

## PCA with just scaling

4a

```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.metrics import confusion_matrix
import seaborn as sns
from pathlib import Path
import os
import string
import math
```

```
↳ /usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning
import pandas.util.testing as tm
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
↳ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.m
```

```
path = Path('/content/drive/My Drive/ml_hw6_q4/places.txt')
```

```
cd /content/drive/My Drive/ml_hw6_q4
```

```
↳ /content/drive/My Drive/ml_hw6_q4
```

```
data = pd.read_csv('places.txt',delim_whitespace=True,na_values='?')
table = data[['Climate', 'HousingCost', 'HlthCare', 'Crime', 'Transp', 'Educ', 'Arts','Rec
```

4b

```
table = np.log10(table)
table.head(5)
```

```
↳
```

	Climate	HousingCost	HlthCare	Crime	Transp	Educ	Arts
<b>Abilene,TX</b>	2.716838	3.792392	2.374748	2.965202	3.605413	3.440437	2.998259
<b>Akron,OH</b>	2.759668	3.910518	3.219060	2.947434	3.688687	3.387034	3.745387
<b>Albany,GA</b>	2.670246	3.865637	2.790988	2.986772	3.403292	3.408240	2.374748
<b>Albany-Schenectady-Troy,NY</b>	2.677607	3.898067	3.155640	2.785330	3.837778	3.531351	3.667920

4c

```
mean = np.mean(table,axis = 0)
print(mean)
```

```
↳ Climate      2.718494
   HousingCost  3.907491
   HlthCare     2.955487
   Crime        2.951627
   Transp       3.597202
   Educ         3.446607
   Arts         3.206209
   Recreat      3.226567
   Econ         3.734182
dtype: float64
```

```
table = table - mean
print(np.mean(table,axis = 0))
```

```
↳ Climate      1.915388e-15
   HousingCost  1.773657e-15
   HlthCare     6.479113e-16
   Crime        3.954959e-16
   Transp       5.007814e-16
   Educ         4.778346e-16
   Arts         1.802003e-15
   Recreat      2.663185e-15
   Econ         8.706308e-16
dtype: float64
```

```
from sklearn.decomposition import PCA
pca = PCA()
pca.fit(table)
print(pca.components_.shape)
```

```
↳ (9, 9)
```

