

The image features a white background framed by blue curtains with red ties. The curtains have a scalloped top edge and are pulled back to reveal the central text.

Harnessing The Natural Language Toolkit for More Productive SEO



Keyword Research

Keyword research is the process of identifying which terms or phrases your target audience use to search for your goods, services or content in search engines and how frequently these terms or phrases are used.

User Behaviour Analysis

This user behaviour analysis helps you identify which keywords to target and evaluate if your content is satisfying the searcher's intent.

- Grouping
- Frequency
- Intent

<https://erudite.agency/insights/how-to-conduct-an-in-depth-keyword-research/>

Keywords	Search Volume	Category	Sub category	Sub category 2
acidophilus capsules	880	Lacto	Acidophilus	Delivery Method
acidophilus tablets	880	Lacto	Acidophilus	Delivery Method
acidophilus pills	140	Lacto	Acidophilus	Delivery Method
acidophilus capsules side effects	70	Lacto	Acidophilus	Delivery Method
acidophilus complex	70	Lacto	Acidophilus	Delivery Method
acidophilus powder	70	Lacto	Acidophilus	Delivery Method
lactobacillus acidophilus tablets	50	Lacto	Acidophilus	Delivery Method
acidophilus bacteria powder	20	Lacto	Acidophilus	Delivery Method
acidophilus capsules review	20	Lacto	Acidophilus	Delivery Method
inserting acidophilus capsules for yeast infection	10	Lacto	Acidophilus	Delivery Method
acidophilus probiotic pills	10	Lacto	Acidophilus	Delivery Method
acidophilus tablets reviews	10	Lacto	Acidophilus	Delivery Method
acidophilus tablets uk	10	Lacto	Acidophilus	Delivery Method
what are acidophilus tablets used for	10	Lacto	Acidophilus	Delivery Method
acidophilus sachet	10	Lacto	Acidophilus	Delivery Method
l acidophilus pills	10	Lacto	Acidophilus	Delivery Method
acidophilus how much to take	10	Lacto	Acidophilus	Dosage
acidophilus medication	10	Lacto	Acidophilus	Dosage
recommended dose of acidophilus	10	Lacto	Acidophilus	Dosage
side effects of acidophilus	390	Lacto	Acidophilus	Side Effects
acidophilus side effects	390	Lacto	Acidophilus	Side Effects
probiotics acidophilus side effects	50	Lacto	Acidophilus	Side Effects
acidophilus probiotic side effects	50	Lacto	Acidophilus	Side Effects
lactobacillus acidophilus la 14	10	Lacto	Acidophilus	Strain
what is acidophilus	260	Lacto	Acidophilus	Summary
how to test acidophilus	90	Lacto	Acidophilus	Summary
is acidophilus a probiotic	70	Lacto	Acidophilus	Summary
what is lactobacillus acidophilus	40	Lacto	Acidophilus	Summary
acidophilus ingredients	10	Lacto	Acidophilus	Summary
define acidophilus	10	Lacto	Acidophilus	Summary
is lactobacillus acidophilus a probiotic	10	Lacto	Acidophilus	Summary
acidophilus supplement	70	Lacto	Acidophilus	Supplement
acidophilus for thrush	110	Lacto	Acidophilus	Thrush
which yogurt contains lactobacillus acidophilus	10	Lacto	Acidophilus	Yoghurt
yogurt that contains live cultures of lactobacillus acidophi	10	Lacto	Acidophilus	Yoghurt
l casei benefits	70	Lacto	Casei	Benefits
lactobacillus casei benefits	50	Lacto	Casei	Benefits
lactobacillus casei shirota	140	Lacto	Casei	Strain
l casei shirota strain	20	Lacto	Casei	Strain
l casei side effects	10	Lacto	Casei	Side Effects

- Cleaning
- Categorization
- Analysis

Clocked hours

28:45:00

10,000 Keywords

1 Day billed

Natural Language Toolkit (NLTK)

Natural language processor with python.

Classification

Tokenization

Stemming

Tagging

Parsing

**Semantic
Reasoning**

Natural language processor with python.

Classification

Tokenization

Pull out various
tokens from text
and classify them
by nouns, verbs

Natural language processor with python.

Instruction

Determine categories
based on word
frequency

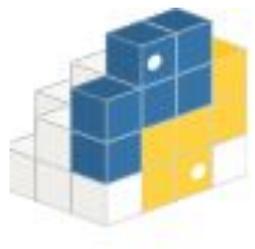
Getting started.

Download and install python library
(preferably version 3.6.4)



[Python.org/downloads/](https://python.org/downloads/)

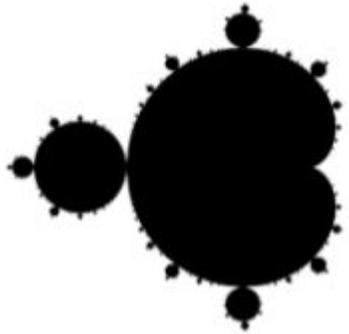
You also need the following
python libraries



pandas 1.2.4

Powerful data structures for data analysis, time series, and statistics.

<https://pypi.org/project/pandas/>



TextBlob

A simple API for common natural language processing tasks such as part-of-speech tagging, **noun phrase extraction**, sentiment analysis, **classification**, translation, and more.

