Audience Text & Sentiment Analysis of Commenting Behavior on Social Platforms:
Informing Brands' Strategy

1

Audience Text & Sentiment Analysis of Commenting Behavior on Social Platforms: Informing Brands' Strategy

ALEX PEAR

KATHERINE PULLY

Abstract

Honing in on the use case for brands attempting to monetize social media platforms and learn more about their audiences, we seek to create an application that specifically focuses on Instagram comments. We will combine this with other platform data or geographic overlays. Applying text and sentiment analysis tools to the comments we scrape (often not exposed through the API), we can define trends of behavior across the site and for particular brands or verticals. For instance, overall positivity/negativity score, indexing across standard metrics, and specific problem or opportunity areas (for instance, complaints about product quality). We also seek to distinguish genuine brand interactions from potentially fake and/or paid responses. The degree of specificity can be massaged as we begin to further explore the data and identify the most insight-rich areas.

Overview

Motivation

Who are the users of this application?

- Social media marketers (agency or in-house)
- Marketing analysts (agency or in-house)
- Companies marketing their products on social media

Who will benefit from this application?

- Marketing agencies: Reduced costs for data science talent, better results for clients
- Companies marketing on social media: Better insights, lower costs
- Social media marketers & marketing analysts: Reduced time spent manually mining text data

### Why is this application important?

- No product addressing this need, likely due to API restrictions
- With Instagram being an increasingly fertile ground for company growth, fully understanding your audience is a significant competitive advantage

4

Industry Focus

To hone in and prove out the use case of this application (to later scale to other industries), we focused on beauty brands.
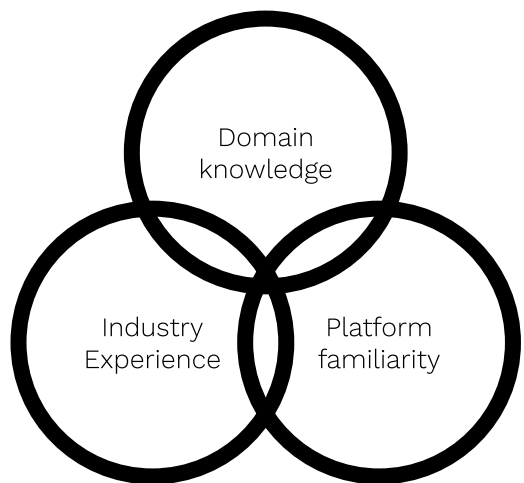
We chose beauty because:

- Domain knowledge of the industry, from work experience with large/small beauty brands
- Grassroots and large brands alike advertise and interact on Instagram as a main (or sometimes sole) channel → highly relevant

5

Goodness

Domain knowledge

Industry Experience

Platform familiarity
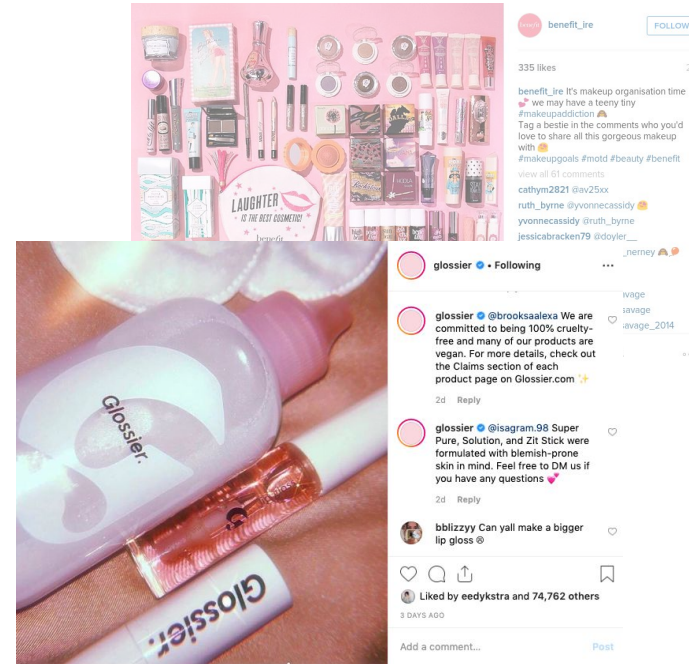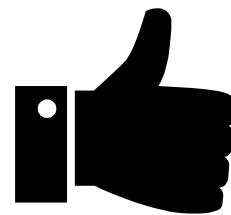
Part of our motivation in pursuing this application was the domain knowledge we had coming into the project. Not only are we familiar with marketing from previous employment (specifically social media marketing), but we have worked with beauty brands in a marketing capacity.

