DataFrame

import numpy as np

arr1=['A']\*5 + ['B']\*5 +['C']\*5+['D']\*5+['E']\*5+['F']\*5+['G']\*5+['H']\*5+['I']\*5+['J']\*5

arr2=['国語','数学','英語','理科','社会']\*10

arr3=[np.random.randint(0,100) for i in range(50)]

print(arr1)

print(arr2)

print(arr3)

import pandas as pd

data=list(zip(arr1,arr2,arr3))

df=pd.DataFrame(data=data,columns=['名前','教科','点数'])

df.info()

print()

df[:5]

列表示

df['点数'][:10]

df['総合']=[np.random.randint(0,100)for i in range(50)]

df[:5]

del df['総合']

df.columns=['氏名','教科','中間','期末']

df[:5]

new\_data=[('K','国語',95,89),('K','数学',87,71),('K','英語',69,93),('K','理科',73,52),('K','社会',54,60)]

new\_df=pd.DataFrame(data=new\_data,columns=['氏名','教科','中間','期末'])

df2=df.append(new\_df,ignore\_index=True)

df2[45:]

行に追加データ

(rows,cols)=df2.shape

df2.loc[rows]=['L','国語',15,20]

df2.loc[rows+1]=['L','数学',25,30]

df2.loc[rows+2]=['L','英語',35,40]

df2.loc[rows+3]=['L','理科',45,50]

df2.loc[rows+4]=['L','社会',55,60]

df2.tail(10)

df2.head(10)

インデックス

df2.index=list('abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ12345678')

df2.head(10)

データの取得

print(df2['氏名'].unique())

print()

print(df2['教科'].unique())

Series

s=pd.Series(list('ABCDEFG'))

s[:3]

print(df2['中間'][:5])

print()

print((df2['中間']\*2)[:5])

print()

print((df2['中間']//10)[:5])

統計

print('合計:%s' % df2['中間'].sum())

print('平均:%s' % df2['中間'].mean())

print('中央:%s' % df2['中間'].median())

print('最小:%s' % df2['中間'].min())

print('最大:%s' % df2['中間'].max())

グループ分けに関する情報

grp.groups

ソート

grp2=srt.groupby('氏名')

grp2.sum().sort\_values(['中間','期末'],ascending=False)

aggで集計

grp.agg(['sum','mean','median','min','max'])

条件検索

df2.query('中間 >= 80 & 期末 >=80')

df2.query('中間+期末 >=170')

ピボットテーブル

df2.pivot\_table(values='中間',index=['氏名'],columns=['教科'])