<https://textblob.readthedocs.io/en/dev/>

xlrd

&

OpenPyXL 

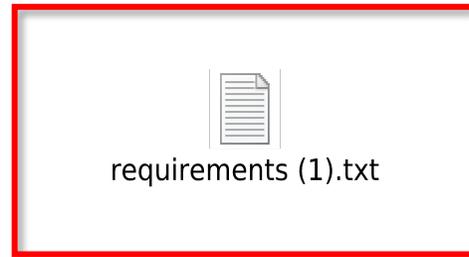
Both libraries for reading
data and formatting
information from Excel
files **.xls** format

<https://xlrd.readthedocs.io/en/latest/>

<https://openpyxl.readthedocs.io/en/stable/>

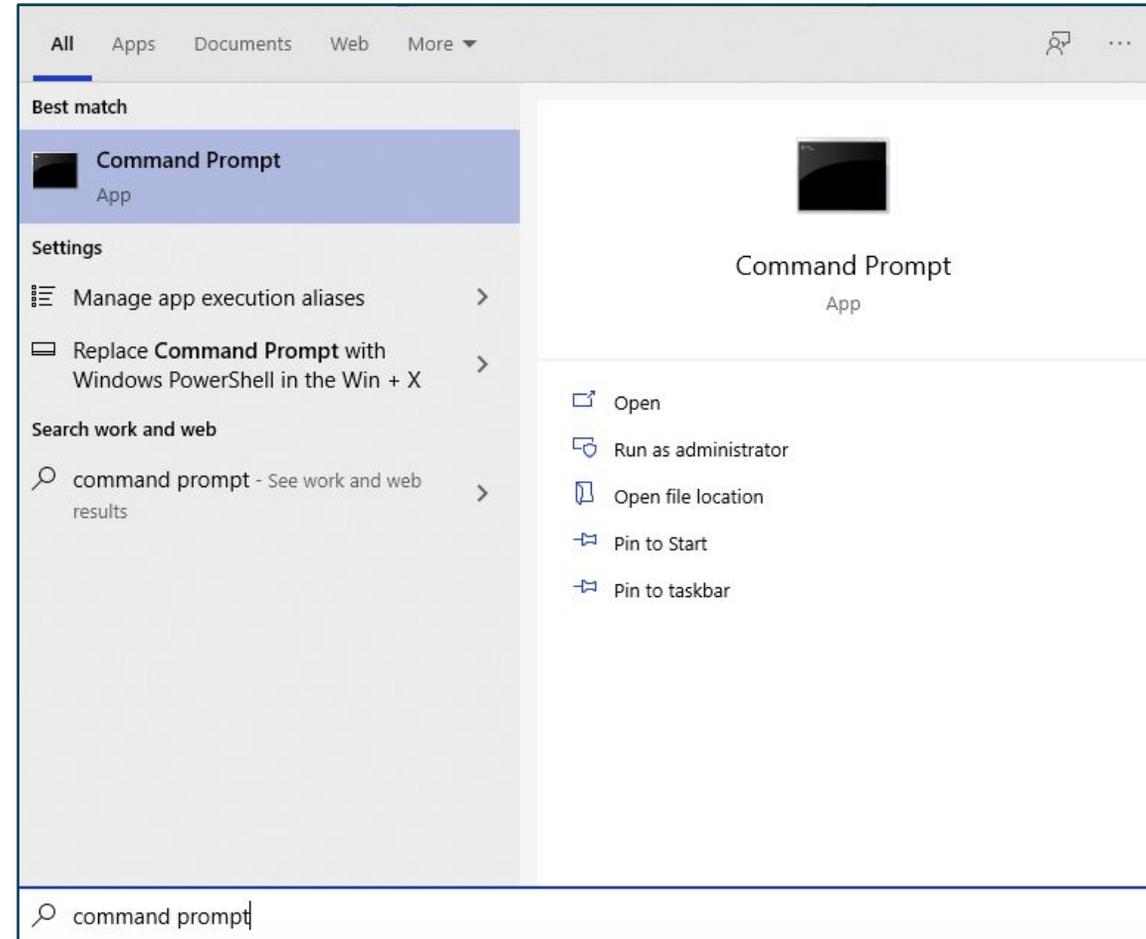
How to pip install the python libraries

Download requirement(1).txt file and save it on your computer (note the file path)



Open the command prompt.

