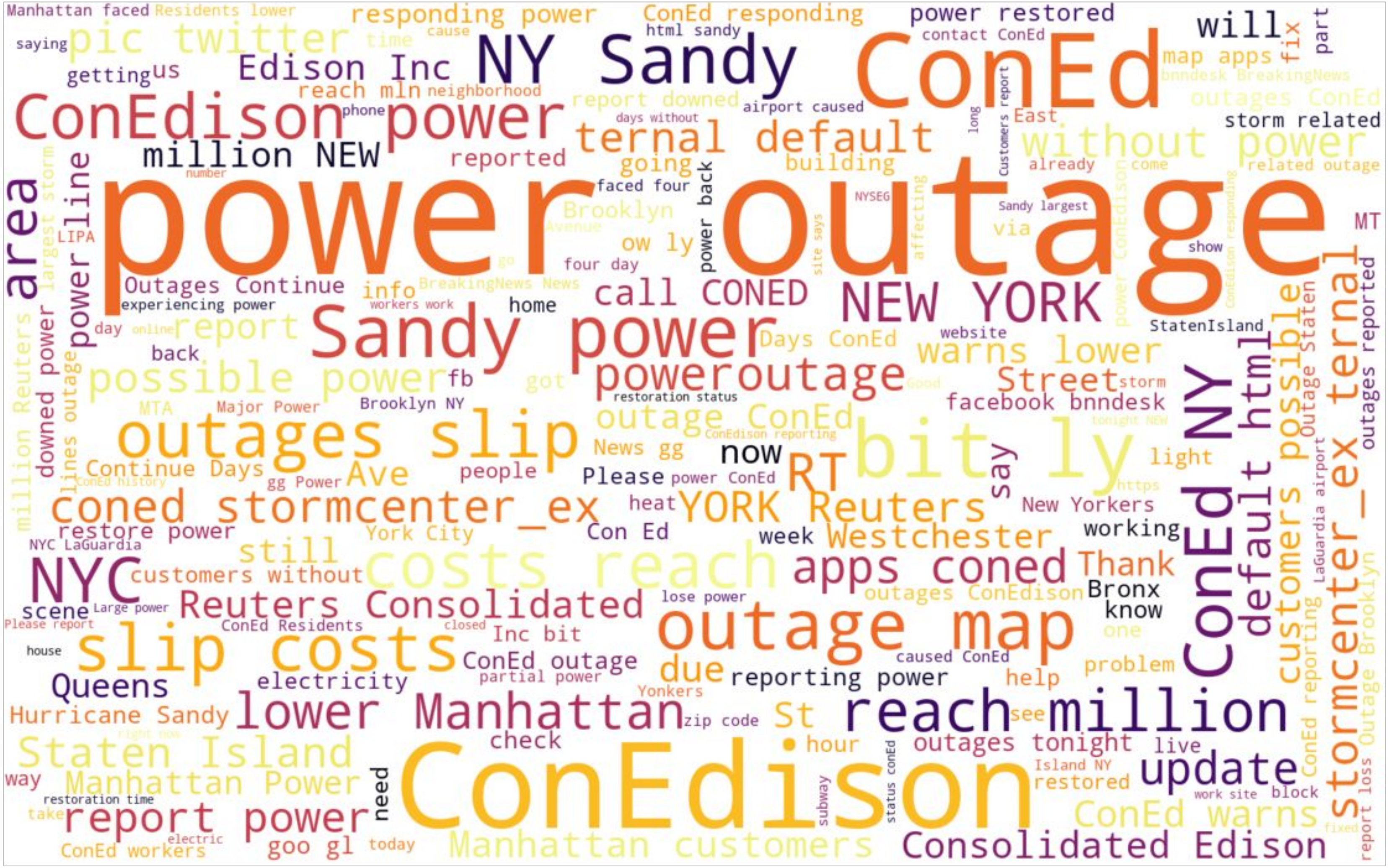
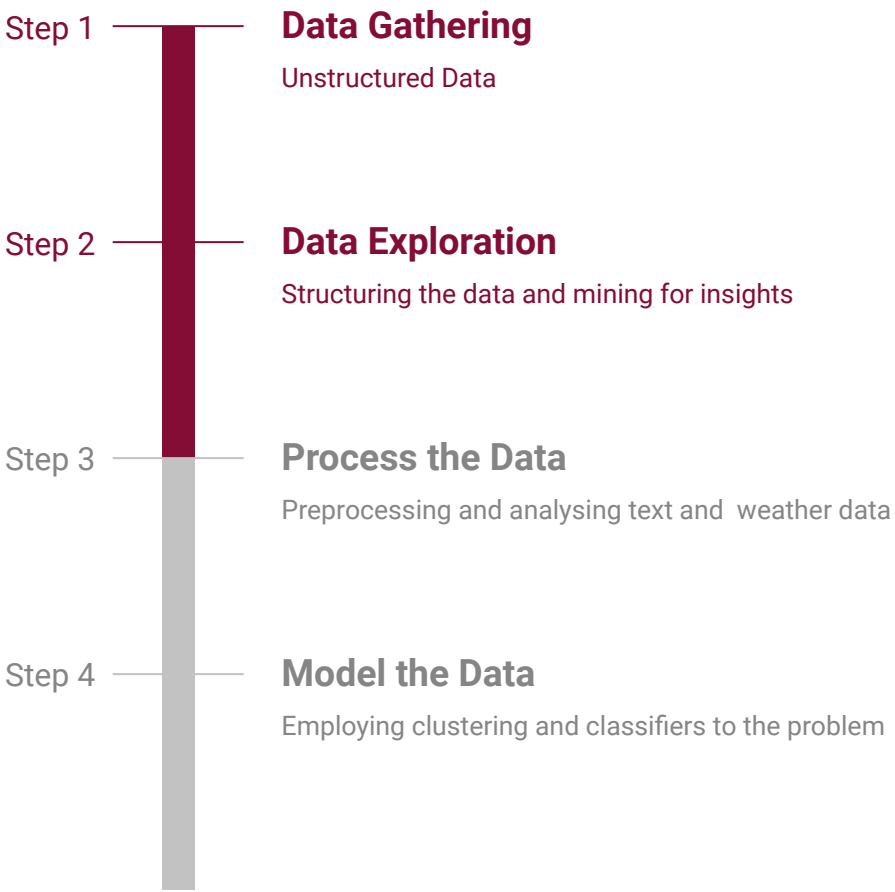
Tweet-Based Identification of Power Outages

