

# Stormy Weather

Predicting Social Media Behavior  
Based on Weather Patterns

What do these long-standing hypotheses have in common?

What do these long-standing hypotheses have in common?

The Earth is Flat

Eating sugar causes a “sugar-rush”

Bulls get angry when they see red

What do these long-standing hypotheses have in common?

**REJECTED**  
The Earth is Flat

Eating sugar causes a “sugar-rush”

Bulls get angry when they see red

What do these long-standing hypotheses have in common?



The Earth is Flat

Eating sugar causes a “sugar-rush”

Bulls get angry when they see red

What do these long-standing hypotheses have in common?

**REJECTED**  
The Earth is Flat

**REJECTED**  
Eating sugar causes a “sugar-rush”

**REJECTED**  
Bulls get angry when they see red

What do these long-standing hypotheses have in common?

**REJECTED**  
The Earth is Flat

**REJECTED**  
Eating sugar causes a “sugar-rush”

**REJECTED**  
Bulls get angry when they see red

Better weather makes people happier (on twitter)

What do these long-standing hypotheses have in common?

The Earth is Flat

Eating sugar causes a “sugar-rush”

Bulls get angry when they see red

Better weather makes people happier (on twitter)

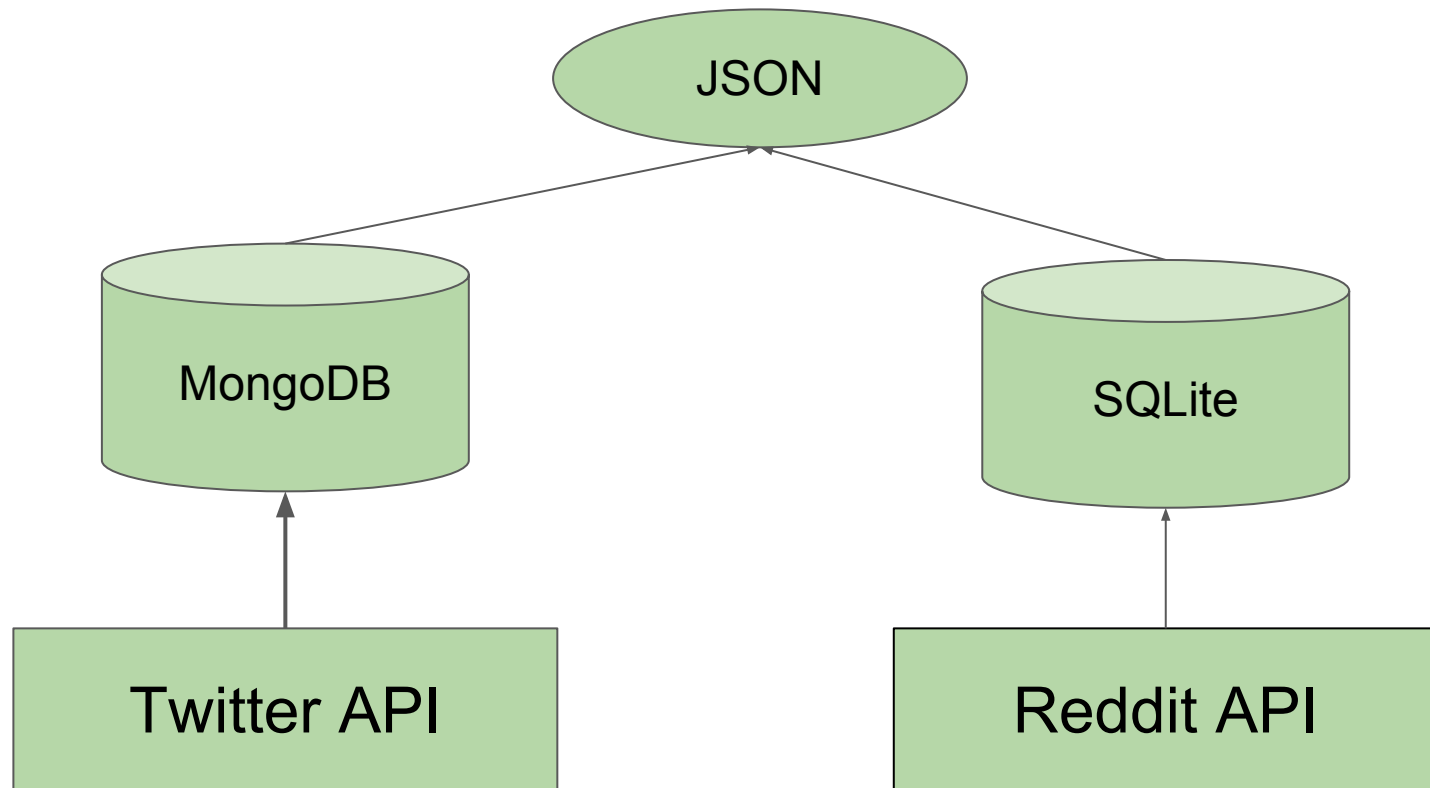


# Problem Overview

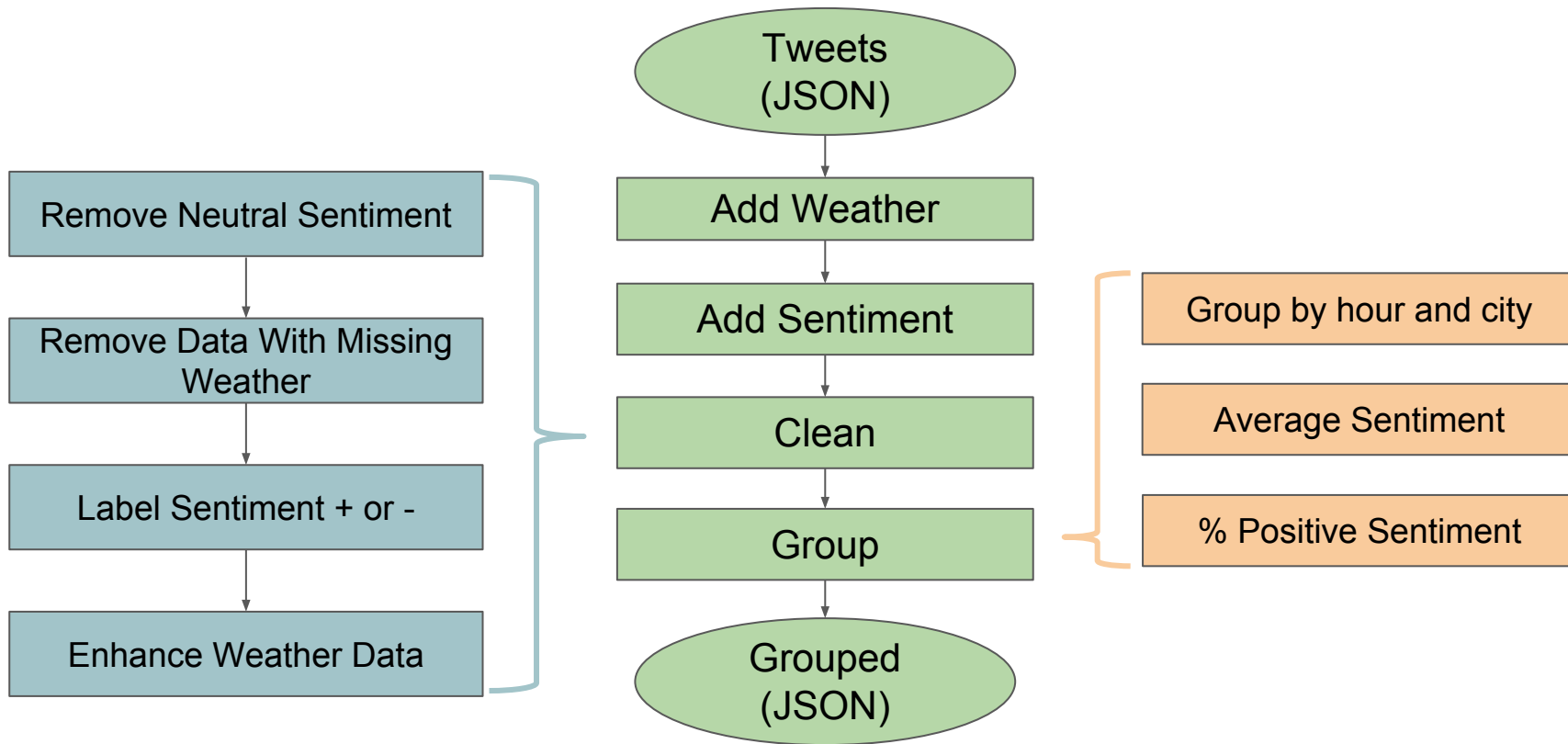
- **Is social media sentiment influenced by weather patterns?**
- Can we measure the impact of weather on the mood of a city by observing sentiment on social media

