

Food Industry

S\$8.3b

annual receipts

S\$4.3b

annual contribution to GDP

48k

workers employed

*Increasing demand due to
growing urban population
& middle class*

Planned by **2020**

(Source: MTI's Food Industry Transformation Map)

4.5%

productivity growth

2,000

new jobs

BIMBO SUPREME PRATA HANDBAG



Stall A, a FoodieBuddie stall



Business Problem



What
should I
sell?

Which
influencer
should I work
with?



Website Demo

<http://gg.gg/IS434FoodieBuddie>



FoodieBuddie

G10 - Group 6

Terence Neo, Gracia Yuwono Kwantalalu,
Kok Jim Meng, Christal Poon

Our Team



TERENCE

Data Analyst
Data Visualisation



CHRISTAL

Business Gatherer
Digital Designer



GRACIA

Data Scraper
Data Visualisation



JIM MENG

Project Scheduler
Data Visualisation

Pipeline

Data Collection

- Data Scraping
 - Instagram posts scraping

Data Sources

Instagram

Data Engineering

Corpus Creation

- Wikipedia scraping
- Web scraping
- Named Entity Extraction
- Wordcloud
- Pareto Analysis

Data Cleaning Sentiment Analysis

Data Analysis

Social Network Graph

- Instagram followers scraping
- Gephi

Data Visualisation

Wikipedia

Food Blogs

Tech Stack

Data Collection

BeautifulSoup
Python

Data Engineering

Wikipedia API
BeautifulSoup
WordCloud
RegEx
nltk

Data Analysis

Tableau
Gephi
Phantombuster

01

Data Collection





Instagram

Top 12 Food Influencers in Singapore

1. danielfooddiary (239k)
2. misstamchiak (154k)
3. ieatishootipost (136k)
4. sethluicious (114k)
5. foodkingnoc (109k)
6. eatbooksg (75.7k)
7. stormscape (73.6k)
8. 8days_eat (64.8k)
9. hungrygowhere (62.7k)
10. burpple (61.4k)
11. sgfoodonfoot (54.9k)
12. sheeatsshecooks(53.6k)



Instagram

Database

Main Dataset

ID	Username	Caption	No. of Likes	No. of Comments
0	danielfooddiary	Katong's first..	326	7
1	danielfooddiary	K-bbq is back..	1439	11
2	danielfooddiary	This chicken..	575	8
...
1005	misstamchiak	Found this...	246	4
...



Instagram

Database

Comments Dataset

ID	Comment
0	@lauren.khoury still can't crack a smile 😊
0	@warren_tqy
0	@viggg @bottombracket @riardsg @pamelatlc oh n...
...	...
1039	😍😍😍
1039	@leeen ❤️



Instagram

Database

Followers Dataset

Followers	Foodies
kralice_ime_ne	danielfooddiary
auroracampp	danielfooddiary
er.nawati8113	danielfooddiary
...	...
mkkb98	misstamchiak
daisydan100	misstamchiak

09

Data Engineering





Corpus Creation: Wikipedia Scrape

Subcategories

This category has the following 51 subcategories, out of 51 total.

A

- ▶ [Afghan cuisine \(2 C, 38 P\)](#)
- ▶ [Armenian cuisine \(6 C, 47 P\)](#)
- ▶ [Azerbaijani cuisine \(6 C, 64 P\)](#)

B

- ▶ [Bahraini cuisine \(10 P\)](#)
- ▶ [Bangladeshi cuisine \(13 C, 61 P\)](#)
- ▶ [Bhutanese cuisine \(12 P\)](#)
- ▶ [Bruneian cuisine \(2 C, 32 P\)](#)
- ▶ [Burmese cuisine \(4 C, 94 P\)](#)

C

- ▶ [Cambodian cuisine \(2 C, 29 P\)](#)
- ▶ [Chinese cuisine \(33 C, 178 P\)](#)
- ▶ [Cypriot cuisine \(4 C, 66 P\)](#)

E

- ▶ [Egyptian cuisine \(5 C, 50 P\)](#)
- ▶ [Emirati cuisine \(5 C, 9 P\)](#)

G

- ▶ [Cuisine of Georgia \(country\) \(4 C, 52 P\)](#)

I

- ▶ [Indian cuisine \(32 C, 164 P, 1 F\)](#)
- ▶ [Indonesian cuisine \(33 C, 221 P\)](#)
- ▶ [Iranian cuisine \(15 C, 93 P\)](#)
- ▶ [Iraqi cuisine \(3 C, 80 P\)](#)

- ▶ [Israeli cuisine \(8 C, 159 P\)](#)

J

- ▶ [Japanese cuisine \(36 C, 244 P\)](#)
- ▶ [Jordanian cuisine \(1 C, 77 P\)](#)

K

- ▶ [Kazakhstani cuisine \(2 C, 31 P\)](#)
- ▶ [Kuwaiti cuisine \(11 P\)](#)
- ▶ [Kyrgyz cuisine \(34 P\)](#)

L

- ▶ [Lao cuisine \(7 C, 47 P\)](#)

M

- ▶ [Malaysian cuisine \(13 C, 172 P\)](#)
- ▶ [Maldivian cuisine \(2 C, 9 P\)](#)
- ▶ [Mongol cuisine \(3 C\)](#)
- ▶ [Mongolian cuisine \(2 C, 15 P\)](#)

N

- ▶ [Nepalese cuisine \(6 C, 92 P\)](#)
- ▶ [North Korean cuisine \(2 C, 7 P\)](#)

O

- ▶ [Omani cuisine \(8 P\)](#)

P

- ▶ [Pakistani cuisine \(33 C, 92 P\)](#)
- ▶ [Palestinian cuisine \(4 C, 84 P\)](#)
- ▶ [Philippine cuisine \(22 C, 184 P\)](#)

Q

- ▶ [Qatari cuisine \(6 P\)](#)

R

- ▶ [Russian cuisine \(20 C, 79 P\)](#)

S

- ▶ [Saudi Arabian cuisine \(1 C, 27 P\)](#)
- ▶ [Singaporean cuisine \(9 C, 94 P\)](#)
- ▶ [South Korean cuisine \(9 C, 12 P\)](#)
- ▶ [Sri Lankan cuisine \(17 C, 19 P\)](#)
- ▶ [Syrian cuisine \(7 C, 88 P\)](#)

T

- ▶ [Taiwanese cuisine \(9 C, 60 P\)](#)
- ▶ [Tajik cuisine \(15 P\)](#)
- ▶ [Thai cuisine \(15 C, 113 P\)](#)
- ▶ [Turkish cuisine \(20 C, 171 P\)](#)
- ▶ [Turkish Cypriot cuisine \(6 P\)](#)
- ▶ [Turkmenistan cuisine \(7 P\)](#)

U

- ▶ [Uzbekistani cuisine \(1 C, 26 P\)](#)

V

- ▶ [Vietnamese cuisine \(16 C, 123 P\)](#)

Y

- ▶ [Yemeni cuisine \(2 C, 41 P\)](#)



Wikipedia

Corpus Creation: Wikipedia API Results

Chinese , Corpus size = 33

```
['bak kut teh', 'beef kway teow', 'bak chang', 'bak chor mee', 'ban mian', 'chai tow kway', 'char kway teow', 'char siu', 'crab bee hoon', 'drunken prawns', 'duck rice', 'fish ball noodles', 'fish soup bee hoon', 'frog leg porridge', 'hae mee', 'ha inanese chicken rice', 'har cheong gai', 'hokkien mee', 'hum chim peng', 'kuay chap', 'mee pok', 'min chiang kueh', "pig's b rain soup", "pig's organ soup", 'popiah', 'shredded chicken noodles', 'sliced fish soup', 'soon kway', 'teochew porridge', 'turtle soup', 'vegetarian bee hoon', 'yong tau foo', 'youtiao']
```

Malay , Corpus size = 27

```
['acar', 'assam pedas', 'ayam penyet', 'bakso', 'begedil', 'curry puff', 'dendeng paru', 'goreng pisang', 'gulai daun ubi', 'keropok', 'ketupat', 'lemak siput', 'lontong', 'nagasari', 'nasi goreng', 'nasi padang', 'otak-otak', 'pecel lele', 'rawo n', 'rojak bandung', 'roti john', 'sambal', 'satay', 'sayur lodeh', 'soto', 'soto ayam', 'tumpeng']
```

Indian , Corpus size = 10

```
['appam', 'dosa', 'murtabak', 'naan', 'roti prata', 'soup kambing', 'soup tulang', 'soup tulang merah', 'tandoori chicken', 'vadai']
```

