Corporate Disclosures on Instagram

ANALYZING THE INSTAGRAM ACCOUNTS OF THE NIFTY50 COMPANIES

Contents :

1. Overview of the Data

2. Topic Modelling Results (Excel + LDA)

3. Technical Details

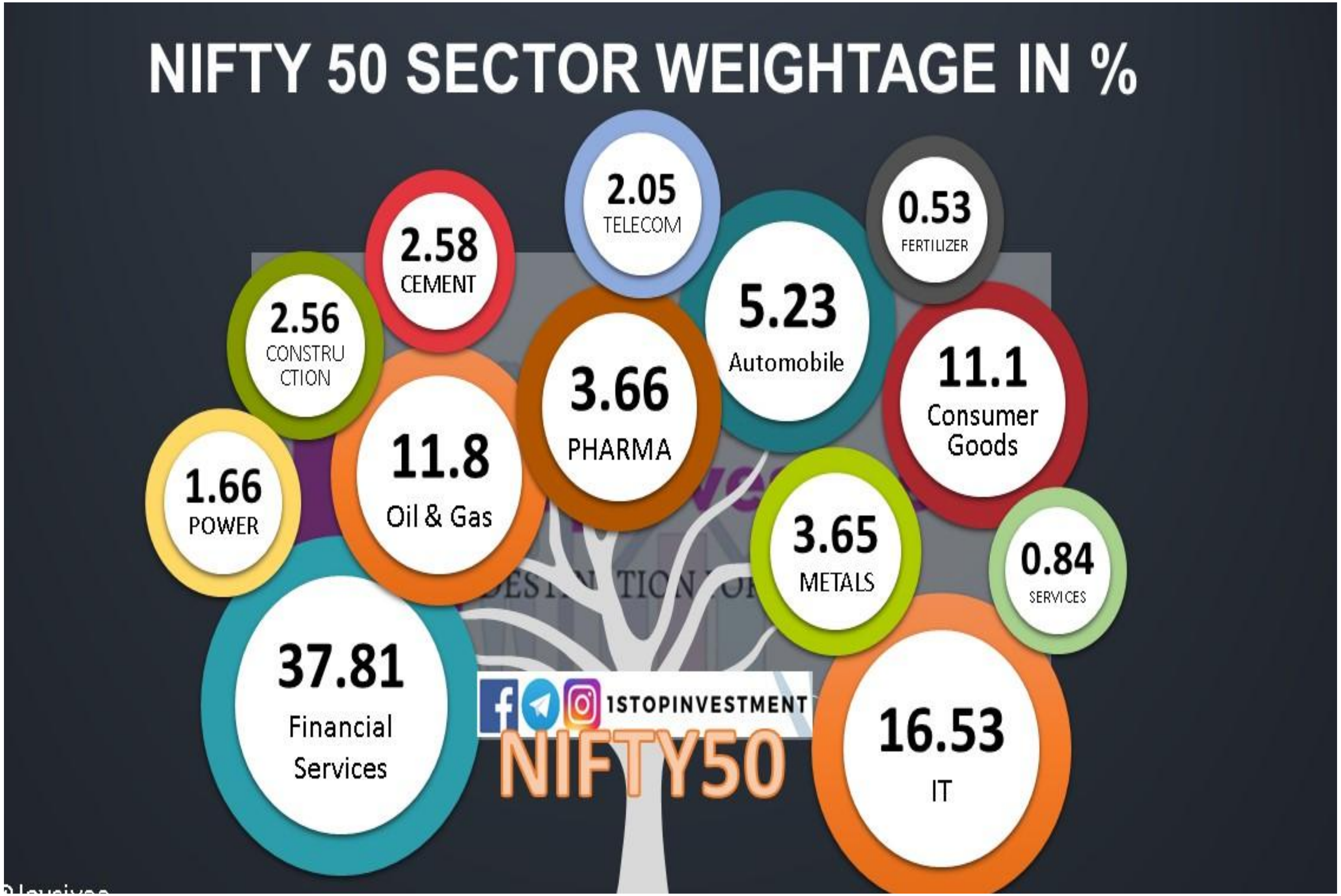
❑ Data Extraction via API calls

❑ Pre-processing of Text

❑ Using LDA for Topic Modelling

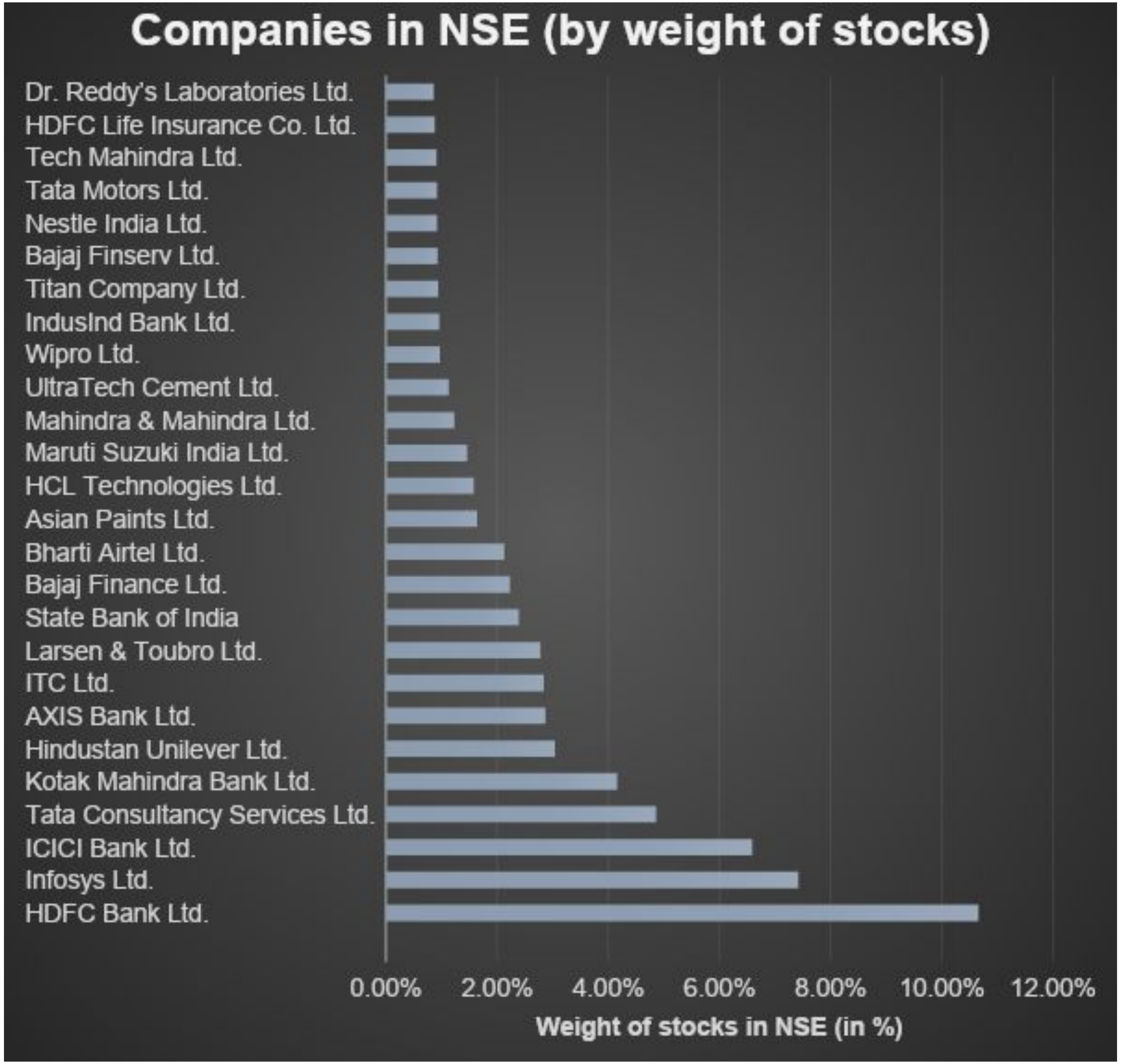
Company-wise Data Plots

Following slides focus on individual companies

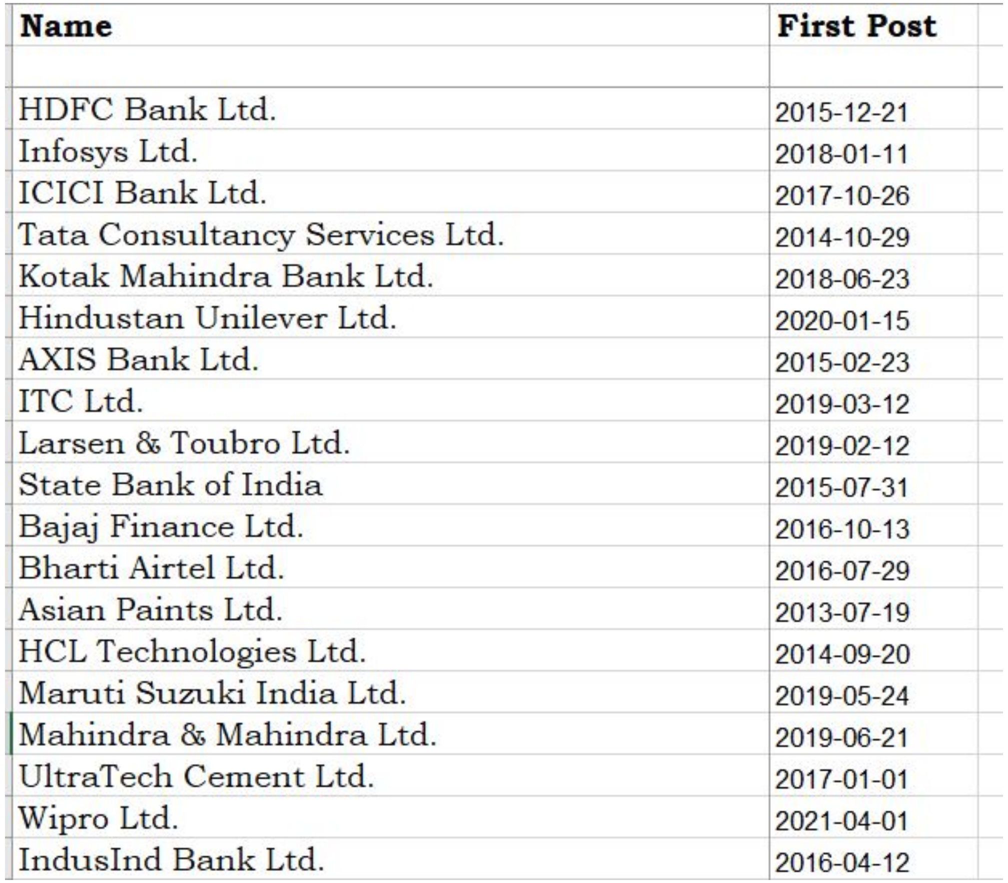


Top Companies
(by weight) in NSE

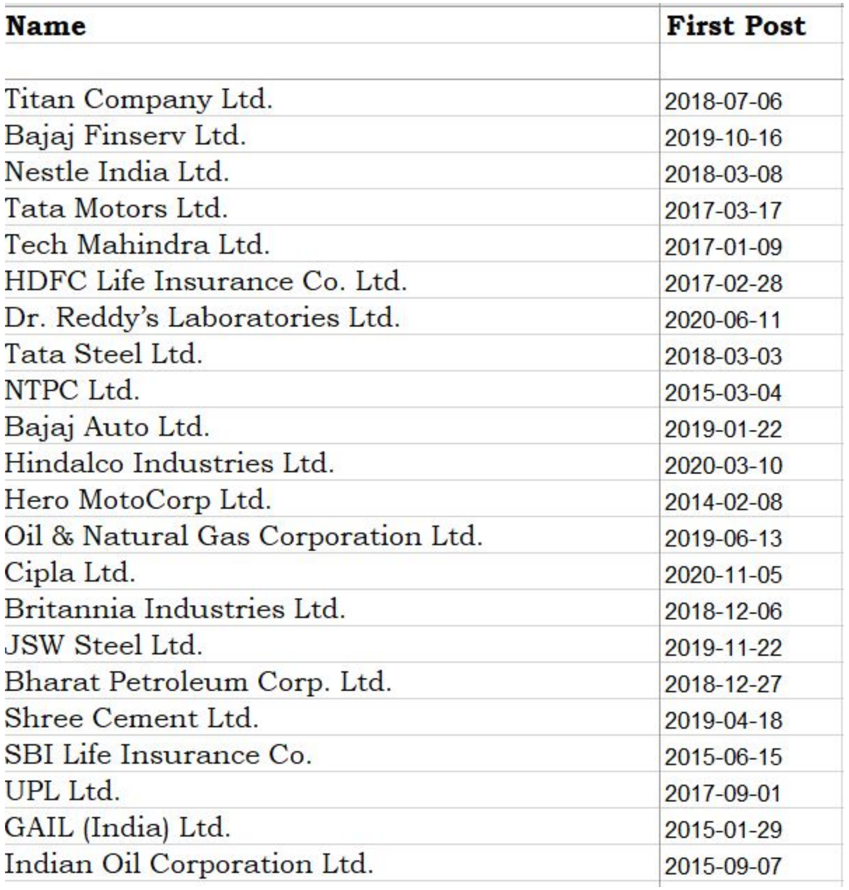
- Banks and IT Companies hold the largest stake in the National Stock Exchange.
- State Bank of India is the only Govt. owned enterprise in NSE.