~~DATA COLLECTION AND ANALYSIS~~

# Data Collection Overview



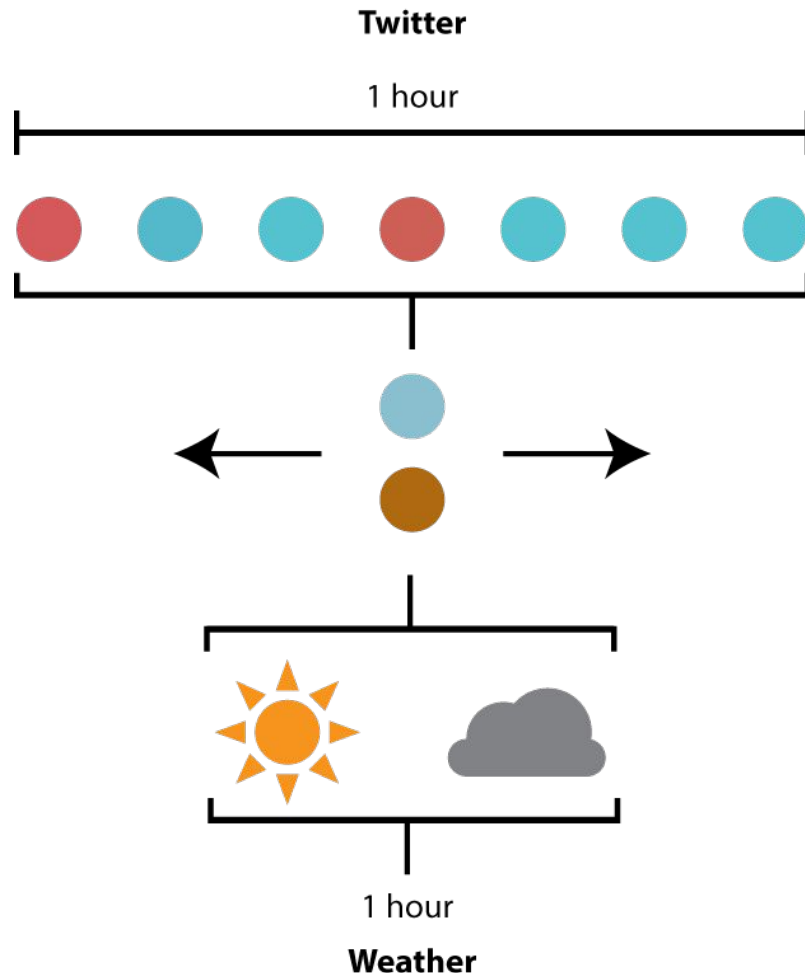
# Data Augmentation Overview



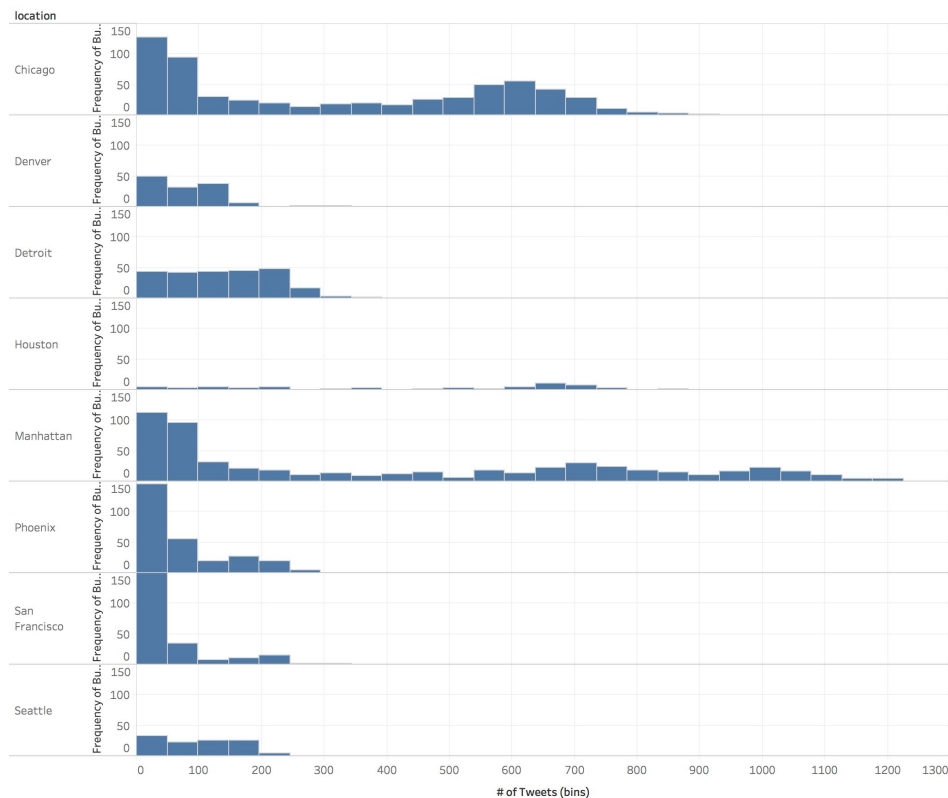
# Data Cleaning

## Training Data Set Creation:

1. Collect 37 days of tweets
2. Add weather
3. Add sentiment
4. Clean
  - a. Non-neutral sentiment
  - b. Weather data available
  - c. Augment weather data
5. Group by hour
6. Average sentiment by hour
7. Enrich tweets with weather data