4d and 4e

```
pa1 = pca.components_[0]
pa2 = pca.components_[1]
print(f"First principal axis = {pa1}")
print(f"Second principal axis = {pa2}")
```

```
↳ First principal axis = [0.03507288 0.09335159 0.40776448 0.10044536 0.15009714 0.0321
 0.87434057 0.15899622 0.01949418]
   Second principal axis = [ 0.0088782  0.00923057 -0.85853187  0.22042372  0.05920111
 0.30380632 0.33399255 0.0561011 ]
```

```
projected = pca.fit_transform(table)
print(f"First principal component = \n {projected[:, 0]}")
print(f"Second principal component = \n {projected[:, 1]}")
```

```

↳ First principal component =
[-4.36677074e-01  6.20957601e-01 -8.73256319e-01  5.02948082e-01
 6.09775036e-01 -7.45633249e-01  5.09710932e-03 -3.93759635e-02
-8.93575198e-01 -1.64904064e-01  7.03776040e-01  1.43875264e-01
-6.68842631e-01 -9.77440004e-01  8.02625339e-01 -8.04526312e-01
-1.54621973e-01 -1.62389595e-02  1.39749114e-01  8.70728224e-01
-3.45375150e-01 -1.22784742e-01  3.88620116e-01  4.06486994e-01
-2.23536863e-01  1.04615102e+00  7.55859984e-02  1.94483334e-01
-5.73823216e-01 -3.97323657e-02 -1.73224976e-01 -2.71937936e-01
-1.25118019e+00  8.87047017e-01 -3.47869328e-01 -4.80455237e-01
-2.70025874e-03  4.54115136e-01 -1.76004109e-01  1.29823120e-02
-3.92728280e-02  5.86171851e-02  1.40182619e+00  4.45607287e-01
-1.07599057e+00 -2.01012631e-01 -7.83826267e-01  8.34867217e-01
-4.76777676e-01  8.23426352e-02 -1.27166483e+00 -1.31549124e-02
 8.64680835e-01 -1.06223777e+00  2.20394326e-01 -3.48957580e-01
-8.00244428e-01 -2.37092714e-01  1.19494634e-01  3.75861896e-01
 1.05731212e-01  4.14729759e-01  5.87183762e-02  4.00918980e-02
 1.46921996e+00 -2.18986874e-01  8.91196920e-01 -1.17987103e+00
 1.11024687e+00  1.92249660e-02  3.44283008e-01  2.83901534e-01
-1.84063166e-02  6.93766317e-01 -2.37017494e-02 -6.73992134e-01
 9.98519103e-01  7.27347027e-01 -1.34091182e+00 -1.34432925e-01
 7.08905584e-01 -2.23373259e-01 -7.08143811e-01  9.17419732e-01
 1.42535024e-01  1.04390164e+00 -1.18825044e+00 -6.86455896e-01
 1.95828961e-01 -3.14030152e-01 -5.14273024e-02  2.10875056e-01
-7.15495184e-01 -2.87181292e-01 -8.95039022e-01 -1.44424997e-02
 1.70124686e-01  1.43223772e-02 -2.65547618e-01  8.01160662e-02
-3.00889783e-01 -4.61710478e-02 -8.37243984e-01  9.10847566e-02
-3.83591506e-01 -4.01020639e-01 -1.34803039e-03  2.14382030e-01
-5.71690587e-01 -6.50367950e-01 -1.07278923e+00 -1.24615314e+00
 2.46684878e-01  6.03191003e-01  2.20624742e-01 -1.14626426e+00
 3.84738101e-01  1.63914719e-01  4.62123684e-01 -9.68228206e-01
 4.84604778e-02  3.28134086e-01 -5.82257012e-01 -2.47512590e-01
 1.60736093e-01  4.83086418e-01  2.31865016e-01 -3.05048845e-01
 1.64255820e-01  4.04451065e-01  5.67316307e-01 -8.85147349e-01
 7.45412296e-01 -4.74517649e-01  9.58400062e-01  4.42203358e-02
-4.63918633e-02  7.15170403e-01  2.73352648e-01 -6.42834616e-01
 3.88400207e-01  2.25356276e-01 -5.13915635e-01 -2.20975873e-01
 7.02146067e-01  5.90552757e-02 -4.60254894e-01  2.95801051e-01
-1.36151672e+00  1.68070053e-01 -1.27223322e+00 -1.22108918e-02
 6.61899469e-01  3.90131912e-01 -4.39195538e-01  2.97952213e-01
-8.62682511e-01 -6.09089132e-02  5.59686422e-02 -2.30104752e-02
-2.26092458e-01  4.08707816e-01 -7.86168311e-01 -1.94288823e-01
 4.55303098e-01 -1.09732465e+00  1.29501231e-01 -1.81916793e-02
 1.72701448e-01  1.66662080e-01 -2.34635545e-01 -3.56755350e-01
 4.36047161e-01 -1.20759143e-01  3.84154622e-01  2.48478692e-01
-7.74989550e-01  1.84702461e-01  1.45880138e+00  6.03877265e-01
 2.49802093e-01  3.09387727e-01 -2.03607436e-01 -1.74820521e-01
 6.31670537e-01 -5.46428689e-01 -6.45100935e-01 -3.63170849e-01
-1.93844578e-01 -5.51918530e-01  5.91550774e-01  7.25312730e-01
 8.09623600e-01 -2.87863626e-01 -8.57806842e-01  9.16311277e-01
 1.05484916e+00  1.39018082e-01 -6.02138606e-01  8.07831941e-01
-2.20793080e-01 -1.26198068e-01  1.49755345e-01 -4.34845843e-01
 6.10066774e-02  5.31850095e-01  9.13403028e-01 -3.97461675e-01
-1.17364202e-03  7.91065630e-01 -1.79129243e-01  8.83968980e-01
 1.90808973e+00  1.14305813e+00 -2.76472076e-01  7.35256407e-01
 7.96101718e-01  7.87383240e-01 -7.46749805e-01 -9.09327628e-01
 5.49033481e-01 -4.97008531e-01  5.45365891e-01  6.56251954e-01
 2.57143453e-01 -3.92757739e-01  2.43170843e-01 -5.53273869e-01
-1.19781617e+00 -1.04084099e+00 -4.16014237e-01  2.42171523e-03
 1.61987331e-01  1.26326903e+00  5.56389731e-01 -7.38200560e-01
 1.00102992e+00 -8.18486519e-01  1.92951320e-01  6.35554911e-01

