

GUIDLINE TO RUN CATEGORY-MAPPING SCRIPTS

https://colab.research.google.com/drive/1ySAqbQM6DR-SqxK_DHKIlwyzsrFSR65#scrollTo=SulkugyoaLVB

1. UPLOAD THE SENTIMENT DEDUP FILE BY CLICKING ON THE UPLOAD FILE SIGN OF COLAB .HERE IS THE IMAGE OF THE ICON WHICH NEED TO BE CLICKED INORDER TO

UPLOAD THE FILE----> 



2. NAME OF THE SENTIMENT DEDUP FILE SHOULD BE IN SMALL CASE "sentiment.csv"
3. STEP TO PREPARE THE KEYWORD FILE.THE KEYWORD FILE HAVE TWO COLUMN.
4. FIRST COLUMN HEADER IS "Categories" IT WILL HAVE THE LIST OF ALL THOSE UNIQUE CATEGORY NAME WHICH NEED TO BE FORMED.HERE IS LIVE EXAMPLE

	A
1	Categories
2	Footwear
3	Headwear and Neckwear
4	Rainwear
5	Bottoms
6	Shirts and Tops
7	Vests
8	Sweaters, Hoodies, Sweatshirts
9	Jackets, Coats, Parkas, and Bombers

3.THE SECOND COLUMN HEADER IS "KEYWORD".IT WILL HAVE THE LIST OF ALL THOSE KEYWORD OF THAT RESPECTIVE CATEGORY WHICH CAN USED TO IDENTIFY EACH CATEGORY.BASICALLY IT IS THE BAG OF WORD OF EACH CATEGORY.THE KEYWORDS SHOULD BE UNIQUE TO EACH CATEGORY .IF SKU NAME CONTAINS THESE MENTION KEYWORD THEN IT WILL GET MAPPED TO CORRESPOND KEYWORD.HERE IS THE LIVE EXAMPLE FOR YOUR REFERENCES.

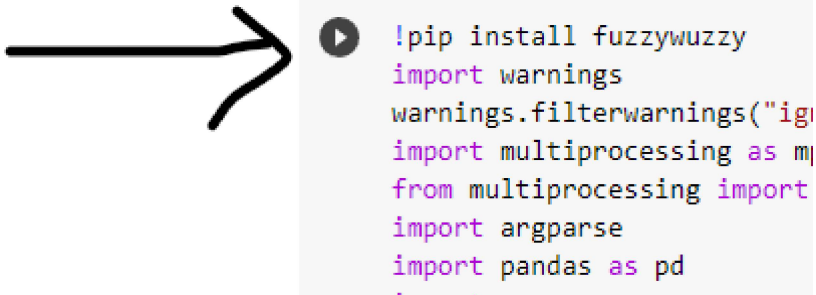
	A	B	C	D	E	F	G	H
1	Categories	Keywords						
2	Footwear	sneaker,sneakers,boot,boots,slider,sliders,sandals,sandal,shoes,shoe,thongs,thong,espadrilles,espadrille,mules,mules,pumps,pump,moccasins,moccasin,loafers,loafer,ballet,flats						
3	Headwear and Neckwear	gloves,glove,scarfs,scarf,hats,hat,beanies,bean,caps,cap,scarves,scarve,mitts,mitt,mitten,mittens,toques,toque						
4	Rainwear	rain,poncho,rains,rain coat,rain coats,raincoat,raincoats,wind breakers,windbreaker,wind threader,windthreader,rainjackets,rainjacket,rain jacket,rain jackets						
5	Bottoms	tracksuits,tracksuit,pants,pant,shorts,leggings,legging,tights,denims,denim,jeans,joggers,jogger,capri,jean,trouser,trousers,bottoms,bottom,skort,skorts,tight						
6	Shirts and Tops	sleeve,tee,tea,shirt,shirts,top,tops,polo,polo shirts,polo-shirt,polo-shirts,t-shirt,t shirt,tshirts,flannels,shirt jackets,tees, jerseys,base layers,mid layer,midlayers,mid-layer,mid-laye						
7	Vests	gilet,vaistcoat,vaist coat,vaist,gilets,waistcoats,vest,vests,waistcoats,waistcoat						
8	Sweaters, Hoodies, Sweat	hoodie,hoodies,sweat hoody,snowsuit,sweaters,hooded,sweatshirts,hood,knitwear,fleece,jumper, jumpers, cardigans,sweater,hoodie,shell, sweaters, hoodies, sweatshirts, knitwear and fleece						
9	Jackets, Coats, Parkas, ani	jacket,overcoat,jackets,coat,coats,capes,puffer,puffers,trench,trenchs,parkas,parka,bombers,bomber,short down jackets,long down jackets,lightweight down jackets,coats & jac						
10	JUNK	face mask,ring,necklace,jewel,jewel,watch,perfume,parfum,colonge,toilete,glasses						
11								

