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


#regex function for data filteration
import re
def clean_text(text):
    text = text.lower()
    text = re.sub(r"\r", "", text)
    text = re.sub(r"\t", "", text)
    text = re.sub(r"[-()\"#/@;:<>{}`+=~|.!?,]", "", text)
    text= re.sub(r'[0-9]', "", text)
    text= re.sub(r' ', "", text)

    return text


#Reading values,txt
import codecs
with codecs.open("/content/drive/MyDrive/cts-main/values.txt", "rb", encoding="utf-8", errors="ignore") as data:
    line = data.read().split("\n")
    dat = []
    for line in line:
        dat.append(line)

values= []
for line in dat:
    text = clean_text(line)
    values.append(text)

values.pop()

print(len(values))
```

```
!pip install openpyxl==3.0.0


#Reading file 21.xlsx(excel file)
import pandas as pd

df = pd.read_excel (r'/content/drive/MyDrive/cts-main/file 21.xlsx',header=None, index_col=False)
print (df)


df.head()


#Appending to a list
dfto_li=[]

for i in range(12):
  dfto_li.append(df[i].tolist())


dfto_li[0]


#Calling regex function
file_1=[]
for x in range(12):
  for line in dfto_li[x]:
    text = clean_text(line)
    file_1.append(text)


len(file_1)


#Resulting values after filteration
flist=[]
for x in file_1:
  if x in values:
    flist.append(x)
```

```
print("Total matching values found in file 21.xlsx:",len(flist))



#reading combin.txt and filtering
my_file = open("/content/drive/MyDrive/cts-main/combin.txt", "r")
content = my_file.read()
#print(content)
content_list = content.split("\n")
my_file.close()
#print(content_list)

"""
with open("/content/drive/MyDrive/cts-main/combin.txt", 'rb') as f:
  contents = f.read()
  """

file_2= []
for line in content_list:
  text = clean_text(line)
  file_2.append(text)

a=True
while a!= False:
  x=len(file_2)
  i=0
  while i<x:
    if(len(file_2[i]))==0:
      i=file_2.index(file_2[i])
      file_2.pop(i)
      x=x-1
    i=i+1
  if ('' not in file_2):
    a = False
```

```
for x in file_2:
  if x.isalpha()== False:
    i=file_2.index(x)
    file_2[i]=x[3:]

flist2=[]
rlist=[]
for x in file_2:
  if x in values:
    flist2.append(x)
  else:
    rlist.append(x)


final_list=[]
final_list = flist + flist2
print("Total values from both excel sheet and txt files:",len(final_list))

#Creating final_list
textfile = open("/content/drive/MyDrive/cts-main/final_list.txt", "w")
for element in final_list:
    textfile.write(element + "\n")
textfile.close()
```

✓ 0s completed at 13:05