1625036252485/work
```

```
[ ]:
```