

Social Media Analysis

Yuwei Zhu, Zirui Zheng, Chenyu Shen

Date: 10/17/2019

Agenda

Overview

- Problem statement
- Background
- Dataset

Exploratory Data Analysis

Topic Modeling

Sentiment Analysis

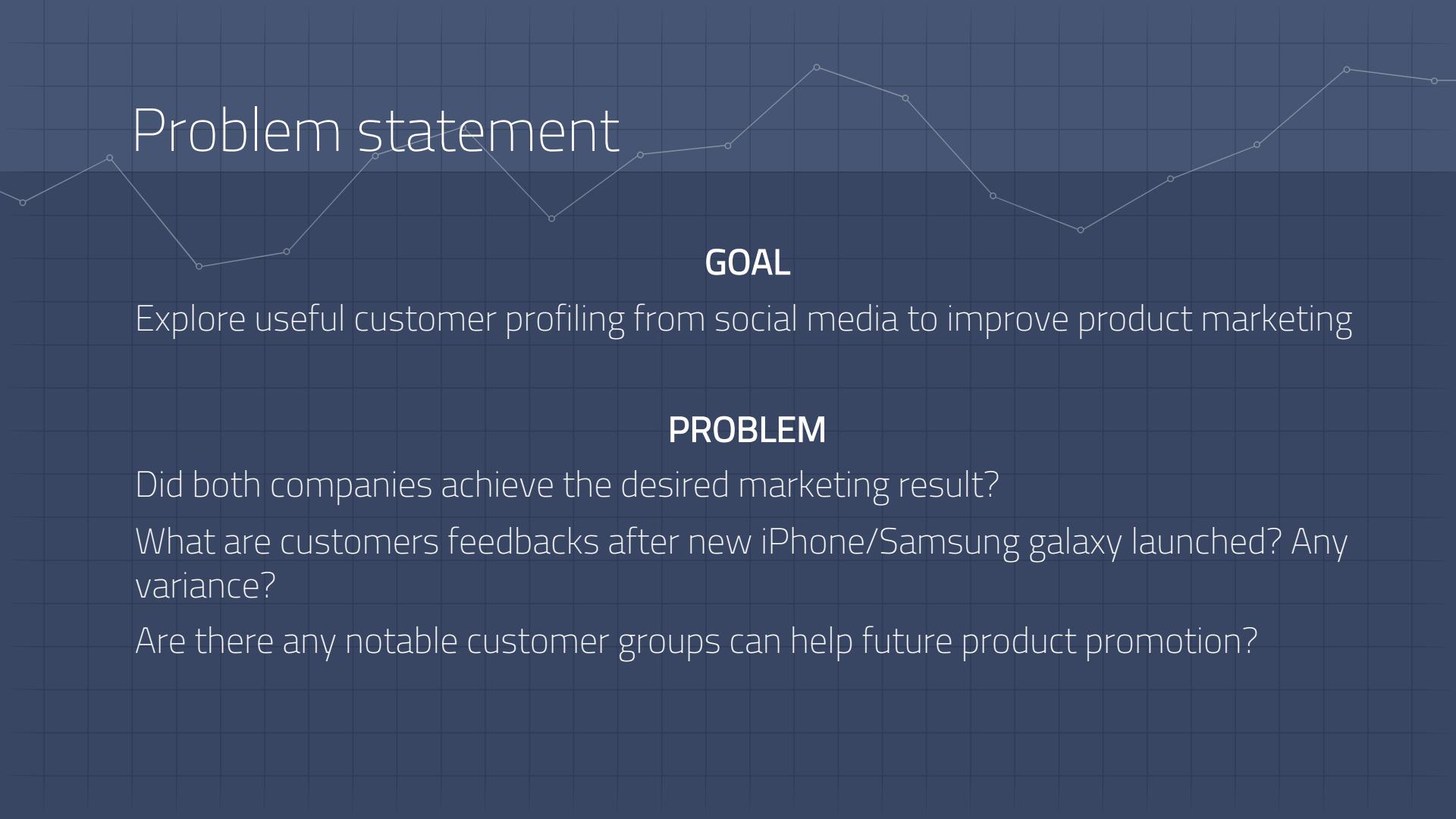
Clustering

Result & Next steps



Overview

- Problem statement
- Background
- Dataset



Problem statement

GOAL

Explore useful customer profiling from social media to improve product marketing

PROBLEM

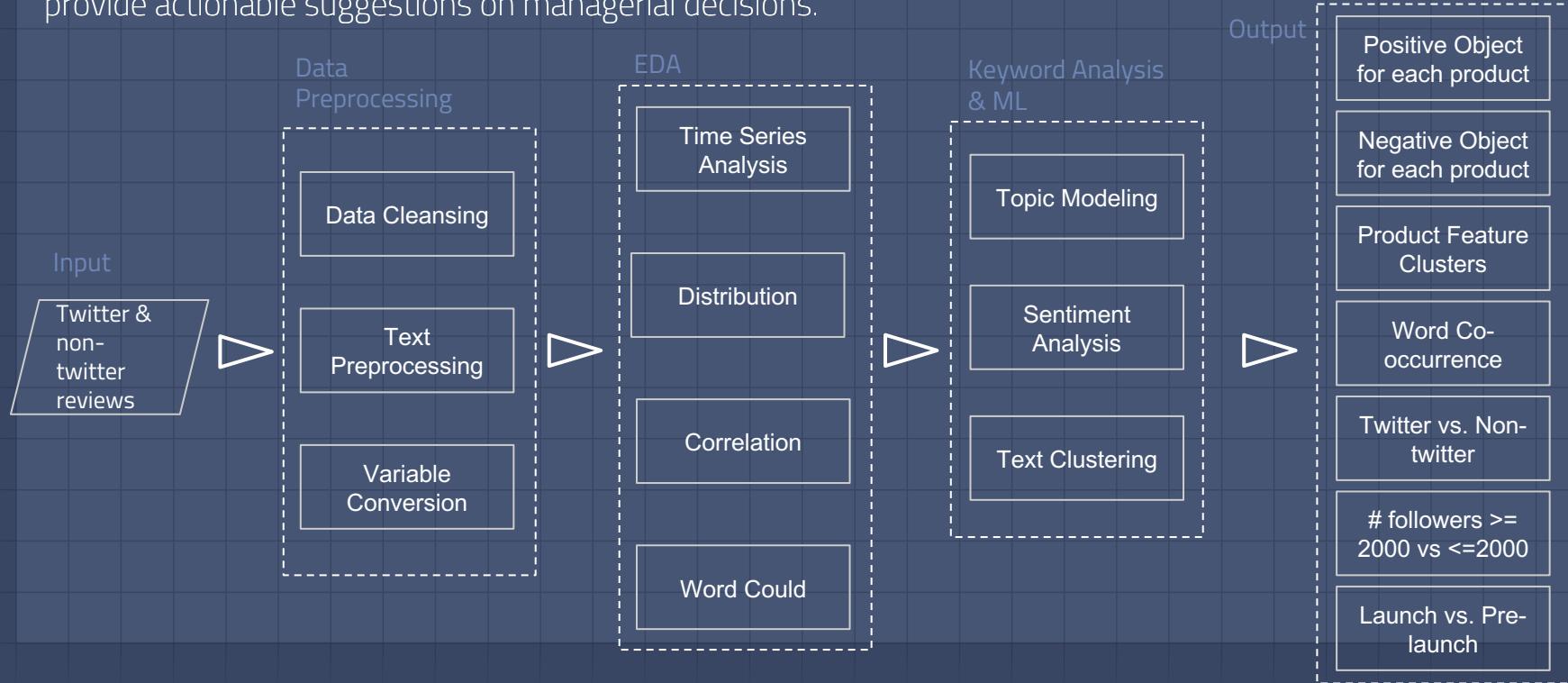
Did both companies achieve the desired marketing result?