```

```

-1.00201802e+00 -4.39937229e-01 7.44202681e-01 -1.59441937e-01
-2.04341556e-01 -1.00951165e-01 7.63243619e-01 -1.41218735e-01
-4.03552058e-01 1.38401285e-01 -6.36332793e-01 7.13513226e-01
6.36725302e-01 -6.16829175e-02 -1.07183975e-01 7.50884042e-01
-2.12246876e-01 4.78148847e-01 4.51345296e-02 -7.20695929e-02
-5.33256733e-01 9.29574069e-01 -3.23688648e-01 3.28705914e-01
-2.61591545e-01 6.68425320e-01 -3.66782512e-01 4.60053529e-01
9.30384017e-01 1.26502281e+00 7.51838384e-01 2.10991448e-01
3.94367304e-01 3.51477955e-01 -2.46996826e-01 2.26659917e-01
4.13875595e-02 1.00664024e+00 -1.46588246e+00 -8.15199182e-01
-1.44263438e+00 2.96734390e-01 -9.74819095e-02 -4.64467530e-01
2.00811654e-01 1.36944066e-01 2.45945937e-01 2.84543680e-01
-8.21374271e-02 8.22965809e-01 -2.23900740e-01 -1.12767013e+00
-1.52885391e-01 5.22244557e-01 3.67616375e-01 -2.65257011e-03
5.63728100e-01 -2.75359954e-01 -1.51418117e+00 5.28677604e-01
4.00579442e-02 5.47050483e-01 5.55056446e-01 4.84355664e-01
-1.22444819e-01 -6.50418220e-01 -4.15458981e-01 2.81234959e-01
-1.24147552e-01 -1.37329797e+00 8.24988588e-02 -6.68710130e-01
-2.62072191e-01 1.35629627e+00 -6.27926571e-01 2.19432220e-01
-3.54050250e-01 1.85975255e-01 -6.36436821e-01 2.14090133e-01
-5.11261542e-01 -9.17309592e-01 5.09380971e-01 -9.05834040e-02
1.08219717e-01 -3.56346894e-01 -3.27268682e-01 2.13432825e-01
-1.29102180e+00]
Second principal component =
[ 4.20163405e-01 5.34579808e-03 -2.12103638e-01 -6.36214530e-02
-7.23262677e-03 -1.88182770e-01 6.34964529e-02 -9.23853756e-02
-9.78616865e-02 4.17715312e-02 -2.81482346e-02 3.31744286e-01
-1.30163191e-01 -4.16159984e-01 -1.26636773e-01 -1.19289184e-01
-2.43710217e-02 -1.79200140e-01 1.62881937e-01 -2.39600454e-01
1.52141581e-01 -4.18409897e-01 1.49547326e-02 1.63683164e-01
2.95596586e-01 -1.18783150e-01 9.32892913e-02 5.17743067e-01
-1.89227519e-01 1.96495826e-01 -1.50352947e-01 5.46005751e-01
-2.32845777e-01 -1.61737631e-01 1.70765147e-01 1.67705766e-01
-1.09594695e-01 -1.35263352e-01 -7.19915014e-02 8.68256380e-02
3.55013576e-03 3.20283407e-01 -1.90429063e-01 2.54291900e-01
5.61431768e-02 2.20344669e-01 1.84864026e-01 -1.43944867e-01
-3.28689170e-01 2.05840074e-02 6.37841377e-01 1.39147569e-01
-8.86132865e-02 -5.31680867e-01 -1.90146194e-01 1.15821987e-01
1.34387663e-01 -7.98896896e-02 1.79313677e-01 8.87178846e-03
4.05172693e-02 9.16187640e-02 -1.71172515e-01 1.87977151e-01
-2.28192718e-01 4.83187865e-01 -1.34115444e-02 -8.55923234e-02
-6.24222621e-02 1.68420145e-01 -2.25372386e-01 -7.08211766e-02
-1.68194083e-02 -3.18819372e-02 3.47510908e-01 -3.03123672e-01
-6.47332676e-02 -2.14767324e-01 -5.02347009e-01 -1.30261614e-01
-1.82428799e-01 1.62885681e-01 -1.96589176e-01 -5.48841882e-02
-7.02944940e-02 -2.15876417e-01 -3.21653451e-01 -1.38114303e-01
-1.50122003e-02 2.18086695e-01 -4.58825929e-02 2.89141196e-01
-1.71949362e-01 6.13368534e-02 -1.88093974e-01 1.69937544e-01
3.49322346e-01 5.98034654e-02 -7.19659545e-02 -4.08456050e-02
-1.74736620e-01 1.30282625e-01 -2.71698966e-01 1.48463891e-01
-6.88452176e-02 -1.57501988e-01 1.92505230e-01 2.51671233e-02
1.20572164e-01 7.08265728e-02 -1.07199530e-01 -1.73280014e-01
-3.26030439e-03 2.30513557e-01 5.79855554e-01 -4.56786320e-01
-1.37886280e-01 -4.90793941e-02 4.50640049e-03 9.62766433e-01
-1.62984833e-01 -1.69353705e-02 -5.07535050e-02 -7.38992467e-02
3.39488116e-02 -1.04347568e-01 -6.73701377e-02 -1.13026430e-01
-5.22802865e-02 -2.66799680e-01 -2.19675115e-01 -4.59672287e-01
1.91974250e-01 2.99593559e-02 -5.16627047e-02 -7.89413478e-02
6.33052945e-02 -1.20691749e-01 -3.60027991e-01 -7.59496010e-02
-1.69433664e-01 5.42222647e-02 -1.23959272e-01 2.38523412e-02
-7.08826226e-02 -1.18294186e-01 -2.20757551e-01 1.98489326e-02
-3.33488403e-01 4.13627206e-02 -3.12904771e-01 -3.19350407e-01