UPLOAD THE KEYWORD FILE BY CLICKING ON THE UPLOAD FILE SIGN OF COLAB . THE NAME OF THE KEYWORD FILE IS IN SMALL CASE "keyword.csv" HERE IS THE IMAGE OF THE ICON WHICH NEED TO BE CLICKED INORDER TO UPLOAD THE FILE---->



7. CLICK ON THE EXECUTOR BUTTON OR ENTER CTRL+ENTER TO RUN THE CODE

HERE IS THE LIVE EXAMPLE FOR REFERNCES



8. AFTER RUNNING THE SCRIPTS IT WILL ASK TO PRESS 1 TO CONTINUE WITH

Press 1 for Category Mapping

9. THEN YOU NEED TO GIVE THE PATH AND NAME OF THE SENTIMENT DEDUP FILE AS MENTION BELOW

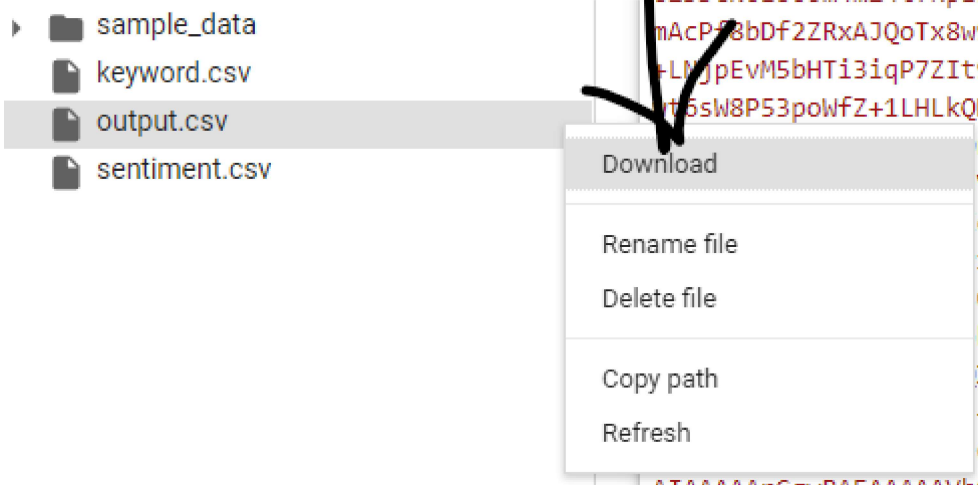
enter the sentiment_deup file out of this
 ['/content/output.csv', '/content/sentiment.csv', '/content/keyword.csv']
 provide the name of sentiment_dedupfile with .csv extension

10. AFTER THAT IT WILL ASK YOU TO GIVE THE FILES NAME ALONG WITH KEYWORD FILE PATH AS MENTION BELOW

Press 1 for Category Mapping
 1
 enter the sentiment_deup file out of this
 ['/content/output.csv', '/content/sentiment.csv', '/content/keyword.csv']
 provide the name of sentiment_dedupfile with .csv extension/content/sentiment.csv
 ['/content/output.csv', '/content/sentiment.csv', '/content/keyword.csv']
 give the keyword file

