

Predicting catastrophic loss from social data

Background:

- **Company:** I work for a risk consulting firm in the Analytics team. One of our specialisations is *natural catastrophe risk analysis* and the impact on insurance.
- **Service:** Our core functions are analytics for reinsurance placements, insurance exposure and portfolio management, **event response** and other risk management products.
- **Clients:** Insurance companies, state/local government, commercial organisations

Who is the **potential audience** of this analysis?

INSURERS



REINSURERS



DISASTER RESPONSE



GOVERNMENT/ COUNCILS



What do our audience **want to know** during a catastrophe?

What is the total loss I can expect from the event?

Which regions are experiencing the greatest damage?

How many claims can I expect?

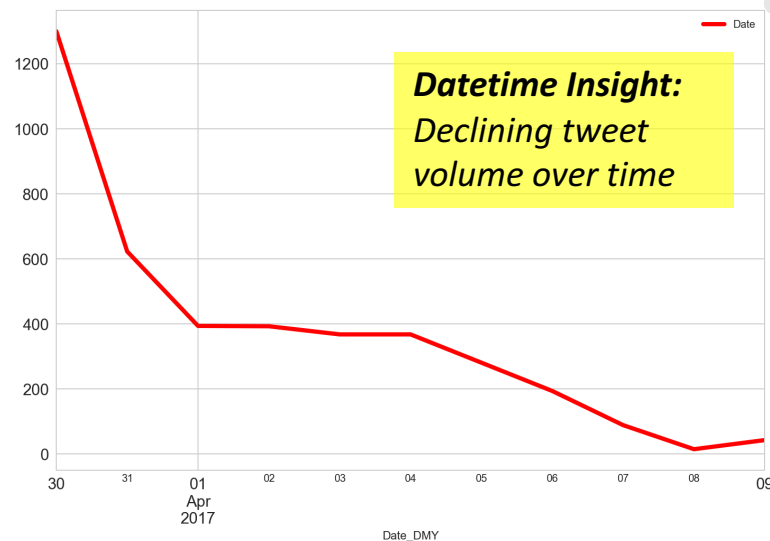


Data behind the Science



Historical Twitter data from periods of major natural catastrophes

- Two major hail events:
 - Brisbane 2014, Sydney 2015
- Cyclone Debbie data for preliminary exploratory data analysis
 - Preliminary EDA exhibits:



```
csv = './data/cyclonedebbie_30Mar.csv'
```

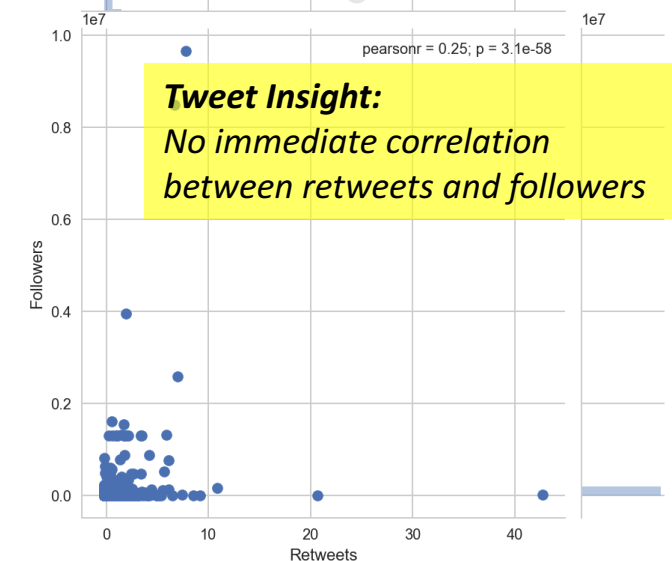
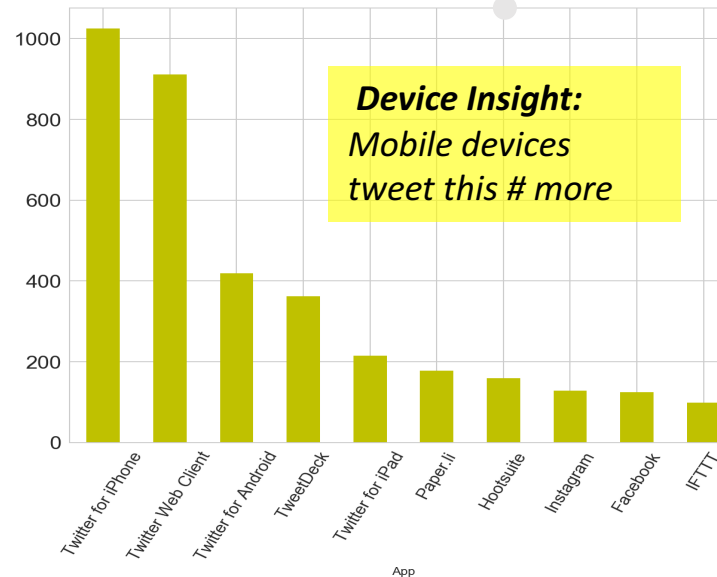
```
cy_deb = pd.read_csv(csv, encoding = "ISO-8859-1")  
cy_deb = pd.DataFrame(cy_deb)  
print (cy_deb.head(5))  
print (cy_deb.shape)
```

	Date	Screen Name	Full Name
0	3/30/2017 3:30:07	@Paul4Eva25	Symantha Martin
1	3/30/2017 3:30:17	@tan2tlc	Tania Lucey
2	3/30/2017 3:30:32	@SaxonTheHound	SaxonTheHound
3	3/30/2017 3:30:51	@lmars111	Luke Marshall
4	3/30/2017 3:30:53	@spellscribe	Amy Hopkins

Data.head()

	Tweet Text	Tweet ID \
0	Welcome To Sydney #CycloneDebbie https://...	8.473960e+17
1	MISSING HORSES Please share! #CycloneDebbie #U...	8.473960e+17
2	RT @ABCemergency: Ex-#CycloneDebbie has hit so...	8.473960e+17
3	Oh well. Luckily for me i haven't been majorly...	8.473960e+17
4	RT @sweirmint: Sharknado; this is how it start...	8.473960e+17

#cyclonedebbie



How will I measure success?

Success will be developing a predictive model that uses social data to indicate the severity of a natural catastrophe:

- Tweet **Language**
- Tweet **volume**
- Social posting **connectivity**



Framing the Hypothesis:

H0: Social data has **no indication** of the severity of a natural catastrophe

H1..n: Social data can **predict the severity** of a natural catastrophe.

The analysis will investigate multiple **machine learning methods** to develop an algorithm that uses Twitter data to indicate how severe a catastrophe is.

The targets of severity will specifically be: **insurance** (financial) **loss**, number of **claims** and **geographic impact**.