First Instagram Post of Companies

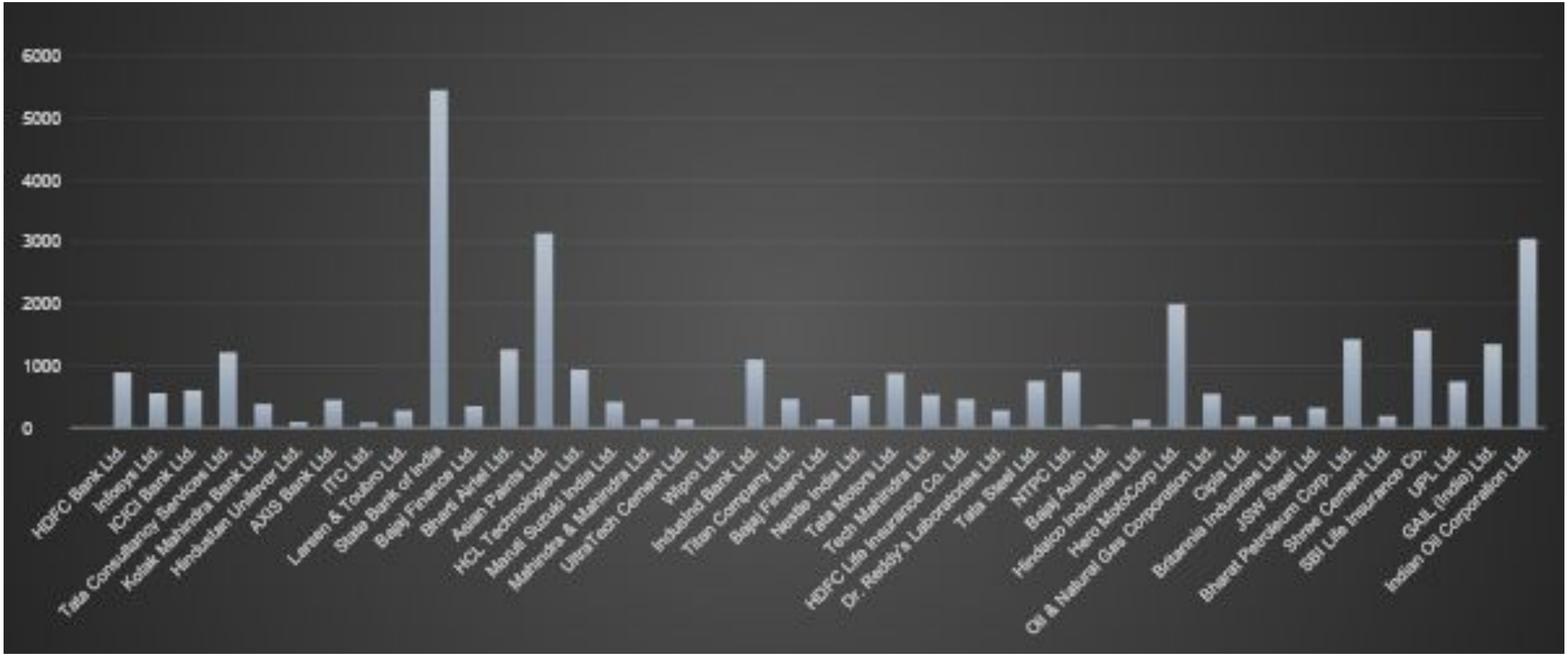


Name	First Post
HDFC Bank Ltd.	2015-12-21
Infosys Ltd.	2018-01-11
ICICI Bank Ltd.	2017-10-26
Tata Consultancy Services Ltd.	2014-10-29
Kotak Mahindra Bank Ltd.	2018-06-23
Hindustan Unilever Ltd.	2020-01-15
AXIS Bank Ltd.	2015-02-23
ITC Ltd.	2019-03-12
Larsen & Toubro Ltd.	2019-02-12
State Bank of India	2015-07-31
Bajaj Finance Ltd.	2016-10-13
Bharti Airtel Ltd.	2016-07-29
Asian Paints Ltd.	2013-07-19
HCL Technologies Ltd.	2014-09-20
Maruti Suzuki India Ltd.	2019-05-24
Mahindra & Mahindra Ltd.	2019-06-21
UltraTech Cement Ltd.	2017-01-01
Wipro Ltd.	2021-04-01
IndusInd Bank Ltd.	2016-04-12

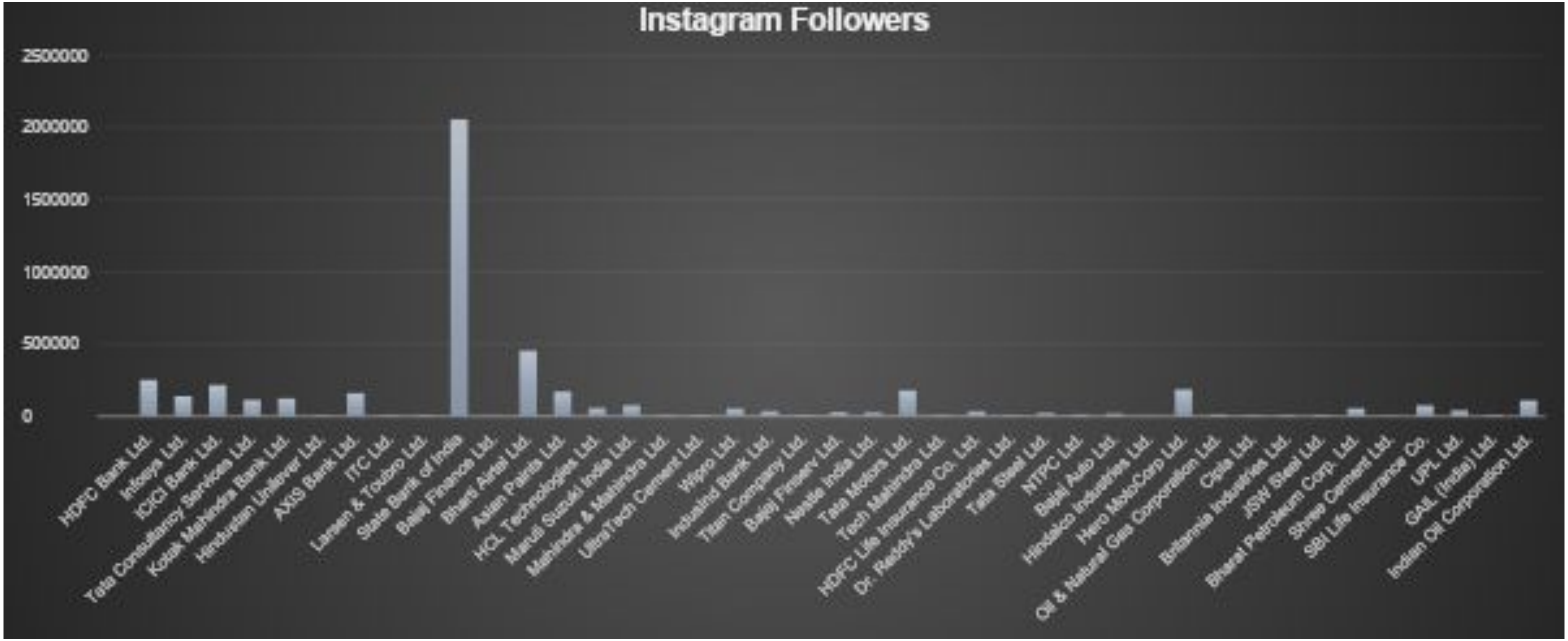


Name	First Post
Titan Company Ltd.	2018-07-06
Bajaj Finserv Ltd.	2019-10-16
Nestle India Ltd.	2018-03-08
Tata Motors Ltd.	2017-03-17
Tech Mahindra Ltd.	2017-01-09
HDFC Life Insurance Co. Ltd.	2017-02-28
Dr. Reddy's Laboratories Ltd.	2020-06-11
Tata Steel Ltd.	2018-03-03
NTPC Ltd.	2015-03-04
Bajaj Auto Ltd.	2019-01-22
Hindalco Industries Ltd.	2020-03-10
Hero MotoCorp Ltd.	2014-02-08
Oil & Natural Gas Corporation Ltd.	2019-06-13
Cipla Ltd.	2020-11-05
Britannia Industries Ltd.	2018-12-06
JSW Steel Ltd.	2019-11-22
Bharat Petroleum Corp. Ltd.	2018-12-27
Shree Cement Ltd.	2019-04-18
SBI Life Insurance Co.	2015-06-15
UPL Ltd.	2017-09-01
GAIL (India) Ltd.	2015-01-29
Indian Oil Corporation Ltd.	2015-09-07

