Diwali: Tradition Through The Twitter Lens

CCT414: Assignment 3

Asiya Atcha (992242373)

University of Toronto

Research Question

In my proposal, my initial question was: What socioeconomic trends become apparent

through an analysis of #Diwali?

Because the nature of the research was socioeconomic, I didn't have a specific

hypothesis in mind. This followed from the assertion in Mejovas's work that questions in

this area are largely exploratory (Mejova, 2015). There is almost no data on this topic,

and the purpose of my analysis was largely to get a sense of what is out there.

For reasons that I will elaborate in the qualitative analysis, this proved to be overly

ambitious. The information that I did uncover, however, opened some fascinating

avenues for further exploration. I remain convinced that the insights garnered through

this research are essential to selling products and services in increasingly localized

market segments in India and the diaspora.

The Theoretical Framework

The Indian Subcontinent is home to one of the largest and fastest growing population

groups on the planet, with over 1.25 billion people living in India. 65% of those people

are under the age of 30 (US Census, 2011). The diaspora of South Asians is also very

large, with the number of card-holding Non-Resident Indians estimated to be around

30,000,000 by the Ministry of Indian Overseas Affairs (MIOA, 2012). This does not

include the millions around the world who identify as being of Indian ethnicity.

This is a highly "wired" demographic. There is evidence to indicate that mobile phone

usage in India will outstrip the United States by this year (Muralidharan et.al, 2015). In

spite of this, there is little marketing research or scholarly literature written on this

subject matter. This may explain why both within and without India, marketing is so shallow and has a tendency to treat the Indian population like a monolith.

Thoughtful investigation of the ways in which Indians use social media has proven to be effective in shaping and directing the success of marketing campaigns in the Indian subcontinent (Kumar et.al, 2012). Twitter, in particular, has been identified as a promising medium (Kuhikar, 2013). It could also prove theories that collectivism is prioritized over individualism (Muralidharan et.al, 2015b) and community outreach are of high value in the Indian culture, helping to direct the decisions of civic organizations during the Diwali season.

Quantitative Analysis

Over the course of the month of November, I collected a total of 45073 tweets for '#Diwali'. My collection was sporadic at best, and were I to run this experiment again, I would definitely be more consistent. That said, I still don't think I could have collected all the tweets related to Diwali, given the sheer volume of them, and I think it would be helpful to be more selective in my queries. Here are some tables illustrating my results:

Screen Name	Mentions	Who Are They?
rameshlaus	922	*"Social Media Strategist"*
Leicester_Merc	848	*British newspaper*
BiscuiteersLtd	670	*British cookie company*
SohoRoadBID	543	*Birmingham business association*
KurKureSnacksLtd	470	*Indian snack company*
leicspolice	396	*Leicester police force*
STR_360	362	*Tollywood actor fanpage*
BSEIndia	334	*Independent trading index*
jassansi	332	*Birmingham based photographer*
Indiancinema360	327	*South Indian film portal*
LeicsPolice	256	*Leicester police force*

fig. 1 - Top 10 screen names

Hashtags	Mentions	What is This?
Diwali	13824	
Vedalam	1671	*Tollywood Movie*
diwali	1469	
Leicester	1364	*UK Town*
Thoongavanam	717	*Tollywood Movie*
Thala	605	*Tollywood Movie*
Vijay59	480	*Tollywood actor*
Greetings	337	
KamalHaasan	283	*Tollywood actor*
Quotes	247	

fig. 2 - Top 10 hashtags

Date of Collection	Tweets Collected
10/31/2015	40
10/31/2015	1868
11/1/2015	4870
11/2/2015	5150
11/3/2015	5228
11/5/2015	5100
11/6/2015	5100
11/9/2015	3768
11/10/2015	1332
11/18/2015	937
11/15/2015	2257
11/16/2015	2226
11/17/2015	1886
11/18/2015	937
11/19/2015	1037
11/20/2015	710
11/21/2015	686
11/22/2015	363
11/24/2015	236
11/25/2015	447
11/26/2015	500