What are customers feedbacks after new iPhone/Samsung galaxy launched? Any variance?

Are there any notable customer groups can help future product promotion?

Flowchart of the analytical pipeline

Leverage Machine Learning and NLP methods for **user profiling** for each product to improve market strategy and provide actionable suggestions on managerial decisions.



Background

Release date:

March 29, 2017

Key features:

- Extra Wide-Screen
- Infinity Display
- Bixby
- Camera
- Samsung Connect
- Dex Dock

Sales:

\$5 million within 1 month

◀ Samsung Galaxy S8

iPhone X & iPhone 8 ▶

Release date:
September 22, 2017

Key features:

- Super Retina OLED display
- Camera
- Face ID
- Home button removed
- Wireless charging

Sales:

\$46 million within 3 month

Data Preprocessing and Dataset Result



Clean Dataset

1. Drop "Sound Bite Text" NaN rows
2. Drop duplicate rows
3. Drop "Sound Bite Text" = "Post deleted by the author"



Processing Text

1. Drop special characters (e.g. punctuations, tags, url)
2. Stop words removal for specific business problems
3. Tokenization
4. Lemmanization



Labeling

1. Product type
2. After/Before product launch

Release	Product	Count
After	Galaxy	103260
	iPhone	55894
	both	23272
Before	Galaxy	20972
	iPhone	62254
	both	4892

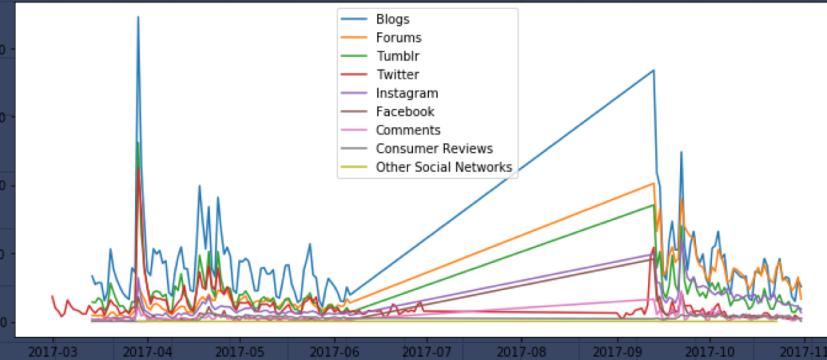
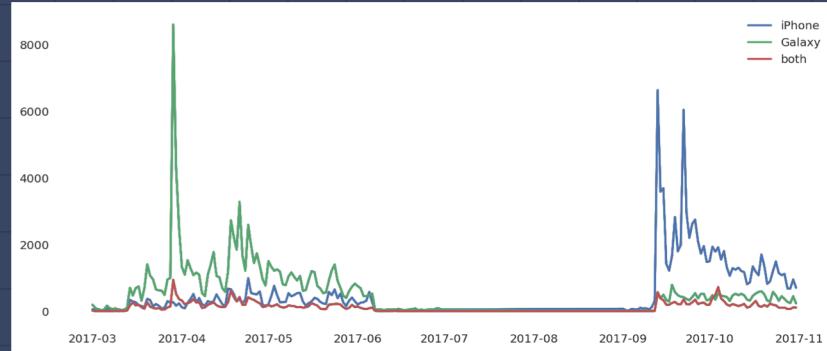


Exploratory Data Analysis

2

EDA - Time Series

2 peak posts on both products
Similar fluctuation for all sources



EDA - User Profile

192 countries

196 professions

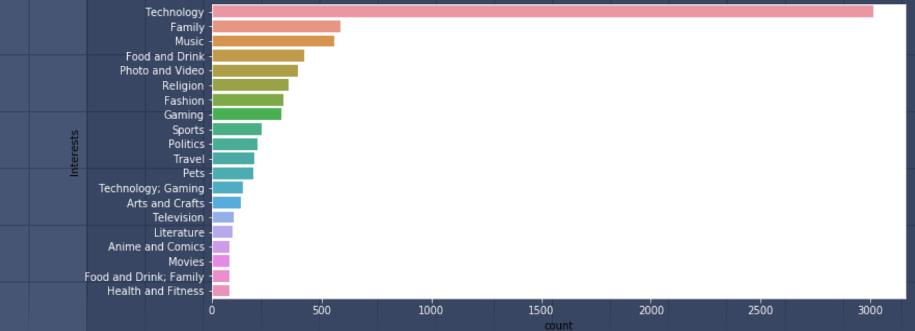
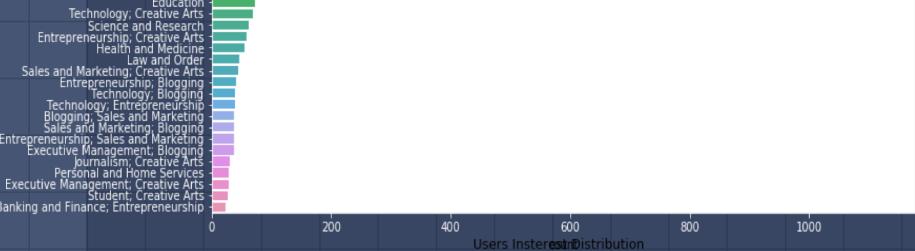
546 interests