```

```

-4.04386503e-02  9.55396016e-02  1.16432011e-01  9.85298618e-02
-2.56233128e-01  9.33436981e-02  1.09682266e-01  2.42647819e-01
 4.16581515e-01  4.08981076e-02  2.56855934e-01 -2.71789688e-01
-3.15204022e-02 -1.84729233e-01  2.65794122e-01  5.08405516e-01
 2.34891177e-01 -4.72514612e-02  2.30670095e-01 -5.21848073e-02
-1.69643021e-01  1.86441810e-02  1.96051316e-01 -2.10122000e-01
-9.27330766e-02 -6.96413177e-02 -9.40596512e-02 -1.38436777e-01
-4.17020307e-02  1.63452896e-01 -2.56743884e-02  7.50282310e-02
-1.57830313e-01 -2.09018127e-01  2.11720498e-01  2.03511115e-01
 4.32439425e-01  1.64082592e-01 -8.62213115e-02  3.67814052e-02
-2.11846397e-01 -1.21062283e-01  6.95916359e-01 -1.57485502e-01
-2.42494050e-01 -9.26277791e-02  1.39021568e-01 -7.81033111e-02
 3.60798614e-01 -6.20179845e-02 -5.36053213e-03  5.45787038e-02
-1.88242658e-02 -1.04807171e-01 -2.65360802e-01  9.53464522e-02
-1.80427851e-01 -1.00533459e-01  1.54711473e-01 -1.29610723e-02
-1.07967798e-01 -2.23136346e-01  1.56035752e-01 -1.47315008e-04
-1.72716386e-01 -6.07155780e-02  1.19396498e-01  3.85098136e-02
-2.12550425e-02  2.21143238e-01 -2.32784964e-01 -1.06211688e-02
 2.82501390e-01 -2.03579634e-01  1.53888870e-02  5.26241156e-01
-3.74240699e-01 -1.83131294e-01 -2.70090059e-01  2.78311945e-01
-6.22349544e-02 -2.93029244e-01  2.50531793e-02 -1.62246125e-01
-2.29135608e-01 -2.22457813e-01  1.80262423e-01  5.52850390e-02
-2.38940445e-01  2.25236844e-01 -1.42814450e-01  6.27091940e-01
 2.31865891e-01 -1.84793604e-02 -3.20863232e-01 -2.11588735e-01
 4.21272294e-01  2.76046819e-01  2.19884997e-01 -1.52102000e-01
-1.12583864e-01 -6.83447351e-02 -6.14860250e-01 -2.70876990e-02
-1.57777329e-01  1.80186735e-01 -7.31278620e-03  7.12288471e-02
 3.97151526e-01 -1.14791103e-01 -2.11782298e-02 -3.56544700e-02
 4.11373244e-01  8.65868149e-02  4.29296789e-01  5.76265055e-03
-1.01123550e-02 -1.64999566e-02 -7.59570084e-02  3.91551785e-01
 8.09537817e-02  8.85792852e-02  2.74666537e-02  8.85588056e-02
-2.15594848e-01  2.11710494e-02 -5.27627721e-01 -1.86257369e-01
-1.94651466e-01 -9.52403291e-02 -4.40487458e-02 -2.87628024e-01
 1.29222642e-01  2.91461970e-01 -2.19992461e-01  2.08507306e-01
 1.74466288e-01 -1.89349563e-01  4.96544909e-04 -2.25623567e-01
 4.08912382e-01  1.56953925e-01  4.78436248e-01  4.28046165e-01
 1.90675355e-01  1.49573982e-01 -2.26050208e-01 -7.74209330e-02
-1.26396259e-01  2.13564946e-02  1.62935897e-01 -2.94884448e-02
 4.73525532e-02 -2.09353697e-01  1.58314906e-01  1.86789656e-02
 1.68149085e-01  6.66868160e-02  8.01305870e-02  4.90143869e-01
 1.23341651e-01 -1.98861770e-01 -4.33381307e-01  5.52505592e-02
-1.42813807e-01  6.01392750e-01 -2.88386208e-01  1.33066022e-01
 1.86651875e-02 -4.64265137e-01  6.18652810e-03 -2.43665070e-02
-2.67956767e-01  2.80679778e-01 -1.57687699e-01 -1.27975778e-02

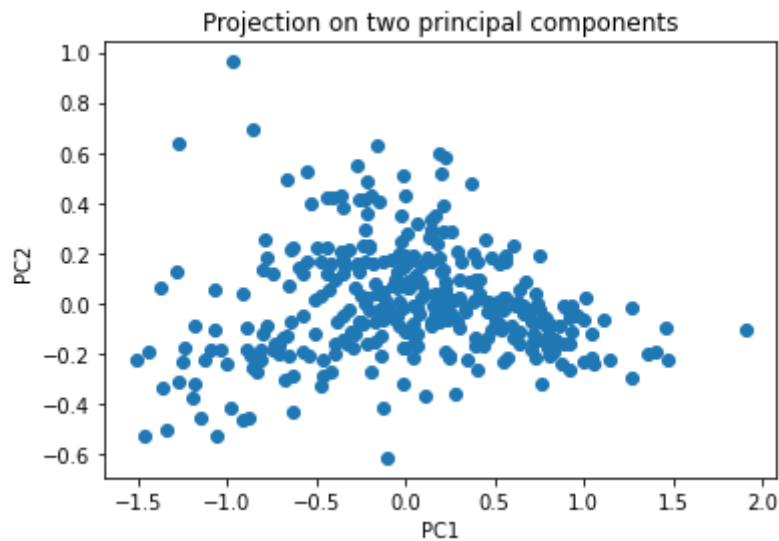
```

```

plt.scatter(projected[:, 0], projected[:, 1])
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.title('Projection on two principal components')
plt.show()

```





```
PC1_outliers = np.where(projected[:, 0]>1.5)
PC1_outliers = np.squeeze(PC1_outliers)
print(PC1_outliers)
```

```
PC2_outliers = np.where(projected[:, 1]>0.63)
PC2_outliers = np.squeeze(PC2_outliers)
print(PC2_outliers)
```

```
↳ 212
   [ 50 119 194]
```

Outliers observed from the above plot

```
outliers = []
outliers.append(data.index[212])
outliers.append(data.index[50])
outliers.append(data.index[119])
outliers.append(data.index[194])
print(outliers)
```

```
↳ ['New-York,NY', 'Brownsville-Harlington,TX', 'Glens-Falls,NY', 'Midland,TX']
```

```
eigenvalues = pca.explained_variance_
print(eigenvalues)
```

```
↳ [0.37746236 0.05105221 0.02791958 0.02296708 0.01677125 0.01195269
    0.0084567  0.00393422 0.00179733]
```

```
from numpy import diag
sq_values = np.sqrt(eigenvalues)
values_diag = diag(sq_values)
cc = np.dot(pca.components_.T, values_diag)
print(cc.T[0])
print(cc.T[1])
```

```
↳
```

```
[0.02154807 0.05735332 0.2505222 0.06171159 0.09221663 0.01975427
 0.53717706 0.09768404 0.01197683]
[ 0.00200601 0.00208562 -0.19398302 0.04980416 0.01337634 -0.01368983
 0.06864424 0.07546474 0.0126759 ]
```

We are only interested in the top correlations with the first and second principal components, which column of cc and select variables with highest absolute values.

