What makes a successful milk tea shop

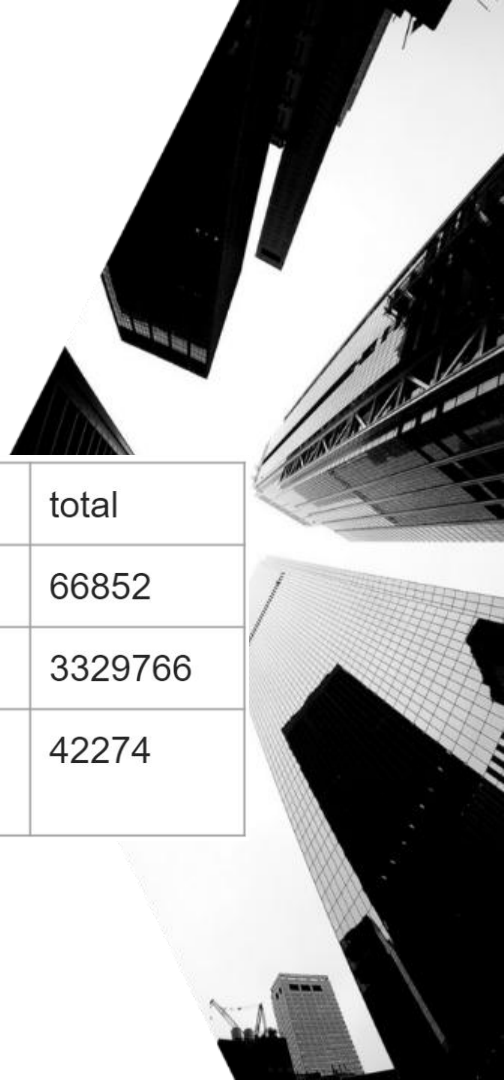
Group 11: Zhengyuan Wen, Qilu Zhou, Jiaying Jia

Data Introduction

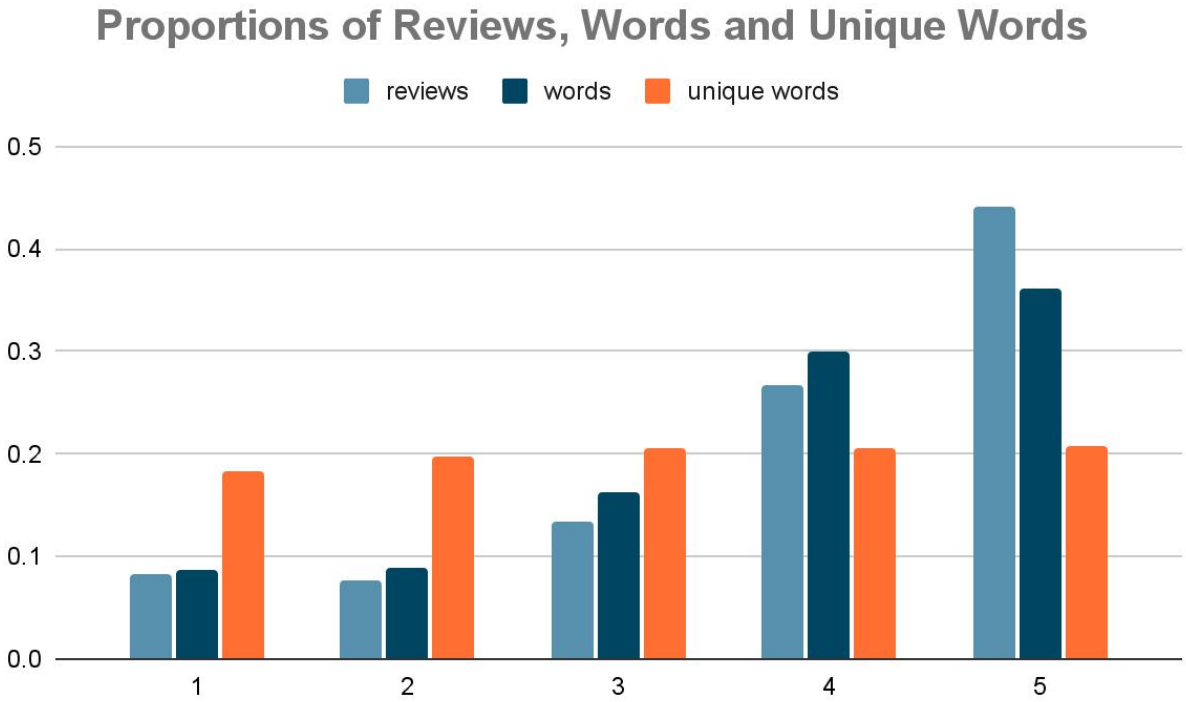
Select data with keyword “bubble tea”

-> 66,852 reviews from 779 shops.

	1	2	3	4	5	total
reviews	5491	5093	8918	17885	29465	66852
words	287153	293087	543443	1001336	1204747	3329766
unique words	7724	8337	8659	8738	8816	42274



Data Introduction



Categories

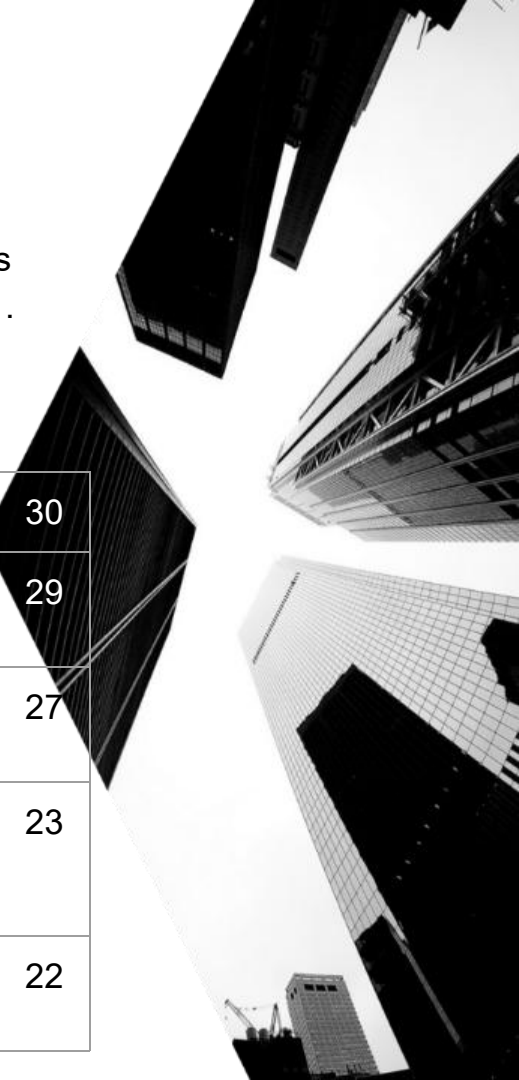
- We want to explore which categories of shops sell bubble tea

Food	397	Desserts	115	Tea Rooms	48	Sandwiches	30
Bubble Tea	378	Cafes	64	Taiwanese	43	FoodBubble Tea	29
Restaurant	242	Chinese	59	Vietnamese	38	Bubble TeaFood	27
Coffee & Tea	186	Asian Fusion	58	Bubble TeaBubble Tea	34	Japanese	23
Juice Bars & Smoothies	141	Ice Cream & Frozen Yogurt	51	FoodFood	32	Noodle	22