Cross-cultural , Corpus size = 19

```
['ayam buah keluak', 'biryani', 'cereal prawns', 'chili crab pasta', 'curry laksa', 'fish head curry', 'kari debal', 'kari l emak ayam', 'katong laksa', 'kueh pie tee', 'kway teow goreng', 'mee rebus', 'mee siam', 'mee goreng', 'mee soto', 'rojak', 'sambal kangkong', 'satay bee hoon', 'tau hu goreng']
```

Seafood , Corpus size = 5

```
['black pepper crab', 'chilli crab', 'oyster omelette', 'sambal lala', 'sambal stingray']
```

Fruit , Corpus size = 0

```
[]
```

Dataframe size = (112, 2)

Desserts , Corpus size = 9

```
['bubur cha cha', 'chendol', 'cheng tng', 'ice kacang', 'kuih or kueh', 'kueh lapis is a rich', 'orh-nee', 'pulut hitam', 'tau suan']
```

Drinks and beverages , Corpus size = 9

```
['bandung', 'chin chow drink', 'kopi', 'horlicks', 'milo', 'sugarcane juice', 'teh halia tarik', 'singapore sling', 'yuenyeung coffee']
```



Wikipedia

Corpus Creation: Wikipedia API Results

Chinese , Corpus size = 33

```
['bak kut teh', 'beef kway teow', 'bak chang', 'bak chor mee', 'ban mian', 'chai tow kway', 'char kway teow', 'char siu', 'crab bee hoon', 'drunken prawns', 'duck rice', 'fish ball noodles', 'fish soup bee hoon', 'frog leg porridge', 'hae mee', 'ha inanese chicken rice', 'har cheong gai', 'hokkien mee', 'hum chim peng', 'kuay chap', 'mee pok', 'min chiang kueh', "pig's b rain soup", "pig's organ soup", 'popiah', 'shredded chicken noodles', 'sliced fish soup', 'soon kway', 'teochew porridge', 'turtle soup', 'vegetarian bee hoon', 'yong tau foo', 'youtiao']
```

Malay , Corpus size = 27

```
['acar', 'assam pedas', 'ayam penyet', 'bakso', 'begedil', 'curry puff', 'dendeng paru', 'goreng pisang', 'gulai daun ubi', 'keropok', 'ketupat', 'lemak siput', 'lontong', 'nagasari', 'nasi goreng', 'nasi padang', 'otak-otak', 'pecel lele', 'rawo n', 'rojak bandung', 'roti john', 'sambal', 'satay', 'sayur lodeh', 'soto', 'soto ayam', 'tumpeng']
```

Indian , Corpus size = 10

```
['appam', 'dosa', 'murtabak', 'naan', 'roti prata', 'soup kambing', 'soup tulang', 'soup tulang merah', 'tandoori chicken', 'vadai']
```

Cross-cultural , Corpus size = 19

```
['ayam buah keluak', 'biryani', 'cereal prawns', 'chili crab pasta', 'curry laksa', 'fish head curry', 'kari debal', 'kari l emak ayam', 'katong laksa', 'kueh pie tee', 'kway teow goreng', 'mee rebus', 'mee siam', 'mee goreng', 'mee soto', 'rojak', 'sambal kangkong', 'satay bee hoon', 'tau hu goreng']
```

Seafood , Corpus size = 5

```
['black pepper crab', 'chilli crab', 'oyster omelette', 'sambal lala', 'sambal stingray']
```

Fruit , Corpus size = 0

```
[]
```

Dataframe size = (112, 2)

Desserts , Corpus size = 9

```
['bubur cha cha', 'chendol', 'cheng tng', 'ice kacang', 'kuih or kueh', 'kueh lapis is a rich', 'orh-nee', 'pulut hitam', 'tau suan']
```

Drinks and beverages , Corpus size = 9

```
['bandung', 'chin chow drink', 'kopi', 'horlicks', 'milo', 'sugarcane juice', 'teh halia tarik', 'singapore sling', 'yuenyeung coffee']
```



Corpus Creation: Alternative Names

	Food	Cuisine	alternative names	
110	kueh lapis	Desserts	kueh lapis	
111	lapis sagu	Desserts	lapis sagu	
112	beef kway teow	Chinese	kway teow	
113	hainanese chicken rice	Chinese	chicken rice	
114	shredded chicken noodles	Chinese	chicken noodles	(129, 3)
115	nasi goreng	Malay	nasigoreng	Chinese 36
116	nasi padang	Malay	nasipadang	Malay 32
117	nasi goreng	Malay	nasi-goreng	Cross-cultural 19
118	nasi padang	Malay	nasi-padang	Drinks and beverages 13
119	roti john	Malay	roti-john	Indian 13
120	roti prata	Indian	rotiprata	Desserts 11
121	roti prata	Indian	roti-prata	Seafood 5
122	roti prata	Indian	prata	Name: Cuisine, dtype: int64
123	kueh	Desserts	kuih	
124	kueh	Desserts	kueh	
125	chin chow drink	Drinks and beverages	chin chow	
126	chin chow drink	Drinks and beverages	chinchow	
127	chin chow drink	Drinks and beverages	chin-chow	
128	yuenyeung coffee	Drinks and beverages	yuenyeung	



Blogs

JAPAN CENTRE

Search from over 2,800 products



► 10% Student & 16-26 Discount >

< 96% Customer Satisfaction >



30 Must-Try Japanese Foods



Over the past few decades, Japan has made significant contributions to the culture of the UK and other Western countries.

As well as revolutionising our cars, computers, and cartoon characters; Japan has also broadened our culinary knowledge with the introduction of its delicious and highly unique cuisine. We at Japan Centre live and breathe Japanese cuisine, so we compiled a list of our top 30 recommendations for Japanese foods that everybody needs to try.

1. Sushi

String Parsing



The first foods that spring to mind when we think about Japanese cuisine. This delicacy was one of the first Japanese dishes to be exported to the US after the Meiji Restoration in 1868, and since then its popularity has steadily increased year after year. The word



South Korean Food: 29 of the Best Tasting Dishes

BeautifulSoup



Blogs

Corpus Creation: Initial Scrape Results

Japanese , Corpus size = 30

```
['sushi', 'udon', 'tofu', 'tempura', 'yakitori', 'sashimi', 'name  
n', 'donburi', 'natto', 'oden', 'tamagoyaki', 'soba', 'tonkatsu',  
'kashipan', 'sukiyaki', 'miso soup', 'okonomiyaki', 'mentaiko', 'ni  
kujaga', 'curry rice', 'unagi no kabayaki', 'shabu shabu', 'onigir  
i', 'gyoza', 'takoyaki', 'kaiseki ryori', 'edamame', 'yakisoba', 'c  
hawanmushi', 'wagashi']
```

Korean , Corpus size = 29

```
['Chili Pickled Cabbage (Kimchi 김치)', 'Samgyeopsal (삼겹살)', 'Por  
k Bulgogi (Daeji Bulgogi 불고기)', 'Korean Barbecue (Gogigui 고기구  
이)', 'Hangover Stew (Haejangguk 해장국)', 'Soft Tofu Stew (Sundubu  
\xa0jjigae 순두부찌개)', 'Mixed Seafood Stew', 'Kimchi Stew (Kimchi  
Jjigae\x00김치찌개)', 'Fish Stew (Saengseon Jjigae 생선찌개)', 'Spic  
y Stir Fried Octopus (Nakji Bokkeum\x00낙지볶음)', 'Korean Ox Bone  
Soup (Seolleongtang\x00설렁탕)', 'Hotpot Mixed Rice (Dolsot Bibimba  
p\x00돌솥 비빔밥)', 'Korean Mixed Rice (Cold Bibimbap 비빔밥)', 'Ste  
amed Mandu Dumplings (Jjinmandu 찐만두)', 'Deep Fried Mandu (Yaki M  
andu)', 'Noodles in Ice Soup (Mul Naengmyeon 둘 냉면)', 'Mixed Cold  
Noodles (Bibim Naengmyeon 비빔 냉면)', 'Kimchi Fried Rice (Kimchi B  
okkeumbap\x00김치 볶음밥)', 'Fried Sweet Potato Noodles (Japchae 잡  
채)', 'Mung Bean Pancake (Bindaetteok 빈대떡)', 'Korean Blood Sausa  
ge (Soondae 순대)', 'Octopus Mixed Plate', 'Gimbap 김밥', 'Korean C  
hicken Skewers (Dakkochi 닭꼬치)', 'Korean Side Dishes (Banchan반  
찬)', 'Tornado Potatoes', 'Gooey Deep Fried Snack (Hotteok 호떡)',  
'Korean Tempura (Twigim 튀김)', 'Red Rice Cakes (Tteokbokki 떡볶  
이)']
```