Kevin Crystal, Ixchel Fragoso, John Ohakim


power outage
ConEdison
ConEd
NY Sandy
Sandy power
ConEdison power
outage map
bit ly
costs reach
outages slip
slip costs
reach million
ConEd NY
NYC
lower Manhattan
NEW YORK
poweroutage
coned stormcenter_ex
YORK Reuters
Reuters Consolidated
Consolidated Edison
apps coned
update
ternal default
without power
stormcenter_ex ternal
default html
customers possible
possible power
report power
Staten Island
Manhattan customers
Manhattan Power
ConEd warns
warns lower
Edison Inc
pic twitter
million NEW
call CONED
area
power line
report
RT
Queens
Hurricane Sandy
Westchester
Street
Thank
outage ConEd
responding power
ConEd responding
power restored
will
map apps
storm related
reported
going
building
Brooklyn
ow ly
four day
info
Outages Continue
experiencing power
home
Days ConEd
facebook bnndesk
now
Please
power ConEd
Ave
Continue Days
people
heat
Con Ed
week
working
New Yorkers
still
customers without
restore power
York City
Bronx
know
scene
ConEd Residents
Inc bit
ConEd outage
due
reporting power
electricity
help
problem
St
check
hour
outages tonight
live
restored
goo gl
today
ConEd workers
need
way
take
say
light
via
fb
got
back
day
getting
us
saying
reach mln
neighborhood
phone
report downed
airport caused
days without
faced four
power back
East
already
Sandy largest
show
MT
StatenIsland
website
restoration status
Brooklyn NY
Major Power
MTA
downed power
lines outage
million Reuters
largest storm
LIPA
NYC LaGuardia
Large power
Please report
house
closed
lose power
outages ConEdison
caused ConEd
partial power
Yonkers
zip code
see
Island NY
block
work site
restoration time
electric
subway
status coned
outages reported
LaGuardia airport
Outage Staten
ConEd reporting
Outage Brooklyn
report loss
part
fix
contact ConEd
html sandy
Manhattan faced
Residents lower
time
cause
outages ConEd
related outage
come
power ConEdison
affecting
NYSEG
site says
go
BreakingNews News
workers work
online
bnndesk BreakingNews
Customers report
long
News gg
gg Power
ConEd history
ConEdison reporting
tonight NEW
https
storm
Good
one
fixed
number
Avenue
right now