6

Remediation

Based on the information our dashboard application provides, social media marketers will be able to more quickly respond to queries posted to brands via social media, and they will be able to easily collect consumer recommendations and feedback from their posts. Without this application, many of these insights would be lost among the thousands of irrelevant and/or spam comments a post normally receives.

reduced cost	faster insights	increased engagement
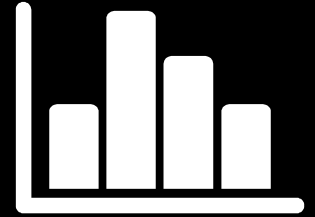
Data

Instagram and Twitter

Data Sources

- ## *Instagram*
    - For each brand, we scraped the comments for the last 100 Instagram posts (image or video)
    - Data includes brand name, timestamps of post and comment, username, user id, and comment text
    - >770K comments from 10+ beauty brands

Data Sources

- ***Twitter***
  - Employing the Twitter API, we collected tweets in the past 7 days (API restriction) directed at the brand (@) and for tweets hashtagging it (#). This data excluded retweets to have a diverse dataset, although kept track of # retweets for relative importance
  - Data includes tweet text, timestamps, user name, user bio, # retweets, # likes, etc.
  - >200k tweets
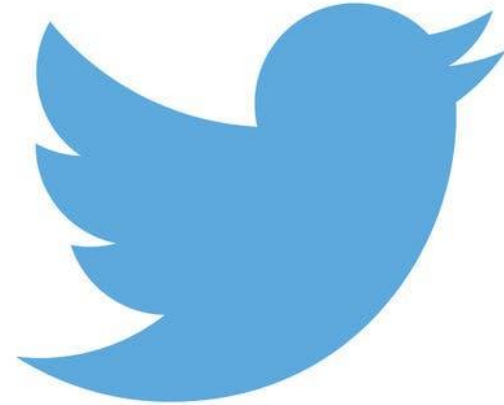
10

