

```
In [1]: import numpy as np
import pandas as pd
import openpyxl
import matplotlib as mp
import matplotlib.pyplot as plt
import sklearn as sl
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.manifold import TSNE
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split
from sklearn.svm import SVR
from sklearn.model_selection import RandomizedSearchCV
from sklearn import neighbors
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.linear_model import Lasso
import xgboost as xg
from sklearn.kernel_ridge import KernelRidge
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.metrics import mean_squared_error
import math
```

```
In [2]: AA="C:/Users/ganes/onedrive/Desktop/AI/EC-CO2 REG/SAC-Data.xlsx"
df=pd.read_excel(AA)
df.head(5)
df=pd.read_excel(AA)
df.head(5)
df.shape
```

Out[2]: (480, 26)

```
In [3]: A=df['NCNF']
B=df['NCNT']
C = df['NG']
D = df['NC']
E = df['Ag']
F=df['Bi']
G=df['Co']
H = df['Cu']
I = df['Fe']
J=df['La']
K=df['Mg']
L = df['MnO2']
M=df['Ni']
N= df['Sn']
O= df['Sb']
P=df['Pd']
```

```

Q=df["Zn"]
R=df["SAC"]
S=df["PT"]
T=df["KE"]
U=df["NE"]
V=df["VO"]
List = [A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V]
List1=pd.concat(List, axis=1, sort=True)
List1.head(5)
List1.shape

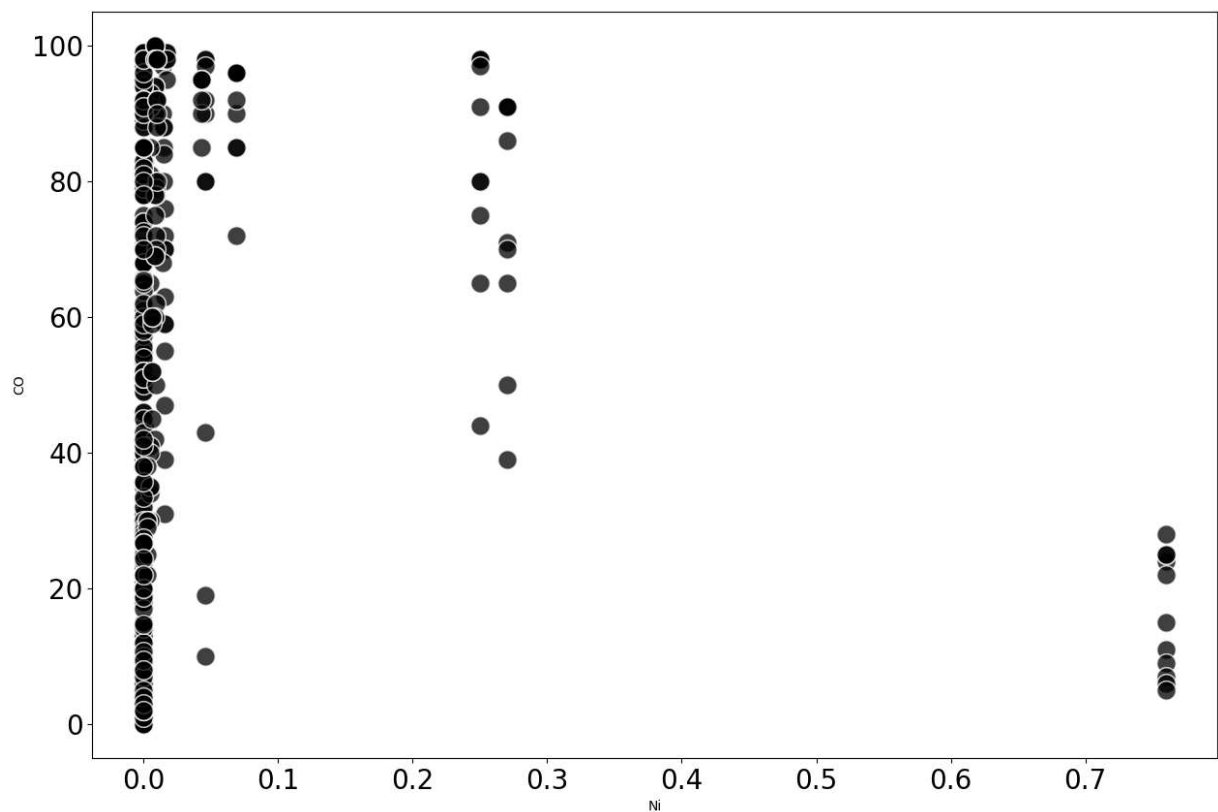
```