- ❑ Open Windows
- ❑ Type command prompt

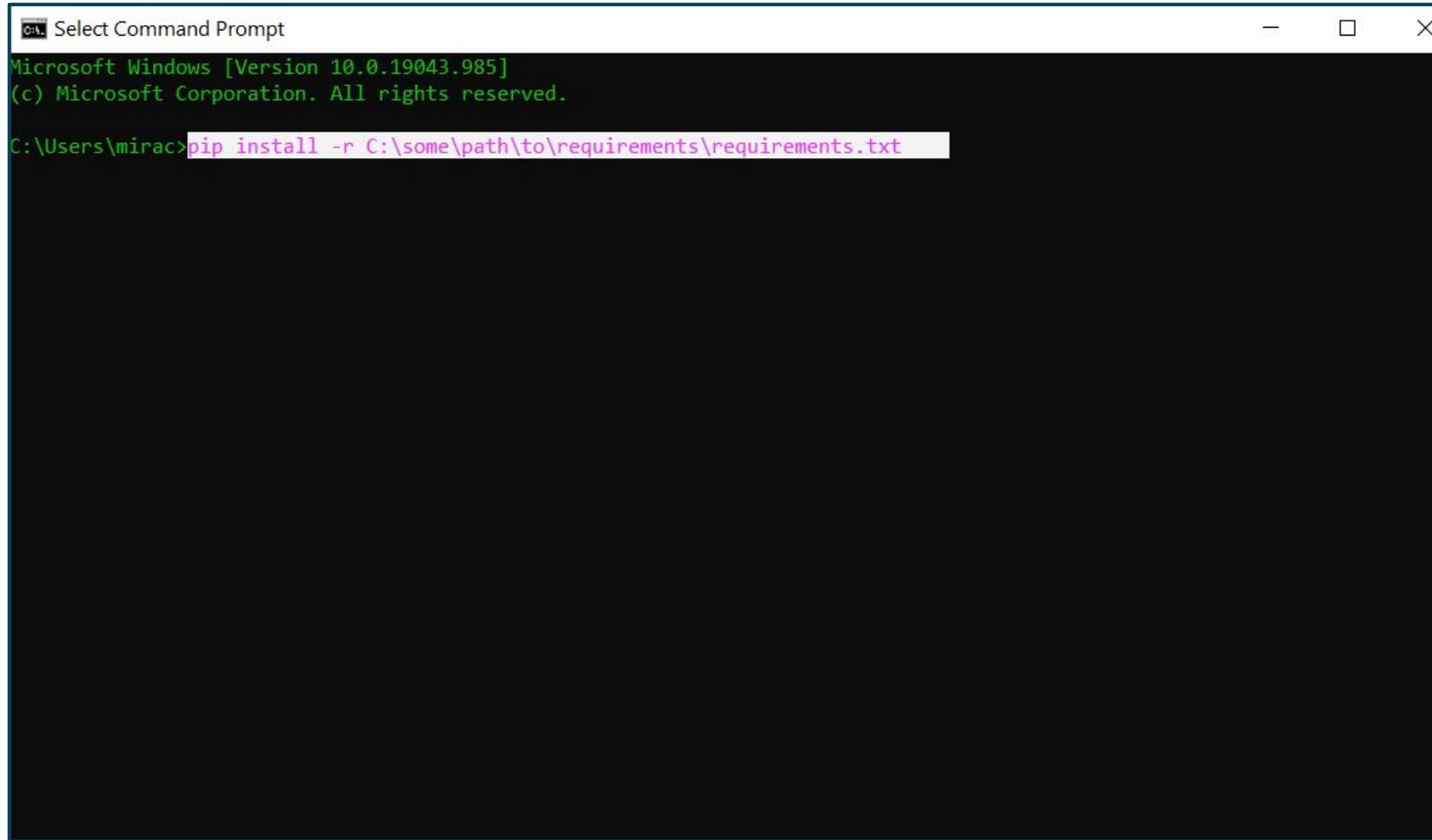


- ❑ In the command prompt
- ❑ Copy and paste - `pip install -r`
`C:\some\path\to\requirements\requirements.txt`

Highlighted in red is the path where your file is saved

Paste the instruction and click enter .

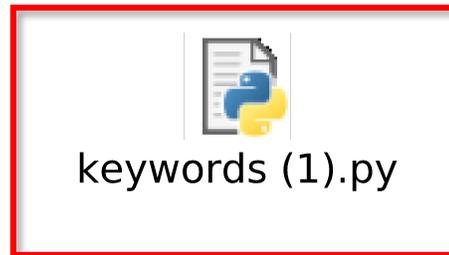
```
pip install -r  
C:\some\path\to\  
requirements\re  
quirements.txt
```

A screenshot of a Windows Command Prompt window titled "Select Command Prompt". The window shows the following text: "Microsoft Windows [Version 10.0.19043.985] (c) Microsoft Corporation. All rights reserved. C:\Users\mirac>pip install -r C:\some\path\to\requirements\requirements.txt". The command is highlighted with a white selection box.

```
Select Command Prompt  
Microsoft Windows [Version 10.0.19043.985]  
(c) Microsoft Corporation. All rights reserved.  
C:\Users\mirac>pip install -r C:\some\path\to\requirements\requirements.txt
```

The analysis script.

Download and save the python script



Prepare your keywords list .

- ❑ Ensure your keywords list is clean has no spaces or empty rows
- ❑ Heading must be keywords
- ❑ File must be .xlsx not CSV

Keywords	Search Volume
acidophilus capsules	880
acidophilus tablets	880
acidophilus pills	140
acidophilus capsules side effects	70
acidophilus complex	70
acidophilus powder	70
lactobacillus acidophilus tablets	50
acidophilus bacteria powder	20
acidophilus capsules review	20
inserting acidophilus capsules for yeast infection	10
acidophilus probiotic pills	10
acidophilus tablets reviews	10
acidophilus tablets uk	10
what are acidophilus tablets used for	10
acidophilus sachet	10
l acidophilus pills	10
acidophilus how much to take	10
acidophilus medication	10
recommended dose of acidophilus	10
side effects of acidophilus	390
acidophilus side effects	390
probiotics acidophilus side effects	50
acidophilus probiotic side effects	50
lactobacillus acidophilus la 14	10
what is acidophilus	260

Analysis.