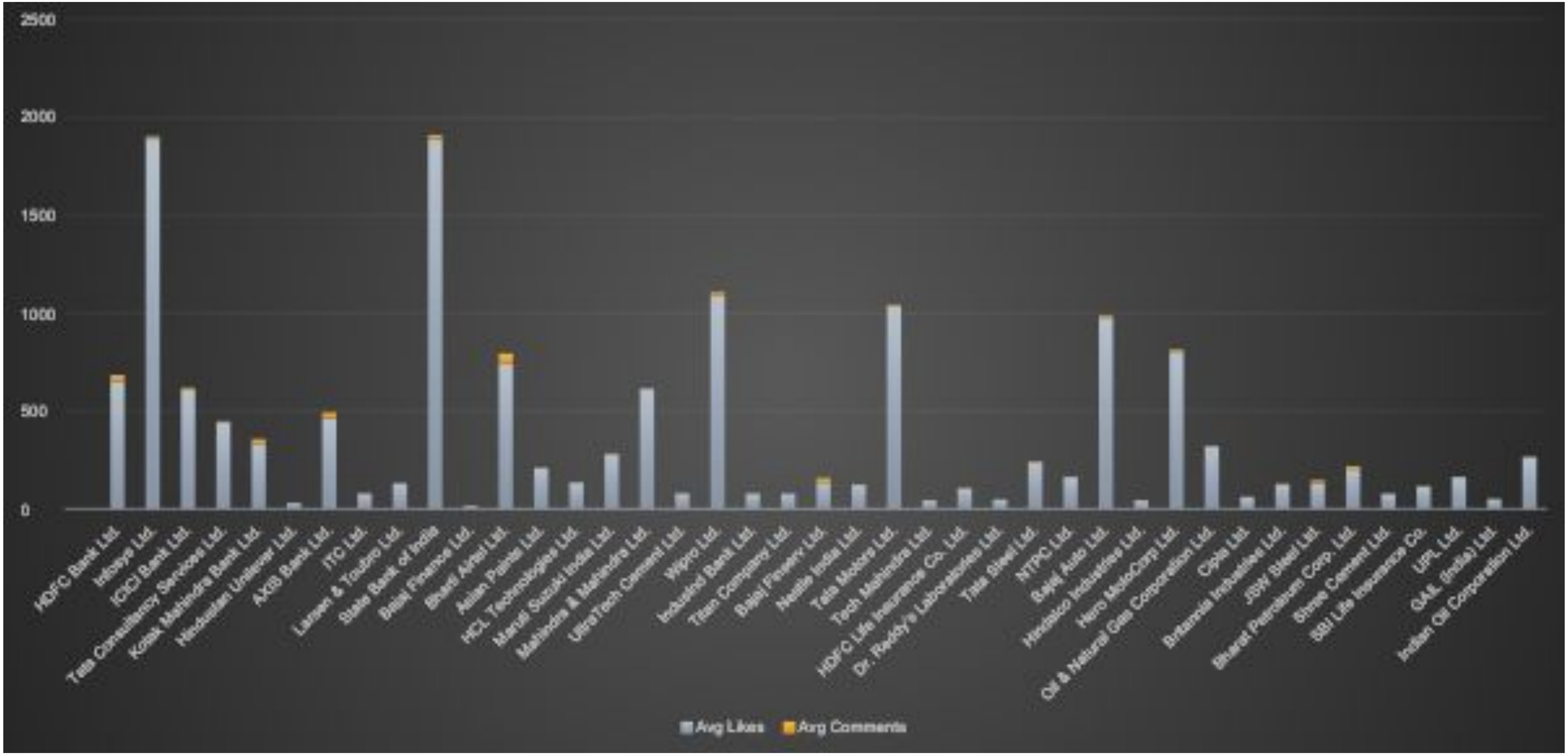
Companies v/s Total no. of Posts



Comparison of Followers



Average Likes & Comments on Instagram Posts

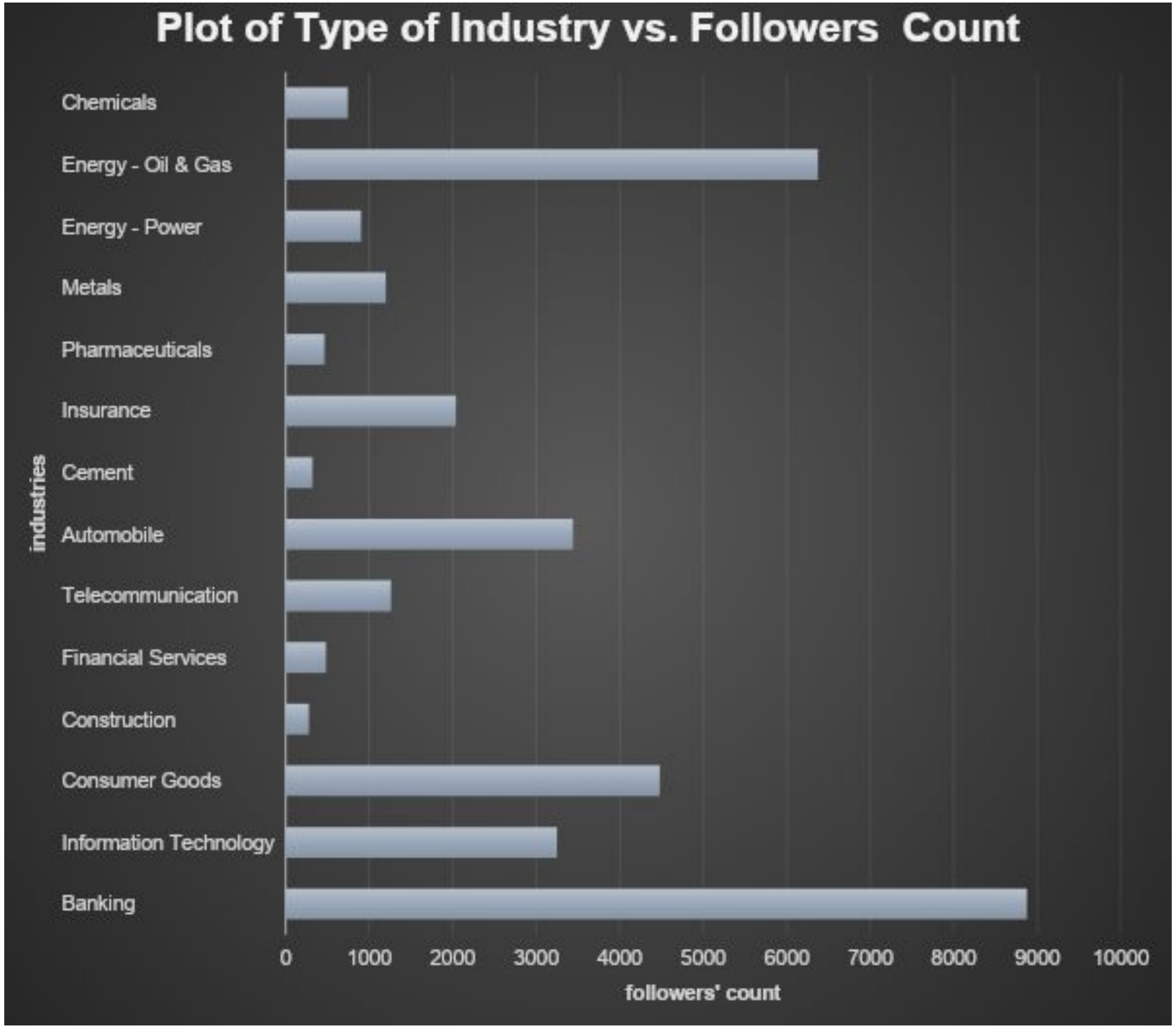


Industry-wise Data Plots

FOLLOWING SLIDES FOCUS ON THE TYPE OF INDUSTRIES

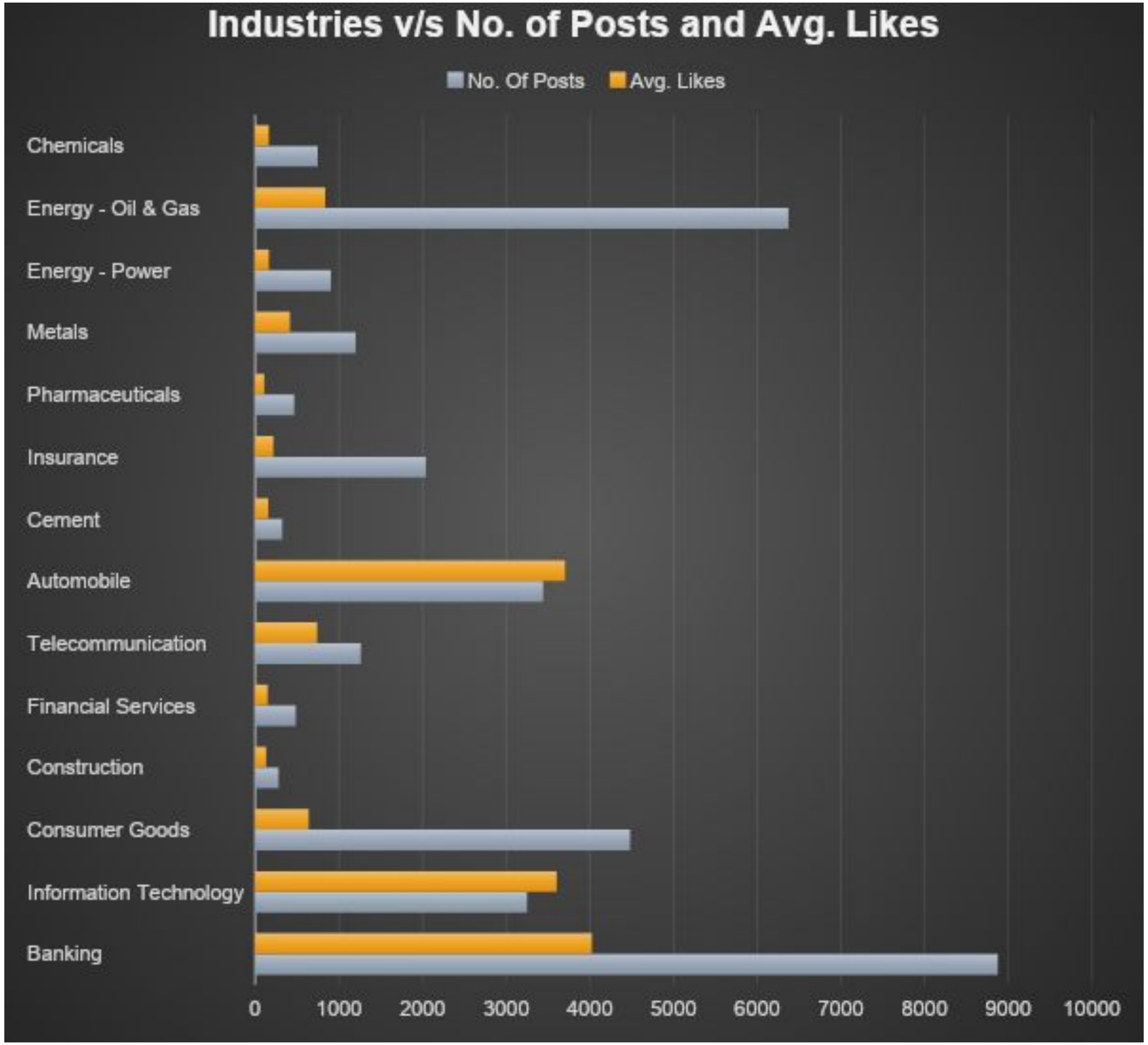
Insights from the Plot

- Banking dominates other industries in terms of followers’ count, followed by Telecommunication and IT.
- Industries such as Energy-Power, Construction and Cement have negligible followers compared to others.



Comparison of Industries in terms of Posts and Likes

- Consumer Industries have more posts and likes compared to others.
- Unusual spike in Energy-Oil & Gas, is due to nearly 4,000 posts of Indian Oil Corp.

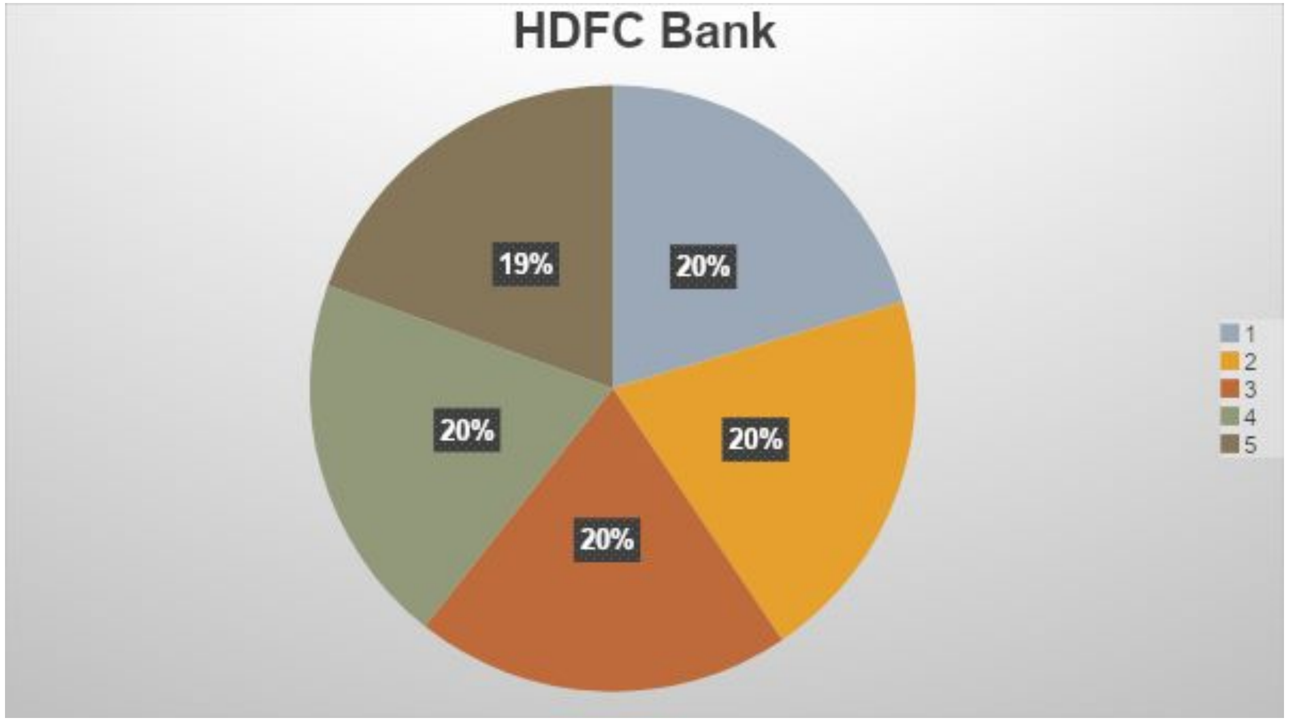


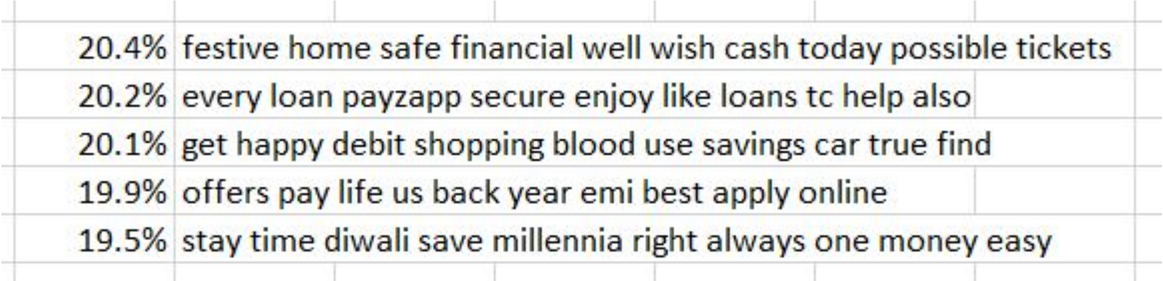
Topic Modelling Results

HDFC Bank

Top Contributing Topics:

1. Festive
2. Loan Payzapp
3. Shopping + Savings
4. EMI Offers
5. Festive offers



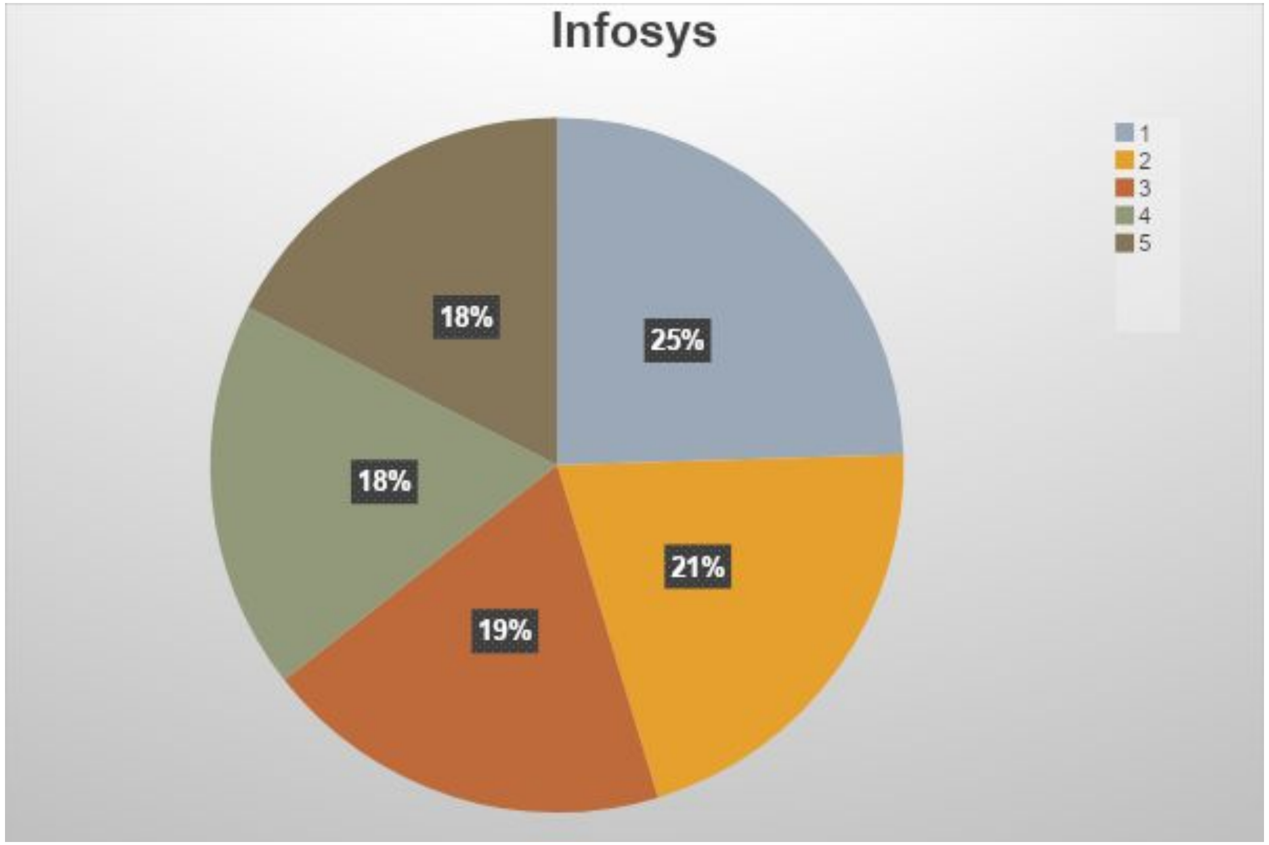


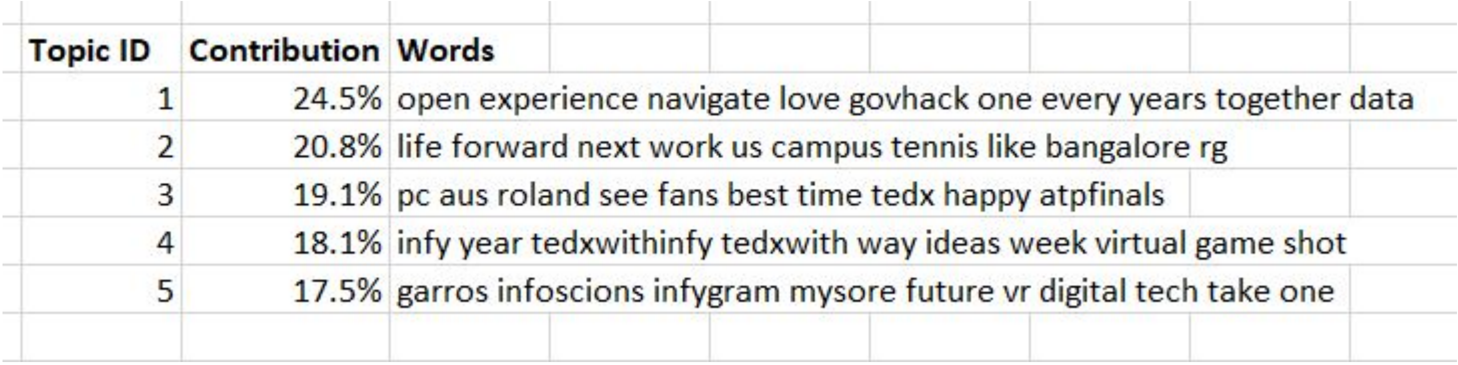
20.4%	festive home safe financial well wish cash today possible tickets
20.2%	every loan payzapp secure enjoy like loans tc help also
20.1%	get happy debit shopping blood use savings car true find
19.9%	offers pay life us back year emi best apply online
19.5%	stay time diwali save millennia right always one money easy

Infosys :

Top Contributing Topics:

1. Data experience
2. Tennis + Work
3. Sponsorship in Roland Garros (Tennis)
4. TedX
5. Digital Future



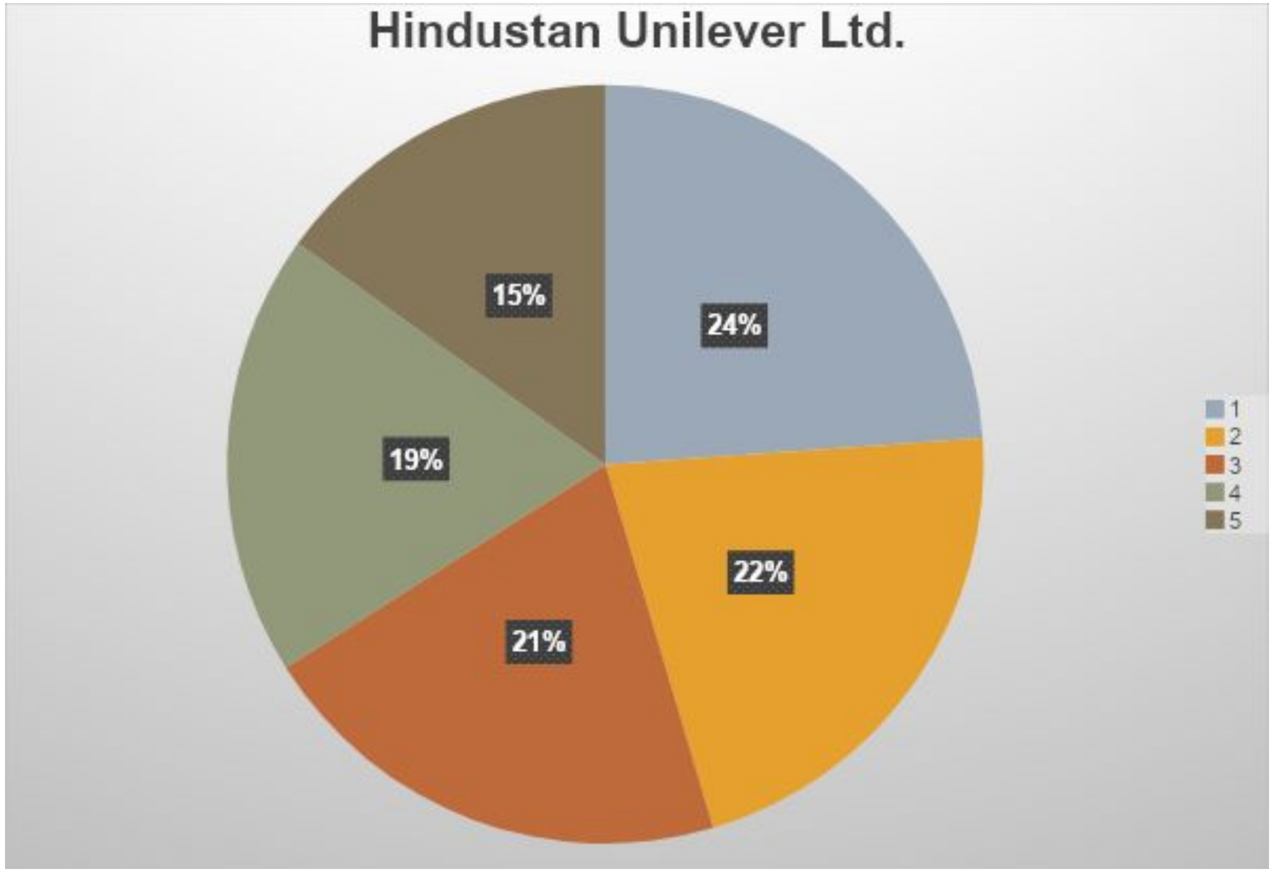


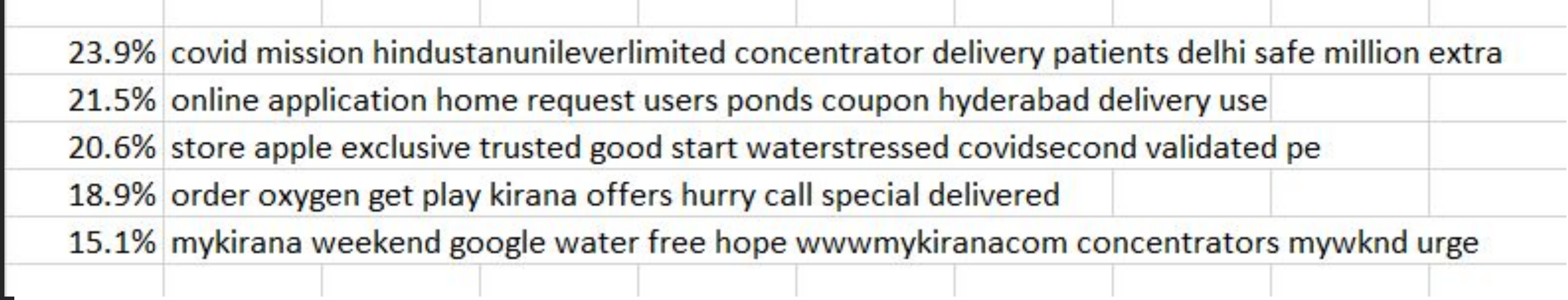
Topic ID	Contribution	Words
1	24.5%	open experience navigate love govhack one every years together data
2	20.8%	life forward next work us campus tennis like bangalore rg
3	19.1%	pc aus roland see fans best time tedx happy atpfinals
4	18.1%	infy year tedxwithinfy tedxwith way ideas week virtual game shot
5	17.5%	garros infoscions infygram mysore future vr digital tech take one

Hindustan Unilever Ltd.

Top Contributing Topics:

1. COVID-19
2. Home Delivery
3. Store offers
4. Kirana Offers
5. MyKirana App



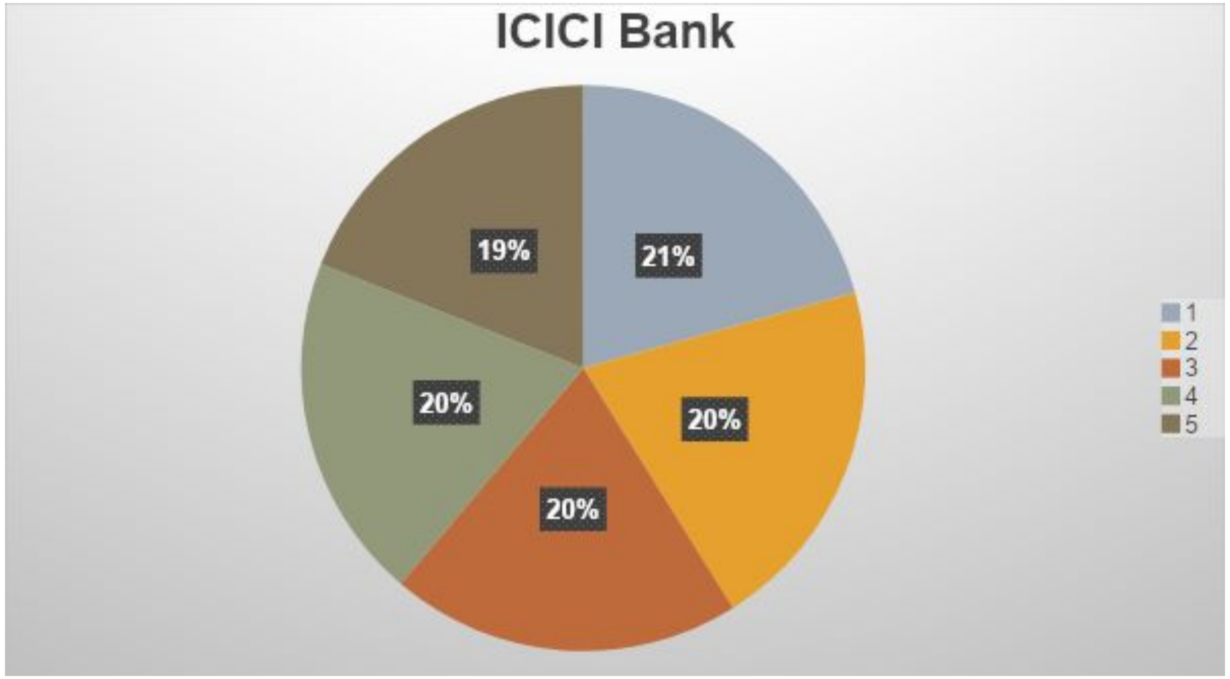


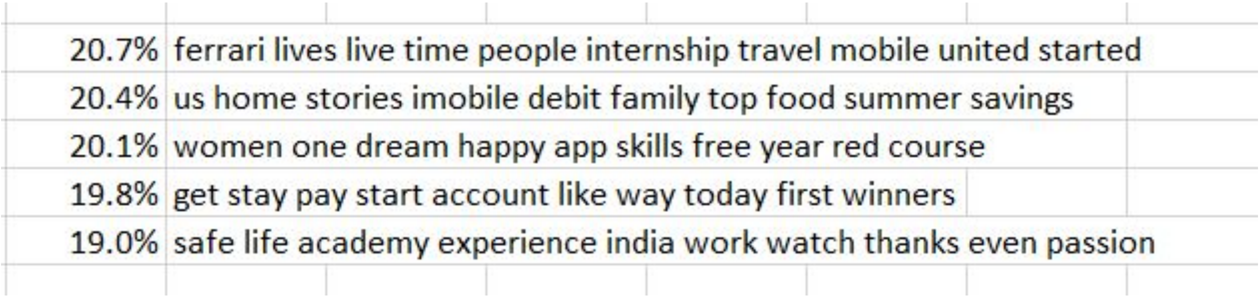
23.9%	covid mission hindustanunileverlimited concentrator delivery patients delhi safe million extra
21.5%	online application home request users ponds coupon hyderabad delivery use
20.6%	store apple exclusive trusted good start waterstressed covidsecond validated pe
18.9%	order oxygen get play kirana offers hurry call special delivered
15.1%	mykirana weekend google water free hope wwwmykiranacom concentrators mywknd urge

ICICI Bank

Top Contributing Topics:

1. Travel
2. IMobile App
3. Women
4. Promotion
5. Safe Life

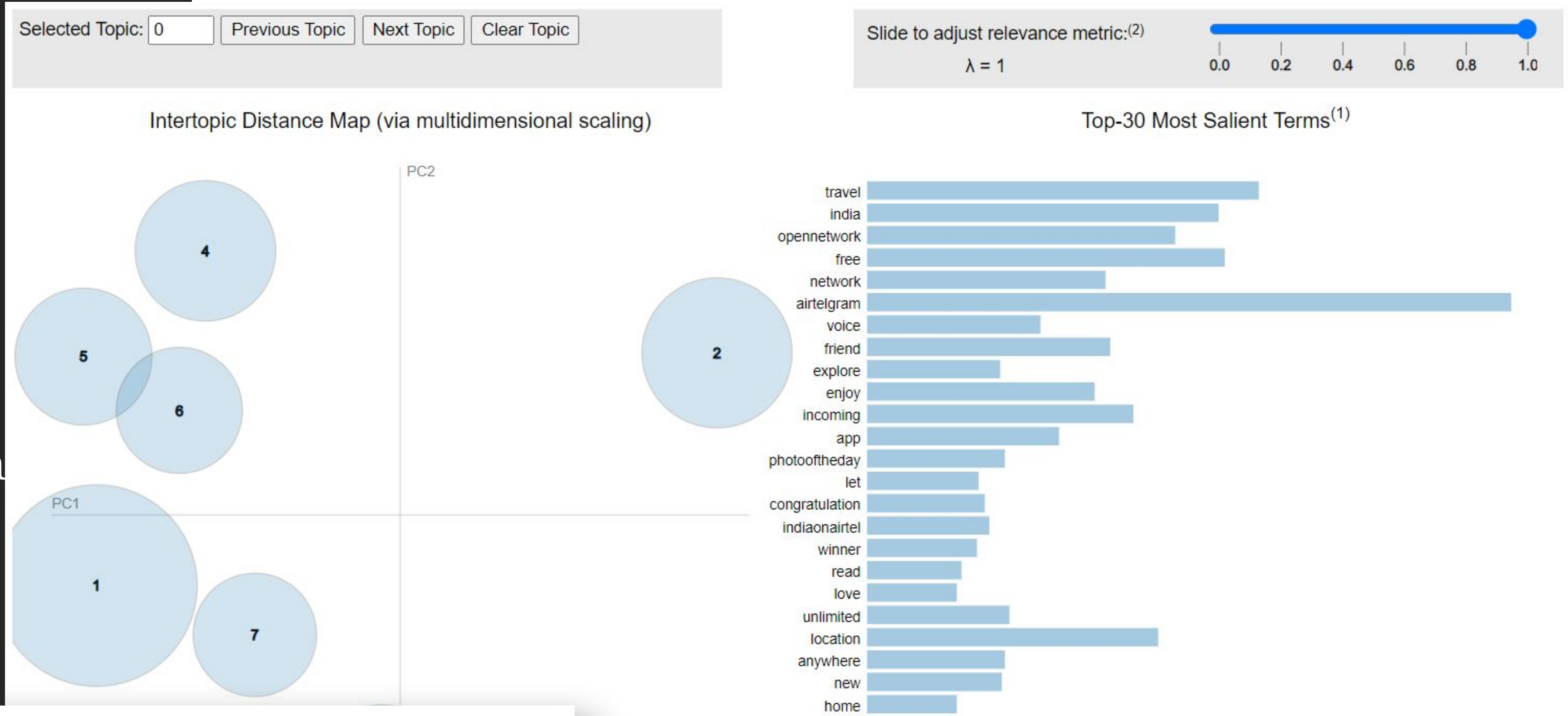




20.7%	ferrari lives live time people internship travel mobile united started
20.4%	us home stories imobile debit family top food summer savings
20.1%	women one dream happy app skills free year red course
19.8%	get stay pay start account like way today first winners
19.0%	safe life academy experience india work watch thanks even passion

Visualizing Output:

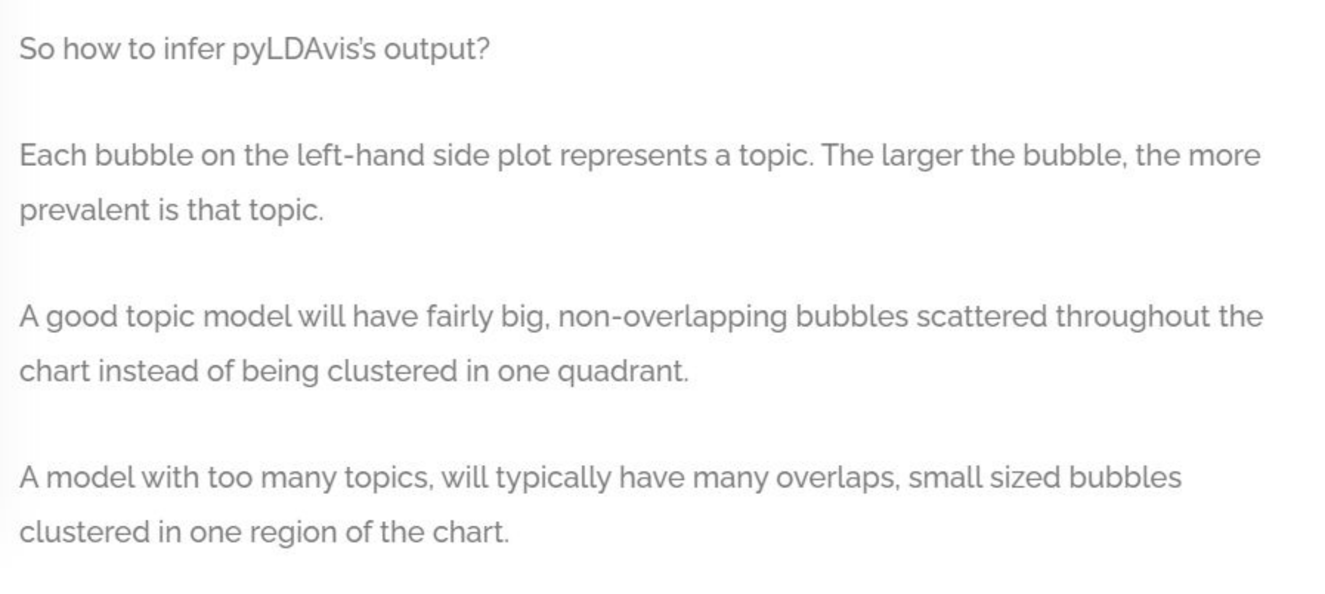
Example for Airtel India


Selected Topic: 0

Slide to adjust relevance metric:(2)