1. For the first component:

From cc.T[0] we can see that PC1 appears to correlate the most with arts feature with a

2. For the second component:

From cc.T[1] we can see that PC2 appears to correlate the most with Healthcare feature

## 4f. PCA with standardizing

```
data = pd.read_csv('places.txt',delim_whitespace=True,na_values='?')
table = data[['Climate', 'HousingCost', 'HlthCare', 'Crime', 'Transp', 'Educ', 'Arts', 'Rec
```

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(table)
mean = scaler.mean_
table = scaler.transform(table)
```

```
mean = np.mean(table,axis = 0)
print(mean)
```

```
↳ [ 2.26768958e-16  3.88746786e-16 -3.23955655e-17  1.29582262e-16
 -2.91560089e-16  2.64563785e-16  1.07985218e-17 -1.61977827e-17
 -3.72549003e-16]
```

```
std = np.std(table,axis = 0)
print(std)
```

```
↳ [1. 1. 1. 1. 1. 1. 1. 1. 1.]
```

```
from sklearn.decomposition import PCA
pca = PCA()
pca.fit(table)
print(pca.components_.shape)
```

```
↳ (9, 9)
```

```
pa1 = pca.components_[0]
pa2 = pca.components_[1]
print(f"First principal axis = {pa1}")
print(f"Second principal axis = {pa2}")
```

```
↳ First principal axis = [0.20641395 0.35652161 0.46021465 0.28129838 0.35115078 0.2752
    0.46305449 0.32788791 0.13541225]
    Second principal axis = [ 0.21783531  0.250624   -0.29946528  0.35534227 -0.17960448
    -0.19478992  0.38447464  0.47128328]
```

```
projected = pca.fit_transform(table)
print(f"First principal component = \n {projected[:, 0]}")
print(f"Second principal component = \n {projected[:, 1]}")
```

```
↳
```



First principal component =

```
[ -1.04176435e+00  4.40483533e-01 -1.87839614e+00  9.12128688e-01
  2.15252133e+00 -1.79068458e+00 -1.05705605e+00 -6.49308564e-01
 -1.80698850e+00 -5.00327920e-01  2.97757642e+00  4.48387275e-01
 -2.32871527e+00 -1.81823707e+00  1.87279151e+00 -2.00908458e+00
 -1.05555704e+00 -2.58499423e-01 -1.48905552e+00  3.02961694e+00
  1.37122389e+00 -1.02856279e+00 -4.96355523e-01  2.45528167e-03
 -4.56453278e-01  4.30390469e+00 -3.35788992e-01 -4.79727748e-01
 -1.12406189e+00 -9.83669345e-01 -1.83352844e+00 -7.64643770e-02
 -1.51418554e+00  2.82394680e+00 -3.55080324e-02 -1.37552095e+00
 -4.20250116e-01  2.67813903e-01 -1.57998708e+00 -1.28456459e+00
 -2.23237082e-01 -3.78309402e-01  6.31065516e+00  2.03324315e+00
 -9.21314225e-01 -1.44916582e+00 -1.13458718e+00  2.25647837e+00
 -8.74523291e-01 -1.48039999e-01 -1.61551967e+00 -7.39241315e-01
  1.87933107e+00 -2.00429375e+00  1.62177989e+00 -1.29296999e+00
 -9.93463453e-01 -9.60813355e-01  7.41296452e-01  9.86604182e-01
 -4.63352388e-01  6.93996248e-01  1.08467560e+00 -1.54025500e+00
  6.47475906e+00 -3.66313337e-01  1.82633616e+00 -1.71053621e+00
  3.59581862e+00 -1.69724310e-02 -3.95639805e-01  6.42073628e-01
 -1.58745238e+00  9.07138196e-01 -5.66519308e-01 -1.29682346e+00
  3.42122682e+00  1.49839943e+00 -2.24207876e+00 -5.90644608e-01
  6.02348941e-01  5.23366362e-01 -1.39359647e+00  3.44166678e+00
 -1.37456922e-01  3.01256732e+00 -2.79309815e+00 -1.62873097e+00
 -1.01482404e+00 -3.55037492e-01 -1.72643599e+00 -7.80571606e-01
 -1.93864472e+00 -1.35305628e+00 -1.42555877e+00 -7.86224125e-01
  9.25976304e-01 -9.29155577e-01 -4.15035291e-01 -1.20261108e+00
 -7.44252524e-01 -1.69355558e+00 -1.81435687e+00 -4.35638913e-01
 -2.65113962e+00 -4.54730410e-01 -8.94018467e-03  1.17101568e+00
  1.97014906e-01  1.48096683e-01 -1.65070821e+00 -1.20465895e+00
 -5.44534684e-01  7.72210235e-01  4.39651767e-01 -2.91984770e+00
  2.94783991e-01  7.46153137e-01 -1.91340773e-02 -2.00883516e+00
 -1.36147209e+00 -4.21787936e-01 -1.19550321e+00 -1.53584268e+00
 -9.87359990e-01  4.75086775e-01  1.08130271e-01 -1.01581023e+00
 -6.55010216e-01  7.78898853e-01  2.84582706e+00 -1.59425875e+00
  2.85703854e+00 -2.29877578e+00  2.54651818e+00 -1.41361256e+00
 -1.64366335e+00  9.15341489e-01 -1.44001802e-01 -9.08920700e-01
  2.34553627e-01  7.37306787e-01 -1.64804737e+00 -1.52163975e+00
  7.95905811e-01 -9.46659857e-01 -1.82120435e+00 -1.55533010e-01
 -2.14728658e+00  2.67921759e-01 -1.26300001e+00  3.01067764e-01
  1.32818222e+00  1.89747813e-01 -1.63923402e+00 -1.36512564e-01
 -2.40380405e+00 -2.49638764e-01 -7.57844866e-01 -2.35506312e-01
 -1.32333592e+00  5.68302409e-01 -7.57746339e-01 -7.68310859e-01
  1.69520547e-01 -1.78816695e+00 -5.85786652e-01  1.35598074e+00
 -9.51907591e-01 -1.16028849e-01 -1.62340250e+00 -1.44013632e+00
  5.10221202e-01 -1.21186326e+00 -1.40588660e-01 -1.24287854e-01
 -1.51666874e+00 -1.02836585e+00  7.25734433e+00  9.85045767e-01
 -5.30689382e-01  5.93397133e-02 -7.82325876e-01 -1.70076192e+00
  1.14564332e+00 -9.30772446e-01 -1.47134438e+00 -2.57560345e+00
 -1.10635190e+00  2.02313275e-01  1.20009115e+00  3.65136374e+00
  2.14153745e+00  3.62085225e-01 -1.09981474e-01  2.18794309e+00
  2.51106754e+00 -2.17306021e-01 -1.44156038e+00  1.92738577e+00
 -1.60414951e+00 -1.70338685e+00 -9.16896069e-01 -5.13966551e-01
 -5.39656982e-01  6.72312594e-01  2.95121899e+00 -1.16265654e+00
  3.86124030e-01  2.26222194e+00  3.69758463e-01  2.62091115e+00
  1.24451792e+01  4.13906526e+00 -7.82142883e-01  1.38428577e+00
  3.02410080e+00  3.25167512e+00 -2.68797776e-01 -9.73478242e-01
  9.09812717e-01 -9.21812963e-01  1.07967830e+00  4.32406371e-01
  1.35896935e+00 -2.06368018e+00  3.09447803e-01 -1.13907637e+00
 -1.58540879e+00 -2.73842204e+00 -6.66542805e-02 -3.24254070e-01
 -1.05245061e+00  4.77262828e+00  1.24925216e+00 -2.33192752e+00
  2.69802893e+00 -8.52403317e-01  5.89913978e-01  2.47126717e+00
```