Blogs

Corpus Creation: Initial Scrape Results

Japanese , Corpus size = 30

```
['sushi', 'udon', 'tofu', 'tempura', 'yakitori', 'sashimi', 'name  
n', 'donburi', 'natto', 'oden', 'tamagoyaki', 'soba', 'tonkatsu',  
'kashipan', 'sukiyaki', 'miso soup', 'okonomiyaki', 'mentaiko', 'ni  
kujaga', 'curry rice', 'unagi no kabayaki', 'shabu shabu', 'onigir  
i', 'gyoza', 'takoyaki', 'kaiseki ryori', 'edamame', 'yakisoba', 'c  
hawanmushi', 'wagashi']
```

Korean , Corpus size = 29

```
['Chili Pickled Cabbage (Kimchi 김치)', 'Samgyeopsal (삼겹살)', 'Por  
k Bulgogi (Daeji Bulgogi 불고기)', 'Korean Barbecue (Gogigui 고기구  
이)', 'Hangover Stew (Haejangguk 해장국)', 'Soft Tofu Stew (Sundubu  
\xa0Jjigae 순두부찌개)', 'Mixed Seafood Stew', 'Kimchi Stew (Kimchi  
Jjigae\x00김치찌개)', 'Fish Stew (Saengseon Jjigae 생선찌개)', 'Spic  
y Stir Fried Octopus (Nakji Bokkeum\x00낙지볶음)', 'Korean Ox Bone  
Soup (Seolleongtang\x00설렁탕)', 'Hotpot Mixed Rice (Dolsot Bibimba  
p\x00돌솥 비빔밥)', 'Korean Mixed Rice (Cold Bibimbap 비빔밥)', 'Ste  
amed Mandu Dumplings (Jjinmandu 찐만두)', 'Deep Fried Mandu (Yaki M  
andu)', 'Noodles in Ice Soup (Mul Naengmyeon 둘 냉면)', 'Mixed Cold  
Noodles (Bibim Naengmyeon 비빔 냉면)', 'Kimchi Fried Rice (Kimchi B  
okkeumbap\x00김치 볶음밥)', 'Fried Sweet Potato Noodles (Japchae 잡  
채)', 'Mung Bean Pancake (Bindaetteok 빈대떡)', 'Korean Blood Sausa  
ge (Soondae 순대)', 'Octopus Mixed Plate', 'Gimbap 김밥', 'Korean C  
hicken Skewers (Dakkochi 닭꼬치)', 'Korean Side Dishes (Banchan반  
찬)', 'Tornado Potatoes', 'Gooey Deep Fried Snack (Hotteok 호떡)',  
'Korean Tempura (Twigim 튀김)', 'Red Rice Cakes (Tteokbokki 떡볶  
이)']
```



Blogs

Corpus Creation: Korean Foods

	Food	Cuisine	alternative names				
0	Chili Pickled Cabbage (Kimchi 김치)	Korean	chili pickled cabbage	21	Mixed Cold Noodles (Bibim Naengmyeon 비빔 냉면)	Korean	cold noodles
1	Samgyeopsal (삼겹살)	Korean	samgyeopsal	22	Kimchi Fried Rice (Kimchi Bokkeumbap 김치 볶음밥)	Korean	kimchi fried rice
2	Pork Bulgogi (Daeji Bulgogi 불고기)	Korean	bulgogi	23	Fried Sweet Potato Noodles (Japchae 짜개)	Korean	sweet potato noodles
3	Korean Barbecue (Gogigui 고기구이)	Korean	korean barbecue	24	Fried Sweet Potato Noodles (Japchae 짜개)	Korean	japchae
4	Korean Barbecue (Gogigui 고기구이)	Korean	korean bbq	25	Mung Bean Pancake (Bindaetteok 빈대떡)	Korean	mung bean pancake
5	Korean Barbecue (Gogigui 고기구이)	Korean	k bbq	26	Korean Blood Sausage (Soondae 순대)	Korean	korean blood sausage
6	Korean Barbecue (Gogigui 고기구이)	Korean	k bbq	27	Octopus Mixed Plate	Korean	octopus mixed plat
7	Korean Barbecue (Gogigui 고기구이)	Korean	k-bbq	28	Gimbap 김밥	Korean	gimbap
8	Hangover Stew (Haejangguk 해장국)	Korean	hangover stew	29	Korean Chicken Skewers (Dakkochi 닭꼬치)	Korean	korean chicken skewers
9	Soft Tofu Stew (Sundubu Jjigae 순두부찌개)	Korean	soft tofu stew	30	Korean Side Dishes (Banchan 반찬)	Korean	korean side dishes
10	Mixed Seafood Stew	Korean	mixed seafood stew	31	Tornado Potatoes	Korean	tornado potatoe
11	Kimchi Stew (Kimchi Jjigae 김치찌개)	Korean	kimchi stew	32	Tornado Potatoes	Korean	tornado potato
12	Fish Stew (Saengseon Jjigae 생선찌개)	Korean	fish stew	33	Gooey Deep Fried Snack (Hotteok 호떡)	Korean	gooey deep fried snack
13	Spicy Stir Fried Octopus (Nakji Bokkeum 낙지볶음)	Korean	fried octopus	34	Korean Tempura (Twigim 튀김)	Korean	korean tempura
14	Korean Ox Bone Soup (Seolleongtang 설렁탕)	Korean	ox bone soup	35	Red Rice Cakes (Tteokbokki 떡볶이)	Korean	red rice cakes
15	Mixed Rice (Bibimbap 비빔밥)	Korean	hotpot mixed rice	36	Red Rice Cakes (Tteokbokki 떡볶이)	Korean	tteokbokki
16	Mixed Rice (Bibimbap 비빔밥)	Korean	korean mixed rice	37	Korean Food	Korean	korean
17	Mixed Rice (Bibimbap 비빔밥)	Korean	bibimbap	38	Army Stew (Budae Jjigae)	Korean	army stew
18	Steamed Mandu Dumplings (Jjinmandu 찐만두)	Korean	mandu dumplings	39	Army Stew (Budae Jjigae)	Korean	budae jjigae
19	Deep Fried Mandu (Yaki Mandu)	Korean	deep fried mandu	40	korean fried chicken	Korean	korean fried chicken
20	Noodles in Ice Soup (Mul Naengmyeon 물 냉면)	Korean	ice soup	41	korean fried chicken	Korean	chir chir



Instagram

Corpus Creation: Dataset Pre-processing

Steps

- 1 Create an unique ID to easily identify posts
- 2 Convert all captions to lower case such that food names can be detected as all the names in our corpus are lower case names.
- 3 Removal of giveaways posts by changing the caption to "giveaway".

*In such posts, the influencer asks followers to participate in a lucky draw by liking and/or commenting on the post. These posts artificially inflate the numbers of likes and comments and hence perceived popularity of the food.

Example: IG Giveaway

[closed] here's a little something to spice up your life. @kebabsfaktory brings us an all-new flavour: mala kebabs. from 29 june 2020, you can turn any kebab into a sizzling hot mala kebab.

the mala lamb shawarma was exceptionally fiery – together with the juicy, gamey lamb meat, the tongue-tingling mala spices made me break out in a sweat. you can also get these mala flavours in their signature rolls, salads and rice boxes.

and for 5 lucky winners, we'll be giving away s\$20 vouchers so that you can enjoy this all-new mala sensation for yourself!

here's how to win:

1. like our post.
2. follow @sethluicious and @kebabsfaktory
3. comment why you'd like to win the vouchers.
4. tag a friend!

giveaway ends at 11.59pm on 1 july 2020. winner will be announced the following day (2 july 2020) at 11am. stay tuned to see if you have won. all the best!