λ = 1

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

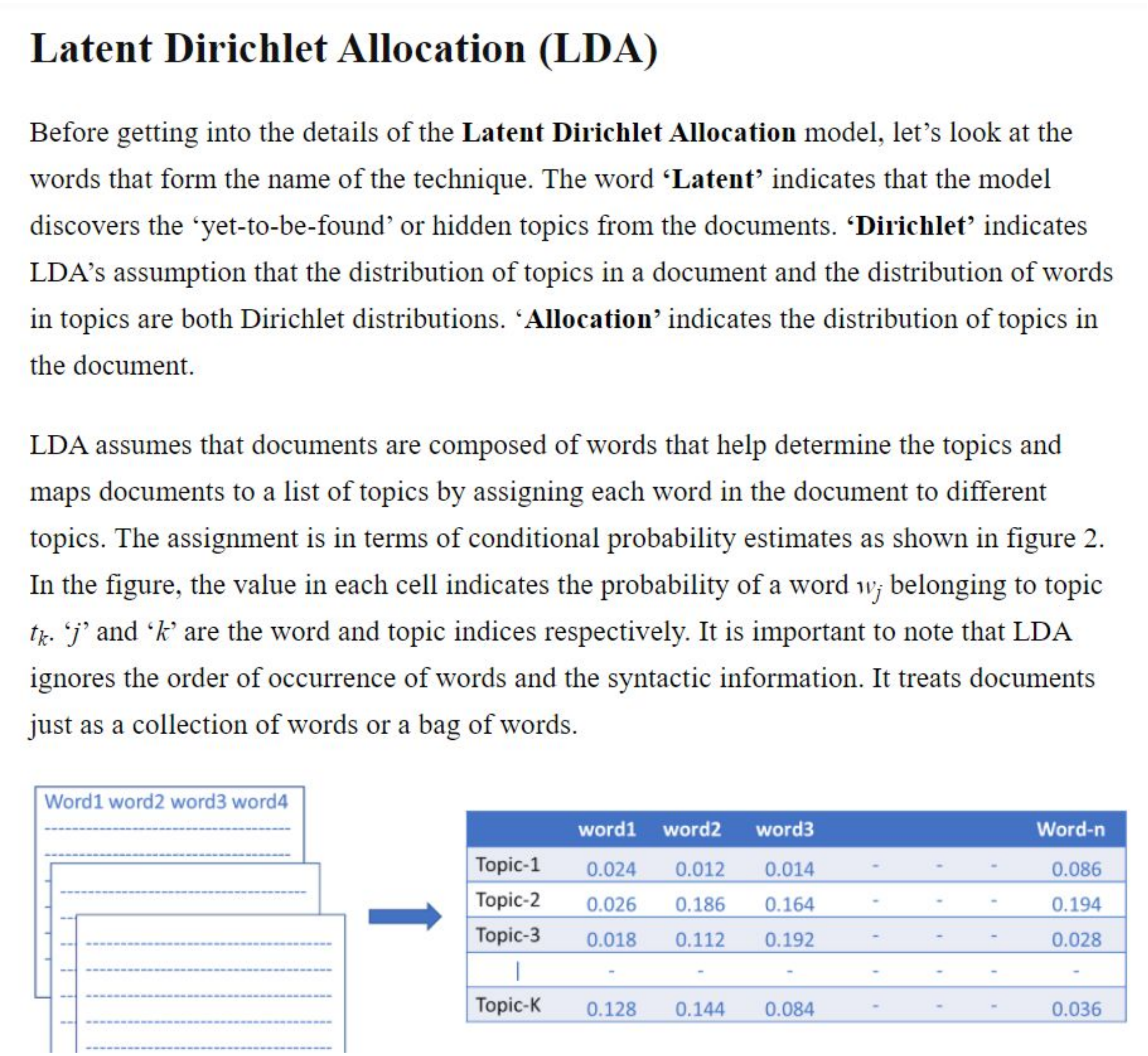

So how to infer pyLDAvis's output?

Each bubble on the left-hand side plot represents a topic. The larger the bubble, the more prevalent is that topic.

A good topic model will have fairly big, non-overlapping bubbles scattered throughout the chart instead of being clustered in one quadrant.

A model with too many topics, will typically have many overlaps, small sized bubbles clustered in one region of the chart.

A little about LDA:



Latent Dirichlet Allocation (LDA)

Before getting into the details of the **Latent Dirichlet Allocation** model, let's look at the words that form the name of the technique. The word **'Latent'** indicates that the model discovers the 'yet-to-be-found' or hidden topics from the documents. **'Dirichlet'** indicates LDA's assumption that the distribution of topics in a document and the distribution of words in topics are both Dirichlet distributions. '**Allocation'** indicates the distribution of topics in the document.

LDA assumes that documents are composed of words that help determine the topics and maps documents to a list of topics by assigning each word in the document to different topics. The assignment is in terms of conditional probability estimates as shown in figure 2. In the figure, the value in each cell indicates the probability of a word w_j belonging to topic t_k. '*j*' and '*k*' are the word and topic indices respectively. It is important to note that LDA ignores the order of occurrence of words and the syntactic information. It treats documents just as a collection of words or a bag of words.

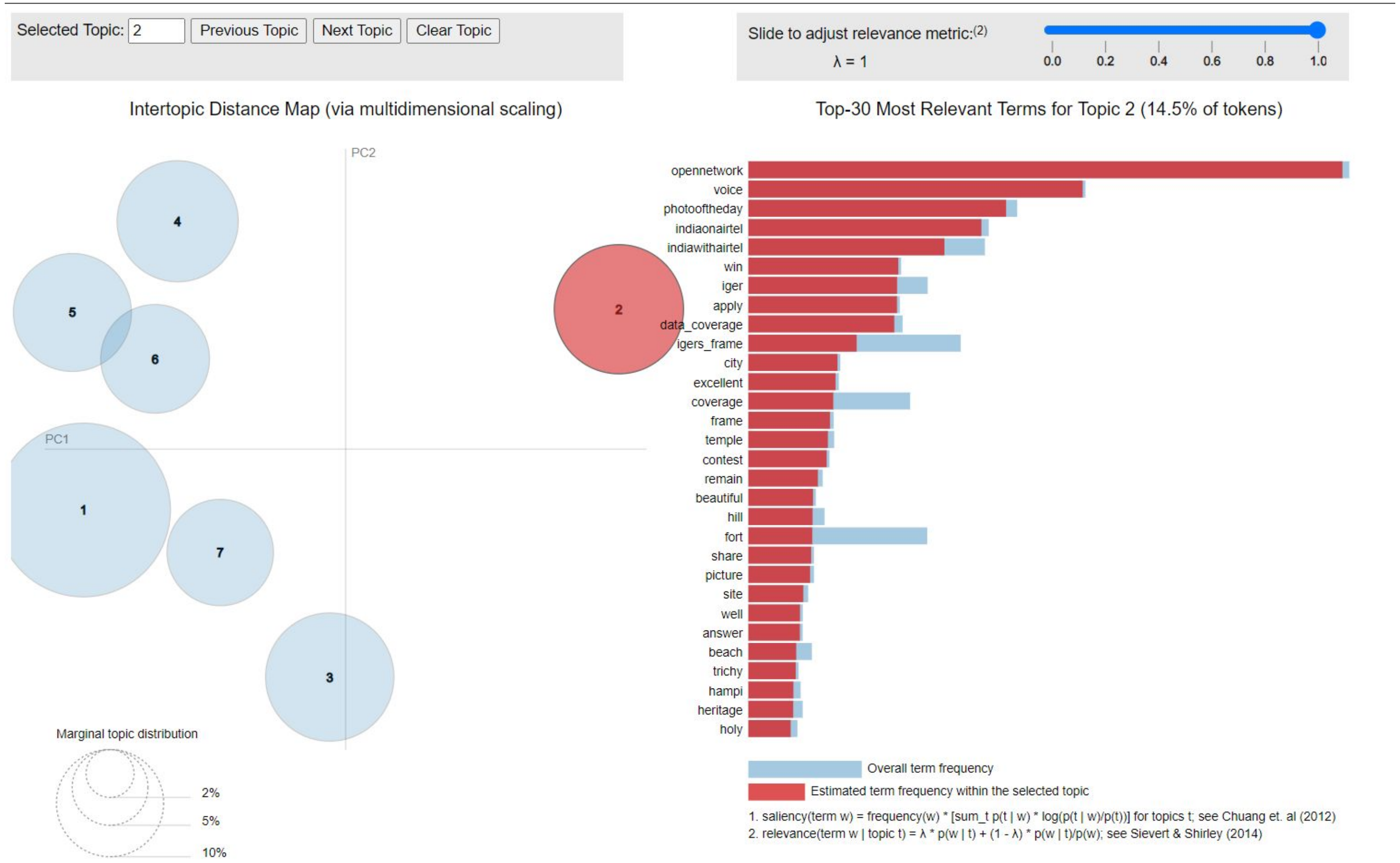
Word1 word2 word3 word4

	word1	word2	word3				Word-n
Topic-1	0.024	0.012	0.014	-	-	-	0.086
Topic-2	0.026	0.186	0.164	-	-	-	0.194
Topic-3	0.018	0.112	0.192	-	-	-	0.028
|	-	-	-	-	-	-	-
Topic-K	0.128	0.144	0.084	-	-	-	0.036

Continuation: Airtel India

Clicking on a bubble will highlight the relevant terms (on the right side) associated with that topic.

To view this (from Google Drive files):
1. Open “Visualization” folder
2. Download “airtelindia.html” and open in a browser (like Google Chrome)



Selected Topic: 2

Slide to adjust relevance metric:(2)

λ = 1

Intertopic Distance Map (via multidimensional scaling)

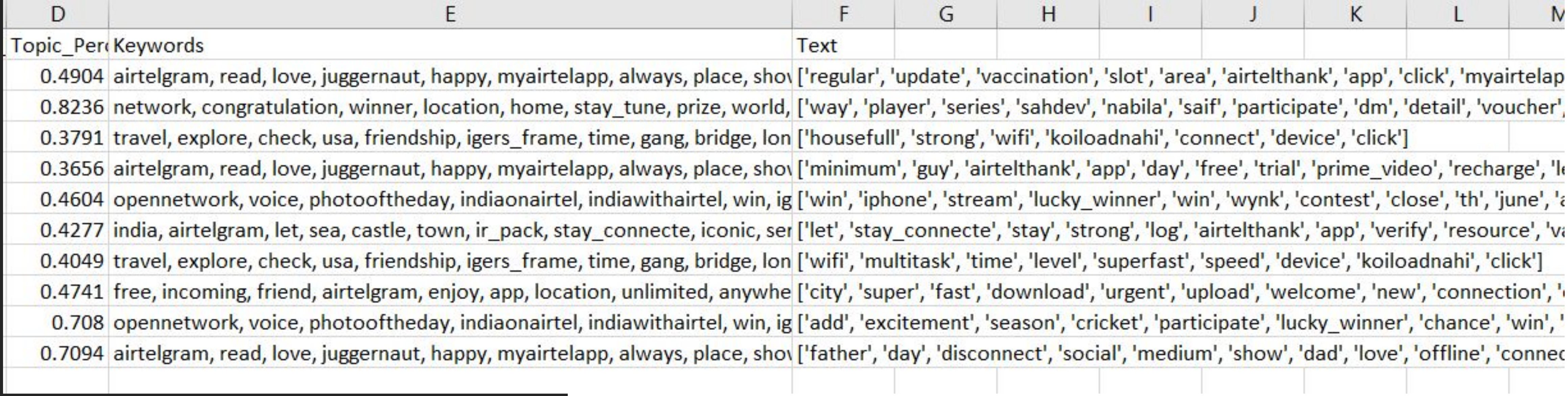
Top-30 Most Relevant Terms for Topic 2 (14.5% of tokens)

🟦 Overall term frequency

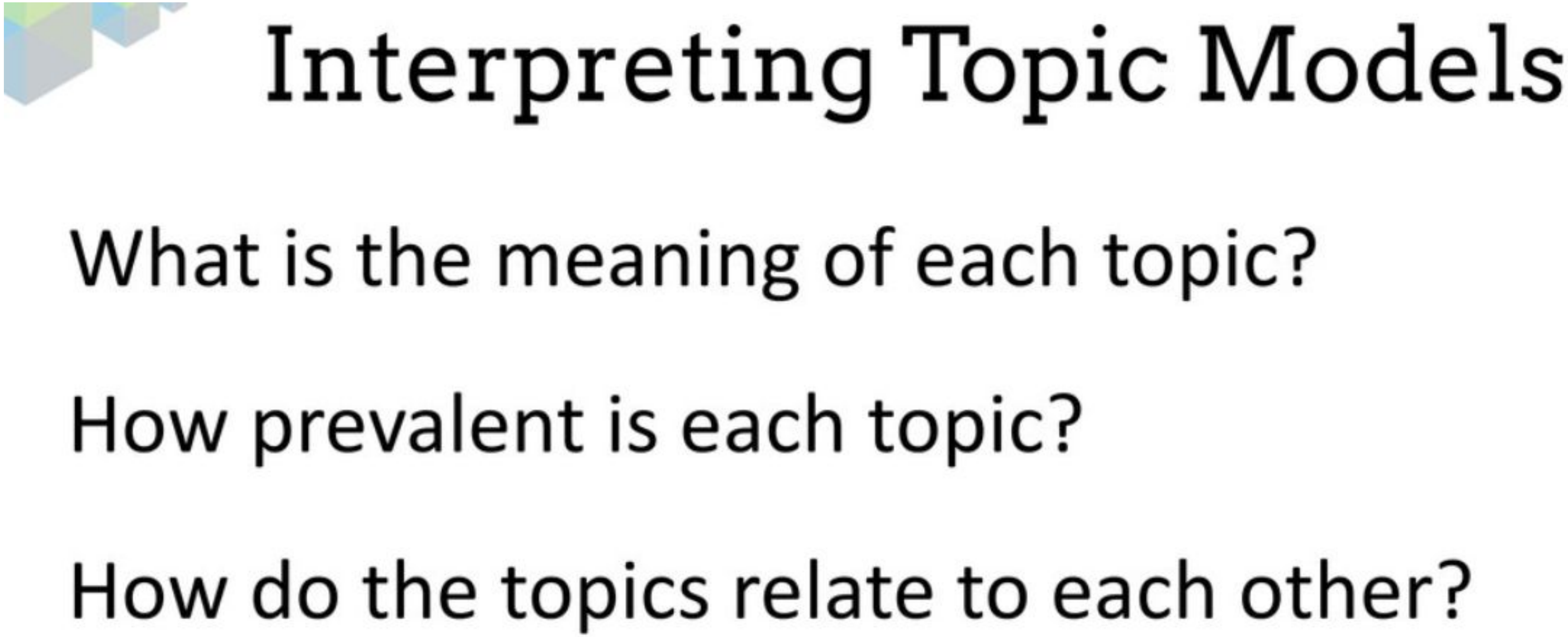
🟥 Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Keywords: Airtel India