- ❑ Run a command prompt
- ❑ Run the script
- ❑ Python C:\some\path\to\keywords.py
C:\some\path\to\keywordlist.xlsx
C:\some\path\to\test_output.xls

Replace highlight with where your file is saved .

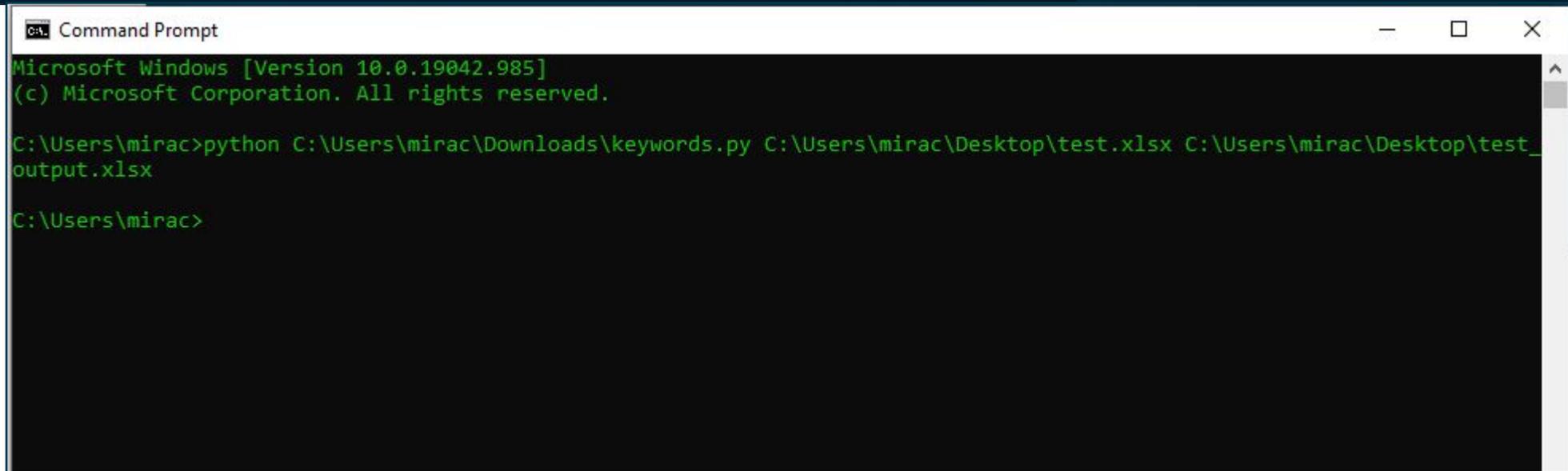
C:\some\path\to\keywords.py-

C:\some\path\to\keywordlist.xlsx

Replace highlight with where you want your test output file to be saved.

C:\some\path\to\test_output.xls

Run the script.



```
Command Prompt
Microsoft Windows [Version 10.0.19042.985]
(c) Microsoft Corporation. All rights reserved.

C:\Users\mirac>python C:\Users\mirac\Downloads\keywords.py C:\Users\mirac\Desktop\test.xlsx C:\Users\mirac\Desktop\test_output.xlsx

C:\Users\mirac>
```

python C:\Users\mirac\Downloads\keywords.py
C:\Users\mirac\Desktop\test.xlsx C:\Users\mirac\Downloads\test_output.xls