Out[3]: (480, 22)

```

In [4]: import seaborn as sns
plt.figure(figsize=(15,10))
sns.scatterplot(x=df["Ni"], y=df["CO"], color="black", s=200, alpha=0.75)
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.show()

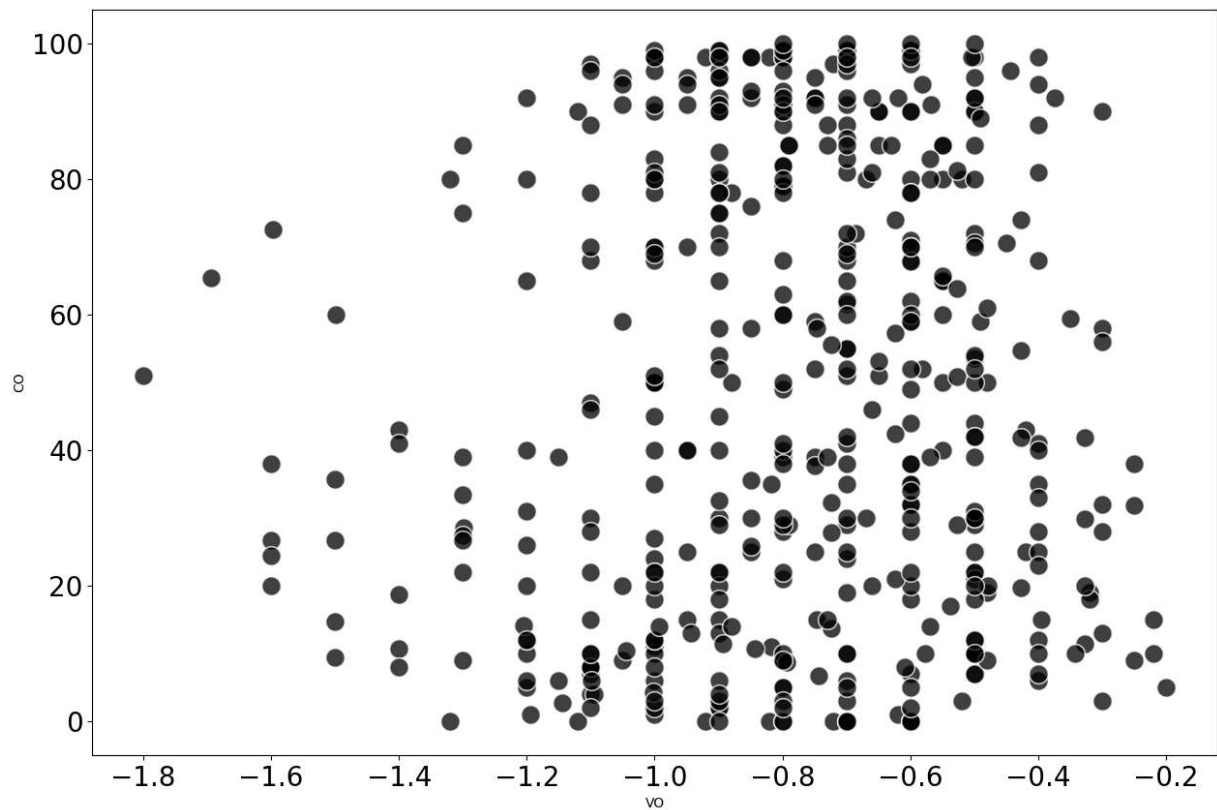
```



