Team 3: YaoChun Hsieh, Nathan Hsu, Ying Fang, Yang Tao, Yuchen Shen

Problem 1: Sampling Twitter Data with Streaming API about a certain topic

Our group choose the key word in problem 1 and 2 after we have decided our case study1 topic to be 'iPhone8/iPhoneX'. The key word is "SteveJobs Theater", because Apple announced iPhone8/iPhoneX at the Steve Jobs Theater, the first keynote there. This must be a very exciting news for both Apple fans and architects, who will express their opinions on Twitter. At 22:00pm, 09/19/2017, there are 508 tweets concerning 'SteveJobs Theater'.

Problem 2: Analyzing Tweets and Tweet Entities with Frequency Analysis

1. Word Count:

Word	Count
RT	326
#AppleEvent	247
SteveJobs	208
#SteveJobs	206
the	186
Theater	183
Theaterの中です！	162
https://t.co/FHODmygoBu	162
@gizmodojapan:	161
at	125
#Apple's	101
for	97
#iPhoneX	90
#iPhone8	71

&	68
@appleinsider:	67
live	64
unveiling	62
Steve	60
event	60
Join	60
Jobs	58
@AppleInsider	56
https://t.co/Omo6UQHrbp	56
#stevejobs	55
http…	55
new	38
to	37
#apple	34
from	34

2. Find the most popular tweets in your collection of tweets

The top 10 tweets with the largest number of retweet:

Count	Screen Name	Text
208	gizmodojapan	RT @gizmodojapan: SteveJobs Theaterの中です！ #AppleEvent https://t.co/FHODmygoBu
62	appleinsider	RT @appleinsider: Join @AppleInsider live at the #SteveJobs Theater for #Apple's #iPhoneX & #iPhone8 unveiling https://t.co/Omo6UQHrbp http…
51	appleinsider	RT @appleinsider: Latest #ApplePark drone video shows completed #SteveJobs Theater ahead of Sept. 12 event https://t.co/2bnK1ad8if https://…
21	yuuuuuiiiii	RT @yuuuuuiiiii: SteveJobs TheaterはSFのワンシーンのようだ。音楽もいい。でも、ここには何にもないの。 #AppleEvent https://t.co/inqftw2qcQ
20	yuuuuuiiiii	RT @yuuuuuiiiii: 人は死んだらここに行くのかな？ってくらいきれい。直径50メートルのSteveJobs Theater 。360度ガラスだから夕焼けに染まって一面ピンク色になるんだって。 #AppleEvent https://t.co/sqWMftEptG
13	appleinsider	RT @appleinsider: .@CityofCupertino approved# SteveJobs Theater occupancy after #iPhoneX event invites sent https://t.co/LQNZbDx5pz https://
7	cpassariello	RT @cpassariello: .@Apple campus opens to the public for the first time for #AppleEvent at #SteveJobs Theater. https://t.co/BgmyWulffq
4	ipadnews	RT @ipadnews: New Apple Park Drone Video of Steve Jobs Theater Ahead of Sept 12 iPhone Launch #applepark #stevejobs #applecampus https://t.…
4	IntuzHQ	RT @IntuzHQ: Apple #iPhoneX event: the 5 most important announcements from #SteveJobs Theater. #AppleEvent https://t.co/n42zqXCjti https://…
3	rajmathai	RT @rajmathai: Ready for takeoff: New #SteveJobs theater @Apple spaceship HQ. #iPhone8 https://t.co/hOhOzKnxpF

3. Find the most popular Tweet Entities in your collection of tweets

(1) The top 10 hashtags, and (2) Top 10 user mentions that are the most popular in collection of tweets.

Hashtag	Count
AppleEvent	257
SteveJobs	212
Apple	136
iPhoneX	92
iPhone8	72
stevejobs	56
apple	34
iPhone	26
iphone8	23
Theater	19

(1)

Screen Name	Count
gizmodojapan	161
appleinsider	123
yuuuuuiiiii	36
Apple	30
CityofCupertino	12
cpassariello	7
tim_cook	6
arstechnica	4
IntuzHQ	4
DHQLounge	4

(2)

Based on the top 30 words in the collection, the top7 is RT, #AppleEvent, SteveJobs, #SteveJobs, the, Theater, Theater の中です！, we know that when comparing with SteverJobs Theater, people concern more about SteveJobs and the AppleEvent. And refer to the most popular tweet, it is a short film about an overall view of SteveJobs theater from different directions, and it has been retwitted 209 times, while the 4th is a picture of the SteveJobs Theater, only be retwitted 20. We can conclude that video advertisement will attract more customers than picture ads. Besides, people are more likely to review the news of Apple through professional website like 'AppleInsider' than others. Thus, Apple can maintain a good relationship with AppleInsider. Finally, five of top ten tweets are from Japanese. We conclude that Japanese like to retweet and they care about both the SteveJobs Theater and Apple event.