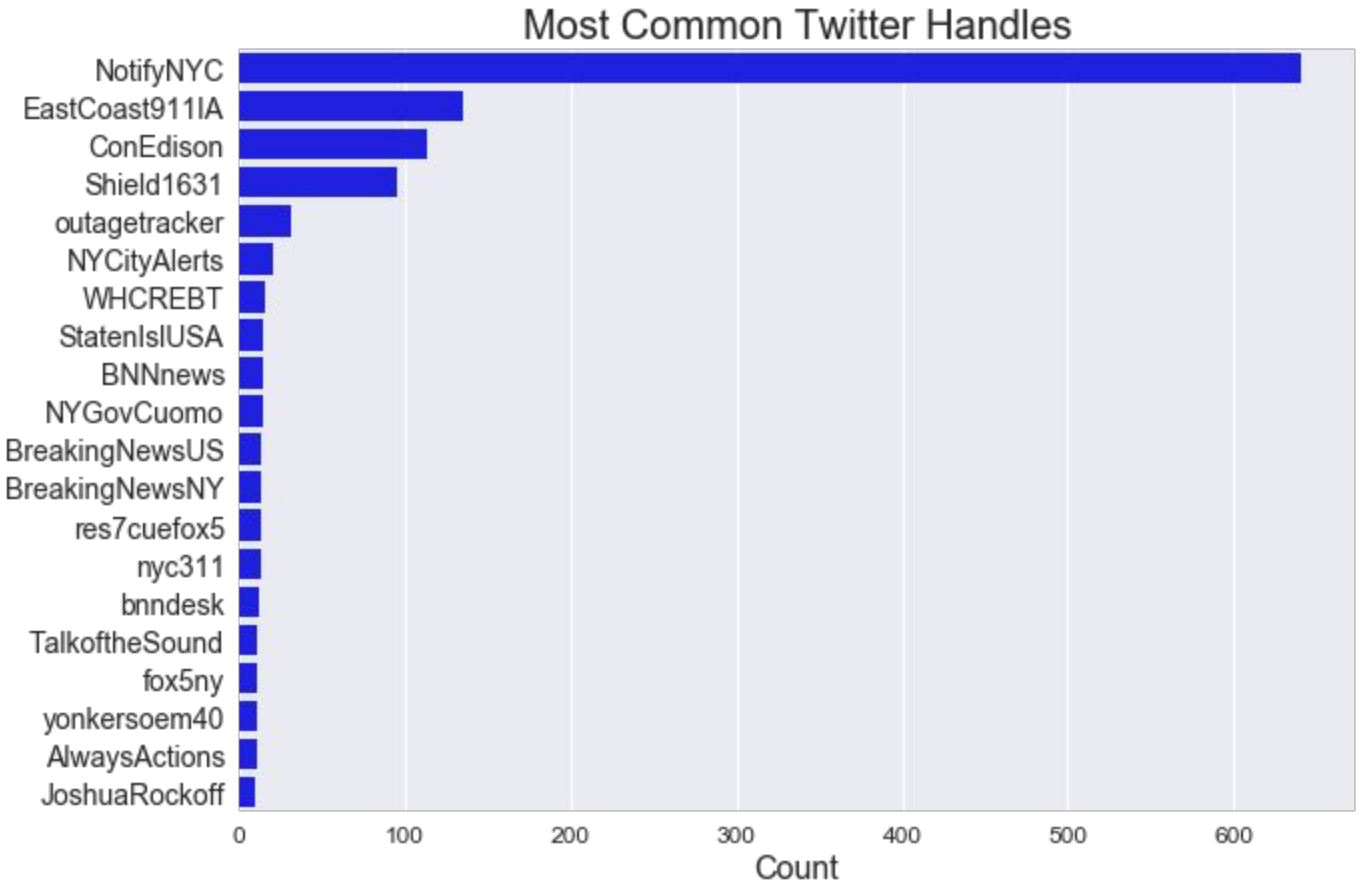
Workflow


Step 1
Data Gathering
Unstructured Data
Step 2
Data Exploration
Structuring the data and mining for insights
Step 3
Process the Data
Preprocessing and analysing text and weather data
Step 4
Model the Data
Employing clustering and classifiers to the problem