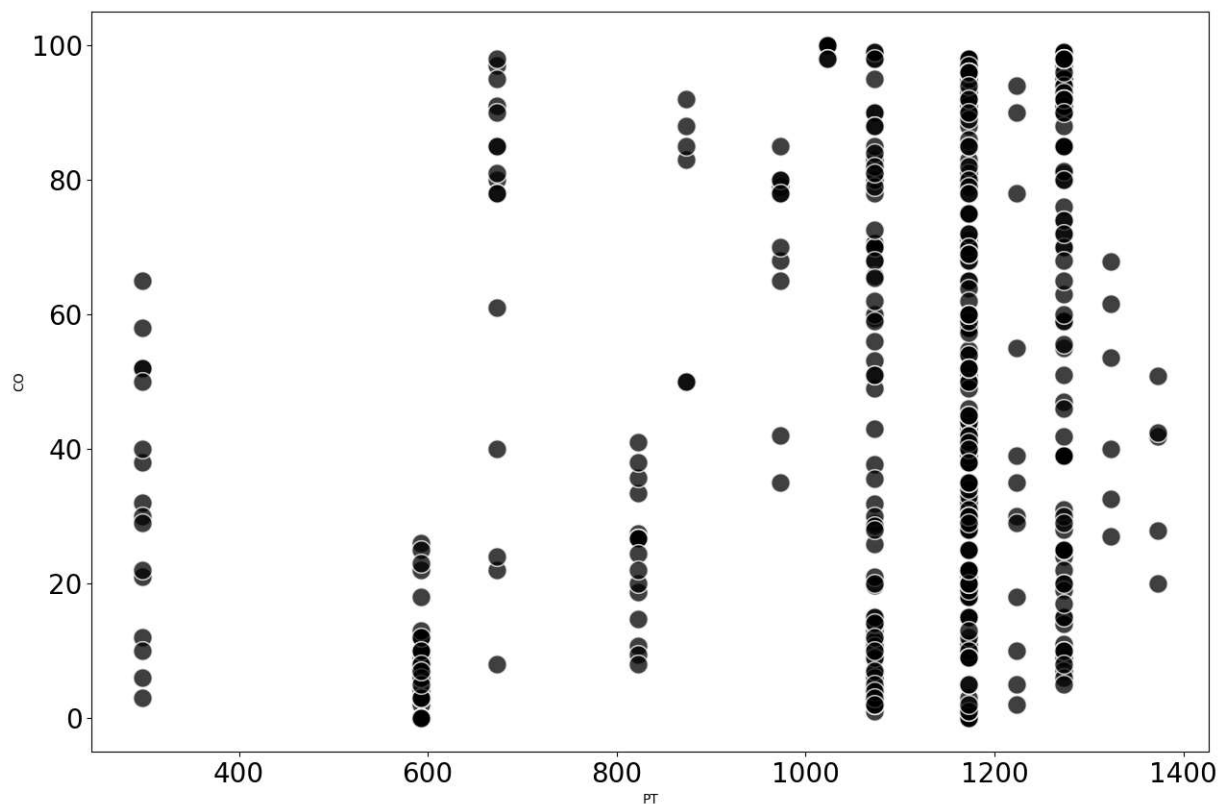
```

In [5]: import seaborn as sns
plt.figure(figsize=(15,10))
sns.scatterplot(x=df["VO"], y=df["CO"], color="black", s=200, alpha=0.75)
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.show()

```



```
In [6]: import seaborn as sns
plt.figure(figsize=(15,10))
sns.scatterplot(x=df["PT"], y=df["CO"], color="black", s=200, alpha=0.75)
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.show()
```



```
In [7]: from sklearn import preprocessing
AB = preprocessing.StandardScaler()
BC=AB.fit_transform(List1)
CD=pd.DataFrame(BC, columns=List1.columns[:])
CD.head(5)
```

```
Out[7]:
```

	NCNF	NCNT	NG	NC	Ag	Bi	Co	Cu
0	-0.197381	-0.225651	-0.45385	0.689075	-0.208514	-0.160128	-0.269681	-0.234838
1	-0.197381	-0.225651	-0.45385	0.689075	-0.208514	-0.160128	-0.269681	-0.234838
2	-0.197381	-0.225651	-0.45385	0.689075	-0.208514	-0.160128	-0.269681	-0.234838
3	-0.197381	-0.225651	-0.45385	0.689075	-0.208514	-0.160128	-0.269681	-0.234838
4	-0.197381	-0.225651	-0.45385	0.689075	-0.208514	-0.160128	-0.269681	-0.234838

5 rows × 22 columns

```
In [8]: from sklearn import preprocessing
AB = preprocessing.MinMaxScaler()
BC=AB.fit_transform(List1)
CD=pd.DataFrame(BC, columns=List1.columns[:])
CD.head(5)
```

```
Out[8]:
```