	keywords	search volume	nouns	results	cat 1	cat 2	cat 3
0	1 tens trainers	110	['tens', 'trainers']	['tens', 'trainers', None]	tens	trainers	
1	200 lb bodybuilder cutting diet	50	['bodybuilder']	['bodybuilder', None, None]	bodybuilder		
2	4 day split gym routine	50	['day', 'routine']	['day', 'routine', None]	day	routine	
3	40 40 20 meal ideas	70	['meal', 'ideas']	['meal', 'ideas', None]	meal	ideas	
4	5 meals a day bodybuilding	70	['meals', 'day']	['day', 'meals', None]	day	meals	
5	7 bridges to recovery news	70	['bridges', 'recovery', 'news']	['recovery', 'bridges', 'news']	recovery	bridges	news
6	9v battery adapter	90	['battery', 'adapter']	['battery', 'adapter', None]	battery	adapter	
7	9v battery charger	480	['battery', 'charger']	['battery', 'charger', None]	battery	charger	
8	9v charger	170	['charger']	['charger', None, None]	charger		
9	9v charger battery	50	['charger', 'battery']	['battery', 'charger', None]	battery	charger	
10	a alpha fibers	50	['alpha', 'fibers']	['fibers', 'alpha', None]	fibers	alpha	
11	a bundle of muscle fibers is called a	90	['bundle', 'muscle', 'fibers']	['muscle', 'fibers', 'bundle']	muscle	fibers	bundle
12	abs stimulator uk	50	['abs', 'stimulator', 'uk']	['stimulator', 'uk', 'abs']	stimulator	uk	abs
13	advanced bodybuilding workout routine	70	['workout', 'routine']	['workout', 'routine', None]	workout	routine	
14	advanced muscle building workout	50	['muscle', 'building', 'workout']	['muscle', 'workout', 'building']	muscle	workout	building
15	after sales service	590	['sales', 'service']	['service', 'sales', None]	service	sales	
16	ale dots	70	['dots']	['dots', None, None]	dots		
17	amount of protein per day bodybuilding	70	['amount', 'day', 'bodybuilding']	['bodybuilding', 'day', 'amount']	bodybuilding	day	amount
18	anaerobic lactic system	50	['system']	['system', None, None]	system		
19	angela watson physiotherapist	90	['angela', 'watson', 'physiotherapist']	['watson', 'angela', 'physiotherapist']	watson	angela	physiotherapist
20	ankle ice pack uk	70	['ice', 'pack', 'uk']	['uk', 'pack', 'ice']	uk	pack	ice
21	ankle sleeve	140	['ankle', 'sleeve']	['sleeve', 'ankle', None]	sleeve	ankle	
22	ankle support products	90	['support', 'products']	['support', 'products', None]	support	products	
23	are pancakes bad for you bodybuilding	50	['pancakes']	['pancakes', None, None]	pancakes		
24	are potatoes good for building muscle	50	['potatoes', 'building', 'muscle']	['muscle', 'building', 'potatoes']	muscle	building	potatoes
25	arm and leg workouts bodybuilding	50	['arm', 'leg', 'workouts']	['workouts', 'leg', 'arm']	workouts	leg	arm
26	arm street workout	90	['arm', 'street', 'workout']	['workout', 'arm', 'street']	workout	arm	street
27	ashley horner	320	['ashley', 'horner']	['horner', 'ashley', None]	horner	ashley	
28	ashley horner military	70	['ashley', 'horner']	['horner', 'ashley', None]	horner	ashley	
29	athlete physiology	70	['physiology']	['physiology', None, None]	physiology		
30	athlete recovery technology	50	['recovery', 'technology']	['recovery', 'technology', None]	recovery	technology	
31	average protein intake for bodybuilders	70	['protein', 'bodybuilders']	['protein', 'bodybuilders', None]	protein	bodybuilders	
32	avocado benefits for bodybuilding	90	['benefits']	['benefits', None, None]	benefits		
33	avoiding lactic acid build up	70	['acid']	['acid', None, None]	acid		
34	back support with gel pack	50	['support', 'pack']	['pack', 'support', None]	pack	support	
35	beactive brace uk	70	['brace', 'uk']	['uk', 'brace', None]	uk	brace	

Output.

Sort the data to identify trends.

keywords	search volume	cat 1	cat 2	cat 3										
lactic acid knots	90	acid	knots											
avoiding lactic acid build up	70	acid												
dot action 2	50	action												
tour de france mobile app	70	app	france	tour										
move free for arthritis	90	arthritis	move	free										
9v battery adapter	90	battery	adapter											
9v battery charger	480	battery	charger											
9v charger battery	50	battery	charger											
spare battery	110	battery												
food for bodybuilding beginners	70	beginners	food											
ben greenfield	2900	ben	greenfield											
beef benefits for bodybuilding	70	benefits	beef											
celebs on benefits	260	benefits	celebs											
sweet potato benefits bodybuilding	70	benefits	potato											
avocado benefits for bodybuilding	90	benefits												
body expo 2018 nec	70	body	nec											
how many calories does a bodybuilder need	70	bodybuilder	calories	need										
how to look like a bodybuilder fast	70	bodybuilder	fast											
look like a bodybuilder perform like an athlete	70	bodybuilder	look	athlet										
200 lb bodybuilder cutting diet	50	bodybuilder												
bodybuilders eating after competition	70	bodybuilders	competition											
why bodybuilders eat egg white only	70	bodybuilders	egg											

Sort

+ Add Level X Delete Level Copy Level ^ v Options... My data has headers

Column	Sort On	Order
Sort by: cat_1	Cell Values	A to Z
Then by: cat_2	Cell Values	A to Z
Then by: cat_3	Cell Values	A to Z
Then by: search_volume	Cell Values	Largest to Smallest

OK Cancel

Spot topics & themes from the data.

keywords	search_volun	cat_1	cat_2	cat_3
how many calories does a bodybuilder	70	bodybuilder	calories	need
how to look like a bodybuilder fast	70	bodybuilder	fast	
look like a bodybuilder perform like an	70	bodybuilder	look	athlete
200 lb bodybuilder cutting diet	50	bodybuilder		
bodybuilders eating after competition	70	bodybuilders	competition	
why bodybuilders eat egg white only	70	bodybuilders	egg	
how many eggs do bodybuilders eat	70	bodybuilders	egg	
what do female bodybuilders eat	70	bodybuilders		
best macro app bodybuilding	70	bodybuilding	app	macro
what do you eat for breakfast bodybuil	90	bodybuilding	breakfast	
best cheese bodybuilding	90	bodybuilding	cheese	
blue cheese bodybuilding	90	bodybuilding	cheese	
amount of protein per day bodybuilding	70	bodybuilding	day	amount
fitness and bodybuilding workouts	90	bodybuilding	fitness	workouts
fitness bodybuilding	110	bodybuilding	fitness	
what to eat after gym bodybuilding	70	bodybuilding	gym	
bodybuilding and powerlifting program	90	bodybuilding	program	
high protein high carb diet bodybuil	70	bodybuilding	protein	carb
different types of bodybuilding	70	bodybuilding	types	
whole egg vs egg white bodybuilding	90	bodybuilding	vs	egg
best way to eat cottage cheese bodybu	70	bodybuilding	way	cottage
bodybuilding workouts	1000	bodybuilding	workouts	

