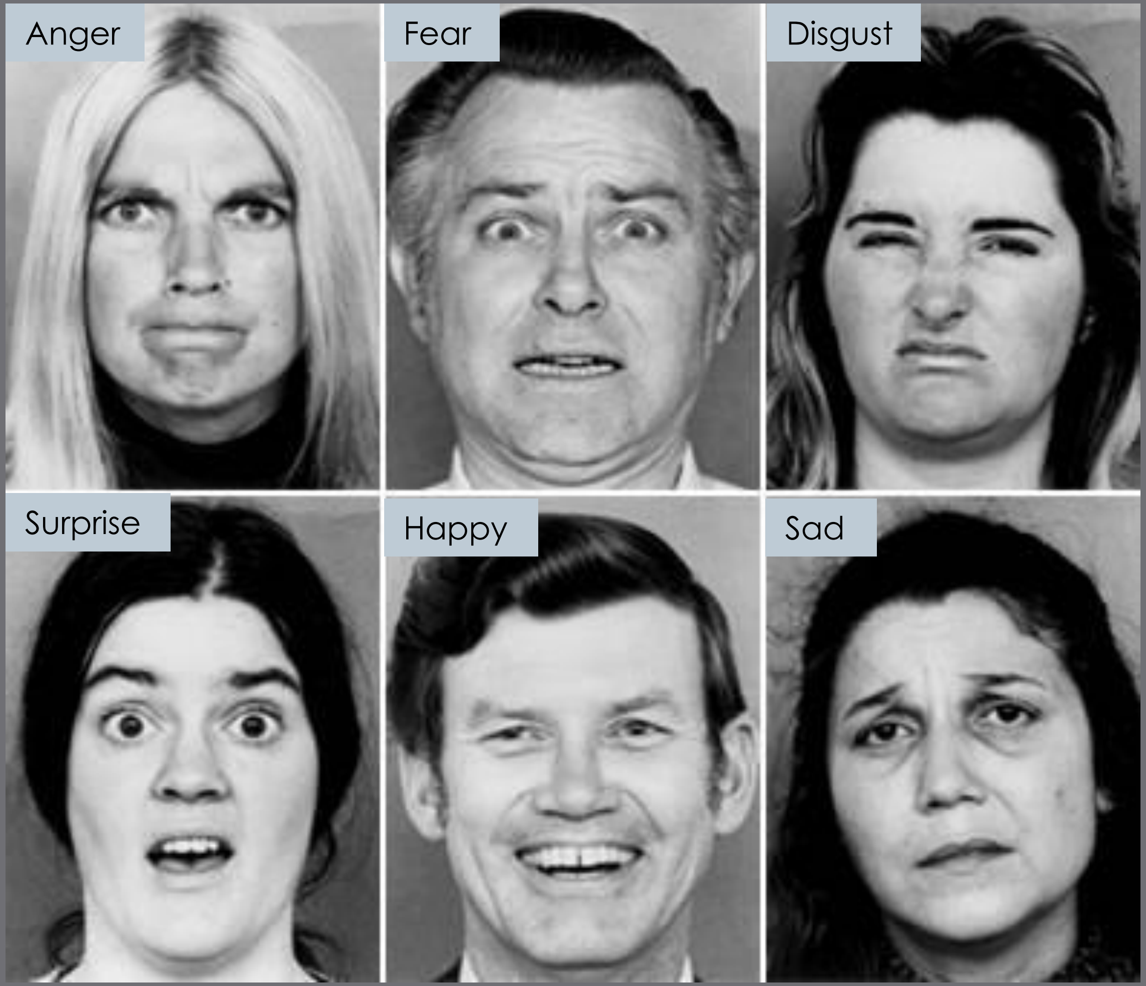
Assessment of Citizen's '**Affect**' from Tweets in different Urban settings

Comparing ATLANTA and BOSTON

By Florina Dutt (PhD Candidate), Dr. Subhrajit Guhathakurta
Georgia Institute of Technology

What do we mean by “affect”?



The outward appearance of feeling and emotions

- Widely used in **psychology**

It can be expressed in :

- A tone
- A facial expression
- A textual expression
-

In this case we are assessing affect of **Twitter microblogs**

Why is assessment of 'affect' important in urban context?

- Cities are associated with **higher rates of most mental health problems**

- 40% higher risk of mood disorder,
- 20% more anxiety issues

(Peen et al., 2010)

- Urban Planners can encourage **designing urban spaces conducive to people's mental wellbeing**

Spaces that are favorable?

Its has been studied that certain **urban spaces, are responsible for mental wellbeing** more than others.

Green Spaces

Roe, J. (2016)

Active places (accessible and public transportation friendly)

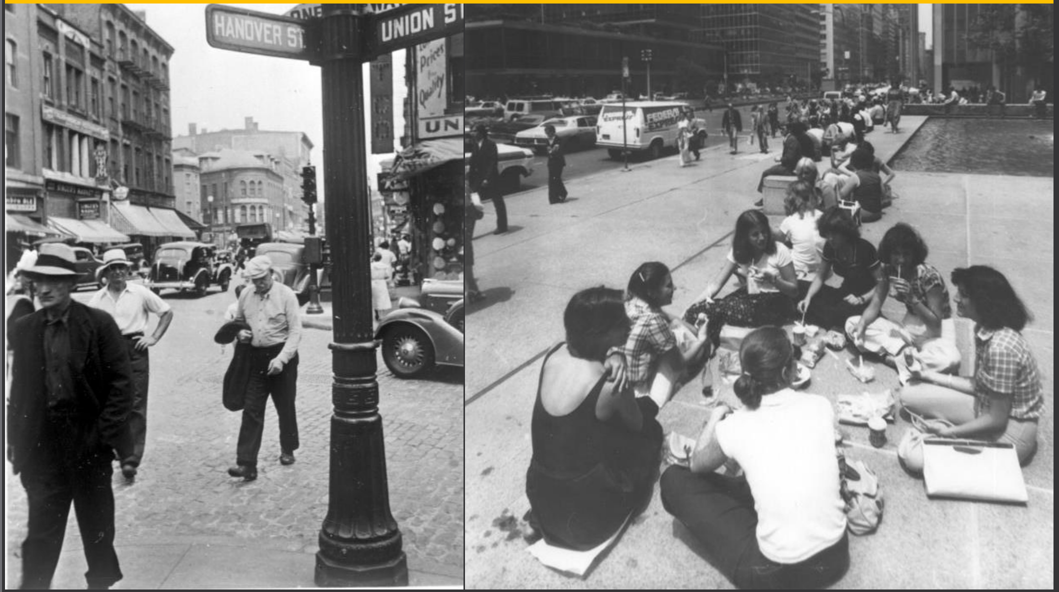
Gehl. (2017)

Pro- social places

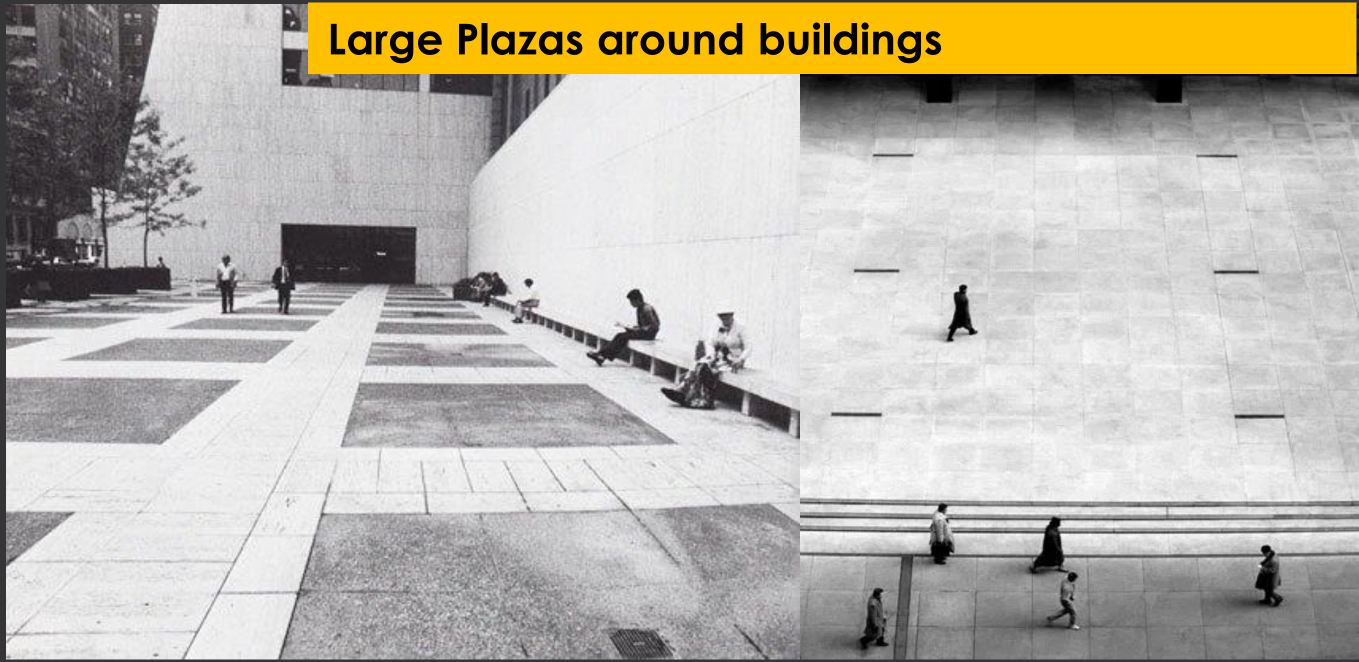
Corcoran R, Marshall G. (2016)

Urban Theory

Small spaces around vital urban Functions



Large Plazas around buildings



"for the foreseeable future the opportunities in the center city are going to be for **small spaces**." (p101,Whyte)

Kevin Lynch talks about the importance of **vitality**, the degree to which urban form supports **vital function**.

Visual Assessment and observation

Advantages:

High Accuracy

Challenges:

Expensive

Covers limited geographical areas

Unable to detect 'affect' of individuals

social media data (LBSN)

Advantages:

Covers extensive geographic areas

Cheaper

Allows 'affect' detection

Challenges:

Noise

Inaccuracies and biases

Literature Review

Last decade has seen a rise in the **LOCATION BASED SOCIAL NETWORK (LBSN)**

Few Key Authors :

- Zheng et. al- LBSN based user recommender system
- Hasan et. al- Urban activity pattern classification
- Fias-Martinez et al – Classification of Urban Landscape
- Naaman et al. – Studying diurnal routines of people
- Sakaki et al.; D'Andrea et al. – Real time event detection (earthquake, traffic etc.)