	NCNF	NCNT	NG	NC	Ag	Bi	Co	Cu	Fe	La	...	Ni	Sn	Sb	Pd	Zn
0	0.0	0.0	0.0	0.9847	0.0	0.0	0.0	0.0	0.0	0.0	...	0.020132	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.9847	0.0	0.0	0.0	0.0	0.0	0.0	...	0.020132	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.9847	0.0	0.0	0.0	0.0	0.0	0.0	...	0.020132	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.9847	0.0	0.0	0.0	0.0	0.0	0.0	...	0.020132	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.9847	0.0	0.0	0.0	0.0	0.0	0.0	...	0.020132	0.0	0.0	0.0	0.0

5 rows × 22 columns

```
In [9]: import seaborn as sns
```

```
In [10]: sse = []
for E in range(1,21):
    kmeans = KMeans(n_clusters=E)
    kmeans.fit(List1)
    score = kmeans.score(List1)
    sse.append([E, kmeans.inertia_])

plt.figure(figsize=(12,9))
plt.xticks(range(1, 21, 2))
plt.rc('xtick', labels=20)
plt.rc('ytick', labels=20)
```

```
plt.tick_params(labelsize=16)
plt.plot(pd.DataFrame(sse)[0], pd.DataFrame(sse)[1], "k", marker="o")
```

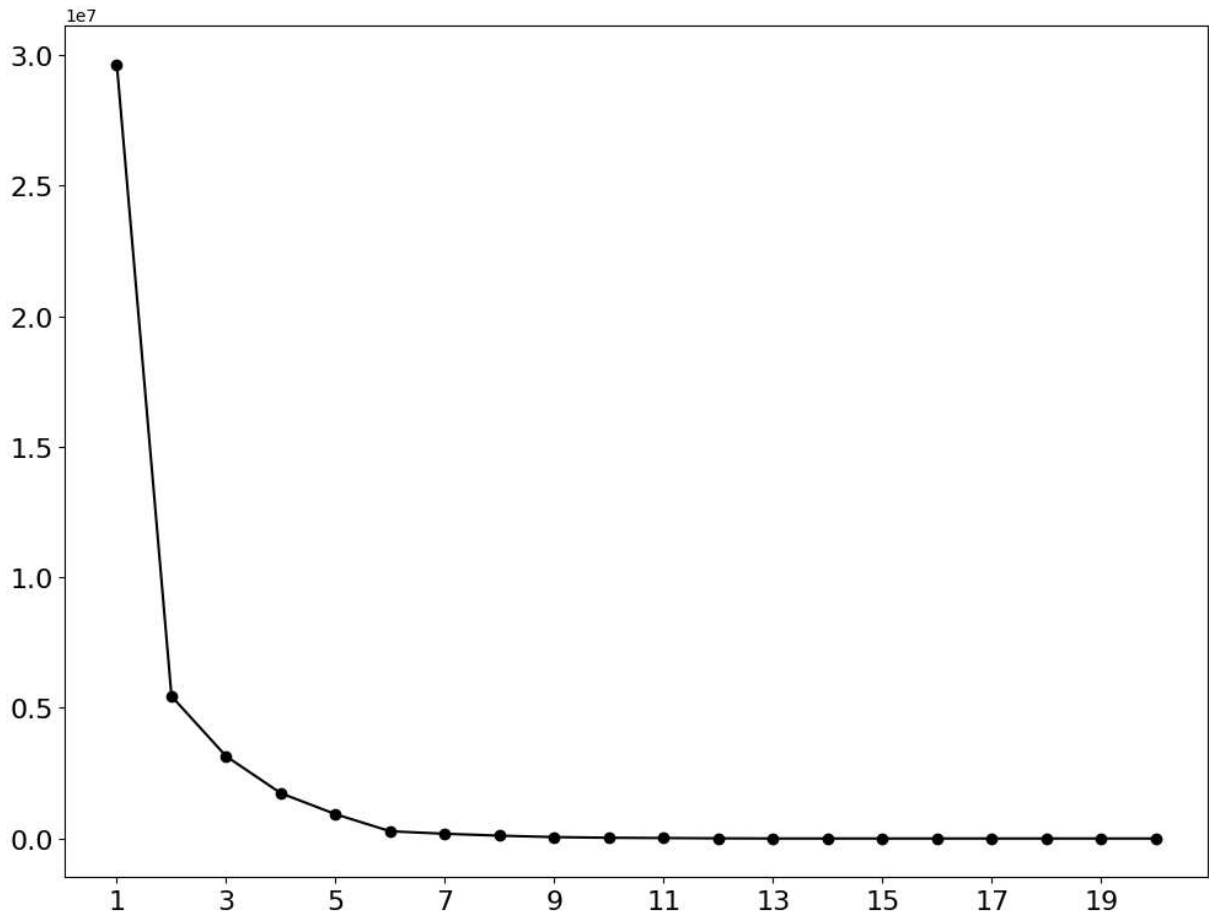
[illegible]


```

C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)