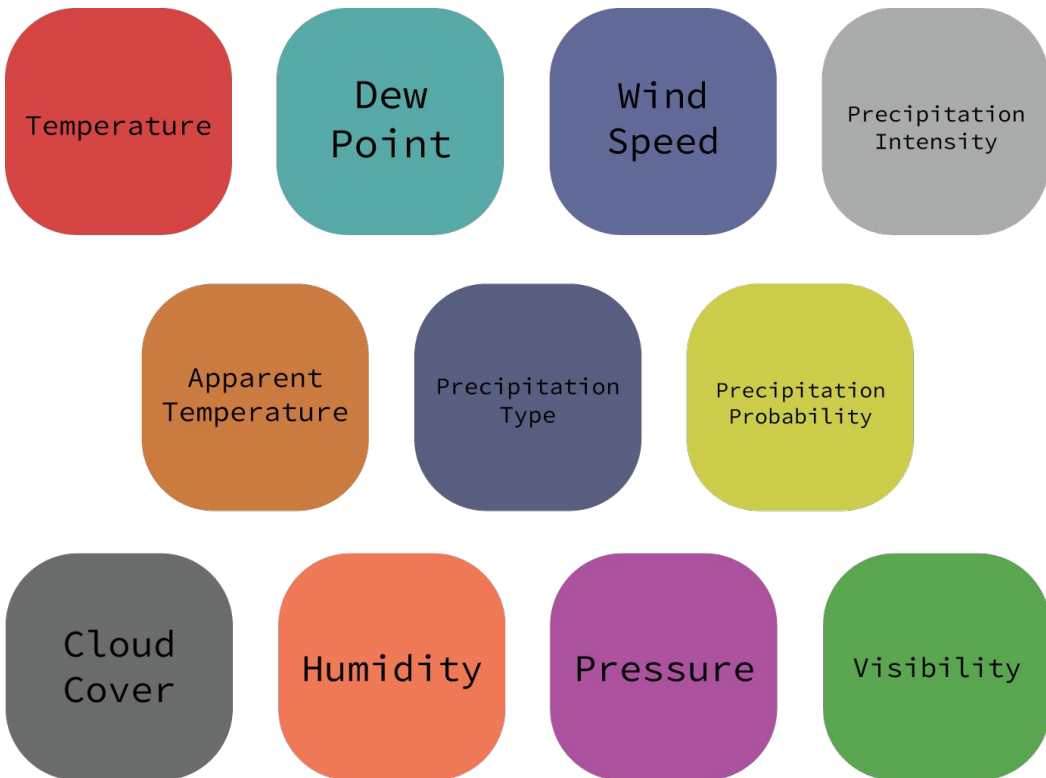


# Data Grouping



- We aggregated our data into a single point for each hour for each city
- In aggregating the data, we calculated the average sentiment and % positive sentiment
- In order for these statistics to be valid, we need enough tweet in each hour for the average to be close to the actual average
- If we had more data we may have excluded any data points with less than ~20 tweets