Problem 3: Getting "All" friends and "All" followers of a popular user in twitter

Our target account is "PriceChopper". From its Tweeter account, we can see 7,205 following, 12.5k followers. Due to the rate limitation, we took a few hours to search all the followers. Each time we can get 200 followers, and 2800 is the cap of search times for a specific time interval, which we are not sure how long it is. Based on our implementation, it seems to be volatile. If one account request too many times, the time interval will increase. Because we have to wait for the rate limit, we write the data we get to the .txt file immediately. Therefore, we can make sure that no result is missing. We list only 20 of them.

List 20 Friend/Followers of pricechopper:

ID	Followers
111054873	my_pricechopper
898870871993049088	tdoyleny
123628824	Ronisueann
285095790	fafachicken
908391860242141185	bikerslove26
598752253	JelloVert
907727295133208576	24newsnet
12253292	LadyBelz40
2602441790	RosieTRiveter
3282986306	Mzladiiblessed
872224595566223360	kawaniiahodawa
899166876256935936	NicoleM16171314
816115727644131330	TaurusBlackCat
707664055452962816	MichaelKayePR
2625388170	RyanShiyz
2780362623	MatthewJM96
21418983	kristenaiken
2310469872	rayruke
4878352511	KarenAColvin4
784481703255416833	NYScertified

ID	Friends
3434218185	sparkshoppe
1870981	garyfarkas
584978602	DruthersBrewing
2188574036	DeCrescenteDist
3282859598	TwitterVideo
27455509	wattisretail
89349631	LaurenAliceC
1924178838	getflipp
219015278	SavingStar
632072700	Checkout51
485416525	IbottaApp
50681506	GP76TMP79
12945712	aleamy
543676654	StoveTop
16752247	OriginalSoupMan
62667032	leadavisonbikes
18058609	FDArecalls
14091384	kevinsilvestris
136719321	kay01267
2174219034	StopandShop

Print mutual friends:

ID	Mutual Friends
86985833	Tammster1993
26676275	SPECNewsAlbany
309263943	slimsterKev
43389299	Bernard321
108701420	StaffordSterner
19970901	AlbanyAquaDucks
37427435	MRMFoods
17812750	thealexion
1130943968	Steep_On_It
557175969	qa0140pc
590125338	UnionCollegeCFR
393311964	rideofhonor
396538338	DBruceLane
254634194	SaratogaRP
62616291	TheMailworks
188446278	cmgrchat
40367630	kct1963
23448443	MySweetSavings
151665448	Aweissman82
52427661	JeepCapone

Problem 4: Business question

Question 1: How is the popularity of iPhone changing through the week of Apple Event?

Reason

It is a coincidence that the Apple event was hold right after the case study was assigned. As one of the most important phone producers, Apple do have everyone's attention. We assume that people will use Twitter to express their opinion about the new iPhone on Twitter, especially during the Apple event. Therefore, we would like to analyze the trend of iPhone on Twitter. The result can be used on plan of advertising by Apple. Also, they can compare the data with the previous publication, as an indicator of the popularity change between different editions of iPhone.

Data

We used Twitter API to collect search results and count them. At the beginning the searching key word is only iphone8. The reason is that iPhoneX had very low results before Apple event, and iPhone8 seemed to perform pretty well. However, the event published two iPhones. Therefore, since Apple event we collected the results for both iPhone8 and iPhoneX. We used "iPhone8" and "iPhoneX" as the key word and get the numbers of search results. Using the free edition of Twitter API, we faced some limitation--rate limit. Every 'for loop', we can have 100 results, and we only search 100 times of loop in 15 minutes. The reason is that the maximum results for every 15 minutes is 18000. Considering the difference of popularity between iPhone and other cell phone, we decide to collect just 10,000 search results at each time. 10,000 search results is quite high, and we don't need to know whether it is "Super Popular" or "Very Popular".