```

Out[10]: [`<matplotlib.lines.Line2D at 0x21be38c8210>`]



```

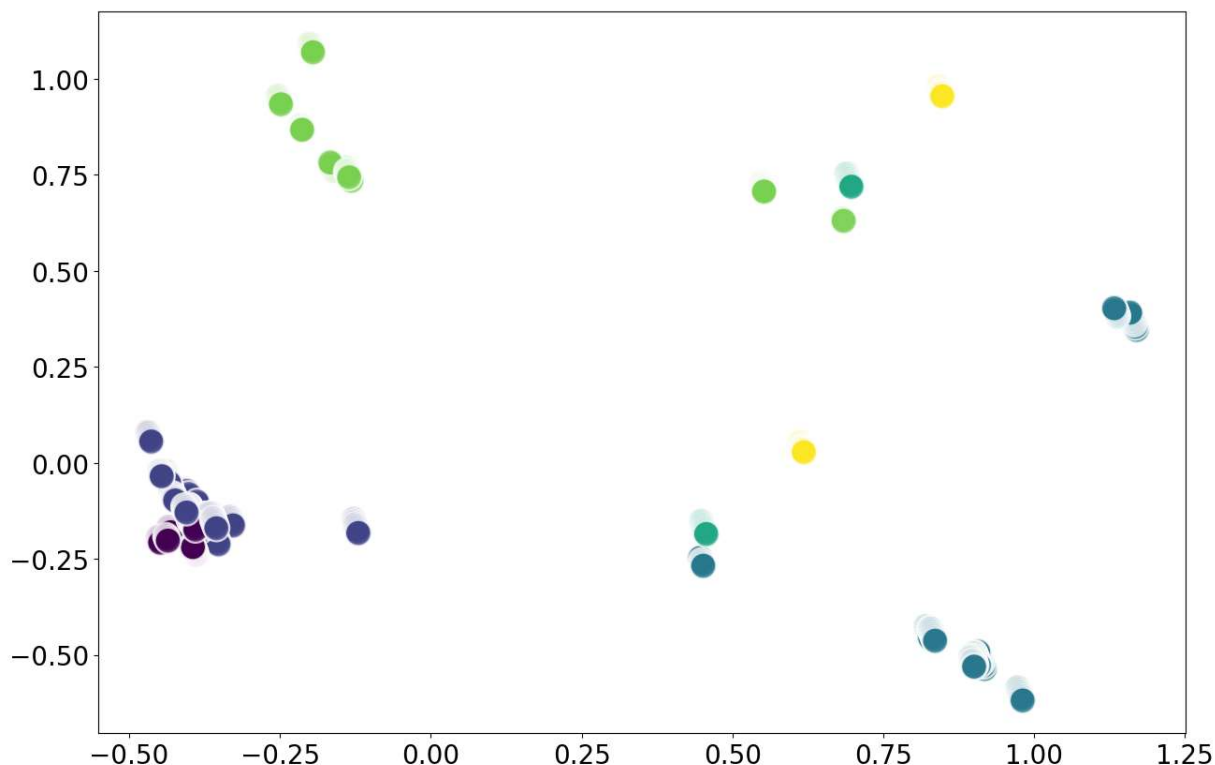
In [19]: from sklearn.decomposition import PCA
         cluster = KMeans(n_clusters=6)