Literature Gaps

Studies primarily focus on predicting events and human activities based on :

- Location
- Time of posts

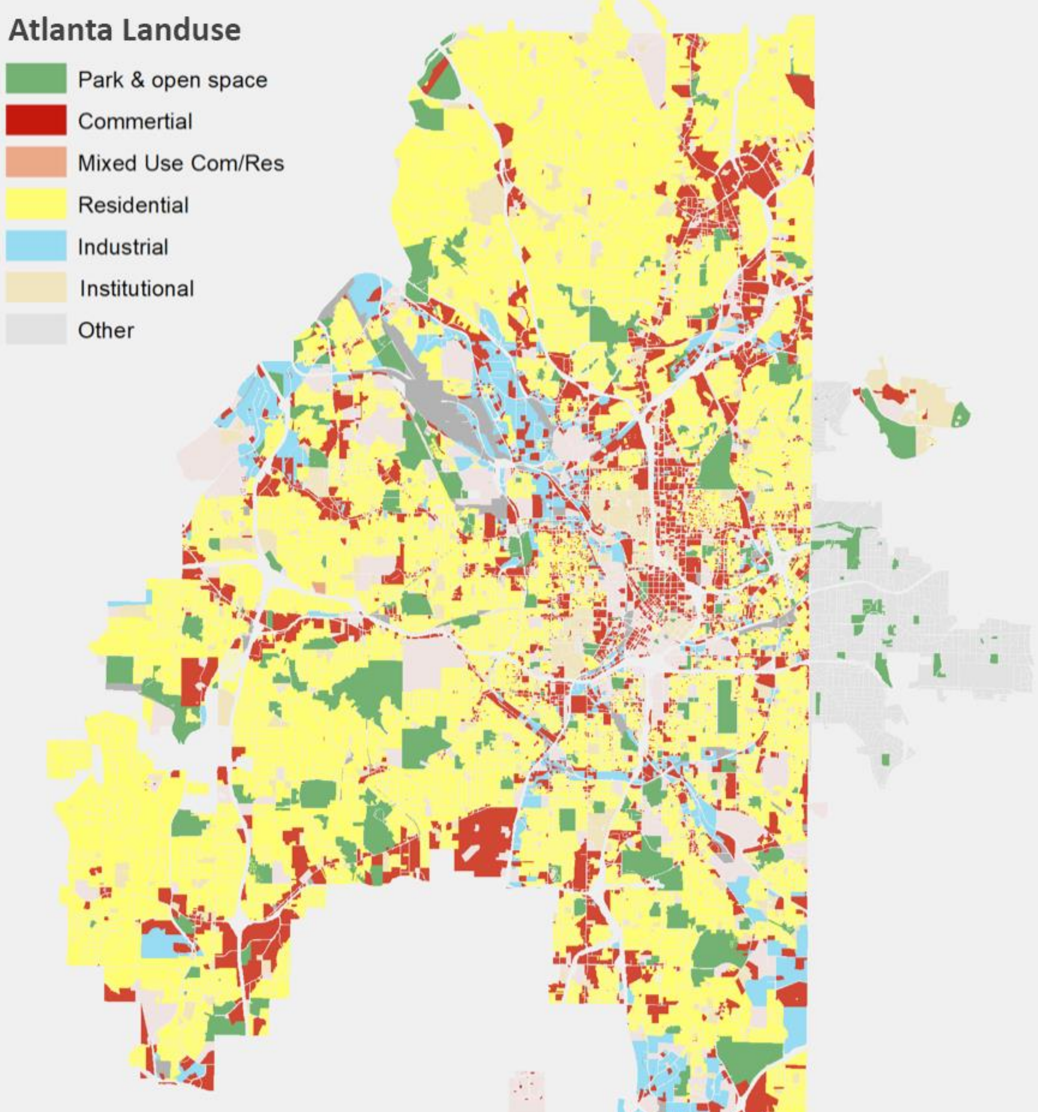
I this research we primarily focussed on:

- the text based '**urban activity**' and '**affect**' assessment
- understanding **'affect'** in relation to the spatial characteristics

Research Goals

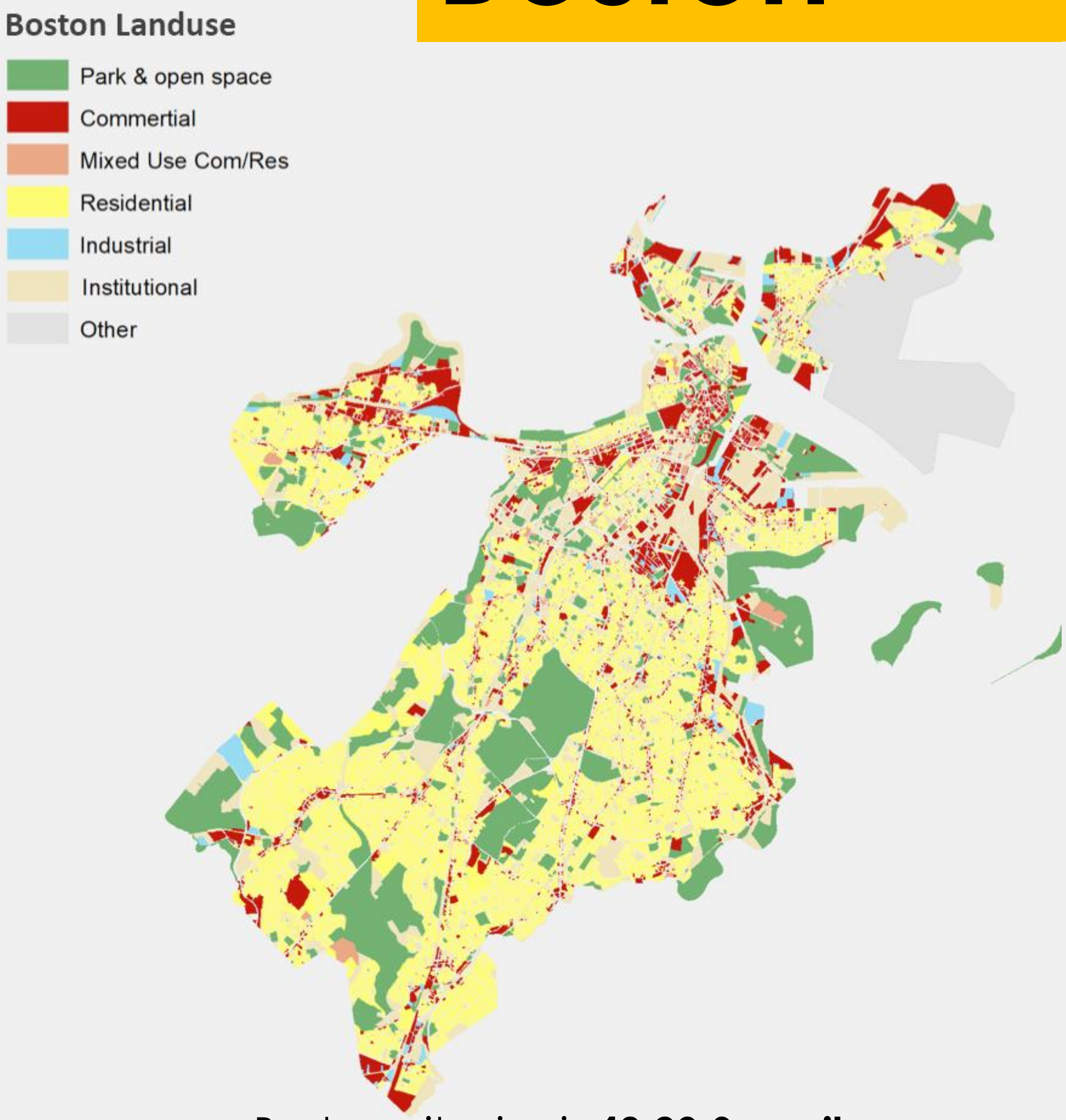
- Predict ‘urban activity’ through linguistic analysis of Tweet text
- Understand “where” people express more “positive affect” while engaging in these urban activities

Atlanta



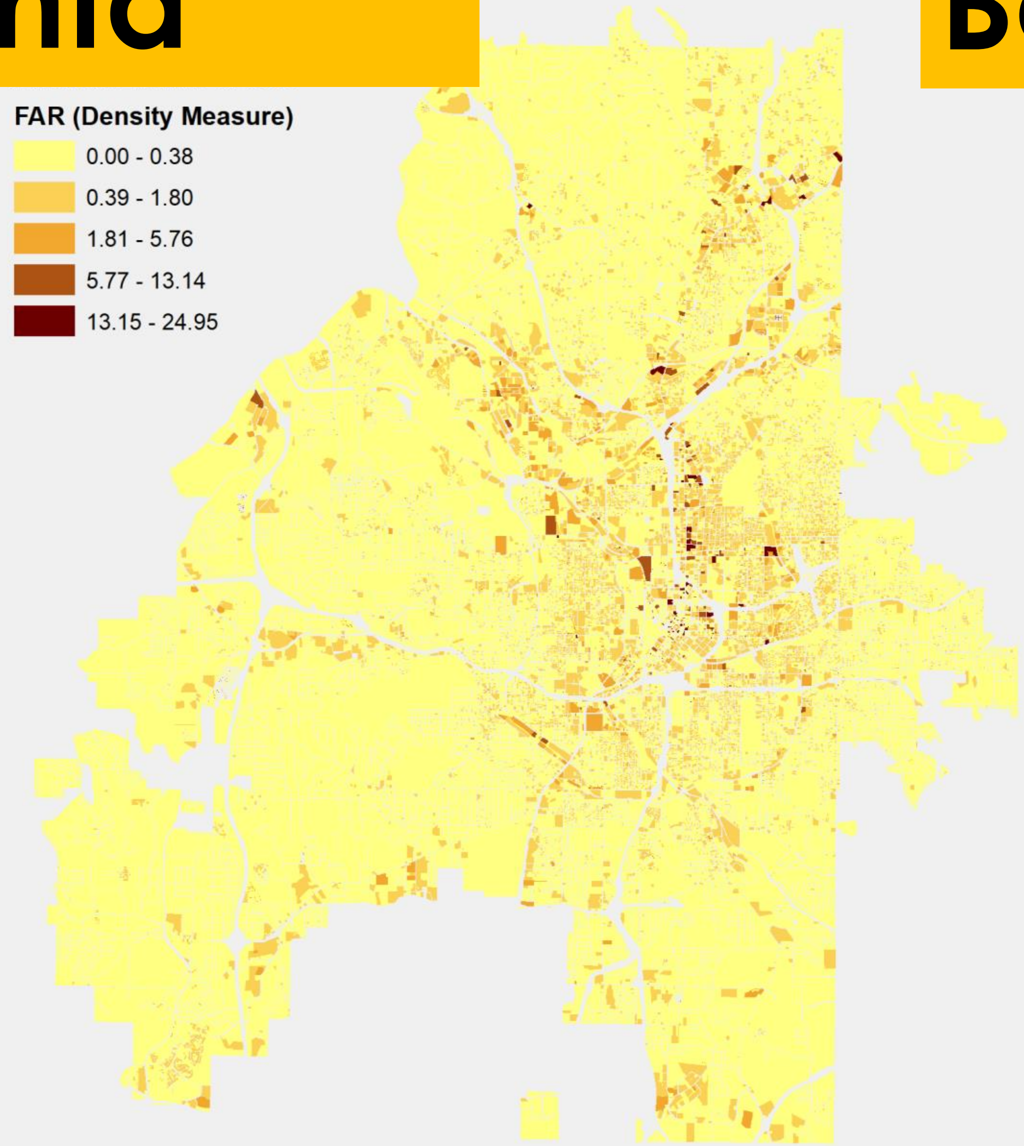
Atlanta city size is **133. 2 Sq. miles**
Number of land parcels- **166,023**
Average land parcel size- **22,958 Sq.feet**