D	E	F
Topic_Perc	Keywords	Text
0.4904	airtelgram, read, love, juggernaut, happy, myairtelapp, always, place, shov	['regular', 'update', 'vaccination', 'slot', 'area', 'airtelthank', 'app', 'click', 'myairtelap
0.8236	network, congratulation, winner, location, home, stay_tune, prize, world,	['way', 'player', 'series', 'sahdev', 'nabila', 'saif', 'participate', 'dm', 'detail', 'voucher',
0.3791	travel, explore, check, usa, friendship, igers_frame, time, gang, bridge, lon	['housefull', 'strong', 'wifi', 'koiloadnahi', 'connect', 'device', 'click']
0.3656	airtelgram, read, love, juggernaut, happy, myairtelapp, always, place, shov	['minimum', 'guy', 'airtelthank', 'app', 'day', 'free', 'trial', 'prime_video', 'recharge', 'l
0.4604	opennetwork, voice, photooftheday, indiaonairtel, indiawithairtel, win, ig	['win', 'iphone', 'stream', 'lucky_winner', 'win', 'wynk', 'contest', 'close', 'th', 'june', 'a
0.4277	india, airtelgram, let, sea, castle, town, ir_pack, stay_connecte, iconic, ser	['let', 'stay_connecte', 'stay', 'strong', 'log', 'airtelthank', 'app', 'verify', 'resource', 'va
0.4049	travel, explore, check, usa, friendship, igers_frame, time, gang, bridge, lon	['wifi', 'multitask', 'time', 'level', 'superfast', 'speed', 'device', 'koiloadnahi', 'click']
0.4741	free, incoming, friend, airtelgram, enjoy, app, location, unlimited, anywhe	['city', 'super', 'fast', 'download', 'urgent', 'upload', 'welcome', 'new', 'connection', '
0.708	opennetwork, voice, photooftheday, indiaonairtel, indiawithairtel, win, ig	['add', 'excitement', 'season', 'cricket', 'participate', 'lucky_winner', 'chance', 'win', '
0.7094	airtelgram, read, love, juggernaut, happy, myairtelapp, always, place, shov	['father', 'day', 'disconnect', 'social', 'medium', 'show', 'dad', 'love', 'offline', 'connec



Interpreting Topic Models

What is the meaning of each topic?

How prevalent is each topic?

How do the topics relate to each other?

Taken from: https://speakerdeck.com/bmabey/visualizing-topic-models?slide=21

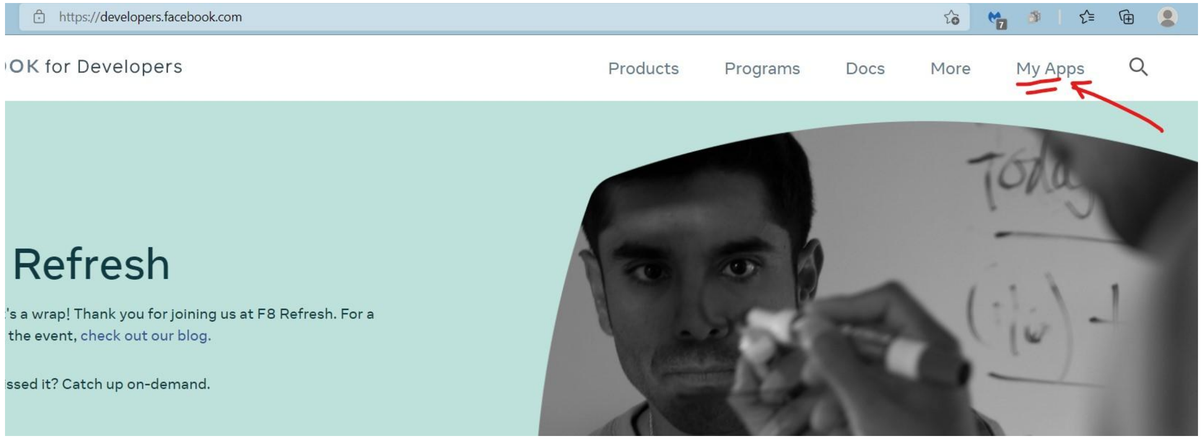
Appendix

Little technical details about the project

Data Extraction via API calls

Making Facebook Developer Account

www.developers.facebook.com

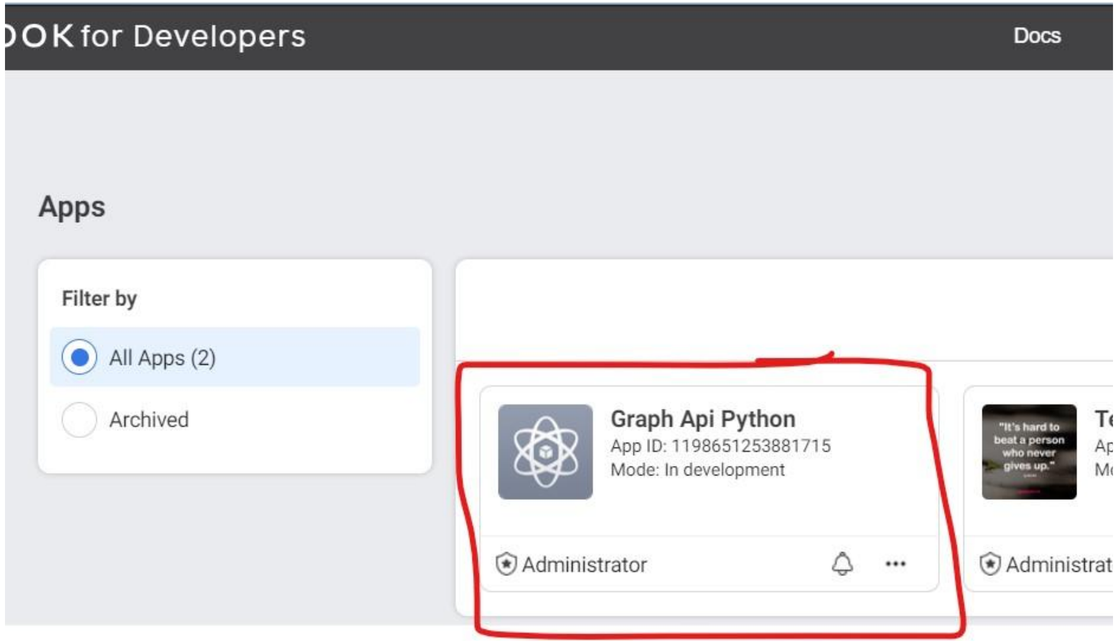


Refresh

's a wrap! Thank you for joining us at F8 Refresh. For a the event, check out our blog.

ssed it? Catch up on-demand.

Creating Test App and Test Page



● All Apps (2)

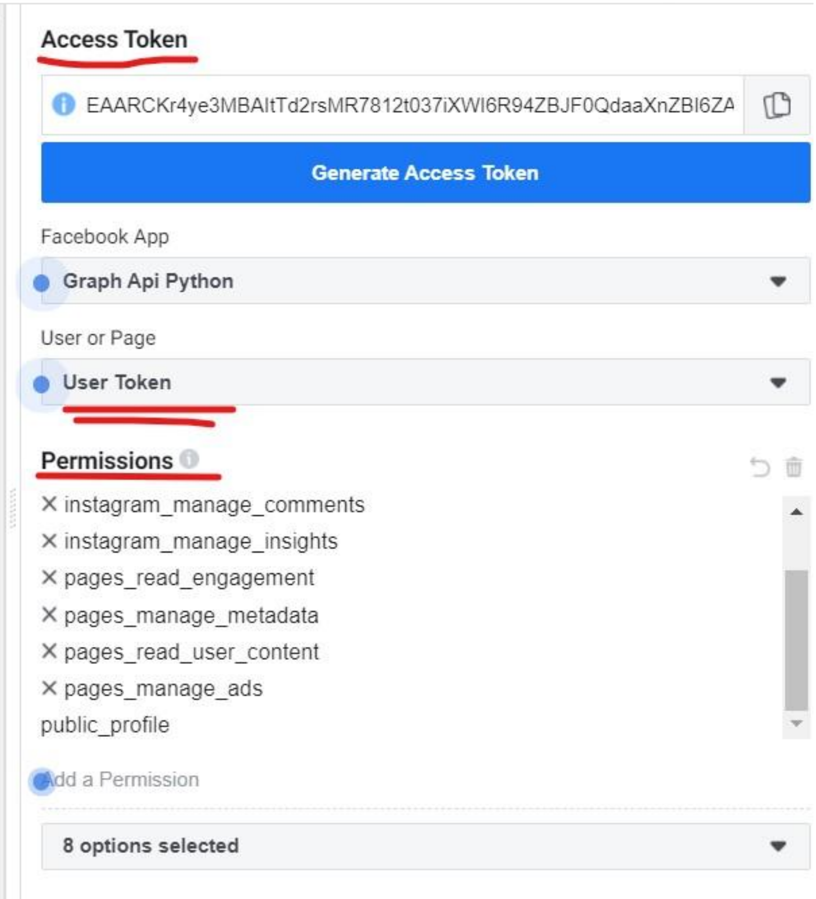
○ Archived

Graph Api Python
App ID: 1198651253881715
Mode: In development

Administrator

Administrat

Generating Access Token and Permissions



Access Token

EAARCKr4ye3MBAItTd2rsMR7812t037iXWI6R94ZBJF0QdaaXnZBI6ZA

Facebook App: Graph Api Python

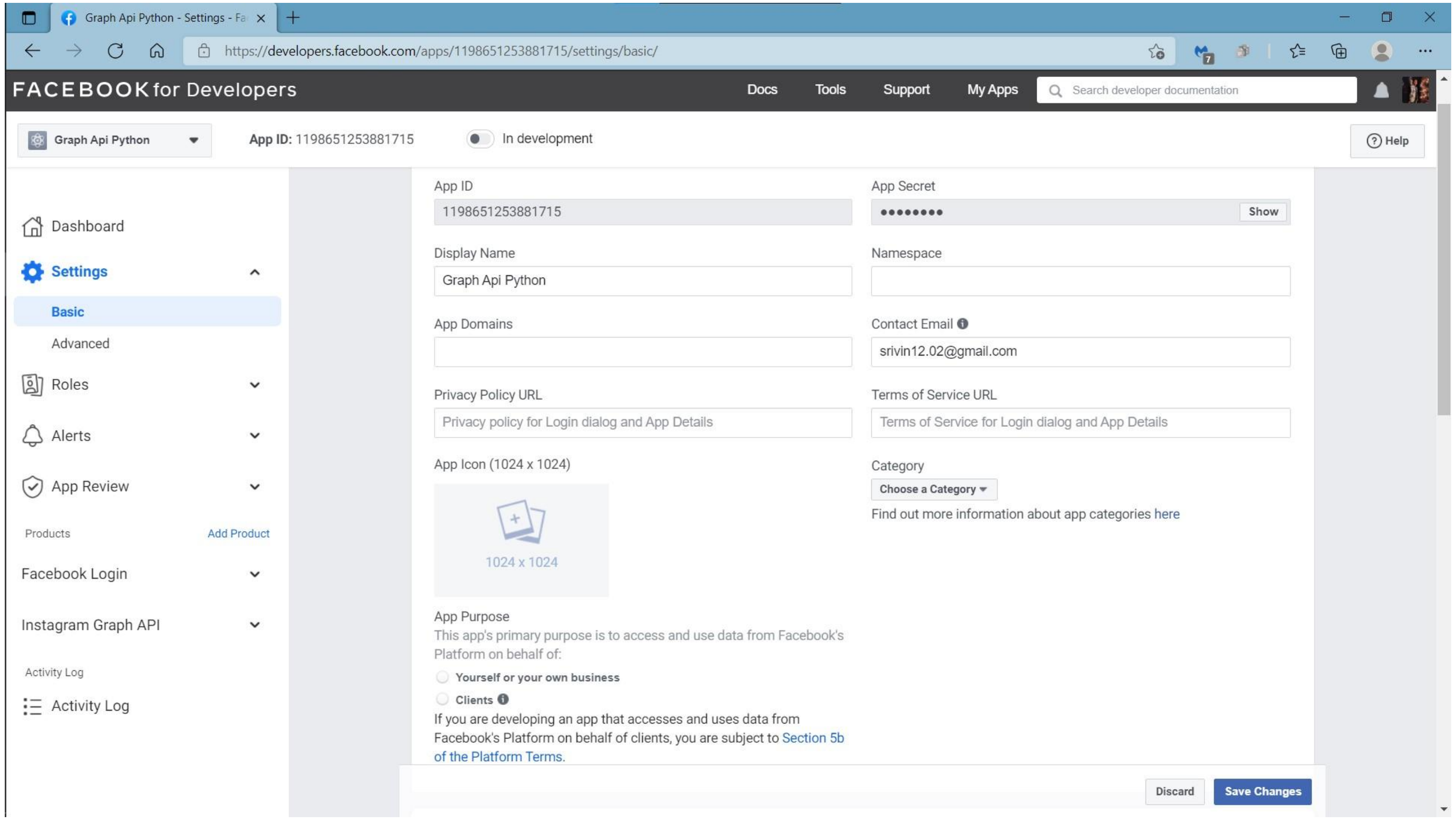
User or Page: User Token

Permissions

- ✕ instagram_manage_comments
- ✕ instagram_manage_insights
- ✕ pages_read_engagement
- ✕ pages_manage_metadata
- ✕ pages_read_user_content
- ✕ pages_manage_ads
- public_profile

Add a Permission

8 options selected



Graph Api Python - Settings - Fa

https://developers.facebook.com/apps/1198651253881715/settings/basic/

Graph Api Python

App ID: 1198651253881715

In development

App ID: 1198651253881715

App Secret: •••••••• Show

Display Name: Graph Api Python

Namespace:

App Domains:

Contact Email: srivin12.02@gmail.com

Privacy Policy URL: Privacy policy for Login dialog and App Details

Terms of Service URL: Terms of Service for Login dialog and App Details

App Icon (1024 x 1024)