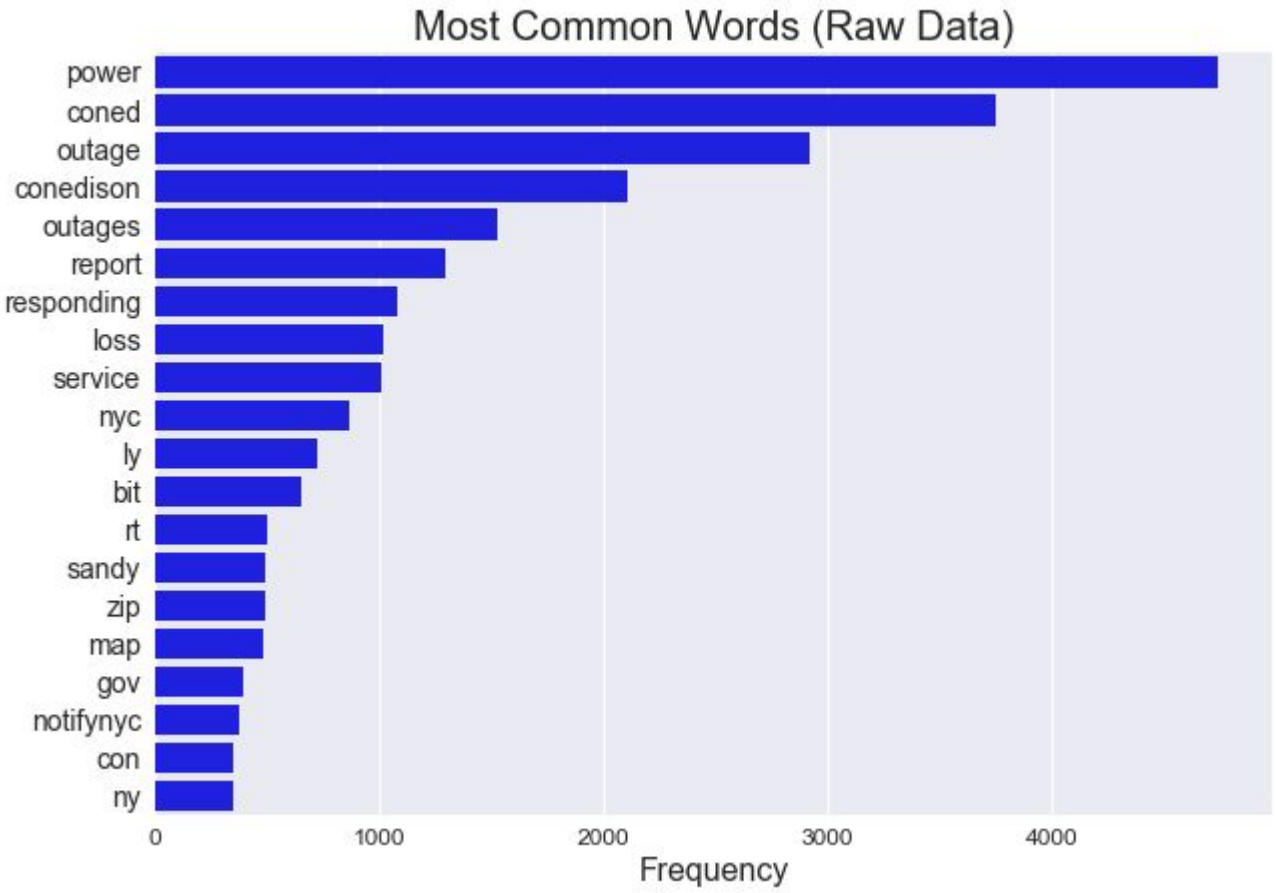
Data Gathering

Challenges			Successes
1 **Access**	✗	✓	**Method**
2 **Cost**	✗	✓	**Historical Data:** - Tweets - Weather Data
3 **Time**	✗	✓	**NYC Open Data**