*** AFTER THE SUCESFULL RUN THE CATEGORY MAPPING OUTPUT FILE WILL AUTOMATICALLY GET DOWNLOADED TO LOCAL SYSTEM***

*** OR IT CAN BE DOWNLOADED BY MANUALLY MENTION BELOW ****



```

!pip install fuzzywuzzy
import warnings
warnings.filterwarnings("ignore")
import multiprocessing as mp
from multiprocessing import Process
import argparse
import pandas as pd
import re
from collections import Counter
import os
import unicodedata
import numpy as np
import itertools
from itertools import combinations
from fuzzywuzzy import fuzz
from fuzzywuzzy import process
import glob

def remove_accented_chars(text):
    new_text = unicodedata.normalize('NFKD', text).encode('ascii', 'ignore').decode('utf-8')
    return new_text

def clean_text(df):
    all_reviews = []
    lines = df["SKU"].values.tolist()
    for text in lines:
        text = text.lower()

        text = text.replace("\\n", "")
        text = text.replace("'s", "")
        text = text.replace("(", "")
        text = text.replace(")", "")
        text = text.replace("\d", "")

```

```

text = text.replace(", ", "")
text = text.replace("\n", "")
text = text.replace(".", "")
# Removing Website Links
pattern = re.compile('http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\(\)\,]|(?:%[0-9a-f
text = pattern.sub('', text)

# Removing emoji's
emoji = re.compile("[
    u"\U0001F600-\U0001FFFF" # emoticons
    u"\U0001F300-\U0001F5FF" # symbols & pictographs
    u"\U0001F680-\U0001F6FF" # transport & map symbols
    u"\U0001F1E0-\U0001F1FF" # flags (iOS)
    u"\U00002702-\U000027B0"
    u"\U000024C2-\U0001F251"
    "]+", flags=re.UNICODE)
text = emoji.sub(r'', text)
text = re.sub(r"'s", "", text)

# Remove Special Characters
text = re.sub(r"[\`!@#$%^&*(){}?;/;`~<>+=]-", "", text)

# Remove accented characters
text = remove_accented_chars(text)
text=text.split(" ")

all_reviews.append(text)

return all_reviews
def category_mapping(data,keyword):
    data["Category"]=" "
    data["Mapped keyword"]="NA"
    for i1,row1 in enumerate(keyword["Keywords"]):
        for word1 in row1.split(","):
            for i2,row2 in enumerate(data["keyword"]):
                if word1 in row2:
                    data.iloc[i2,2]=keyword.iloc[i1,0]
                    data.iloc[i2,3]=word1
    data["Category"].fillna("NA",inplace=True)
    return data
if __name__ == '__main__':
    print("Press 1 for Category Mapping")
    s=input()
    if s==str(1):
        path=os.getcwd()
        files = glob.glob(path + "/*.csv")
        print("enter the sentiment_deup file out of this")
        print(files)
        d=input("provide the name of sentiment_dedupfile with .csv extension")

        data=pd.read_csv(d)
        print(files)
        k=input("give the keyword file")
        keyword=pd.read_csv(k)
        l=clean_text(data)

```

```

data["keyword"]=1
d=data
k=keyword
output=category_mapping(d,k)
print("Category Mapping Sucessfull")
print("Here is the frequency distribution of each category that been mapped to eac
print("-----")
print(data["Category"].value_counts())
print("-----")
from google.colab import files
output.to_csv('output.csv', encoding= 'utf-8-sig')
print("file downloading started")
files.download('output.csv')

```

Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/>
Requirement already satisfied: fuzzywuzzy in /usr/local/lib/python3.7/dist-packages
Press 1 for Category Mapping

1

enter the sentiment_deup file out of this

```
['/content/output.csv', '/content/sentiment.csv', '/content/keyword.csv']
```

provide the name of sentiment_dedupfile with .csv extension/content/sentiment.csv

```
['/content/output.csv', '/content/sentiment.csv', '/content/keyword.csv']
```

give the keyword file/content/keyword.csv

Category Mapping Sucessfull

Here is the frequency distribution of each category that been mapped to each sku

```

-----
Jackets, Coats, Parkas, and Bombers      445
                                           201
Footwear                                132
Shirts and Tops                          120
JUNK                                     83
Sweaters, Hoodies, Sweatshirts          74
Bottoms                                 43
Headwear and Neckwear                   38
Vests                                   24
Rainwear                                12
Name: Category, dtype: int64

```

```

-----
file downloading started

```

Double-click (or enter) to edit

[Colab paid products](#) - [Cancel contracts here](#)

✓ 24m 40s completed at 1:07 AM ● ✕