Time to collect data

We collect data from 9/10 to 9/15. Every day we choose 07:30, 11:00, 12:30, 15:00, 18:00, 21:00 and 00:00 to collect data. We want to collect data when there's higher chance for people using twitter. That will be at breakfast, lunch, afternoon, and before going to bed. Also, we considered the time difference between East Coast and West Coast. Therefore, we have 11:00 trying to catch record of people from West Coast having breakfast. The time of 15:00 has two reasons: lunch time for West Coast, and the break time for people at work from East Coast. So is 0:00 trying to collect data of people from West Coast at the time before they go to bed. The real collecting time will not be exactly on time. And if we collect more than one data in an hour we will average them. However, for the difference between 11:00 and 12:30, we think it's actually doesn't matter when analyzing the data. If a company want to know the best time to put Ads, they don't need to know where the user is from. All they need to know is that whether it's a good time or bad time. As a result, we average the data of 11:00 and 12:30 and only show them as 12:00. The final times of our data is 08:00, 12:00, 15:00, 18:00, 21:00 and 0:00 referring to different period during a day.

Analyze
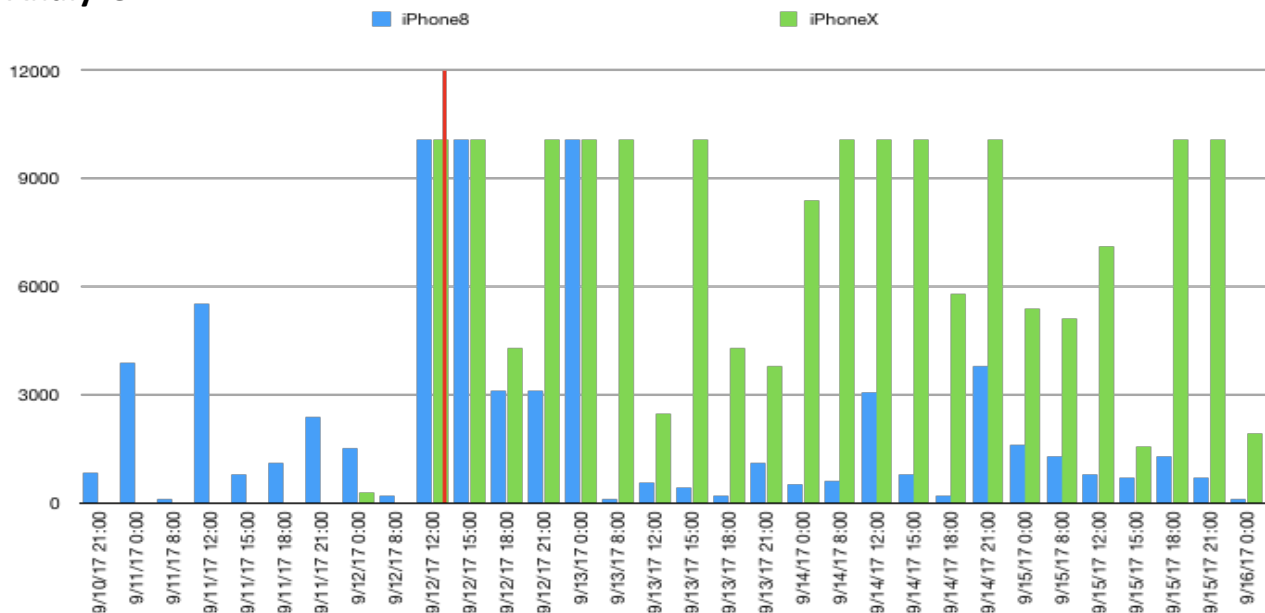
Before the Apple event we don't really search "iPhoneX" since the number of search result is really small. However, after Apple event the results of iPhoneX dominate iphone8 all the time. During a day, we see the results of iPhoneX over 10,000 at 08:00, 12:00, 15:00, 21:00 and 0:00. The results show that a hot trend can occur at any time during a day. From the table, we can see that it's more likely to gain a search result over 10,000 at 15:00 and 21:00 for iPhoneX. 15:00 is an interesting time. We suggest that it's the time for lunch at West Coast; also, a break time at East Coast. Therefore, people will intend to use Twitter to see if there's something fun or interesting.

iPhone8	~5000	5001~10000	10001~	TOTAL
8:00	5	0	0	5
12:00	3	1	1	5
15:00	4	0	1	5
18:00	5	0	0	5
21:00	6	0	0	6
0:00	5	0	1	6
TOTAL	28	1	3	32

iPhoneX	~5000	5001~10000	10001~	TOTAL
8:00	0	1	2	3
12:00	1	1	1	3
15:00	1	0	3	4
18:00	2	1	1	4
21:00	1	0	3	4
0:00	2	2	1	5
TOTAL	7	5	11	23