# Look at all the weather parameters!



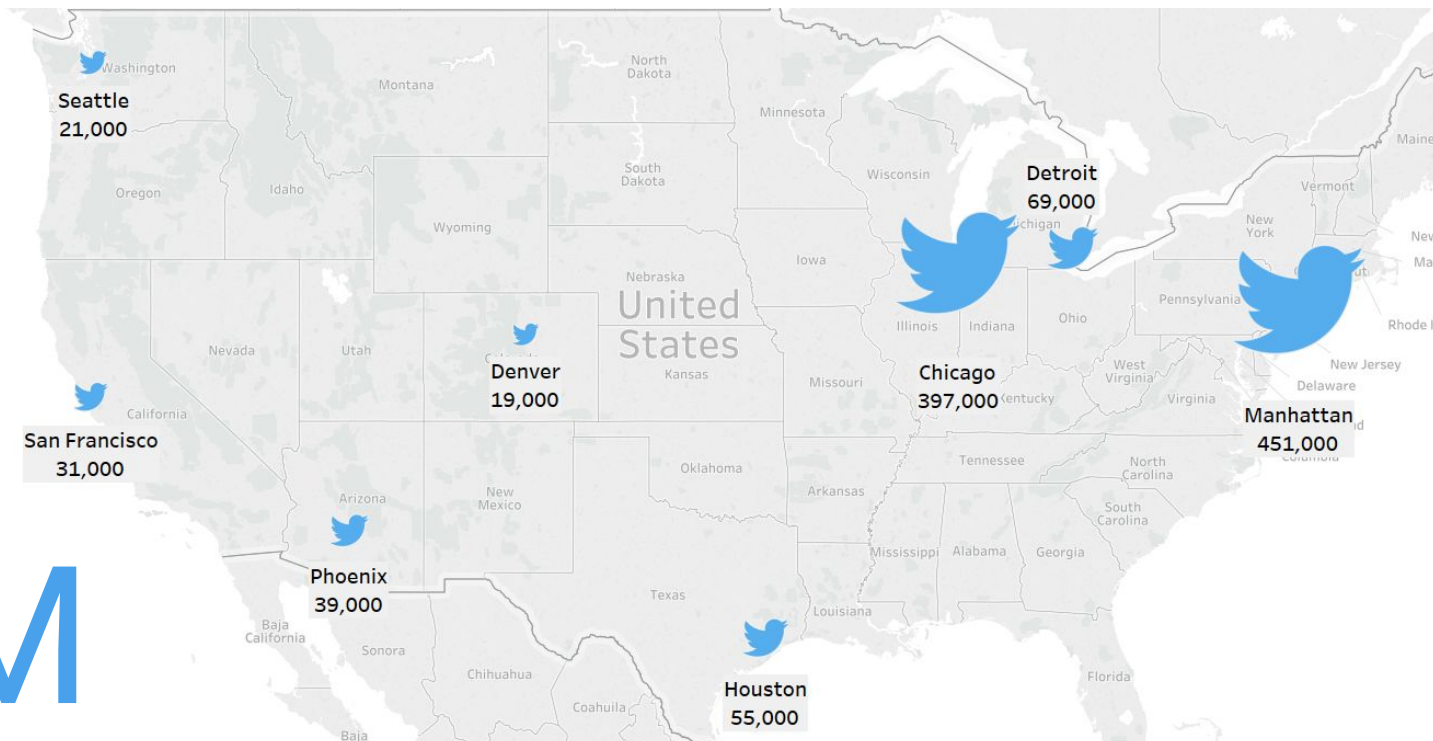
# Twitter Sentiment Analysis

8

Cities

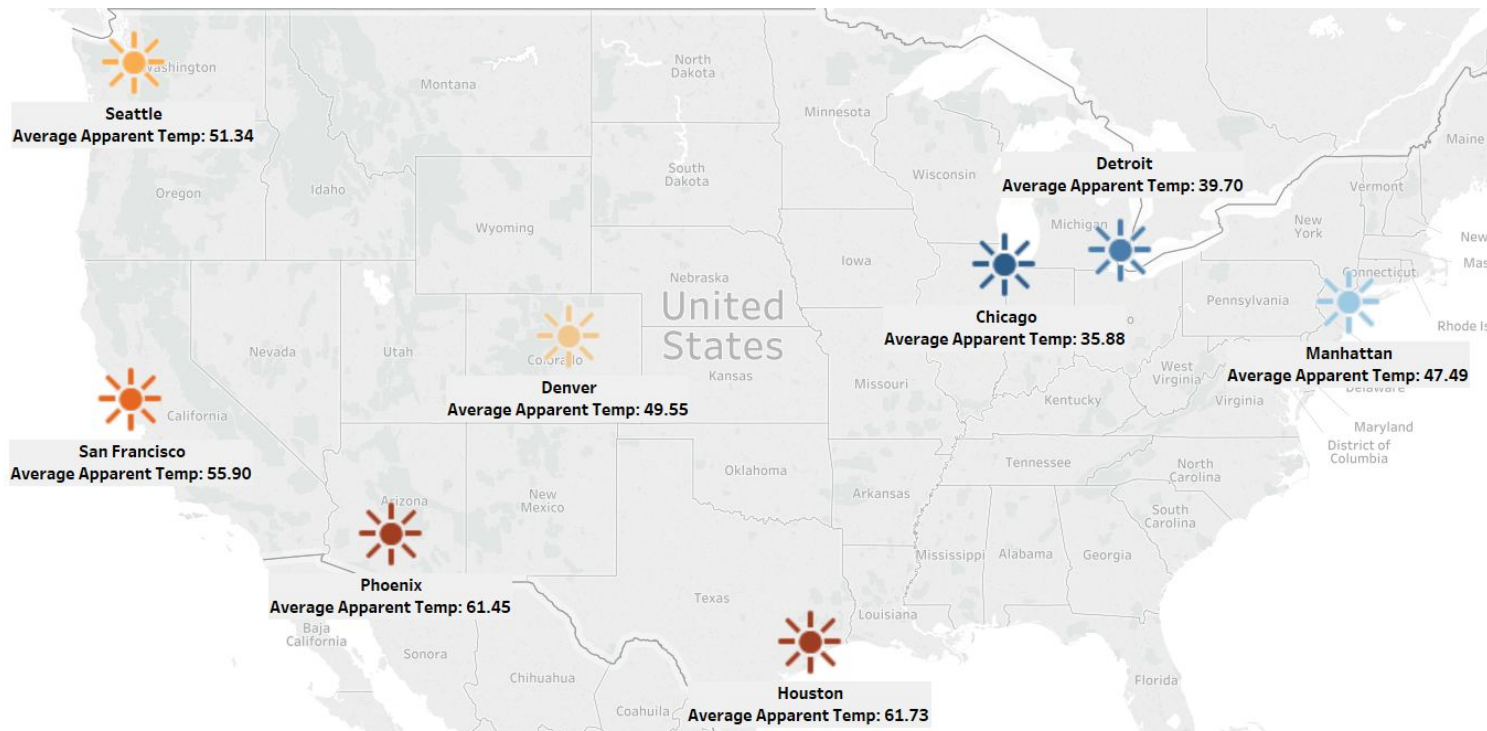
1.1M

Tweets



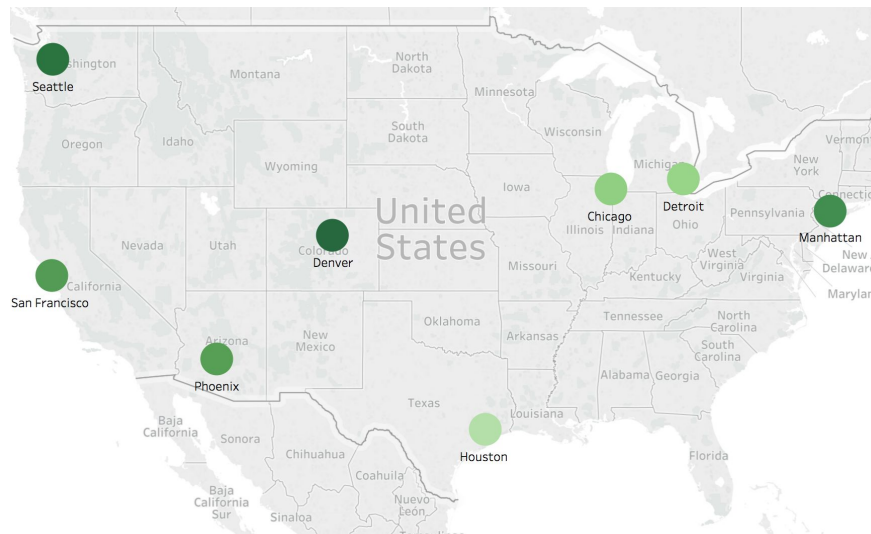


# Twitter Cities' Average Temperature

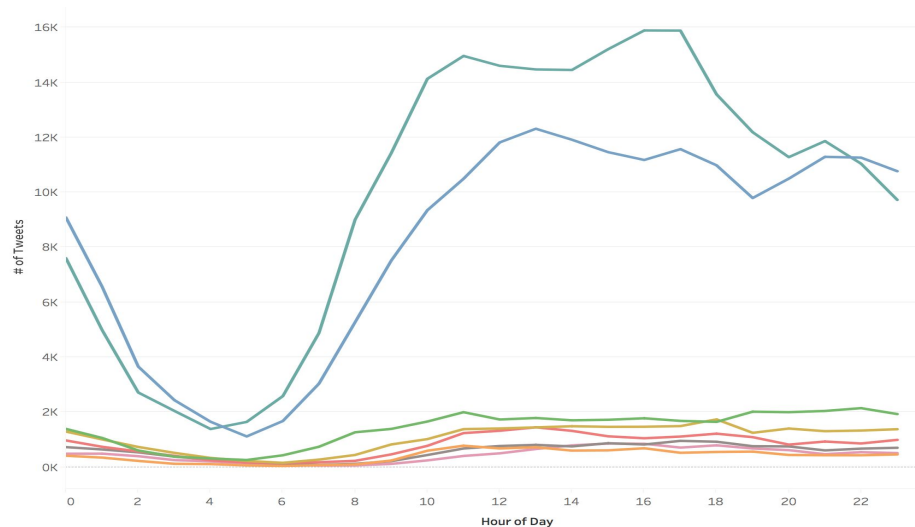


# Twitter Data Insights

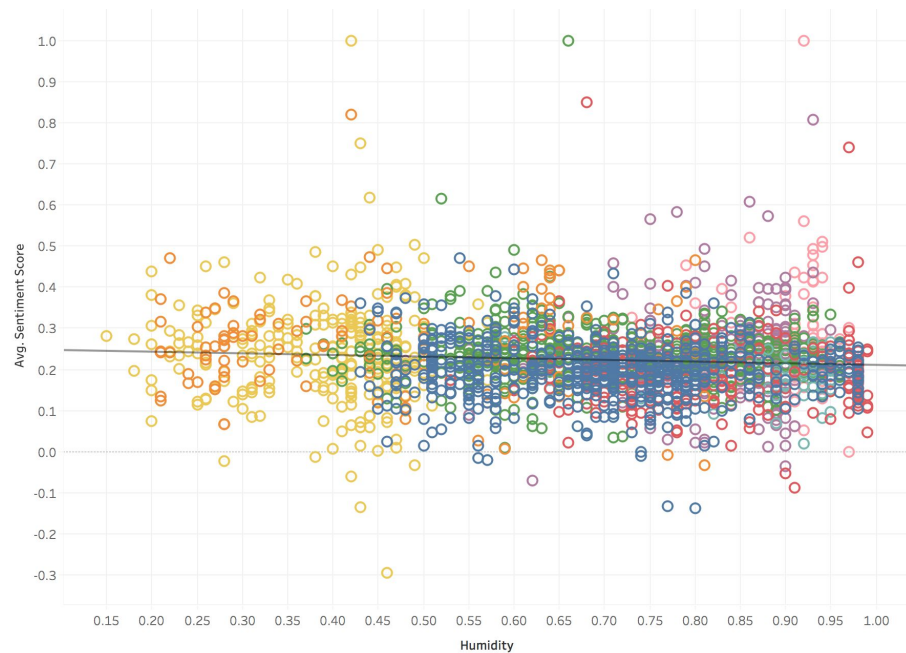
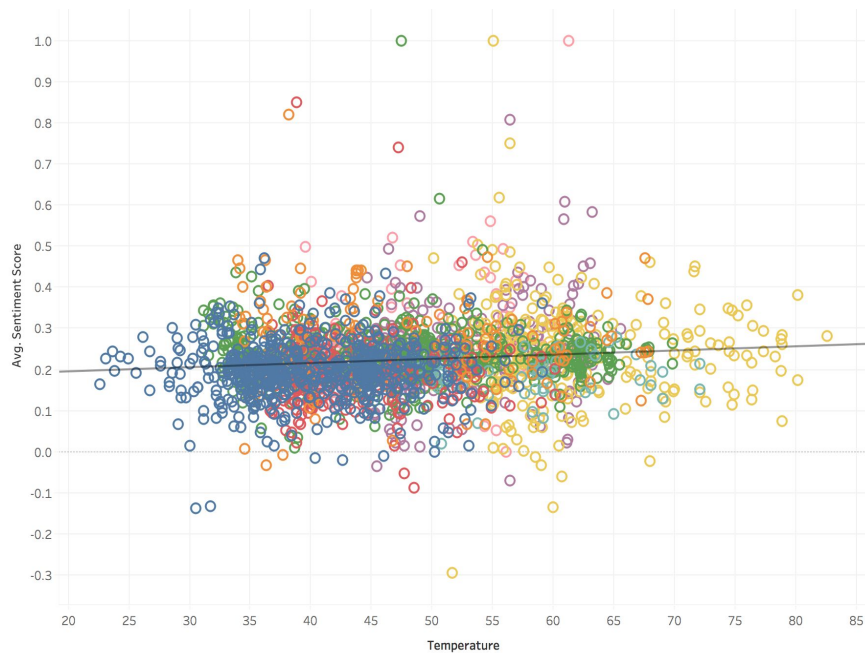
## Average Sentiment across America