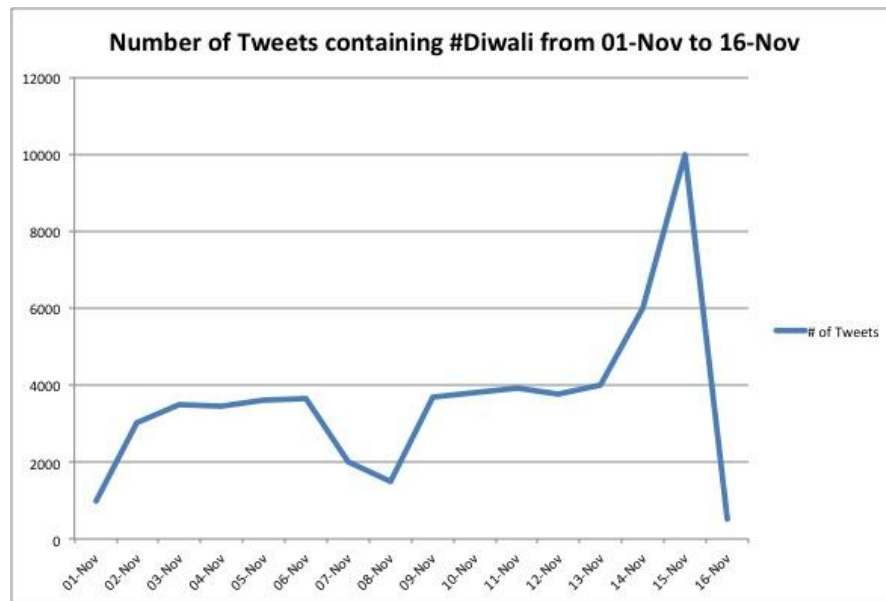| 11/27/2015	449
11/28/2015	180
11/29/2015	206
11/30/2015	190
TOTAL TWEETS COLLECTED	45703

fig. 3 - Total tweets collected

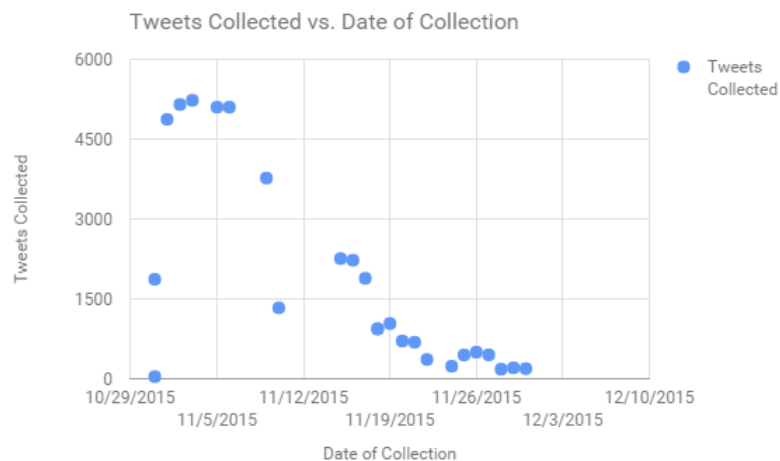fig. 4 - my initial prediction for number of tweets over the course of a month

fig. 5 - my actual results for number of tweets over the course of a month

The validity of the data in the above two graphs is tenuous at best, given the sporadic

intervals at which I collected the data, and the fact that some of files wouldn't run or

were missing.

Filename	Total Words	Positive Words	Negative words	Neutral Words
asiya-diwali_saved-tweets1 (corrupt)	-	-	-	-
asiya-diwali_saved-tweets1b	77585	2033	195	75357
asiya-diwali_saved-tweets2	79196	1528	202	77466
asiya-diwali_saved-tweets3b	0	1552	200	-1752
asiya-diwali_saved-tweets4	81520	1628	288	79604
asiya-diwali_saved-tweets5	85758	1992	235	83531
asiya-diwali_saved-tweets6	85990	1985	241	83764
asiya-diwali_saved-tweets7	83693	2198	403	81092
asiya-diwali_saved-tweets8 (corrupt)	-	-	-	-
asiya-diwali_saved-tweets9	83285	2337	371	80577
asiya-diwali_saved-tweets12	74902	2238	272	72392
asiya-diwali_saved-tweets13	153410	3434	509	149467
asiya-diwali_saved-tweets14	277	1	0	276
asiya-diwali_saved-tweets15 (corrupt)	-	-	-	-
asiya-diwali_saved-tweets16	45135	696	341	44098
TOTALS	929928	21622	2916	905049

fig. 6 - Total words in each file plus sentiment analysis of the total words

Qualitative Analysis

I think one of the main takeaways from this has been that I asked the wrong questions. The sentiment analysis, for example, revealed very little to me, since it wasn't nuanced enough to go beyond positive vs. negative comparison. Given that Diwali is the biggest national holiday in the country, it seems fairly obvious that the tweets would be overwhelmingly positive. There is already some evidence to suggest that people are more likely to tweet positive things (Ferrara & Yang, 2015).

