Rogelio Zuniga R.

Sentiment Analysis

March 2016 | DAT-30

The Question

- * Is there a possible *relationship* between the perception of environmental factors expressed in social media comments, such as beach cleanliness, to airline revenue?
- * In English: if we look at comments about a specific travel location, Bermuda for example, on Twitter and TripAdvisor and score them for negative or positive sentiment, do the results over time suggest anything about travel volume to that place?

The Data: Twitter

* Twitter API search: Geolocation coordinates and key words determined to return relevant comments:

"https://twitter.com/search?q=beach%20trash %20jamaica&src=typd"

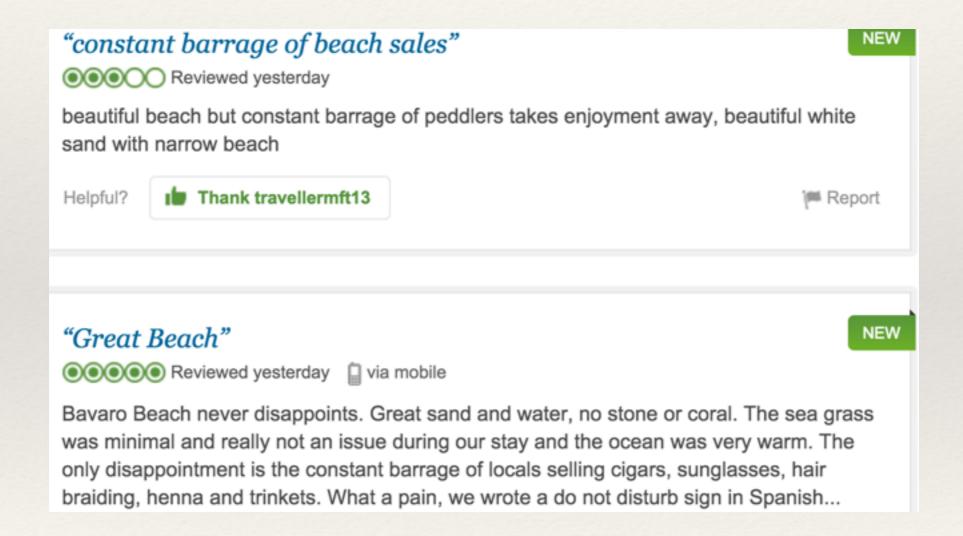


Dr. Claire Nelson @DrClaireNelson · 12 Nov 2015

Now that would be nice.. I am sick of seeing people leave their **trash** on the **beach** in **Jamaica**. We mustban Styrofoam

The Data: TripAdvisor

* TripAdvisor: TA "place" codes



Exploring Social Media is Hard!

Twitter: search API using TwitterSearch

- Geocodes don't help
- Keywords lead to irrelevant returns
- Very hard to search anything except new tweets

Exploring Social Media is Hard!

TripAdvisor: direct from the website using Scrapy

- Keywords lead to irrelevant returns
- Huge results!
- Need to organize a lot of results
- Not allowed to use API for mining data!

Yikes!

- * Without being able to search Tweets far into the past I can't compare sentiment scores with volume data!
- * On the other hand TripAdvisor reviews are far better qualitatively!

Better Quality but More Work

8,842 Reviews from our TripAdvisor Community				
Read reviews that	at mention:		Search reviews	Q
All reviews can	walk for miles long wa	plenty of palm trees	s seaweed problem white powder	
walk forever gre	eat sand cleaned every	y day washed ashore	great walking lots of water sports	
seaweed floating	cleaned everyday g	entle surf great resorts	lovely ocean washing up	
plenty of sun beds	nice temperature c	elean rooms		
Traveler rating	Traveler type	Time of year	Language	
Excellent (25)	☐ Families (12)	☐ Mar-May (7)	○ All languages	
☐ Very good (15)	Couples (24)	☐ Jun-Aug (15)	English (48)	
Average (4)	☐ Solo (1)	□ Sep-Nov (11)	○ Spanish	
Poor (4)	☐ Business (0)	Dec-Feb (15)	OPortuguese	
Terrible (0)	Friends (4)			

```
▼ <div class="entry">
▼ p id="review_301296294">
This is a lovely beach when the "
    <span class="searchHit">seaweed</span>
    " isn't bad. For most days of our trip, it w
    the last day, it was so bad the hotels just one of the resorts next to ours was using a entire time I was ankle deep in it and had t always so bad, but due to it, I would not ch
```

Decisions, Decisions ...

- * Use only TripAdvisor?
- * How to deal with reviews that include keywords but are not relevant? Naive Bayes?
- * How long will it take to gather enough data, scrape it and process it?