```

```
cluster.fit(CD)
cluster_predictions = cluster.predict(CD)
pca = PCA(n_components=2)
pca.fit(CD)
x_pca = pca.transform(CD)
plt.figure(figsize=(15,10))
sns.scatterplot(x=x_pca[:,0],y=x_pca[:,1], c=cluster_predictions, s=400, alpha=0.75)
```

C:\Users\ganes\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster_kmeans.py:1416: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
super()._check_params_vs_input(X, default_n_init=10)

Out[19]: <Axes: >



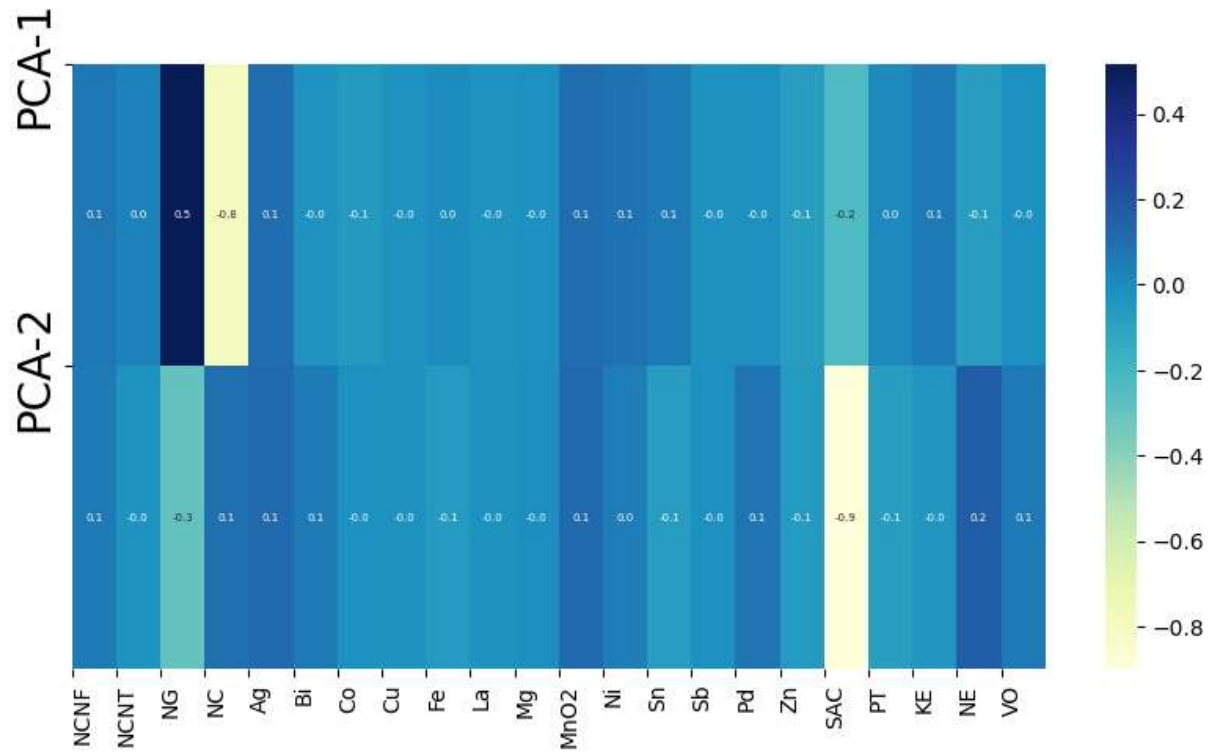
```
In [20]: pca_ep_Df = pd.DataFrame(data = x_pca
                                   , columns = ['principal component 1', 'principal component 2'])
print('Explained variation per principal component: {}'.format(pca.explained_varian
```

Explained variation per principal component: [0.28557293 0.16658367]