Another stumbling block was the scope of the search. Doing a twitter crawl for the word Diwali, given the context of the occasion, is a little like asking what trends become apparent through a search of "#Christmas". The types of tweets, the potential audiences, the different social and cultural contexts are far too broad to be covered in one query. It is ironic that in attempting to combat the tendency to treat Indians like a monolith, I approached their celebration of an international holiday as though it were a very homogenous experience.

My data analysis did reveal some promising avenues for further exploration. One was the South Indian movie industry, Tollywood. If I exclude '#Diwali' and '#diwali', three of the most common hashtags in my top 5 were all related to Tollywood movie releases. Although I didn't expect this, it is unsurprising, since the Tollywood industry is highly prolific, producing more films than even it's more affluent cousin, Bollywood (Hendrix, 2010). The hashtags were '#Vedalam' '#Thala' and 'Thoongavanam'. When I did a general search for '#Diwali' on the twitter dashboard, I got a lot of tweets about a Bollywood movie opening on Diwali. It's interesting to me to see that there is a thriving

subculture elsewhere in India, producing their own products and marketing them the same way.

The above tweet is an example of the kinds of tweets that are common in relation to South Indian cinema. It might interesting, for example, to do a twitter miner search specific to regions in south India, and analyze the use of '#Diwali' in that region. Perhaps looking for the types of users who tweet the most about the occasion and looking for intersections between the hashtag and promotional hashtags for movie releases.

Another interesting and potential avenue of exploration. In my sample, the tweets related to Indian Bollywood actor, Salman Khan accounted for only 98 of my mentions, compare this to 605, for '#Thala', a hashtag for the character played by South Indian actor, Vijay Ajith in the movie "Vedalam". India is a deeply feudal society; religious and caste divides still play a significant role in shaping the undercurrents of

interaction. Salman Khan, being a Muslim Bollywood actor, is unlikely to trend around a

Hindu celebration, and the only reason he had as many mentions as he did may only be

because of his movie release. Outside of that, he may not have shown up at all. I'm

hypothesizing this based on the fact that an even more popular Muslim actor from

Bollywood, ShahRukh Khan, had no mentions in my sample, in spite of the fact that

both actors have a following of 15 million and 16.5 million each, respectively.

ShahRukh Khan had no movie releases during the Diwali season. Perhaps a side by

side comparison of the mentions of two actors over the course of the holiday period,

based on their economic activity, could reveal some interesting trends.

        The second intriguing intersection of data in my sample was situated in the UK,

namely the city of Leicester. As I mentioned in my theoretical framework, the diaspora

is expansive. When I kept seeing Leicester's civic organizations turn up in my results I

was frustrated because I couldn't draw the connection, and it didn't look authentic, since

"traditional" Indian media outlets weren't represented. Again, I was making the mistake

of treating Indians like a monolith. I was expecting to find tweets from a major economic

hub, such as London or Manchester, or even from businesses that catered to the British

Indian community. Instead Leicester kept turning up. A quick google search revealed

some information as to why this might have been the case. In 2011 UK census data,

Indians in the city of Leicester accounted for 28.3% of the city's population, the highest

concentration of any ethnic minority in the country (UK Census, 2011). And it turns out

that as of last year, the city was on track to become the first Asian majority city in

Europe, with Indians and Pakistanis accounting for up to 45% of the demographic in

some boroughs (Ghosh, 2014). In light of that, it makes sense that such a large

sampling of my tweets would include mentions from there. It is possible that community

outreach and civic engagement would be a higher priority than economic activity in the

diaspora, where establishing roots is a priority. A geolocation-based search of Diwali

tweets in the region could shed some light on this question.