Categories

- Bubble tea shops often also offer food, and some are restaurants
- They sell many kinds of drinks, e.g. coffee&tea, juice&smoothies.
- They also sell desserts, for example ice cream and yogurt
- Asian shops sell bubble tea the most

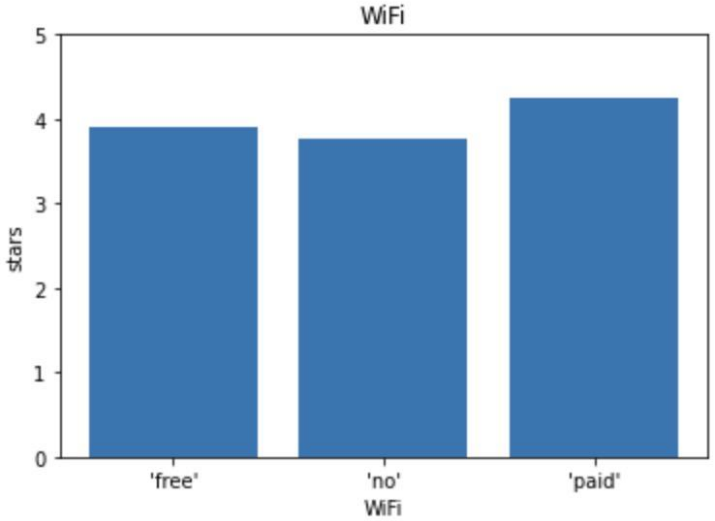
Food	397	Desserts	115	Tea Rooms	48	Sandwiches	30
Bubble Tea	378	Cafes	64	Taiwanese	43	FoodBubble Tea	29
Restaurant	242	Chinese	59	Vietnamese	38	Bubble TeaFood	27
Coffee & Tea	186	Asian Fusion	58	Bubble TeaBubble Tea	34	Japanese	23
Juice Bars & Smoothies	141	Ice Cream & Frozen Yogurt	51	FoodFood	32	Noodle	22



Attributes

RestaurantsTakeOut	589	DogsAllowed	319
BusinessParking	560	GoodForKids	299
RestaurantsDelivery	538	NoiseLevel	294
WiFi	501	Alcohol	262
Caters	480	HasTV	257
BikeParking	471	RestaurantsReservations	255
RestaurantsPriceRange2	461	Ambience	243
BusinessAcceptsCreditCards	424	RestaurantsGoodForGroups	223
OutdoorSeating	411	GoodForMeal	218
WheelchairAccessible	368	RestaurantsAttire	195

WIFI

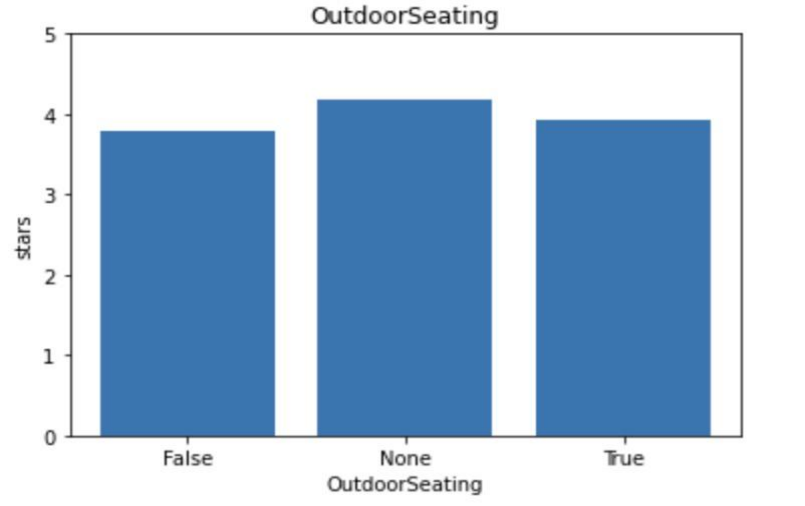


WiFi	stars
'free'	3.911964
'no'	3.757576
'paid'	4.250000

WiFi	
'free'	443
'no'	99
'paid'	2

- Paid WIFI has the highest stars, but only 2 shops offer paid WIFI
- Free WIFI has higher stars than no WIFI

Outdoor Seating

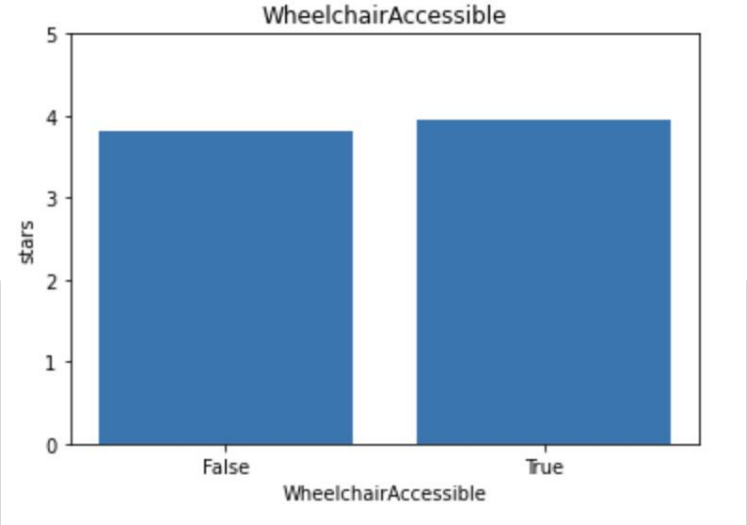


OutdoorSeating	stars
False	3.78053
None	4.17346
True	3.92198

OutdoorSeating	
False	262
None	49
True	141

- Customers prefer shops that offer outdoor seating
- Now most of the shops don’t offer that

Wheelchair Accessible

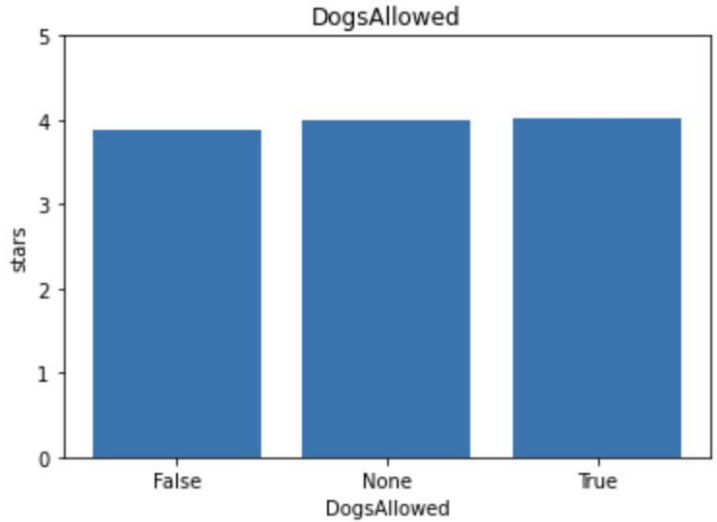


WheelchairAccessible	stars
False	3.812500
True	3.940233

WheelchairAccessible	
False	56
True	343

- Shops that allow wheelchair to come in offer more convenience and have relatively higher stars