Boston



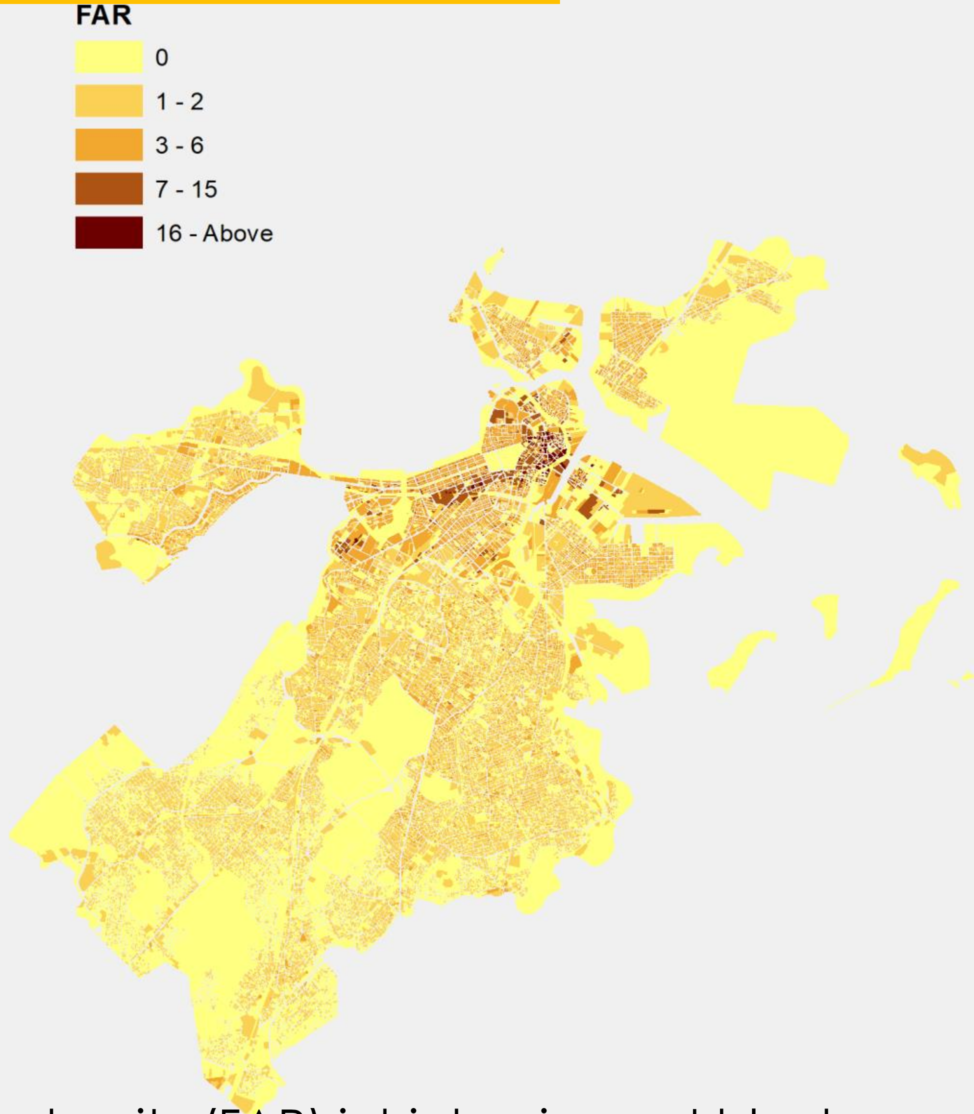
Boston city size is **48.23 Sq. miles**
Number of land parcels -**166,248**
Average land parcel size **8099 Sq.feet**

Atlanta



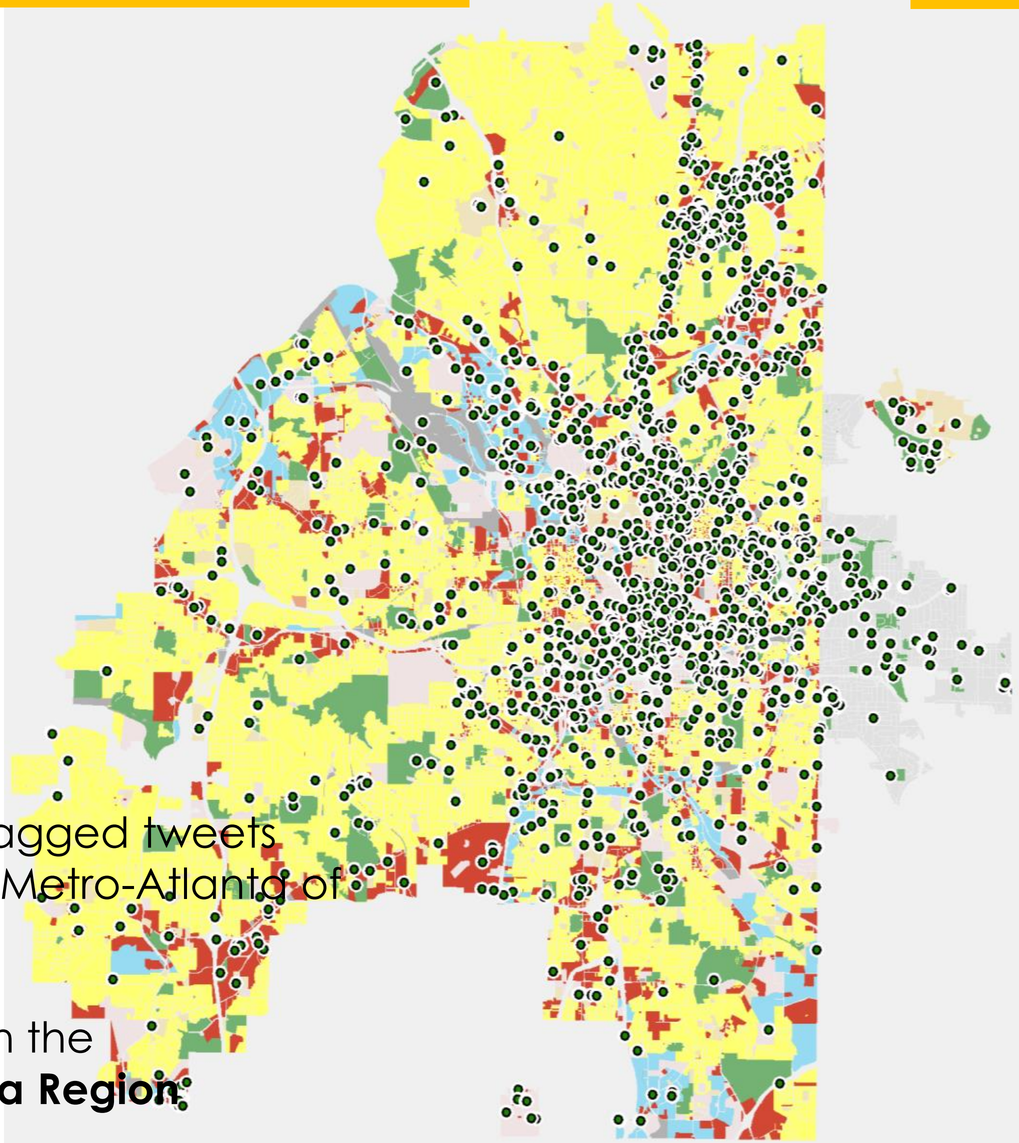
Building Density (FAR) is high in the city center
Street Network Density : **317 nodes/Sq Miles**

Boston



Building density (FAR) is higher in most blocks
Street Network Density : **552 nodes/Sq. miles**