```

-9.35939773e-01 -5.77147483e-01 2.20551991e+00 -6.21485677e-01
-1.18106351e+00 -5.23591402e-01 2.62189843e+00 -7.39871134e-01
-2.80238551e-01 9.86107539e-01 -1.67896655e+00 1.79236589e+00
9.37285638e-01 -4.47033638e-01 -7.08443524e-01 2.04036596e+00
-1.29686827e+00 1.02283685e+00 -7.47715733e-01 -1.93670631e+00
-1.63514499e+00 3.01215474e+00 2.27493515e-01 3.32529932e-01
1.08321724e+00 1.46265558e+00 -1.37044969e+00 3.34850517e-01
3.67712462e+00 7.40266214e+00 2.94550853e+00 2.22070703e+00
1.34173252e+00 8.38665753e-01 -2.20380728e-01 4.50249926e-01
-8.01244642e-01 3.99021285e+00 -2.02197001e+00 -1.87721504e+00
-2.00955673e+00 -2.10301615e-02 -1.43035125e+00 -1.05211084e+00
-1.74216416e-01 -5.06013925e-01 8.72367610e-01 7.36201048e-01
-9.53694653e-01 3.45134930e+00 -1.13335135e+00 -2.45190983e+00
-6.19787271e-01 1.04988206e+00 1.04090329e+00 -1.95059103e-01
1.49865560e+00 -1.49900125e+00 -2.22760577e+00 5.05113203e-01
-4.24369002e-01 1.46841226e+00 1.20346221e+00 2.44072083e-01
-1.39090038e+00 -7.86510575e-01 -1.11144340e+00 3.74123978e-01
-9.23947445e-01 -1.61231987e+00 -4.60063989e-01 -1.12340818e+00
-1.28446737e+00 6.19589473e+00 -7.26092856e-01 -1.00098878e+00
-1.77842069e+00 2.20593284e+00 -1.82625009e+00 -2.97674980e-01
-1.34237442e+00 -1.88952256e+00 8.62551585e-01 -9.03023290e-03
-1.41197659e-01 -1.13434894e+00 -1.71080288e+00 -7.66476745e-01
-1.72858951e+00]

```

Second principal component =