Beside analyzing iPhone through the week, we also want to see how it works during the event. However, during the event there are too many requests, which caused some unexpected limitation of request; So, the data is really unstable. Below is the chart we have on the day Sep 12, from 12:00 to 19:00.
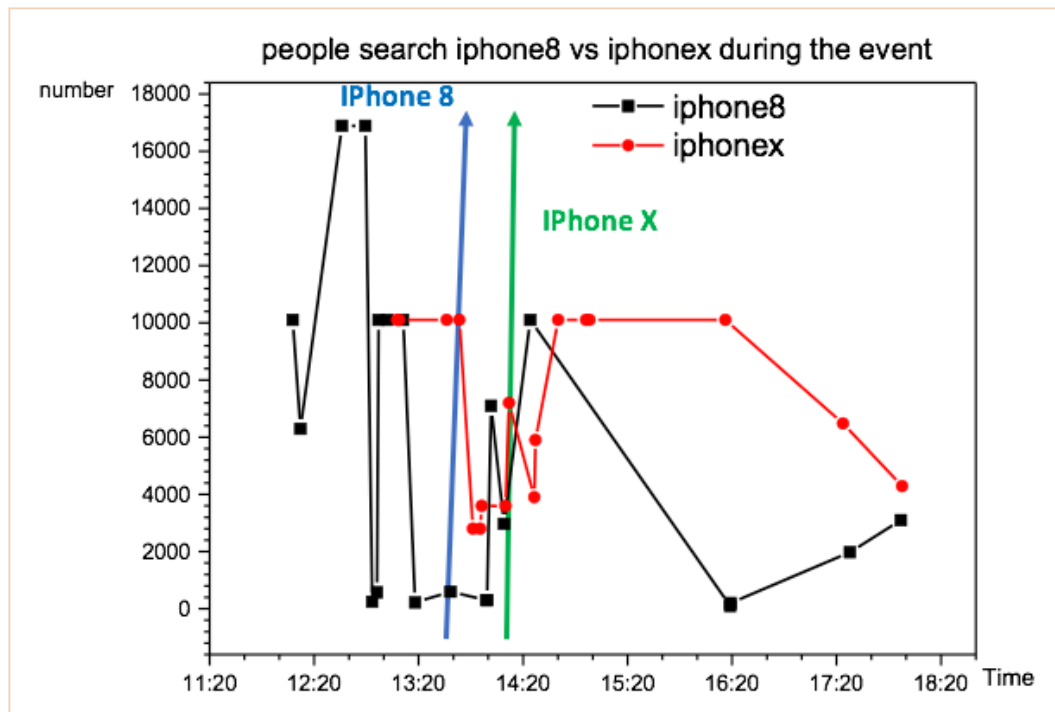
We can see that before Apple event started, iPhone8 suddenly drop drastically, and then is the first time we see a large search result of iPhoneX. We suggest that people started "knowing" iPhoneX will be more important than iPhone8, or it's more likely to use iPhoneX as the name. Thus, we started adding iPhoneX search results into our data set. One thing to notice is that we have two observations of iPhone8 before the event, which were both 16,886, are the highest number we can get under the rate limit. There's a hump between the start of Apple event and the time iPhone8 came out. We thought that people start getting online to check if there's any news of iPhone online, however then is the time for Apple Watch and Apple TV, so we can see drops of both iphone8 and iPhoneX after that.

There's some other interesting finding. The exact time of publishing iPhone8 is 13:55, we can see the searching result started climbing. The time of iPhoneX publishing was 14:19. IPhoneX raised and iPhone8 dropped. This is obvious that it takes time for the news to be spread. After the event, both of them dropped. However, iPhoneX is the one that lasted longer. Its energy came back at night, and have stayed there for quite some time.

We conclude that for a big event like this, people started surfing since morning. And it did take time to spread the heat. For both iPhone8 and iPhoneX, we don't see a suddenly rise to over

10,000 after they were published. The behavior may be explained that most people was waiting for famous people to Tweet, then read and retweet. And we also need to be aware of the drop after the event. If Apple wants to put some Ads here, it's a bad time. However, after 18:00 it will be perfect to bring the topic up again. By our observation, it can last at least 3 days.

Question 2: Finding the users who have the highest business value for Apple

In this part, we focus on searching most popular persons who have large amounts of retweets about iPhone products, such as iPhone8 and iPhoneX. By doing these, we can find famous people who concern about Apple Special Event. Apple may cooperate with them to attract their followers or collect comments.