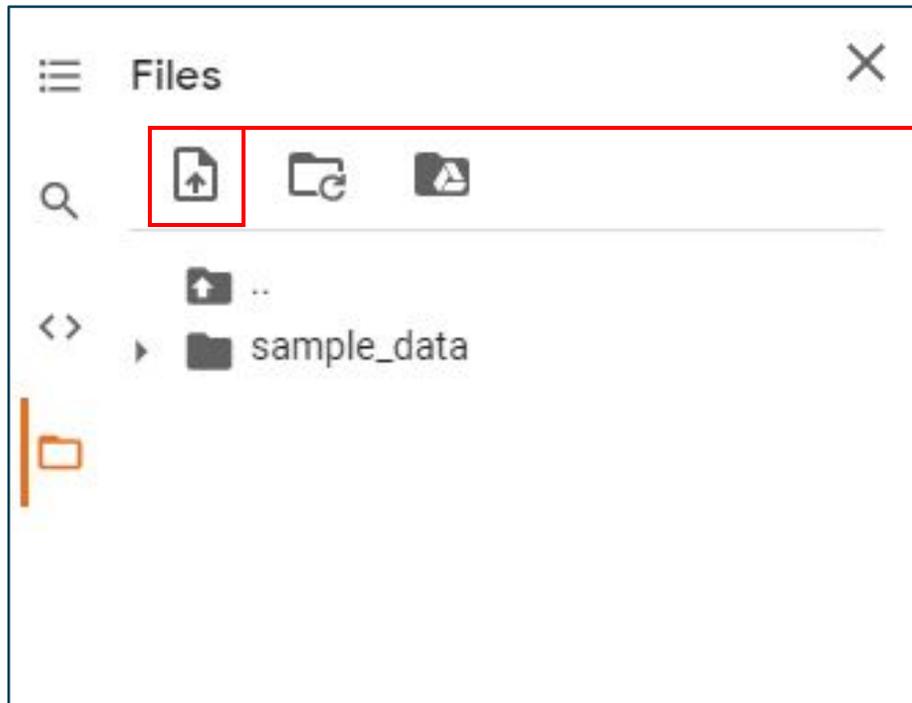
Running the script on Google Collab

colab

Make a copy and save on your
google drive.

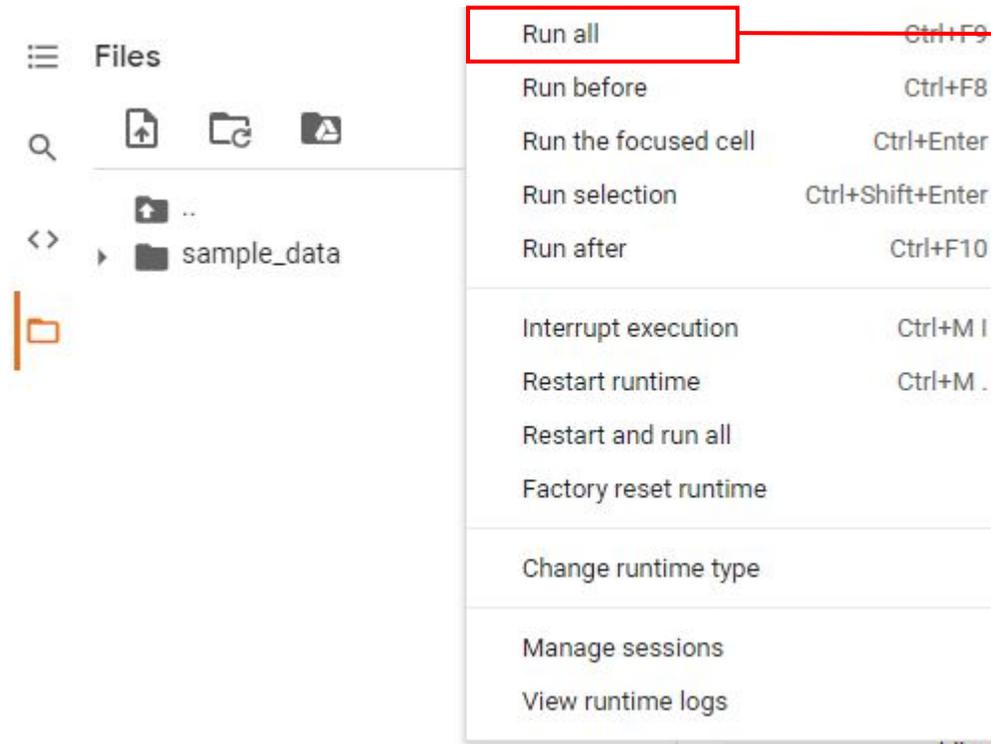
<http://bit.ly/keywordclassification>

Upload keywords data .