Category: Choose a Category

Find out more information about app categories here

App Purpose

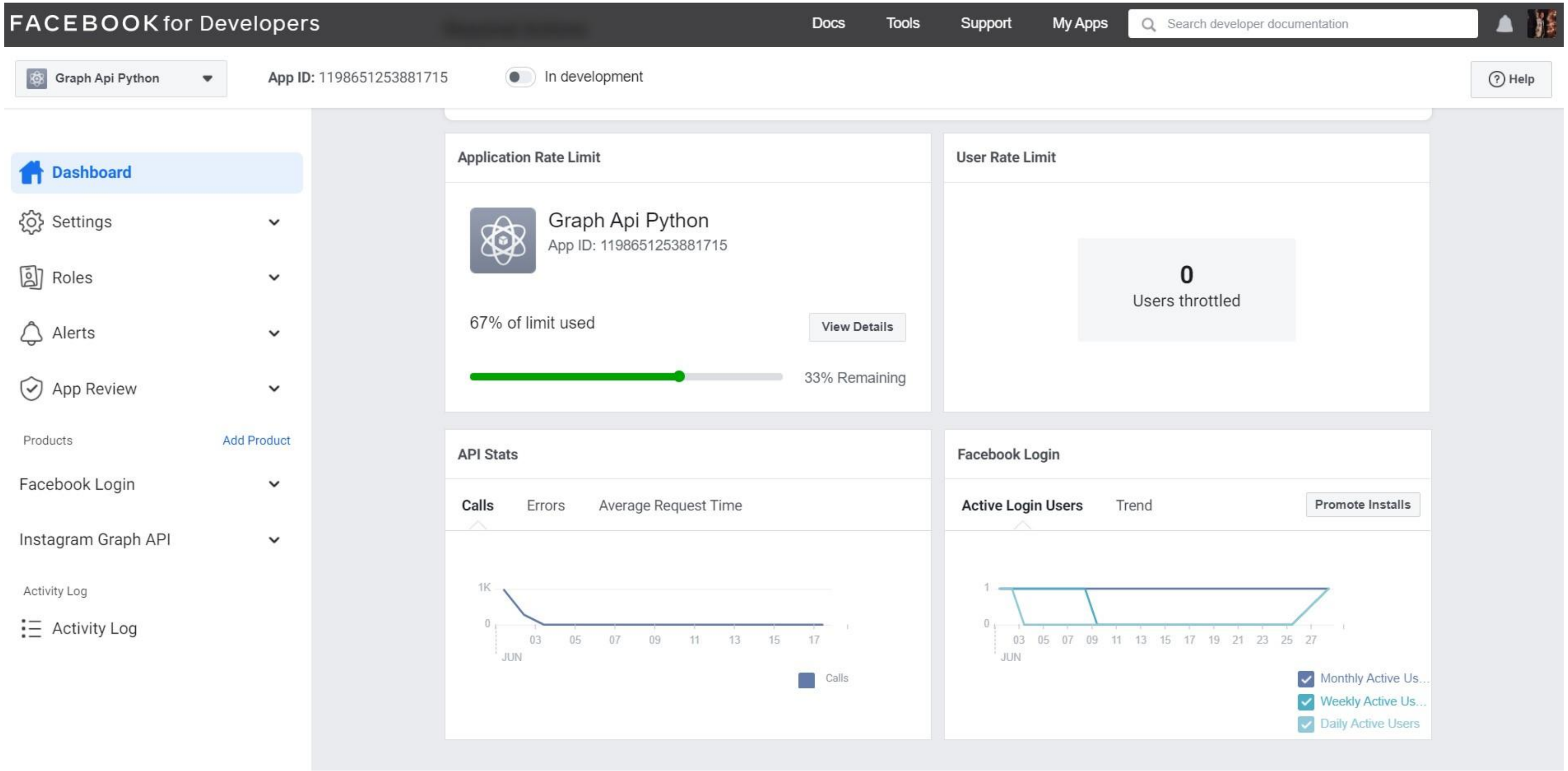
This app's primary purpose is to access and use data from Facebook's Platform on behalf of:

○ Yourself or your own business

○ Clients

If you are developing an app that accesses and uses data from Facebook's Platform on behalf of clients, you are subject to Section 5b of the Platform Terms.

Developer Home Page



Graph Api Python

App ID: 1198651253881715

In development

Application Rate Limit

Graph Api Python
App ID: 1198651253881715

67% of limit used

33% Remaining

User Rate Limit

0
Users throttled

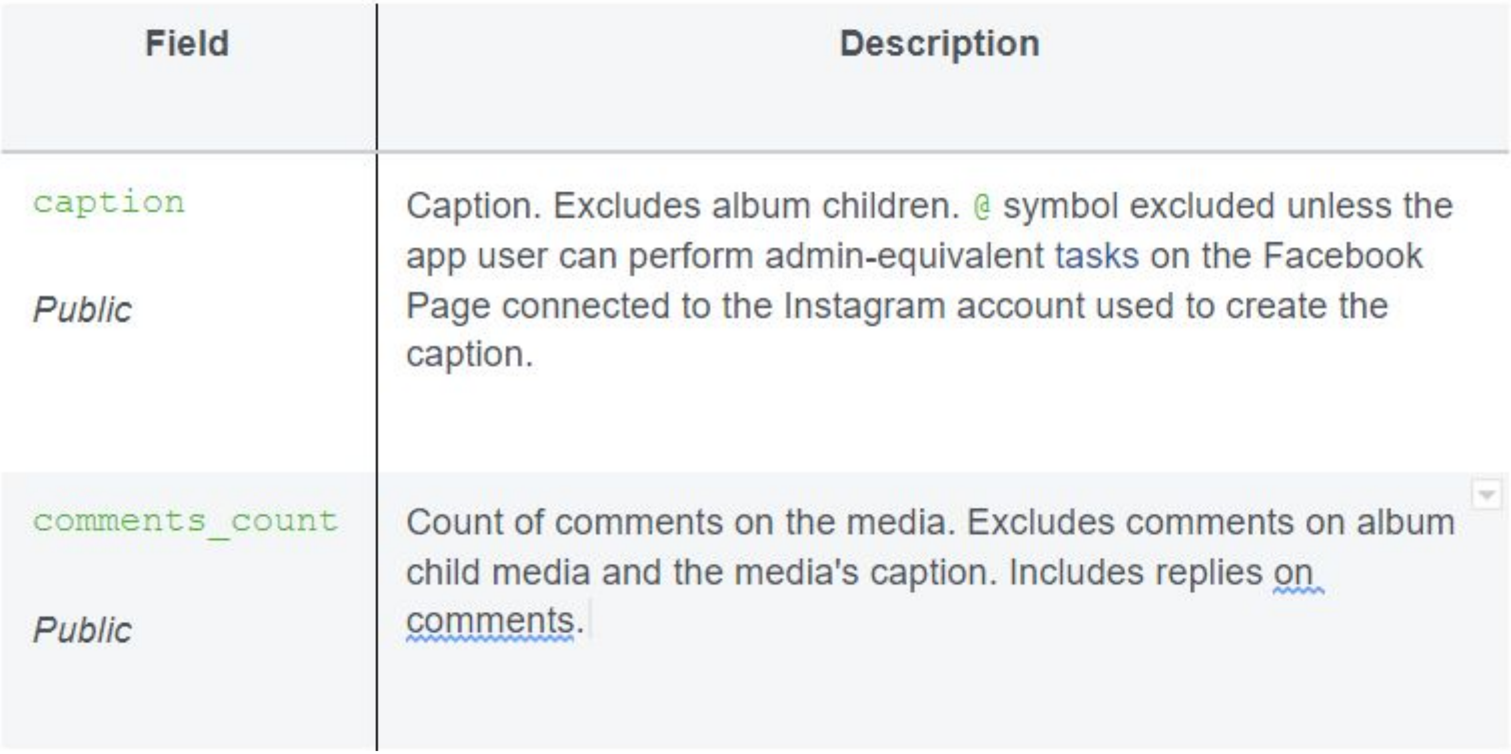
Instagram Graph API

An API delivers a user response to a system and sends the system's response back to a user

Returnable Fields for IG User
[IG User - Instagram Platform (facebook.com)]

- biography*
- id*
- ig_id
- followers_count*
- follows_count
- media_count*
- name
- profile_picture_url
- username*
- website*

Fields marked with an asterisk (*) are public field which means they can be returned by an edge using field expansion.

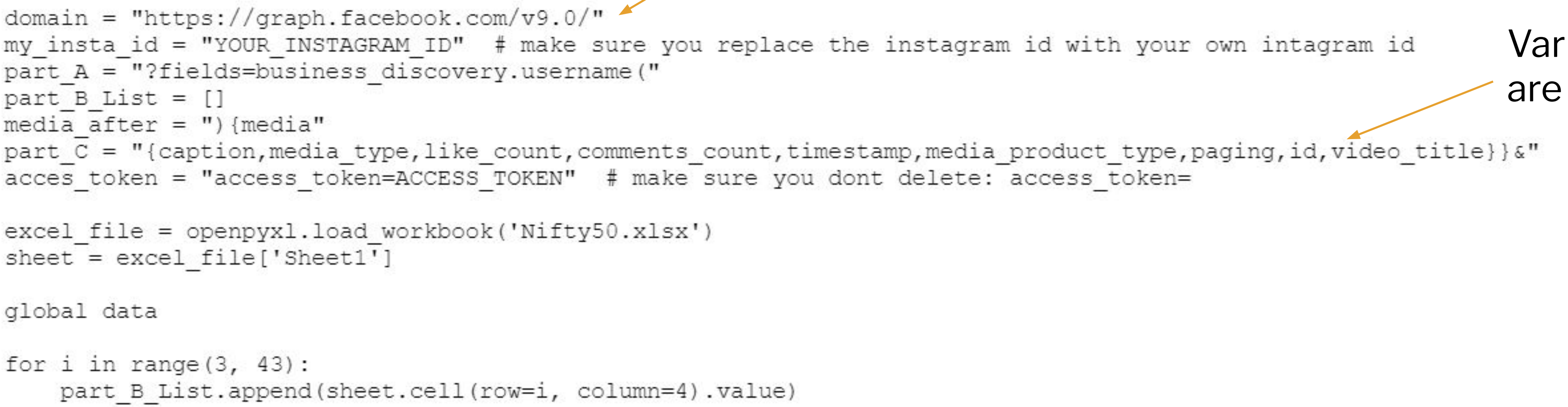


Field	Description
caption *Public*	Caption. Excludes album children. @ symbol excluded unless the app user can perform admin-equivalent tasks on the Facebook Page connected to the Instagram account used to create the caption.
comments_count *Public*	Count of comments on the media. Excludes comments on album child media and the media's caption. Includes replies on comments.

Making API Calls

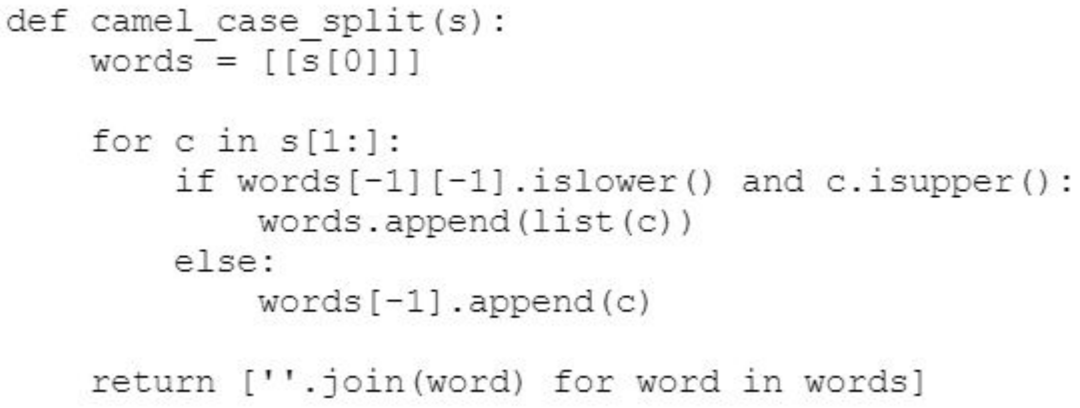
Facebook Graph API server

Variables that are required



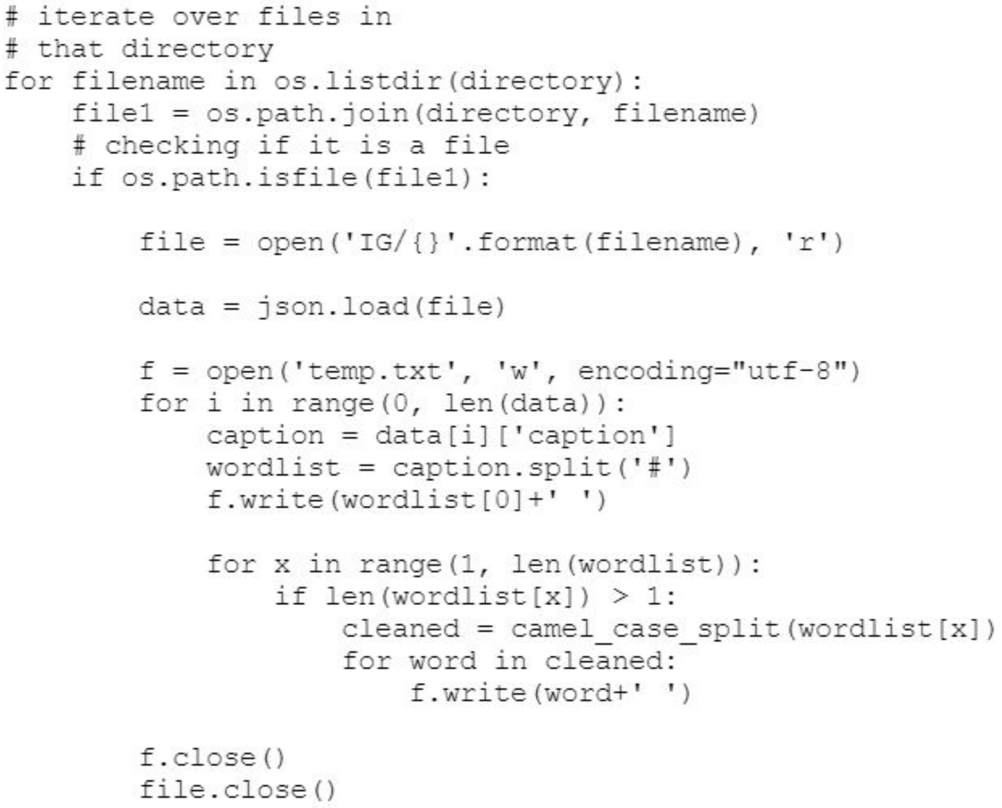
```
domain = "https://graph.facebook.com/v9.0/"
my_insta_id = "YOUR_INSTAGRAM_ID"  # make sure you replace the instagram id with your own intagram id
part_A = "?fields=business_discovery.username("
part_B_List = []
media_after = "){media"
part_C = "{caption,media_type,like_count,comments_count,timestamp,media_product_type,paging,id,video_title}}&"
acces_token = "access_token=ACCESS_TOKEN"  # make sure you dont delete: access_token=

excel_file = openpyxl.load_workbook('Nifty50.xlsx')
sheet = excel_file['Sheet1']

global data

for i in range(3, 43):
    part_B_List.append(sheet.cell(row=i, column=4).value)
```