Atlanta

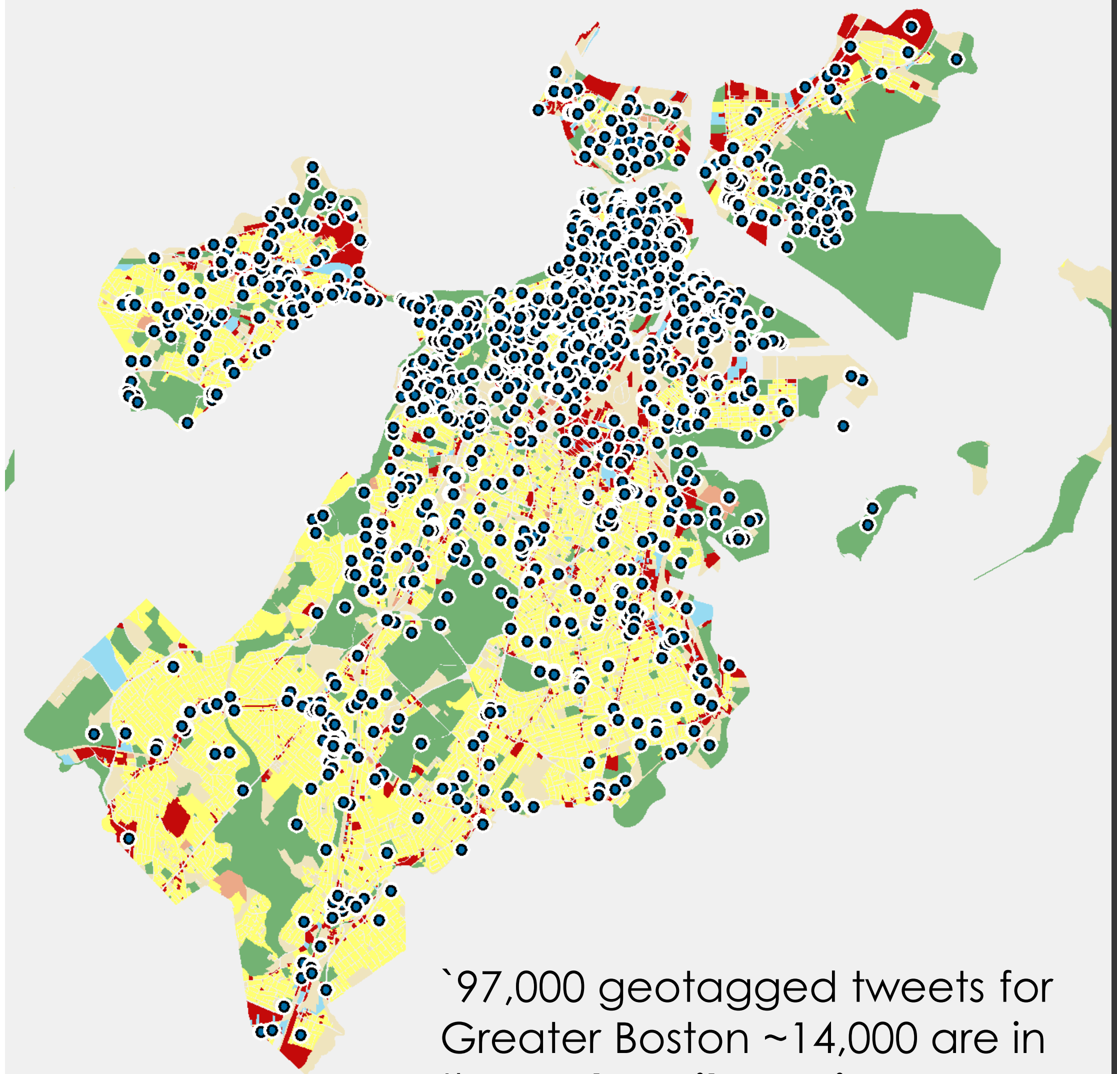


`50,000 geotagged tweets analyzed for Metro-Atlanta of which only

~16,000 are in the **City of Atlanta Region**

Tweet Density **121** per sq. miles

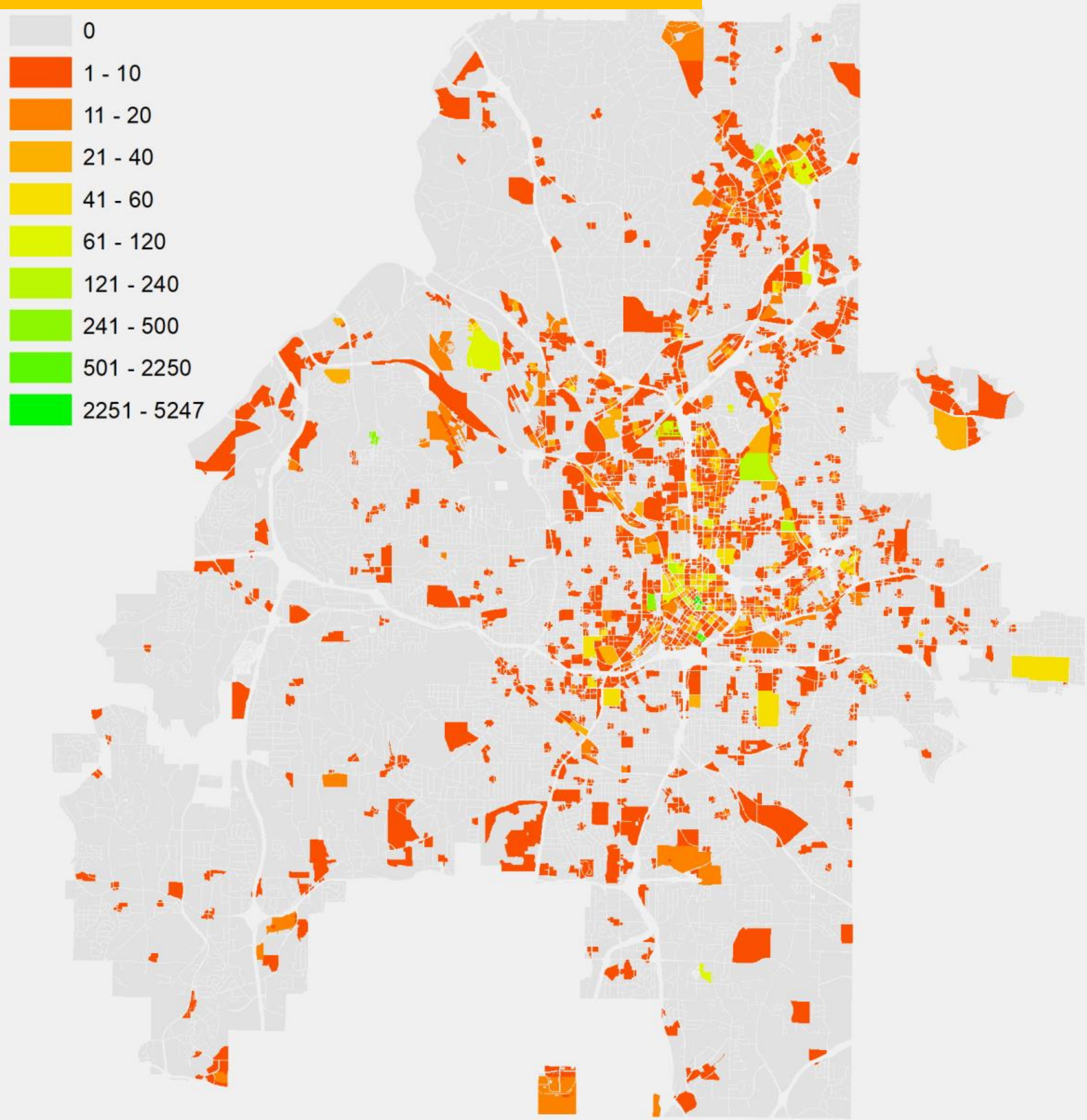
Boston



`97,000 geotagged tweets for Greater Boston ~14,000 are in the **Boston city Region**

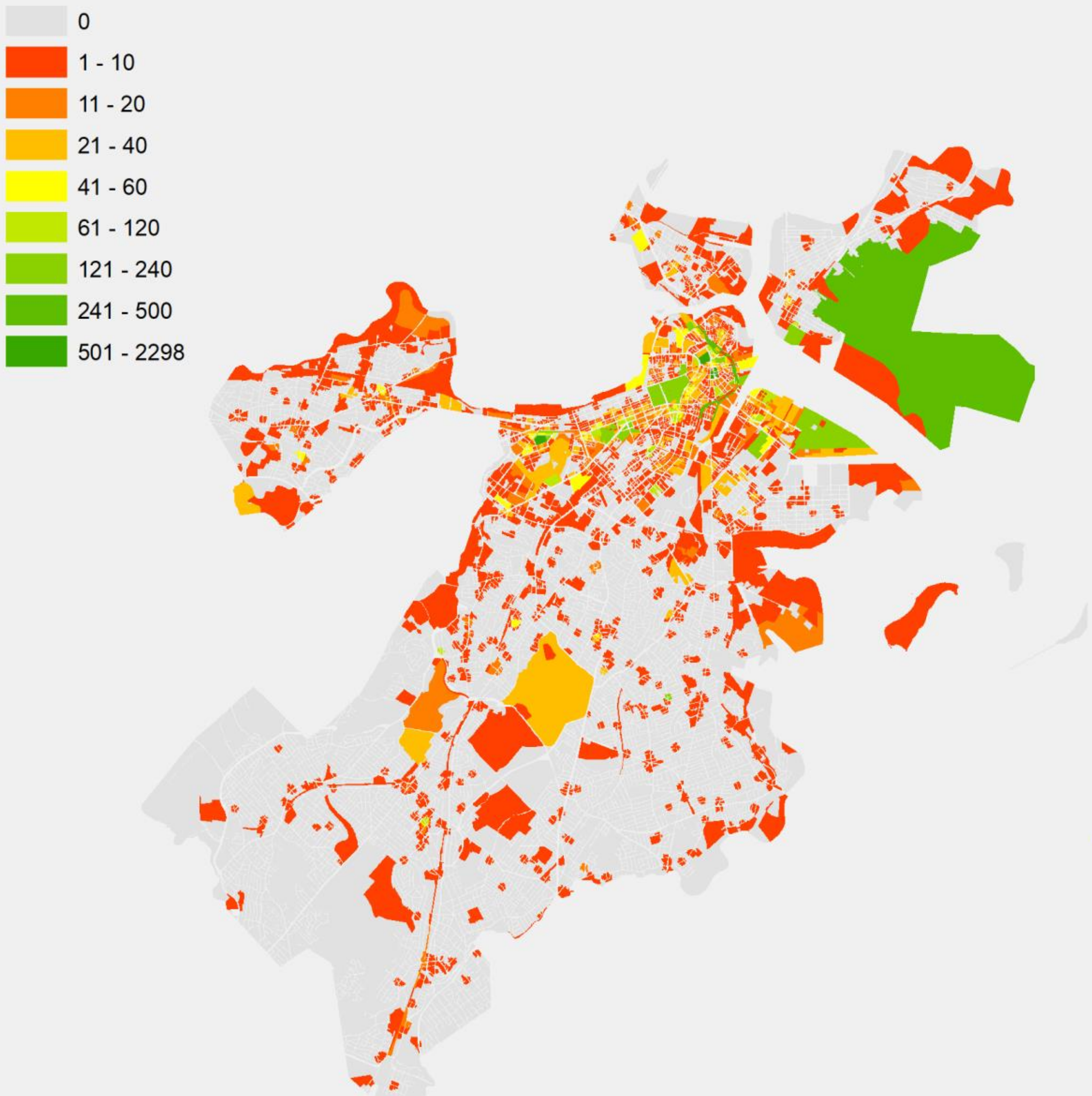
Tweet Density **290** per sq. miles

Atlanta



~16,000 are in the (in 6 months) **City of Atlanta**
Tweet Density **121** per Sq. miles

Boston



~14,000 are in the **Boston city** (6 months)
Tweet Density **290** per Sq. miles

Hypothesis

H1. Urban blocks with high street network density and building density (FAR) produce more activities, and induce more positive affect.