Corpus Creation: Evaluation of Labelling

```
▶ print(df["food"].unique())
print(df["food"].nunique())
```

```
['hokkien mee' None 'acar' 'kopi' 'yong tau foo' 'hainanese chicken rice'
 'bak chang' 'tofu' 'kuih' 'ramen' 'pulut hitam' 'edamame' 'milo'
 'biriyani' 'korean fried chicken' 'char kway teow' 'gyoza' 'mee pok'
 'sashimi' 'tonkatsu' 'duck rice' 'roti prata' 'teochew porridge'
 'chai tow kway' 'satay' 'chilli crab' 'assam pedas' 'appam' 'sambal'
 'oden' 'hum chim peng' 'dosa' 'udon' 'char siu' 'curry puff'
 'shabu shabu' 'bubur cha cha' 'sukiyaki' 'keropok' 'sushi'
 'har cheong gai' 'cereal prawns' 'rawon' 'soto' 'oyster omelette'
 'bandung' 'bak kut teh' 'yakitori' 'popiah' 'frog leg porridge'
 'ban mian' 'curry rice' 'horlicks' 'crab bee hoon' 'fish head curry'
 'mee siam' 'chawanmushi' 'tempura' 'ketupat' 'rojak']
```

59

```
▶ print(df["cuisine"].unique())
print(df["cuisine"].nunique())
```

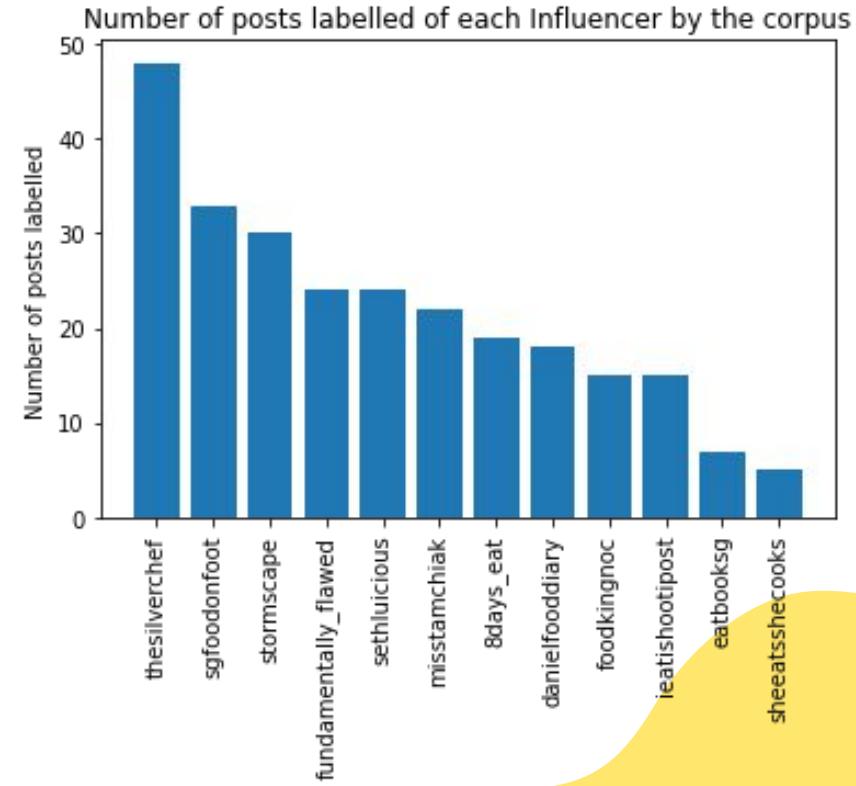
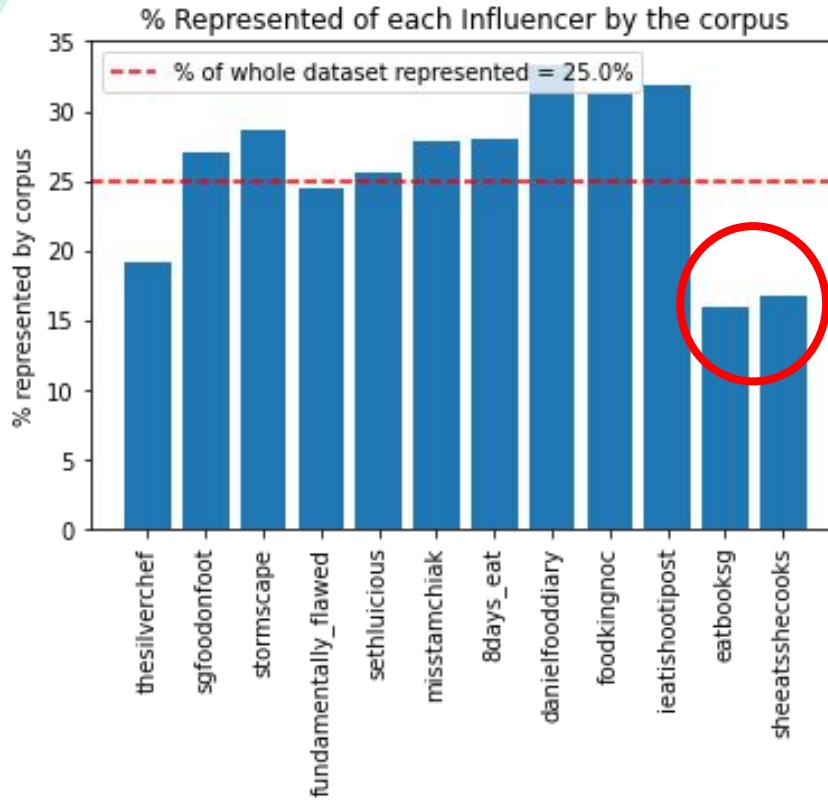
```
['Chinese' None 'Malay' 'Drinks and beverages' 'Japanese' 'Desserts'
 'Cross-cultural' 'Korean' 'Indian' 'Seafood']
```

9



Instagram

Corpus Creation: Evaluation of Labelling





Corpus Creation: Data Cleaning

Steps

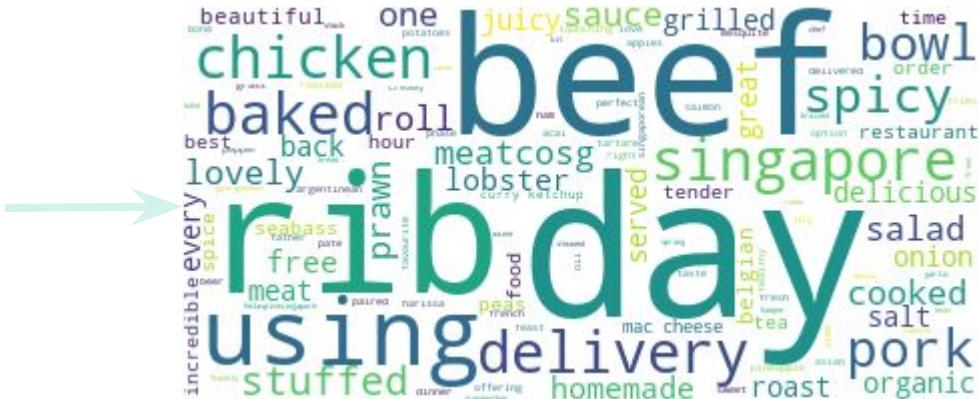
- 1 Regular expression to remove punctuation
- 2 Check that all text are already in lower case
- 3 Remove stop words with NLTK



Pareto Analysis

Named Entity Extraction

Corpus Creation: Word Frequency



word	freq
day	1.000000
rib	0.666667
beef	0.666667
using	0.555556
delivery	0.555556
chicken	0.555556
singapore	0.555556
baked	0.555556
one	0.444444
roll	0.444444
prawn	0.444444
cooked	0.444444
lovely	0.444444
sauce	0.444444
salad	0.444444
stuffed	0.444444
bowl	0.444444
spicy	0.444444
pork	0.444444
meatcosg	0.444444



Pareto Analysis

word	freq
day	1.000000
rib	0.666667
beef	0.666667
using	0.555556
delivery	0.555556
chicken	0.555556
singapore	0.555556
baked	0.555556
one	0.444444
roll	0.444444
prawn	0.444444
cooked	0.444444
lovely	0.444444
sauce	0.444444
salad	0.444444
stuffed	0.444444
bowl	0.444444
spicy	0.444444
pork	0.444444
meatcsg	0.444444

1. Select top food words
2. Get full food names from captions
3. Check if food names are already in corpus but named differently