Dogs Allow

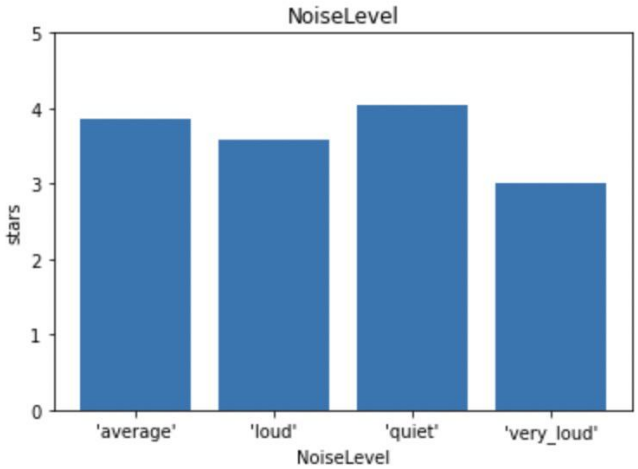


DogsAllowed	stars
False	3.888889
None	4.000000
True	4.022989

DogsAllowed	
False	252
None	1
True	87

- Although most shops don’t allow dogs to come in, dogsallowed shops do have a better stars performance

Noise Level



NoiseLevel	stars
'average'	3.860360
'loud'	3.583333
'quiet'	4.033333
'very_loud'	3.000000

NoiseLevel	
'average'	222
'loud'	12
'quiet'	75
'very_loud'	2

- Quiet shops have the highest stars
- Most of the shops are average quiet

Reviews Tokenization

- Tokenizer: Subword Tokenization based on the review texts.

Original Sentences: It was just I had a really hard time communicating with the cashier because of a language barrier (which I am not faulting her for).

After Tokenized:

communicating -> ‘com’, ‘##m’, ‘##un’, ‘##icating’

faulting -> ‘fault’, ‘##ing’

TF-IDF

Term Frequency:

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

Inverse Document Frequency:

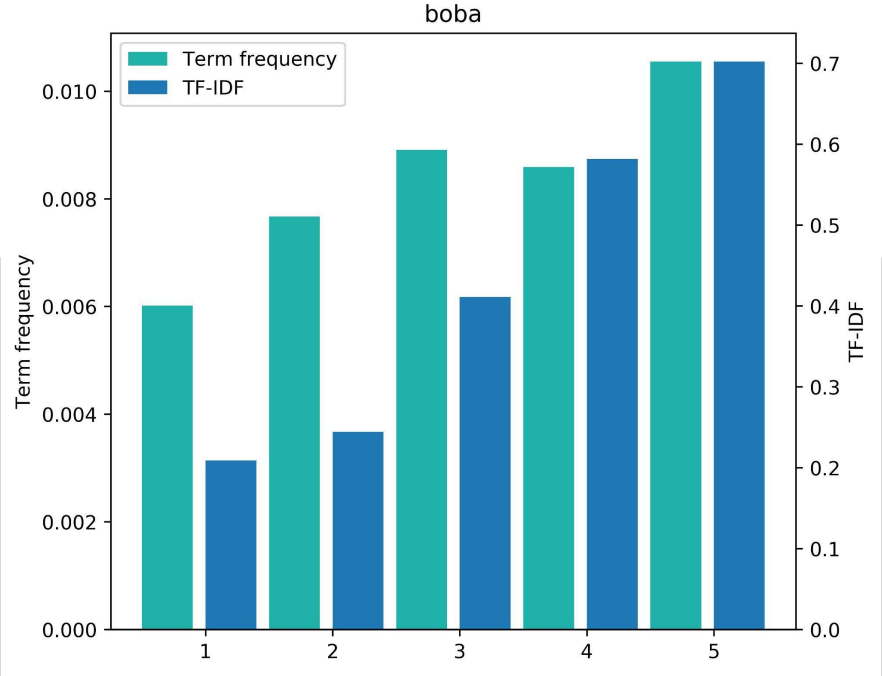
$$IDF = \log \frac{N}{1 + |\{d \in D : t \in d\}|}$$