Upload the file you want to run. Same rules about uploading a clean file still apply see **slide 24**

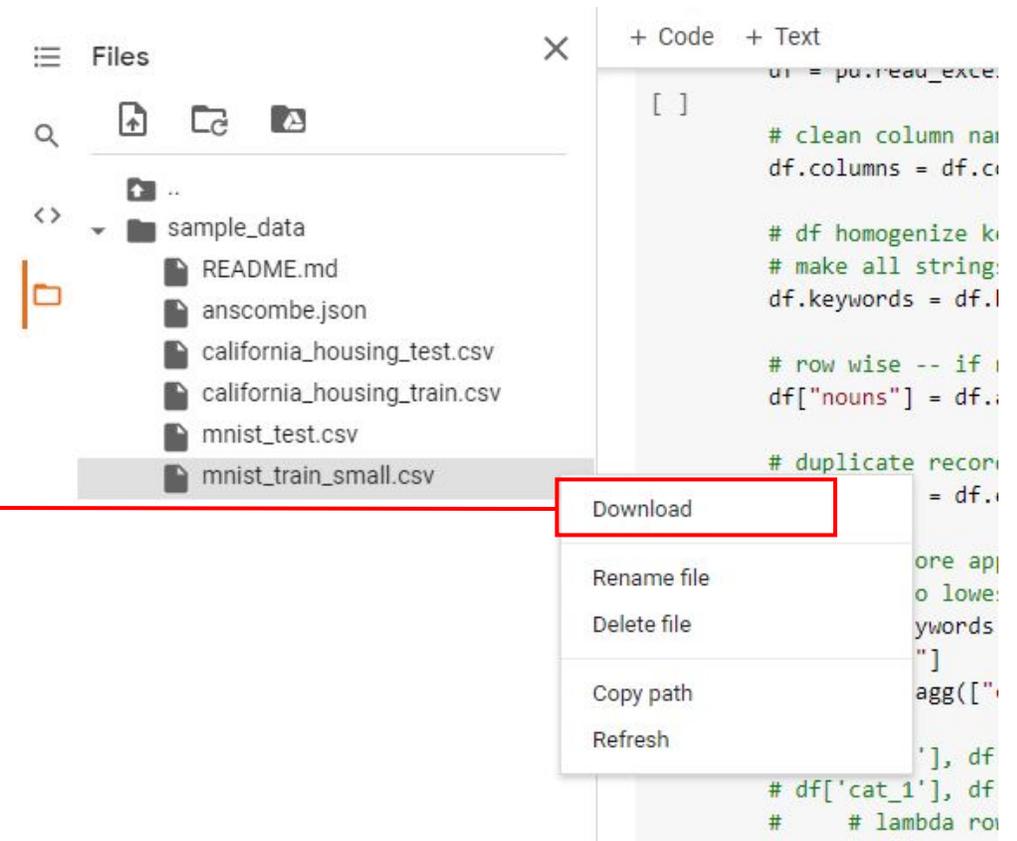
Click run all or Ctrl+F9.



Run all

Download the output file.