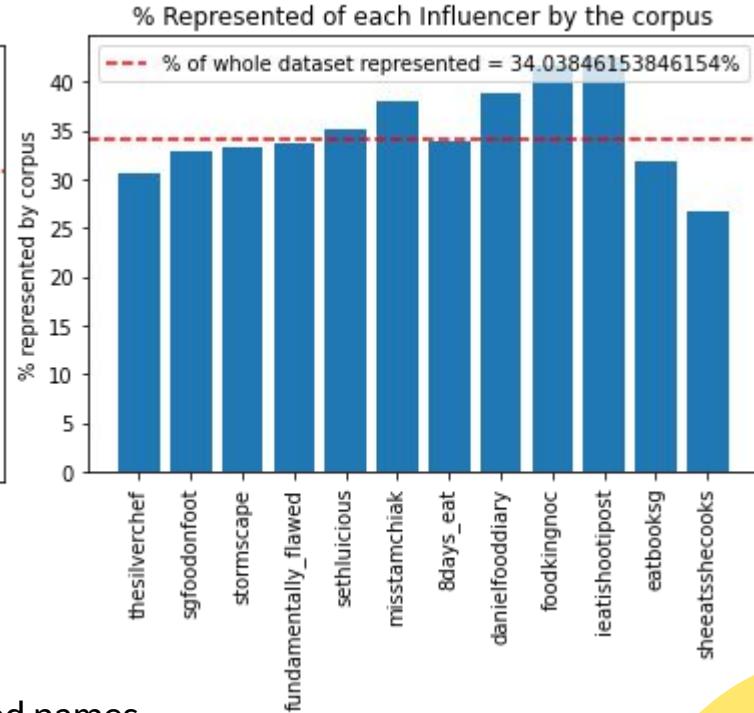
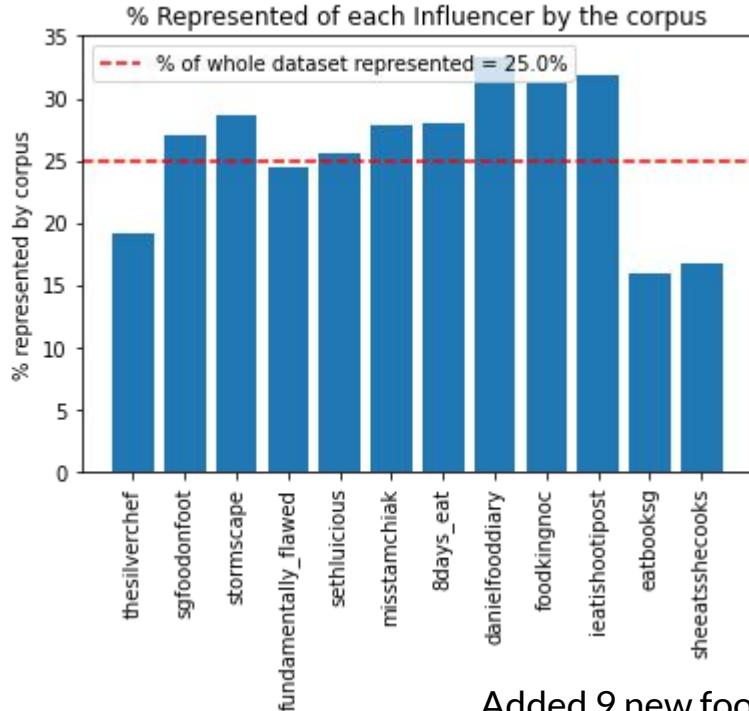
	Food	Cuisine	alternative names
0	beef tartare	Western	beef tartare
1	rib	Western	rib
2	roast chicken	Western	roast chicken
3	wagyu	Japanese	wagyu
4	scones	Snacks	scones
5	cookies	Snacks	cookies
6	salted egg	Snacks	salted egg
7	poke bowl	Japanese	poke bowl
8	unagi no kabayaki	Japanese	unagi
9	rib	Western	ribyte
10	rib	Western	ribs
11	roast chicken	Western	fried chicken

Increased corpus size
to 213 food names



Pareto Analysis

Corpus Creation: Results



Added 9 new food names

Overall representation:	25%	to	34%
Eatbooksg:	15.9%	to	31.8%
Sheeatsshecooks:	16.7%	to	26.7%



Instagram

Data Labelling!

Main Dataset

ID	Username	Caption	No. of Likes	No. of Comments	Food	Cuisine
0	danielfooddiary	Katong's first..	326	7	Teriyaki	Japanese
1	danielfooddiary	K-bbq is back..	1439	11	Steak	Western
2	danielfooddiary	This chicken..	575	8	Dumpling	Chinese
...
1005	misstamchiak	Found this...	246	4	Chicken Rice	Chinese
...



Instagram
Comments

Sentiment Analysis

Comments Dataset

ID	Comment
0	@lauren.khoury still can't crack a smile 😊
0	@warren_tqy
0	@viggg @bottombracket @riardsg @pamelatlc oh n...
...	...
1039	😍😍😍
1039	@leeen 😍



Instagram Comments

Sentiment Analysis: Comment Score

ID	comment	neg	neu	pos	compound	ID	neg	neu	pos	compound
0	@lauren.khoury still can't crack a smile 😢	0.000	0.615	0.385	0.3612	count	6676.000000	6676.000000	6676.000000	6676.000000
0	@warren_tqy	0.000	1.000	0.000	0.0000	mean	497.138107	0.012212	0.748302	0.200391
0	@viggg @bottombracket @riardsg @pamelatlc oh n...	0.000	1.000	0.000	0.0000	std	291.553523	0.060263	0.320179	0.282378
0	@in_vincentble luckily we went there last week...	0.000	0.615	0.385	0.7835	min	0.000000	0.000000	0.000000	-0.817600
0	It is my go-to char kway teow bee Hoon spot. A...	0.149	0.810	0.041	-0.7333	25%	250.000000	0.000000	0.516750	0.000000
...	50%	465.000000	0.000000	1.000000	0.000000
1039	Wow yummy	0.000	0.000	1.000	0.8020	75%	742.250000	0.000000	1.000000	0.398250
1039	Thank you for the recipe!	0.000	0.589	0.411	0.4199	max	1039.000000	1.000000	1.000000	1.000000
1039	Wahh a lot of work... but indeed worth the has...	0.000	0.773	0.227	0.3291					
1039	😊😊😊	0.000	1.000	0.000	0.0000					
1039	@leean 🍔	0.000	1.000	0.000	0.0000					



Instagram Comments

Sentiment Analysis: Post Score

ID	Username	Caption	No. of Likes	No. of Comments	Sentiment Score
0	danielfooddiary	Katong's first..	326	7	0.082280
1	danielfooddiary	K-bbq is back..	1439	11	0.000000
2	danielfooddiary	This chicken..	575	8	-0.371407
...
1005	misstamchiak	Found this...	246	4	0.456950
...

count 1040.000000
 mean 0.228807
 std 0.186286
 min -0.170500
 25% 0.076422
 50% 0.218460
 75% 0.337104
 max 0.992100
 Name: sentiment score