$$TF\text{-}IDF = TF \times IDF$$

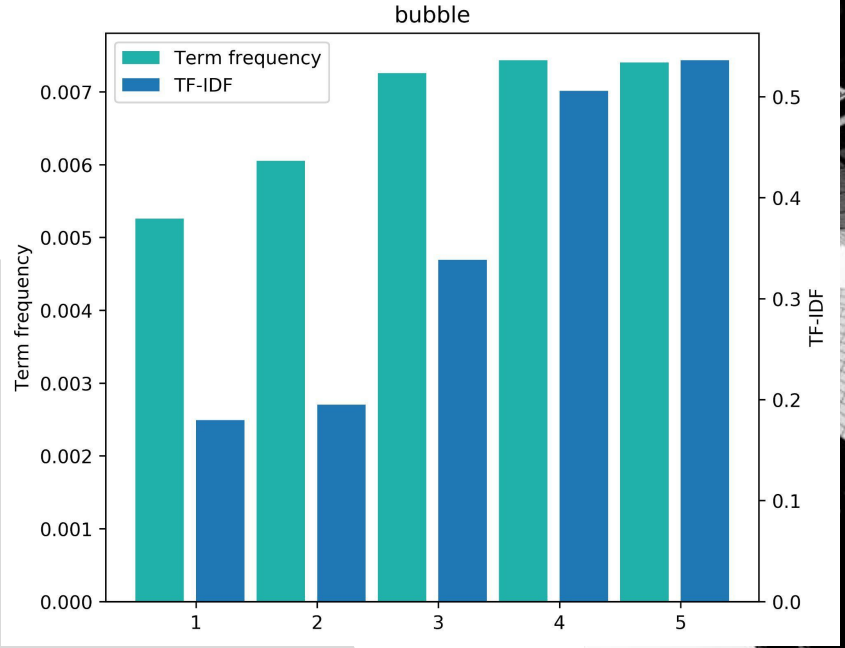
Wordcloud Based on word frequency

Ingredients

Boba

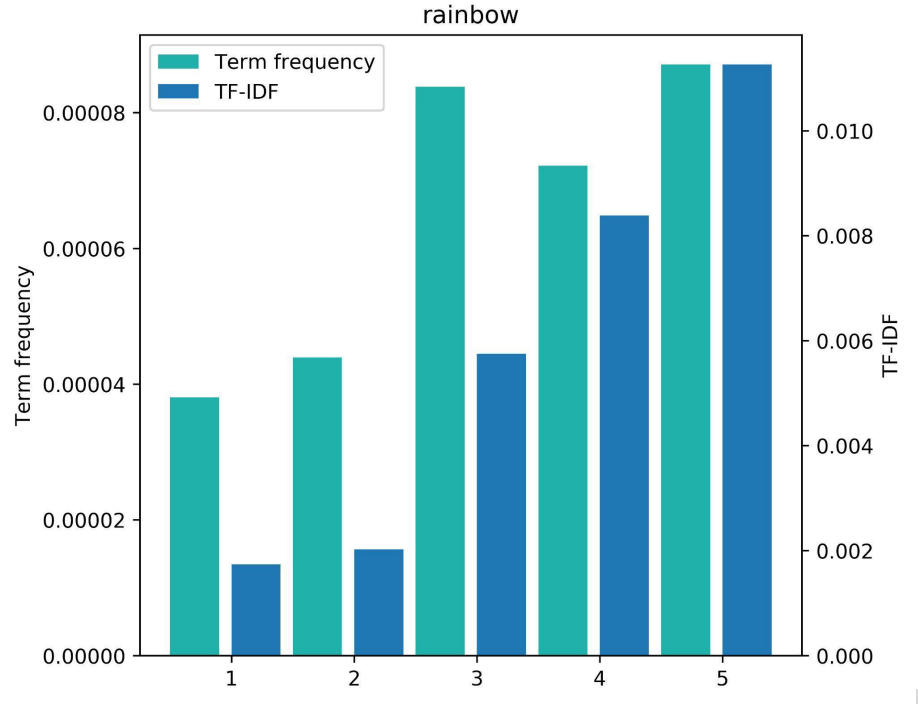


Bubble

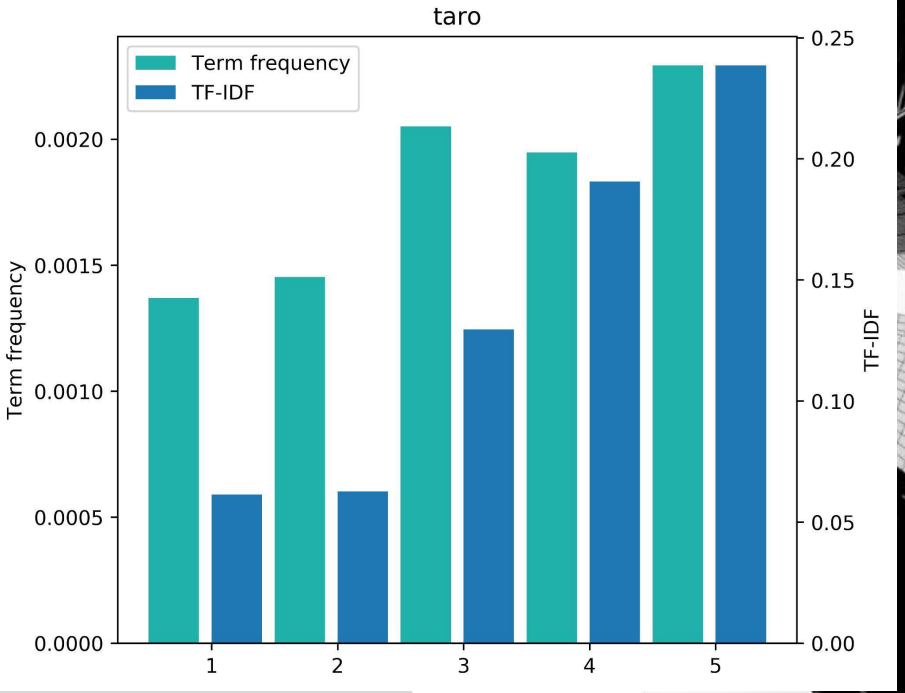


Ingredients

Rainbow

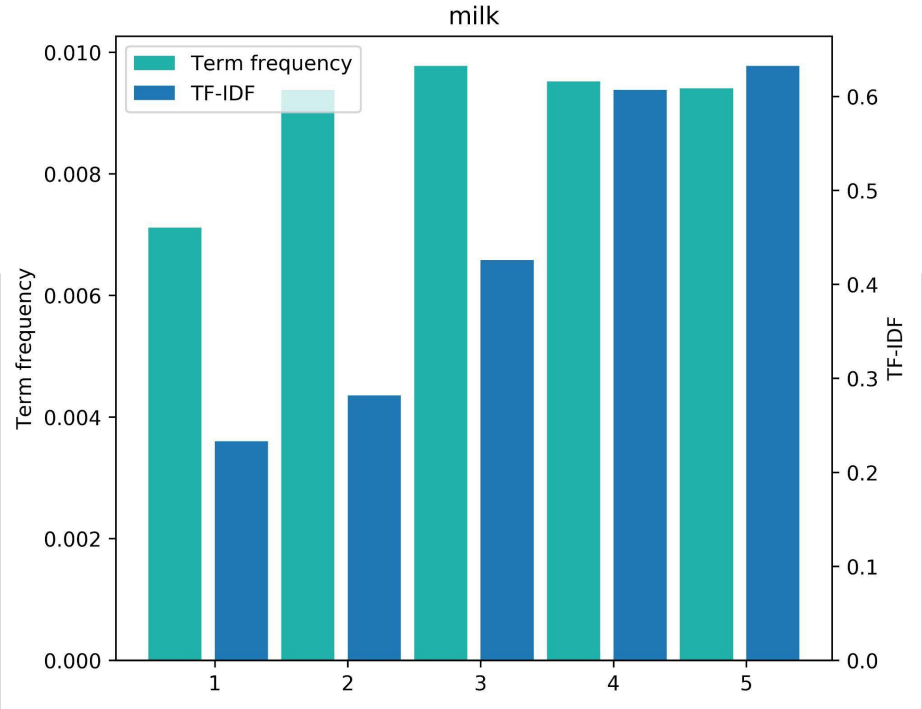


Taro

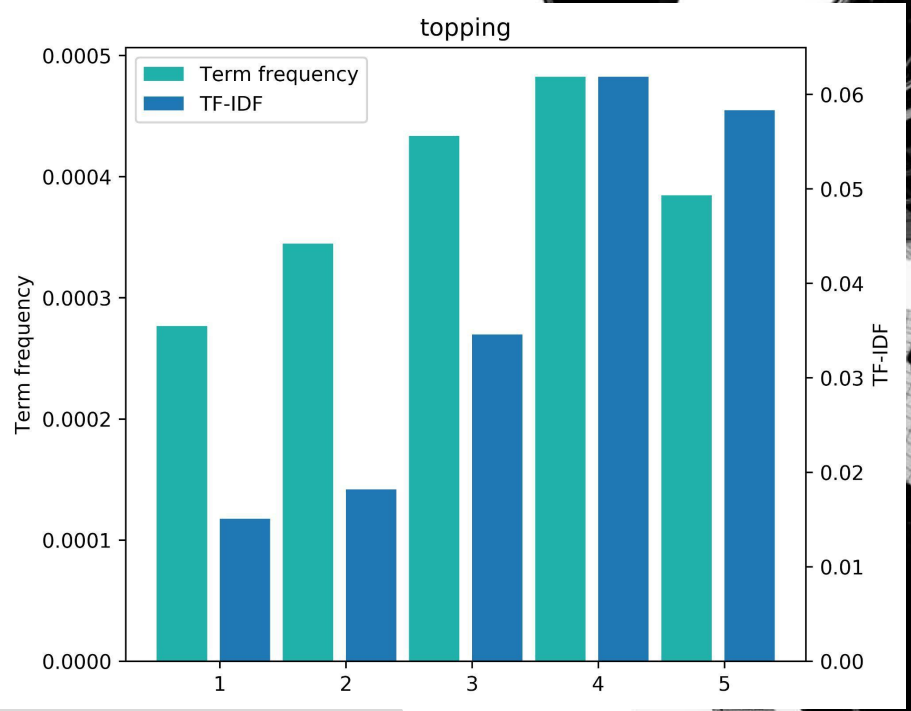


Ingredients

Milk

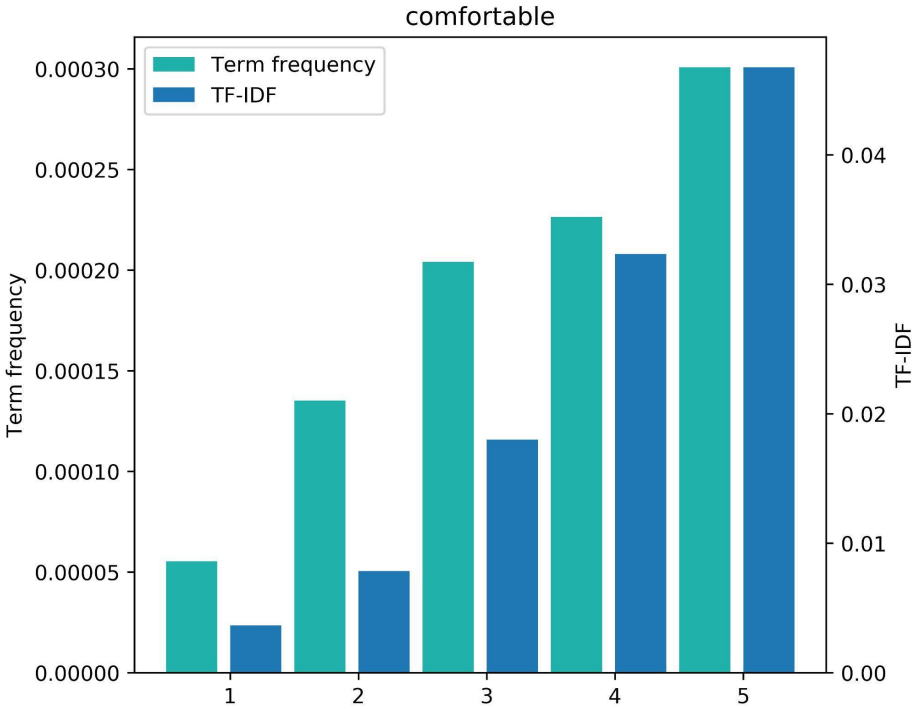


Topping

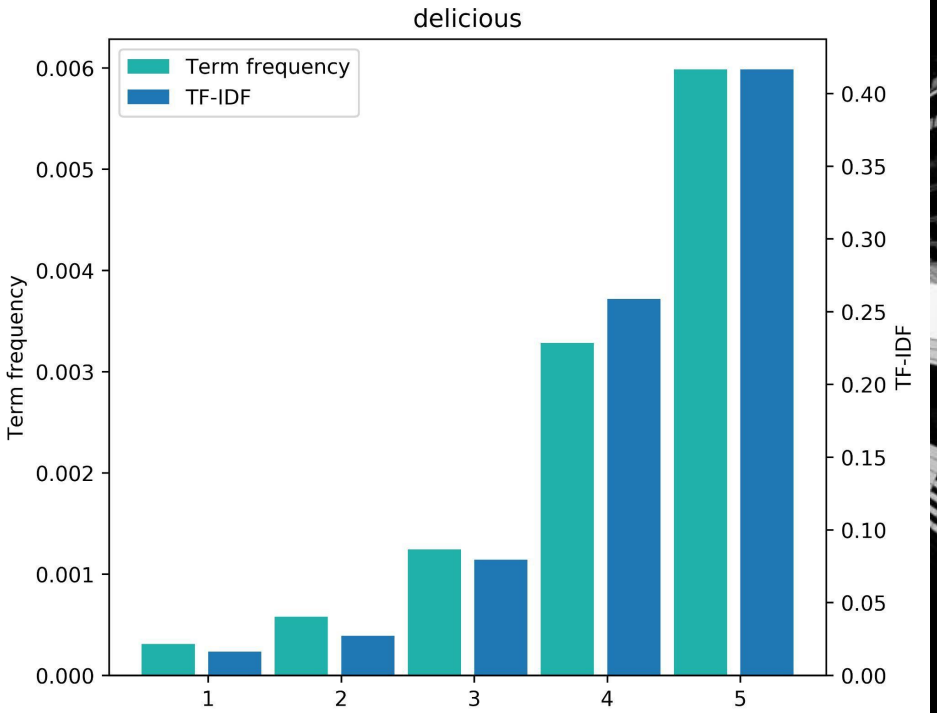