Followers range: 11 ~

127,774,881 (127 M)

M: 47319 F: 14058

Unknown: 265851



Topic Modeling

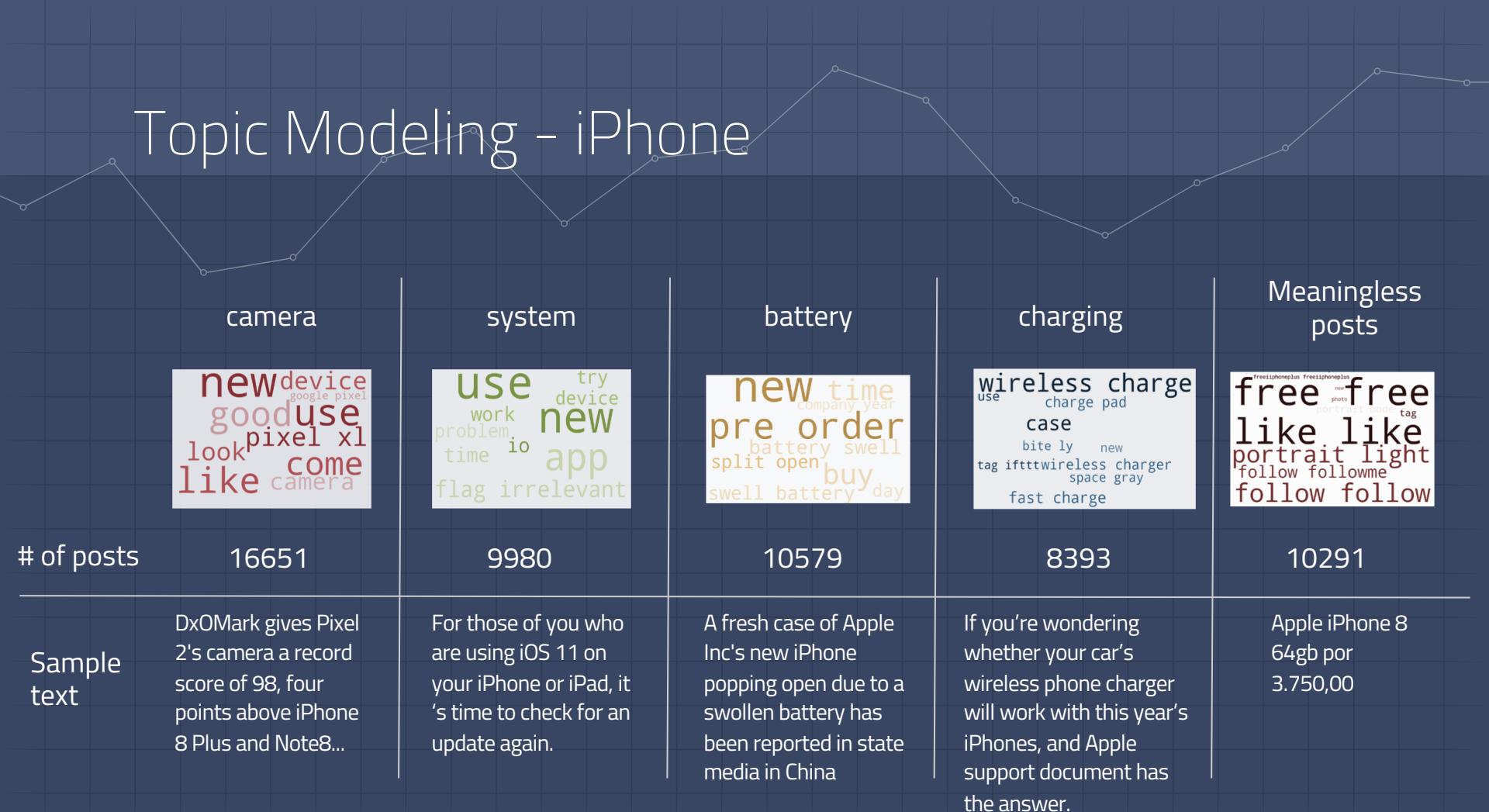
Use grid search to find the best LDA model for reviews on iPhone, Galaxy and both, respectively