## Twitter Usage during an Average American Day



# Twitter Correlation Analysis - Average Sentiment



location

Chicago

Denver

Detroit

Houston

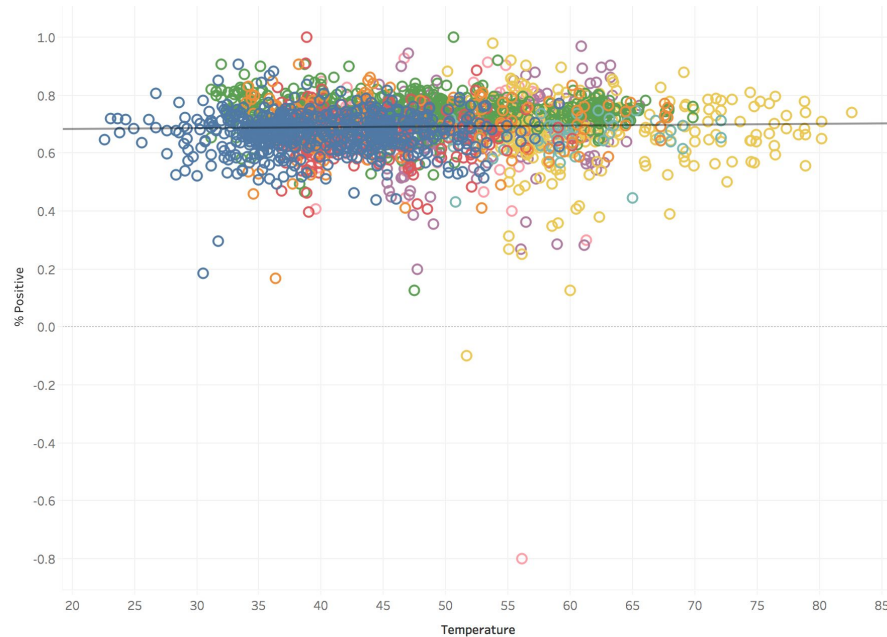
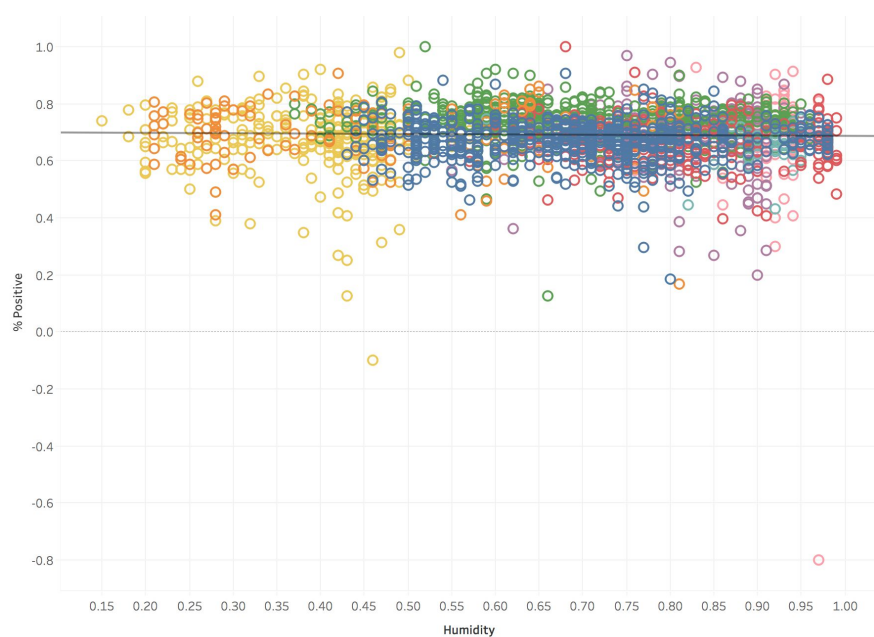
Manhattan