H2. Certain land use categories encourage more activities that induce positive affect.

H3. Boston due to its inherently different urban characteristics than Atlanta will show differences in activity pattern and affect values.

Urban ACTIVITY

AFFECT

SPACE

- Knowing activity indicated by a Tweet
  - Entertainment
  - Shopping
  - Outdoor activity
  - Home based
  - Work based
  - Mobility

- Assessing Affect of Tweets
  - Positive
  - Negative

Getting the built environment Metrics
- Land use
- Network density
- Building density etc.

"If you have to be stuck in traffic a #rainbow is a great view"

29 likes

APRIL 26

Add a comment...

Coming through Atlanta tonight during rush hour was a challenge. Stop and go bumper to bumper traffic traffic during rain was a challenge. The answer, but a Peachtree pass that lets one... instagram.com/p/BoNn2IOgUXF/ ...

Atlanta

Mobility-Drive

"Only in this city on a rainy day stuck in traffic lead to a great moment"

8 likes

APRIL 25

Boston is truly an intellectual powerhouse. A city of bustling nightlife and diverse cultures. Although traffic jam is common, the surrounding architecture... instagram.com/p/Bpl06o2Bc51u ...

10:19 AM - 20 Oct 2018 from Boston, MA

Boston

Boston is truly an intellectual powerhouse. A city of bustling nightlife and diverse cultures. Although traffic jam is common, the surrounding architecture will surely keep your eyes busy! #eastcoast #exploringboston

dzenirow Nice 😍

Boston is truly an intellectual powerhouse. A city of bustling nightlife and diverse cultures. Although traffic jam is common, the surrounding architecture… instagram.com/p/Bpl06o2Bc51u …

10:19 AM - 20 Oct 2018 from Boston, MA

Mobility-Drive

3

Boston

Only in this city on a rainy day stuck in traffic lead to a great moment #ShotOniPhone X

8 likes

APRIL 25

Stuck in traffic with a pretty view

7 likes

OCTOBER 14

11AM start time!

Traffic got us jammed up in deez ATL Streets!!

"Nothing stops a hustle like traffic!!! We will be done at 11am today!!! "

23 likes
OCTOBER 18
Add a comment...

Coming through Atlanta tonight during rush hour was a challenge. Stop and go bumper to bumper traffic traffic during rain was a challenge. The answer, but a Peachtree pass that lets one...
instagram.com/p/BoNn2lOgUXF/ ...

10:30 AM - 27 Sep 2018 from Atlanta, GA

"If you have to be stuck in traffic a #rainbow is a great view"

29 likes
APRIL 26
Add a comment...

"just saw three accidents within a mile of each other. traffic is crazy"
Sat May 05 10:47:43 2018

Atlanta

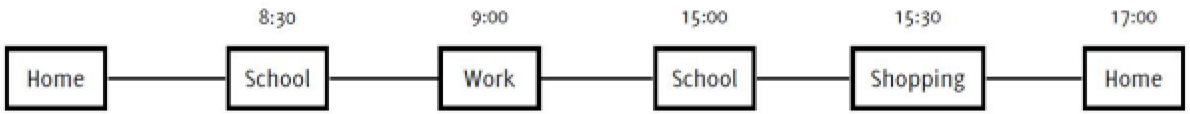
Mobility-Drive

Urban ACTIVITY

Predict 'Urban Activity'

Tweet topics

Activities primarily focus on recreation and work in outdoor / indoor spaces.



Key activities selected

- **Entertainment** (food and drinks, events , shows , games , movies etc.)
- **Recreation** (primarily outdoor activity - workout , beach activity etc.)
- **Shopping** (buying grocery or any other goods)
- **Home based** (primarily outdoor indoor activity at home)
- **Work related** (outdoor/indoor work related activity)

+

Mobility (walking , biking, taking transit , driving)

STEP- 1

- **Data Cleaning**

~14,000 Tweets for Greater Boston Area

~16,000 Tweets for Metro Atlanta

STEP- 2

- **Hand labeling**

Hand labelled ‘k’ tweets
Memos for ‘k’ tweets
Creating a Training Dataset
~1000 hand labeled data points

STEP- 3

- **Clustering & Spot checking**

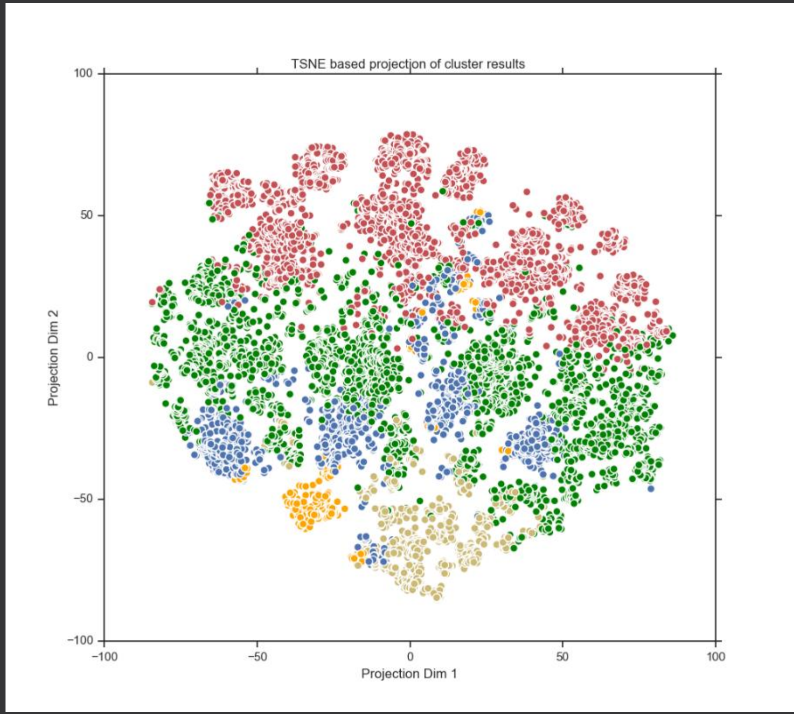
- After iterative clustering ~4000 semi-automated labeled data points are generated

Clean Tweets → Labeled + Relevant Tweets → Increasing labeled Tweets

Method

- **Cluster Training Data in K – groups**
  - Review K groups, used Silhouette co-eff as metric
  - Update K –Tweets with M- examples
  - Review and Update Memos
  - Iterate

- **Output = semi – unsupervised data labeling**
  - began with ~ 1000 hand labelled data points , k =5 ,m =200
  - After iterative clustering ~4000 semi-automated labelled data points



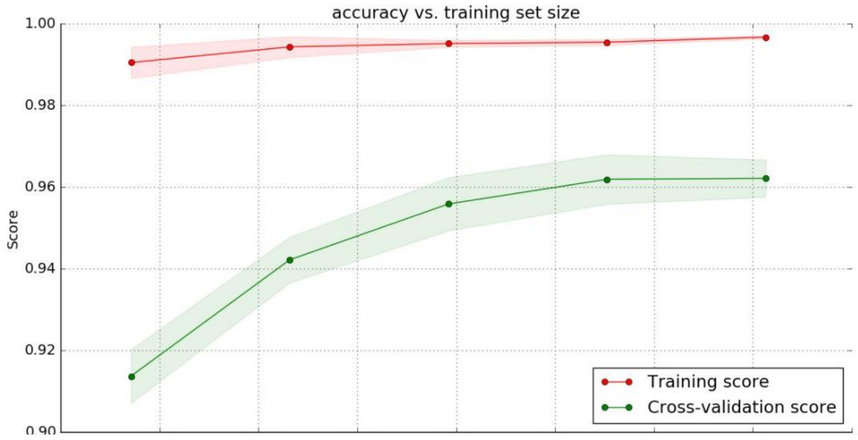
Clustering Diagram showing 5 key categories

STEP- 4 •Model Building

Random Forest Machine Learning Model for Tweet Classification

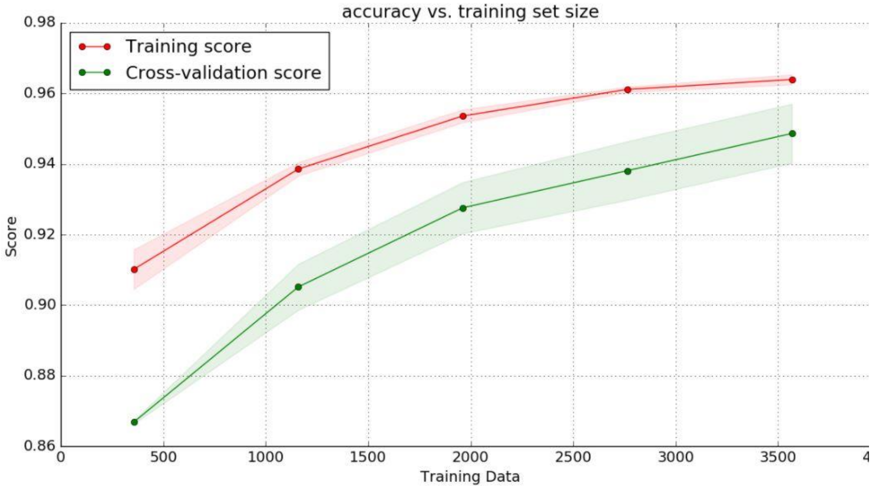
Atlanta
94.89 %

Activities	precision	recall	f1-score
educational	1.00	0.67	0.80
entertainment	0.93	1.00	0.96
home	1.00	0.80	0.89
mobility	1.00	0.62	0.77
recreation	1.00	0.75	0.86
shopping	1.00	0.50	0.67
work	1.00	1.00	1.00



Boston
93.63%

Activities	precision	recall	f1-score
educational	0.85	0.88	0.87
entertainment	0.99	0.99	0.99
home	1.00	0.78	0.88
mobility	1.00	0.91	0.95
recreation	1.00	0.93	0.97
shopping	1.00	0.82	0.90
work	0.94	1.00	0.97



For the entire dataset the accuracy is 72.14%

Urban ACTIVITY

Example Tweets :

Outdoor Activity

Enjoyed a beautiful afternoon exploring The Freedom Trail.

Mobility

@mbta mechanical issues with bus 676? Was very slow so I bailed and took a different route

Shopping

Doing a little Sunday shopping @whitebarnfarm for sides for Burger night tomorrow

Entertainment

We just had a delicious breakfast at the redarrowdiner of @foodnetwork Dinners Drive-Ins

AFFECT

Assessing ‘Affect’ of Each Tweet

STEP- 5 Affect Assessment

Assessment based on the words in the tweets

- words that reflect **positive and negative affect are assessed**
  - Positive Affect – e.g. nice, love, wonderful, fun, sweet etc.
  - Negative Affect – e.g. nasty , bad , ugly , hurt etc.