03

Data Analysis

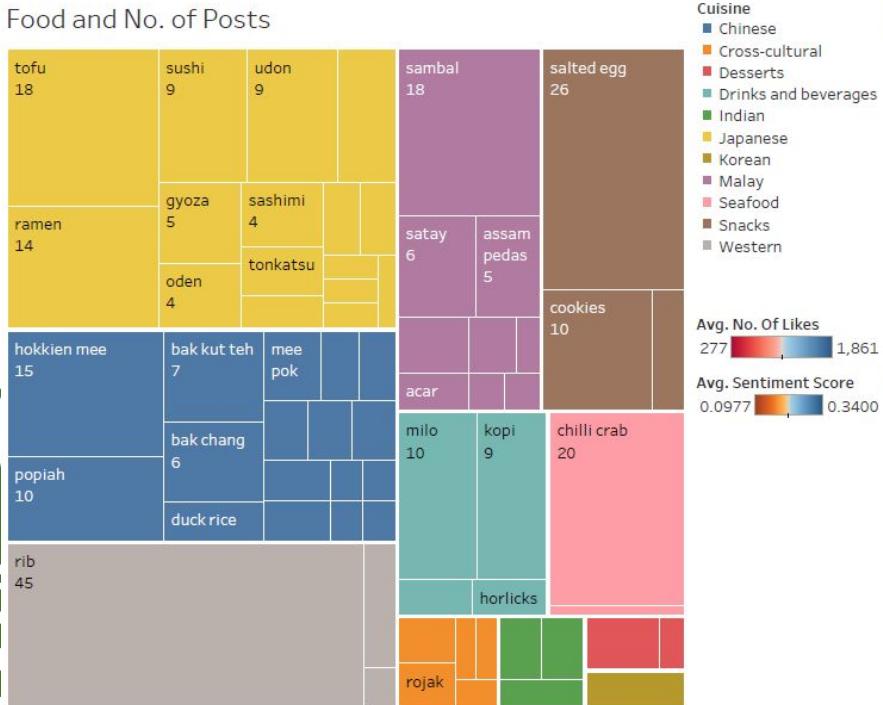




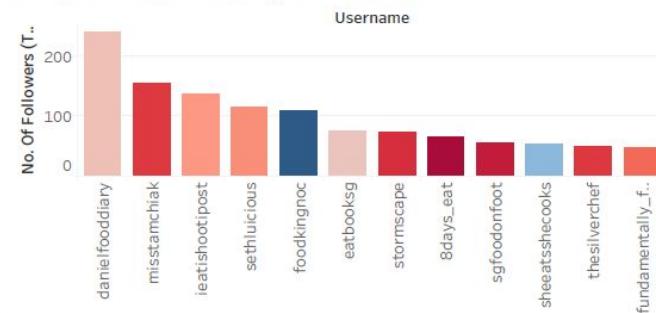
New Business Owner

Tableau

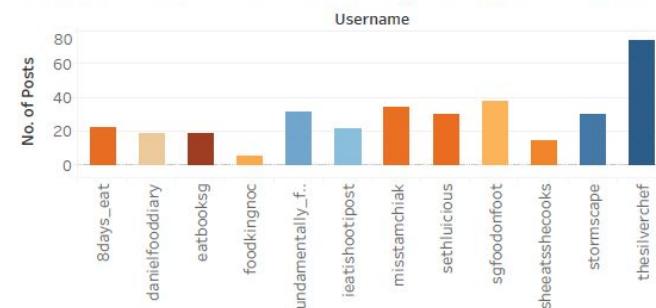
Food and No. of Posts



Number of followers per influencer



Influencers' No. of Posts and Avg. Sentiment Scores

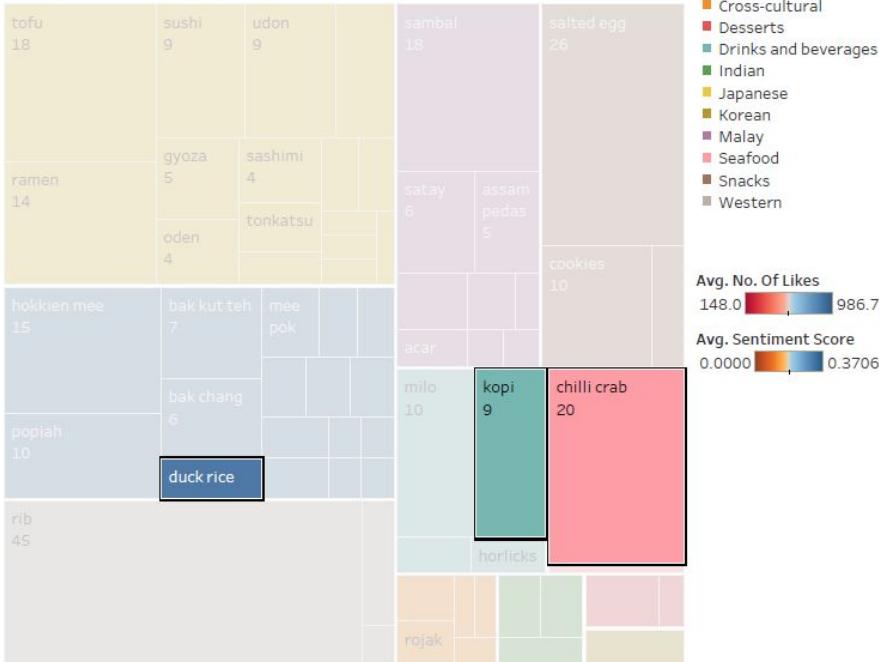




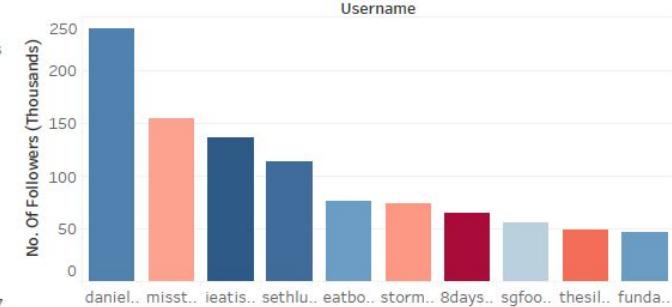
New Business Owner

Tableau

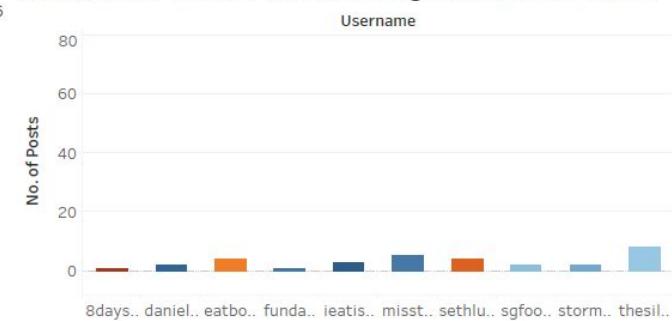
Food and No. of Posts



Number of followers per influencer



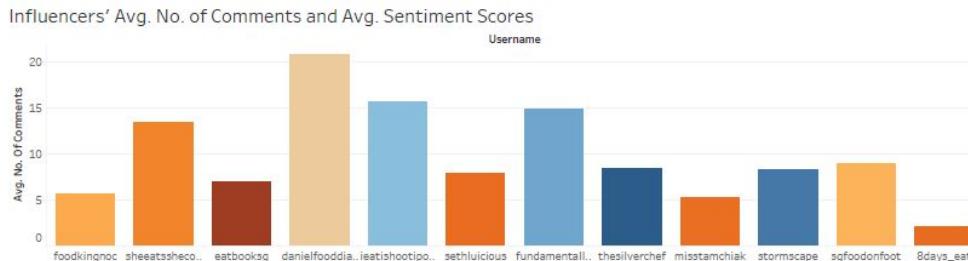
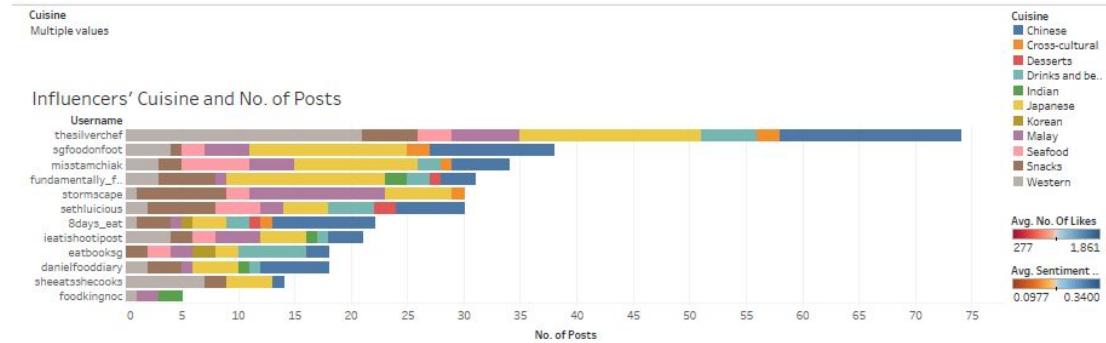
Influencers' No. of Posts and Avg. Sentiment Scores