Notice that we have collected twitter data about iPhone product from Sep.10 to Sep.15, we got lots of users with a high amount of retweet. We want to find out the users with large followers, and his tweet is retweeted a lot. To make our result more reasonable, we used two steps to retrieve data we are interested in. The first step is to list top 15 retweet numbers. We choose 15:30 on Sep.13 as an example when iPhoneX has been released for a day, and there will be enough information about it.

No.	Users	Retweets	Followers
1	JonhCena	267219	10091654
2	TopAchat	52357	396409
3	nakurubox	23732	697
4	IGN	18472	6367708
5	Cerberus_Ignite	16795	53
6	_Do_t	14361	12601
7	ceemeagain	9160	335885
8	priwpriww	7733	956948
9	biggie_teethara	7633	47
10	cowboy751	7236	192
11	its_kachi	7202	997
12	ElijahMiano	7002	2931
13	_Do_t	6910	12600
14	KTHopkins	6381	811405
15	SuperSaf	6149	94780

Users

JohnCena

IGN

priwpriww

KTHopkins

TopAchat

We notice that although some users have large amount of retweets, they have only small amount of followers. To kick out this kind of case, we then list out 5 accounts who have more followers as our second step. We believe people who have high number in both retweets and followers are the most valuable users for Apple.

The final top 5 users above is the result we got. *JohnCena*、*priwpriww* and *KTHopkins* are popular users in different area such as sports and press. IGN and TopAchat are known as an account sending technology news.

We run the procedure on the entire data set, and get a summary for the type of users:

1. Vlogger or users who often provide entertaining tweets, like *DavidDobrik* who is a vlogger with 1.25 million followers and his tweet about iPhone was retweeted about 60 thousand times, and *Ellen DeGeneres* who uploads short videos has 75 million followers and 40 thousand retweets.
2. Famous artists, athletes, etc. *BTS_twt* from Korea is a group of artists, their followers are in passion on replying and retweeting messages sent by their idols.
3. Third is users specialized in technology like TopAchat, which has 400 thousand followers, they post news and deliver professional test on new released product.

As you can see, not all users we got are professional in technology area, but still their tweets are wide-spread and have deep influence on their followers. Their comments on each product can lead to a more positive or even worse situation. Therefore, Apple should pay attention to these users and check their own evaluation. Through this way, company can make response to negative comments, or even have a better design in their future product.

Overall, users who should be focused by companies need to have three properties:

1. They should concern about Apple company and their products which means many Apple related tweets.
2. They should post useful information or comments which means lots of retweets.
3. They should have stable users reading their tweets which means lots of followers.

Question 3: The tweets trends of other phone brands around Apple Event

We also collected the tweets data of other smart phone brands from 11th Sep to 15th Sep. Apart from iPhone, we have chosen five brands to study, which are Motorola, LG, Samsung, Xperia (Sony) and Huawei. We use their name to search on Twitter, and save the results as problem 1. To make sure the fairness of data collection, we extracted tweets of each brand at the same time with iPhone every day. They are 07:30, 11:00, 12:30, 15:00, 18:00, 21:00 and 00:00. We recorded the number of tweets related with each brand, and save the contents of these tweets we extracted. After gathering and studying all these tweets, we got some interesting conclusions.

1. Tweets Trend of other brands

To better compare the tweets trends between brands, we use the average number of tweets of each brand by date. The figure below shows the trend around Apple Event. (Before 12th Sep there wasn't too many search results of iPhoneX, so we started to collect iPhoneX tweets from then.)

