import pandas as pd

import numpy as np

import random

# read data

d=pd.read\_excel("C:\\Users\S.Nivedhitham\\SNivedhitha\\clarivative\\Sample Data.xlsx")

# get unique no. of values in betwen 2-6 clumns

unq=pd.DataFrame(list(set().union(\*[list(df.iloc[:,i].unique()) for i in range(2,7)])))

# ordinal encode betwen 2-6 clumns

df=d.replace({'Neither Agree nor Disagree':'None of them','Strongly Agree':'Strongly Agre','Strongly Disagree':'Strongly Disagre'},regex=True)

df.replace({'Strongly Disagre':-2,'Disagree':-1,'None of them':1,'Agree':2,'Strongly Agre':3},regex=True,inplace=True)

### Custom Encoding: calculate sum of occurances of unique values and divide by count of total occurances of values

## between 9th column to 15th column

ce=pd.concat([(df.iloc[:,i].value\_counts()/len(df))\*random.sample(range(1,df.iloc[:,i].nunique()+1),df.iloc[:,i].nunique()) for i in range(9,len(df.columns))]).sort\_values(0,True)

## replace duplicate values

ce1=[ce[0]]

temp=ce[0]

cnt=2

frq=0

for i in range(1,len(ce)):

if ce[i]!=temp:

temp=ce[i]

ce1.append(ce[i])

cnt=0

else:

cnt+=2

ce1.append(ce[i]+(ce[i]/cnt))

# concat

d1=pd.concat([df.iloc[:,:7],df.iloc[:,9:].replace(to\_replace=ce.index,value=ce1,regex=True)],axis=1)

# fill na

d1.fillna(0,inplace=True)

# performance calculated by calculating the sum of all 5 rating question respesctive to ID/row wise and then buketing them into 4 groups

d['perf']=np.where(d1.iloc[:,2:7].sum(axis=1) < 3, 0, np.where(d1.iloc[:,2:7].sum(axis=1) <8, 1, np.where(d1.iloc[:,2:7].sum(axis=1) <13, 2,3)))

# labeling the groups

d['perf\_desc']=np.where(d.perf == 0, 'Poor', np.where(d.perf ==1, 'Average', np.where(d.perf ==2, 'Good','Excellent')))

# 2019 vs 2020 overall performance

op=pd.DataFrame(d1.groupby(by=["Survey Year","overall\_perf"]).sum()).reset\_index().iloc[:,:7]

# data transform

op1=pd.melt(op,id\_vars=["Survey Year","overall\_perf"])

op2=op1.pivot\_table(index=['Survey Year','variable'],values='value',columns='overall\_perf').reset\_index()

op2['perf\_metric']=op2.iloc[:,2:7].sum(axis=1)

# exporting the results

d.to\_excel("C:\\Users\S.Nivedhitham\\SNivedhitha\\clarivative\\modf\_data.xlsx")

op2.to\_excel("C:\\Users\S.Nivedhitham\\SNivedhitha\\clarivative\\op.xlsx")