```

[ 8.95130384e-01 7.51805200e-02 6.98980010e-02 -1.82035075e+00
3.29359005e-01 -7.82391621e-01 -1.07759473e+00 -7.60597247e-01
-1.21532481e+00 1.02441832e-01 2.01834512e+00 2.23021777e+00
-8.10935744e-01 -8.19399856e-01 -1.08764164e+00 6.19758044e-01
-3.69075831e-01 -8.76777282e-01 6.53697547e-02 -9.32394753e-01
3.05361972e+00 -8.35863319e-01 6.63745153e-01 1.31601459e-01
1.45415756e+00 -9.69494406e-01 -7.73527421e-01 1.09389957e+00
-2.68137236e-01 6.34035637e-01 -1.08281644e+00 8.42787211e-01
-1.62251194e-02 -1.03828874e+00 -2.83984361e-01 8.49329631e-01
-8.32393775e-01 -8.16225630e-01 -7.57402083e-01 -2.42729713e-01
-4.01631801e-01 1.28767812e-01 -1.61119210e+00 2.11748506e+00
1.12373610e+00 3.24147167e-01 1.47821304e+00 6.48943515e-01
6.96876830e-02 1.03701647e+00 1.02416228e+00 8.30249578e-01
-1.10375211e+00 1.33676976e-01 -6.22127941e-01 -5.78549442e-01
9.71185603e-01 -5.21776414e-01 -7.30668017e-01 6.33014212e-02
-8.42599175e-01 -4.03591022e-01 -9.93816026e-01 1.30500550e-03
-3.09183688e+00 -2.18753663e-01 -6.51256266e-01 2.73657350e-01
-8.69862170e-01 1.70834684e+00 1.12322589e-01 -2.16323134e-01
-5.21183970e-01 -4.66783997e-01 8.03043090e-01 -2.01714583e+00
1.19679766e-01 1.54409826e-01 -1.69467470e+00 -5.08274022e-01
-1.13696284e+00 2.33413529e+00 -6.14931460e-01 -4.40643995e-02
-4.54203867e-02 -1.40270895e+00 1.86864696e-01 -9.31238851e-01
-1.22243726e+00 3.86985272e-01 -1.38062975e+00 5.41247095e-01
-8.98647010e-01 -9.62660213e-01 1.07292415e+00 -7.32148807e-01
-3.29052417e-01 -2.01339252e-01 6.60870080e-01 -1.02968363e+00
-8.38543386e-01 5.91694587e-04 -5.14076207e-01 1.22637269e-01
-3.47542373e-01 -9.15049634e-01 2.47002350e+00 2.03597499e+00
1.70566036e+00 1.85044528e+00 -7.04962974e-01 4.80393963e-01
-1.15134489e+00 1.31065143e+00 2.11354446e+00 -5.25772681e-01
-3.57923055e-01 1.38792530e+00 -6.96621161e-01 1.32697818e-01
-1.54870003e+00 -1.51093873e-01 -2.85193324e-01 -1.49571599e-01
-1.40365380e-01 -8.45094532e-01 -8.10712072e-01 -9.26883345e-01
-3.24926421e-01 -1.71202548e+00 -1.06948119e+00 -5.57930390e-01
1.96177101e+00 1.90854112e+00 -1.44579785e-01 -6.79955192e-01
1.11036175e+00 -6.92594127e-01 -7.23615884e-01 -4.00925057e-01
-1.30959123e+00 1.99761196e-01 6.70001160e-01 -1.43927157e-01
-1.30192398e-01 -5.65284665e-01 -1.40641917e+00 -7.14606048e-01
-6.96679467e-01 -2.59202544e-01 -1.90252930e-01 -4.21261005e-01

```

```

-5.37729759e-01 -2.62909644e-01 -8.59628338e-01 1.82945874e-01
-1.63318354e+00 -8.54344121e-01 -1.26071912e+00 2.54572339e+00
1.14725336e+00 1.08791733e+00 3.64412337e-01 -1.26533646e+00
-1.05240576e+00 -5.75071177e-01 6.42982156e-01 3.34415199e+00
-3.75817073e-02 1.25858113e+00 1.21652143e+00 -3.37573545e-01
9.50657412e-02 -5.01204232e-01 -9.28420339e-01 7.89772263e-02
3.85807106e-01 -8.94190247e-01 4.77878392e-01 -8.76272890e-01
7.09683019e-01 1.21284946e+00 -5.79857434e-01 -1.84271675e-01
-1.25462091e+00 1.23989300e-01 -3.90812486e-01 -4.24319811e-01
-5.89036892e-01 2.32699565e+00 -2.96675428e-01 2.54541233e+00
-1.32247530e+00 6.05964543e-02 3.03460743e+00 -1.04857712e+00
-1.73772079e+00 6.90197948e-02 6.69060555e-01 1.08527545e+00
4.22322460e-01 -1.27679267e-01 -1.47553675e+00 6.83487615e-01
6.00623633e-01 -1.49761762e-02 3.79945810e-01 7.56648166e-01
-2.60485406e-01 -1.49810998e+00 2.44061761e-01 8.37818144e-01
-2.06488660e+00 -7.90926230e-01 -1.31639557e-01 3.42992540e-01
8.85665919e-01 1.47717294e+00 1.39489625e+00 1.89665665e+00
9.46257974e-01 3.24346041e-01 -1.07337386e+00 4.81804817e-02
1.55343107e+00 1.13877129e-01 2.47283776e+00 1.57537259e+00
-2.56194460e-01 1.38350743e+00 -6.33400374e-01 9.23642796e-01
-9.09033062e-02 -3.07796833e+00 6.11955435e-01 -2.73812744e-01
-2.45964805e+00 -1.05458215e+00 5.27203971e-01 -2.01878330e-01
7.97011728e-01 4.28760951e-01 -1.53045430e+00 -4.82474293e-02
9.73022024e-02 4.22705922e-02 -1.57692775e+00 -5.84522742e-01
6.94263764e-01 1.22746454e+00 1.61694283e+00 -5.85441400e-01
6.37931124e-01 2.28035915e-01 -1.91704833e-01 -9.87882625e-01
-3.49441923e-01 6.66231667e-01 -4.27019194e-01 -8.44176868e-01
-6.88925050e-02 -1.60752687e+00 -8.76249898e-01 7.95045803e-01
2.87799507e+00 1.08842107e+00 1.06495335e+00 -1.70524706e-01
1.40571193e+00 1.22886866e+00 1.90889911e+00 2.39512291e+00
1.25674876e+00 1.33334621e+00 1.67655415e+00 1.11155073e+00
-1.69905975e+00 1.58968765e+00 -1.45286841e+00 -5.10390272e-01
5.32677975e-01 7.96625613e-02 -9.07828619e-01 -1.15696536e+00
-5.21809497e-01 -5.73469413e-01 -8.19627764e-01 -1.38244255e+00
3.07846964e-01 1.08992065e+00 -1.12665083e+00 -8.03220370e-01
1.84981584e+00 -1.17801862e+00 1.86122586e+00 3.39773452e-01
9.35758537e-01 -1.23158235e+00 -9.71465080e-01 -8.53678107e-01
-5.59086659e-01 -3.09032076e-01 1.51030164e+00 9.63961828e-01
-9.90738787e-02 -3.25676175e-01 -8.34416685e-01 1.25579170e+00
7.23051489e-01 1.14093598e+00 7.40134250e-02 2.16696638e+00
1.21145604e-01 -2.26832376e+00 -2.54608572e-01 -1.03741434e+00
-1.23450513e+00 2.49801998e+00 -1.27850466e+00 -4.85748532e-02
6.52881847e-01 -1.85784304e+00 -9.08932836e-02 8.64934812e-01
-1.61161109e+00 7.10956991e-01 -9.06311170e-01 -1.07696175e+00

