

Understanding the Needs of People in Big Cities through Data Science

Harvey Alférez, Ph.D.

Global Software Lab

School of Engineering and Technology

Universidad de Montemorelos, Mexico

www.harveyalferez.com





Cities are growing fast

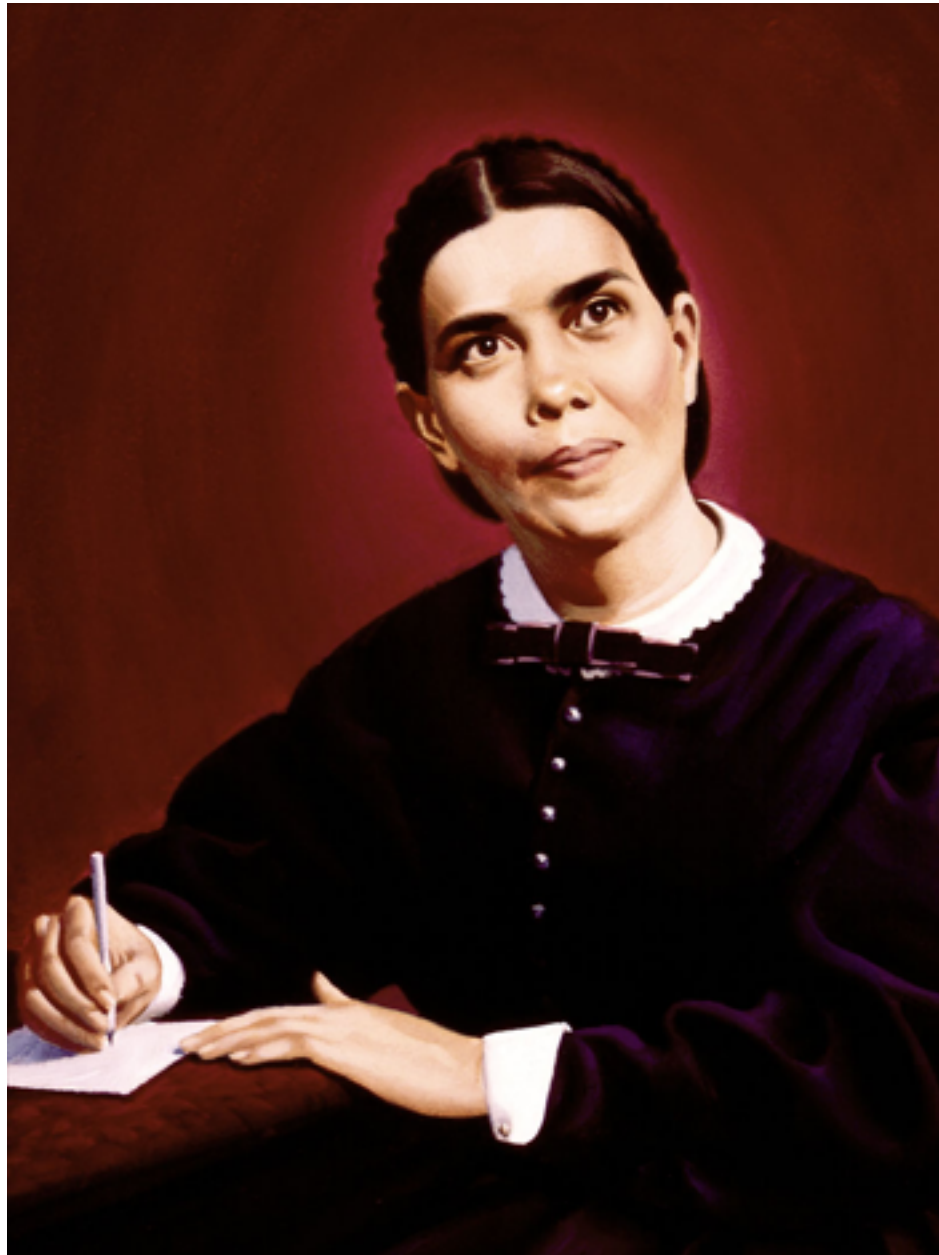
- **66%** of the world's population will live in urban areas by **2050** [1].
- There are more than **500** cities with a population of **1 million or more people**. However, these cities have an average of **1 Adventist congregation** for every **89,000 people!** [2].

1. Department of Economic and Social Affairs, United Nations, "World's Population Increasingly Urban with More than Half Living in Urban Areas," *United Nations* (July 10, 2014) <https://www.un.org/development/desa/en/news/population/world-urbanization-prospects.html>; retrieved November 10, 2015.
2. A. Oliver, "Adventist Church Implements Assessment Plan for Urban Mission," *Adventist News Network* (October 25, 2013) <http://news.adventist.org/en/all-news/news/go/2013-10-25/adventist-church-implements-assessment-plan-for-urban-mission/>; retrieved November 11, 2015.

“The work in the cities is the essential work for this time. When the cities are worked as God would have them, the result will be the setting in operation of **a mighty movement such as we have not yet witnessed**” [1].


1. E. G. White, Medical Ministry (Pacific Press Pub, 1963), p. 304.





“The importance of making our way in the great cities is still kept before me. For many years the Lord has been urging upon us this duty, and **yet we see but comparatively little accomplished in our great centers of population” [1].**

1. E. G. White, A Call to Medical Evangelism and Health Education (TEACH Services, Inc., 1997), p. 14.



“the Savior mingled with men as one who desired their good. He showed His sympathy for them, ministered to their needs, and won their confidence. Then He bade them, ‘Follow Me.’” [1]

1. E. G. White, *The Ministry of Healing* (Review & Herald, 1905), p. 143.



Use **data science** to understand the **needs** of people in **New York City**.

The background of the slide is a complex, abstract network of nodes and connections. The nodes are represented by circles of varying sizes and shades of gray, some appearing as bright white highlights. These nodes are interconnected by a dense web of thin, light gray lines, creating a sense of a large-scale, interconnected system. The overall aesthetic is technical and data-driven, typical of network science or data visualization.

Data Science can be defined as the study of the generalizable extraction of knowledge from data [1].

1. V. Dhar, "Data science and prediction," *Commun. ACM* , 56 (12, 2013), pp. 64-73.

Why do we need a new term like **data science** when we have had statistics for centuries?

1. The raw material, the “**data**” part of data science, is increasingly **heterogeneous** and **unstructured**.
2. Traditional database methods are *not* suited for **knowledge discovery**.

Unlike database querying, which asks “What data satisfies this pattern (query)?”

discovery asks “What **interesting** and **robust patterns satisfy** this **data**?”

The Digital Universe is Huge

- The digital universe is **doubling in size every two years**.
- By **2020** it will reach **44 zettabytes**, or **44 trillion gigabytes** [1].
- These facts have motivated **companies** and **scientists** in the last years to find new ways to understand **big data** in the digital universe.

1. IDC, "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," *EMC Corporation* (April, 2014) <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>; retrieved January 27, 2015.

- **Big data** is a term that can be used to describe data sets so **large** and **complex** that they become difficult to work with using standard techniques [1].
- **Big data is the next big thing. The new oil** [2].

1. C. Snijders, U. Matzat, and U. D. Reips, “‘Big Data’: Big Gaps of Knowledge in the Field of Internet Science,” *International Journal of Internet Science* 7, no. 1 (2014): 1-5.

2. P. Rotella, “Is Data the New Oil?,” *Forbes* (April 2, 2012) www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/; retrieved January 28, 2015.

My Way Towards Research on Data Science

2014

2015

2016

2017

Understanding Data

Software (IJSC, SERP 2014)

Health (IUPESM 2015)

Geoscience (ICAI 2015)

Smart Cities (ICAI 2015)

*Full references are available on
www.harveyalferez.com*

My Way Towards Research on Data Science

2014

2015

2016

2017



**Is it possible to use data science to
understand the needs of people in big
cities?**

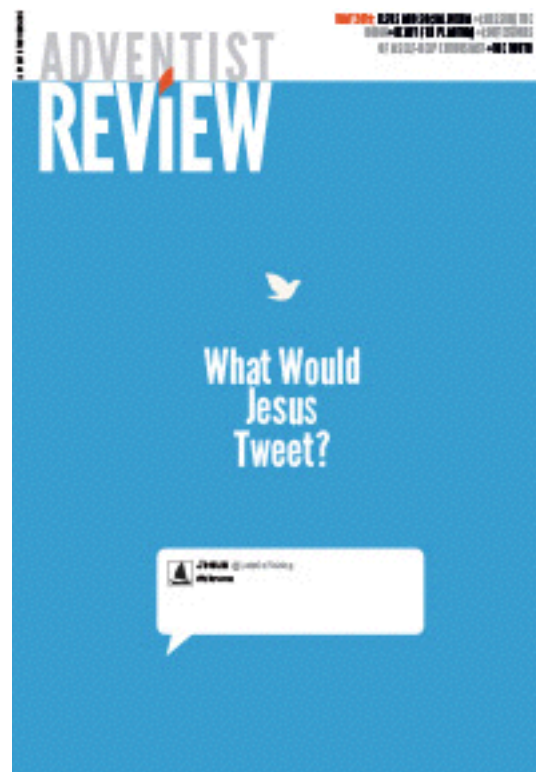
My Way Towards Research on Data Science

2014

2015

2016

2017



**Big Data for
Reaching
a Big World**



My Way Towards Research on Data Science

2014

2015

2016

2017



**Tweeting in New
York City, Data
Science Can
Teach Us to
Sympathize**

Which *data* to use to understand the needs of people in big cities?



Twitter is the largest searchable archive of human thought, that's public, that's ever existed [1].

1. T. Simonite, "Twitter Boasts of What It Can Do with Your Data," *MIT Technology Review* (October 21, 2015) <http://www.technologyreview.com/news/542711/twitter-boasts-of-what-it-can-do-with-your-data/>; retrieved November 10, 2015.

Reaching People's Tweets

Sentiment analysis was used to discover the **needs** of people from tweets.

The **computational study** of **opinions**, **sentiments**, and **emotions** expressed in text [1].

Sentiment analysis has been **satisfactory** used to classify users' sentiments in tweets [2].

1. B. Ling, "Sentiment Analysis and Subjectivity," in N. Indurkha, & F. J. Damerau, *Handbook of Natural Language Processing*, 2nd ed., (Boca Raton, FL: Chapman & Hall, 2010), pp. 627-665.
2. A. Tumasjan, T. O. Sprenger, & P. G., Sa. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. AAAI, (2010), pp. 178-185.

Reaching People's Tweets

- **Tweets** are **classified**
 - as ***positive*** when they communicate a positive sentiment, such as happiness;
 - as ***negative*** when a negative sentiment is attached to them (e.g. sadness);
 - and as ***neutral*** when no emotions are implied.

Reaching People's Tweets

Machine learning [1] was used as a tool to differentiate tweets with *positive*, *negative*, and *neutral* sentiments.

Machine learning explores the study and construction of **algorithms** that can **learn from** and **make predictions on data**.

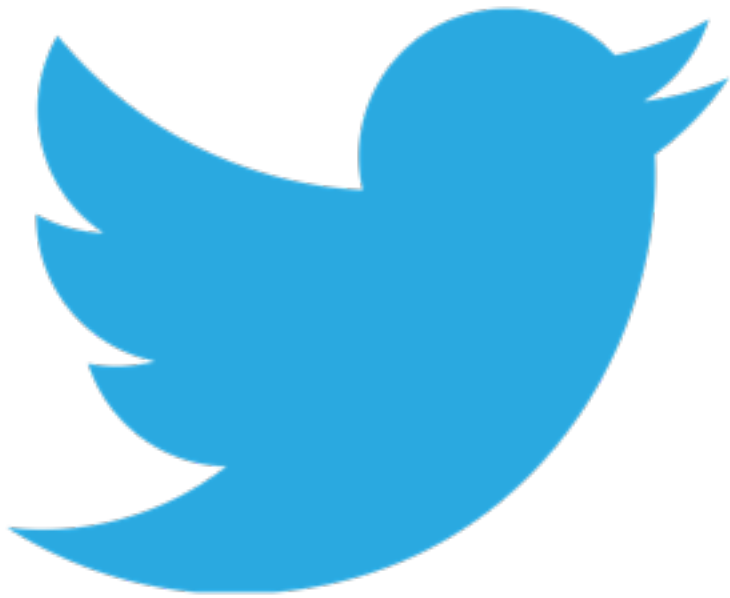
1. A. Go, R. Bhayani, & L. Huang, *Twitter Sentiment Classification using Distant Supervision* (Stanford University, 2009)

Reaching People's Tweets

“Most of us are trained to believe theory must originate in the human mind based on prior theory, with data then gathered to demonstrate the validity of the theory. **Machine learning turns this process around.** Given a large trove of data, the computer taunts us by saying, ‘If only you knew what question to ask me, I would give you some very interesting answers based on the data.’” [1]

1. V. Dhar, “Data science and prediction,” *Commun. ACM*, 56 (12, 2013), pp. 64-73.

Listening Closely to the Birds



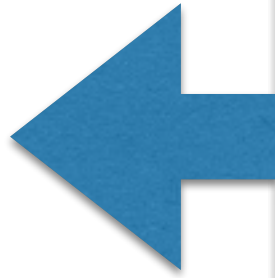
Over a period of six weeks (September 22 to November 3, 2015), we collected 2,084 tweets from New York City, 1,633 of them bearing positive sentiments and 451 expressing negative sentiments. Tweets with neutral sentiments were not collected.

Listening Closely to the Birds

30 specified keywords:

Adventist, addiction, Bible, children, Christ, church, contamination, divorce, education, elderly, exercise, family, God, health, Jesus, obesity, peace, poverty, religion, rest, safety, salvation, Savior, stress, teenagers, teens, terrorism, vegetarian, violence, youth

1. Collects



```
def main(argv):
    global user_city
    global EXPRESSION
    global radio

    try:
        opts, args = getopt.getopt(argv, "h:c:r:l:", ["expr=", "city=", "ra"])
    except getopt.GetoptError:
        print os.path.basename(__file__) + ' -e <expression> -c <city> -r <radio>'
        sys.exit()

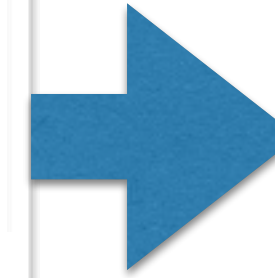
    for opt, arg in opts:
        if opt == '-h':
            print os.path.basename(__file__) + ' -e <expression> -c <city> -r <radio>'
            sys.exit()
        elif opt in ("-e", "--expr"):
            EXPRESSION = arg
        elif opt in ("-c", "--city"):
            user_city = arg
        elif opt in ("-r", "--radio"):
            radio = arg

    tweets = collect_tweets(user_city, EXPRESSION, radio)

    return negative_tweets, positive_tweets
```



2. Classifies



+ and -

3. Stores



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Positive	455	friends	new family and...	https://t.co/123GfNwDQ	71.99278645	40.75052798													0.725
2	Positive	455	hater and family	near and far and my	jetblue family@noo lets...	https://t.co/0uQjyobA	73.77555711	40.64579177												0.725
3	Positive	455	1003442230351360	*248009151*	AliceLaurissa	2015-10-05 11:58:10	*Hanging with my	nieces for my family	bdy dinner. Love these	guys. #AliceLaurissa	#BirthDaygirl...	https://t.co/13GL5d8qic	73.74631052	40.70295328						0.725
4	Positive	455	0915559132000254	*2472332402*	cityscapesny	2015-10-05 06:08:57	*TOMORROW NIGHT	#ADAMOSS	@theraalkiss	#THOTUNEBUG	HNY Family	invades	#rooperstarmondays	with The...	https://t.co/luQjYlenAg	73.91181951	40.725			0.725
5	Positive	455	my man	my family. I love	you SO much	Mama Nancy- you are	a wonderful mom- we	celebrate YOU...	https://t.co/1qQCKCufNv	73.8831482	40.7555008									0.725
6	Positive	455	08460212287590400	*50217891*	ALoveLyToy	2015-10-05 02:29:01	*Love my family!!!!	HAPPY BIRTHDAY TO	MY BABY SISTER!!!!	LOVE U!!	DEATH DO US AND	BEYOND!!!!	LIBRA...	https://t.co/F4HGxEOemc	73.9832231	40.7449989				0.725
7	Positive	455	0848490111653376	*67786749*	MannyR999	2015-10-05 01:42:26	*This is my beautiful	family God bless my	Beautiful family I love	you...	https://t.co/1T4PLH6Qku	74.0064	40.7342							0.725
8	Positive	455	0847671297536001	*110518597*	WJBLA	2015-10-05 01:39:11	*Suite life w/ the	family! Thanks for the	great memories! Mets	won to put the cherry	on top...	https://t.co/5M4EGQ45	73.84604275	40.75685645						0.725
9	Positive	455	0826880317128704	*86459013*	EmreDemyel	2015-10-05 00:16:34	*I always love	spending time with my	family. @ Tamashi	Ramen https://t.co/owd8qEDZ	73.92671723	40.76262602								0.725
10	Positive	455	0826929963655168	*2658278567*	FortheLoveof	2015-10-04 23:50:32	*My family loves	this crest	Roughshield it is	the best!! With	cleaner teethes and	freshier breath for...	https://t.co/w794PD8Q25	74.0064	40.7142					0.725
11	Positive	455	0818621313302529	*277773125*	...wavyday	2015-10-04 21:43:45	*Happy Birthday	Cinays #family @	Sweet Chick	https://t.co/1qQCKCufNv	73.9573975	40.758399								0.725
12	Positive	455	0795826457870396	*1571019229*	intoTheRowlog	2015-10-04 22:13:39	*Happy Birthday	to october_bobi	Enjoying another	round of drinks just	for you #family	#instacool...	https://t.co/4yftzAK798	74.2596588	40.7449684					0.725
13	Positive	455	0785188311413504	*15436343*	the_captain66	2015-10-04 21:30:54	*What a perfect	day to watch the	FLAHOIT SOUND	NY METS with my	favorite meta	loving family	ginting...	https://t.co/1uXhDncrCid	73.84604275	40.75685645				0.725
14	Positive	455	0760792846036993	*753249602*	Cinderella_mia	2015-10-04 19:53:58	*Sunday tapping	with my New York	family @	Champion Dance	Studios https://t.co/21G4b57PK5	73.9903564	40.7552299							0.725
15	Positive	455	0745063175405472	*851804250*	swellingtonmids	2015-10-04 18:32:19	*Happy Sunday!!	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	#family #Sunday	0.725
16	Positive	455	0745035844100099	*14768046*	baristatnet	2015-10-04 18:51:21	*Love this family	that turned out to	honor Yogi @	Yogi Berra Museum	& Learning Center	https://t.co/EB5nQ14gA3	74.19464988	40.86826079						0.725
17	Positive	455	073051783433089	*14301804*	travellinghild	2015-10-04 18:17:30	*#family with the	family! So excited!!!	@ Richard	Rodgers Theatre -	#niederlanderbay	for Hamilton (NY)	https://t.co/wkxBOu472A	73.98070732	40.75930971					0.725
18	Positive	455	0725601553305600	*71097356*	BRONXGURL12	2015-10-04 17:34:07	*#aboutlastnight	Family Fun with	My Brother My	Daughter And My	Godkids...	https://t.co/52swKyY8N	73.98872539	40.73639096						0.725
19	Positive	455	07188311413504	*15436343*	the_captain66	2015-10-04 21:30:54	*What a perfect	day to watch the	FLAHOIT SOUND	NY METS with my	favorite meta	loving family	ginting...	https://t.co/1uXhDncrCid	73.84604275	40.75685645				0.725
20	Positive	455	069463791505409	*255800296*	djordanmetz	2015-10-04 15:50:16	*Best morning	ever!! Hangin	with my niece	in Central Park!	#family #NYC	#newyork	#outdoors @	Bufo... https://t.co/1yrCGjv6P8	73.9701224	40.77005404				0.725
21	Positive	455	068401374631040	*274570873*	kingjellyd	2015-10-04 14:48:52	*Blessed that I	can worship	along side my	family watching	this wonderfu...	https://t.co/1x5dXUvsn	73.83593728	40.8957819						0.725
22	Positive	455	0674928929386496	*302211068*	JohnTripp59	2015-10-04 14:12:45	*Sunday morning	NEOA walk	with the family	TeamBec @	Foley Square	https://t.co/12qOCSKt	74.00297031	40.7146602						0.725
23	Positive	455	06442094460028	*290453632*	felipeart80	2015-10-04 08:17:11	*Happy birthday	to me!! #birthday	#family #sister	#brother #nyc	@ Therapy NYC	https://t.co/1uXhDncrCid	73.98070732	40.7640419						0.725
24	Positive	455	0634401423360	*105627214*	rlang3r	2015-10-04 07:31:51	*Great to have	role models	of family and	relationship like	President and	Mrs Obama!!!	Happy Anniversary!!!	https://t.co/1uXhDncrCid	73.98070732	40.71022649				0.725
25	Positive	455	060138496081920	*248939854*	CANTHIMANBABY	2015-10-04 06:16:38	*SQUAD UP	FAMILY RIGHT	HERE HAPPY	BDAY RICK I	LOVE YOU KID	#UnionGpns	#UnionGpnsManagement...	https://t.co/1uXhDncrCid	73.98070732	40.71022649				0.725
26	Positive	455	054826630054784	*31569621*	jeffbeacher	2015-10-04 05:49:27	*We are live	from New York	!!! On Saturday	night !!! #s	#family the	queen hosting @	mileyevans @...	https://t.co/1uXhDncrCid	73.97891732	40.75882475				0.725
27	Positive	455	054826630054784	*31569621*	jeffbeacher	2015-10-04 05:49:27	*We are live	from New York	!!! On Saturday	night !!! #s	#family the	queen hosting @	mileyevans @...	https://t.co/1uXhDncrCid	73.97891732	40.75882475				0.725
28	Positive	455	05060069916672	*103855993*	nun04758dva	2015-10-04 03:53:24	*I swear ain't	nothing like	family!!! I	LOVE MY	COUSINS!!!	#familyfirst...	https://t.co/02nmvCfNME	73.96116732	40.71674569					0.725
29	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
30	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
31	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
32	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
33	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
34	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
35	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
36	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
37	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
38	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
39	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725
40	Positive	455	0503477429411840	*2867179032*	FUP_NYC	2015-10-04 02:51:29	*Family after	happy 50th	Simon and	Barak's	happy 200th	old man...	https://t.co/1uXhDncrCid	73.98070732	40.7406578					0.725