Existing Business Owner

Tableau



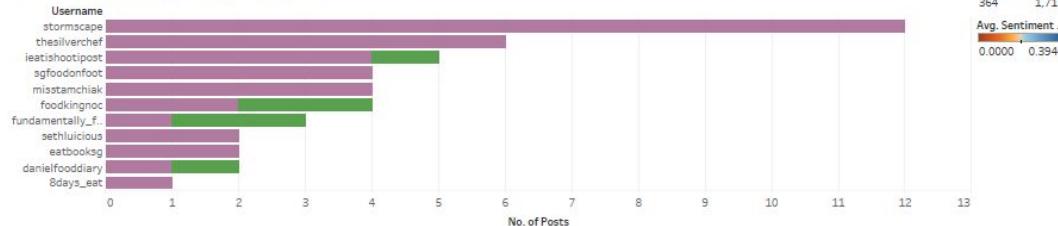


Existing Business Owner

Tableau

Cuisine
Multiple values

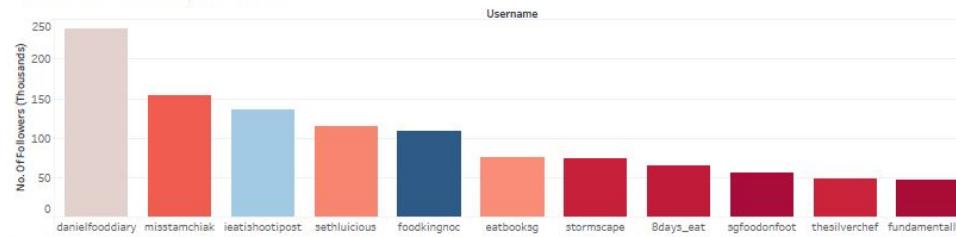
Influencers' Cuisine and No. of Posts



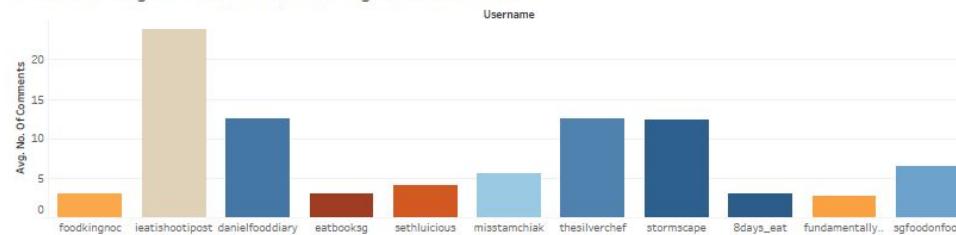
Cuisine
Indian
Malay

Avg. No. Of Likes
364 1.717
Avg. Sentiment ..
0.0000 0.3946

Number of followers per influencer



Influencers' Avg. No. of Comments and Avg. Sentiment Scores



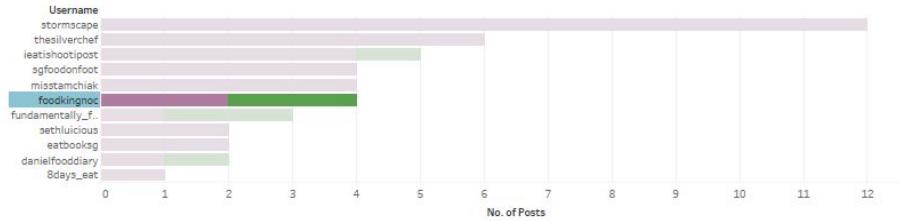


Existing Business Owner

Tableau

Cuisine
Multiple values

Influencers' Cuisine and No. of Posts



Cuisine
Indian
Malay

Avg. No. Of Likes
1.717
Avg. Sentiment ..
0.1553

Number of followers per influencer



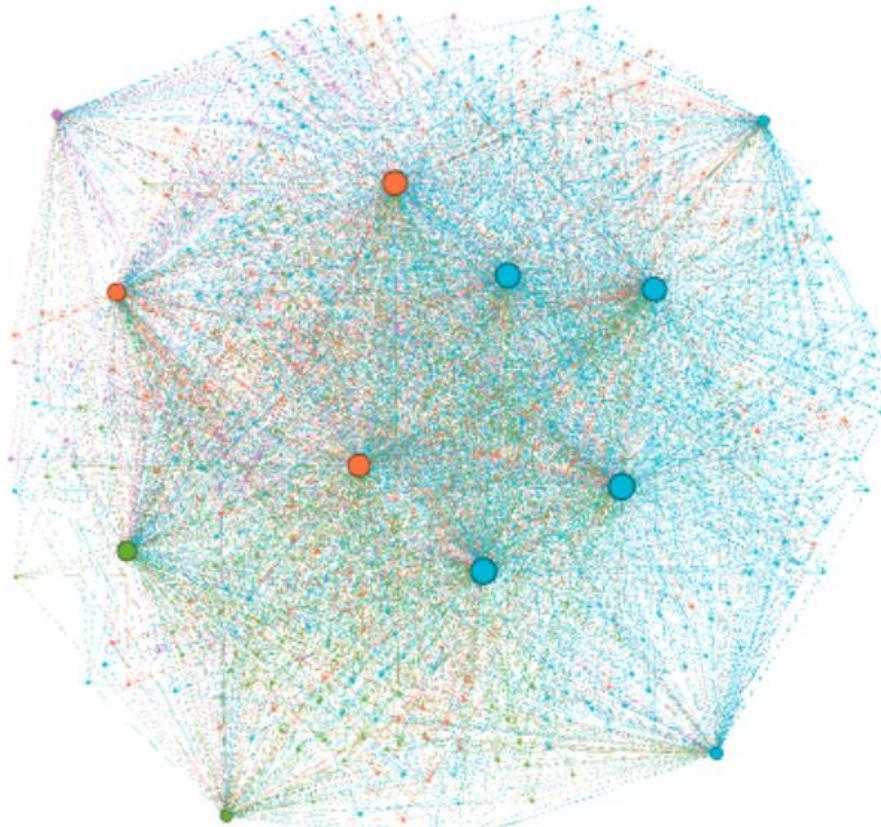
Influencers' Avg. No. of Comments and Avg. Sentiment Scores



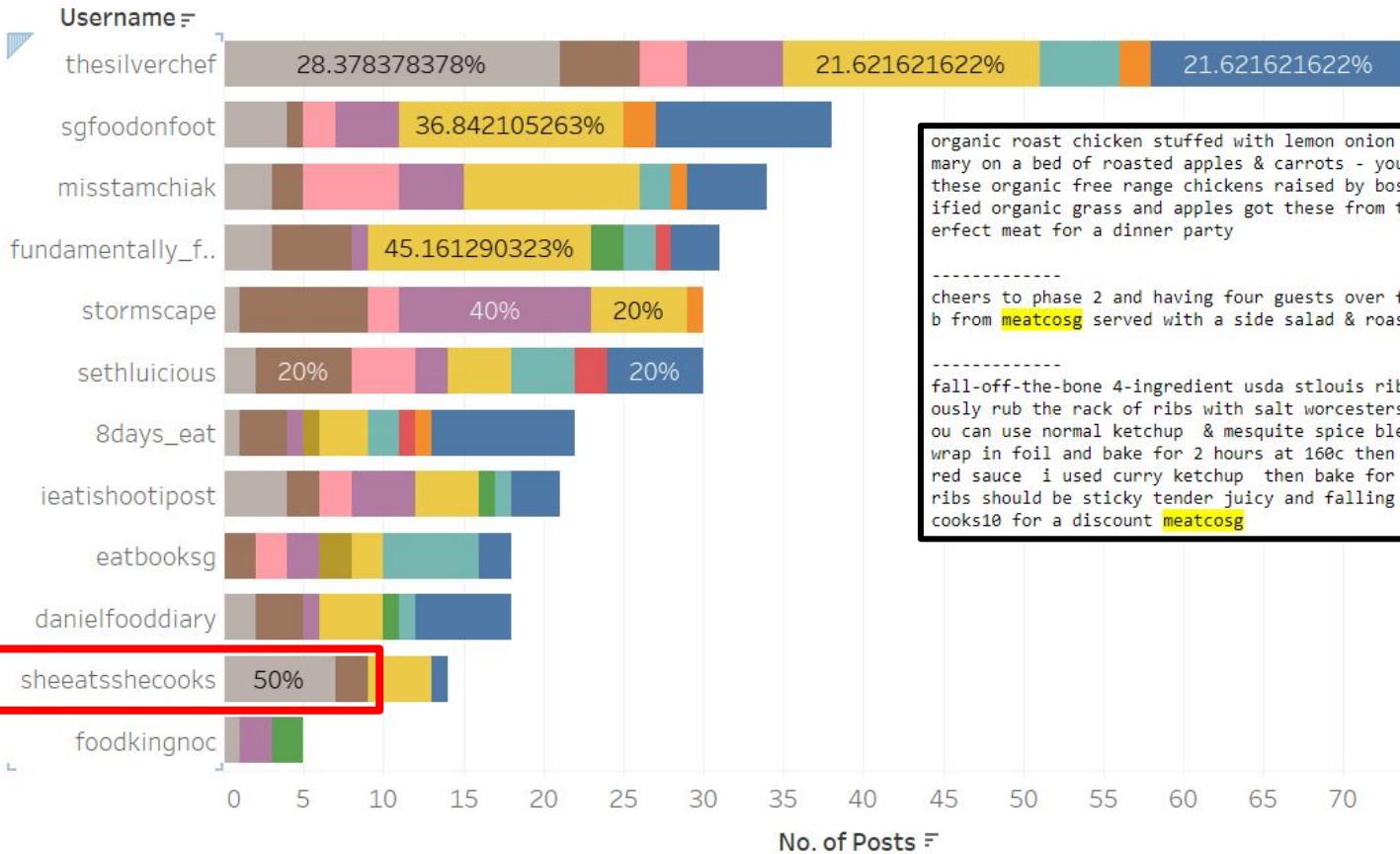
Social Network Analysis



Gephi



Influencers' Cuisine and No. of Posts



organic roast chicken stuffed with lemon onion bacon juniper berries & rosemary on a bed of roasted apples & carrots - you can taste the difference in these organic free range chickens raised by bostockbrothers grazing on certified organic grass and apples got these from the guys meatcog & are the perfect meat for a dinner party

cheers to phase 2 and having four guests over for dinner with this prime rib from meatcog served with a side salad & roasted sweet potatoes