post_owner	post_ts	comment_ts	comment_id	comment_owner_username	comment_owner_id	comment_text
glossier	1563812235	1563812251	17849044111510980	asapaverie	15394683299	cuteee😍
glossier	1563812235	1563812267	18052789741177250	camilalunares	1266969796	💖💖💖
glossier	1563812235	1563812274	18085690753021820	genevievetorres_	1638857330	❤️❤️
glossier	1563812235	1563812292	18042856258170559	guavaa_k	4067123338	If only I had money 😭😭
glossier	1563812235	1563812293	18054755776134657	gidley.brooklyn	2955496378	maybe instead of getting rid of stickers actually reduce the plastic in your product packaging
glossier	1563812235	1563812316	18057281611094225	makeupbyjudyy	11146038306	😍😍😍
glossier	1563812235	1563812330	18081646153064648	potatoegyal	4721865066	Best gloss bee
glossier	1563812235	1563812334	18015774496219928	potatoegyal	4721865066	Ever
glossier	1563812235	1563812473	18068437426096590	giavaladez	9779483628	where can i get this again @glossier
glossier	1563812235	1563812515	17898202651356199	nicholeperris	1992876259	😍
glossier	1563812235	1563812552	18085226566014493	dolandobrev	5703522657	Can I use it on the eyes too?
glossier	1563812235	1563812842	18040236727198041	glowtoxnyc	7129839124	🔥🔥🔥🔥
glossier	1563812235	1563812857	18086420467058011	kidder.me	6304642840	Kissed by the sun😊
glossier	1563812235	1563812957	17896041247360275	_ava_is_a_cowgirl_	8782461098	I just bought some 😂😂😂
glossier	1563812235	1563813055	17892142585367454	glossier	1415714040	@dolandobrev Lip Gloss was developed for use on lips only so we don't recommend use around the e
glossier	1563812235	1563813157	17871162331424974	glossier	1415714040	@giavaladez Lip Gloss is available on Glossier.com and at both our retail locations in NYC and LA ✨
glossier	1563812235	1563813602	17847214636524732	karticajingga	271579274	💜
glossier	1563812235	1563813872	17848178800508799	anna.enfp	3951730324	@gidley.brooklyn they're working on it! they've been responding to a lot of comments like this for mont
glossier	1563812235	1563813956	18056632015191697	yayo649	8313690229	@guavaa_k no worries they're overpriced 😂
glossier	1563812235	1563814222	17907228022326727	razha123455	8471635460	I love this glossy lips ❤️😢
glossier	1563812235	1563814532	18051820591149418	anagonpe92	1658145636	Glossier x Spain pleaaaaaaaase 😮😮😮😄😮😮😮😮😮😮
glossier	1563812235	1563814811	18007285426235142	sydneymargolis_	196416622	Please update us on Glossier Boston 💜
glossier	1563812235	1563815183	17848028881517393	gidley.brooklyn	2955496378	@anna.enfp cool i'm glad they're listening! i hope they also reduce plastic in their containers also. like