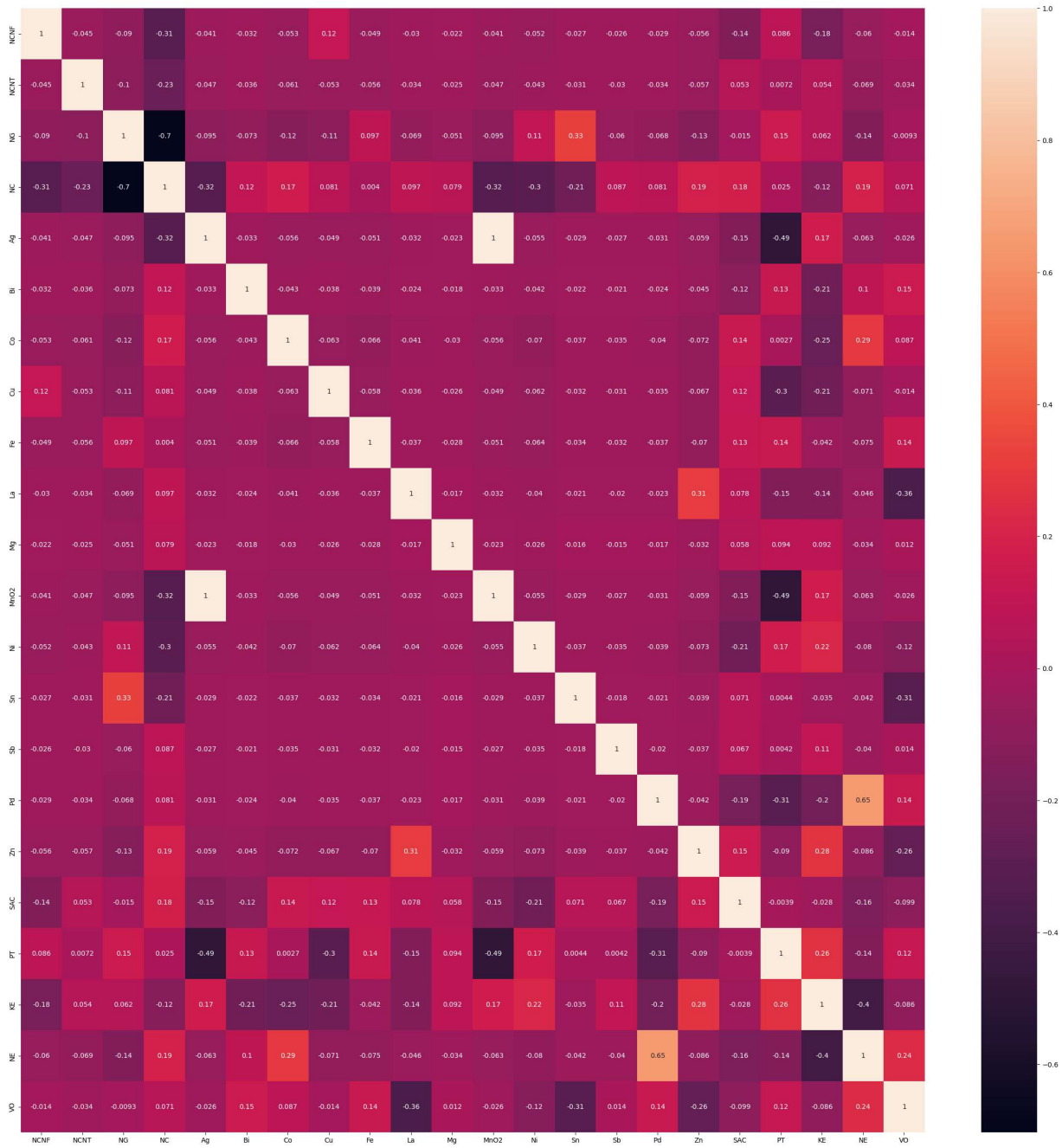
```
In [30]: plt.figure(figsize=(10,5))
sns.heatmap((pca.components_ ), annot = True, annot_kws={'size': 5}, fmt=".1f", cma

plt.rc('xtick', labels=10)
plt.rc('ytick', labels=10)
plt.xticks(range(len(List1.columns)), List1.columns, rotation=90, ha='left')
plt.yticks([0,1],[ 'PCA-1', 'PCA-2'], fontsize=20)
```

Out[30]: ([<matplotlib.axis.YTick at 0x21bec1aef50>, <matplotlib.axis.YTick at 0x21bec1b5910>], [Text(0, 0, 'PCA-1'), Text(0, 1, 'PCA-2')])



```
In [31]: corr=CD.corr()
plt.subplots(figsize=(30, 30))
sns.heatmap(corr,annot=True, annot_kws={'size': 10})
plt.rc('xtick', labels=10)
plt.rc('ytick', labels=10)
```



In []: