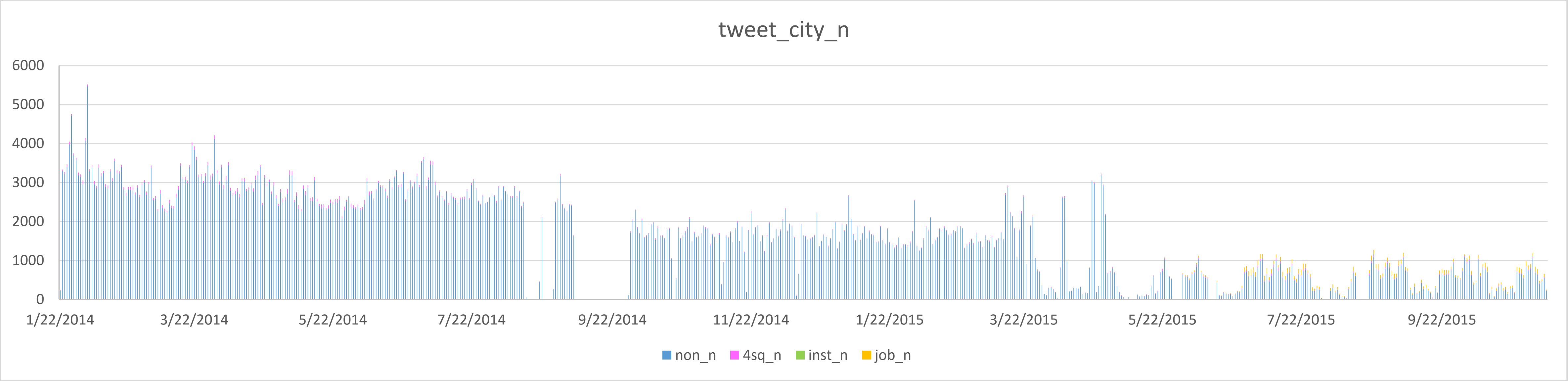
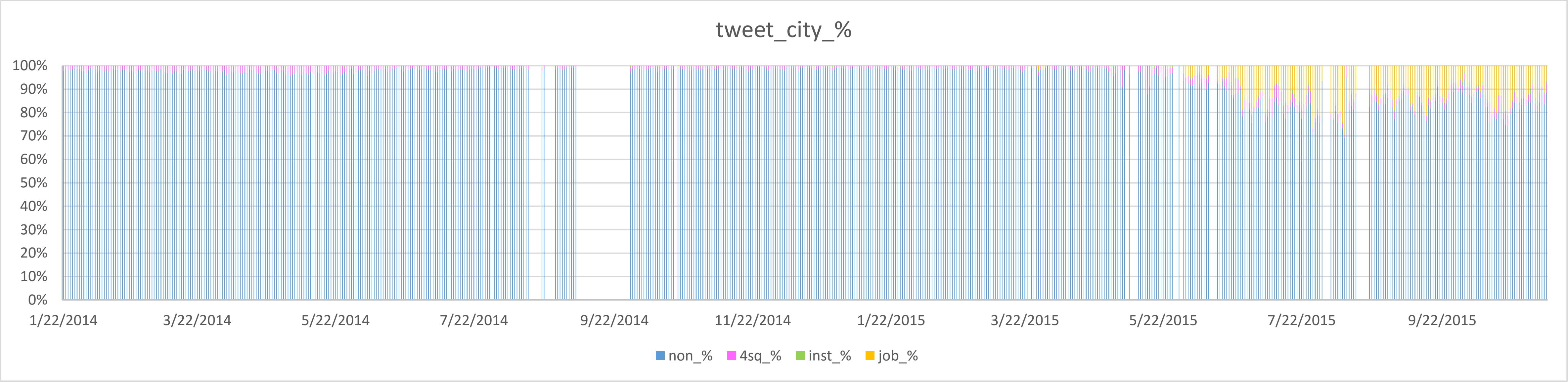
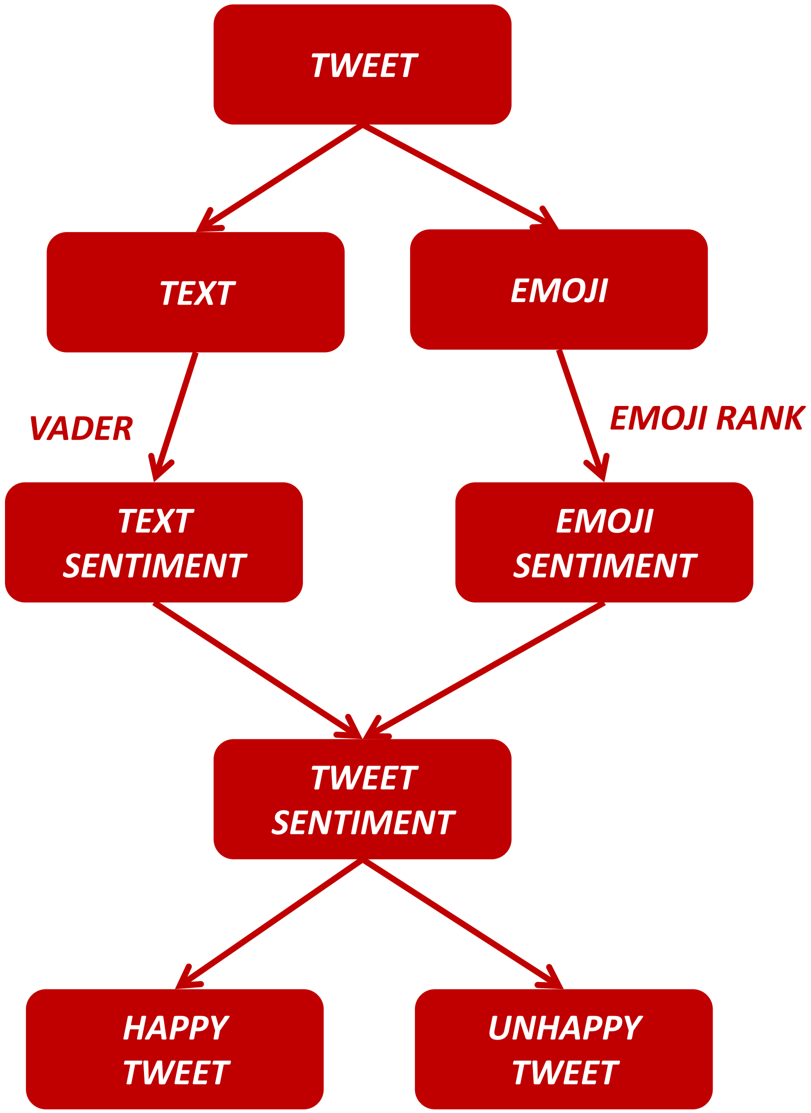
CITY

# of geo-tweets by category…

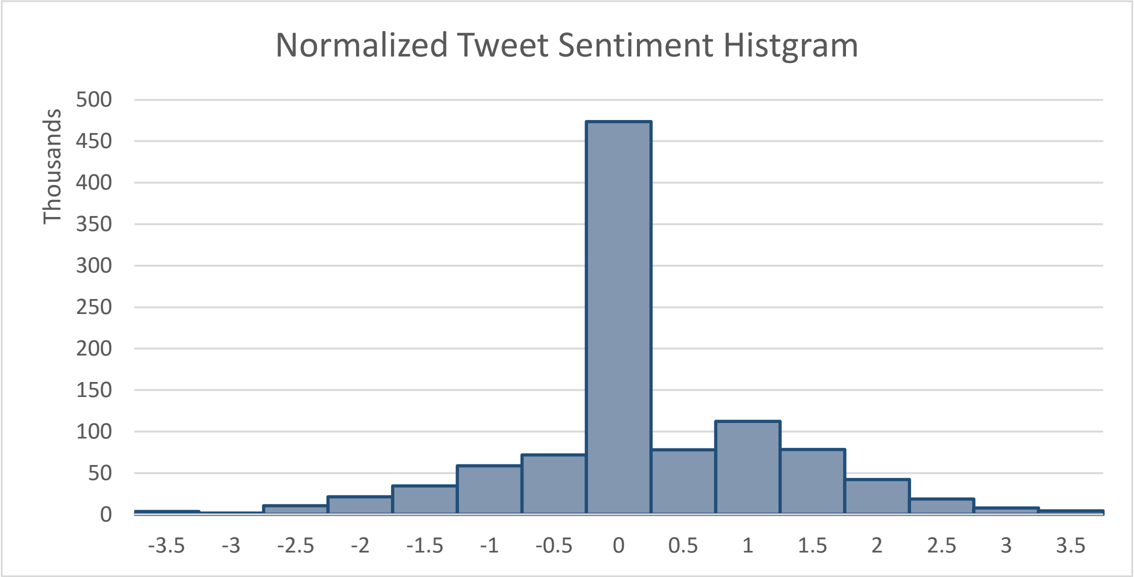


% of geo-tweets by category…



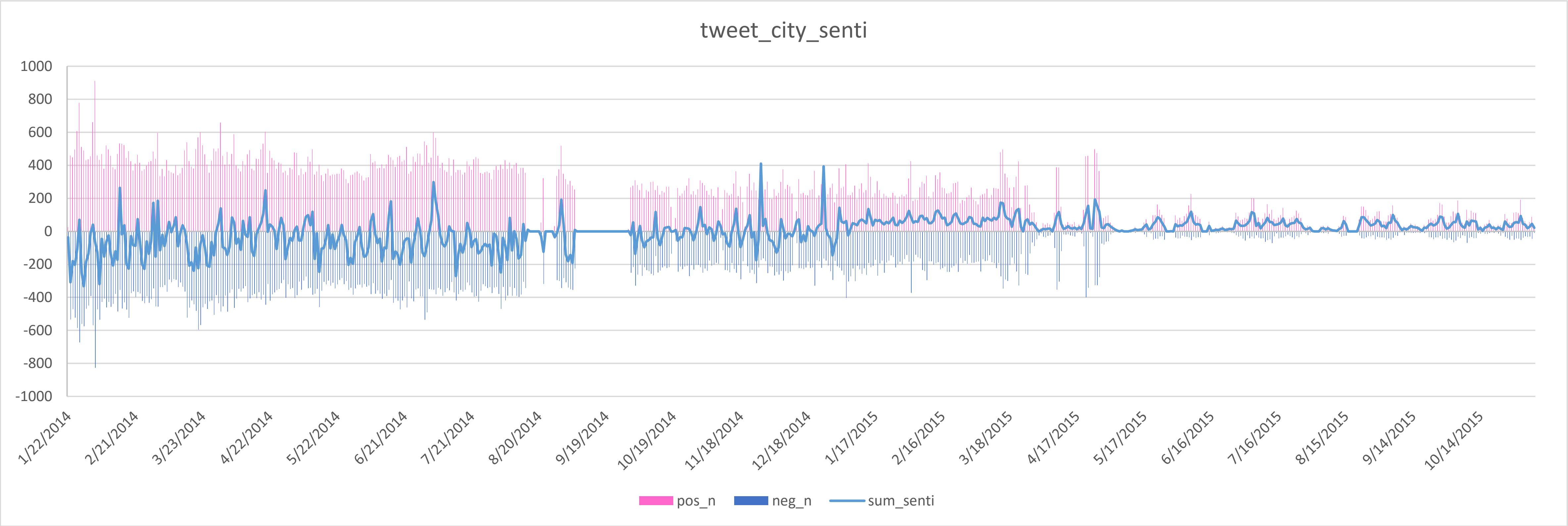


1,017,040 valid tweets
in range [(-80.10,40.36),(-79.86,40.51)]
from 01/22/2014 to 11/7/2015

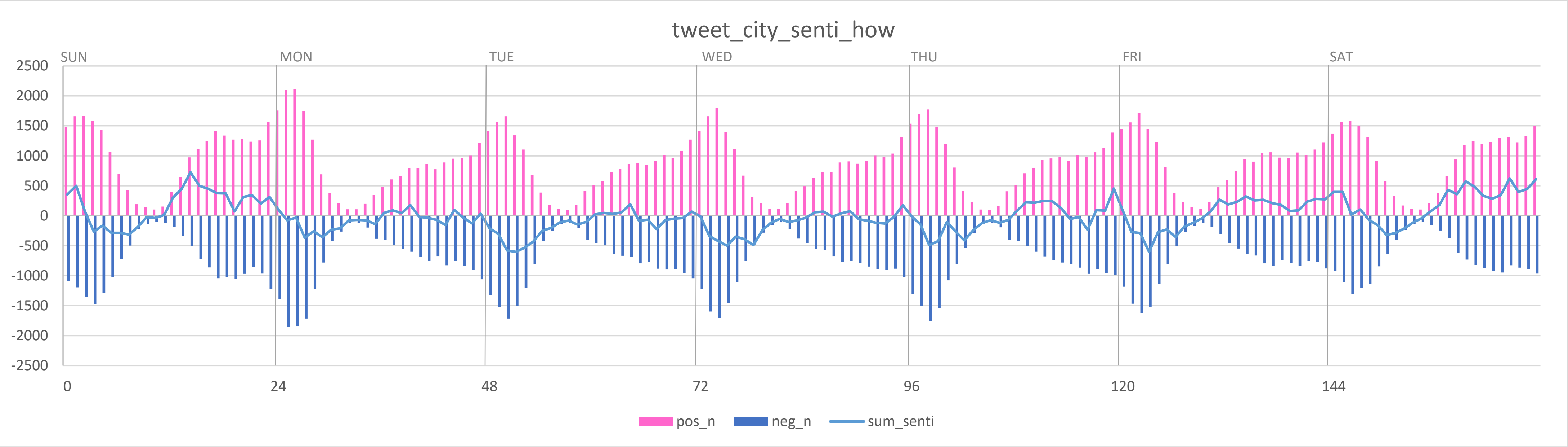


Dichotomized by stddev=1
- #_happy tweet 151,761 14.92%
- #_unhappy tweets 129,863 12.77%

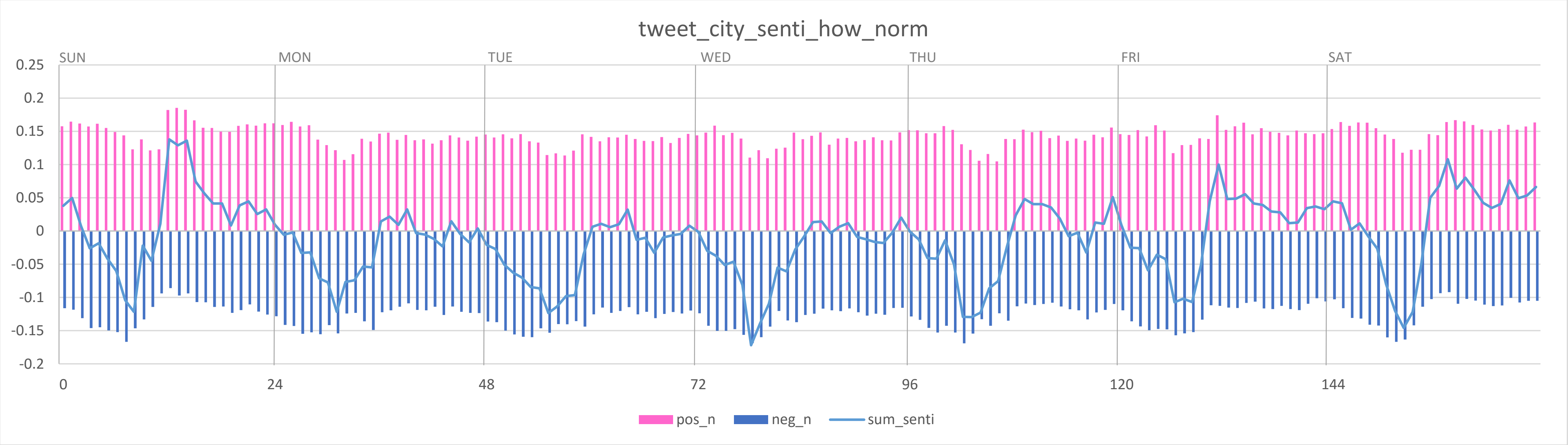
of happy/unhappy genuine tweets and sum sentiment value…



of happy/unhappy genuine tweets and sum sentiment value by hour of week…



% of happy/unhappy genuine tweets and sum sentiment value by hour of week…



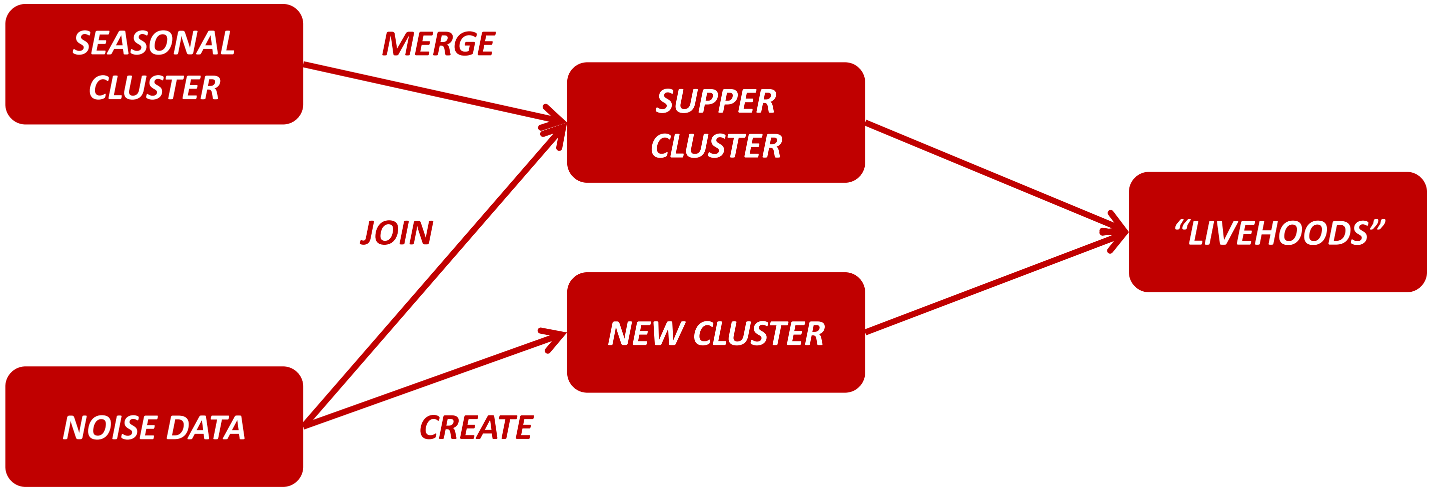
Some findings:

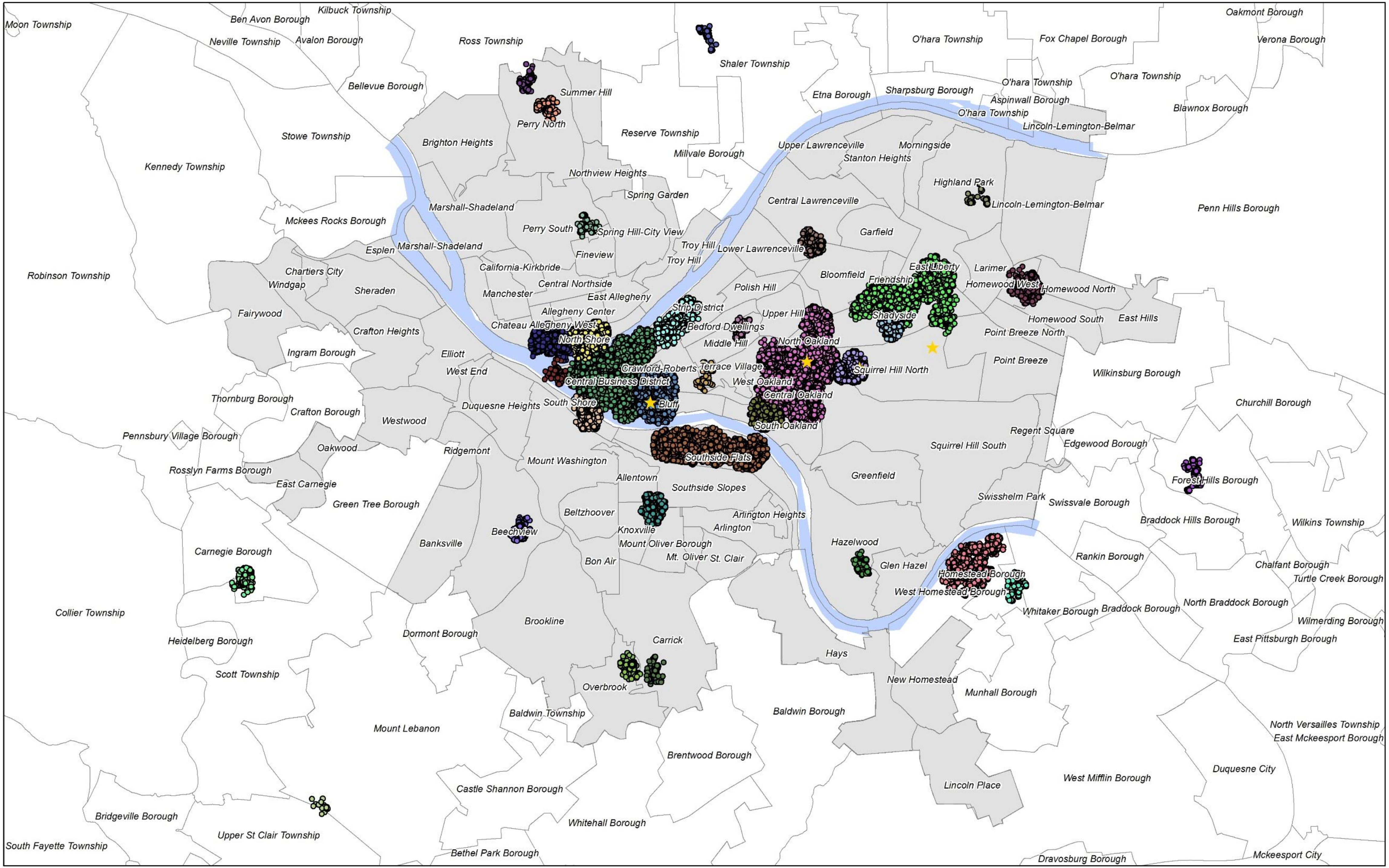
- *Text and emoji could be used to evaluate the tweet sentiment, in turn to quantify the happiness of a city*
- *There is a daily pattern of tweeting. But when people are posting more happy tweets, they are also posting more unhappy tweets.*
- *The temporal pattern of happiness shows peaks at noon and midnight. And people are much happier during weekends than weekdays.*

URBAN AREAS

Identify Pittsburgh "livehoods" from tweet clusters with DBSCAN algorithm

- *40% data set*
- *Define season by day of year [(60,151),(152,243),(244,344),(345,59)]*
- *Minimum points 250 within 0.0015 coordinate distance (~170m)*
- *Minimum cluster of 500 points*





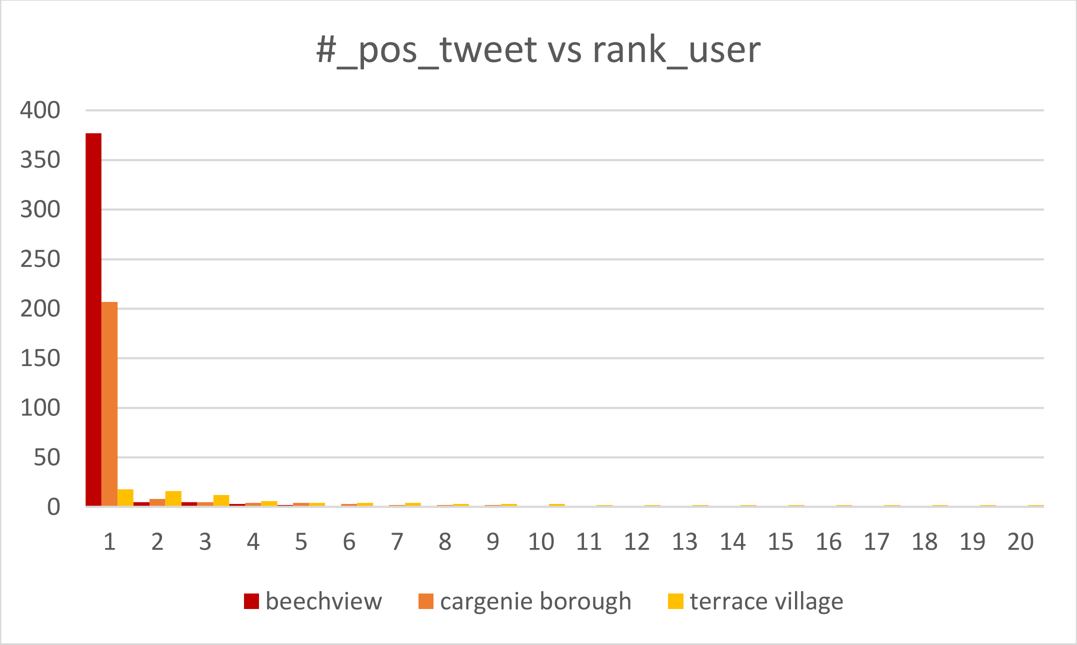
Moon Township
Ben Avon Borough
Kilbuck Township
Neville Township
Avalon Borough
Ross Township
Shaler Township
O'hara Township
Fox Chapel Borough
Oakmont Borough
Verona Borough
Bellevue Borough
Summer Hill
Perry North
Etna Borough
Sharpsburg Borough
O'hara Township
Aspinwall Borough
O'hara Township
O'hara Township
Blawnox Borough
Lincoln-Lemington-Belmar
Stowe Township
Reserve Township
Brighton Heights
Upper Lawrenceville
Morningside
Stanton Heights
Millvale Borough
Kennedy Township
Northview Heights
Spring Garden
Highland Park
Lincoln-Lemington-Belmar
Penn Hills Borough
Marshall-Shadeland
Central Lawrenceville
Mckees Rocks Borough
Perry South
Spring Hill-City View
Garfield
Esplen
Marshall-Shadeland
Troy Hill
Lower Lawrenceville
Fineview
Troy Hill
California-Kirkbride
Bloomfield
East Liberty
Larimer
Robinson Township
Chartiers City
Windgap
Central Northside
Friendship
Homewood West
Homewood North
Sheraden
Manchester
East Allegheny
Polish Hill
Fairywood
Allegheny Center
Strip District
Upper Hill
Shadyside
Homewood South
East Hills
Chateau
Allegheny West
Bedford Dwellings
Crafton Heights
North Shore
Middle Hill
North Oakland
Point Breeze North
Ingram Borough
Elliott
Crawford-Roberts
Terrace Village
Point Breeze
West End
Squirrel Hill North
Wilkinsburg Borough
Central Business District
West Oakland
Thornburg Borough
Central Oakland
Churchill Borough
Crafton Borough
Duquesne Heights
South Shore
Bluff
Westwood
South Oakland
Regent Square
Pennsbury Village Borough
Edgewood Borough
Oakwood
Ridgemont
Squirrel Hill South
Southside Flats
Mount Washington
Rosslyn Farms Borough
Greenfield
Allentown
Forest Hills Borough
East Carnegie
Southside Slopes
Swisshelm Park
Swissvale Borough
Green Tree Borough
Beltzhoover
Arlington Heights
Braddock Hills Borough
Wilkins Township
Arlington
Beechview
Knoxville
Banksville
Mount Oliver Borough
Hazelwood
Carnegie Borough
Mt. Oliver St. Clair
Rankin Borough
Bon Air
Glen Hazel
Chalfant Borough
Homestead Borough
Turtle Creek Borough
West Homestead Borough
North Braddock Borough
Collier Township
Whitaker Borough
Braddock Borough
Wilmerding Borough
Brookline
Dormont Borough
Heidelberg Borough
Carrick
East Pittsburgh Borough
Hays
Scott Township
Overbrook
New Homestead
Munhall Borough
Baldwin Township
Baldwin Borough
Mount Lebanon
North Versailles Township
East Mckeesport Borough
Brentwood Borough
Duquesne City
Castle Shannon Borough
West Mifflin Borough
Lincoln Place
Bridgeville Borough
Whitehall Borough
Upper St Clair Township
South Fayette Township
Bethel Park Borough
Mckeesport City
Dravosburg Borough

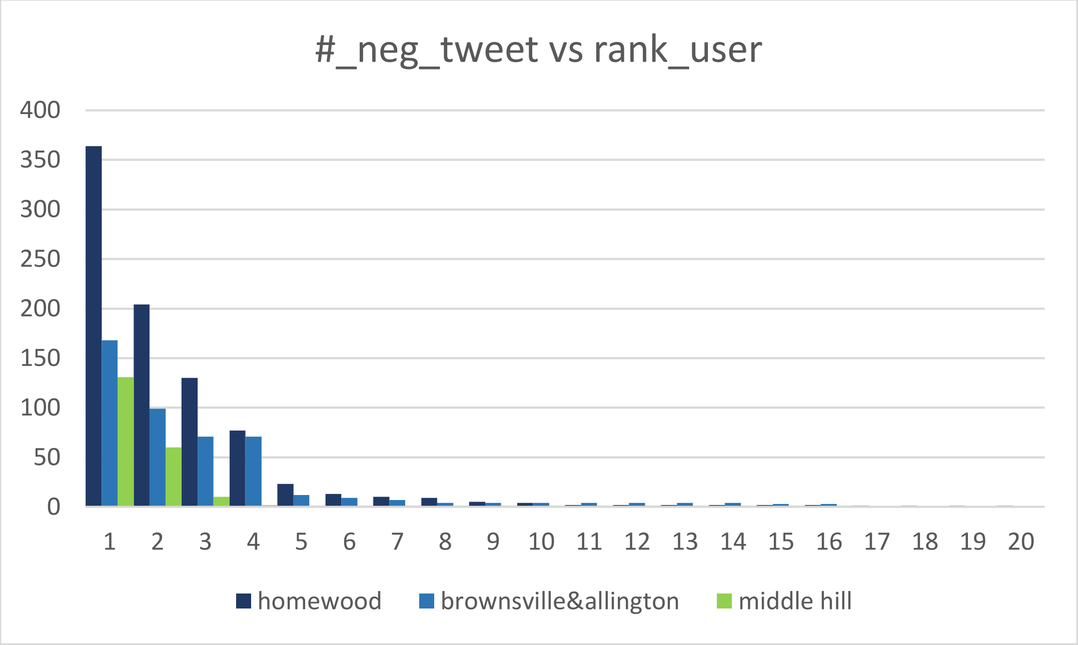
cluster	name	#_tweet	cluster	name	#_tweet	cluster	name	#_tweet	cluster	name	#_tweet
0	downtown	40707	8	homewood	3787	16	beechview	2025	24	carnegie borough	1257
1	oakland	34123	9	homestead-waterfront	3643	17	upper st clair township-mount lebanon	1890	25	foresthills borough	1203
2	southside	17343	10	pittsburgh tech center	2995	18	terrace village	1768	26	homestead	1170
3	console energy center	14483	11	cmu	2513	19	lawrenceville-liberty & bloomfield	1742	27	middle hill	1060
4	northside-pnc park	11300	12	southside-brownsville & allington	2362	20	shaler township	1606	28	hazelwood	1000
5	northside-heinz field	8274	13	overbrook	2280	21	perry north-ross township	1487	29	carrick	955
6	shadyside-east liberty	6821	14	station square	2278	22	shadyside-walnut	1348	30	highland park	919
7	strip district	4731	15	point state	2094	23	perry north	1266	31	perry south	527

If we simply consider all tweets of each cluster…

pos_%_top3	***beechview***	***carnegie borough***	station square
	20.1	19.41	18.66
pos_%-neg_%_top3	***terrace village***	***carnegie borough***	shadyside-walnut
	11.71	10.34	10.16
senti_avg_top3	***carnegie borough***	shadyside-walnut	***terrace village***
	0.21	0.18	0.18
neg_%_top3	***homewood***	***southside-brownsville & allington***	perry north
	22.92	21.34	19.43
pos_%-neg_%_bottom3	***homewood***	***southside-brownsville & allington***	***middle hill***
	-11.78	-9.44	-9.15
senti_avg_bottom3	***homewood***	***southside-brownsville & allington***	***middle hill***
	-0.28	-0.25	-0.22

However these clusters just seem to be happier/unhappier because of several users…





If the user posts more happy/unhappy tweets than 90th percentile users who post happy/unhappy tweets, the weight of his/her tweets in happy/unhappy tweet groups should be devalued.
The weight of happy/unhappy tweets of each user is defined as…

$$w_{user} = \begin{cases} \exp\left((tweet_{90th\ per} - n\_tweet)\right] & \text{n}\_tweet > tweet_{90th\ per} \\ 1 & else \end{cases}$$

pos_%_top3	***station square***	***terrace village***	***strip district***
	16.7103	14.256	13.1334
pos_%-neg_%_top3	***terrace village***	***station square***	***strip district***
	10.27	9.4921	7.7766
senti_avg_top3	***terrace village***	***station square***	***strip district***
	0.1573	0.1348	0.1085
neg_%_top3	***perry south***	upper st clair township-mount lebanon	***highland park***
	7.9506	6.5126	6.2006
pos_%-neg_%_bottom3	***perry south***	***highland park***	***perry north***
	-0.9076	-0.5091	-0.3027
senti_avg_bottom3	point state	***perry south***	***highland park***
	-0.203	-0.0231	-0.0134

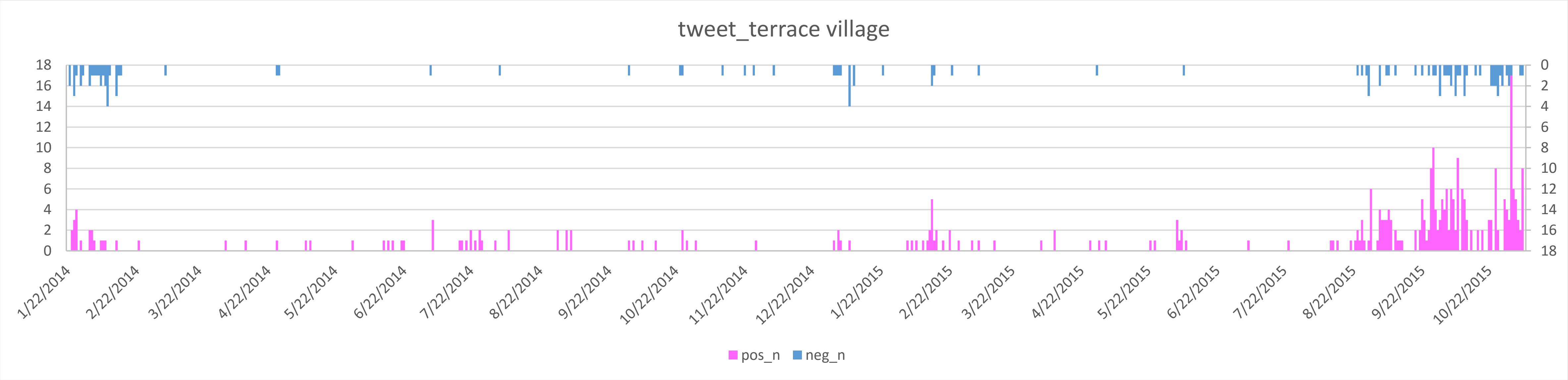
T-test on average cluster tweet sentiment with significant level of 0.1 (sample size=500, iteration=10)

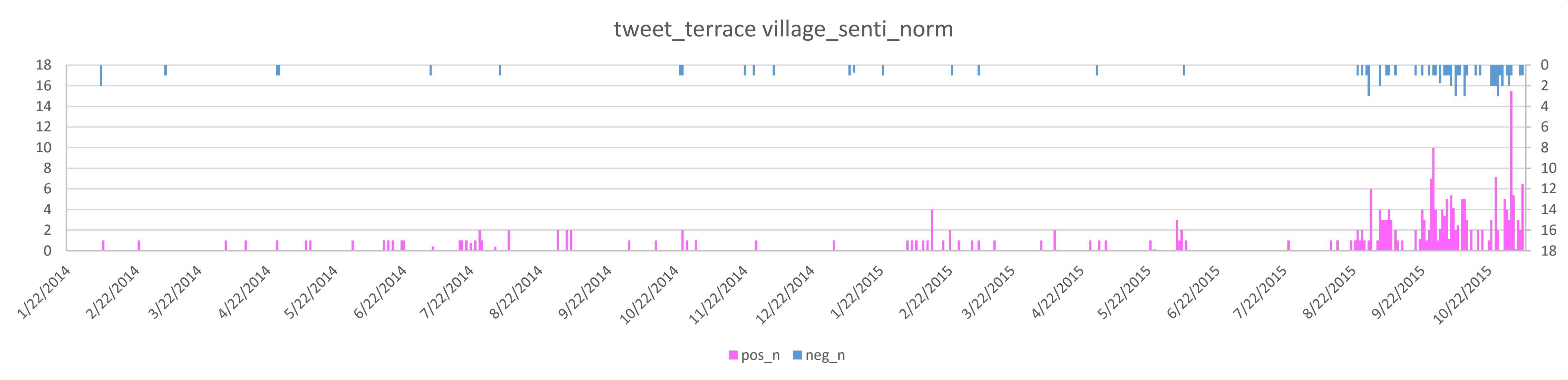
- *“Y” means the null hypothesis is accepted; “N” means the null hypothesis is rejected*
- *a simple statistical test of the difference of happiness among clusters*

(following table shows t-test results of top/middle/bottom clusters ordered by average tweet sentiment)

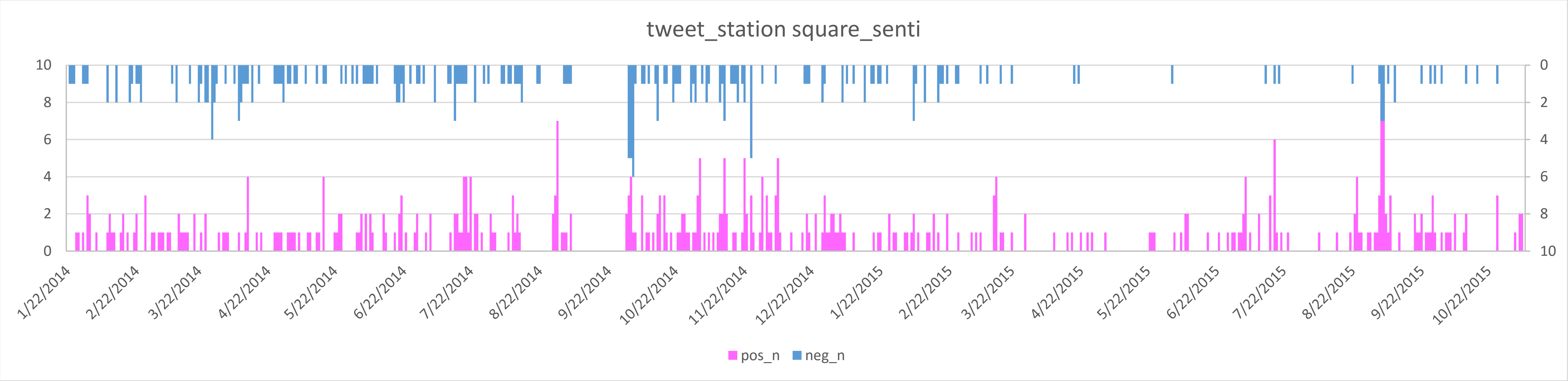
		14	7	13	6	28	16	19	26	30	31	15
18	terrace village	N	N	N	N	N	N	N	N	N	N	N
14	station square		Y	N	N	N	N	N	N	N	N	N
7	strip district			N	N	N	N	N	N	N	N	N
13	overbrook				N	N	N	N	Y	Y	N	N
6	shadyside-east liberty					N	N	Y	N	N	N	N
28	hazelwood						N	N	N	N	N	N
16	beechview							Y	N	N	N	N
19	lawrenceville-liberty & bloomfield								N	N	N	N
26	homestead									Y	N	N
30	highland park										N	N
31	perry south											N
15	point state											

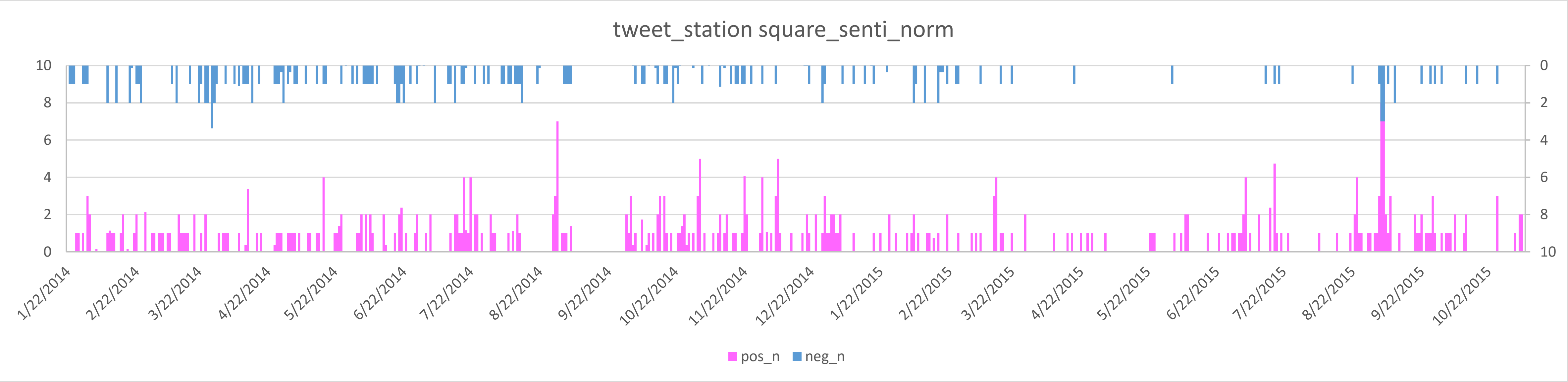
Temporal tweet sentiment of the happiest and least happy clusters, equal vs unequal weighted…



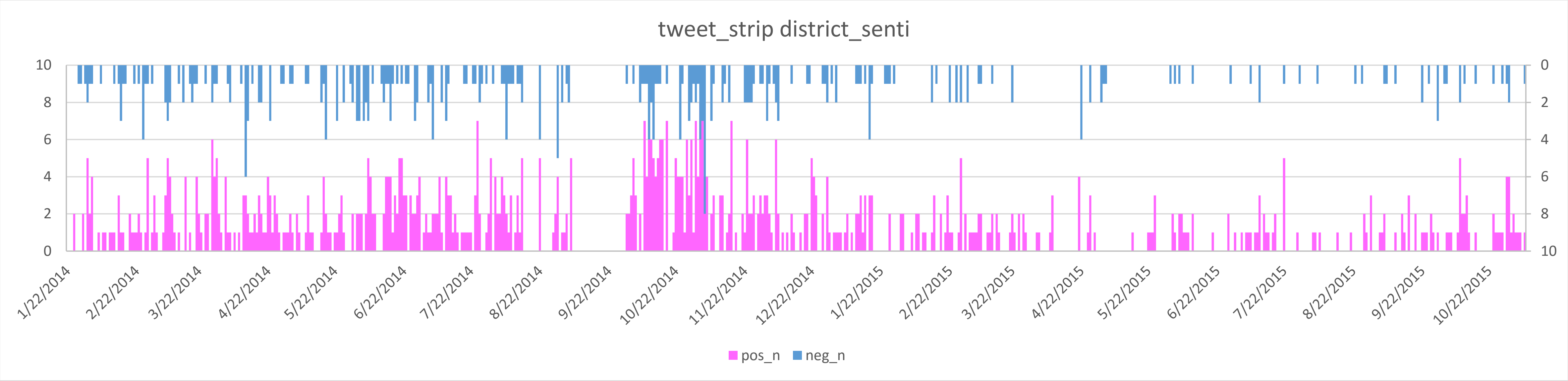


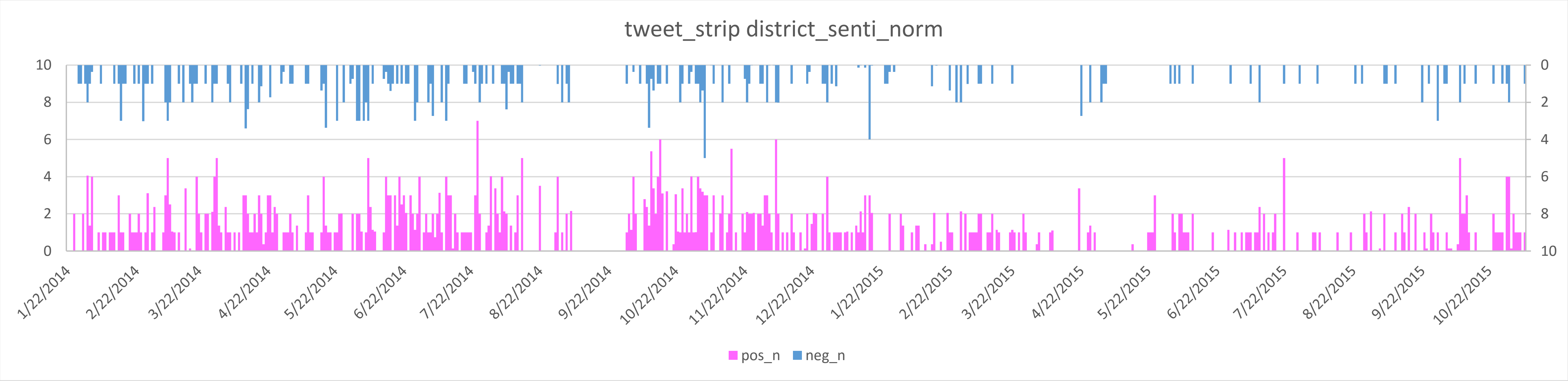
Temporal tweet sentiment of the happiest and least happy clusters, equal vs unequal weighted…



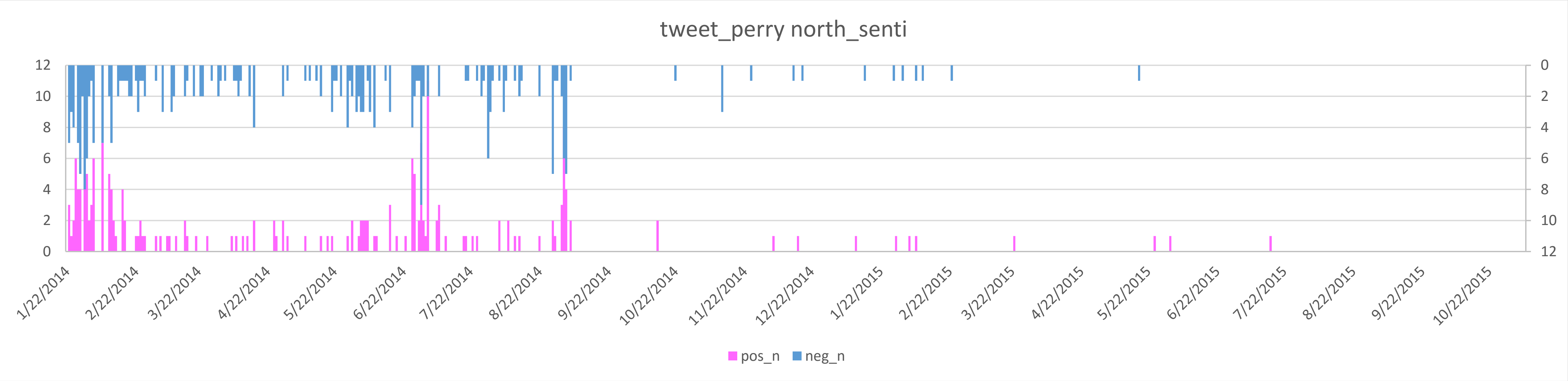


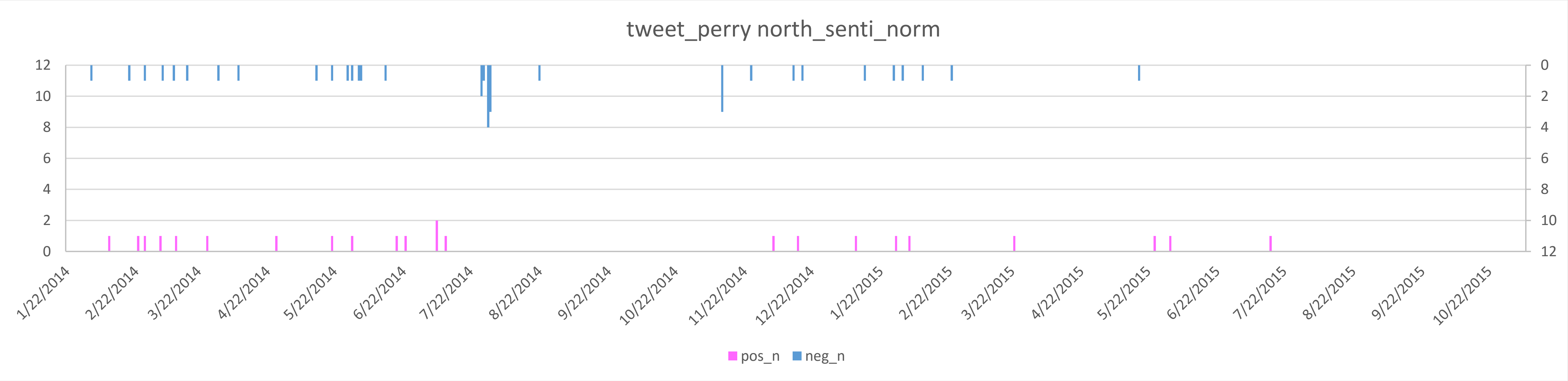
Temporal tweet sentiment of the happiest and least happy clusters, equal vs unequal weighted…



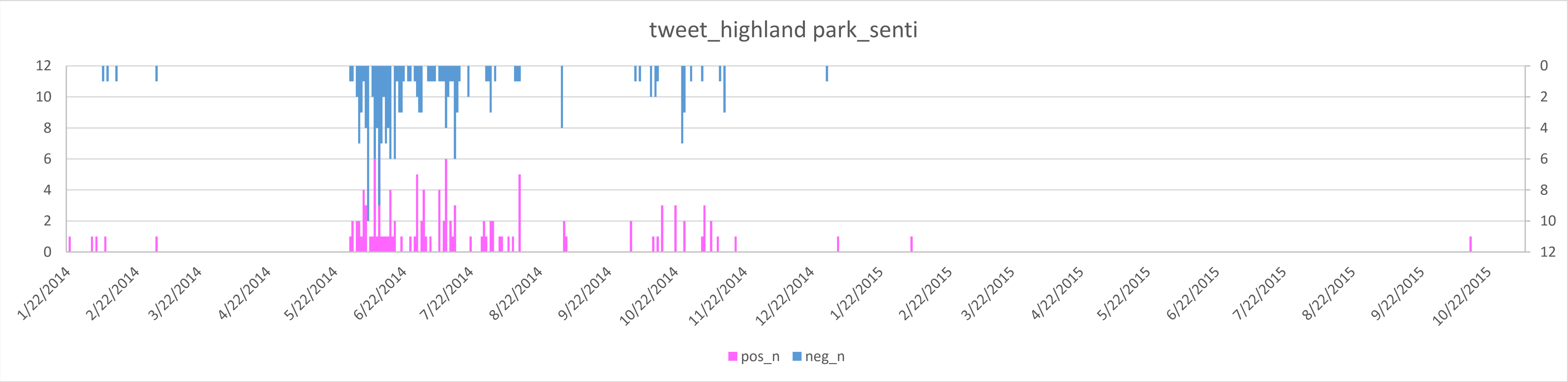


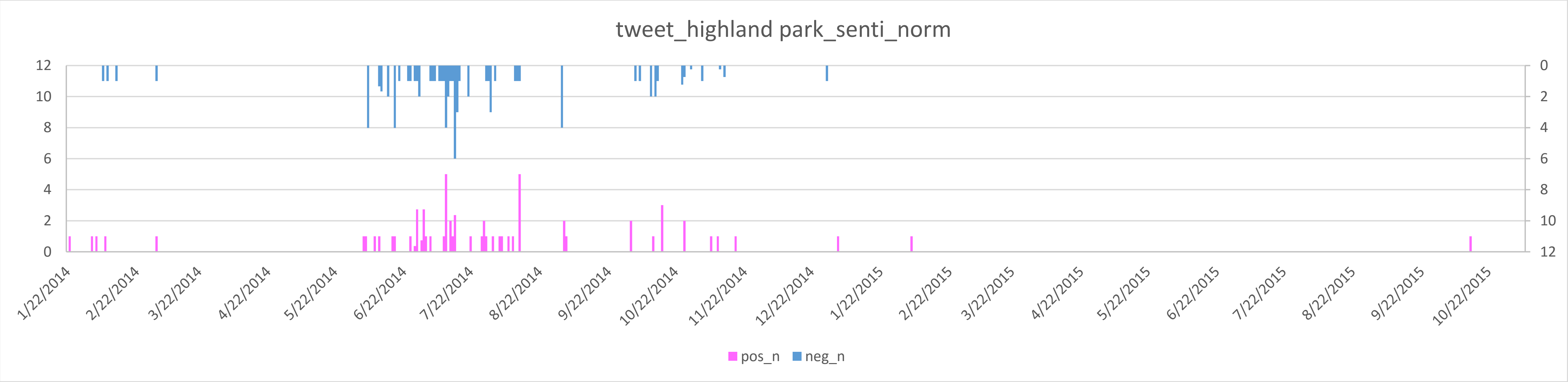
Temporal tweet sentiment of the happiest and least happy clusters, equal vs unequal weighted…



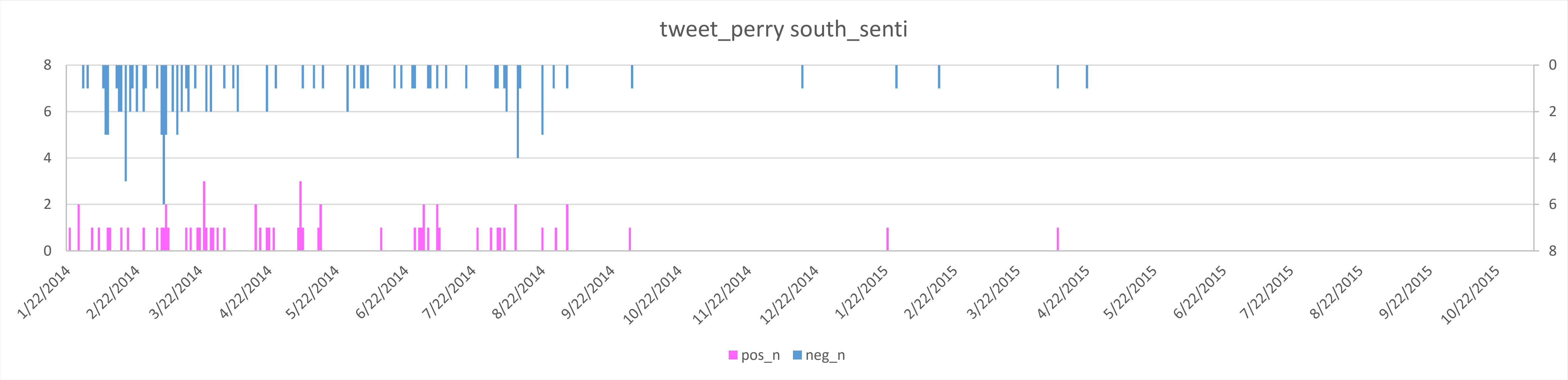


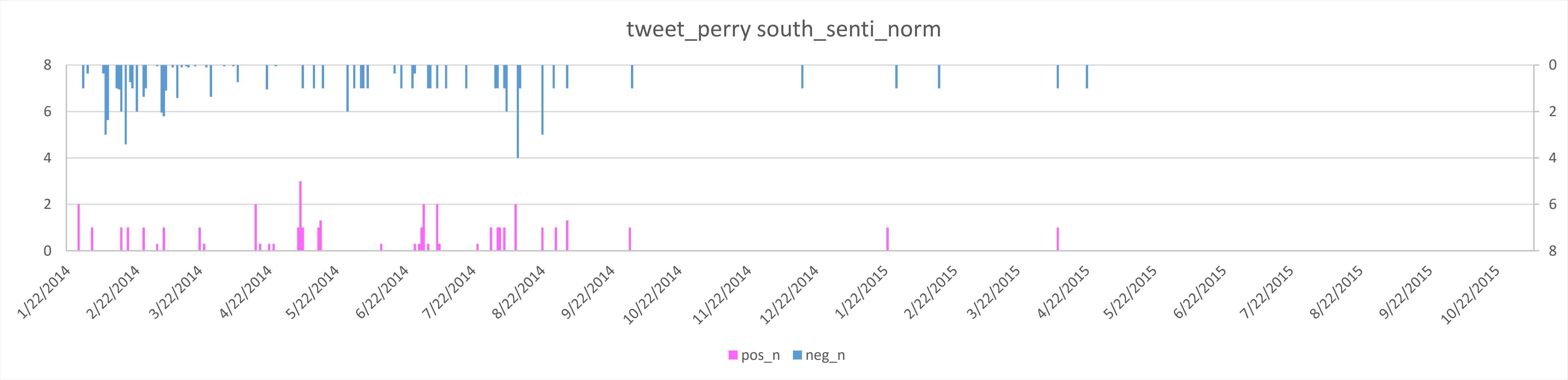
Temporal tweet sentiment of the happiest and least happy clusters, equal vs unequal weighted…





Temporal tweet sentiment of the happiest and least happy clusters, equal vs unequal weighted…

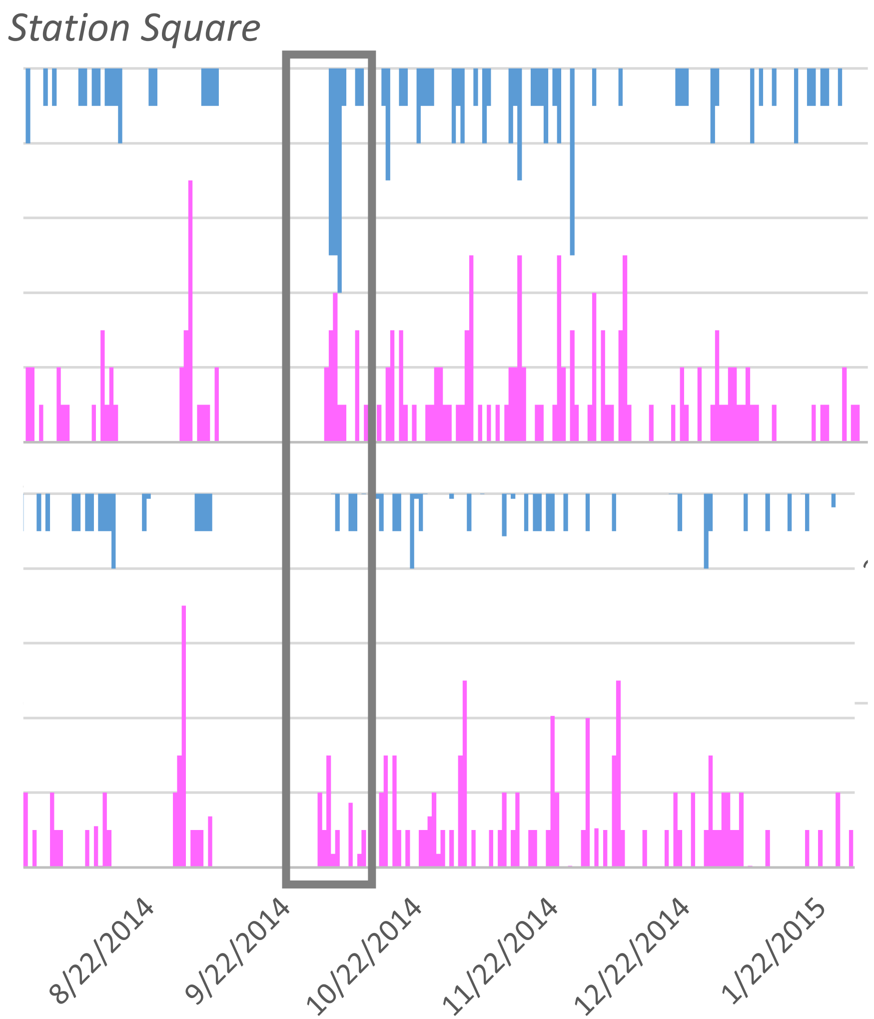




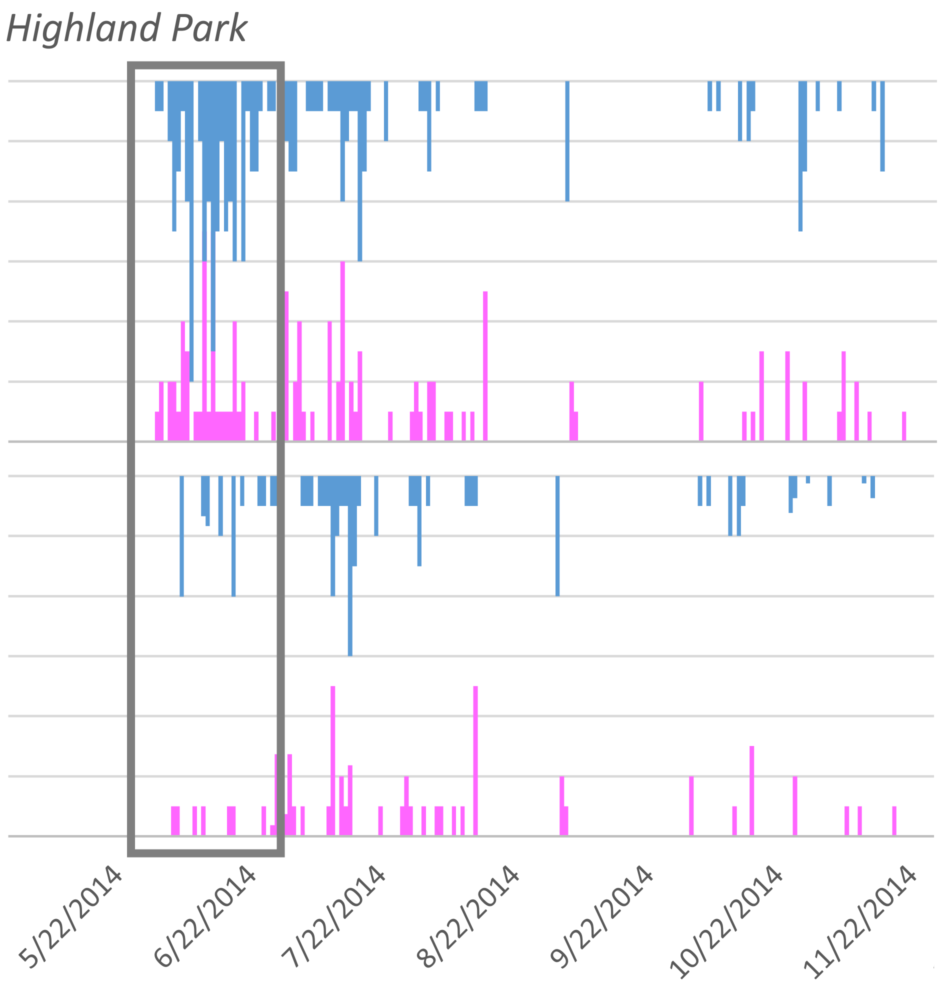
Some findings:

- *For small clusters (e.g. Terrace Village, Perry North, Perry South and Highland Park), the tweet sentiment is much more event-related, while larger clusters (e.g. Station Square and Strip District) show some temporal patterns.*
- *Several users do have impacts on the overall sentiment level, especially for small clusters.*
- *Case study: some “unhappy” days for Station Square and Highland Park…*

2015-10-02	BelowThaaBelt_	He pissing me off
	BelowThaaBelt_	I ain't had my coffee this morning...so I'm half dead rn
	BelowThaaBelt_	I am too wtf
	BelowThaaBelt_	Ion tell my boys shit bout no problems. They TOO hot already
	BelowThaaBelt_	I was mad af sending that text
	jimmyb304	My shit right herrrr Bar Louie Station Square
2015-10-03	BelowThaaBelt_	Can I not like you for a while ?
	BelowThaaBelt_	Why do you worry? You know I'm the same.
	BelowThaaBelt_	Bad bitches is the only thing that I like
	BelowThaaBelt_	You lazy bittches is fucking up the economy
	BelowThaaBelt_	Hate when get too attached to me
2015-10-04	BelowThaaBelt_	Fuck around & act like I NEVER met to ass
	Mrstevecooper	Blah blah blah. No one cares bitch. Shut up.
	Mrstevecooper	*throws chair at the bitch in Human Relations*
	Mrstevecooper	Imma cut you.
	Mrstevecooper	I seriously hate using my MasterCard. Creditcard shit sucks.
	Zizi_Softtail	Gonna chill with furries and shit this weekend



2015-06-10	311WestOtterman	Lil b blows dead"
	311WestOtterman	Or nah o rly fag"
	311WestOtterman	sorry my dad wasn't Arthur yeah he wad a pedophile"
	311WestOtterman	im pissed
	311WestOtterman	imma be on clay ave sellin dis lawwwwwng dick
	311WestOtterman	not much left to do but drowned
2015-06-12	311WestOtterman	fuck toby & smokey bear
	311WestOtterman	bored txt me
	311WestOtterman	Why should I even waste my breath anymore.. fwm"
	311WestOtterman	GUYS WHO IS ALIANDO??? DONT IGNORE ME PLS what the fuck!
	311WestOtterman	Do u rlly no dick bukkis"
	311WestOtterman	on &on, reckless abandon
	311WestOtterman	hannah, you're actually really dumb
	XoNyny	Esha be talking shit
2015-06-13	XoNyny	I forever calls esha miserable
	311WestOtterman	i fight animals
	TylerNolder	fuck this rain
	311WestOtterman	look at snap chat dead"
	311WestOtterman	kevin harts getting annoying
2015-06-15	311WestOtterman	World Cup > baseball gayest shit ive seen all night"
	311WestOtterman	no well cry about it no"
	311WestOtterman	no well cry about it no k" hockeys on bye"
	311WestOtterman	shinners :c
	311WestOtterman	Gbg is probably the most lamest city ever fwm"



Some findings:

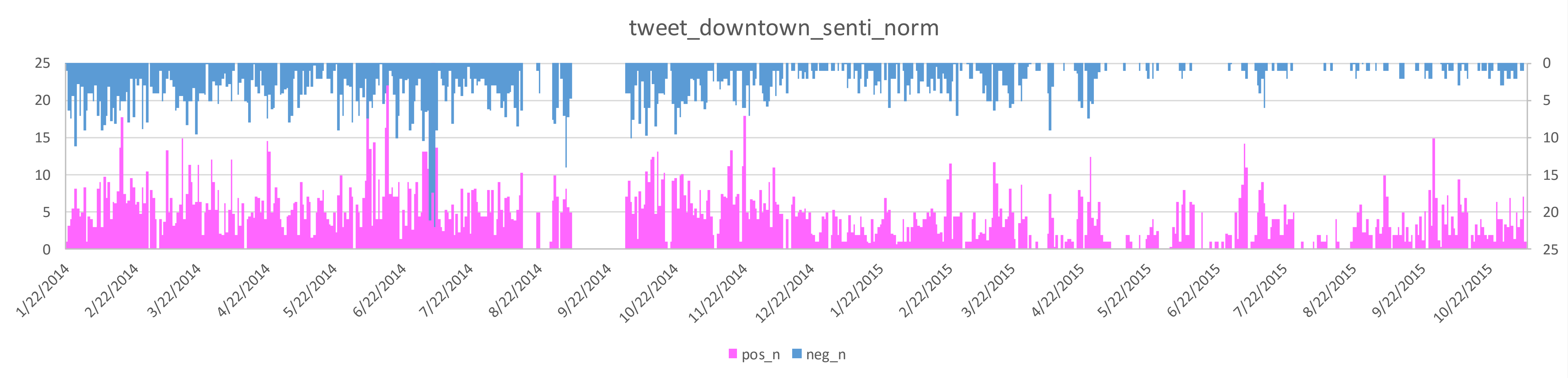
- *'Livehoods' in a city could identified by clustering tweets.*
- *Happiness is different among urban areas, and could be highly affected by a small group of users.*
- *Devaluing users who post much more happy/unhappy tweets than others in each cluster could help get a better understanding of the overall wellbeing.*

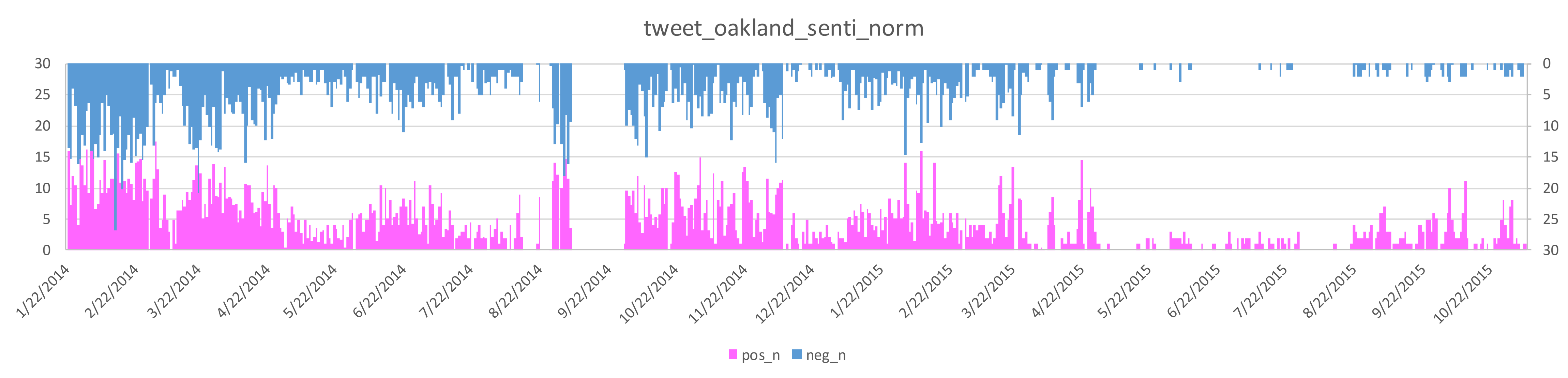
Questions:

- *Is there any sampling issue that causes this problem?*
- *Is tweet or user being randomly chosen?*

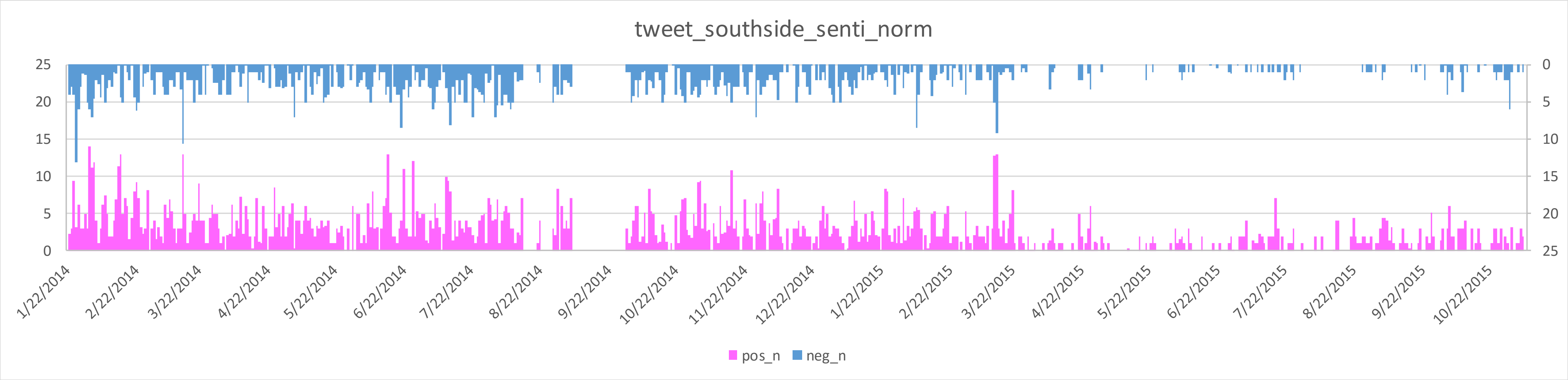
Considering the lack of data and user effect, the following analysis focuses on large clusters.

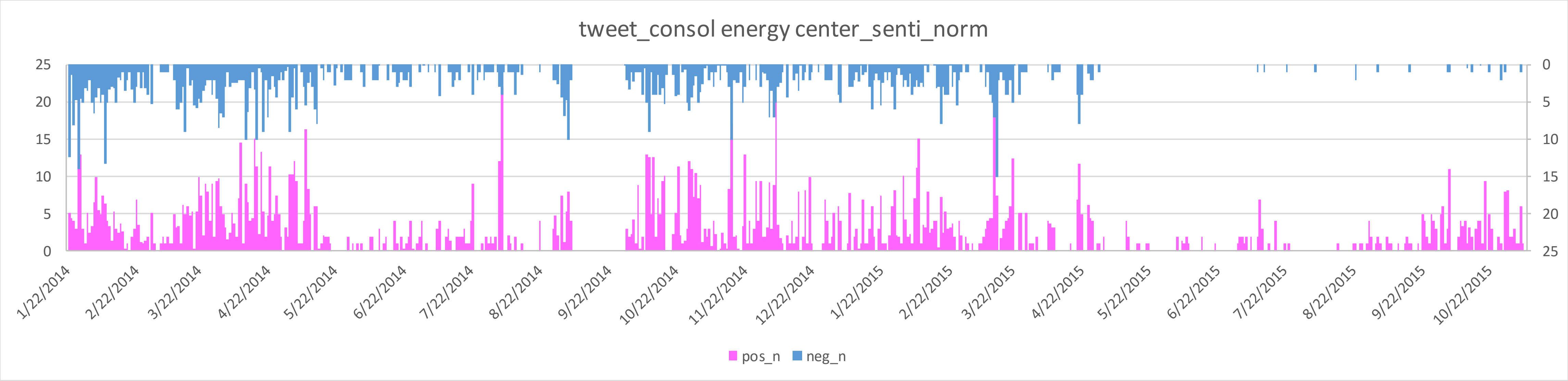
Temporal tweet sentiment considering user-weight…



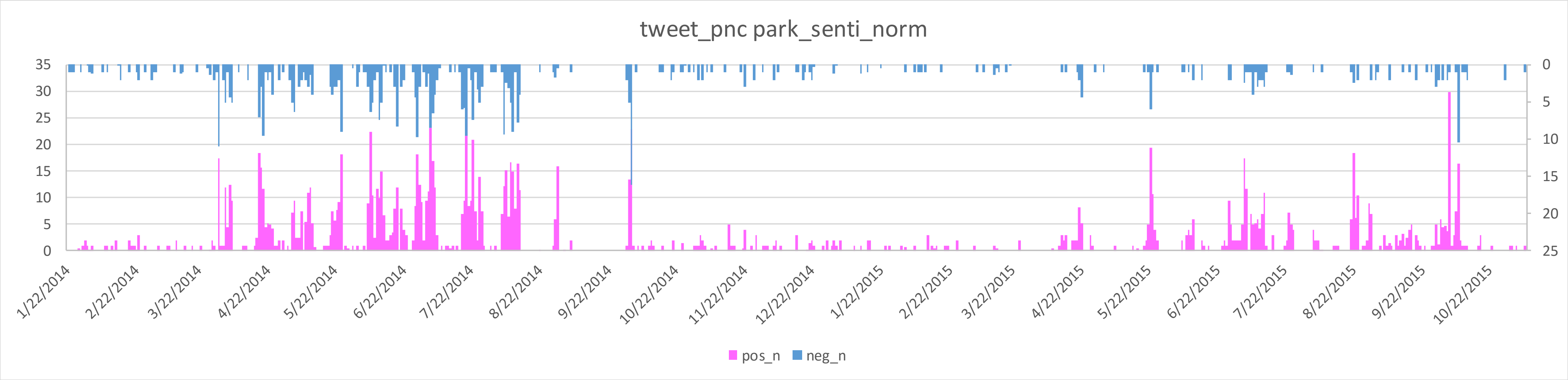


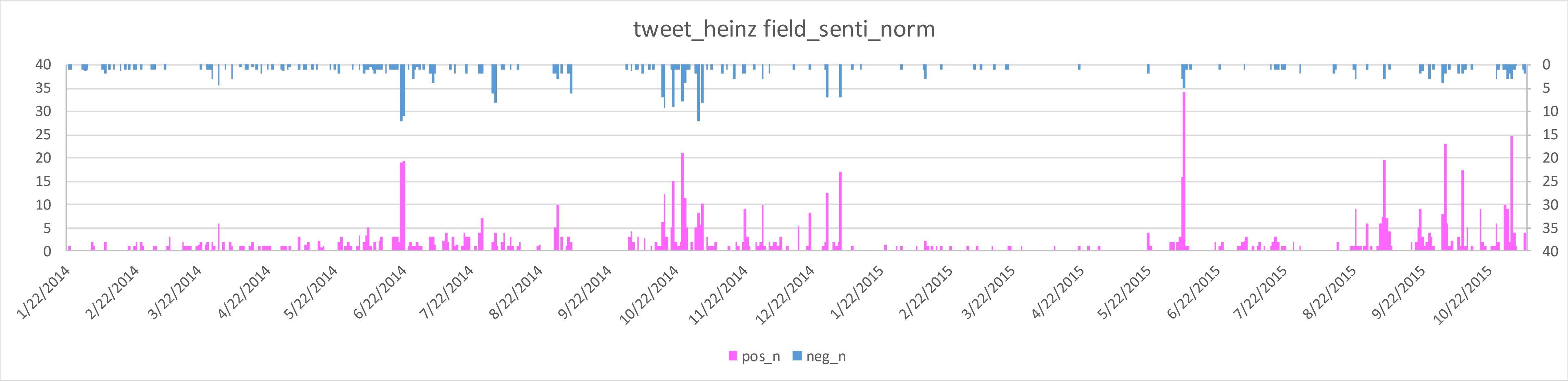
Temporal tweet sentiment considering user-weight…



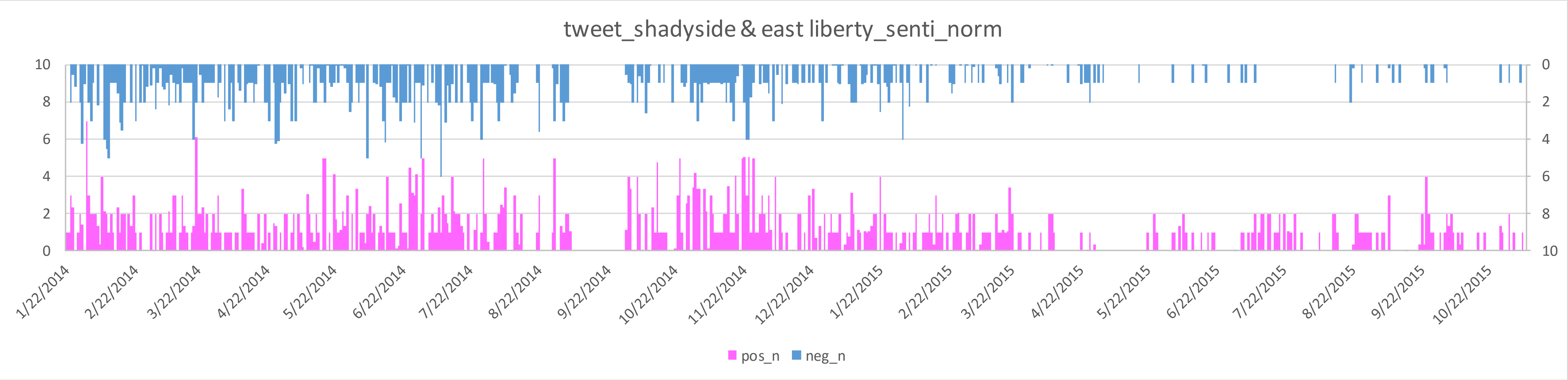


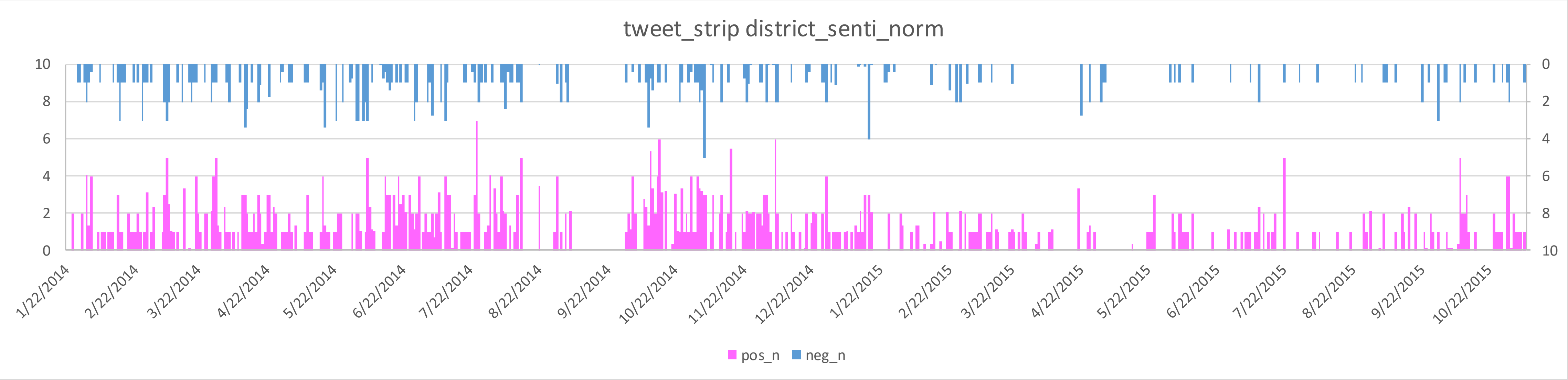
Temporal tweet sentiment considering user-weight…





Temporal tweet sentiment considering user-weight…





A simple classification based on the temporal sentiment pattern…

Event-based — *e.g. PNC Park, Heinz Field*

Hybrid — *e.g. Downtown, Oakland, Southside, Console Energy Center*

Function-based — *e.g. Shadyside & East Liberty, Strip District*

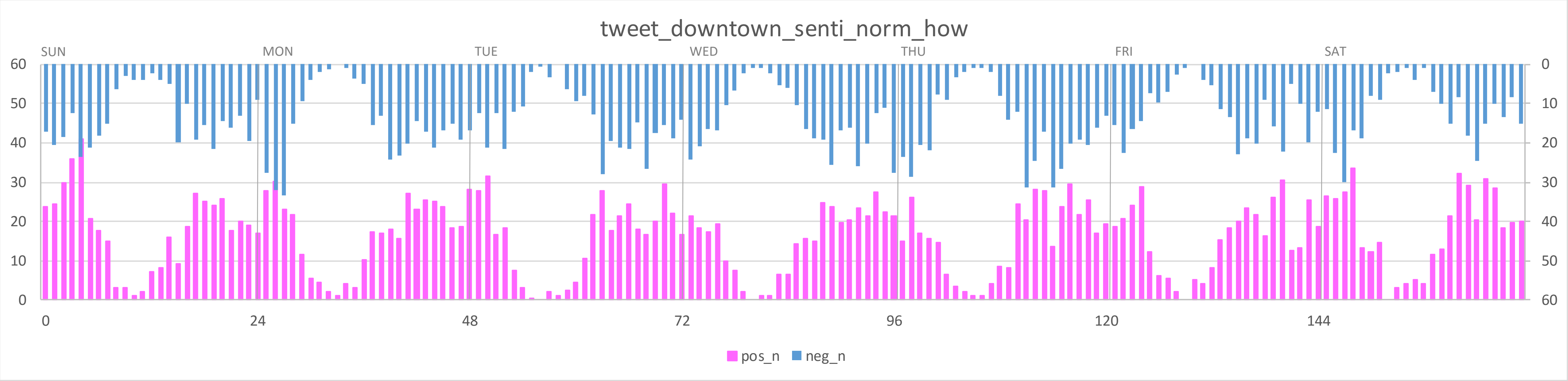
⇨

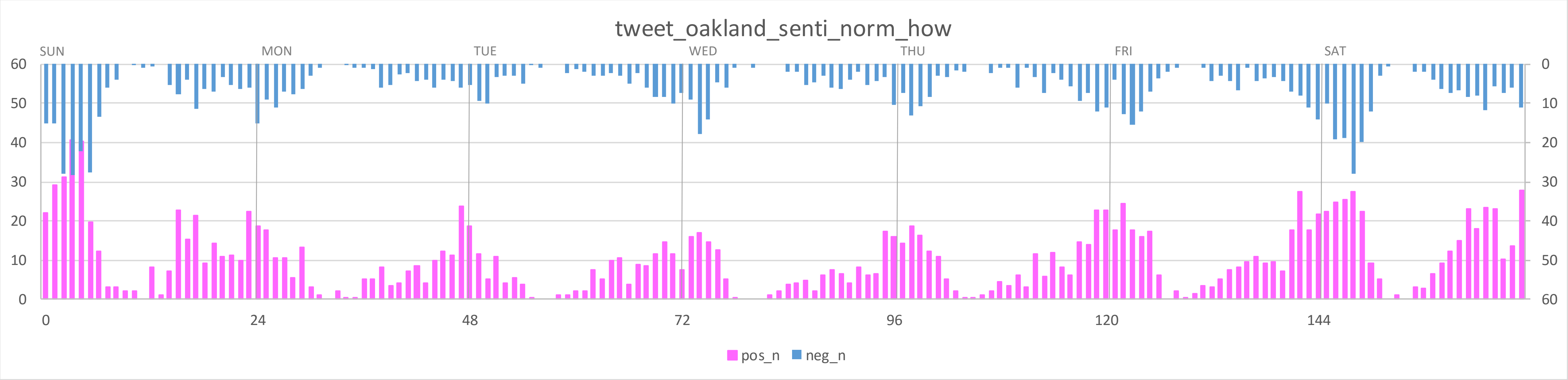
Event Detection

Urban Function

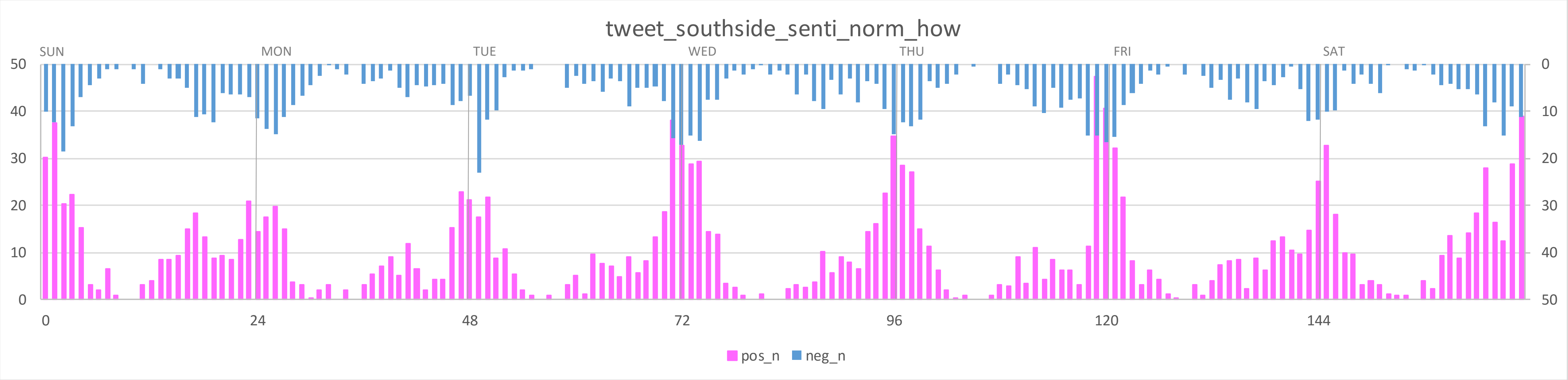
Emotion Prediction

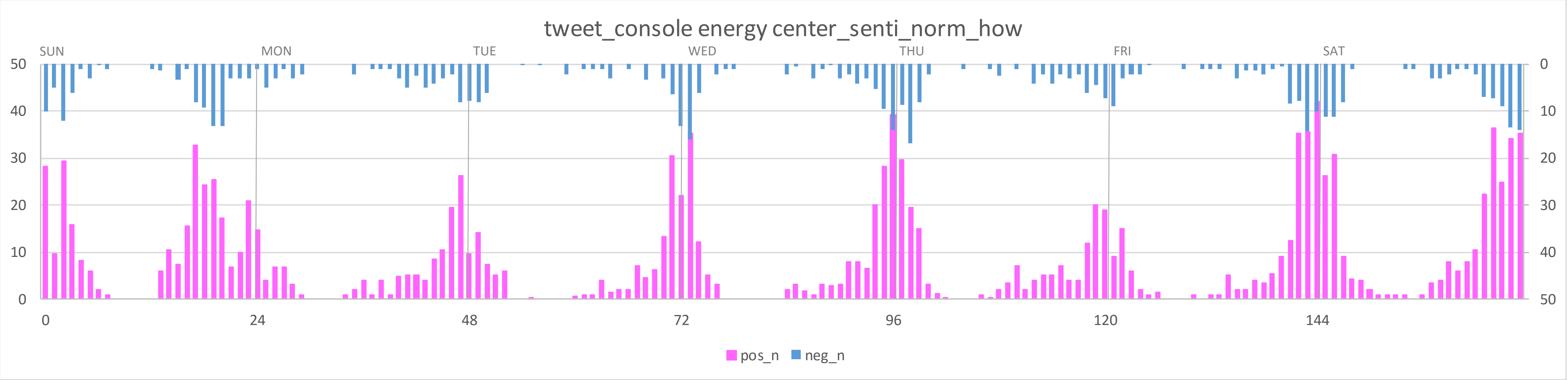
Weekly tweet sentiment patterns considering user-weight…



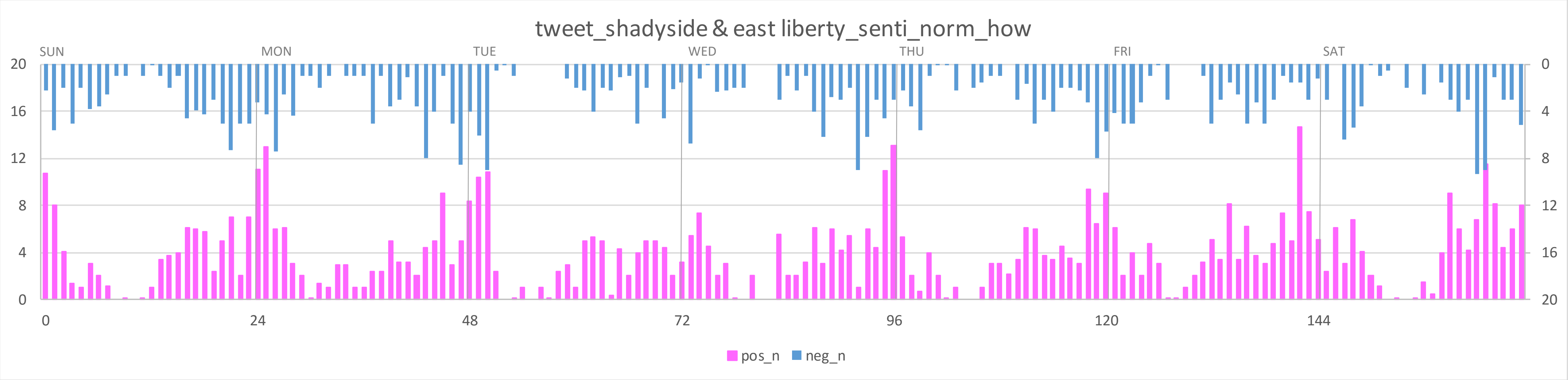


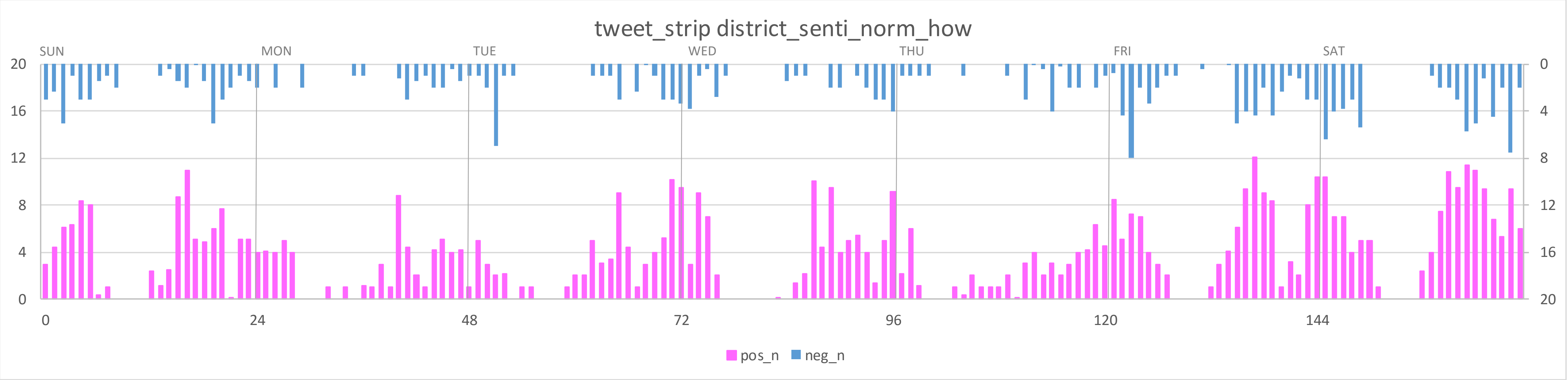
Weekly tweet sentiment patterns considering user-weight…





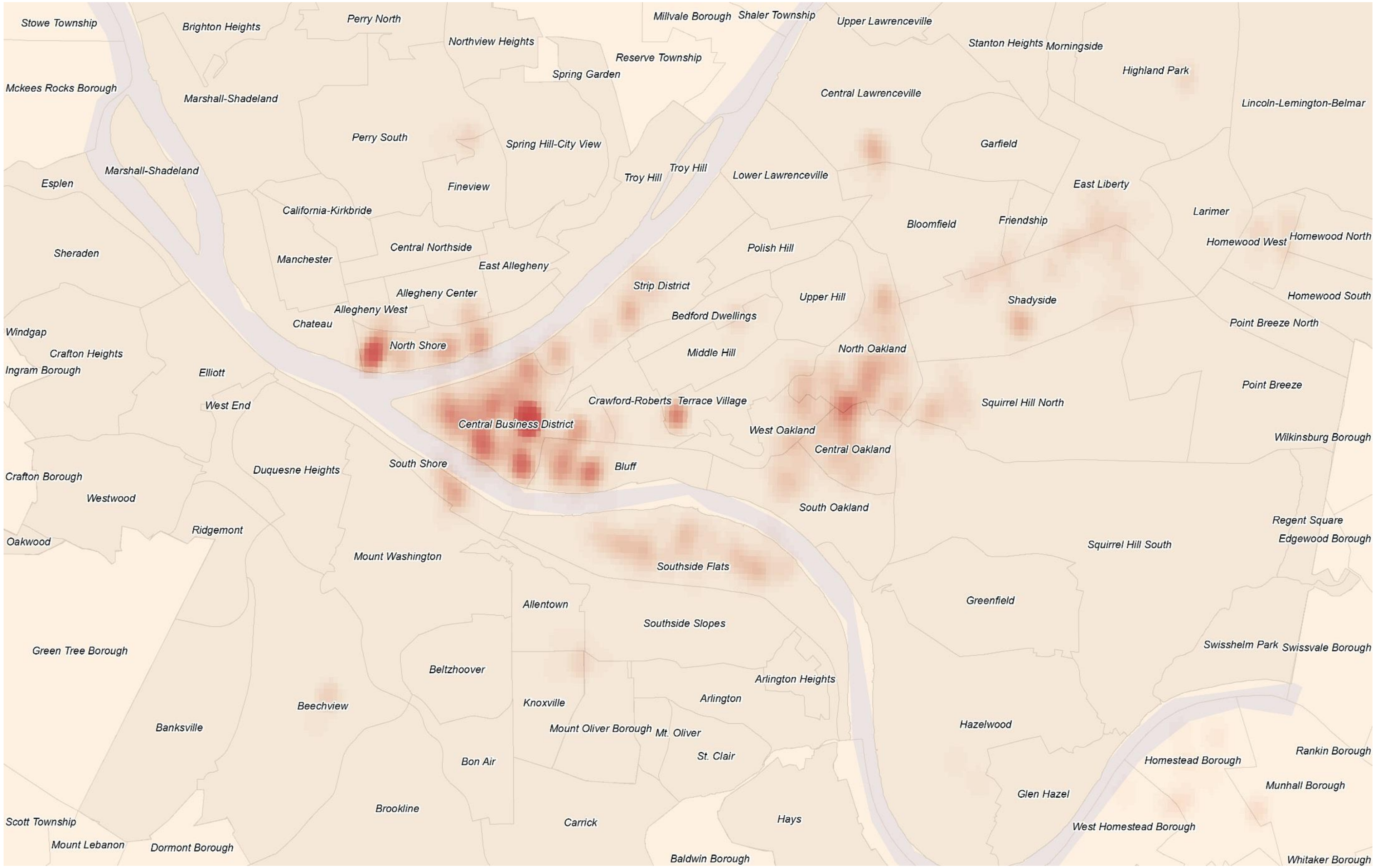
Weekly tweet sentiment patterns considering user-weight…





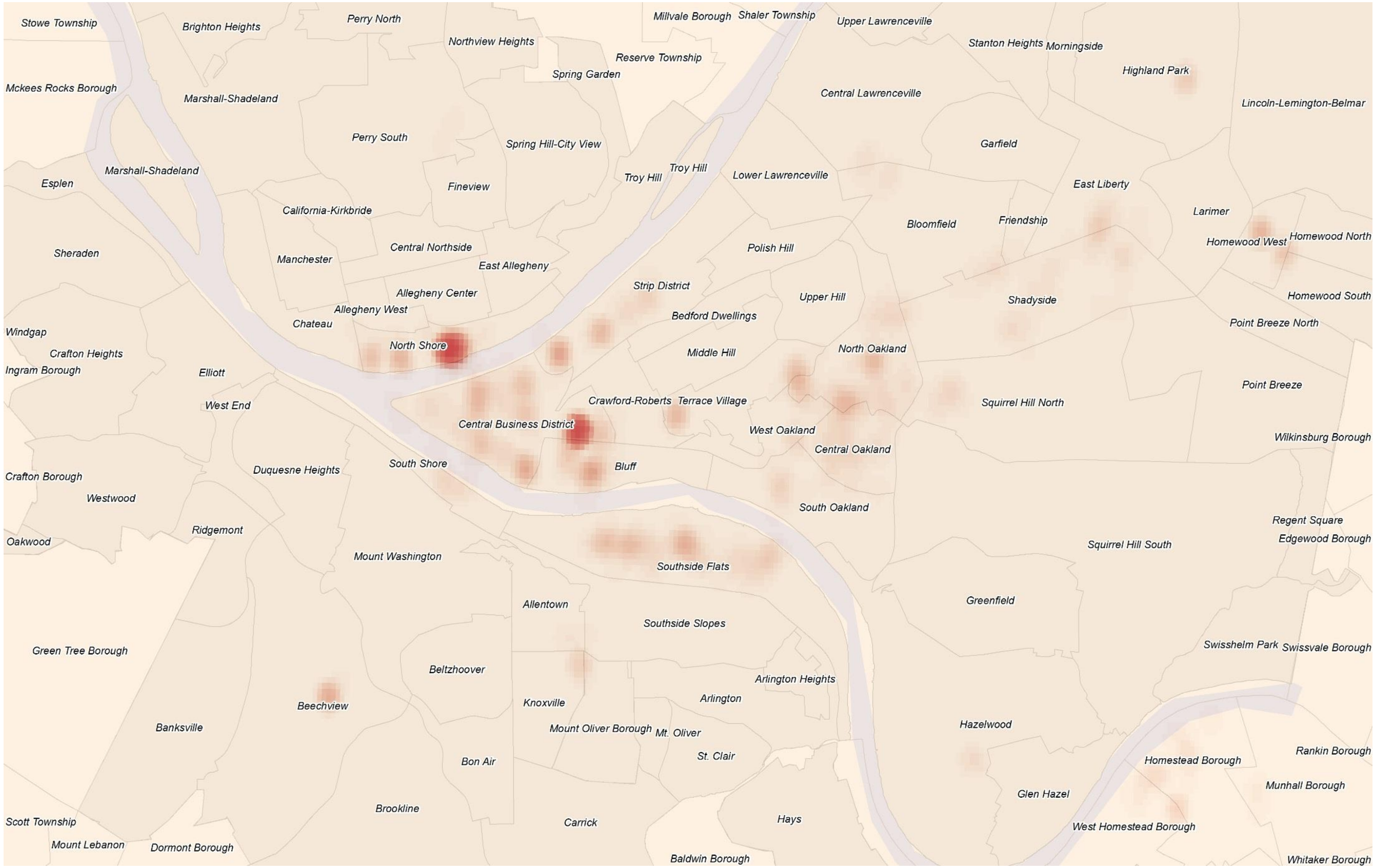
Heat-maps showing spatial-temporal patterns of happy/unhappy tweets…

- *Weekday midday happy tweets*



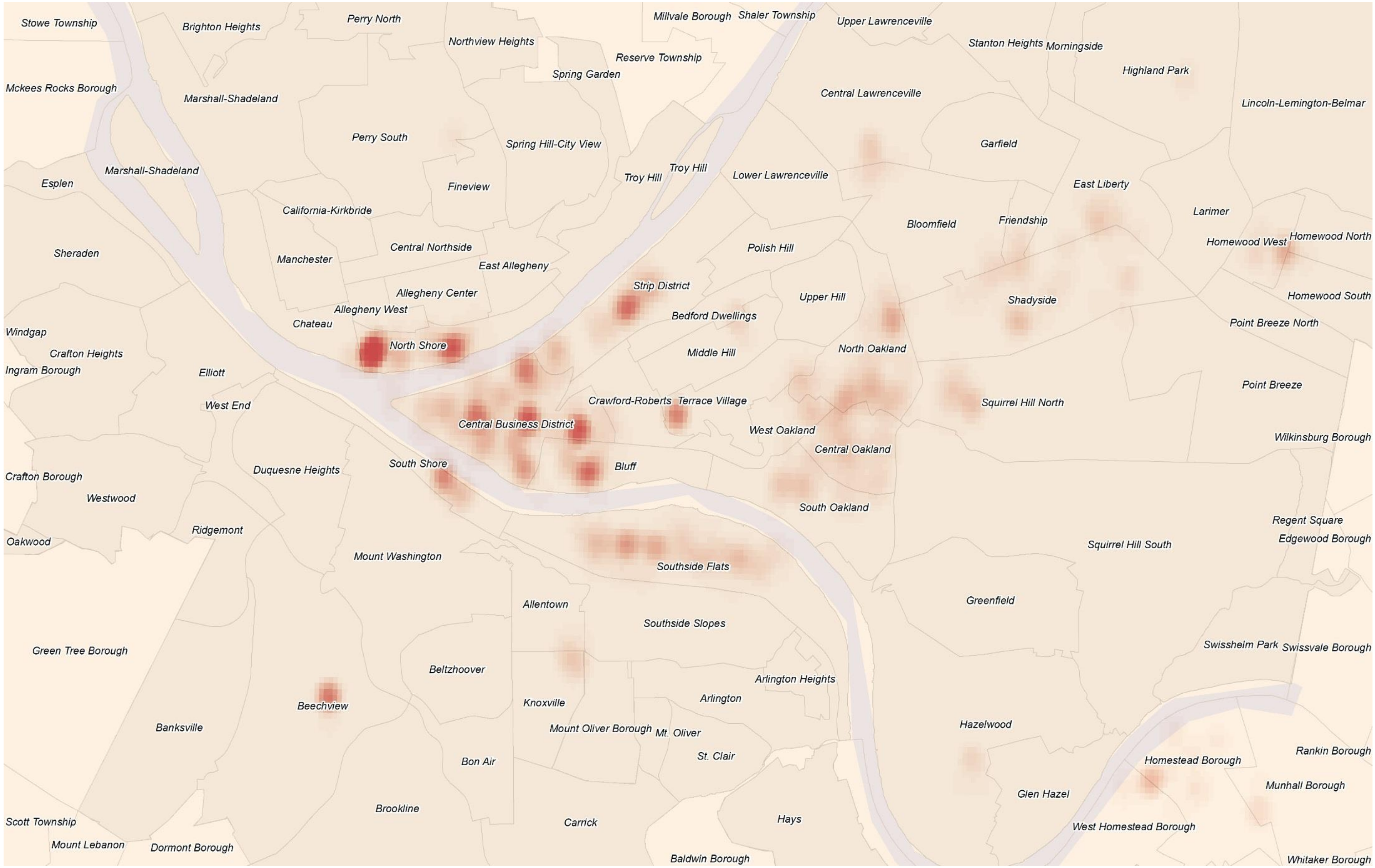
Heat-maps showing spatial-temporal patterns of happy/unhappy tweets…

- *Weekday midnight happy tweets*



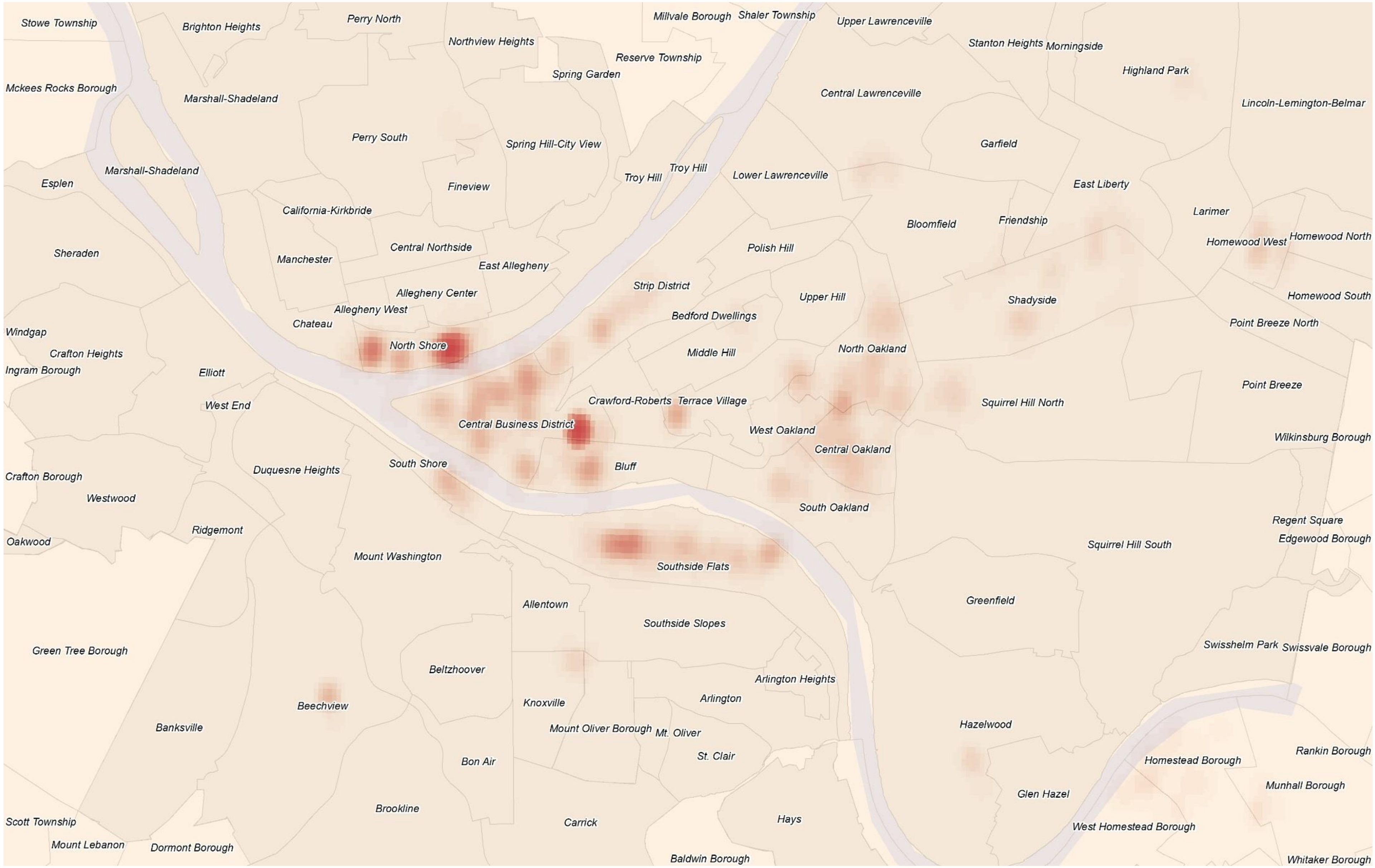
Heat-maps showing spatial-temporal patterns of happy/unhappy tweets…

- *Weekend afternoon happy tweets*



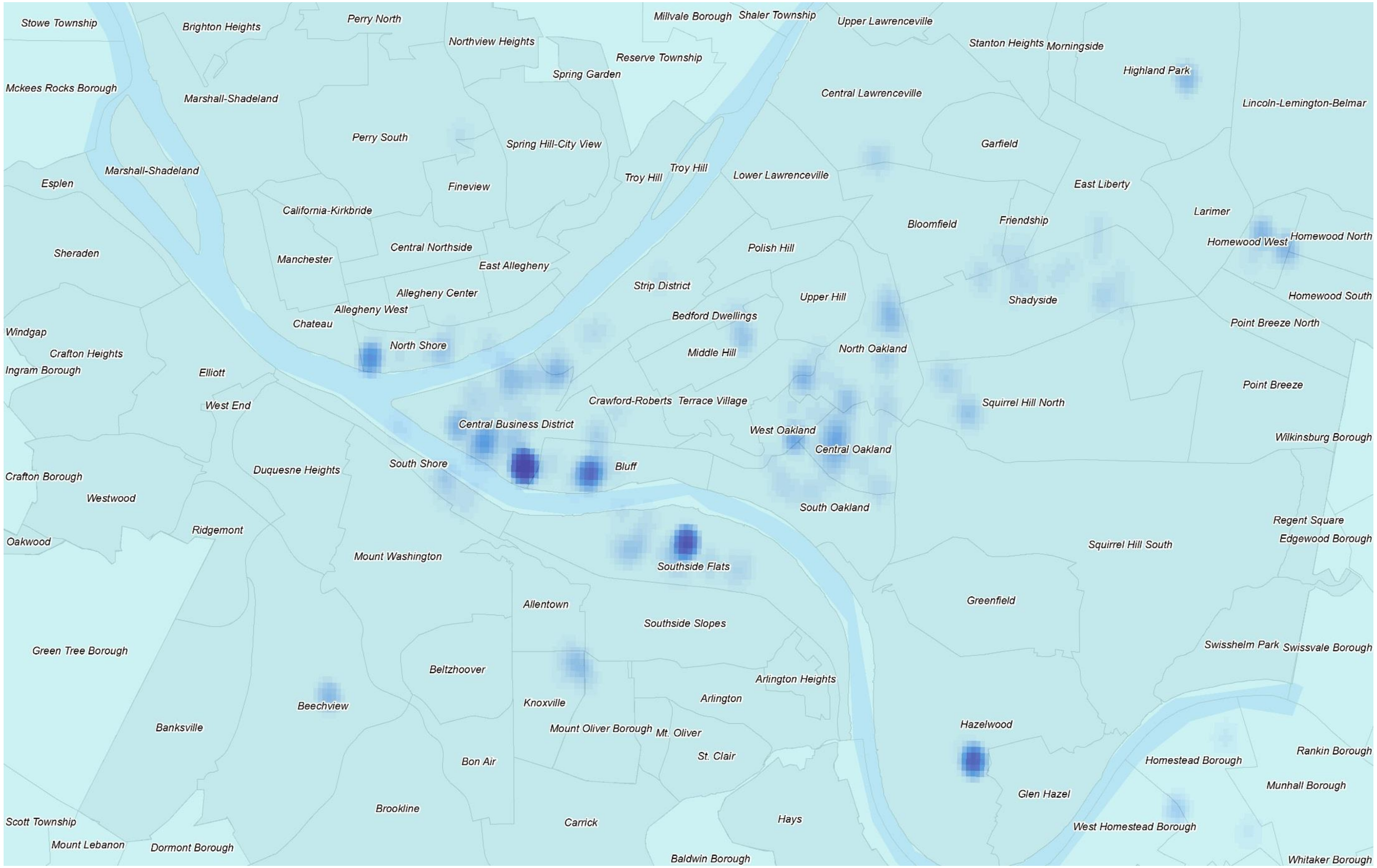
Heat-maps showing spatial-temporal patterns of happy/unhappy tweets…

- *Weekend midnight happy tweets*



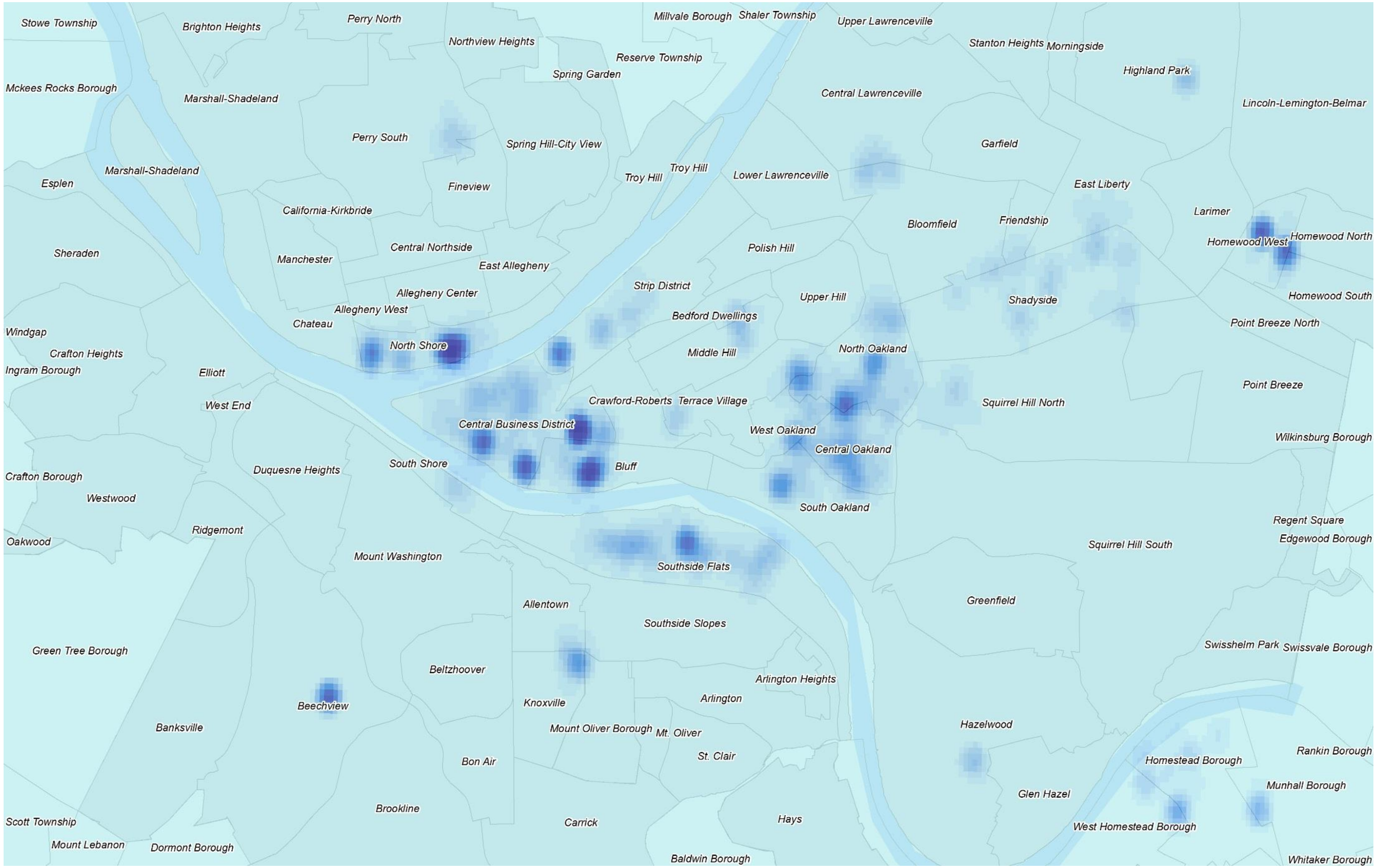
Heat-maps showing spatial-temporal patterns of happy/unhappy tweets…

- *Weekday morning unhappy tweets*



Heat-maps showing spatial-temporal patterns of happy/unhappy tweets…

- *Weekday midnight unhappy tweets*



Some findings:

- *Urban areas could be classified based on their temporal patterns of happiness. The urban happiness is either event-related or function-related.*
- *Further analysis on the happiness spikes and patterns could help the study of event detection and emotion prediction.*
- *These temporal patterns also give us insights on how different areas perform their urban functions.*

What are the features that generate happiness in each cluster?

- *Evaluate the importance of a term for a cluster by tf-idf score (term frequence – inverse document frequence)*
- *Considering uni-gram, bi-gram and tri-gram*
- *Merge n-grams after sorting tf-idf score (same word-stem/inclusion/intersection)*
- *Visualize term tf-idf score and correlated tweet sentiment by wordcloud (lighter color and larger size mean higher tf-idf and sentiment value)*

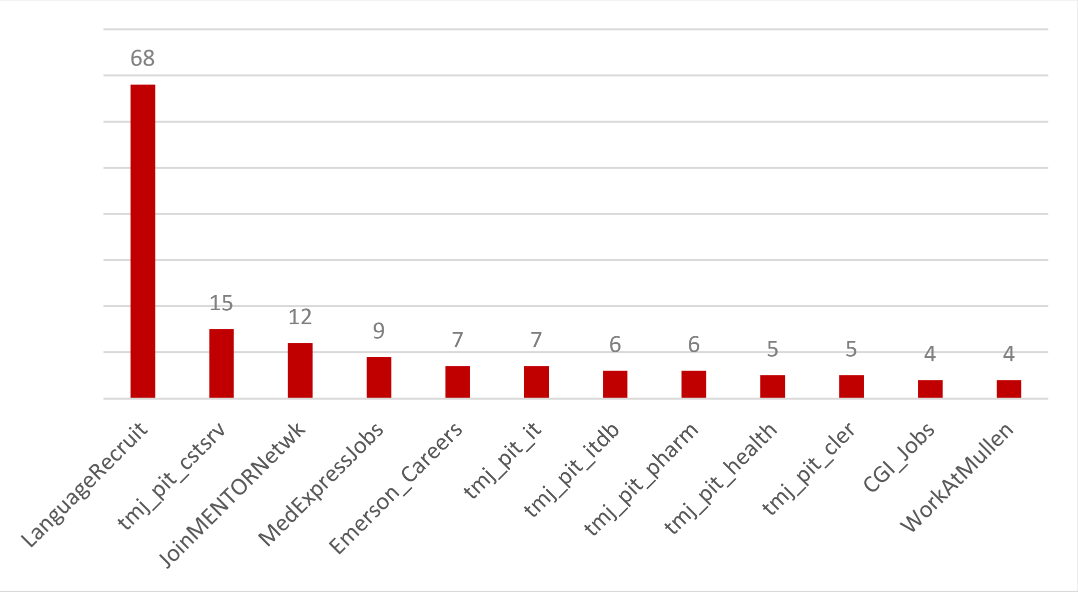
Tweet WordCloud Downtown

However…

- *Why 'alert' has such tf-idf score?*
- *Why 'interpreters languageline solutions' is associated with so many happy tweets?*

If we search happy tweet in Downtown with keyword 'alert' and 'Interpreters'…

- *Most tweets containing 'alert' are job recruitment posts*



- *Vader gives high sentiment value to "Interpreters … | LanguageLine Solutions | PA"*

	username text	txt text	senti_val real
1	LanguageRecruit	alert: Hindi Interpreters | LanguageLine Solutions | PA	1.612
2	LanguageRecruit	alert: Korean Interpreters | LanguageLine Solutions | PA	1.612
3	LanguageRecruit	alert: Albanian Interpreters | LanguageLine Solutions | PA	1.612
4	LanguageRecruit	alert: Hindi Interpreters | LanguageLine Solutions | PA	1.612
5	LanguageRecruit	alert: Tagalog Onsite Interpreters in Houston, Texas | LanguageLine Solutions |	1.023
6	LanguageRecruit	alert: Russian Onsite Interpreters in New York... | LanguageLine Solutions | PA	1.023
7	LanguageRecruit	alert: Vietnamese Interpreters | LanguageLine Solutions | PA	1.612
8	LanguageRecruit	alert: Karenni Interpreters | LanguageLine Solutions | PA	1.612
9	LanguageRecruit	alert: Cambodian (Khmer) Interpreters | LanguageLine Solutions | PA	1.413
10	LanguageRecruit	alert: Anuak Interpreters | LanguageLine Solutions | PA	1.612

To address this issue, another tf-idf is considered, which evaluates the importance of a term for a specific user. Therefore, the user-effect could excluded by dividing cluster tf-idf with user tf-idf standard deviation…

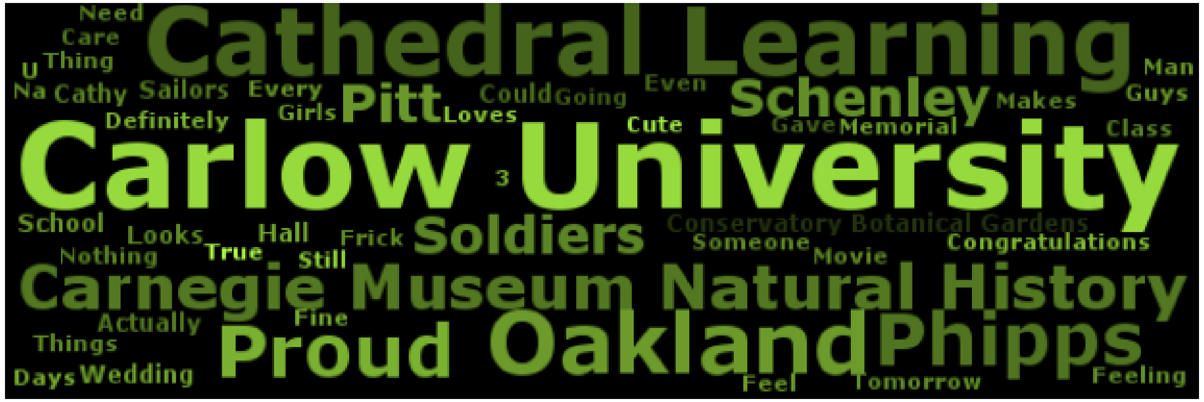
Tweet WordCloud Downtown

Tweet WordCloud Downtown

- *Great! 'Alert' and 'Interpreters…' are no longer popping up.*
- *More terms give identify to Oakland.*

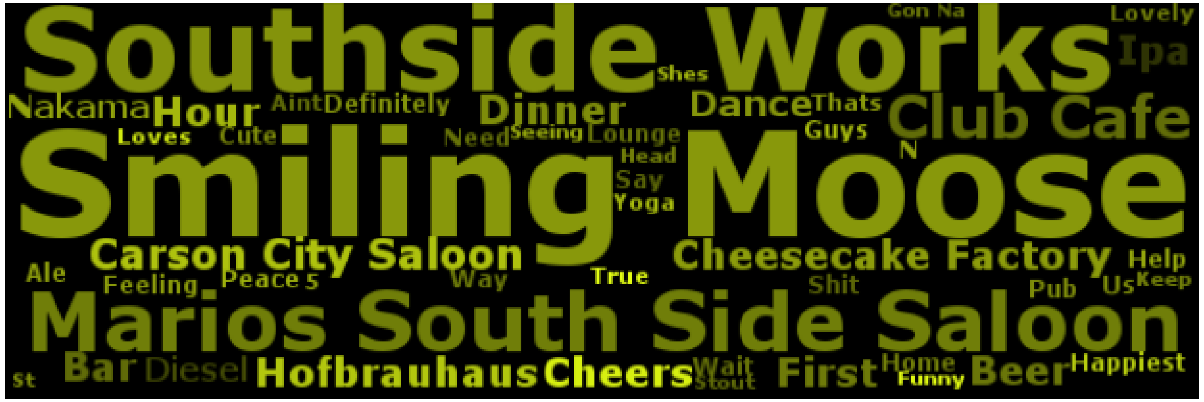
Tweet WordCloud Oakland

Tweet WordCloud Oakland

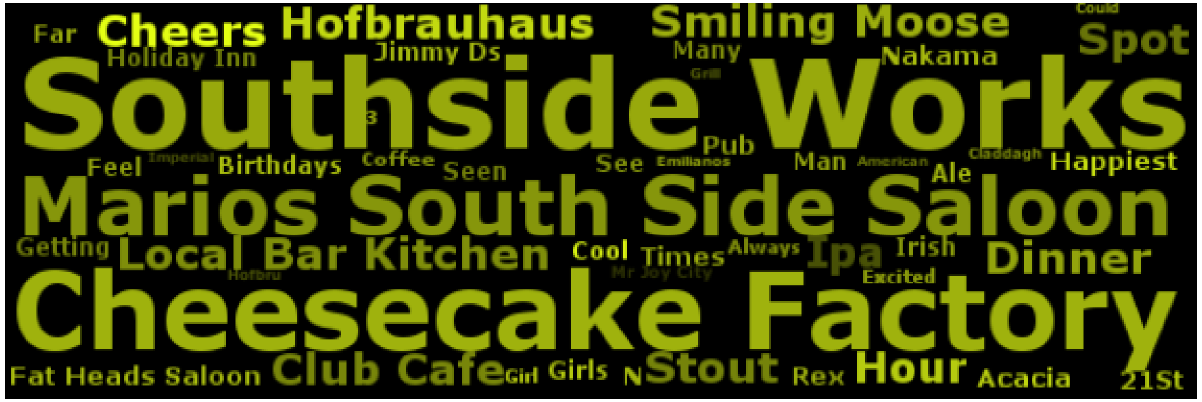


To address this issue, another tf-idf is considered, which evaluates the importance of a term for a specific user. Therefore, the user-effect could excluded by dividing cluster tf-idf with user tf-idf standard deviation…

Tweet WordCloud Southside



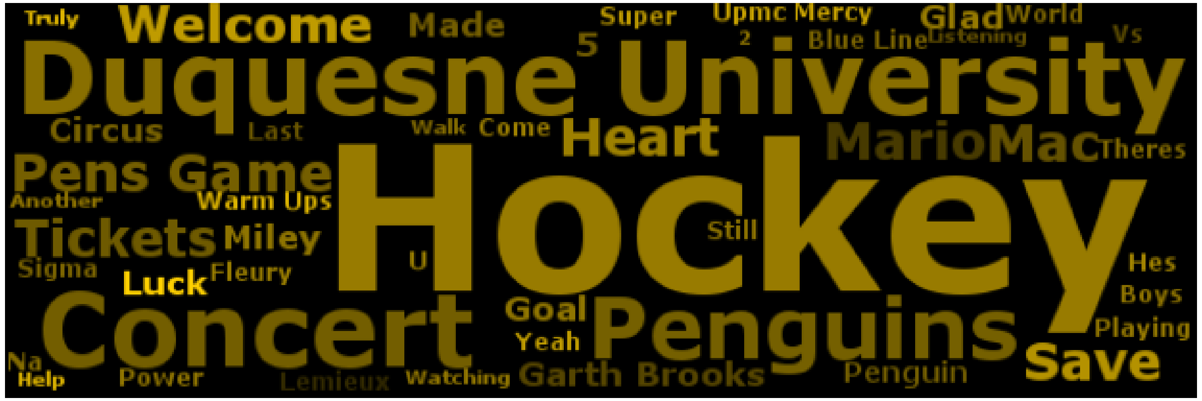
Tweet WordCloud Southside



- *'Smiling Moose' may not be as popular as it is at first glance.*
- *Hockey is still the main purpose for most people heading to Console Energy Center*

Tweet WordCloud Console Energy Center

Tweet WordCloud Console Energy Center



To address this issue, another tf-idf is considered, which evaluates the importance of a term for a specific user. Therefore, the user-effect could excluded by dividing cluster tf-idf with user tf-idf standard deviation…

Tweet WordCloud Pnc Park

Tweet WordCloud Pnc Park

- *It seems many people enjoyed Taylor Swifts concert last year.*
- *Most tweets containing ‘steeler’ were posted by ‘SteelTownUsa’.*

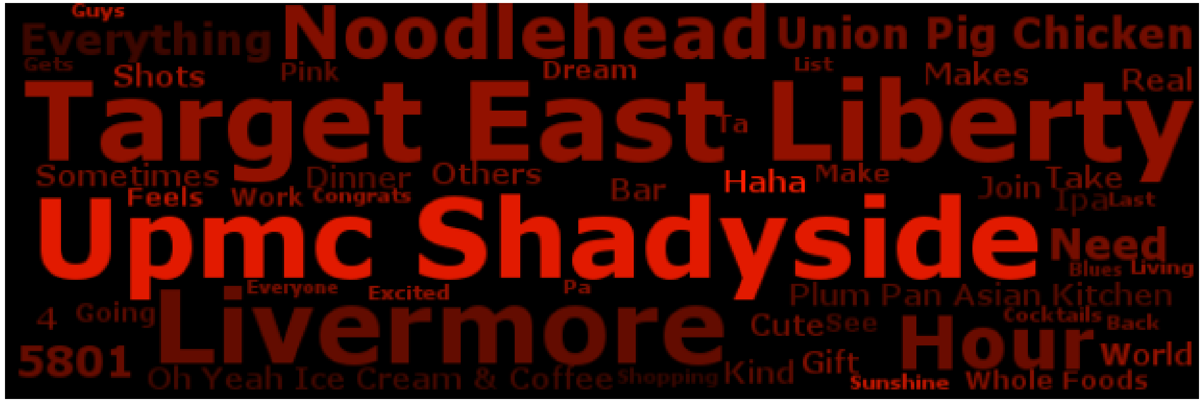
Tweet WordCloud Heinz Field

Tweet WordCloud Heinz Field

To address this issue, another tf-idf is considered, which evaluates the importance of a term for a specific user. Therefore, the user-effect could excluded by dividing cluster tf-idf with user tf-idf standard deviation…

Tweet WordCloud Shadyside & East Liberty

Tweet WordCloud Shadyside & East Liberty



- *UPMC is not in the original wordcloud, not shows up after considering user impact.*
- *More good restaurants in Strip District pop up, but the lightest color of 'Peace Love Little Donuts' is due to its tricky name.*

Tweet WordCloud Strip District

Tweet WordCloud Strip District

Some findings:

- *The wordcloud based on tf-idf score and sentiment value visualizes hot topics, popular places and events in different urban areas. However, user impacts on the term tf-idf score should be considered.*
- *Terms that are popular among most people could be better recognized if the cluster-based tf-idf is normalized by user-based tf-idf.*
- *This method could also be used in user classification, differentiating organizations from individuals.*

How about unhappy tweets?
Wordclouds of unhappy tweets are created following the same method…

Tweet WordCloud Downtown

Tweet WordCloud Oakland

Tweet WordCloud Southside

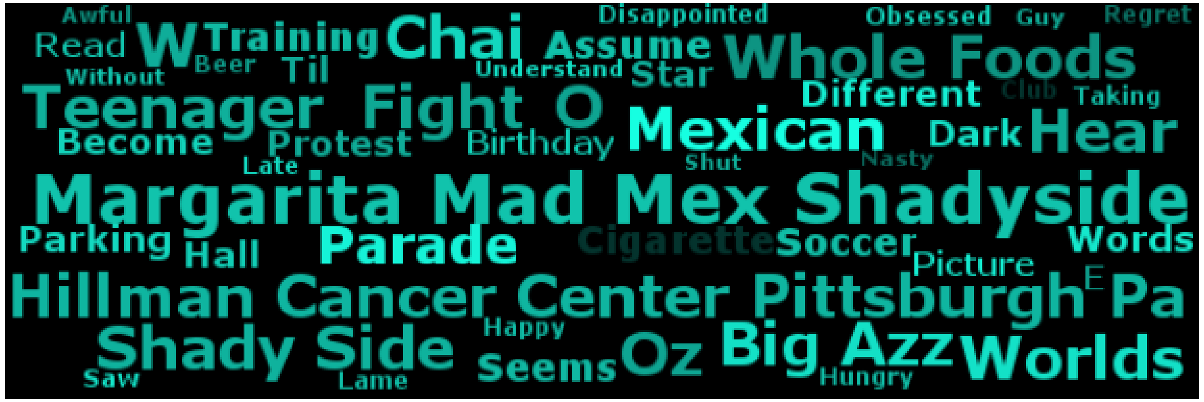
Tweet WordCloud Console Energy Center

How about unhappy tweets?
Wordclouds of unhappy tweets are created following the same method…

Tweet WordCloud Pnc Park

Tweet WordCloud Heinz Field

Tweet WordCloud Shadyside & East Liberty



Tweet WordCloud Strip District

What if we generalize happy and unhappy tweets into emotional tweets?
Wordclouds of emotional tweets (happy/unhappy) are created following the same method…

Tweet WordCloud Downtown

Tweet WordCloud Oakland

Tweet WordCloud Southside

Tweet WordCloud Console Energy Center

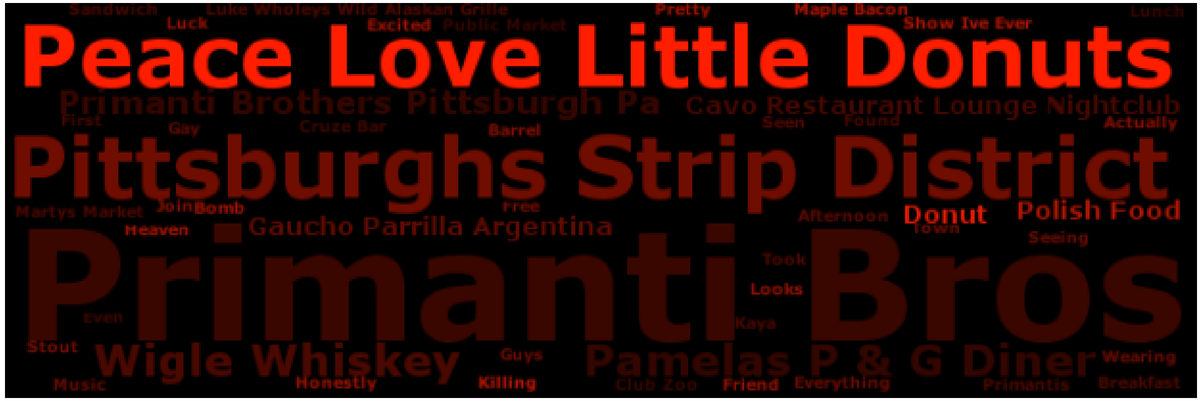
What if we generalize happy and unhappy tweets into emotional tweets?
Wordclouds of emotional tweets (happy/unhappy) are created following the same method…

Tweet WordCloud Pnc Park

Tweet WordCloud Heinz Field

Tweet WordCloud Shadyside & East Liberty

Tweet WordCloud Strip District

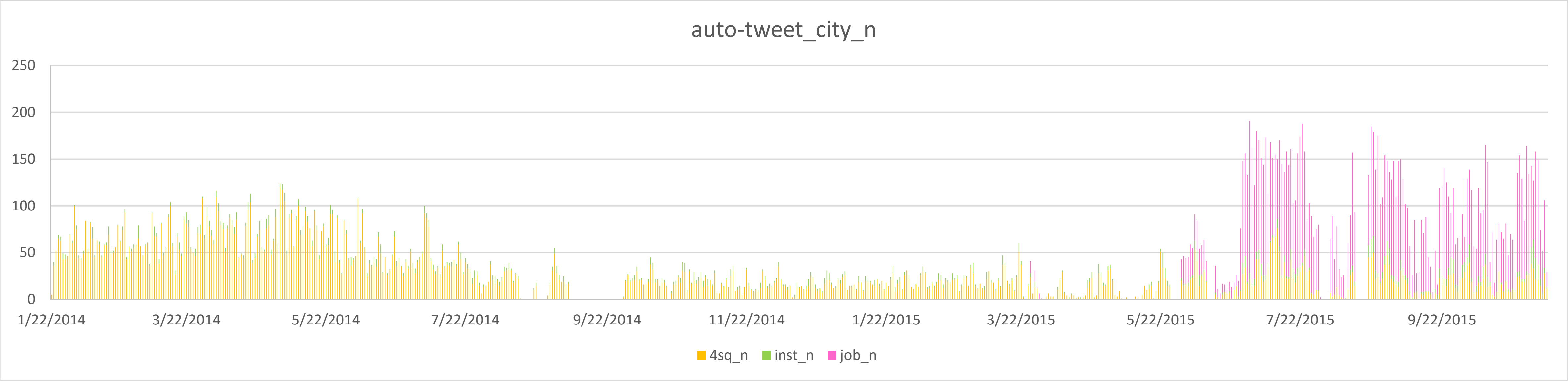


Some findings:

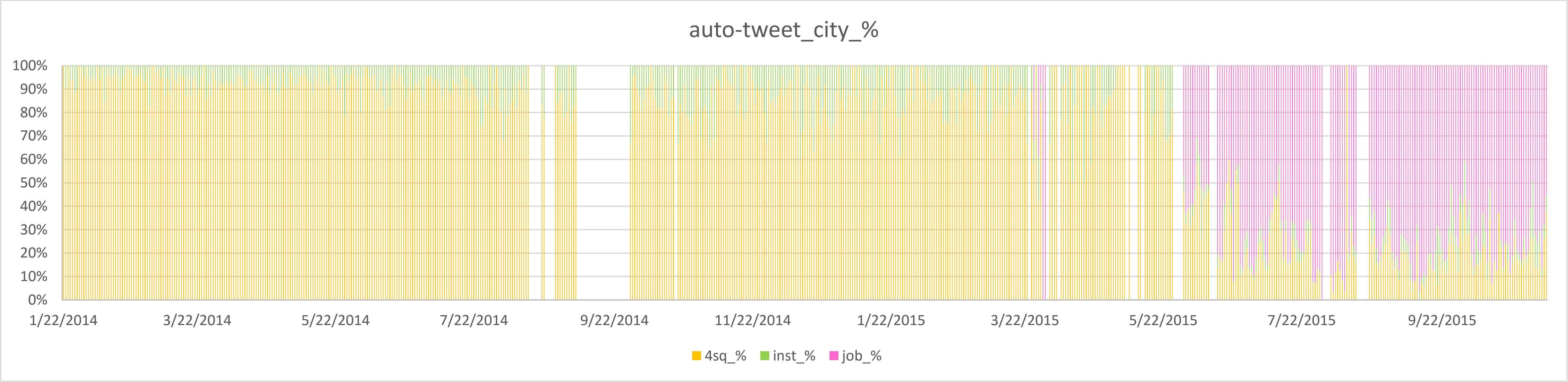
- *Even though more noisy, local venues are also popping up from unhappy tweet wordclouds. Some of them appear in wordclouds of happy tweets as well. Therefore, it is hard to say these places are making people happy or unhappy*
- *However, we may conclude that local venues are more likely to cause emotion, either make people feel happy or unhappy.*
- *Therefore, it becomes interesting to study the relationship between tweet sentiment strength and urban functions in certain areas.*

URBAN FUNCTIONS

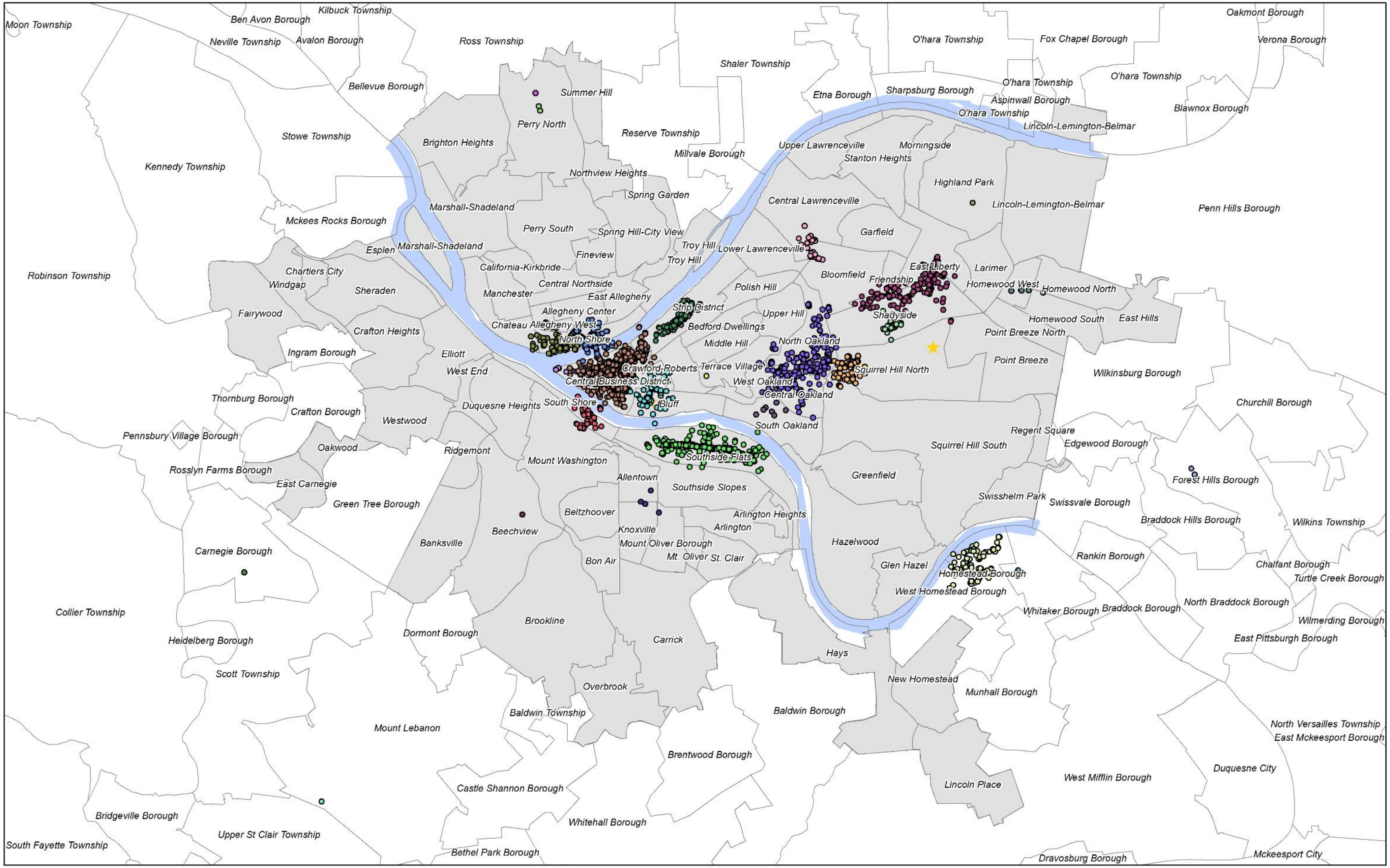
# of auto-tweets by category…



% of auto-tweets by category…



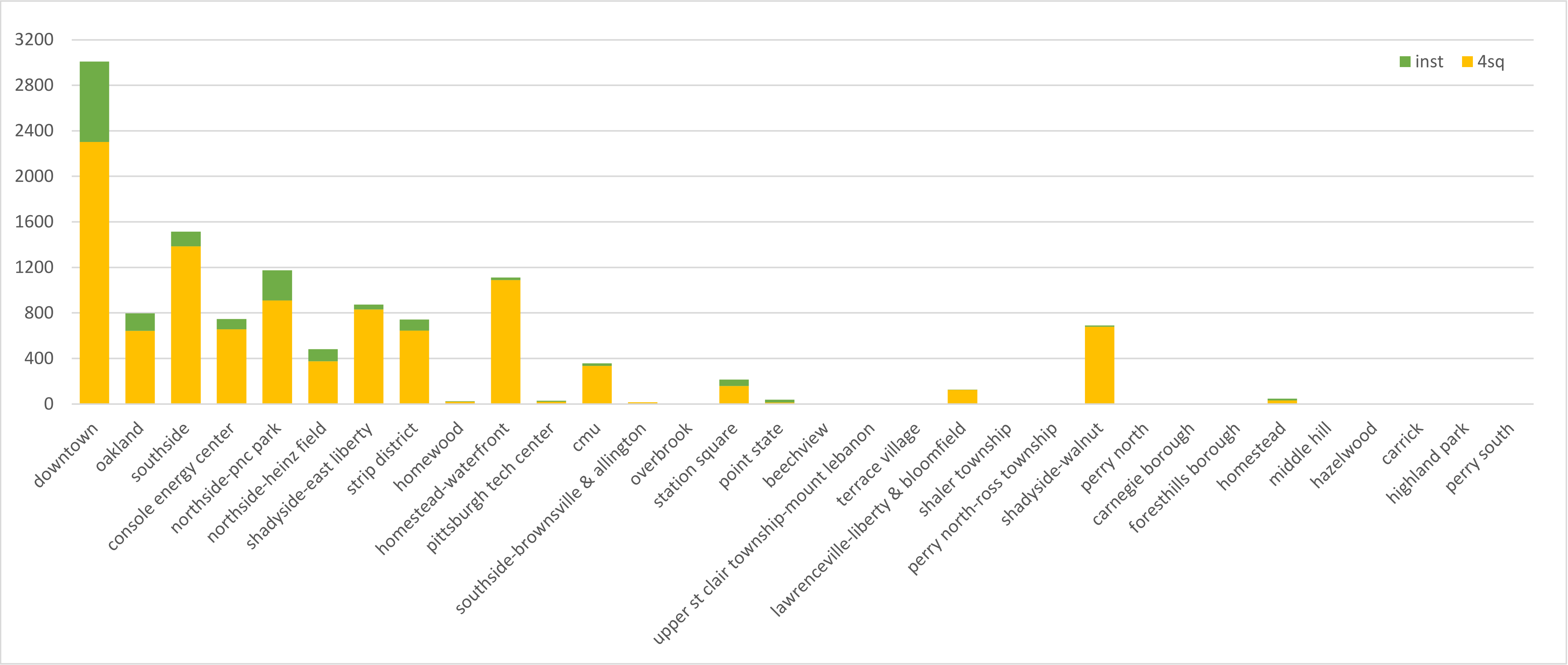
Extract venue check-ins from Foursquare and Instagram auto-tweets, and then merge into clusters…



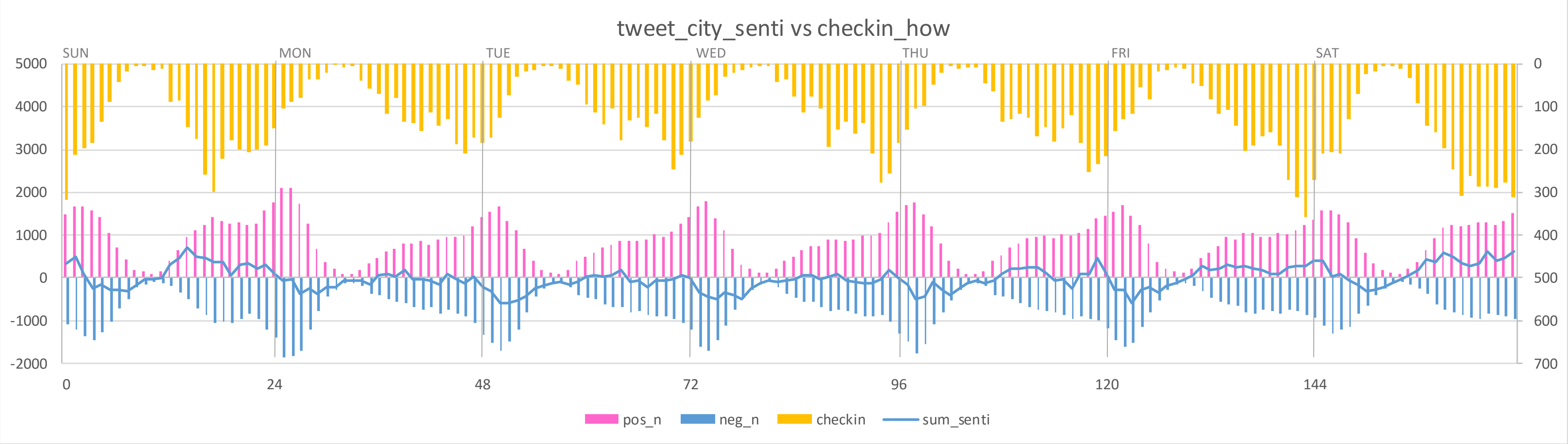
Numbers of venue extracted from Foursquare and Instagram auto-tweets…

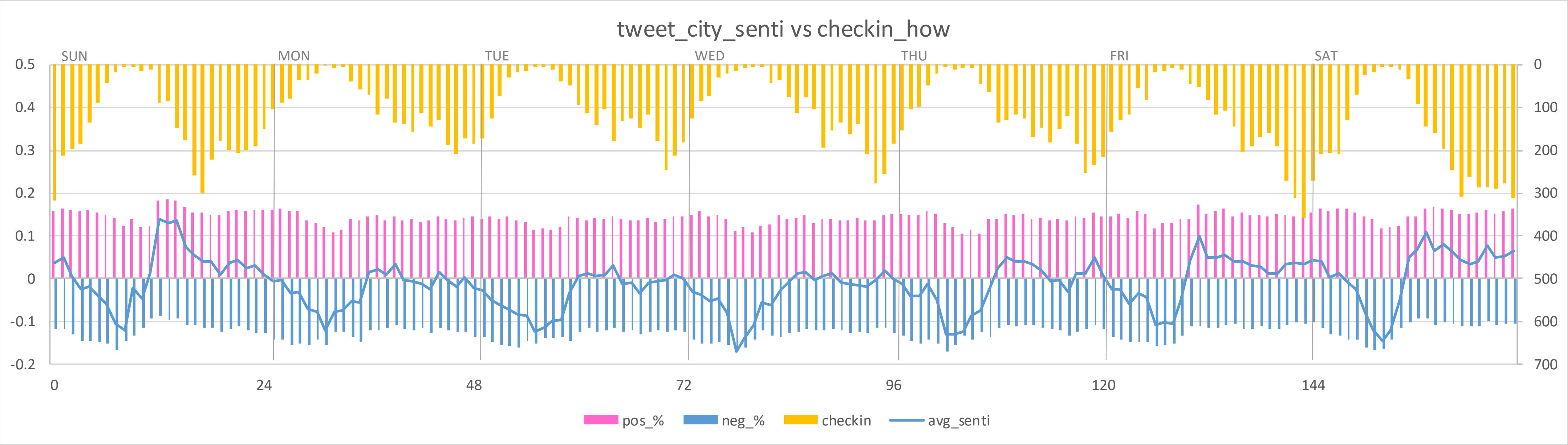
clid		all	4sq	inst	clid		all	4sq	inst
0	downtown	3008	2302	706	16	beechview	1	0	1
1	oakland	797	642	155	17	upper st clair township-mount lebanon	1	0	1
2	southside	1514	1386	128	18	terrace village	1	0	1
3	console energy center	747	657	90	19	lawrenceville-liberty & bloomfield	127	123	4
4	northside-pnc park	1174	910	264	20	shaler township	0	0	0
5	northside-heinz field	481	375	106	21	perry north-ross township	1	0	1
6	shadyside-east liberty	873	831	42	22	shadyside-walnut	691	679	12
7	strip district	741	644	97	23	perry north	2	2	0
8	homewood	24	20	4	24	carnegie borough	1	1	0
9	homestead-waterfront	1112	1088	24	25	foresthills borough	3	3	0
10	pittsburgh tech center	29	19	10	26	homestead	47	31	16
11	cmu	358	335	23	27	middle hill	0	0	0
12	southside-brownsville & allington	15	15	0	28	hazelwood	0	0	0
13	overbrook	0	0	0	29	carrick	0	0	0
14	station square	215	157	58	30	highland park	1	0	1
15	point state	38	13	25	31	perry south	0	0	0

Numbers of venue extracted from Foursquare and Instgram auto-tweets…



Temporal venue checkins and tweet sentiment by hour of week…





Is there any correlation between urban emotion and venue checkin?
If so, is venue checkin correlated with urban happiness or sentiment strength?

- *Pearson correlation coefficients between tweet sentiment and venue checkin by hour of week…*

clid		**senti**	**senti_abs**	**pos_n**	**neg_n**
	pittsburgh	0.6151	0.5677	0.6723	0.4839
0	downtown	0.5866	0.7949	0.7885	0.6195
2	southside	0.3865	0.8046	0.7907	0.6672
4	northside-pnc park	0.5691	0.7197	0.7037	0.5858
9	homestead-waterfront	0.2125	0.6338	0.5991	0.5084
6	shadyside-east liberty	0.1873	0.6256	0.5446	0.4118
1	oakland	0.2751	0.5016	0.5254	0.3557
3	console energy center	0.5401	0.4994	0.5076	0.3422
7	strip district	0.4587	0.6708	0.6273	0.3678
22	shadyside-walnut	0.1503	0.3316	0.2431	0.1404
5	northside-heinz field	0.2891	0.48	0.4333	0.4345

Similar to tweet wordclouds, we could visualize popular venues of each cluster based on their checkin and users…

Venue checkin WordCloud Downtown

Venue user WordCloud Downtown

Venue checkin WordCloud Southside

Venue user WordCloud Southside

Similar to tweet wordclouds, we could visualize popular venues of each cluster based on their checkin and users…

Venue checkin WordCloud Pnc Park

Venue user WordCloud Pnc Park

Venue checkin WordCloud Waterfront

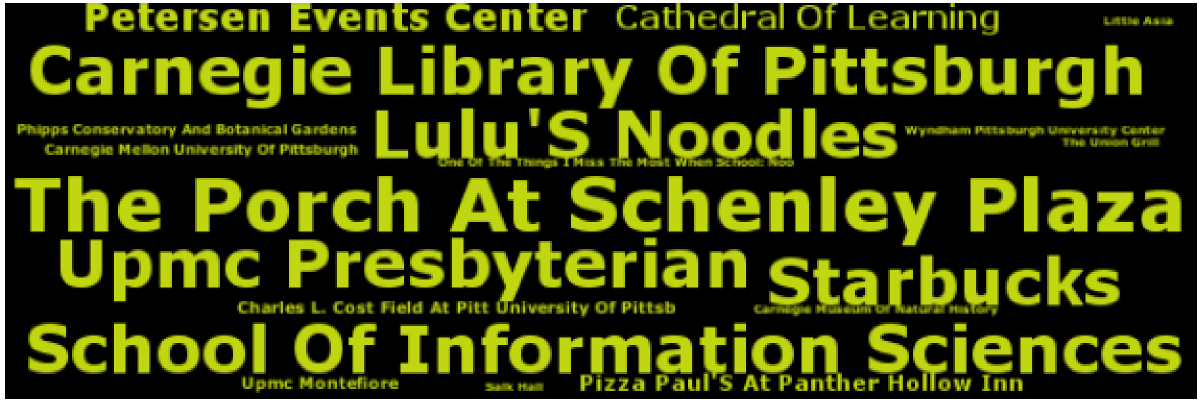
Venue user WordCloud Waterfront

Similar to tweet wordclouds, we could visualize popular venues of each cluster based on their checkin counts…

Venue checkin WordCloud Shadyside & East Liberty

Venue user WordCloud Shadyside & East Liberty

Venue checkin WordCloud Oakland



Venue user WordCloud Oakland

Similar to tweet wordclouds, we could visualize popular venues of each cluster based on their checkin counts…

Venue checkin WordCloud Console Energy Center

Venue user WordCloud Console Energy Center

Venue checkin WordCloud Strip District

Venue user WordCloud Strip District

Some findings:

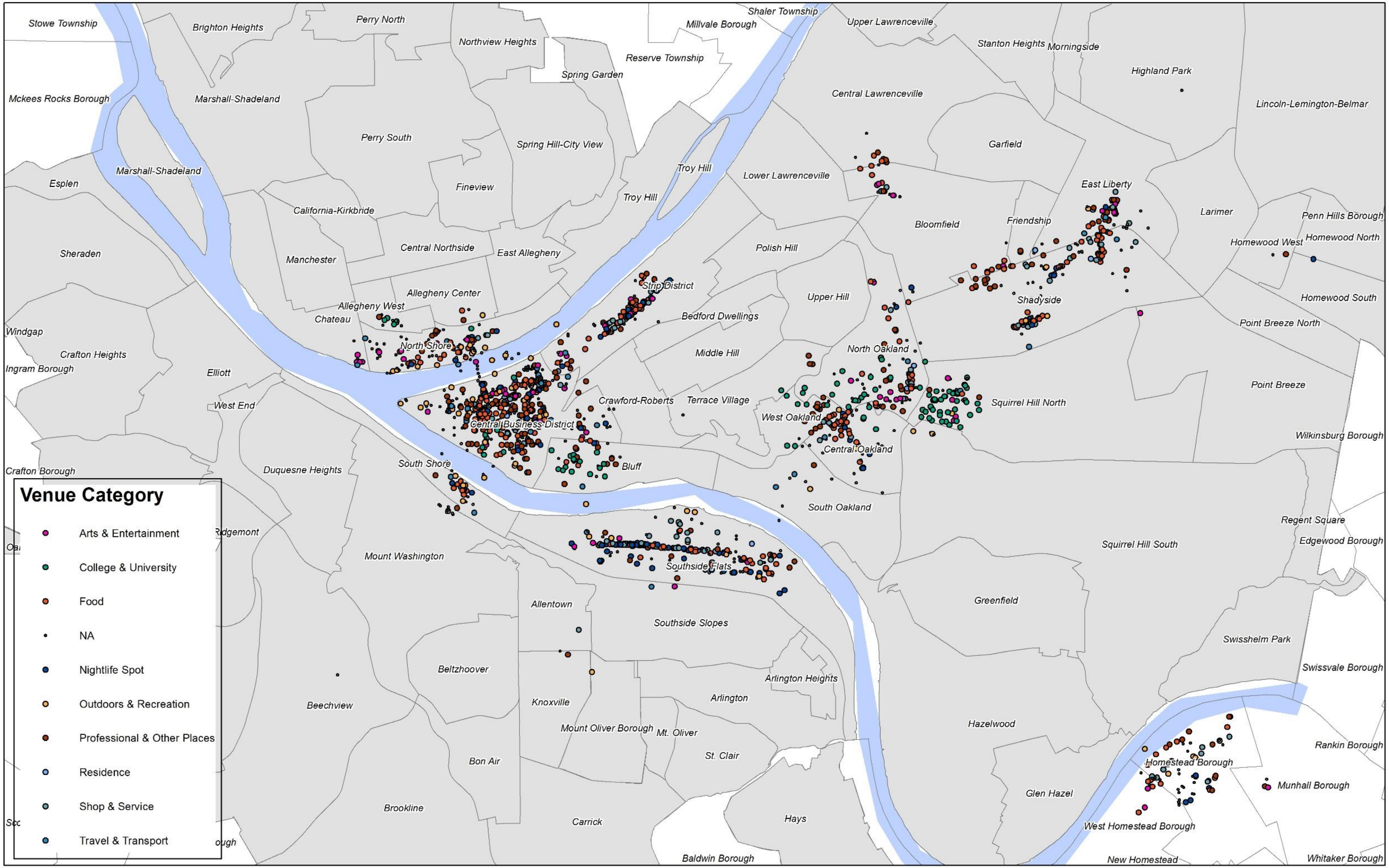
- *Venues could be extracted from Foursquare and Instgram auto-generated tweets.*
- *Temporal venue checkins are positively correlated with tweet happiness over the city and several clusters.*
- *For other clusters, venue checkins are only correlated with tweet sentiment strength, which means certain urban activities make people feel more emotional or tend to express emotion.*
- *Venues distribute unevenly in the city, and each cluster has different type of popular venues. By comparing wordclouds generated from checkins and users, we could learn whether certain venues are popular over the city, or just attract a specific group of people*

Question:

- *Is there any correlation between tweet happiness and certain venue category?*
- *If we generalize venue categories as different urban use, what is the temporal pattern of urban functions?*

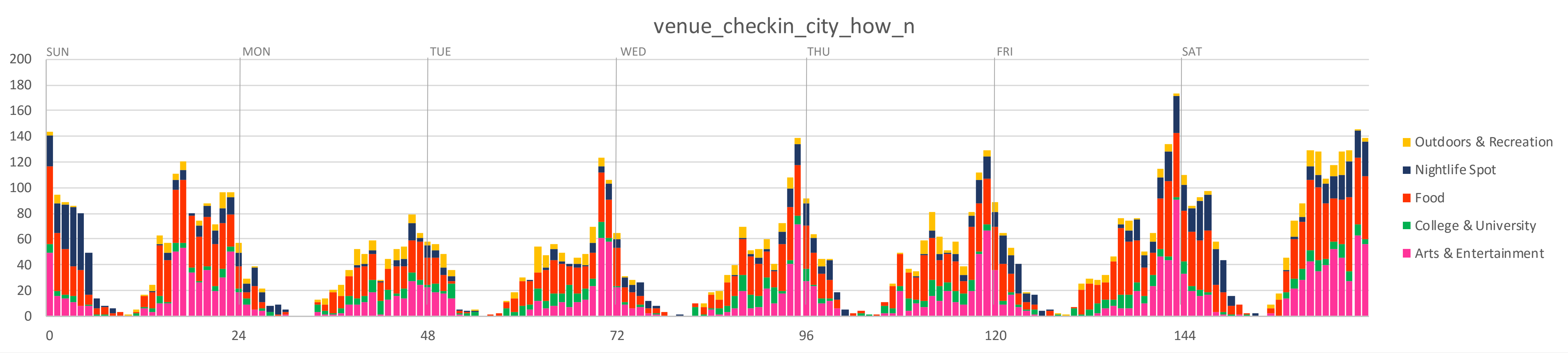
Foursquare gives 9 main categories. By searching venue name and coordinates through Foursquare API, we could get the categories of most venues.

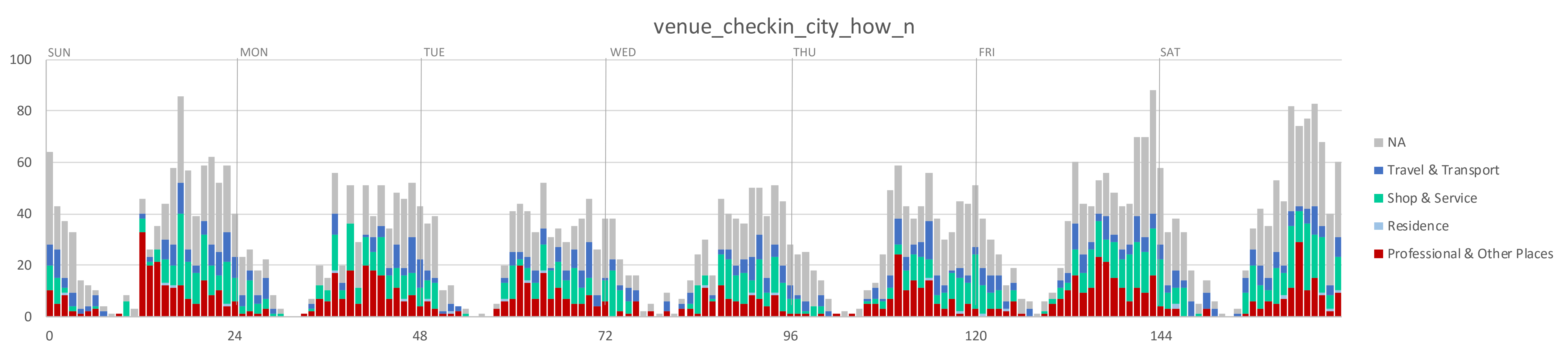
clid	**pitts**	**0**	**2**	**4**	**9**	**6**	**1**	**3**	**7**	**22**	**5**
Arts & Entertainment	88	25	10	7	3	6	11	3	6	0	9
College & University	101	9	0	1	0	0	25	17	0	0	5
Event	0	0	0	0	0	0	0	0	0	0	0
Food	360	118	54	10	18	30	47	10	36	15	4
Nightlife Spot	130	23	46	6	4	7	11	7	14	5	1
Outdoors & Recreation	80	27	9	10	4	4	4	1	2	3	5
Professional & Other Places	209	80	9	7	8	32	26	11	5	4	5
Residence	10	3	3	0	0	2	2	0	0	0	0
Shop & Service	143	22	37	2	10	19	13	2	22	12	0
Travel & Transport	50	25	3	3	0	3	3	2	0	1	2
Others	522	153	59	26	29	37	63	20	27	16	37
Total	**1693**	**485**	**230**	**72**	**76**	**140**	**205**	**73**	**112**	**56**	**68**



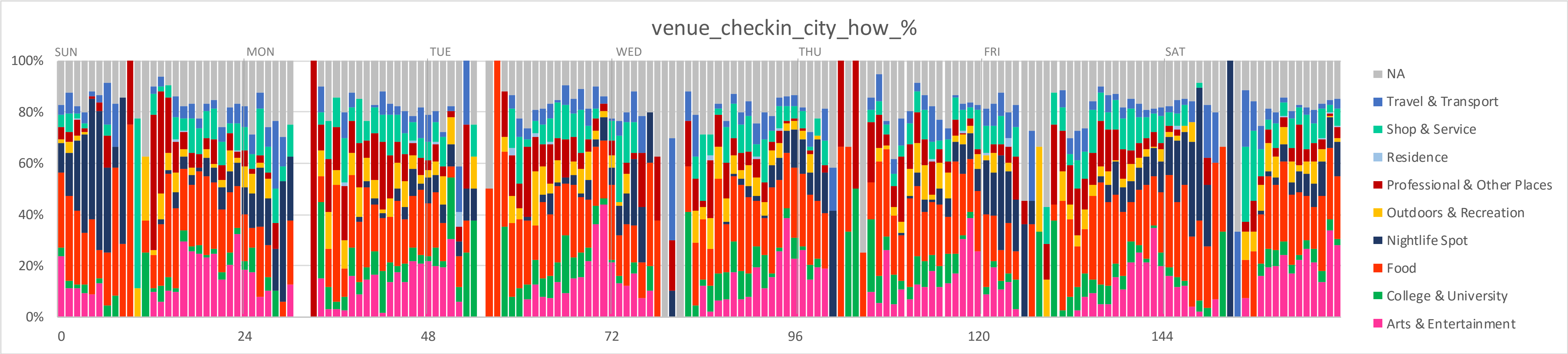
Stowe Township
Brighton Heights
Perry North
Northview Heights
Shaler Township
Millvale Borough
Upper Lawrenceville
Reserve Township
Stanton Heights
Morningside
Spring Garden
Highland Park
Mckees Rocks Borough
Marshall-Shadeland
Central Lawrenceville
Lincoln-Lemington-Belmar
Perry South
Spring Hill-City View
Garfield
Troy Hill
Esplen
Marshall-Shadeland
Lower Lawrenceville
Fineview
East Liberty
Troy Hill
California-Kirkbride
Bloomfield
Friendship
Larimer
Penn Hills Borough
Sheraden
Central Northside
East Allegheny
Polish Hill
Homewood West
Homewood North
Manchester
Strip District
Allegheny Center
Upper Hill
Shadyside
Homewood South
Allegheny West
Chateau
Bedford Dwellings
Point Breeze North
Windgap
North Shore
North Oakland
Crafton Heights
Middle Hill
Ingram Borough
Elliott
Point Breeze
Crawford-Roberts
Terrace Village
Squirrel Hill North
West End
West Oakland
Central Business District
Wilkinsburg Borough
Central Oakland
South Shore
Bluff
Crafton Borough
Duquesne Heights
South Oakland
Regent Square
Ridgemont
Squirrel Hill South
Edgewood Borough
Mount Washington
Southside Flats
Greenfield
Allentown
Southside Slopes
Swisshelm Park
Swissvale Borough
Beltzhoover
Arlington Heights
Beechview
Knoxville
Arlington
Hazelwood
Mount Oliver Borough
Mt. Oliver
Rankin Borough
St. Clair
Bon Air
Homestead Borough
Munhall Borough
Glen Hazel
Brookline
Carrick
Hays
West Homestead Borough
Baldwin Borough
New Homestead
Whitaker Borough
Venue Category
Arts & Entertainment
College & University
Food
NA
Nightlife Spot
Outdoors & Recreation
Professional & Other Places
Residence
Shop & Service
Travel & Transport

Temporal venue checkins of each category by hour of week…

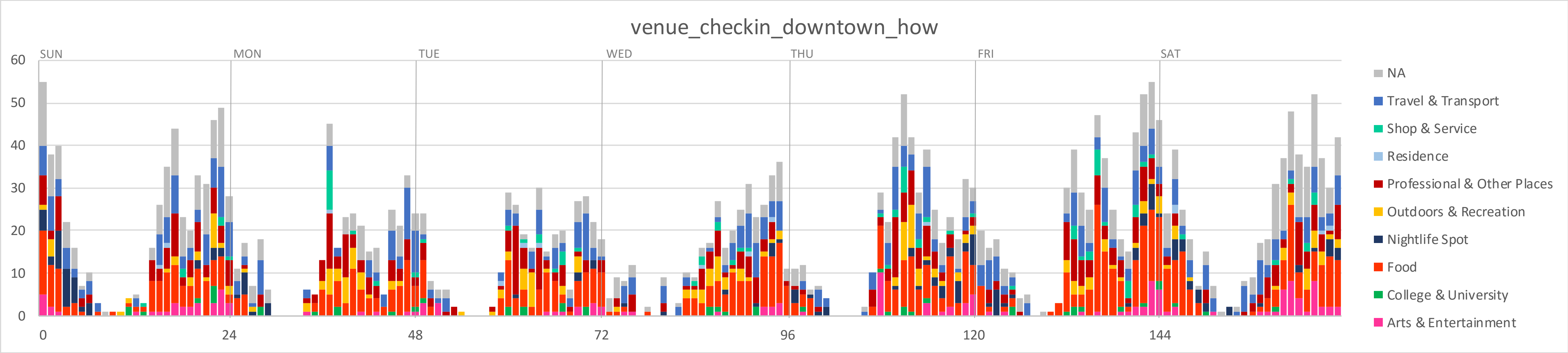


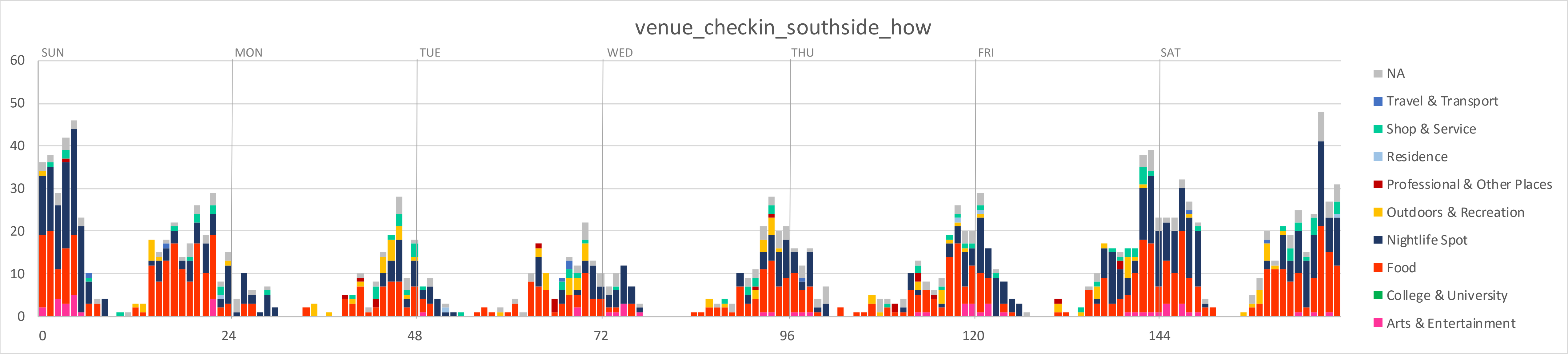


% of temporal venue checkin of each category by hour of week…

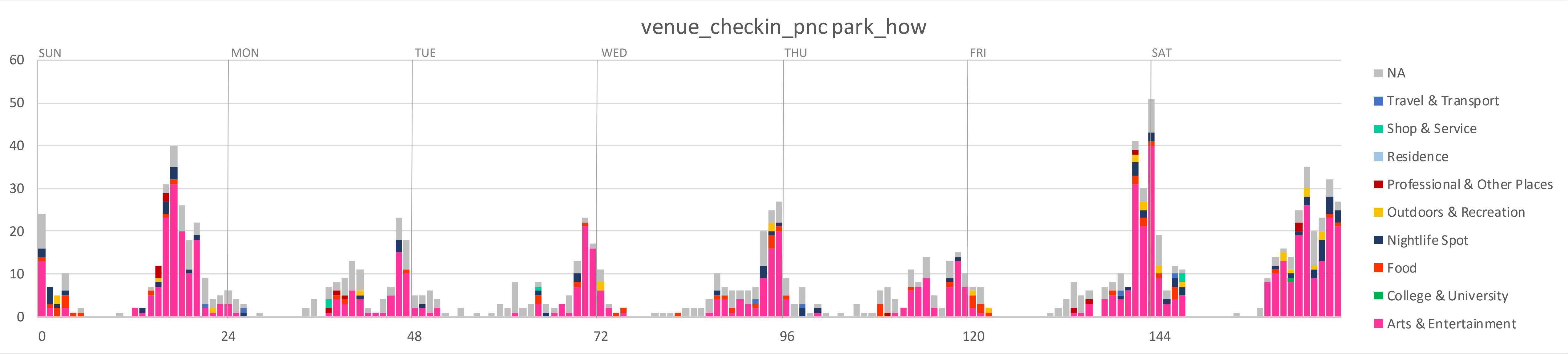


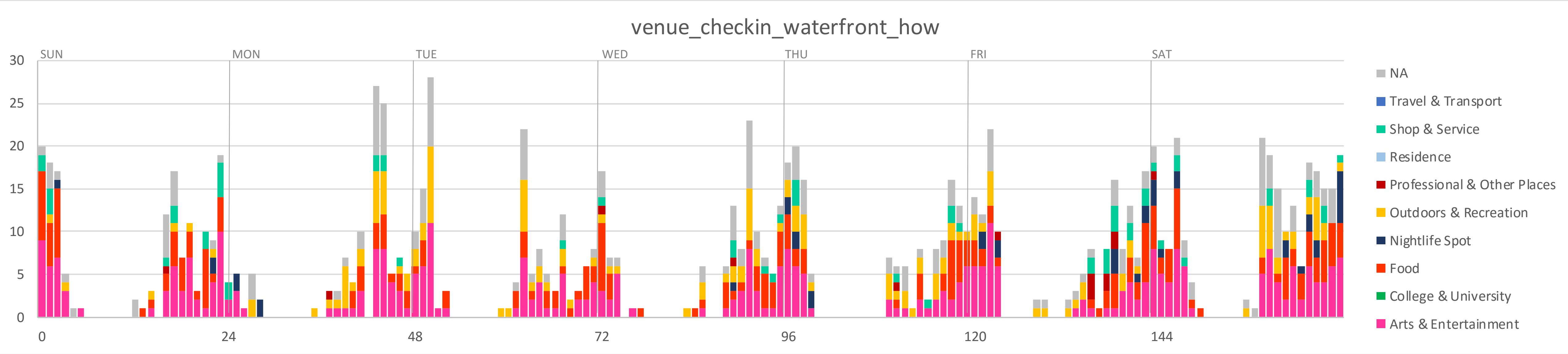
Temporal venue checkins of each category by hour of week…





Temporal venue checkins of each category by hour of week…





Is there any correlation between urban happiness and urban use?

- *Pearson correlation coefficients at city level…*

use	senti	senti_abs	pos_n	neg_n
College & University	0.355	0.1722	0.2379	0.1414
Food	0.6791	0.48	0.5937	0.3858
Residence	0.1242	0.1855	0.2043	0.166
Travel & Transport	0.3511	0.4253	0.4809	0.3765
Outdoors & Recreation	0.4529	0.2407	0.3172	0.1877
Arts & Entertainment	0.5079	0.4407	0.5288	0.3699
Shop & Service	0.5197	0.3795	0.4688	0.3152
Nightlife Spot	0.1535	0.6175	0.6386	0.5756
Professional & Other Places	0.5371	0.0253	0.1161	-0.036
Others	0.5699	0.5278	0.6231	0.4507

- *Pearson correlation coefficients at cluster level…*

	downtown				**southside**				**pnc park**				**waterfront**			
use	**senti**	**senti_abs**	**pos_n**	**neg_n**	**senti**	**senti_abs**	**pos_n**	**neg_n**	**senti**	**senti_abs**	**pos_n**	**neg_n**	**senti**	**senti_abs**	**pos_n**	**neg_n**
College & University	0.098	0.1817	0.1845	0.116	--	--	--	--	0.0495	-0.0078	-0.0008	-0.0416	--	--	--	--
Food	0.514	0.7182	0.6996	0.592	0.4245	0.5902	0.6023	0.4193	0.3525	0.3715	0.3754	0.3354	0.1968	0.687	0.5971	0.568
Residence	0.0732	0.1942	0.1793	0.1789	0.3092	0.1218	0.1508	0.0084	--	--	--	--	--	--	--	--
Travel & Transport	0.3252	0.4622	0.4581	0.3805	-0.0425	0.1165	0.0856	0.136	0.0346	0.1139	0.0893	0.1672	--	--	--	--
Outdoors & Recreation	0.2556	0.4705	0.4273	0.3836	0.0622	-0.0431	-0.0312	-0.0864	0.3525	0.3672	0.3939	0.3243	0.1414	0.1695	0.2418	0.1131
Arts & Entertainment	0.3809	0.3651	0.4095	0.2203	0.1001	0.6078	0.5776	0.5865	0.4991	0.6496	0.6329	0.5123	0.2524	0.6483	0.6133	0.5066
Shop & Service	0.0872	0.3485	0.265	0.3313	0.3856	0.28	0.3572	0.1189	-0.0936	-0.0304	-0.0481	0.0363	0.0566	0.3349	0.3048	0.2942
Nightlife Spot	0.4149	0.2668	0.3719	0.1588	0.236	0.8219	0.7857	0.7456	0.3887	0.5222	0.5101	0.4716	0.0476	0.3351	0.3089	0.2352
Professional & Other Places	0.2293	0.5475	0.4656	0.4938	-0.0713	-0.0254	-0.0659	-0.0186	0.0015	-0.031	-0.0225	-0.0471	0.0012	-0.0343	-0.0384	-0.059
Others	0.6157	0.6599	0.7083	0.4443	0.2843	0.5207	0.5099	0.4175	0.3898	0.49	0.4587	0.4351	0.1138	0.2379	0.2446	0.1927

Some findings:

- *Urban use could be represented by checkins of different venue categories.*
- *Temporal patterns of urban use are different among clusters.*
- *Some clusters have dominant urban functions while some are more mixed-use.*
- *Some urban functions are correlated with tweet happiness or emotion strength.*

Using wordclouds, we could better visualize the mixed-use level of each cluster…

Using wordclouds, we could better visualize the mixed-use level of each cluster…

checkin WordCloud Downtown

checkin WordCloud Southside

checkin WordCloud Pnc Park

checkin WordCloud Waterfront

How to evaluate the mixed-use level of each cluster?

- *Overall entropies among venue categories…*

clid		entro	clid		entro	clid		entro
1	oakland	2.8293	19	lawrenceville-liberty & bloomfield	1.6097	17	upper st clair township-mount lebanon	0
0	downtown	2.768	10	pittsburgh tech center	1.5538	18	terrace village	0
9	homestead-waterfront	2.3624	12	southside-brownsville & allington	1.4295	20	shaler township	0
5	northside-heinz field	2.3446	23	perry north	1	21	perry north-ross township	0
14	station square	2.3356	11	cmu	0.9374	24	carnegie borough	0
6	shadyside-east liberty	2.3233	25	foresthills borough	0.9183	27	middle hill	0
7	strip district	2.194	8	homewood	0.549	28	hazelwood	0
2	southside	2.1176	15	point state	0.5328	29	carrick	0
22	shadyside-walnut	2.0238	26	homestead	0.2678	30	highland park	0
3	console energy center	1.7553	13	overbrook	0	31	perry south	0
4	northside-pnc park	1.7045	16	beechview	0			

How to evaluate the mixed-use level of each cluster?

- *Sum of entropies among venue categories at each hour of week…*

clid		entro	clid		entro	clid		entro
0	downtown	295.774	14	station square	63.9341	20	shaler township	0
1	oakland	188.4001	11	cmu	29.3334	21	perry north-ross township	0
2	southside	178.528	19	lawrenceville-liberty & bloomfield	14.2636	23	perry north	0
9	homestead-waterfront	169.3661	10	pittsburgh tech center	4.8113	24	carnegie borough	0
6	shadyside-east liberty	154.1201	15	point state	1.8366	25	foresthills borough	0
22	shadyside-walnut	130.1147	12	southside-brownsville & allington	1	26	homestead	0
7	strip district	124.6809	8	homewood	0	27	middle hill	0
4	northside-pnc park	114.1051	13	overbrook	0	28	hazelwood	0
3	console energy center	102.0079	16	beechview	0	29	carrick	0
5	northside-heinz field	93.298	17	upper st clair township-mount lebanon	0	30	highland park	0
			18	terrace village	0	31	perry south	0

Some findings:

- *The mixed-use level of each cluster could be visualized by wordclouds.*
- *The mixed-use level of each cluster could be quantified either by calculating the overall entropy among venue categories or by summing up entropies by hour of week.*

(However, the latter measurement is based on the assumption of identical weekly pattern of urban use for each cluster.)