**Problem Statement:** Suppose your marketing team is working with a client - who are trying to rank their website in Yoga segment. They are looking for some insights on rankings for the most popular keywords - and hand over a dump of keywords. (Out of which, top 50 keywords with highest search volumes are critical).

**Python Code:**

conda install -c plotly plotly=4.12.0

pip install advertools==0.8.0

import pandas as pd

pd.options.display.max\_columns = None

import plotly

import plotly.graph\_objs as go

from plotly.offline import init\_notebook\_mode, iplot, plot

init\_notebook\_mode(connected=True)

yoga\_keywords=pd.read\_csv("C://Users//Arjun//Desktop//Yoga\_Keywords\_DS\_JR\_Evaluation.csv")

yoga\_keywords.shape

yoga\_keywords\_new=yoga\_keywords[ yoga\_keywords["Global Search Volume"].str.contains("-")==False]

yoga\_keywords\_new.shape

yoga\_keywords\_new

yoga\_keywords\_new['Global Search Volume'] = pd.to\_numeric(yoga\_keywords\_new['Global Search Volume'], errors='coerce')

yoga\_keywords\_new.sort\_values(by=['Global Search Volume'], inplace=True, ascending=False)

yoga\_keywords\_top50=yoga\_keywords\_new.head(50)

queries=['how to get '+i for i in yoga\_keywords\_top50['Keyword'].tolist()]

# getting data through google search engine

cx='5225d4511bd83a831'

key='AIzaSyDdd4vLTsuBGhwwjNuTG0APIetU3X6rX-Y'

yoga=adv.serp\_goog(cx=cx,key=key,q=queries)

type(yoga)

import pathlib

pathlib.Path().absolute()

import os

os.chdir("C://Users//Arjun//Desktop")

yoga.to\_csv('yoga\_another.csv', index = False)

yoga\_detailed=pd.read\_csv("C://Users//Arjun//Desktop//yoga\_another.csv")

summary=yoga\_detailed.groupby(['displayLink'],as\_index=False).agg(({'rank': ['count', 'mean']})).sort\_values(('rank', 'count'),ascending=False)

summary.columns=['displayLink', 'count', 'avg\_rank']

summary['displayLink'] = summary['displayLink'].str.replace('www.', '')

summary['avg\_rank'] = summary['avg\_rank'].round(1)

summary.head(20).reset\_index(drop=True)

top\_domains = yoga\_detailed['displayLink'].value\_counts()[:10].index.tolist()

top\_df = yoga\_detailed[yoga\_detailed['displayLink'].isin(top\_domains)]

rank\_counts = top\_df.groupby(['displayLink', 'rank']).agg({'rank': ['count']}).reset\_index()

rank\_counts.columns = ['displayLink', 'rank', 'count']

rank\_counts.head()

fig = go.FigureWidget()

fig.add\_scatter(x=top\_df['displayLink'].str.replace('www.', ''),

y=top\_df['rank'], mode='markers',

marker={'size': 35, 'opacity': 0.035,})

fig.add\_scatter(x=rank\_counts['displayLink'].str.replace('www.', ''),

y=rank\_counts['rank'], mode='text', text=rank\_counts['count'])

fig.layout.hovermode = False

fig.layout.yaxis.autorange = 'reversed'

fig.layout.yaxis.zeroline = False

fig.layout.yaxis.tickvals = list(range(1, 11))

fig.layout.height = 600

fig.layout.title = 'Top Domains for yoga Keywords - Google'

fig.layout.yaxis.title = 'SERP Rank (number of appearances)'

fig.layout.showlegend = False

fig.layout.paper\_bgcolor = '#eeeeee'

fig.layout.plot\_bgcolor = '#eeeeee'

iplot(fig)

serp\_word\_freq = adv.word\_frequency(yoga\_detailed['title'],

rm\_words=list(adv.stopwords['english']) + ['-', '', '|', '–', '&'])

serp\_word\_freq.head(10)

serp\_word\_freq\_snippet = adv.word\_frequency(yoga\_detailed['snippet'].fillna(''), rm\_words=list(adv.stopwords['english']) + ['-', '', '|', '–', '&'])

serp\_word\_freq\_snippet.head(10)

emoji\_summary = adv.extract\_emoji(yoga\_detailed['title'])

emoji\_summary['emoji\_freq']

emoji\_summary['top\_emoji'][:20]

yoga\_detailed.iloc[:, :8].corr(method ='pearson')

**Question and Answers:**

1. Which domain/website ranks prominently in the search results.

Answer: Youtube.

2. Text Analysis - absolute and weighted frequency (From Title, Meta Description).

Answer:

Foe variable: ‘Title’

|  |  |  |  |
| --- | --- | --- | --- |
| word | abs\_freq | wtd\_freq | rel\_value |
| yoga | 559 | 559 | 1 |
| online | 323 | 323 | 1 |
| classes | 244 | 244 | 1 |
| free | 240 | 240 | 1 |
| best | 156 | 156 | 1 |
| corepower | 82 | 82 | 1 |
| 2020 | 66 | 66 | 1 |
| home | 65 | 65 | 1 |
| videos | 52 | 52 | 1 |
| demand | 50 | 50 | 1 |
|  |  |  |  |

For variable: ‘snippet’

|  |  |  |  |
| --- | --- | --- | --- |
| word | abs\_freq | wtd\_freq | rel\_value |
| yoga | 715 | 715 | 1 |
| classes | 342 | 342 | 1 |
| free | 312 | 312 | 1 |
| online | 312 | 312 | 1 |
| 2020 | 227 | 227 | 1 |
| access | 140 | 140 | 1 |
| Â· | 135 | 135 | 1 |
| videos | 123 | 123 | 1 |
| â€” | 110 | 110 | 1 |
| unlimited | 86 | 86 | 1 |

3. Is there any co-relation between rankings and emojis? What are the most frequently used emojis? Is there a special pattern?

Answer: There are no emoji found.

4. Is there any co-relation of numbers or dates to the ranking?

Answer: there no co-relation exists.