Positive Tweet about Vegetarian Food

- Positive
- her*
- 2015/10/02 02:08:16
- I want to be vegetarian. I really do. @arrogantwine @ East Williamsburg Brooklyn <https://t.co/rpatPGyhXw>
- -73.939 (longitude)
- 40.714 (latitude)

Negative Tweet about Family

- Negative
- And*
- 11/10/15 18:48
- My ex has made them hate me, but I still see the children in my dreams.
- -73.74663446 (longitude)
- 40.69729011 (latitude)

Listening Closely to the Birds

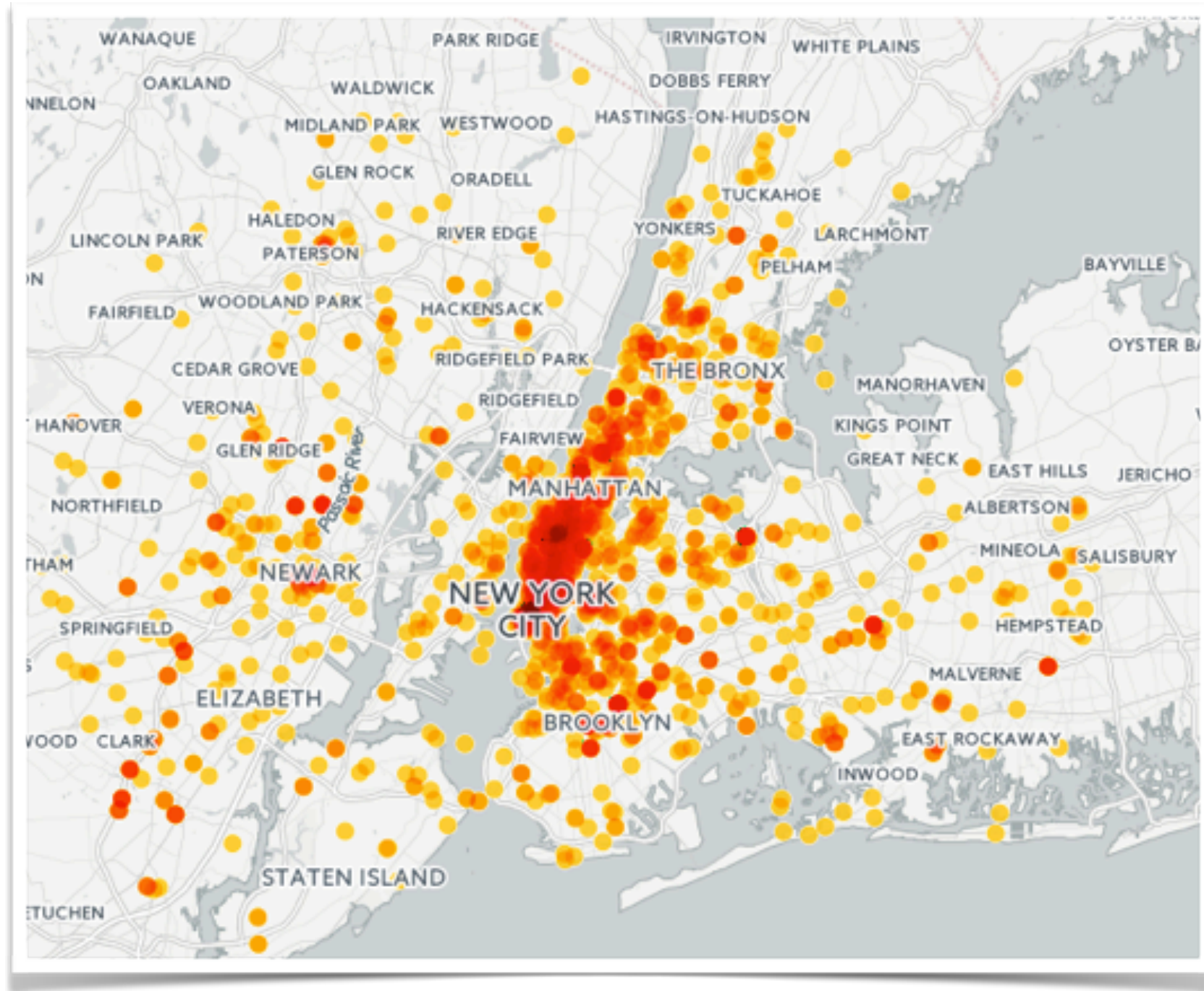


Figure 1. Intensity of tweets in New York City