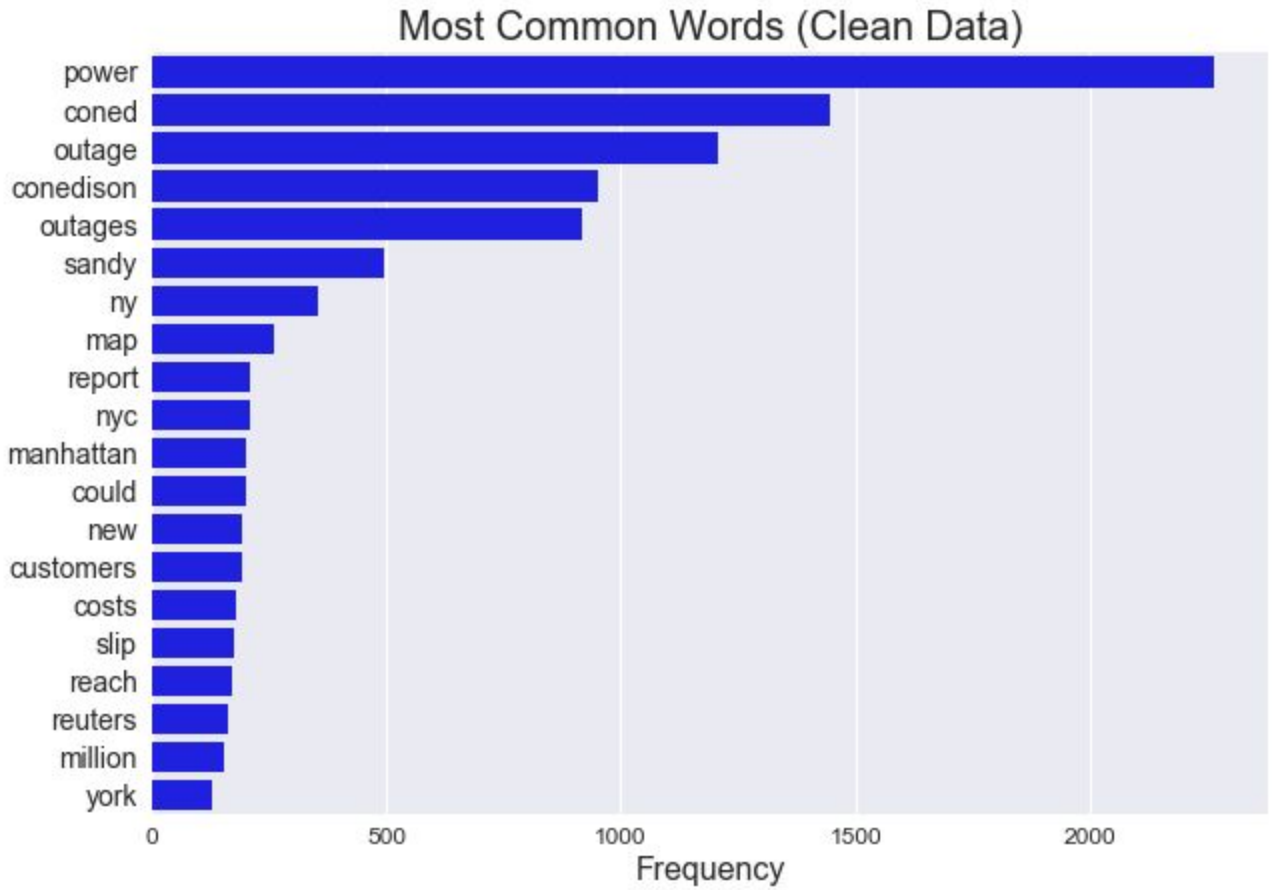
Data Exploration



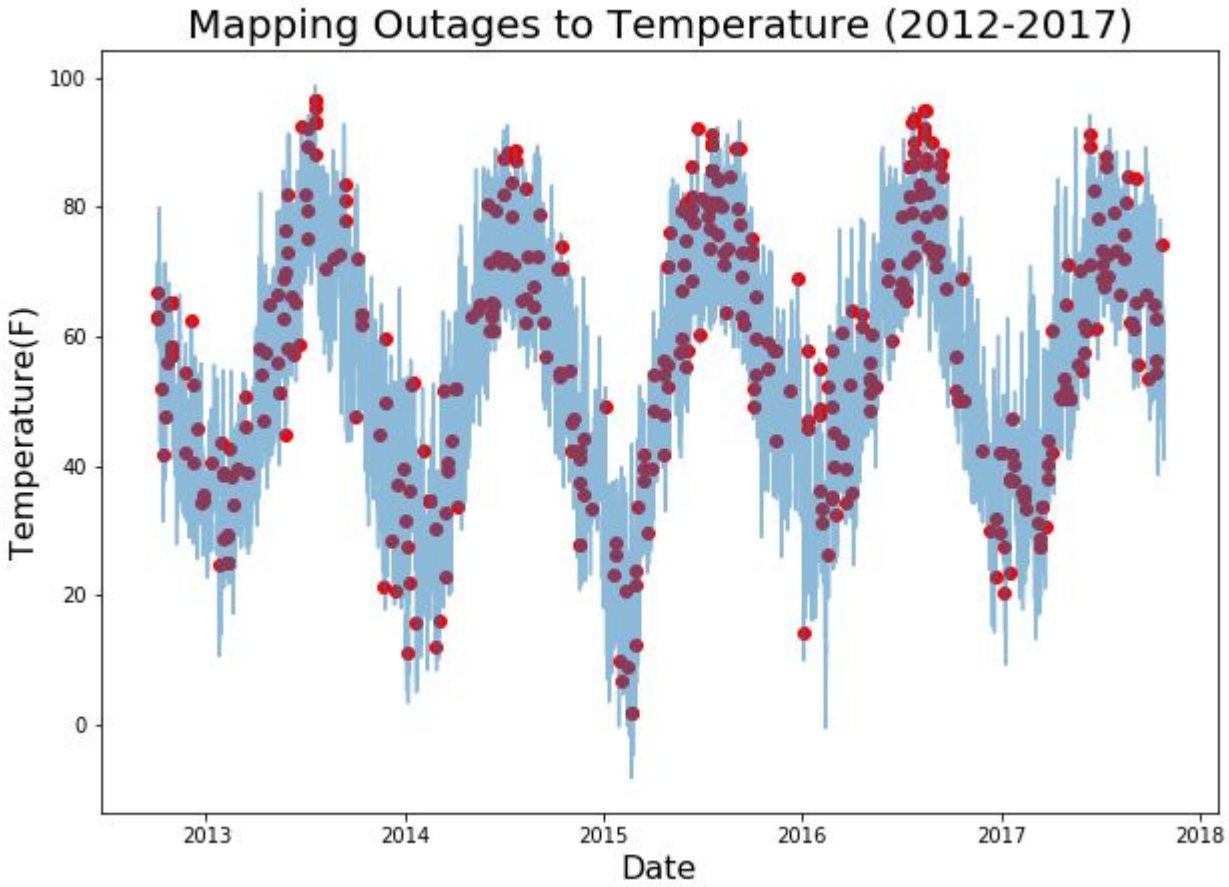