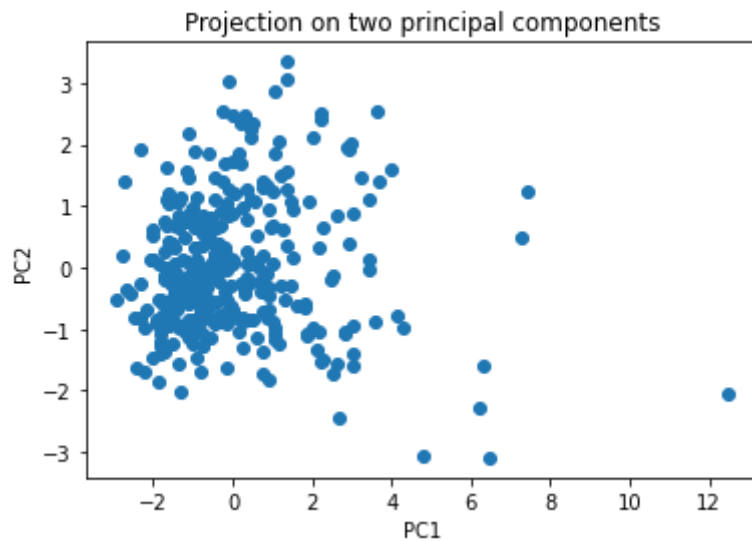
```

```

plt.scatter(projected[:, 0], projected[:, 1])
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.title('Projection on two principal components')
plt.show()

```





Outliers observed from the above plot

```
PC1_outliers = np.where(projected[:, 0]>12)
PC1_outliers = np.squeeze(PC1_outliers)
print(PC1_outliers)
```

```
PC2_outliers = np.where(projected[:, 1]<-3)
PC2_outliers = np.squeeze(PC2_outliers)
print(PC2_outliers)
```

```
↳ 212
   [ 64 233]
```

```
outliers = []
outliers.append(data.index[212])
outliers.append(data.index[64])
outliers.append(data.index[233])
print(outliers)
```

```
↳ ['New-York,NY', 'Chicago,IL', 'Philadelphia,PA-NJ']
```

```
eigenvalues = pca.explained_variance_
print(eigenvalues)
```

```
↳ [3.41868293 1.21767731 1.14495927 0.9237255  0.75558148 0.63248434
    0.49455091 0.31900812 0.12076916]
```

```
from numpy import diag
sq_values = np.sqrt(eigenvalues)
values_diag = diag(sq_values)
cc = np.dot(pca.components_.T, values_diag)
print(cc.T[0])
print(cc.T[1])
```

```
↳
```

```
[0.38165281 0.65919707 0.85092219 0.52011172 0.64926658 0.50900729
 0.85617297 0.60625427 0.25037293]
[ 0.24037781 0.27655962 -0.3304552 0.39211457 -0.1981907 -0.53340449
-0.21494759 0.42426168 0.52005364]
```

We are only interested in the top correlations with the first and second principal components, which column of cc and select variables with highest absolute values.

1. For the first component:

From cc.T[0] we can see that PC1 appears to correlate the most with arts feature with a

2. For the second component:

From cc.T[1] we can see that PC2 appears to correlate the most with Education feature

The consistent outlier found in cases of scaling(mean = 0) and standardizing(mean=0, var = 1) is s

Principal component analysis will tend to give more emphasis to those variables that have high variances. In effect the results of the analysis will depend on the units of measure each variable. That would imply that a principal component analysis should only be used if all variables have the same units of measure. And even in this case, only if you wish to give those variables with higher variances more weight in the analysis.

From the above reasoning we can conclude that it is necessary to standardize the variables before the above solution we can see that the outliers are different in both the cases, except for the outlier

From the above plots we see that the results for standardizing and not standardizing vary. PCA changes the plot this in turn changes the outliers.