Comments

Comfortable



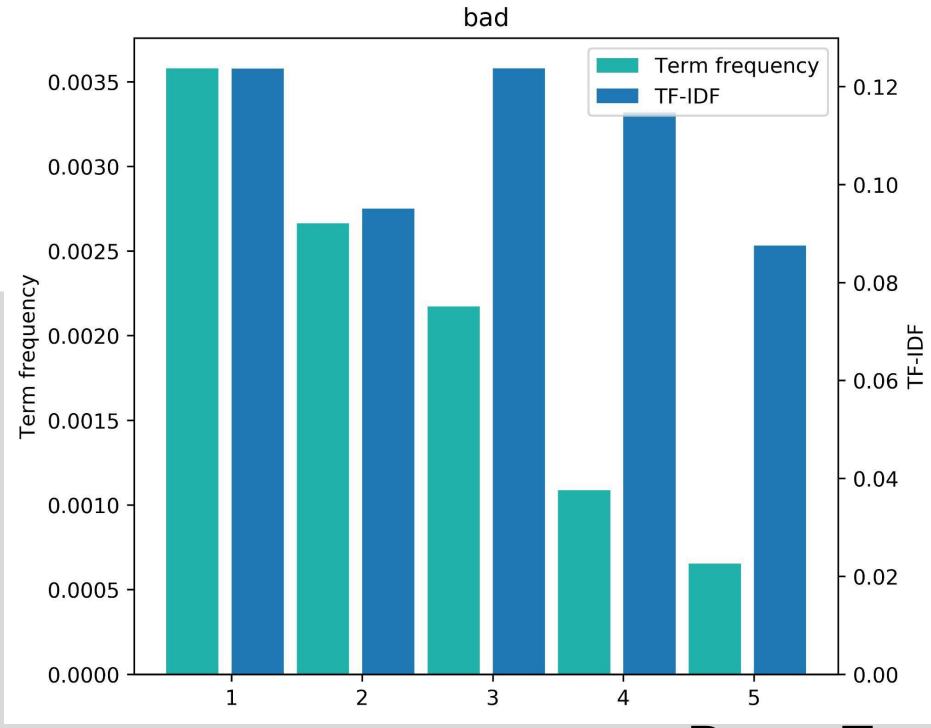
Delicious



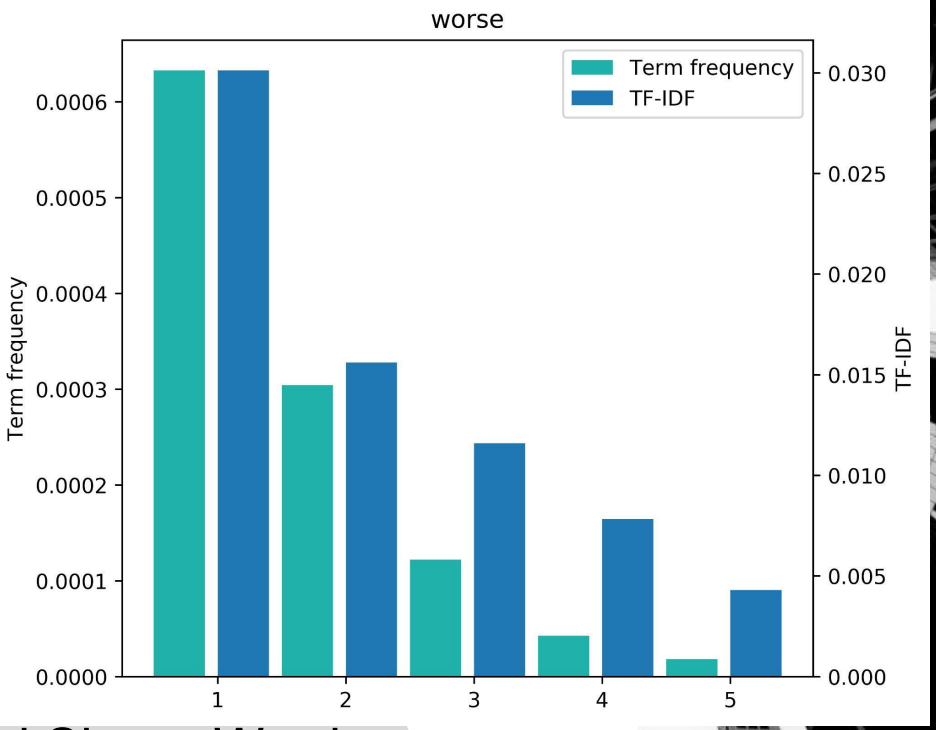
Up-Trend Shape Words

Comments

Bad



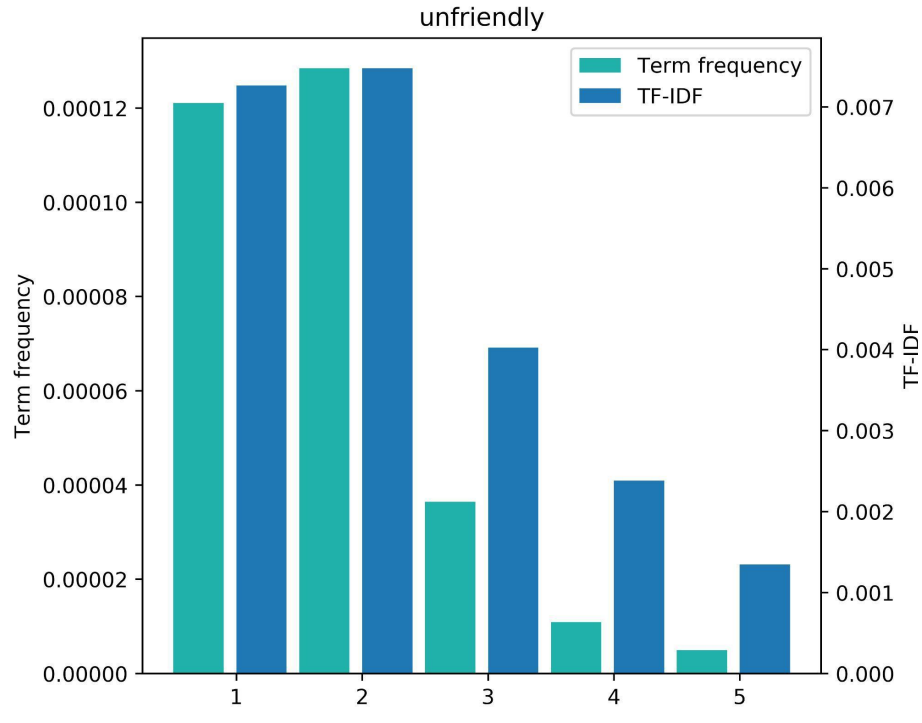
Worse



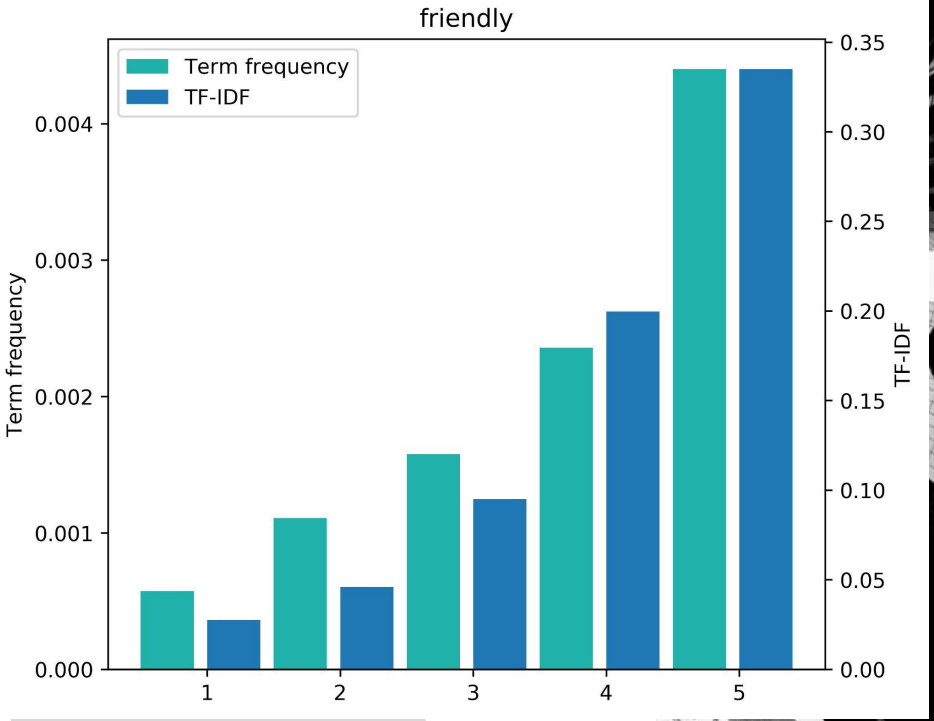
Down-Trend Shape Words

Comments

Unfriendly

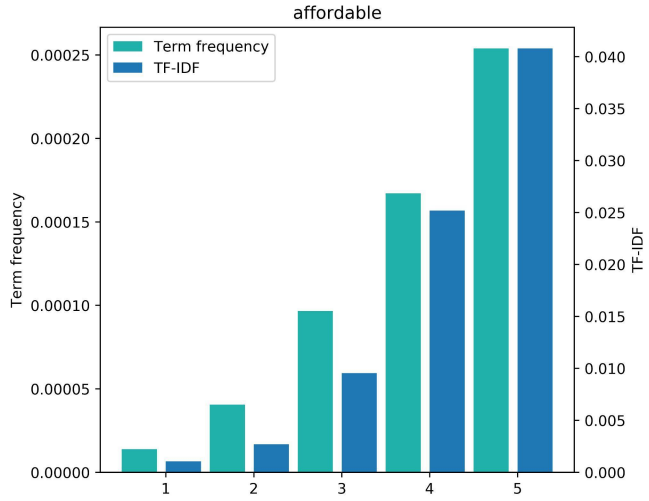


Friendly

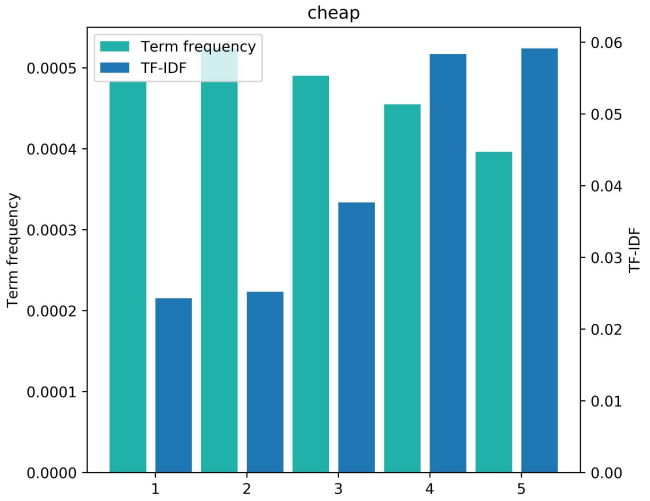


Price

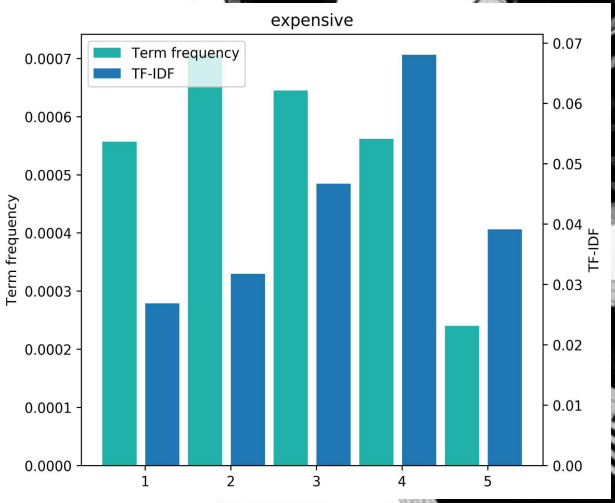
Affordable