Trends By DAY
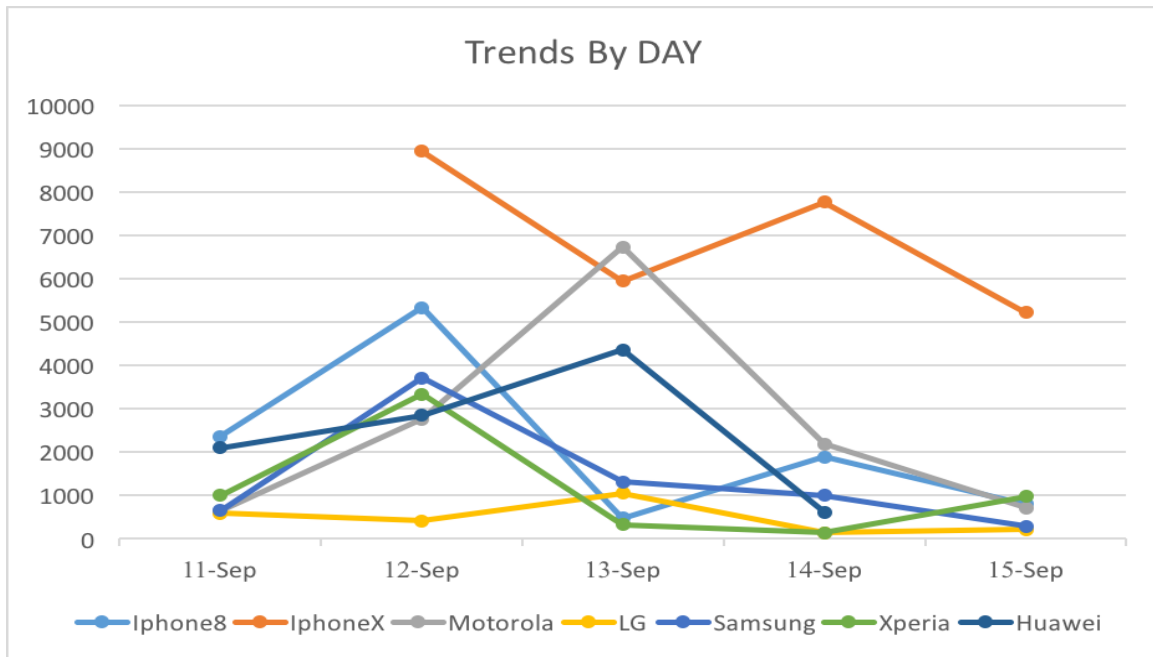
As shown in the graph above, iPhoneX is the highest among all the brands. Following iPhoneX, iPhone8, Samsung and Huawei also have good performance relatively. Tweets number of LG keeps in low through the period we studied. There is an obviously increase of all brands on 12th Sep, the Apple Event Day. The trends of Samsung and Xperia are similar, although Samsung is higher than Xperia. To some extension, it shows how popular a brand is. We can easily conclude that iPhone was the hottest topic in this period.

To find out how these trends generated, we studied some contents of the tweets we collected. We mainly focus on two questions. First, all brands tweets number have increased on 12th Sep. Does it have any relationship with Apple Event? Second, how do people compare iPhone with other smartphones in tweets? We will discuss further in the following section.

2. Content of tweets

By analyzing tweets, we found that Apple Event do have positive effect on search results of other brands. In general, we summarized it into two aspects. First, there is a great amount of tweet for advertising their online phone shop and adding "iPhone" in order to gain more exposure, like these:

Sep 11 00:57:41 @AlaaddinExpress:
#HtcU11 #Huawei #Ipad #Iphone #Oneplus #Samsung #Xiaomi(#Htc U11 [Slickcandy] [Black/Black] Heavy Duty ...)Link: https://t.co/gYblPuUta1

Sep 12 13:55:09 @chocoberu RT @syugerboy:
iPhoneX iPhone8 iPhone7s iPhone7sPlus Galaxy SC-04J Xperia [] https://t.co/MPU8EZBv5H #RakutenIchiba https://t.

Moreover, there are many tweets compared iPhone with other smartphones. For all these brands, Samsung phone is the one which are most often being compared with iPhone, then is Huawei. Please see some comments we picked up from tweets below.

Sep 12 16:27:42 @cuteplums:
I thought I was gonna change to Samsung now that the iPhone 8 will cost me my kidney but I can't deal with those ugly ass messy phones

Sep 12 16:28:24 @KikiEnso:
The new iPhone looks like the old Samsung.

Sep 12 18:26:30 @_ondary:
iPhone X is basically a Samsung with an Apple logo

Sep 12 13:34:47 @Cute_Fab1:
RT @annamends_: Design do iPhone 8 ficou mt feio, igualzinho Motorola (iPhone 8 design was ugly, just like Motorola)

Sep 11 23:06:28 @GCP1989:
New iPhones about to come out so I've gone bought a 100 Huawei off eBay. Probs have a better camera and battery life anyway

Sep 12 22:55:22 @WillowMouse:
iPhone x looks nice but that $1000 price point? No thanks. I could buy a huawei phone with the same tech for half the price

We found that many people were disappointed about new iPhones. Most of the tweets that make comparison between iPhones and other phones showed disappointment on new iPhones. Many people thought other brand phones already had new iPhones' features before, so the new iPhones are not attractive enough.