Text Processing



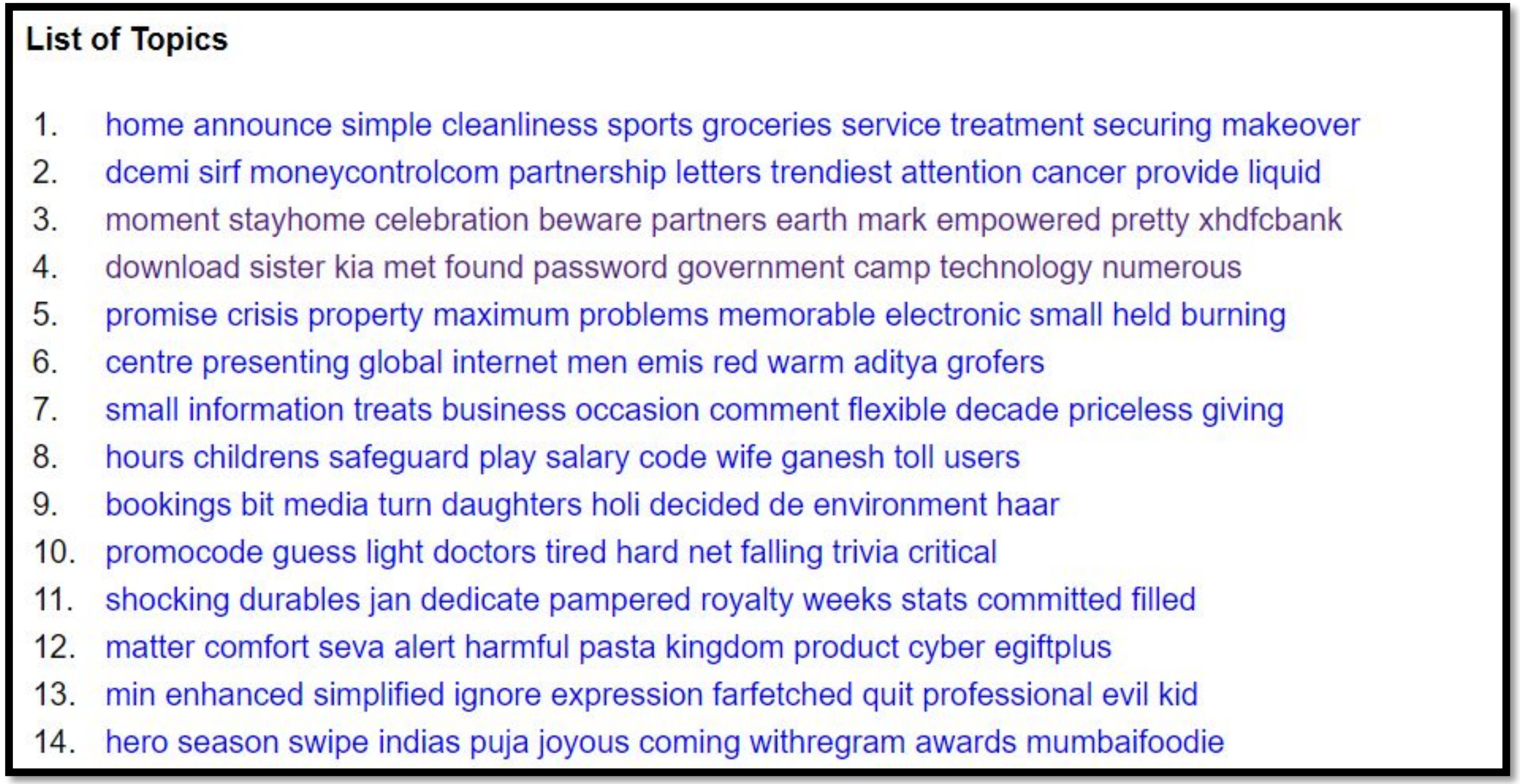
```
def camel_case_split(s):
    words = [[s[0]]]

    for c in s[1:]:
        if words[-1][-1].islower() and c.isupper():
            words.append(list(c))
        else:
            words[-1].append(c)

    return [''.join(word) for word in words]
```

Example : #WeStandFor



```
# iterate over files in
# that directory
for filename in os.listdir(directory):
    file1 = os.path.join(directory, filename)
    # checking if it is a file
    if os.path.isfile(file1):

        file = open('IG/{}'.format(filename), 'r')

        data = json.load(file)

        f = open('temp.txt', 'w', encoding="utf-8")
        for i in range(0, len(data)):
            caption = data[i]['caption']
            wordlist = caption.split('#')
            f.write(wordlist[0]+' ')

            for x in range(1, len(wordlist)):
                if len(wordlist[x]) > 1:
                    cleaned = camel_case_split(wordlist[x])
                    for word in cleaned:
                        f.write(word+' ')

        f.close()
        file.close()
```

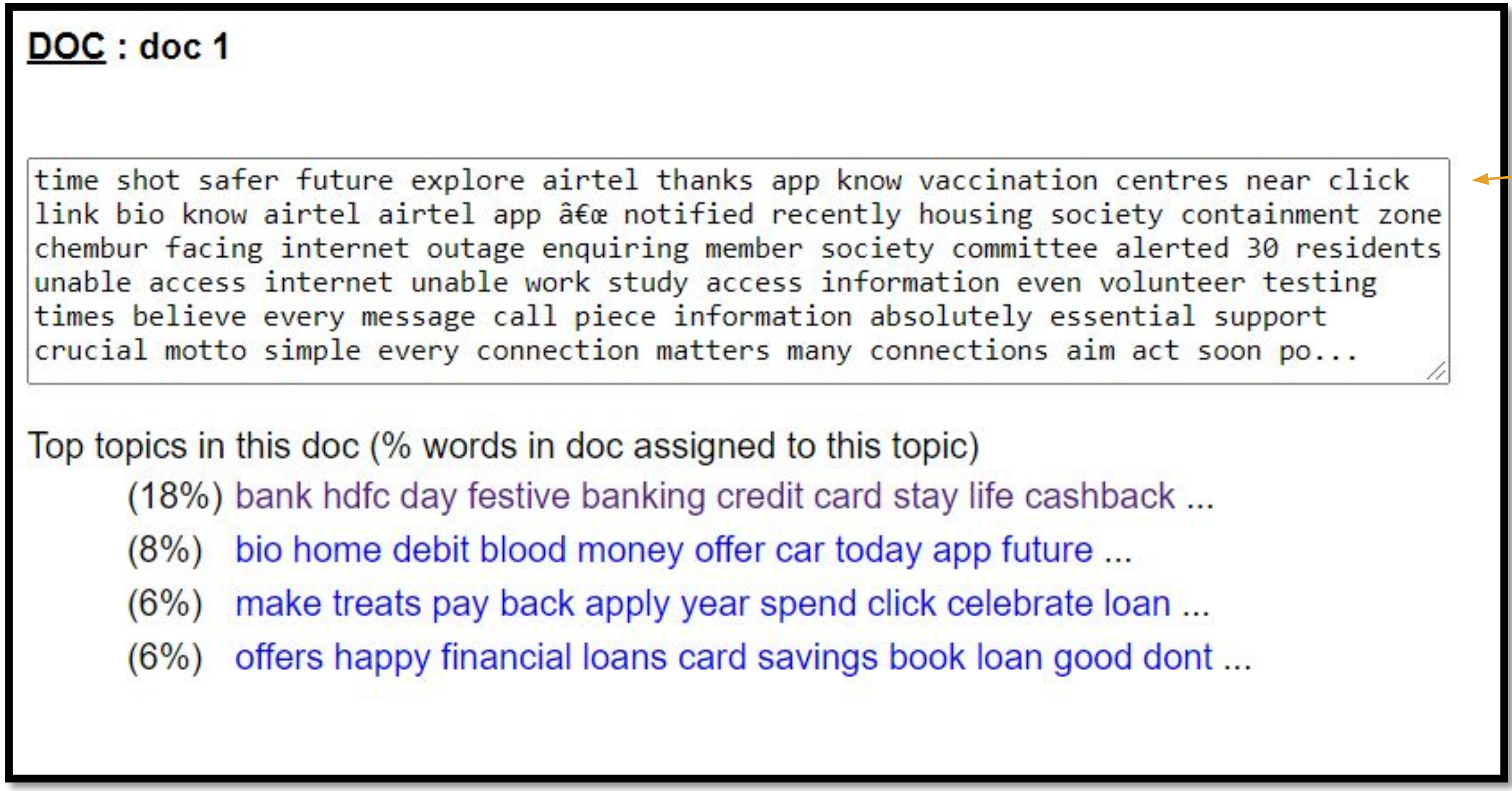
Sample Example from HDFC Bank



List of Topics

1. home announce simple cleanliness sports groceries service treatment securing makeover
2. dcemi sirf moneycontrolcom partnership letters trendiest attention cancer provide liquid
3. moment stayhome celebration beware partners earth mark empowered pretty xhdfcbank
4. download sister kia met found password government camp technology numerous
5. promise crisis property maximum problems memorable electronic small held burning
6. centre presenting global internet men emis red warm aditya grofers
7. small information treats business occasion comment flexible decade priceless giving
8. hours childrens safeguard play salary code wife ganesh toll users
9. bookings bit media turn daughters holi decided de environment haar
10. promocode guess light doctors tired hard net falling trivia critical
11. shocking durables jan dedicate pampered royalty weeks stats committed filled
12. matter comfort seva alert harmful pasta kingdom product cyber egiftplus
13. min enhanced simplified ignore expression farfetched quit professional evil kid
14. hero season swipe indias puja joyous coming withregram awards mumbaifoodie

Processed text



DOC : doc 1

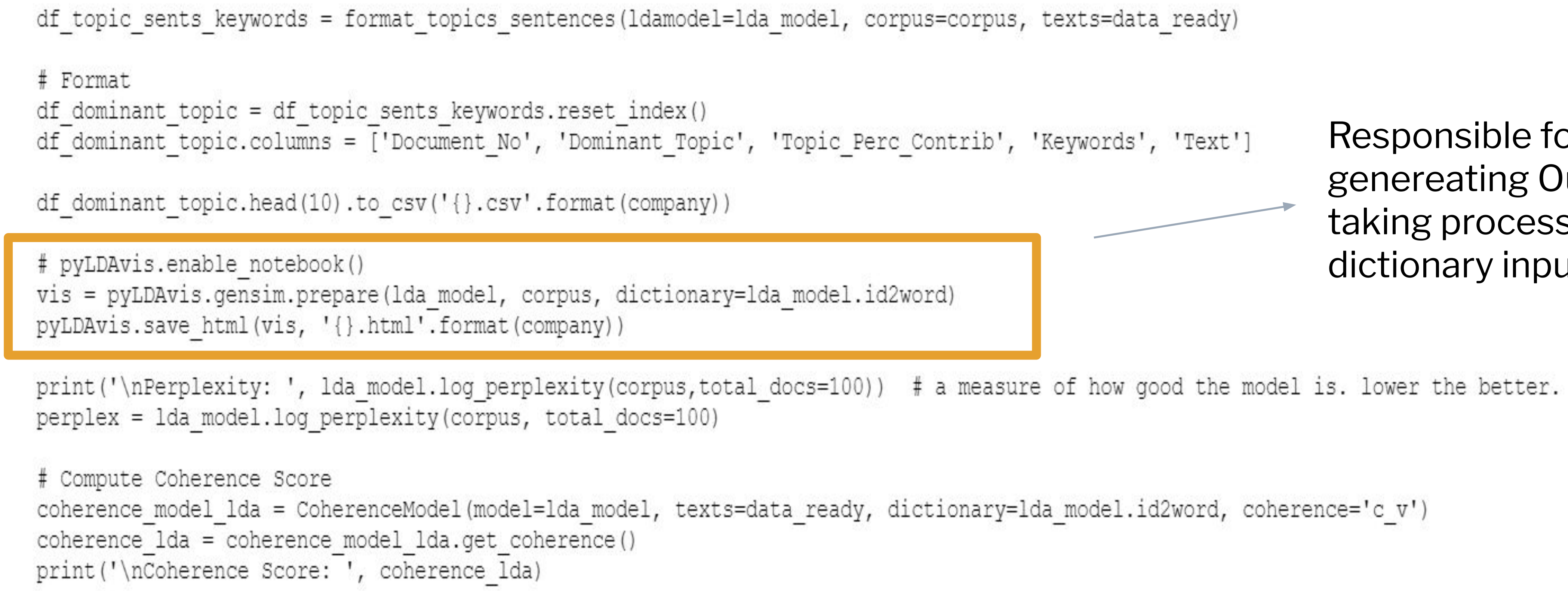
time shot safer future explore airtel thanks app know vaccination centres near click link bio know airtel airtel app â€œ notified recently housing society containment zone chembur facing internet outage enquiring member society committee alerted 30 residents unable access internet unable work study access information even volunteer testing times believe every message call piece information absolutely essential support crucial motto simple every connection matters many connections aim act soon po...

Top topics in this doc (% words in doc assigned to this topic)

(18%) bank hdfc day festive banking credit card stay life cashback ...
(8%) bio home debit blood money offer car today app future ...
(6%) make treats pay back apply year spend click celebrate loan ...
(6%) offers happy financial loans card savings book loan good dont ...

This contains the processed text from the captions of Airtel India

Visualizing Output



```
df_topic_sents_keywords = format_topics_sentences(ldamodel=lda_model, corpus=corpus, texts=data_ready)

# Format
df_dominant_topic = df_topic_sents_keywords.reset_index()
df_dominant_topic.columns = ['Document_No', 'Dominant_Topic', 'Topic_Perc_Contrib', 'Keywords', 'Text']

df_dominant_topic.head(10).to_csv('{}.csv'.format(company))

# pyLDAvis.enable_notebook()
vis = pyLDAvis.gensim.prepare(lda_model, corpus, dictionary=lda_model.id2word)
pyLDAvis.save_html(vis, '{}.html'.format(company))

print('\nPerplexity: ', lda_model.log_perplexity(corpus,total_docs=100))  # a measure of how good the model is. lower the better.
perplex = lda_model.log_perplexity(corpus, total_docs=100)

# Compute Coherence Score
coherence_model_lda = CoherenceModel(model=lda_model, texts=data_ready, dictionary=lda_model.id2word, coherence='c_v')
coherence_lda = coherence_model_lda.get_coherence()
print('\nCoherence Score: ', coherence_lda)
```

Responsible for genereating Output by taking processed text as its dictionary input