Best model -- number of topics = 5

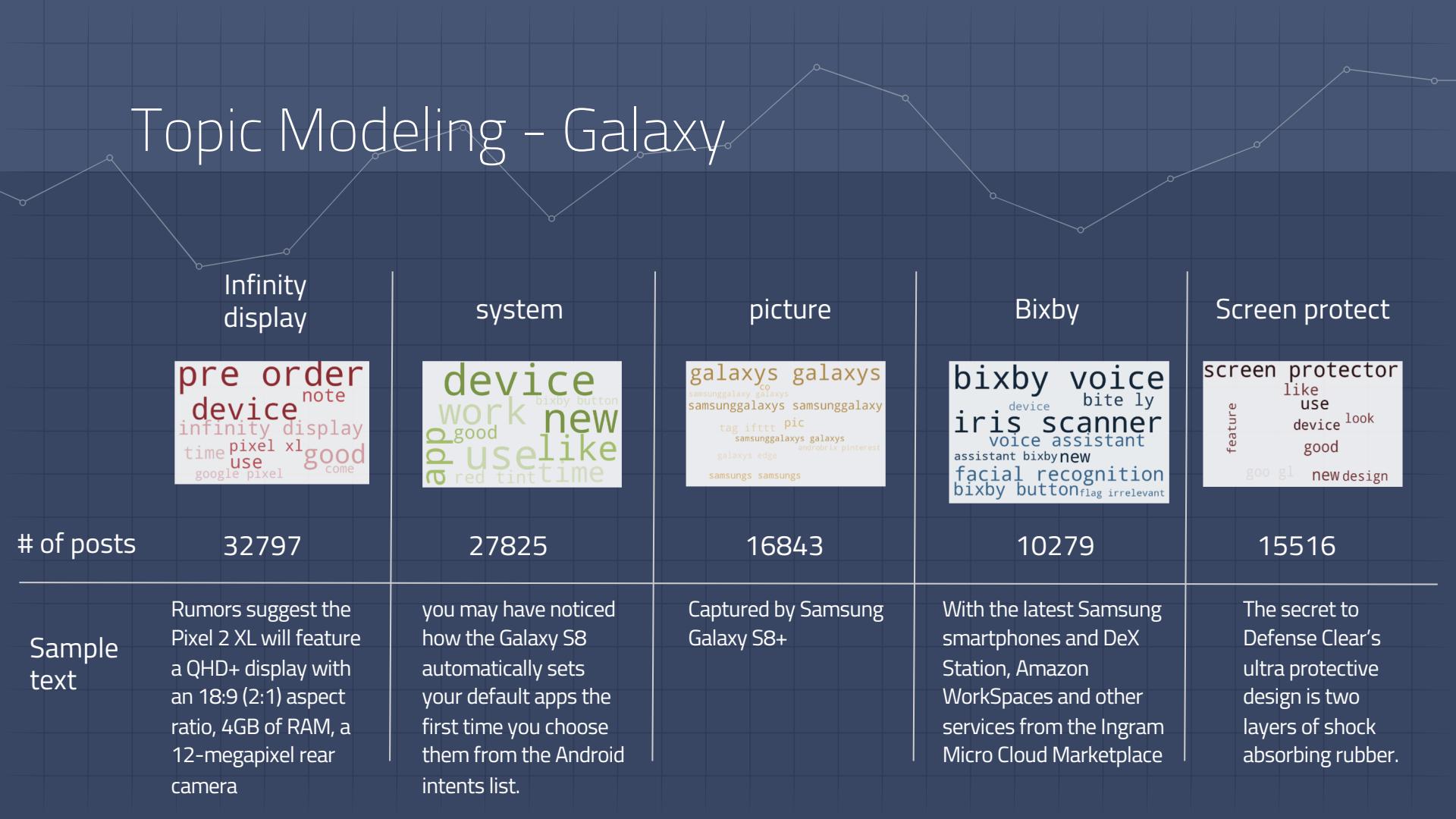


3

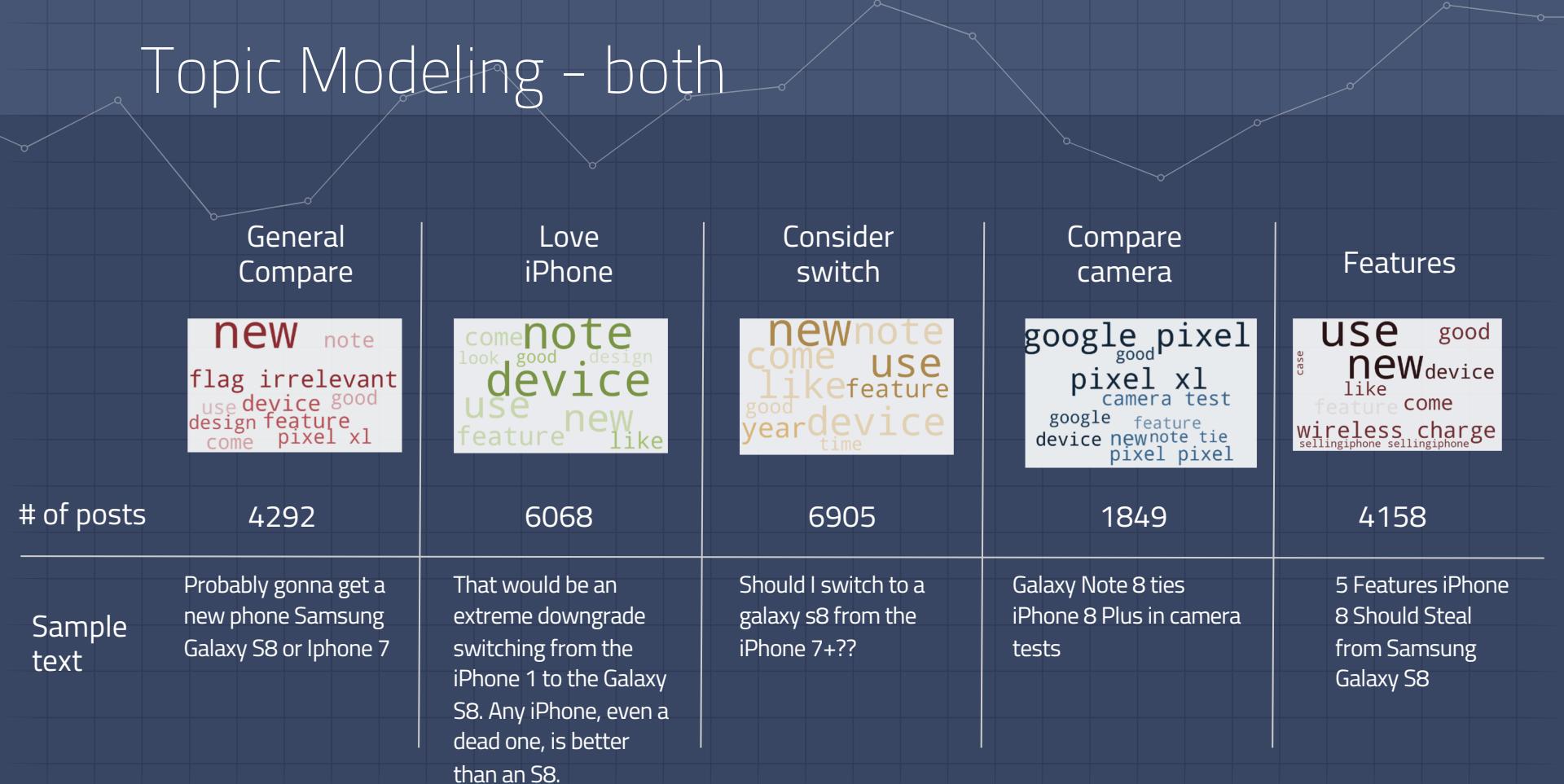
Topic Modeling - iPhone



Topic Modeling - Galaxy



Topic Modeling - both



Sentiment Analysis

Goals:

1. Identify the impact of product launch
1. Analyze sentiment to find product adoption patterns for different segments of customers

Sentiment Analysis - Preparation



Sentiment Analysis tools in Python - TextBlob (polarity)

spaCy, NLTK, coreNLP, Gensim, ...