From each tweet a score is generated that reflects the positive and negative affective measures.

USED – Modified dictionary of the the Linguistic Inquiry and Word count (LIWC 2015) and, Vader python Library

AFFECT

Example Tweets :

Positive – Happy Tweet

Enjoyed a beautiful afternoon exploring The Freedom Trail. **+20.1**

Negative – Angry Tweet

Very frustrating weekend in the first lap of the main. I hit a bike in the sand and was stuck **-25.0**

Negative – Anxious Tweet

I hate driving down here. He was fine. Ran out in front of me as I was starting to take a right turn. So no damage to my car. Or him. **-12.5**

Urban "Activity" comparison

Linear Regression Model	Atlanta Coefficients	Atlanta Sig.		Boston Coefficients	Boston Sig.	
	B					
1 (Constant)	5.372	0.011		5.793	0.119	
FAR	**0.122**	**0.043**	*	**1.572**	**0.000**	***
Network Density	**0.585**	**0.001**	**	0.037	0.696	
Parcel Area	-2.85E-06	0.477		5.25E-07	0.414	
LU=Commercial	3.961	0.101		**10.727**	**0.003**	***
LU=Mixed Use Com/Res	-3.913	0.669		2.938	0.451	
LU= Institutional	**13.151**	**0.001**	**	3.748	0.324	
LU= Park & Open Space	**18.195**	**0.000**	***	**6.373**	**0.048**	*
LU=Residential	-1.15	0.574		-0.407	0.912	

Dependent Variable: Urban Activity (function of Tweet volume)

	Atlanta	Boston
R-Square	**0.1**	**0.32**

p-value < .05

Activity & Landuse Distribution in Atlanta

	Commercial	Institutional	Industrial	Mixed Use Com/Res	PublicPark & Open Space	Residential
educational	0.15	0.52	0.02		0.04	0.03
entertainment	23.44	31.16	3.68	0.17	2.60	6.23
home	0.20	0.23	0.02		0.02	0.07
mobility	0.32	0.47	0.04		0.10	0.10
recreation	0.16	0.31	0.03		0.41	0.02
shopping	0.04	0.03	0.01			
work	0.31	0.46	0.03		0.07	0.09
	24 %	33%	3.8%	0.2 %	3.2%	6.5%

SUM([Number of Records]*100/16679) broken down by LU Final vs. True Label. Color shows sum of Number of Records. The marks are labeled by SUM([Number of Records]*100/16679). The view is filtered on True Label and LU Final. The True Label filter excludes others and posting. The LU Final filter excludes No Landuse Info.

Number of Records

1 5,198

Atlanta

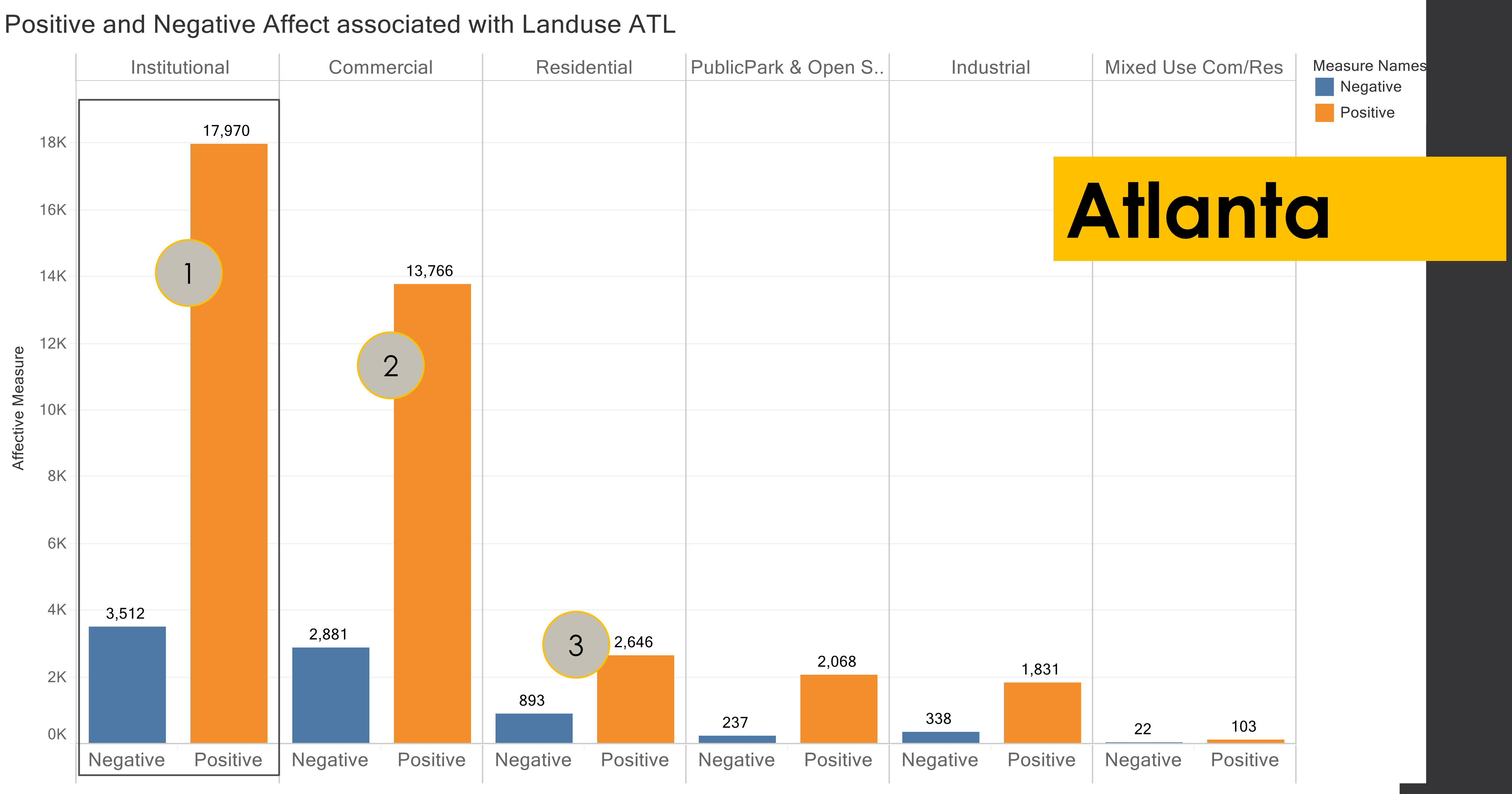
True Labels	Commercial	Institutional	Industrial	Mixed Use Com/Res	Park & open space	Residential
	Landuse Final					
educational	4.40	1.18	0.15	0.49	1.34	1.15
entertainment	35.08	8.67	1.25	4.17	8.03	6.68
home	0.31	0.08	0.01	0.02	0.05	0.10
mobility	0.73	0.20	0.01	0.13	0.96	0.21
recreation	1.18	0.16	0.01	0.02	0.17	0.26
shopping	0.11			0.02		0.02
work	0.11	0.02	0.02	0.02	0.02	0.02
	42%	10%	1.5%	5%	10%	8.5%

SUM([Number of Records]*100/13320) broken down by Landuse Final vs. True Labels. Color shows sum of Number of Records. The marks are labeled by SUM([Number of Records]*100/13320). The view is filtered on Landuse Final and True Labels. The Landuse Final filter excludes Null. The True Labels filter excludes others and posting.