Data Exploration



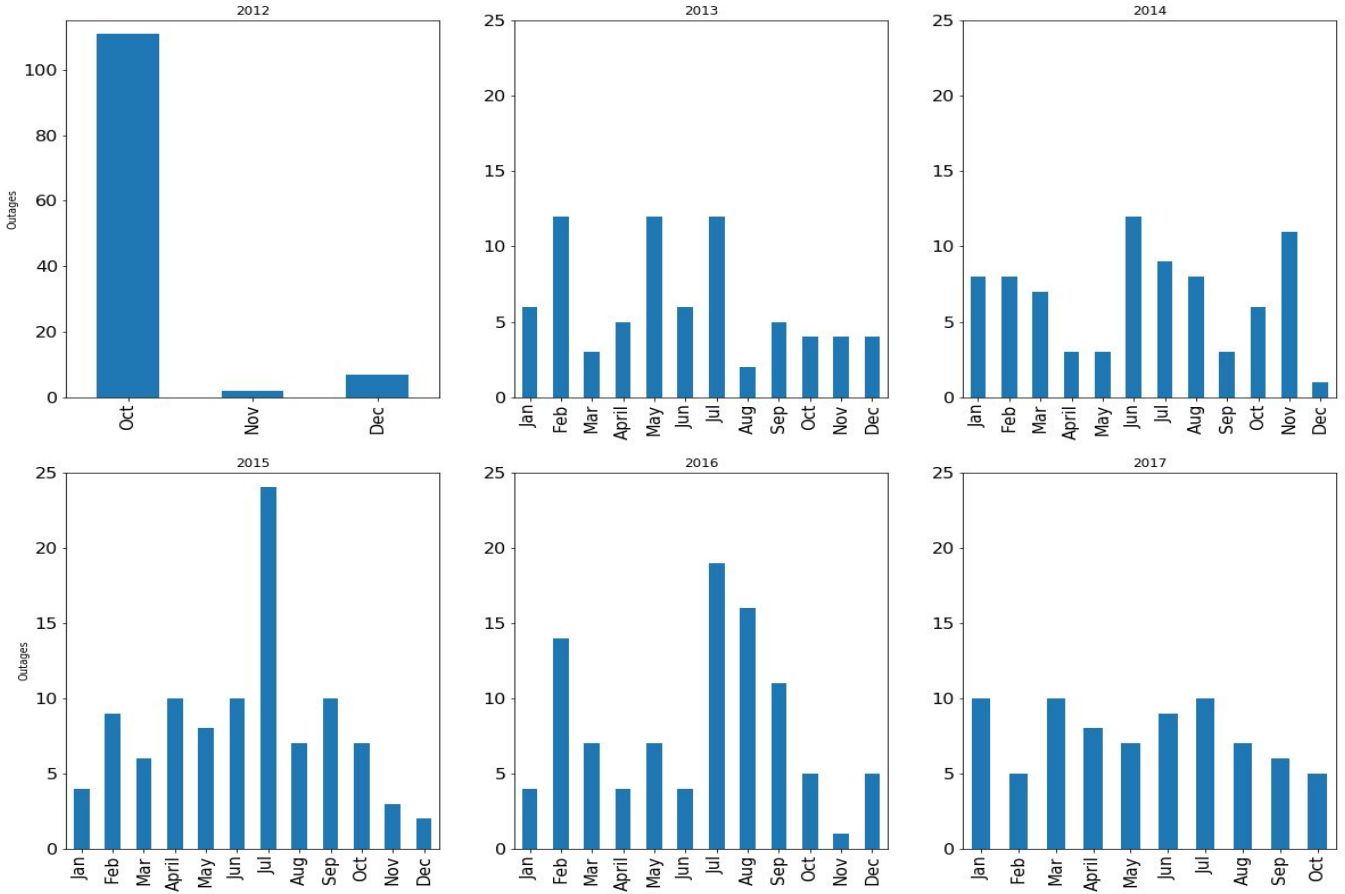
Data Exploration



Data Exploration



Number of Outages (2012-2017)



Clustering

- K-Means
- DBSCAN

- Any insights in the tweets?
- Any prominent patterns?