Use CountVectorizer and LDA to group the data into 2 topics

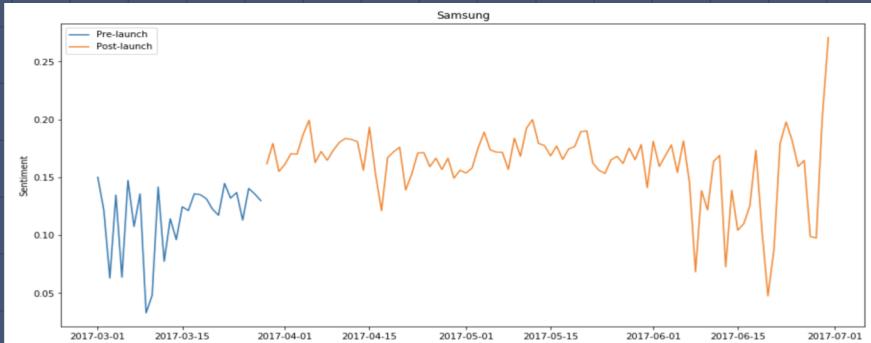
Top 10 words in each topic:

```
[Topic 0]
['s8', 'galaxy', 'samsung', 'new', 'phone', 'android', 'plus', 'screen', 'note', 'smartphone']

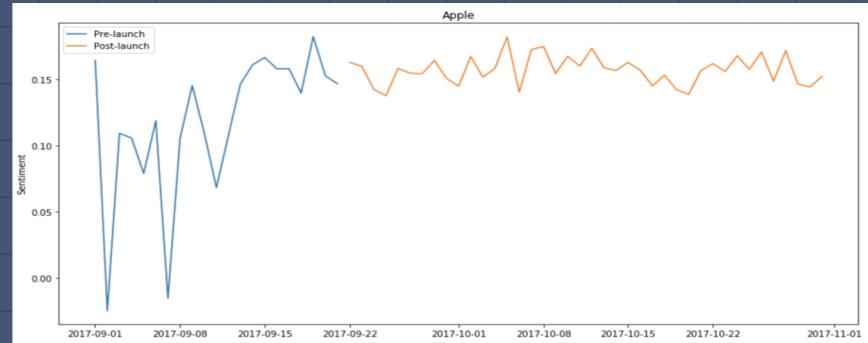
[Topic 1]
['iphone', 'apple', 'plus', 'new', 'com', 'charging', 'phone', 'wireless', '2017', 'camera']
```

Sentiment Analysis - Product Launch Impact

Samsung ↑32%



Apple ↑1%



Pre-launch

Post-launch

Samsung beats the expectation, while Apple shows no surprise.

Sentiment Analysis - Product Launch Impact

SAMSUNG

The first release since its global recall of the fire-prone Galaxy Note 7

Unprecedented design and infinity display

Enhanced software



Samsung Dex gadget transforms the Galaxy S8 into a PC

Geoffrey A. Fowler  @geoffreyfowler

The #GalaxyS8 makes the iPhone 7 look absolutely ancient. It's gonna be a long few months, Apple fans. on.wsj.com/2ogWBoW



THE WALL STREET JOURNAL.
English Edition | October 16, 2019 | Print Edition | Video

179 11:11
162 people like this

TECH | PERSONAL TECHNOLOGY: REVIEW
Galaxy S8 First Look: Samsung Goes to Greater Lengths to Impress

Sentiment Analysis - Product Launch Impact

SAMSUNG

Good Sentiment



Bad Sentiment



Sentiment Analysis - Product Launch Impact



iPhone is more expensive
iPhone X is too expensive?

Most Expensive

iPhone 8: muted reaction and small queues lead to questions over demand



Exactly as leaked...

iPhone 8?

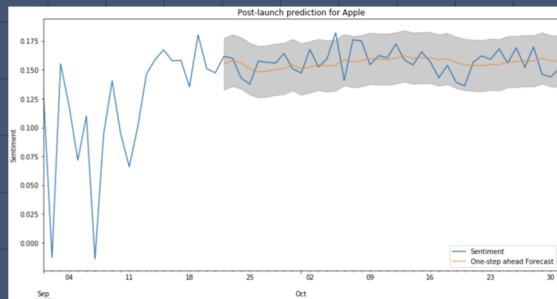
Too expensive...

Questions on the new technologies: no home button, face ID, etc...

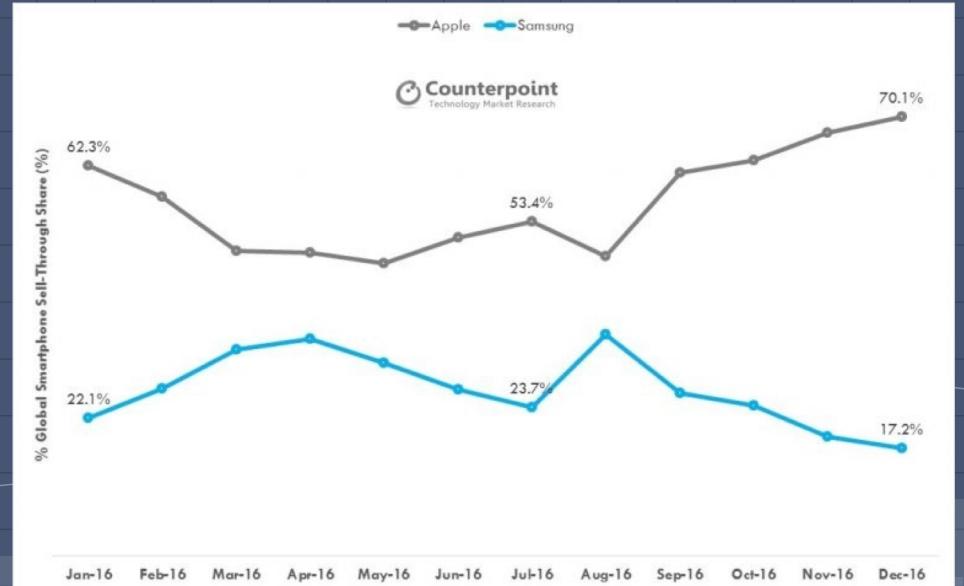
Sentiment Analysis - Time Series Forecast