Brand	Query	user_id	status_id	created_at	screen_name	text
bumbleandbumble	@bumbleandbumble	983154954	1159151461965545474	2019-08-07 17:16:59	AgeofIrony	@MarienneMaid @Trixie_Ren @starwars @bumbleandbumble Shampoo, conditioner, dry shampoo, sea s
bumbleandbumble	@bumbleandbumble	1042162672266080257	1159151861825335296	2019-08-07 17:18:34	MarienneMaid	@AgeofIrony @Trixie_Ren @starwars @bumbleandbumble Yessss, the Crait salt spray tie-ins are ready m
bumbleandbumble	@bumbleandbumble	1042162672266080257	1159141433711812608	2019-08-07 16:37:08	MarienneMaid	The only @starwars marketing crossover I want for IX is a Ben Solo hair care products line @bumbleandb
bumbleandbumble	@bumbleandbumble	954416312134008832	1159149330714038272	2019-08-07 17:08:31	An_Absent_Mind	@Rtndpds @MarienneMaid @starwars @bumbleandbumble Just screaming into the mic.
bumbleandbumble	@bumbleandbumble	954416312134008832	1159141986735144961	2019-08-07 16:39:20	An_Absent_Mind	@MarienneMaid @starwars @bumbleandbumble Ben starts up a YT channel about his hair routine and he
bumbleandbumble	@bumbleandbumble	48364817	1159149281577721859	2019-08-07 17:08:19	ReyloMel	@MarienneMaid @starwars @bumbleandbumble Perfect!
bumbleandbumble	@bumbleandbumble	755526210294280192	1159148658031566848	2019-08-07 17:05:50	thesupremejedi	@giaciak2 @MarienneMaid @starwars @bumbleandbumble petition for Kylo to give hair advice to all of u
bumbleandbumble	@bumbleandbumble	384404209	1159145657594257408	2019-08-07 16:53:55	pinkfish_daria	@MarienneMaid @starwars @bumbleandbumble I have some feelings about his hair. Like jealous maybe.
bumbleandbumble	@bumbleandbumble	2182598810	1159144898559242240	2019-08-07 16:50:54	Rtndpds	@An_Absent_Mind @MarienneMaid @starwars @bumbleandbumble Imagining Kylo saying 'Smash that l
bumbleandbumble	@bumbleandbumble	17604456	1159106934408208384	2019-08-07 14:20:02	POPSUGARBeauty	.@GeorgiaAFowler shares her favorite products from @bumbleandbumble and more. https://t.co/dYqXC
bumbleandbumble	@bumbleandbumble	193763632	1159068210387464192	2019-08-07 11:46:10	numberthreebath	Beautiful cut and style by Cristina. We love to use the @bumbleandbumble Curl range to compliment the
bumbleandbumble	@bumbleandbumble	14834876	1158878851067338753	2019-08-06 23:13:43	foodfinery	@sadglamour @naturallyfled @bumbleandbumble Yeah. I have to be careful with brown roots and ash en
bumbleandbumble	@bumbleandbumble	14834876	1158850420019564544	2019-08-06 21:20:45	foodfinery	@naturallyfled @sadglamour @bumbleandbumble My best girlfriend is a hairdresser and gave me a bottle
bumbleandbumble	@bumbleandbumble	14834876	1158844192849715200	2019-08-06 20:56:00	foodfinery	@sadglamour @naturallyfled This one tho 😍😍@bumbleandbumble https://t.co/i6L5OagK9E
bumbleandbumble	@bumbleandbumble	2902618321	1158871801948020737	2019-08-06 22:45:43	sadglamour	@foodfinery @naturallyfled @bumbleandbumble I love this one too! Great for blonde hair https://t.co/VzEHzoHBPv
bumbleandbumble	@bumbleandbumble	1106275344523259904	1158844852433350656	2019-08-06 20:58:37	naturallyfled	@foodfinery @sadglamour @bumbleandbumble i love that one! i have their blow dry heat protectant 😍
bumbleandbumble	@bumbleandbumble	18138157	1158422416114233345	2019-08-05 17:00:01	FashionCanada	\I didn't realize it would be so successful until it was copied.\"--Michael Gordon
bumbleandbumble	@bumbleandbumble	82373001	1157337799969951744	2019-08-02 17:10:08	mitchellssalon	Your perfect summer set, to go! @bumbleandbumble eliminates summertime hair woes (sun, chlorine, an
bumbleandbumble	@bumbleandbumble	2369465382	1157218867560833024	2019-08-02 09:17:32	keshia_east	@missmja930809 Hi darling!! I use a brand called cantu, @bumbleandbumble and @love_amika ! Those a
bumbleandbumble	@bumbleandbumble	840975093325000705	1156997534834987010	2019-08-01 18:38:02	TeamatNylon	T E X T U R E D L O B ✨ Lightly tonged + spritz with Bb HIO Dry Oil Spray. Cut | Michelle | Colour Jay
bumbleandbumble	@bumbleandbumble	840975093325000705	1156627444989923329	2019-07-31 18:07:26	TeamatNylon	H O N E Y B L O N D E ☀️ Bangs and Balayage. Use Bb Repair Creme before applying heat. Colour | Fl
bumbleandbumble	@bumbleandbumble	840975093325000705	1155917399931183105	2019-07-29 19:05:58	TeamatNylon	B L O N D E B L E N D S 💯 Style with Bb Go Big Treatment! Colour | Shannon #beforeandafter @WellaP

```scala
val user_brands = beauty.map(row=>(row._1, row._5)).distinct.groupByKey().map(row=>(row._1, row._2.toSet))

def get_union(brand: String, users1: scala.collection.immutable.Set[String]) = {
    var cross_counts = collection.mutable.Map[String, Int]()
    for (brand <- distinct_brands) {
        val overlap = users1.intersect(user_brands_map(brand)).size
        cross_counts += (brand -> overlap)
    }
    cross_counts
}
val user_brands_intersect = user_brands.map(row=>(row._1,get_union(row._1,row._2)))
val df = user_brands_intersect.toDF("col1","col2")
val exploded_rdd = fm.select(fm("col1"),explode(fm("col2"))).filter(row=>row(0)!=row(1))
```