Right click the 3 dot and download the output file

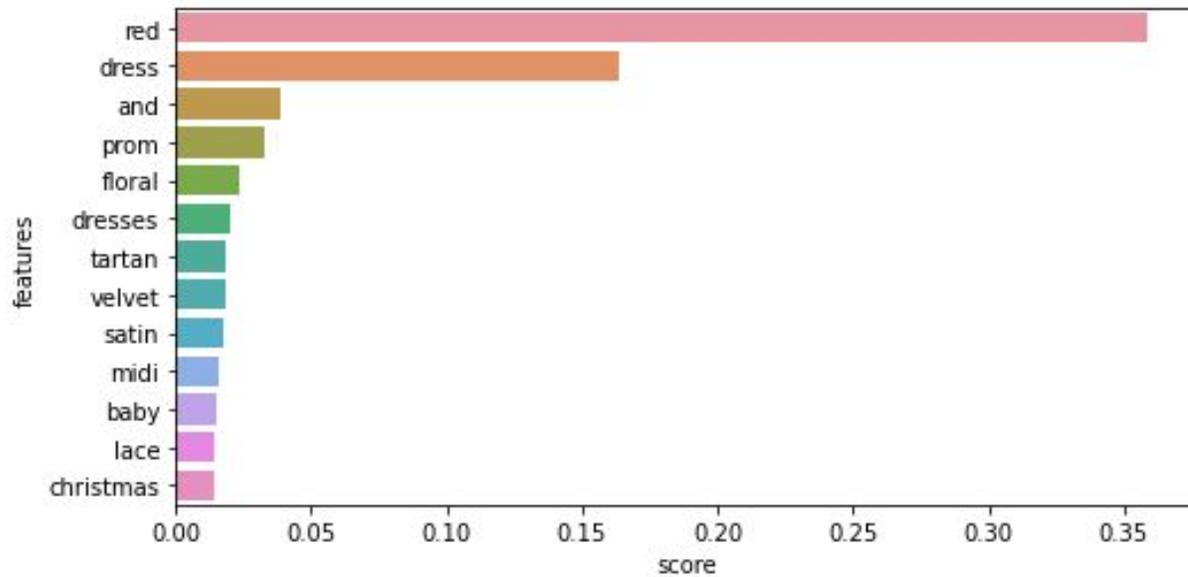


Cluster Classification

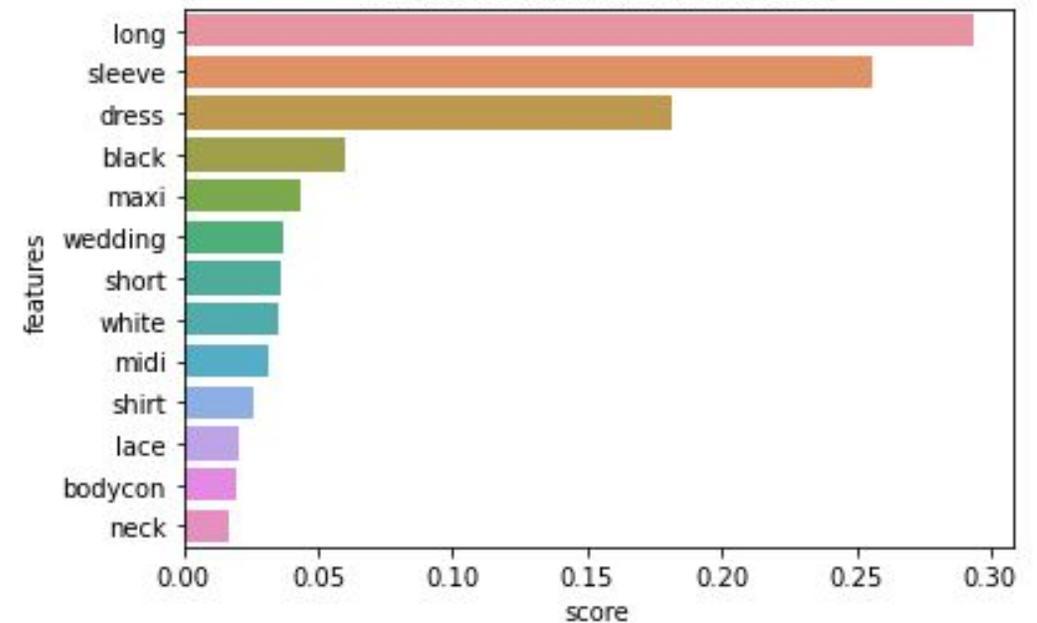
colab

Pulls out the most common clusters by frequency.

Most Common Words in Cluster 0



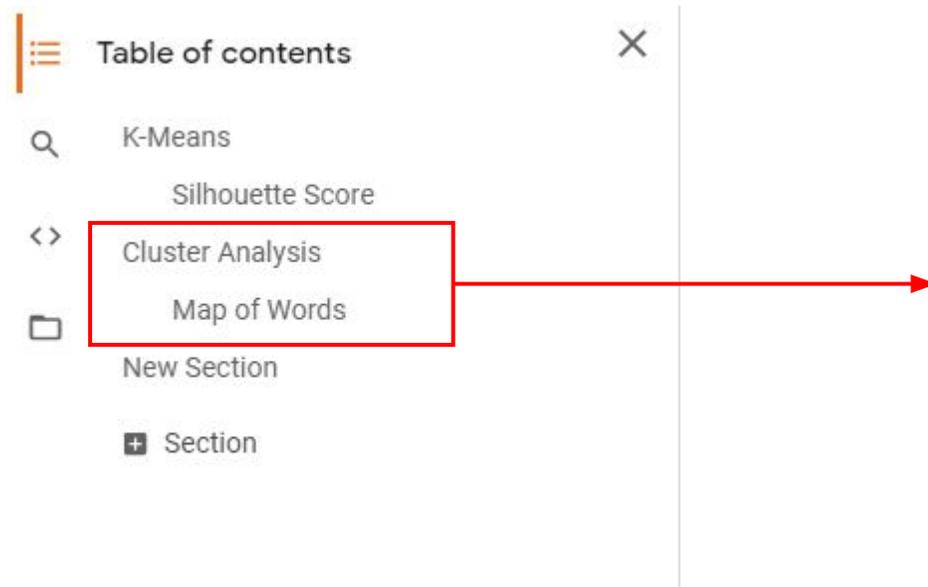
Most Common Words in Cluster 2



Make a copy and save on your
google drive.

<http://bit.ly/keywordcluster>

Increase the number of clusters depending on the size of the data.



Click on cluster analysis to view the code.

Increase the number of clusters depending on the size of the data.

```
▶ best_result = 5  
kmeans = kmeans_results.get(best_result)  
  
final_df_array = final_df.to_numpy()  
prediction = kmeans.predict(final_df)  
n_feats = 20  
dfs = get_top_features_cluster(final_df_array, prediction, n_feats)  
plotWords(dfs, 13)
```

→ Increase or decrease the number of clusters

How to use the cluster file .

The logo for Google Colab, featuring the word "colab" in a lowercase, rounded font. The "co" is yellow with a white outline, and the "lab" is orange with a white outline.

Same way to
upload data and
run as slide 32-35

Feedback.

What additional functions would
you like to see?
Send me a message @mira_inam

Thanks.

Eniola Olaleye @galileoeni
&
Mahdi Mairza

Thank You!