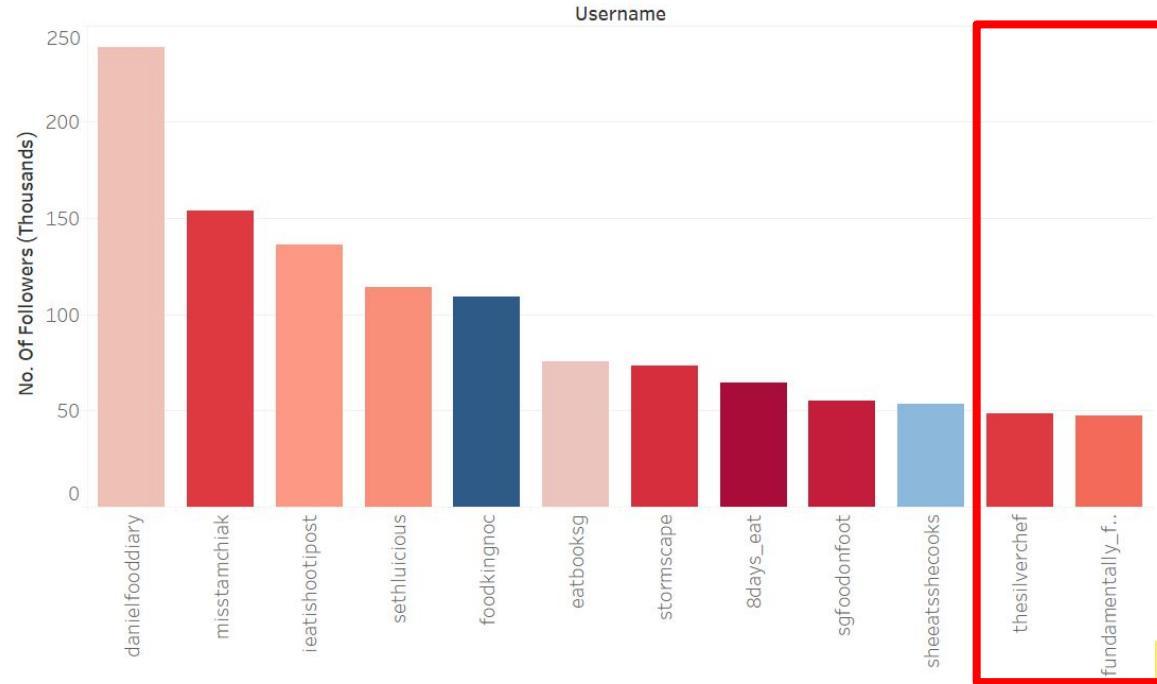
fall-off-the-bone 4-ingredient usda stlouis ribs from meatcog simply generously rub the rack of ribs with salt worcestershire sauce curry ketchup you can use normal ketchup & mesquite spice blend from helaspicesingapore wrap in foil and bake for 2 hours at 160c then unwrap brush a layer of desried sauce i used curry ketchup then bake for another 1 hour uncovered the ribs should be sticky tender juicy and falling off the bones use sheatsshe cooks10 for a discount meatcog

ooks

thesilverchef

fundamentally_flawed

Number of followers per influencer



sgfoodonfoot

stormscape

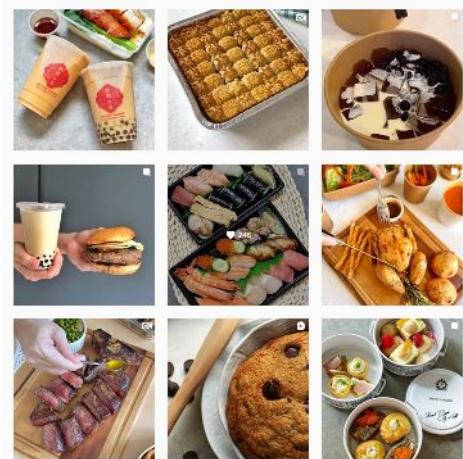
8days_eat



sgfoodonfoot

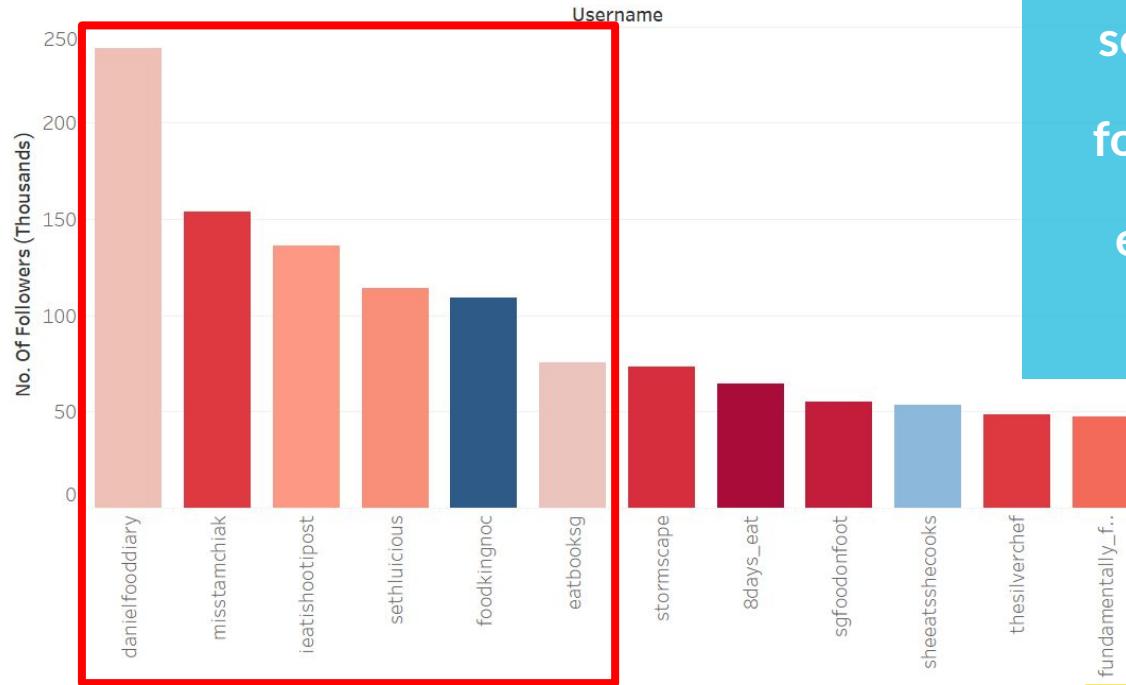


stormscape



8days_eat

Number of followers per influencer



danielfooddiary

misstamchiak

ieatishootipost

sethluicious

foodkingnoc

eatbooksg

Limitation

Inaccuracy of sentiment scores due to misspelled words

```
@viggg @bottombracket @riardsg @pamelatlc oh nooooooo!!! Captain Lard!!  
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

```
@in_vincentble luckily we went there last week! @albertee01 there goes yr f  
av CKT!  
{'neg': 0.0, 'neu': 0.615, 'pos': 0.385, 'compound': 0.7835}
```

ID	comment	neg	neu	pos	compound
1	Awesomeeeeee neeed this now	0.0	1.0	0.0	0.0
1	@becauseitsdone hei huili lets go heytea	0.0	1.0	0.0	0.0

Possible Extension

- Food/not food classifier from transfer learning on **en_core_web_sm**

Type	Description
PERSON	People, including fictional.
NORP	Nationalities or religious or political groups.
FAC	Buildings, airports, highways, bridges, etc.
ORG	Companies, agencies, institutions, etc.
GPE	Countries, cities, states.
LOC	Non-GPE locations, mountain ranges, bodies of water.
PRODUCT	Objects, vehicles, foods, etc. (Not services.)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART	Titles of books, songs, etc.
LAW	Named documents made into laws.
LANGUAGE	Any named language.
DATE	Absolute or relative dates or periods.
TIME	Times smaller than a day.
PERCENT	Percentage, including "%".
MONEY	Monetary values, including unit.
QUANTITY	Measurements, as of weight or distance.
ORDINAL	"first", "second", etc.
CARDINAL	Numerals that do not fall under another type.

<http://gg.gg/IS434FoodieBuddie>

Thank You

Especially to Prof Kyong and TAs
for your guidance