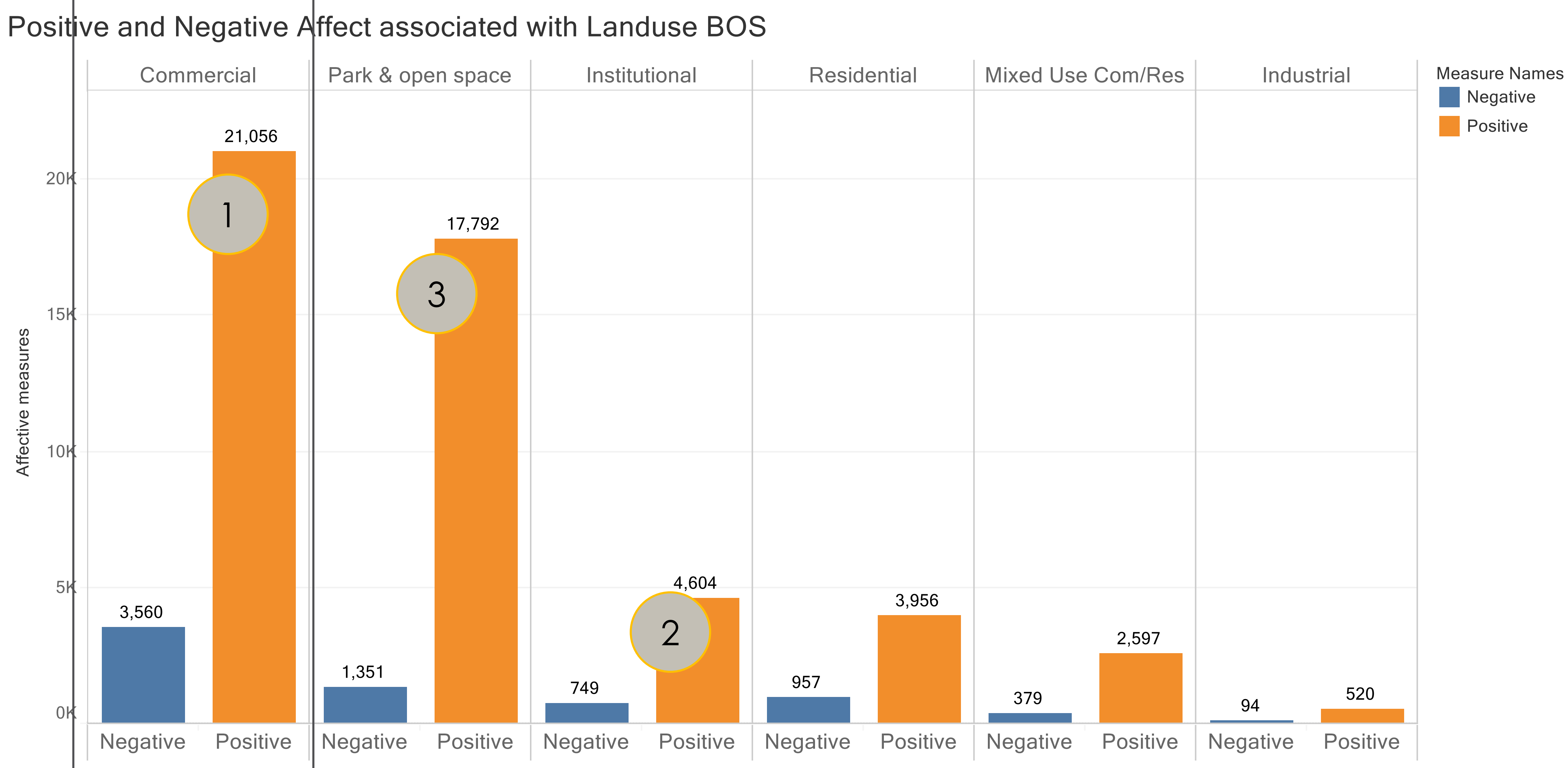
Number of Records

1 4,673

Boston

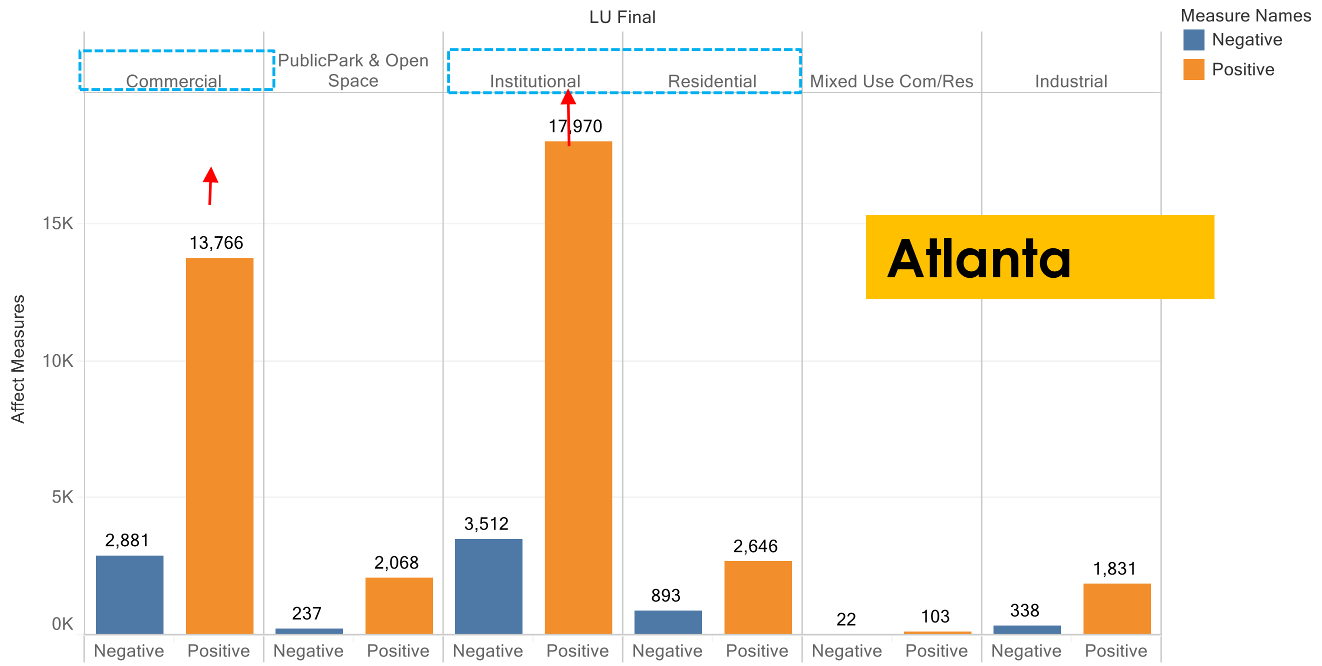

Positive and Negative Affect associated with Landuse ATL
Institutional
Commercial
Residential
PublicPark & Open S..
Industrial
Mixed Use Com/Res
Measure Names
Negative
Positive
Affective Measure
18K
16K
14K
12K
10K
8K
6K
4K
2K
0K
17,970
3,512
13,766
2,881
893
2,646
237
2,068
338
1,831
22
103
1
2
3
Negative
Positive
Atlanta

Boston

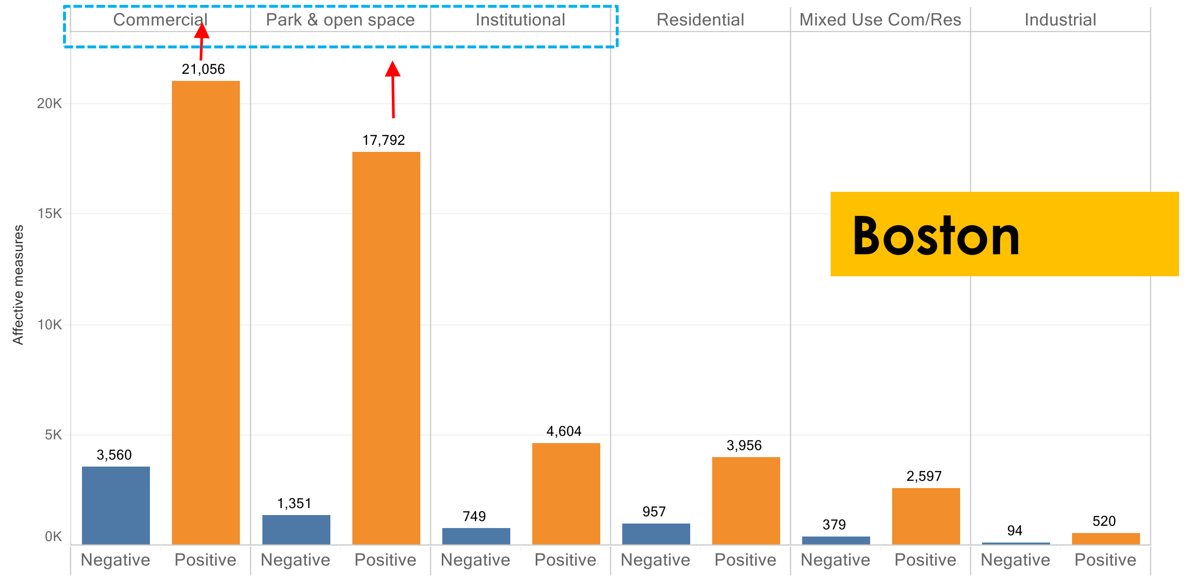


Negative and Positive for each Landuse Final. Color shows details about Negative and Positive. The view is filtered on Landuse Final, which excludes Null.

Positive and Negative Affect associated with Landuse ATL



Positive and Negative Affect associated with Landuse BOS



Negative and Positive for each Landuse Final. Color shows details about Negative and Positive. The view is filtered on Landuse Final, which excludes Null.

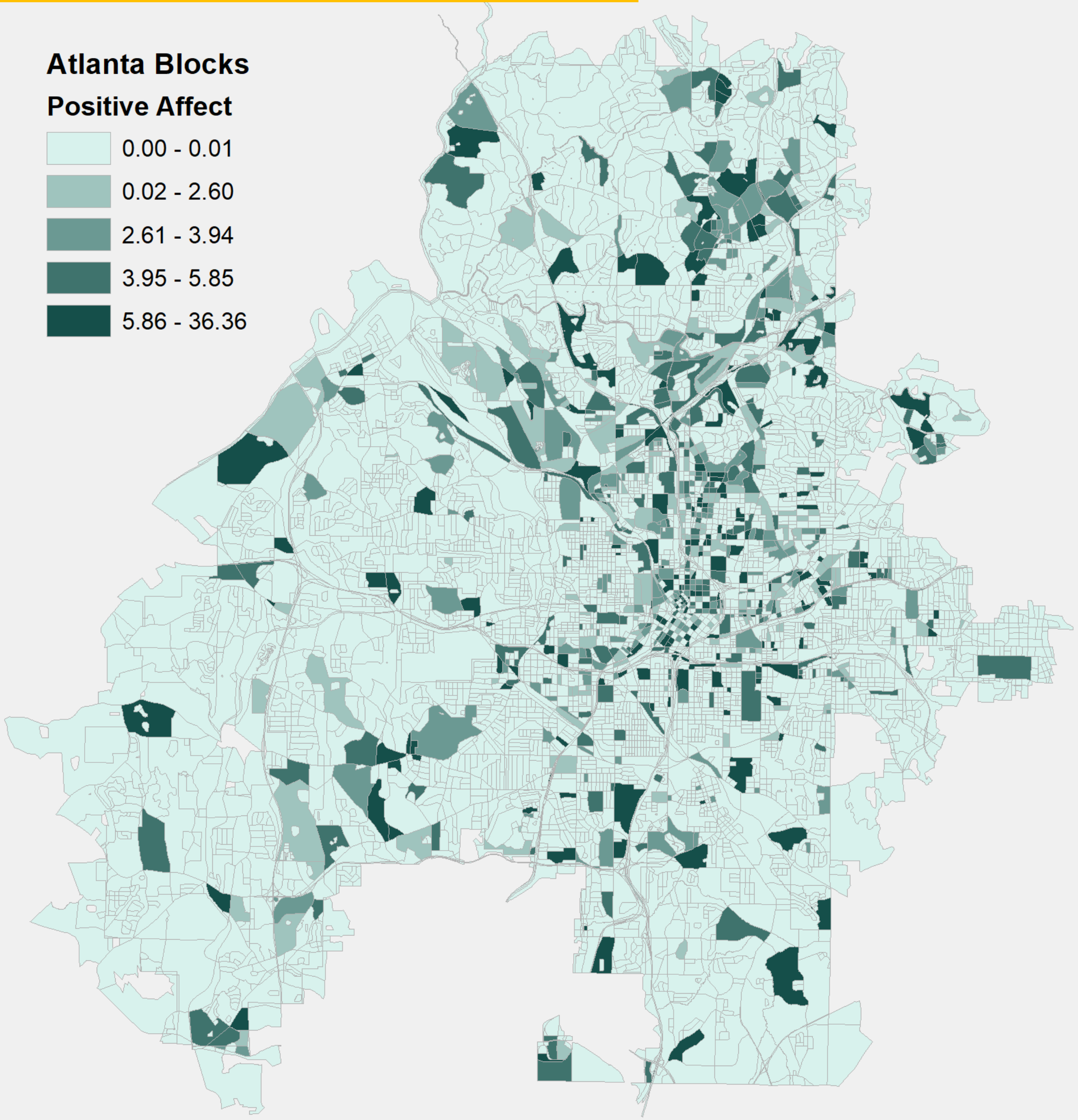
Boston in Comparison to Atlanta

Higher Positive Affect in :