Phoenix

San Francisco

Seattle

# Twitter Correlation Analysis - % Positive



location

Chicago

Denver

Detroit

Houston

Manhattan

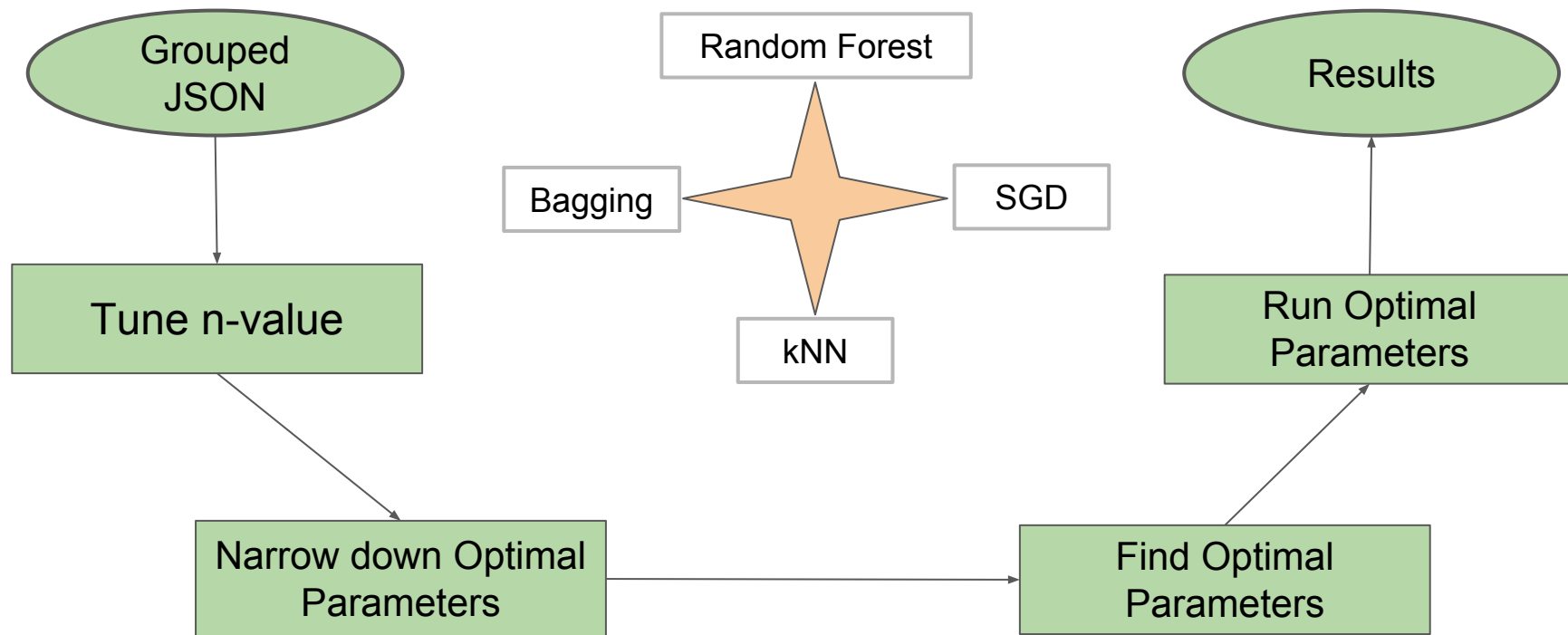
Phoenix

San Francisco

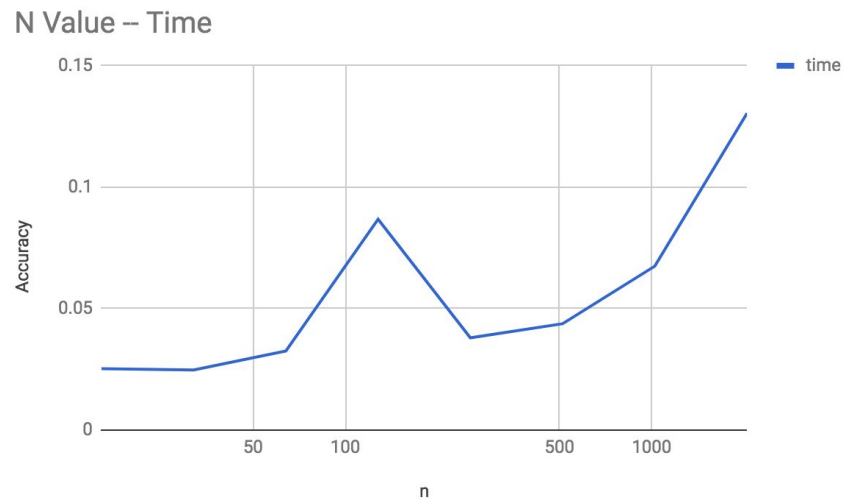
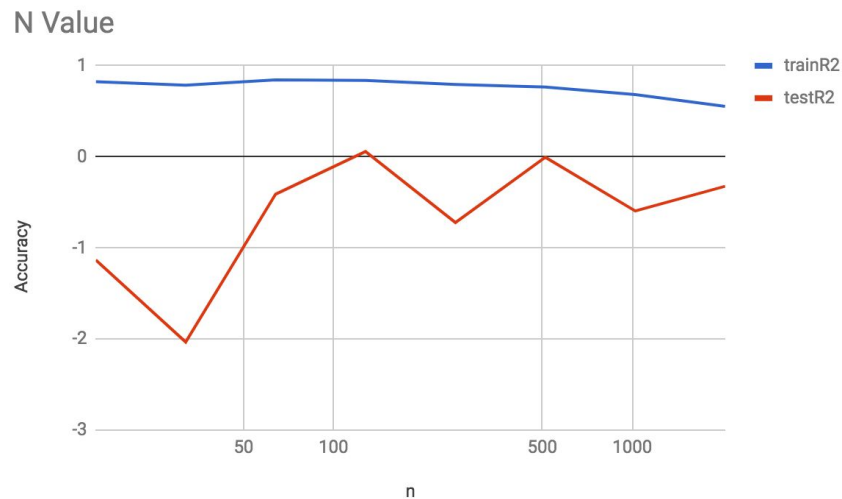
Seattle