Cheap



Expensive

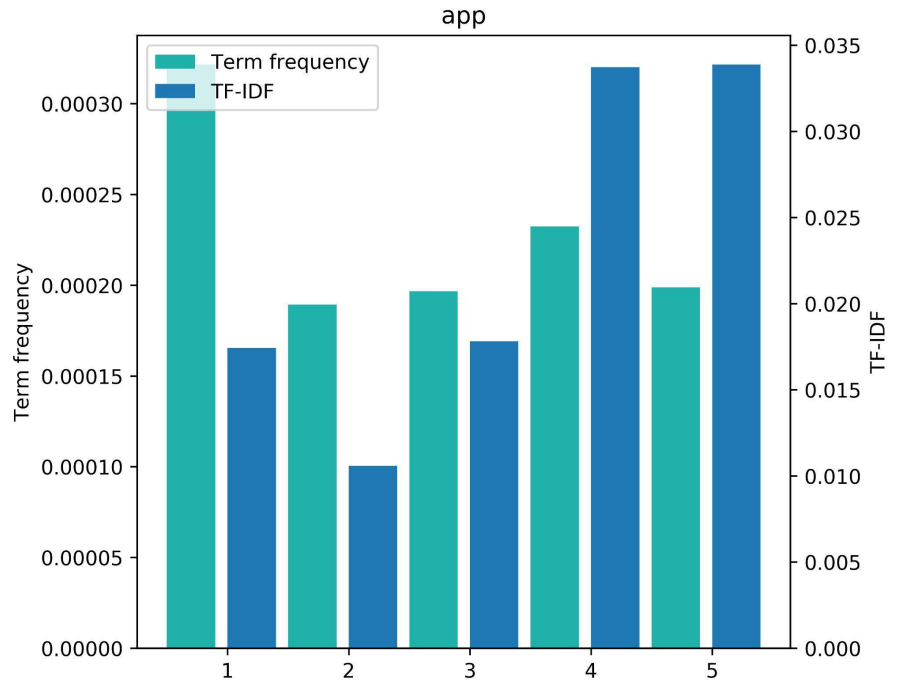


$\Uparrow$

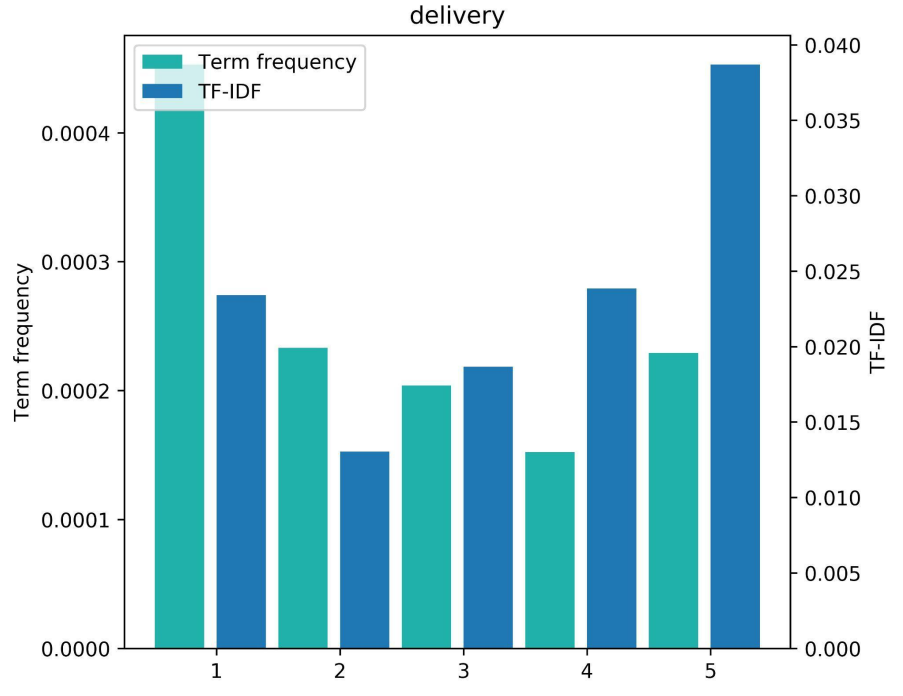
X - Shape Word

“U-Shape” words

App

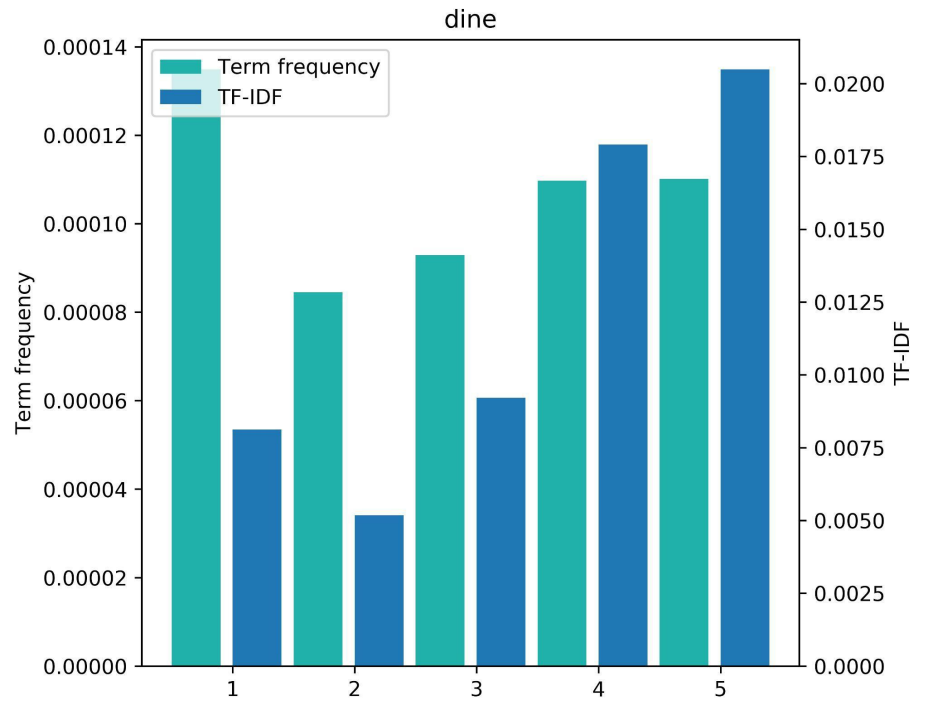


Delivery

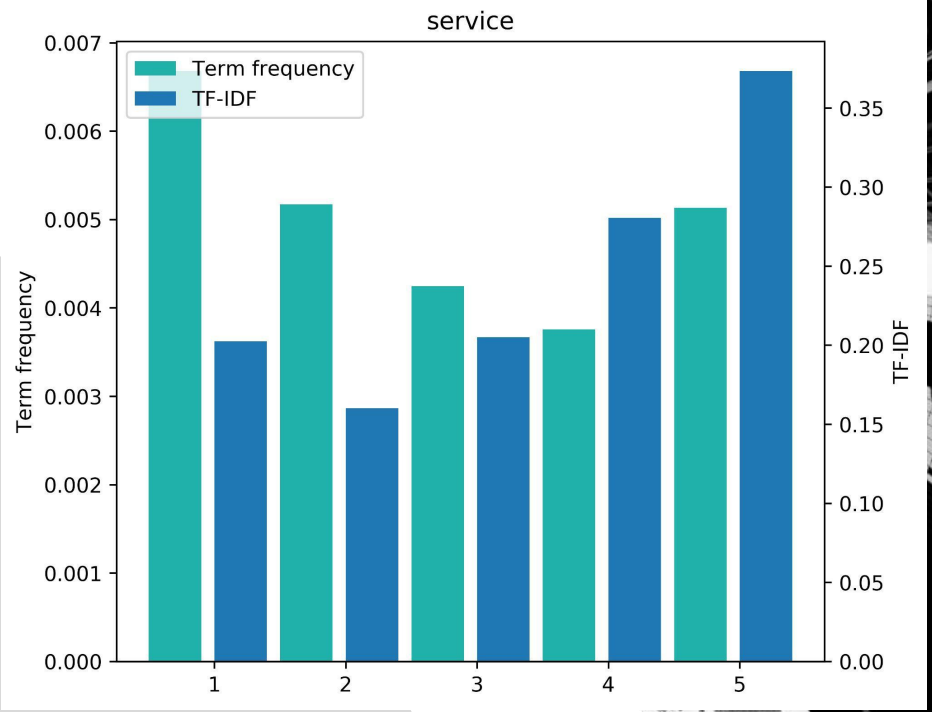


“U-Shape” words

Dine



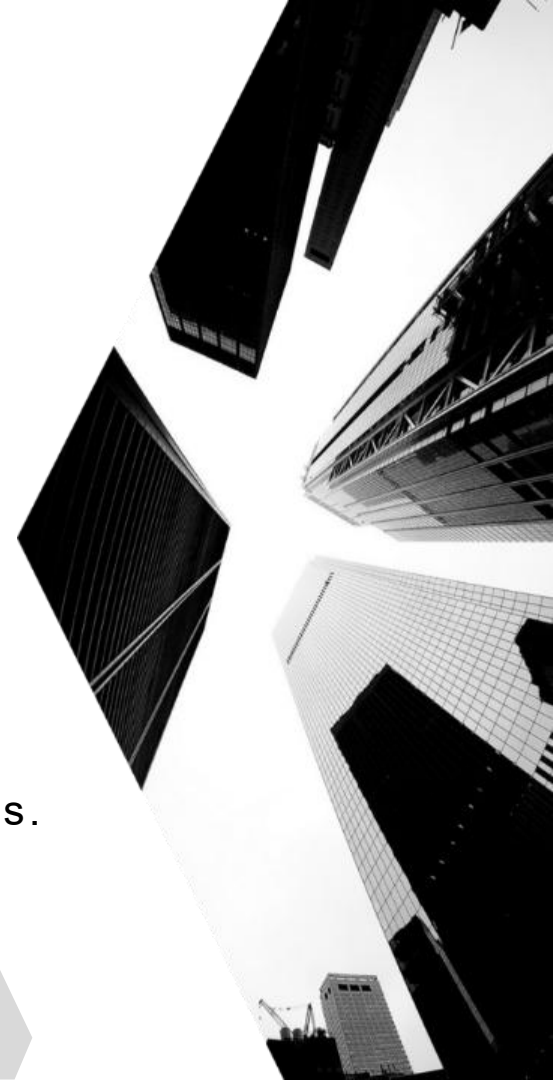
Service



Preliminary Conclusions

Based on four types of word:

- **Up-Trend Shape Words**: Good things, commonly used in good rating reviews.
- **Down-Trend Shape Words**: Bad things, commonly used in bad rating reviews.
- **“X - Shape” Words**: Often seen in the common words like: Cheap, Time, Food. They have different meanings under different situations.
- **“U - Shape” Words**: Controversial things or items. People have strongly different opinions on these things or items.



Some other Commonly-seen Words

- **Up-Trend Shape Words:** Coconut, Coffee, Dessert, Environment, Fresh, Jelly, Location, Music, Quiet, Snack, Sweet.

- **Down-Trend Shape Words:** Money, Un##

- **“X - Shape” Words:** ##able, ##ly, Cold, Food, Wait, Time.

- **“U - Shape” Words:** Ever, Never, Togo.

Future works

- Construct an overall level model combining reviews with business attributes.
- Improve the results of tokenization of subword tokenizer.
- Dig out the relationship between the words, not only just looking at word frequency.