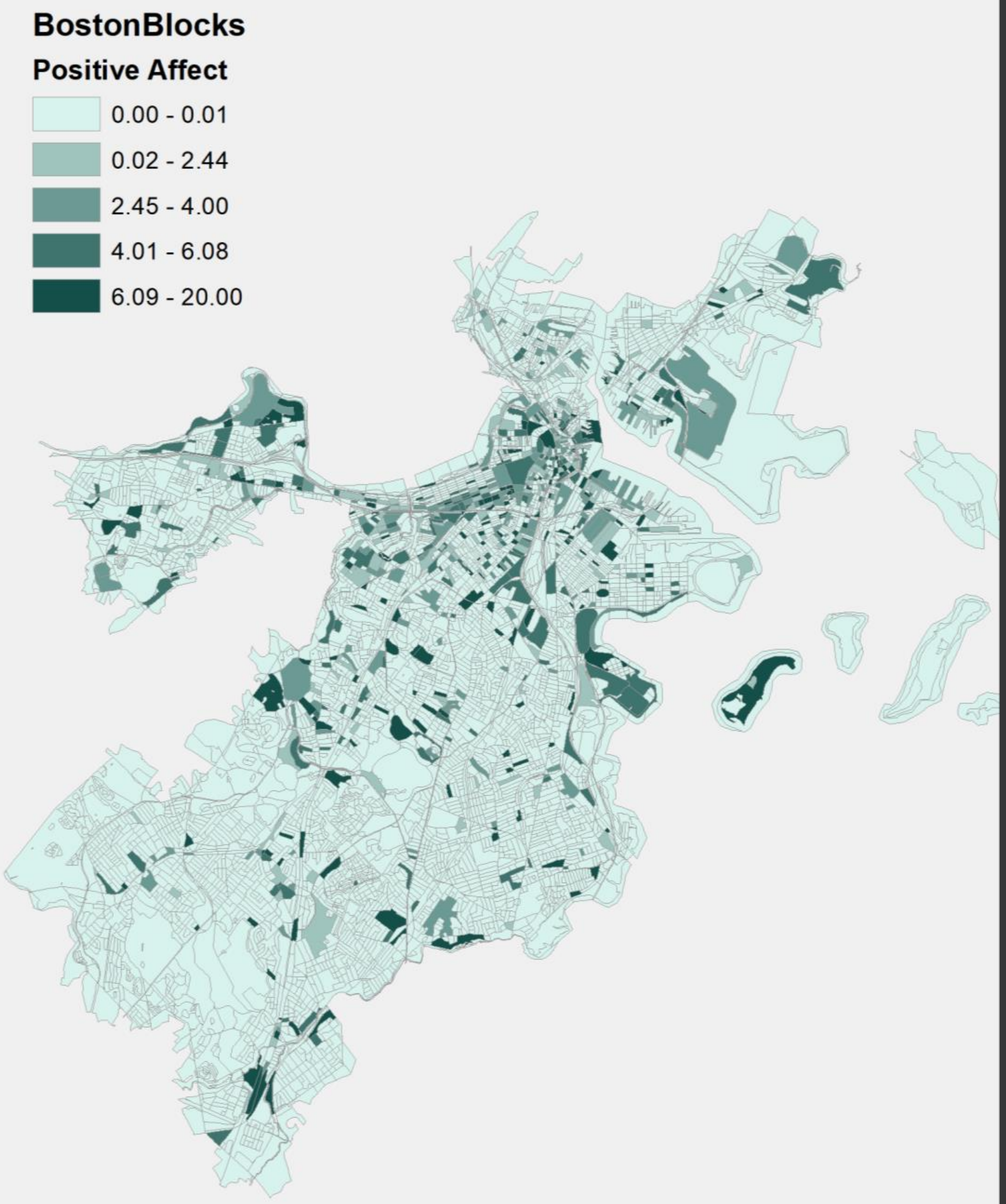
Commercial
Park and open spaces
Institutional

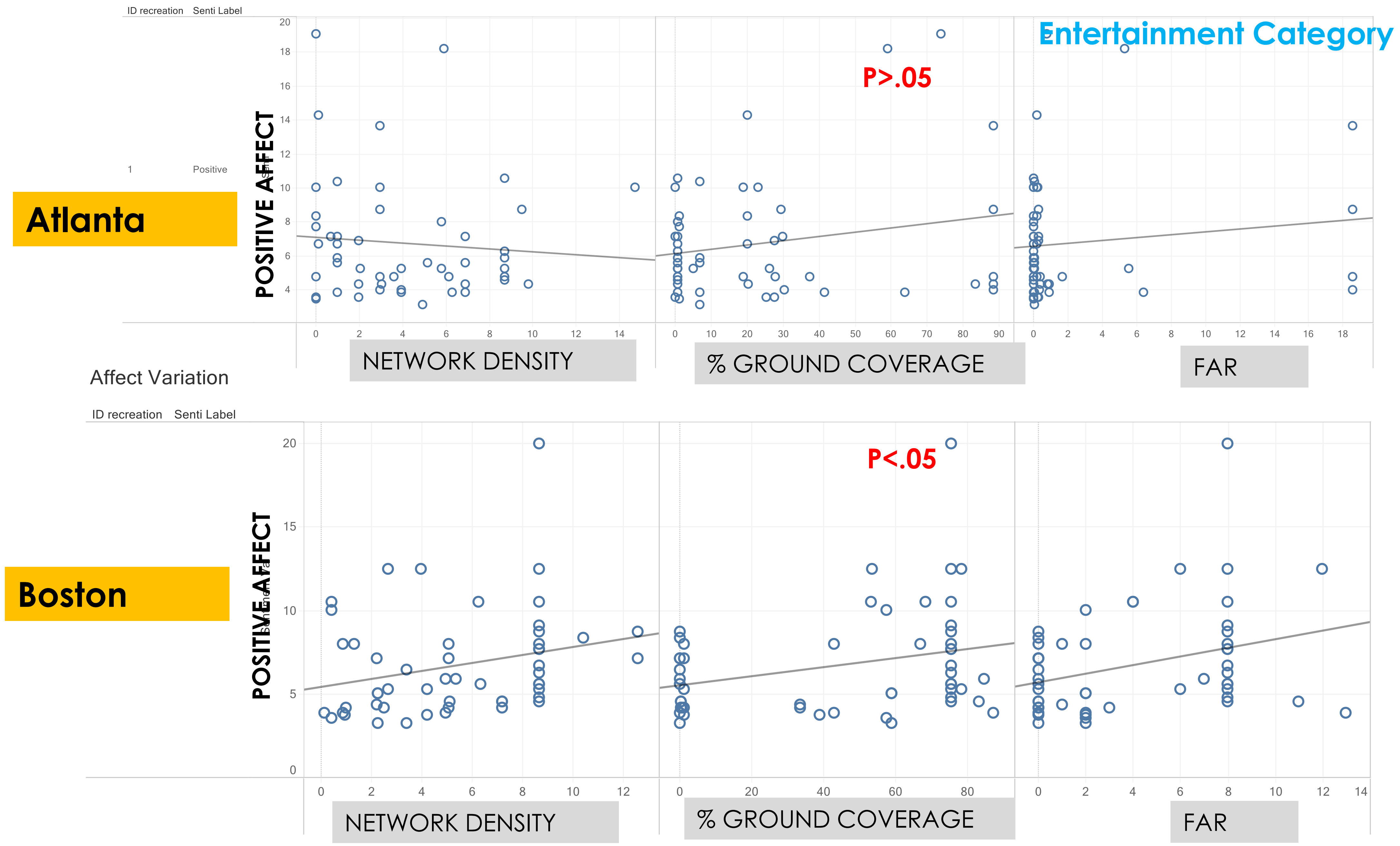
Higher positive Affect in :
Commercial
Institutional
Residential

Atlanta



Boston





Entertainment Category
Atlanta
ID recreation Senti Label
1 Positive
POSITIVE AFFECT
P>.05
NETWORK DENSITY
% GROUND COVERAGE
FAR
Affect Variation
Boston
ID recreation Senti Label
POSITIVE AFFECT
P<.05
NETWORK DENSITY
% GROUND COVERAGE
FAR

Findings

H1. Small blocks, high street network density and building density (FAR) produce **more activities**, and more induce **more positive affect.**

Urban Activities tend to to originate in the areas with High Building Density (FAR)

- For **Both Atlanta** and **Boston**

In Boston- Positive affect is positively correlated with Network density, FAR , % ground coverage

In Atlanta - Positive affect not correlated with Network density, FAR , % ground coverage

Findings

H2. Certain Land Use categories produces more urban activities and associated with more positive affect

Urban activities are higher in the, Park and Open Spaces – For both **Atlanta** *and* **Boston**

More 'positive affect' is associated with certain Land use categories such as **commercial** , **institutional** , **park and open spaces**

Method

- Text Pre-processing
  normalization,
  Stemming
  Tokenization

- Feature Construction
  Bag of word (bow)
  ngram analysis
  Tf-idf score
  Hand Labeling
  #Hashtags
  Post tagging

  Keyword Dictionary , memo making

Findings

H3. Boston City due to its inherent urban characteristics (mix of land use smaller blocks, accessibility) may induce more positive affect when compared to the City of Atlanta

- Spatial variables are better predictor of urban activities in **Boston**

- **Boston** shows more activities in Commercial land use , and **Atlanta** shows more activities in Institutional land use

- Boston shows highest '**positive affect**' associated with , commercial, park and open spaces, and institutional land use

- Atlanta shows highest **positive affect** in Institutional , commercial, and residential land use

Findings

H3. Boston City due to its inherent urban characteristics (mix of land use smaller blocks, accessibility) may induce more positive affect when compared to the City of Atlanta

Beyond the city center ,

- In **Atlanta** more urban activities and positive affect are situated in the commercial malls
- In **Boston** activities and positive affect beyond the city center is observed in urban parks , river walks, and small commercial establishments near transit stations.

Summary

- ‘Affect’ values in both cities primarily varies with land use types. Land use constituting prosocial places are usually associated with higher positive affect.

- Higher density encourage urban activities (such as entertainment , outdoor activity, mobility etc.)

- Higher density is not always associated with ‘positive affect’

Future directions

- **Current limitation**
  - Refined land use categories is not used in this study
  - More spatial variables can be used
  - Geo-locations for tweets have high margin of error

- **Analyze non-geo-tagged** tweets and **image data** will give us better insight

- **Study user behavior** by tracking few users and who are frequently posting on Twitter can give better insights.