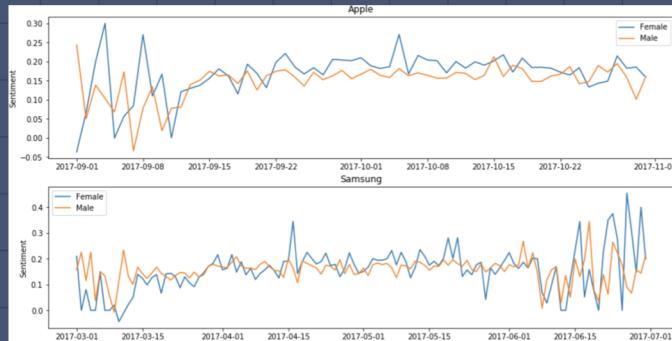
Might not be a good dataset



2017 Smartphone Market Share



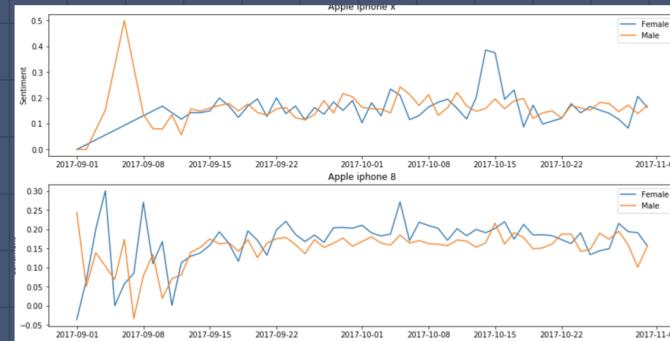
Sentiment Analysis - Gender Difference



Apple

Female Male

Samsung

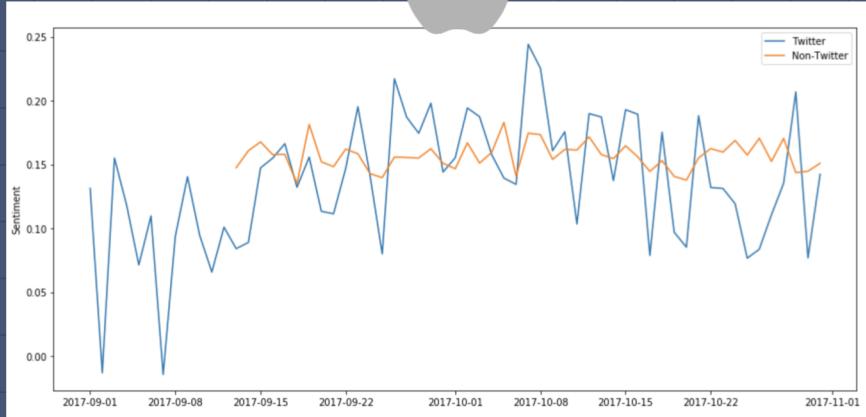


iPhone X

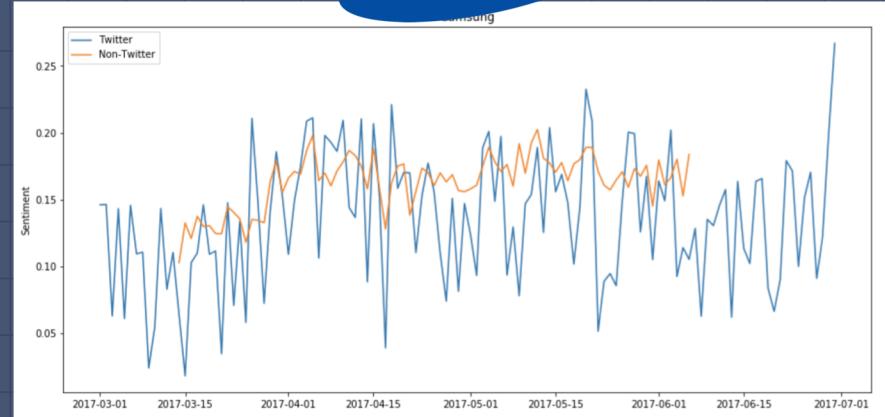
iPhone 8



Sentiment Analysis - Platform Difference

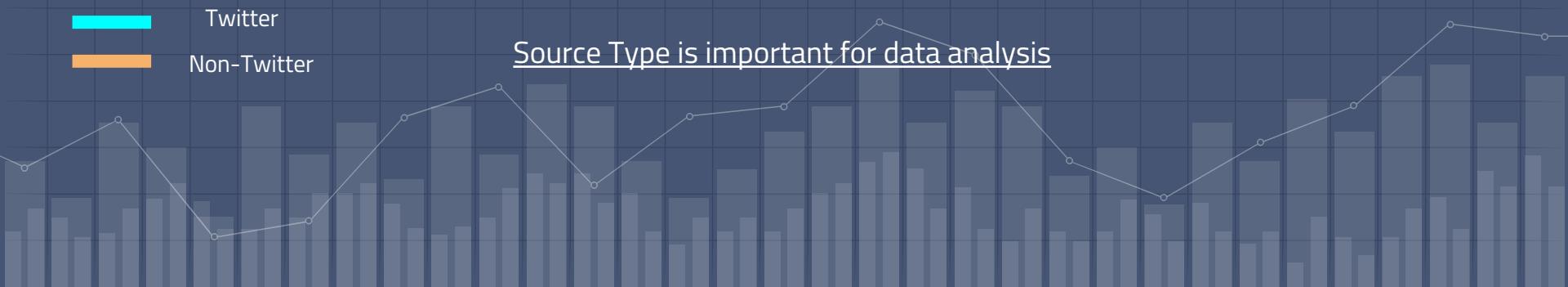


SAMSUNG



Twitter
Non-Twitter

Source Type is important for data analysis



Sentiment Analysis - More Discoveries



Professionals and celebrities (has 2,000+ followers) tend to have lower sentiment mean and higher standard deviation than normal users.

A post with video yields the highest sentiment for both Samsung & Apple, followed by images.



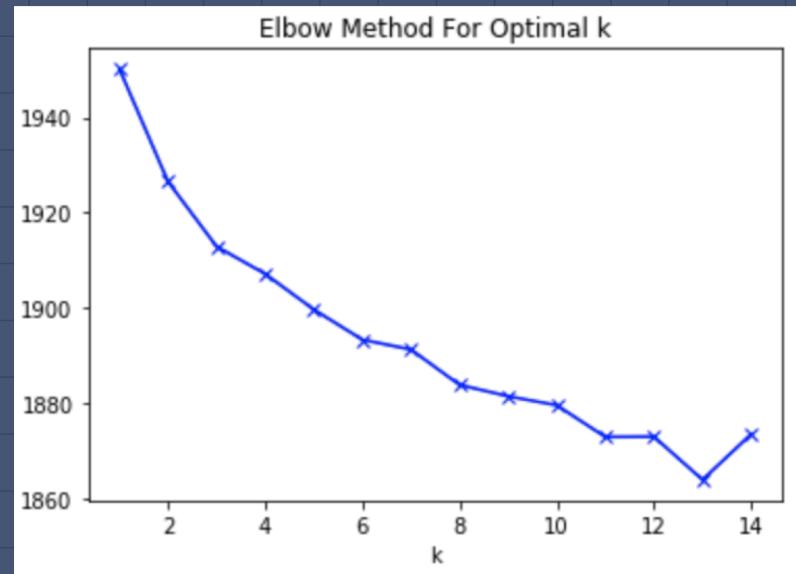
Clustering

Goals:

1. Identify the quality, price, and value perceived by users
1. Segment market based on the user's preference for each product and conduct user profiling analysis on market segmentation

Methodology for k-means Clustering on Review Text

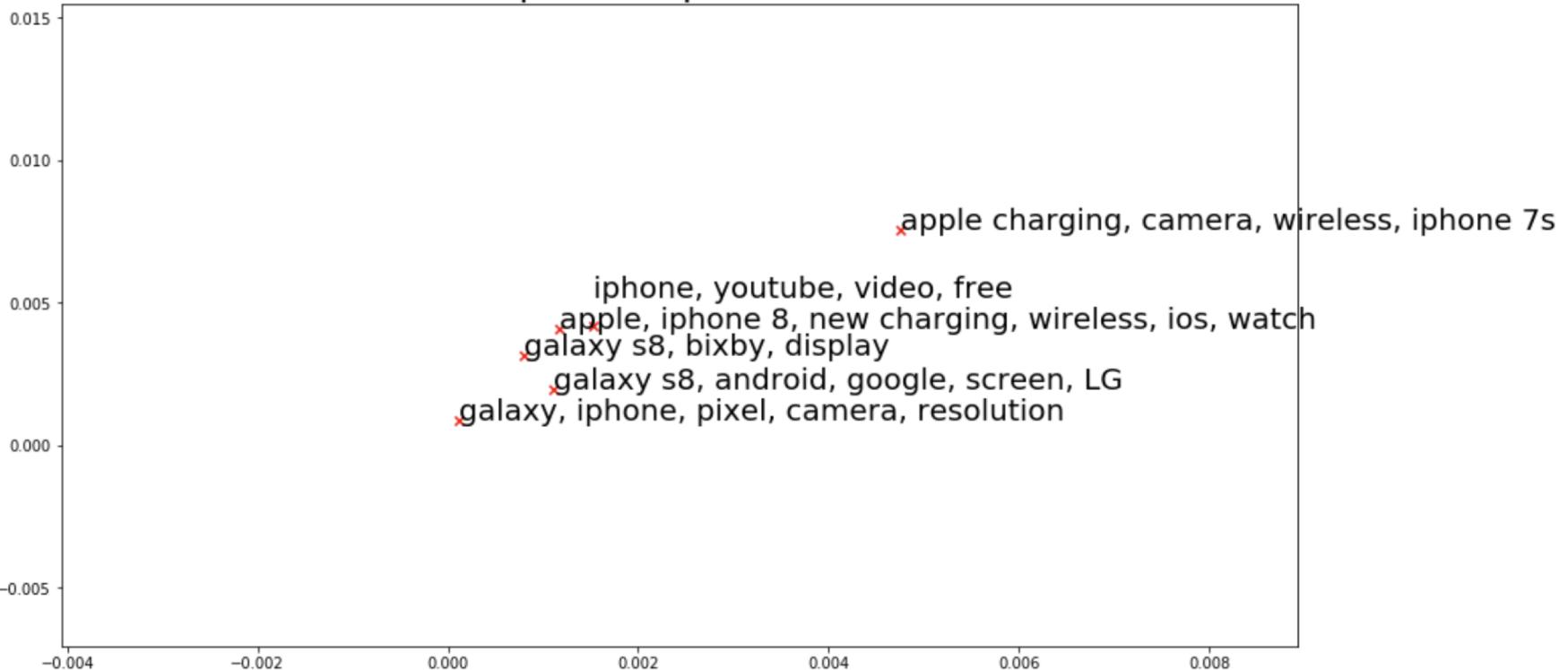
- ❖ Assumption of k
 - Three products: Samsung Galaxy, Iphone x, Iphone 8.
 - Three dimensions that reviewers care about the most price, quality, and value.
- ❖ Feature selection
 - Features are set to be the review text itself by using their tf-idf vectors
- ❖ Goal
 - Cluster customers based on the features for each product, in the dimensions of price, quality, and value perceived by customers.



Optimal K = 6

Top Terms per Cluster

Top Terms per Cluster



Quality, price, and value dimensions for each cluster

	 Positive Objects	 Negative Objects	Products
Cluster 0	Wireless charging, OLED (Super Retina Display)	New battery, cracking design	iPhone x
Cluster 1	Camera, resolution, screen, Pixel	Red, display, exploding, bixby, battery, schematic	iPhone, Galaxy
Cluster 2	Camera, Pixel, portrait	Battery, camera, Pixel, design, price	iPhone 8 plus
Cluster 3	Camera, screen, case, Pixel, LG	Red, device issue	Samsung Galaxy, Apple iPhone
Cluster 4	Flagship, Bixby, display, camera, button, device, brand	Bixby, issue, voice, image	Samsung Galaxy
Cluster 5	Best camera, brand, wireless charging, case, upgrade	Price, battery, charging, GB, upgrade, series	iPhone