The Data-Information-Knowledge-Wisdom Hierarchy

One observation is definitely that the shift from data to information was far more gradual

than the more immediate move from knowledge to wisdom. At the beginning of this

exercise, I was at the point of having some knowledge, with little information or data to

support my understanding. I knew when I started, for example, that there was a large

Indian British community in the UK and I had some experience of this fact. I had no

data to back up my conclusions and the information available to me was secondhand.

In terms of data, I had a rudimentary amount. I had the most basic grasp of the

python language, and no understanding of what tweets looked like. As I collected raw

data, it made little and less sense over the collection period. I got names and entities

that I could not recognize, and even the json for each file made little sense to me.

Reading them over and working out what things meant took time, and even a little bit of

knowledge about the code or the results took hours of work to get through.

Once I had all the data however, and I started compiling it into lists, things started

to make a little sense, in that I could see some organization. The number of tweets for

specific users could be organized by how many mentions they had, the tweets I

collected could be sorted by date, and there were hashtags that were repeatedly appearing. But as to what that information meant, I still could not understand. Then came the information to knowledge stage. I did some basic twitter searches to find the meaning of some of the hashtags in my results. This led to me drawing parallels and connections. Between Leicester and representation of the diaspora, for example.

The rise to wisdom would require more data, this time on more specific questions. For example, an analysis of the social media performance of various Bollywood and Tollywood actors over the course of the Diwali season could provide insight for producers into which stars would draw the biggest crowds over the Hollywood season, this could lend itself to making informed casting decisions.

Closing Thoughts

This has been, in no uncertain terms, a very painful process. I was not confident in my ability to do the data analysis given my technical proficiency. I was confused by my results and frustrated by my inability to make any sense of them. In the end, however, the way in which all my work came together to give me some actionable insight, has been immensely rewarding.

I can see real trends and avenues for exploration. I've learned something about how a bit of information with little knowledge and less data can result in a knowledge vacuum. I learned that thoughtful forays into the world social media analytics can help to carve new understandings and opportunities to build cultural, social and economic networks grounded in the many ways that a nation of billions experiences and expresses its history and traditions.

References

Belinger, G., Durval, C., & Mills, A.. (2004). Data, Information, Knowledge and Wisdom.

　　Systems Thinking. Retrieved from:

　　http://www.systems-thinking.org/dikw/dikw.htm

Ghosh, P. (2014). The golden mile: could Leicester, the most ethnically diverse place in

　　Britain, become the UK's first Asian majority City? *IBI Times.* Retrieved from:

　　http://www.slate.com/articles/arts/movies/2010/09/superstar_rajinikanth.html

Hendrix, G. (2010) SUPERSTAR Rajnikanth! *Slate.* Retrieved from:

　　http://www.slate.com/articles/arts/movies/2010/09/superstar_rajinikanth.html

Kuhikar, P. (2013). Social media as a future marketing tool in india: An overview.

　　PRIMA, 3(2), 15-21. Retrieved from

　　http://search.proquest.com/docview/1478065466?accountid=14771

Kumar, V., Bhaskaran, V., Mirachandani, R., & Shah, M.. (2012). Creating a Measurable

　　Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles

　　and Tangibles for Hokey Pokey. 32(2), 194-212. Retrieved from:

　　http://simplelink.library.utoronto.ca/url.cfm/483426

Mejova, Y., Weber, I., & Macy, M. W., (Eds). (2015) Twitter: A Digital Socioscope.

　　Retrieved from:

　　http://dx.doi.org.myaccess.library.utoronto.ca/10.1017/CBO9781316182635

Ministry of Indian Overseas Affairs. (2012). India and its diaspora. Retrieved from:

　　http://moia.gov.in/accessories.aspx?aid=10

Muralidharan S., La Ferle, C., & Sung Y.. (2015).How culture influences 'social' in social

media: Socializing and advertising on smartphones in India and the USA.

Cyberpsychology, Behavior, and Social Networking 18 (6), 356-360. Retrieved

from: http://www.ncbi.nlm.nih.gov/pubmed/25977980

Office for National Statistics. (2011). 2011: Key Statistics for Local Authorities in the UK

and Wales. Retrieved from:

http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorit

ies-in-england-and-wales/rft-table-ks201ew.xls

United States Census Bureau. (2011). United States Census Bureau - International

Database. Retrieved from:

http://www.census.gov/ipc/www/idb/informationGateway.php