MACHINE LEARNING

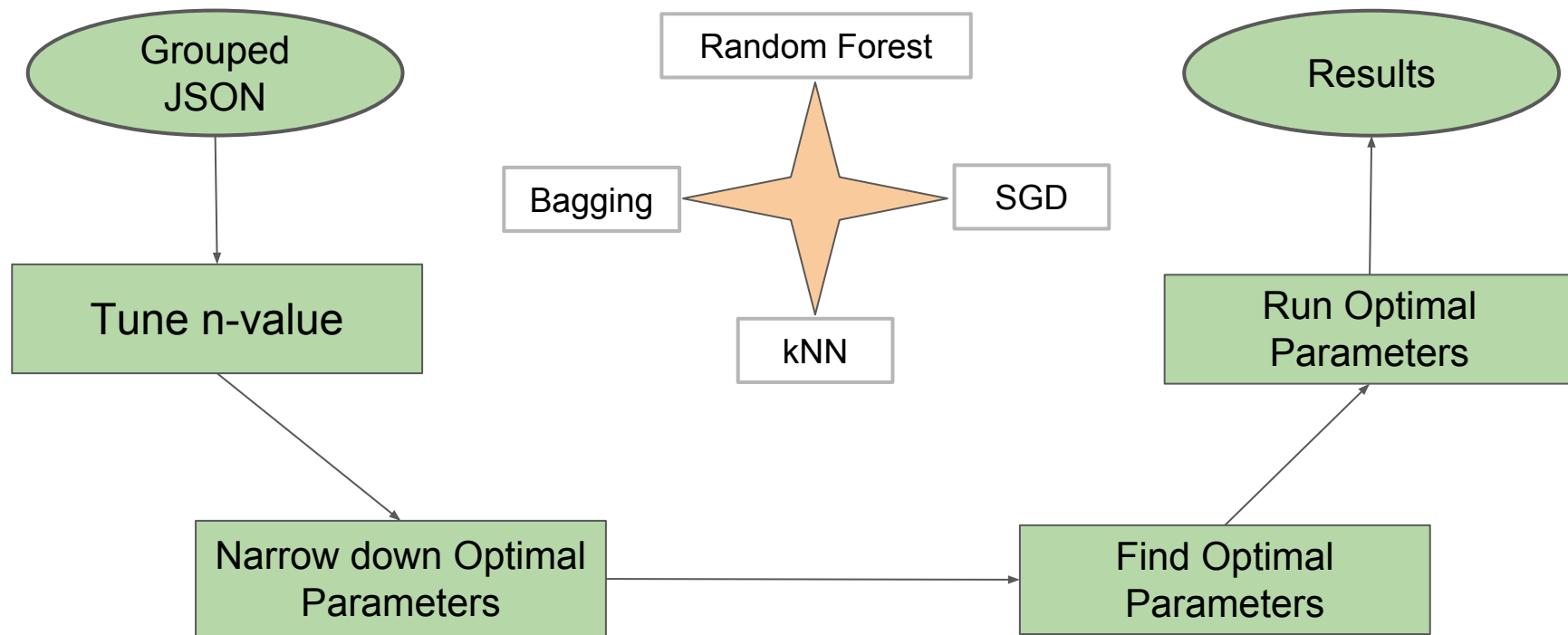
# Machine Learning Overview



# Twitter N-value Optimization



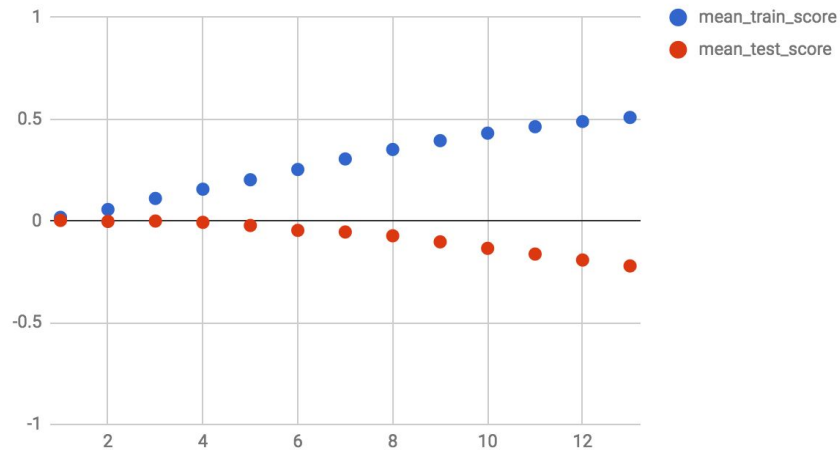
# Machine Learning Overview



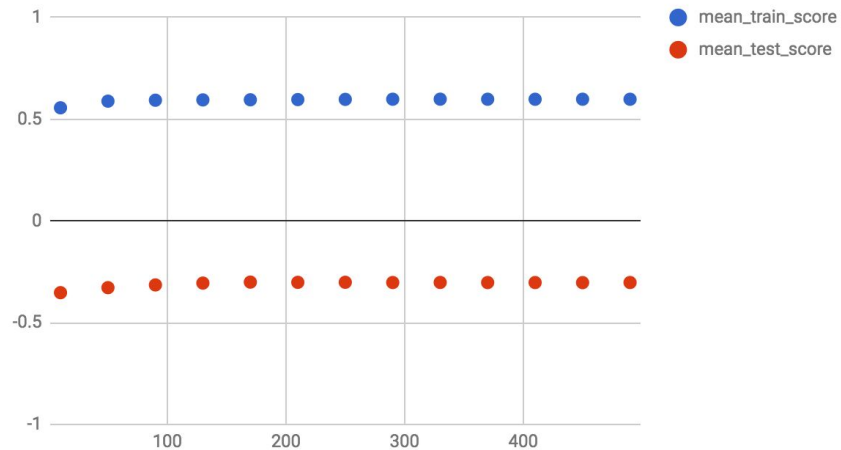


# Twitter Parameter Optimization

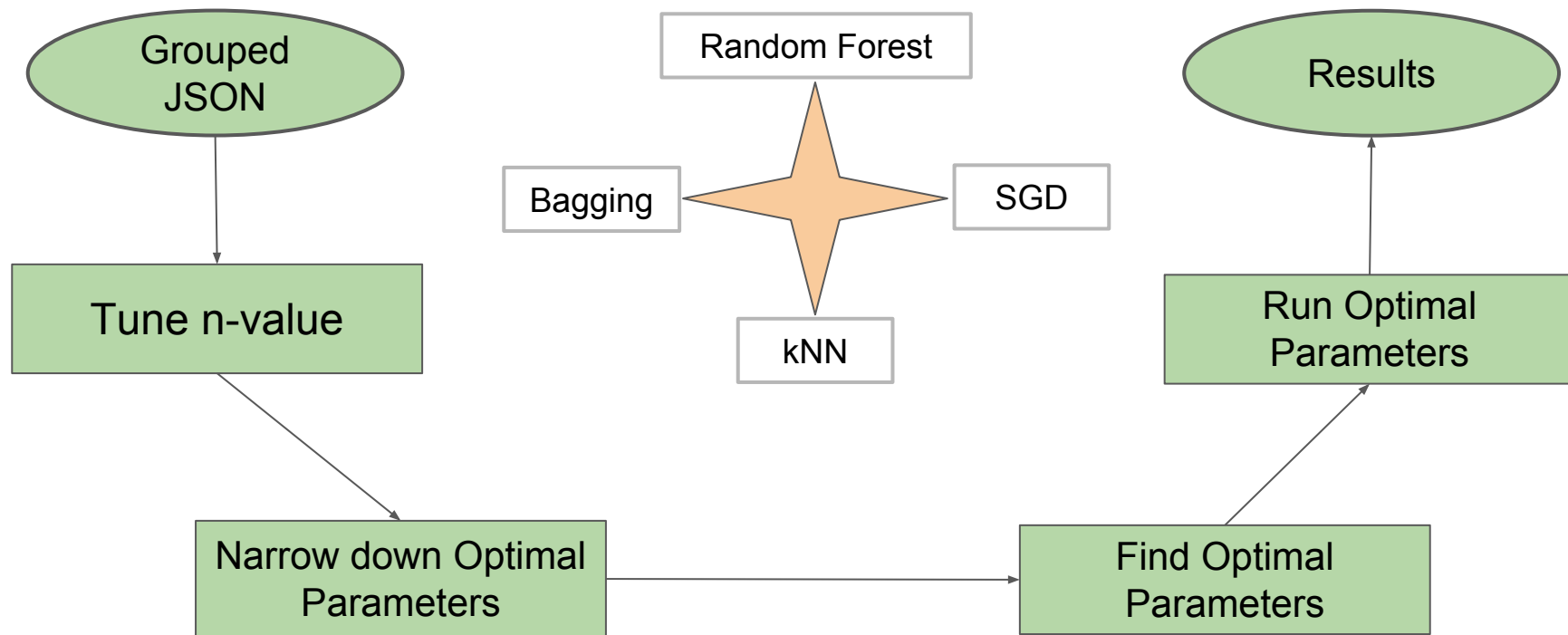
Max Depth (Random Forest)



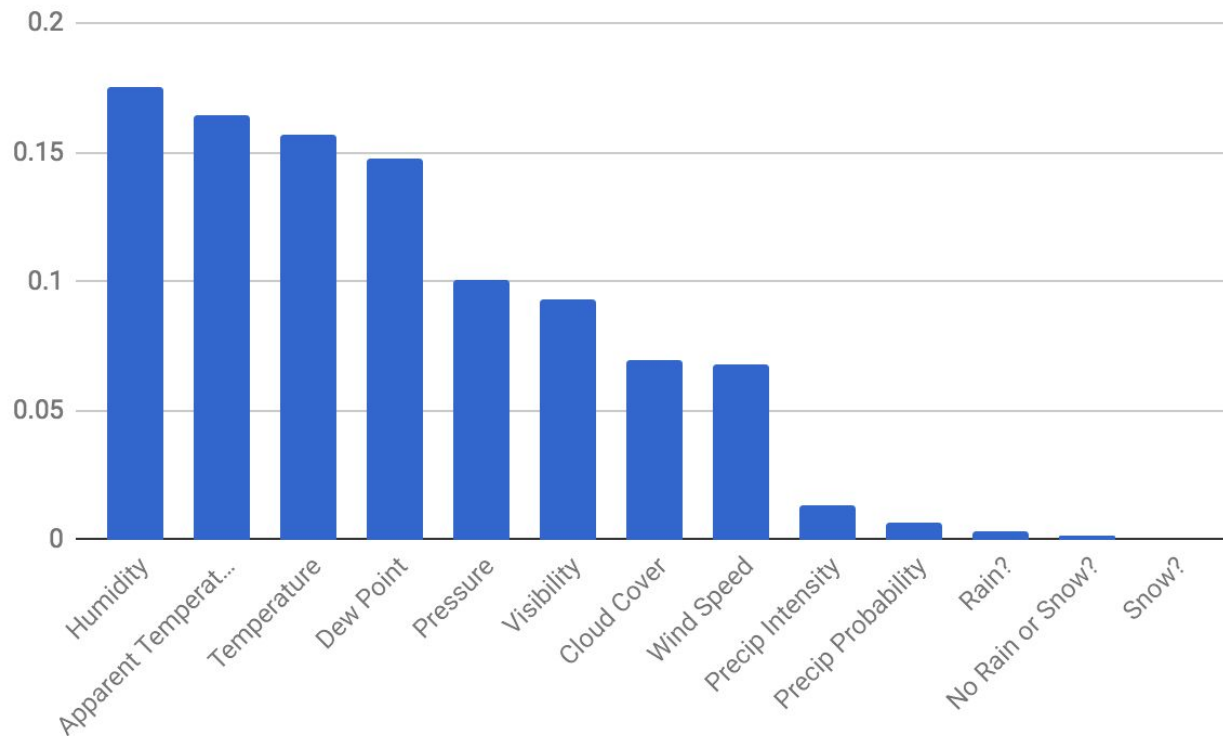
N Estimators (Random Forest)