Result of user_brands_intersect:

```
scala> user_brands_intersect.take(10).foreach(println)
(anastasiabeverlyhills,Map(kyliecosmetics -> 1295, benefitcosmetics -> 1315, chanelo
fficial -> 153, sephora -> 729, anastasiabeverlyhills -> 43802, fentybeauty -> 1085,
 bumbleandbumble -> 26, honest -> 18, post_owner -> 0, bobbibrown -> 236, maccosmeti
cs -> 707, urbandecaycosmetics -> 999, skinlaundry -> 9, patmcgrathreal -> 581, glos
sier -> 209, madisonreedllb -> 1))
(glossier,Map(kyliecosmetics -> 174, benefitcosmetics -> 205, chanelofficial -> 51,
sephora -> 175, anastasiabeverlyhills -> 209, fentybeauty -> 232, bumbleandbumble ->
 23, honest -> 34, post_owner -> 0, bobbibrown -> 105, maccosmetics -> 118, urbandec
aycosmetics -> 110, skinlaundry -> 28, patmcgrathreal -> 203, glossier -> 29105, mad
```

Insights
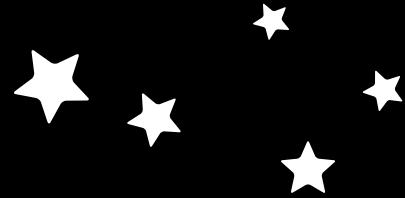
1. If user base overlaps in one social media platform, it is likely to overlap on another.

2. Overlap of user base depends less on product offered and more on company branding (i.e., hair care companies do not necessarily share a user base if their positioning is not the same).

3. Number of replies on a post does not correspond to number of queries. In other words, some brands choose to reply to "statements" with words of encouragement rather than just questions in order to increase engagement.

4. More established, higher-end brands tend to have less social media engagement despite similarly established presence on social media platforms.

Obstacles

- **Data comparability:** while brand posts on Instagram get thousands of comments, Twitter posts primarily garner likes and retweets. Instead of pulling Twitter posts responses, we pulled "#[brand name]
- **Emojis 😒:** social media responses often include emojis with text or no text at all! We could not get the emojis to render in HDFS for analysis (display as '?')
- **Local code development:** Initially we found it time consuming to copy/paste or transfer code back and forth to Dumbo, but we came up with a modular solution whereby we wrote small, encapsulated bits of code into separate files and ran them as scripts in our spark shell

Summary

- By processing large amounts of text data in real time, companies can keep track of their branding across multiple social media accounts, respond to queries quickly, and measure the health of their platform marketing presence.

17

Future Work

- ■ Incorporate **additional data sources**: geographic overlays → census data, other social platforms → reddit
- ■ Collect additional data and modify the application to run in **real-time**, and **collect** data (i.e. tweets) continuously from moment of application launch
- ■ Two separate UI versions: brand-specific and industry overview

18

References
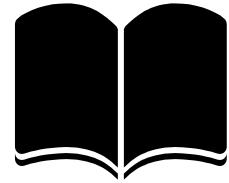
1. T. White. Hadoop: The Definitive Guide. O'Reilly Media Inc., Sebastopol, CA, May 2012.

2. I. Sen, S. Singh, A. Aggarwal, P. Kumaraguru, S. Mian, A. Datta. 2018. Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram. WebSci '18 Proceedings of the 10th ACM Conference on Web Science. DOI=https://dl.acm.org/citation.cfm?id=3201105.