- 4 Clusters
- Very similar vocabulary
- Multiple clusters, depending on the data
- Over 2000 as noise

Classification

- Logistic Regression
- SVC

- Less than 1% positive class
- Class Weights
- Decision threshold

- Poor performance based on Recall
- Sharp tradeoffs
- Moderate tradeoffs
- Extensive search

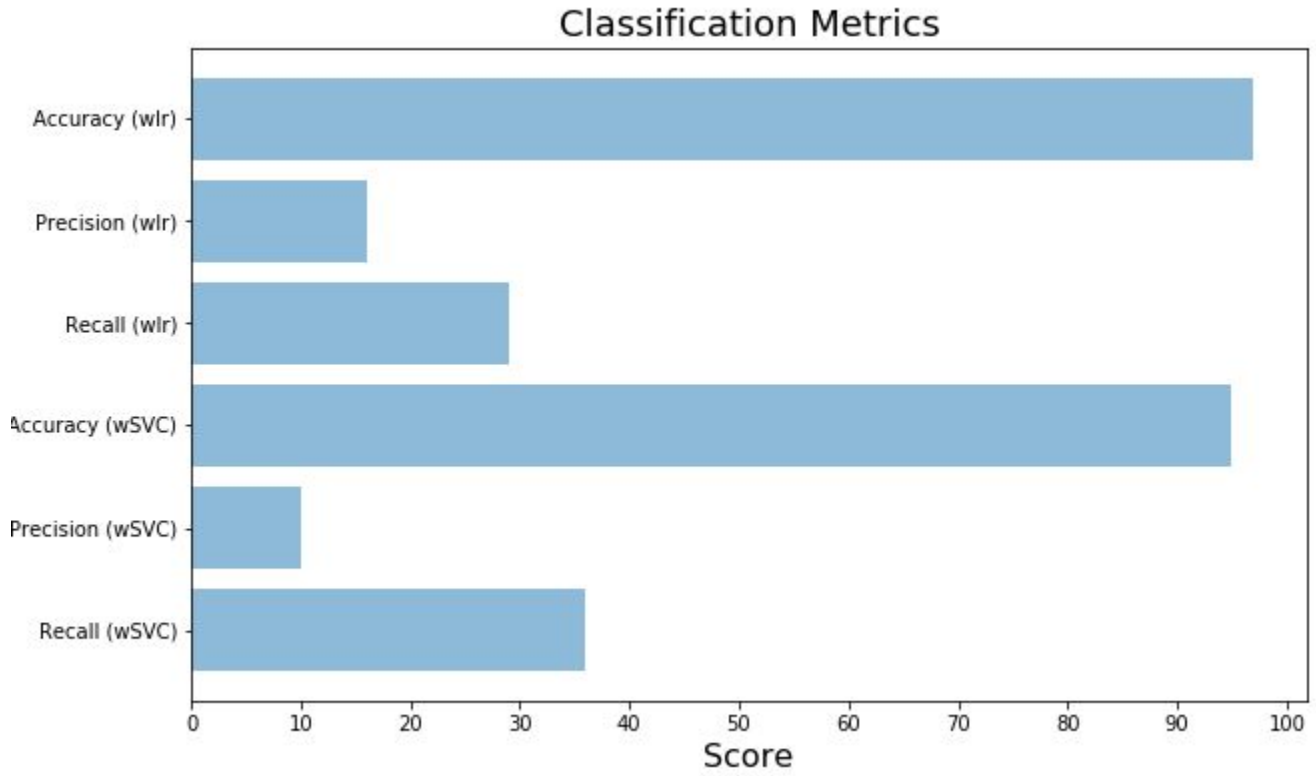
PCA

- Reduced Dimensions

- Weather a good predictor of outages
- Clumsy to employ all features

- No insights yet
- *Apriori*, no significant changes

Model Evaluation



Concluding Remarks

01	**Model**	• Employ other classification algorithms • Predict using weather data • Map outages to locations
02	**Evaluation**	• Identify other evaluation metrics • Imbalanced class problem
03	**Deployment**	• Capture new tweets • Model that predicts in real time