# Machine Learning Overview



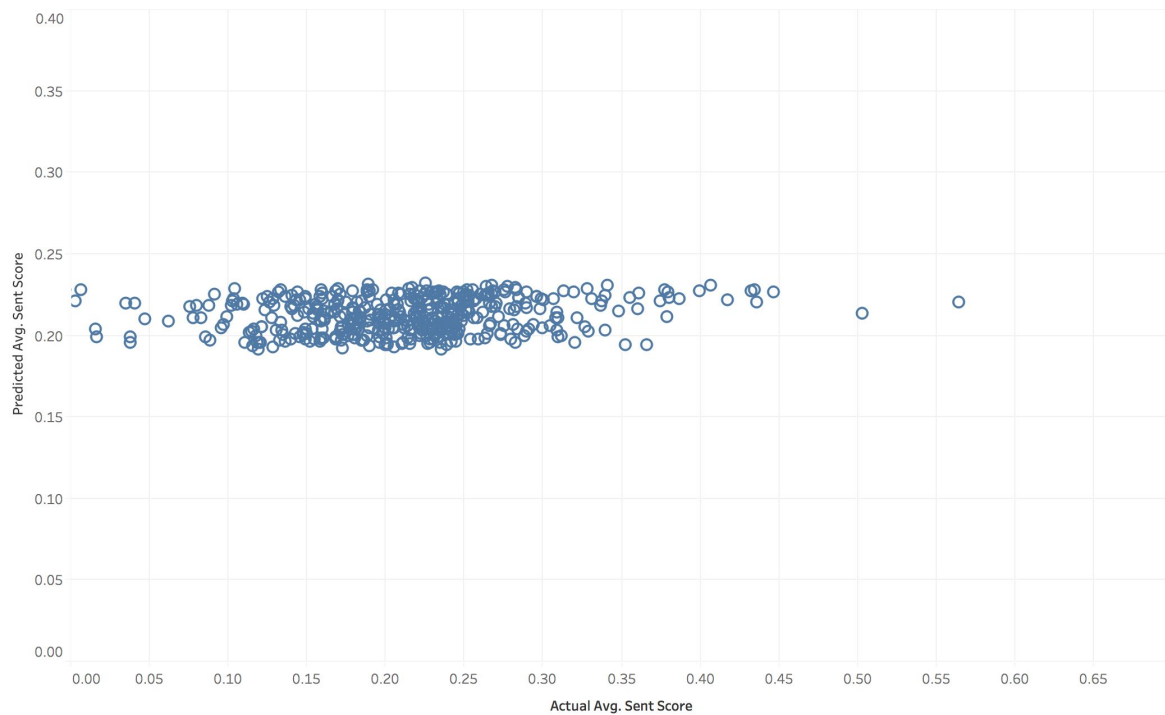
# Twitter Feature Importance (Random Forest)



# Twitter Machine Learning Results

Algorithm	R2 Train	R2 Test
RandomForest	0.0424	0.0203
K-Neighbors	0.0419	0.0322
Stochastic Gradient	-0.1220	-0.1619
Bagging	0.1000	0.0042

# Twitter Machine Learning Results



OTHER CONSIDERATIONS

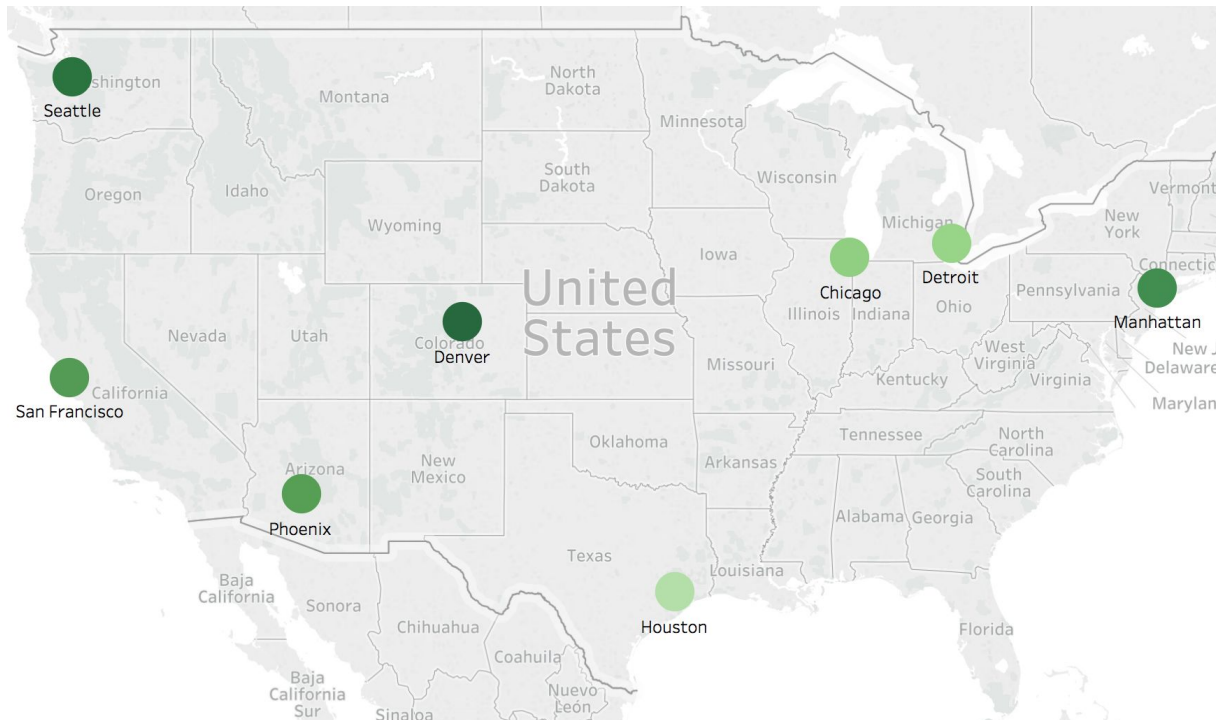
# Reddit Sentiment Analysis Extension

8

Cities

3.4M

Comments



## Reddit Comments

$$r^2 = .002$$





## Twitter % Positive Sentiment

$$r^2 = .03$$



## Conclusion

$$H_0 : r^2 > 0$$

$$H_A : r^2 \leq 0$$

We **reject** the null hypothesis that we can predict sentiment using weather forecasts.

## Conclusion

$$H_0 : r^2 > 0$$

$$H_A : r^2 \leq 0$$

We **reject** the null hypothesis that we can predict sentiment using weather forecasts.

**DENIED**

A man in a blue raincoat and hat is running through the rain, holding an umbrella. He has a joyful expression on his face, with his mouth open as if shouting or laughing. The background is a brick wall and a doorway. The text "Happiness in the rain?" is overlaid in large, bold, yellow letters.

**Happiness  
in the  
rain?**



Happiness

in the

rain?

APPROVED

# Reflection and Next Steps

## Reconsiderations

- More data (Year-round Twitter data)
- Explore “change in weather” rather than absolute weather
- Filter tweets for specific topics (i.e. tweets about weather)
- Other than weather, what else could impact sentiment?

## Controlled Experiment

- Have a group of people tweet 5 times a day at pre-set times
- Clean data!

THE END