User Profiling of each Cluster

Cluster 0: iPhone 8 plus, iPhone x:
design

wireless charging, OLED

bad

Cluster 1: Galaxy, iPhone, Pixel:
unfavorable Bixby

camera, resolution, image

Cluster 2: iPhone 8 plus, Pixel:
high price

portrait, design

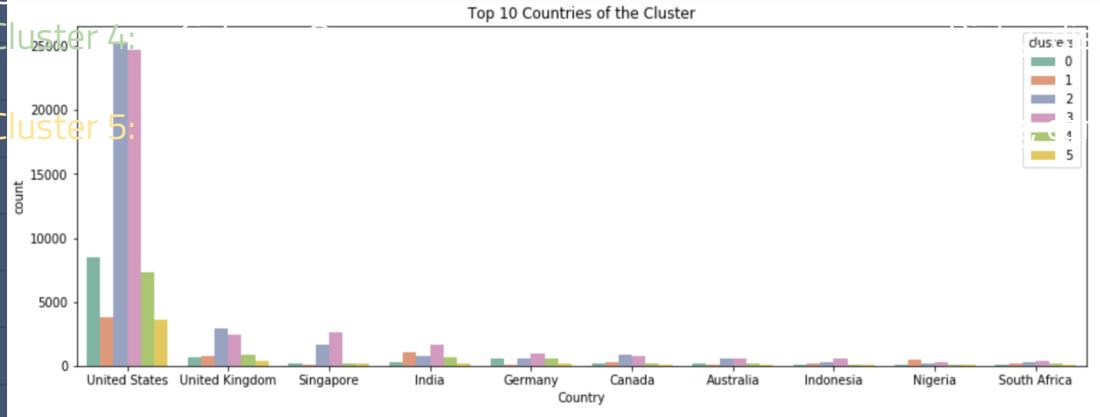
Cluster 3: Galaxy s8, Google, LG:
issue

camera, screen, case

device

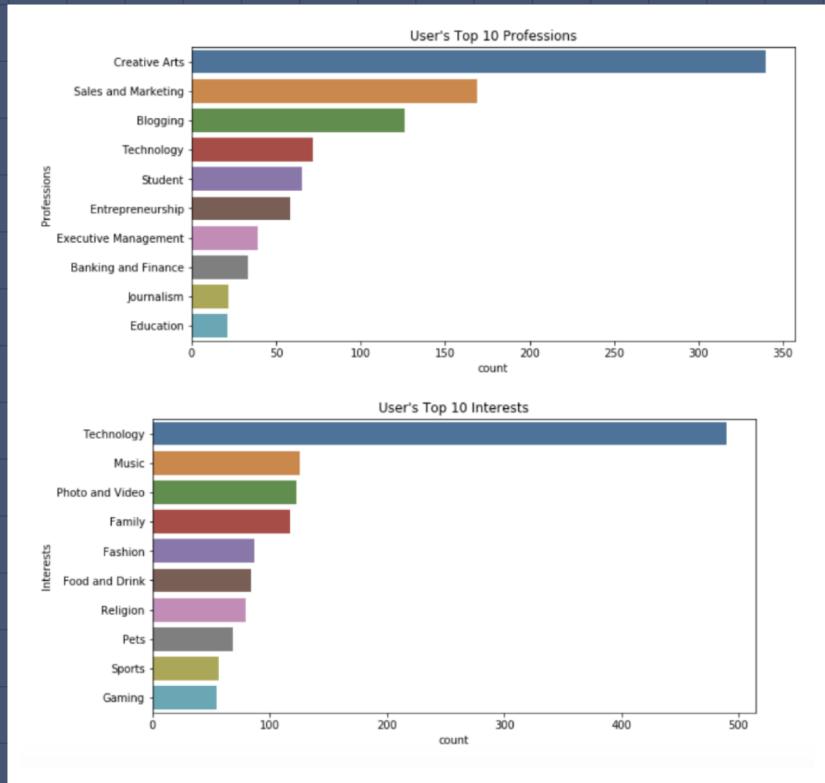
Cluster 4:
Cluster 5:

Top 10 Countries of the Cluster

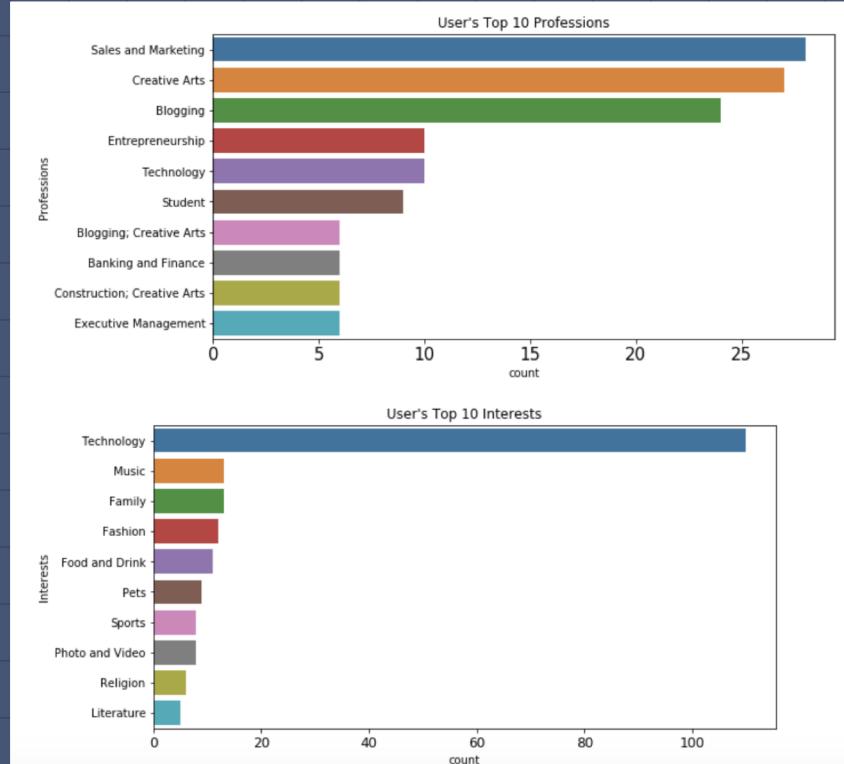


China, India, followed by the United States are the world's largest market for smartphone sales since 2017 (newZoo, 2019).

- **US, UK, Singapore** market consists mostly user **cluster 2 and 3**, whose main concern for iPhone is the **high price**, and the main critique for Galaxy is **device issue**.
- Meanwhile, the most thing they like about iPhone is the **design**, and the favorite feature of Galaxy is its **camera**.
- Cluster 2 and 3 has similar distribution of professions and interests, with the top three professions in **creative arts, sales and marketing, and blogging**. Interests are mostly in **technology**.



- **India** market consists most of cluster 1 and 3, whose reviews majorly focus on **Samsung Galaxy** and the comparison with other **android** smartphones (Google's Pixel, LG).
- They like about Galaxy's **camera**, **image resolution**, and **display**, but have a mixed review about **Bixby** and some of them do not like the **device issues**.
- Similar to cluster 3, their top three professions in creative arts, sales and marketing, and blogging. Interests are mostly in technology.



Results & Next Steps

6

Next Steps

- Reach out to professionals and celebrities to overview the benefits of key features regarding the review of product features as their sentiment is either very good or very bad
 - professionals and celebrities (has 2,000+ followers) tend to have lower sentiment mean and higher standard deviation than normal users.
- Encourage social media posts with hashtags and posts with video and images
 - a post with video yields the highest sentiment for both Samsung & Apple, followed by images.
- Collect qualitative data on design for iPhone 8 plus and iPhone x
 - users have a mixed review on the design of iPhone 8 plus and iPhone x
- Consider more discount activities for iPhone in US market
 - US users are concerned about the high price of iPhone
- Marketing strategy with Pixel as competitor for both Galaxy and iPhone, focusing on the camera feature
 - Google's Pixel is brought up many times as comparison with both products for each user cluster, especially on the camera feature
- ...



Q & A