3. Pak, Alexander & Paroubek, Patrick. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Proceedings of LREC. 10.

4. Emilio Ferrara, Roberto Interdonato, and Andrea Tagarelli. 2014. Online popularity and topical interests through the lens of instagram. In Proceedings of the 25th ACM conference on Hypertext and social media (HT '14).   ACM, New York, NY, USA, 24-34. DOI=http://dx.doi.org/10.1145/2631775.2631808

5. Georgios Paltoglou and Mike Thelwall. 2012. Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media. ACM Trans. Intell. Syst. Technol. 3, 4, Article 66 (September 2012), 19 pages. DOI=http://dx.doi.org/10.1145/2337542.2337551

6. C. W. Seah, H. L. Chieu, K. M. A. Chai, L. Teow and L. W. Yeong, "Troll detection by domain-adapting sentiment analysis," *2015 18th International Conference on Information Fusion (Fusion)*, Washington, DC, 2015, pp. 792-799.

7. D. Abdullah, N. Sakinah. 2017. Sentiment Analysis of Online Crowd Input towards Brand Provocation in Facebook, Twitter, and Instagram. Proceedings of the International Conference on Big Data and Internet of Things. DOI=https://dl.acm.org/citation.cfm?id=3175689.

8. N. Carah, M. Shaul. 2015. Brand and Instagram: Point, tap, swipe, glance. Mobile Media & Communication. DOI=https://doi.org/10.1177/2050157915598180.

9. W. Tripiawan, H. Amani, A T Wijaya. 2019. Effectiveness Analysis of Social Media Ads as A Promotional Media (Case Study: Instagram Taya.Id). 1st International Conference on Industrial and Manufacturing Engineering. DOI=https://iopscience.iop.org/article/10.1088/1757-899 X/505/1/012095/pdf.

10. N. Segev, N. Avigdor, E. Avigdor. 2018. Measuring Influence on Instagram: a Network-oblivious Approach. DOI=https://arxiv.org/pdf/1806.00881.pdf

Demo!
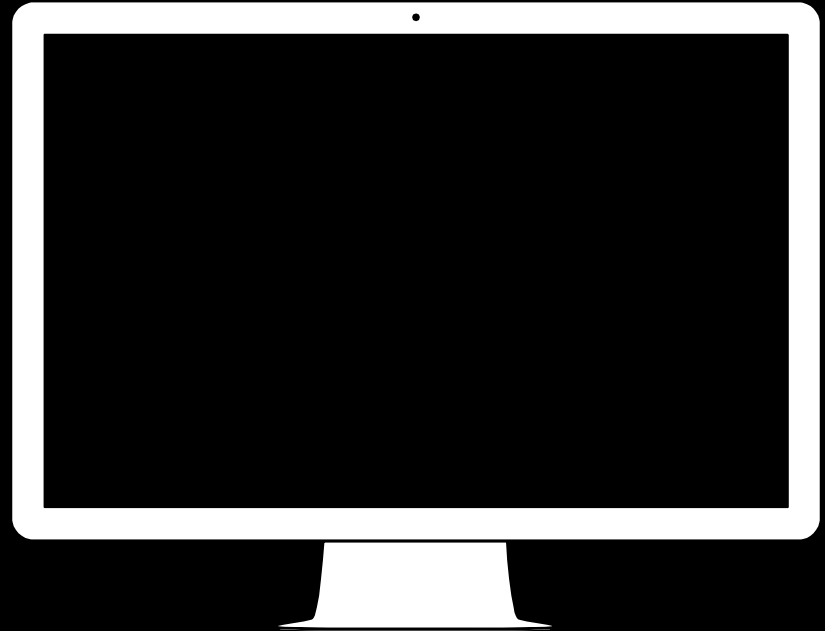
Credits

Special thanks to all the people who made our project easier:

- HPC
- Tableau
- Twitter (Open source API)
- Instagram / Instagram scraper library

Thanks!

Any questions?

You can find us at

- kp1276@nyu.edu
- aap703@nyu.edu

22