Listening Closely to the Birds

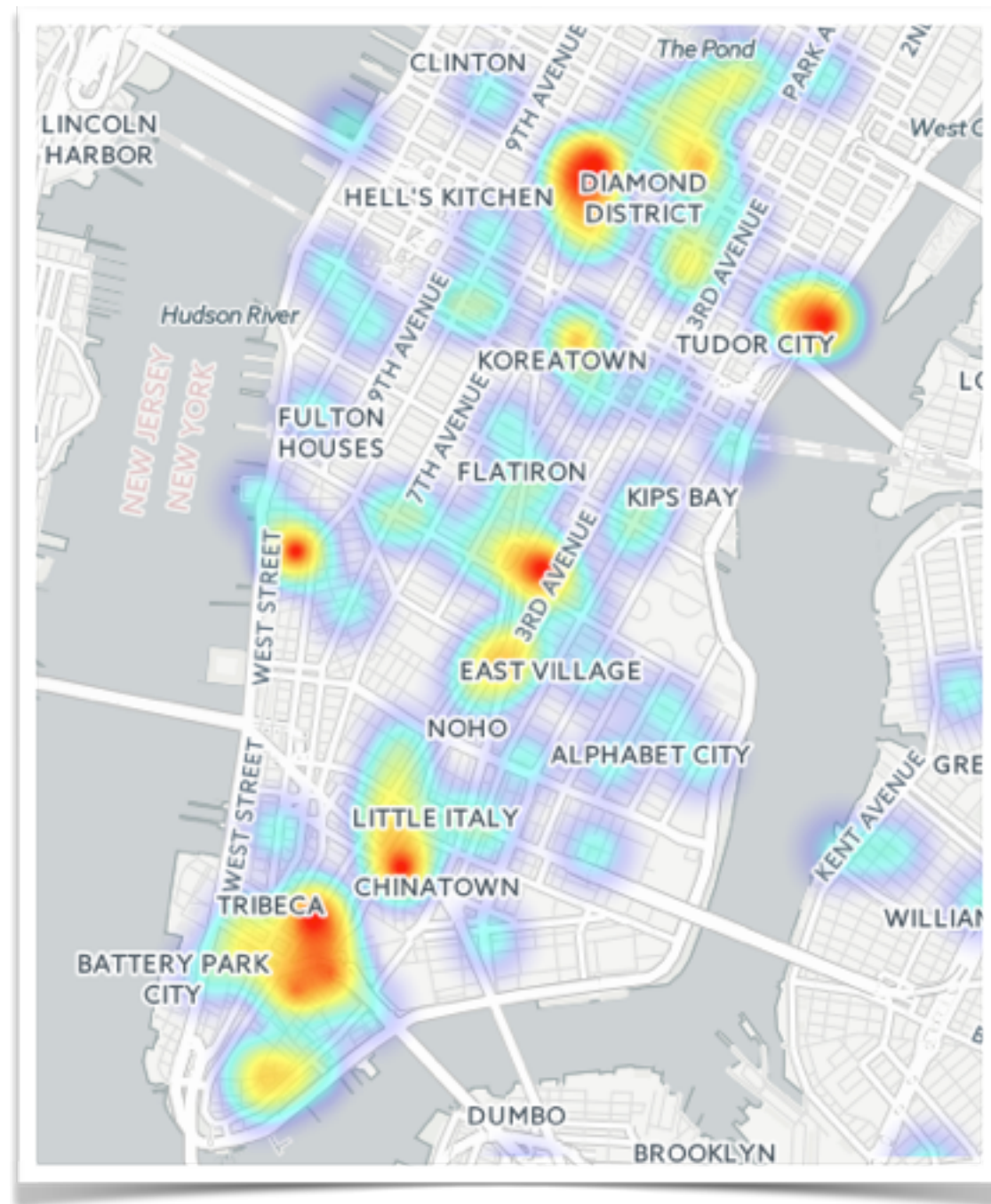


Figure 2. Areas with negative tweets in Manhattan

Upbeat and Downbeat

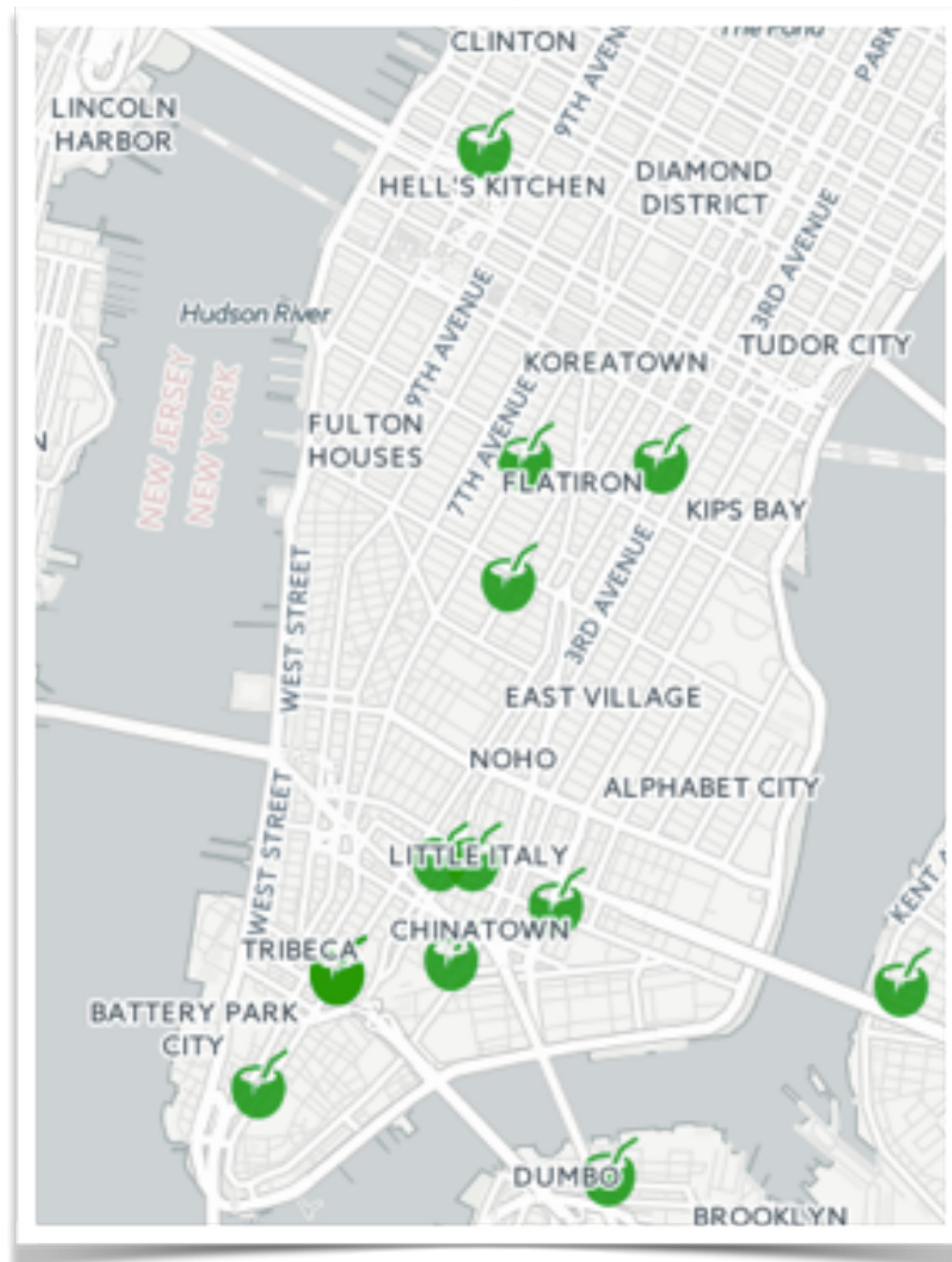


Figure 3. Positive tweets about vegetarian food in Manhattan

Upbeat and Downbeat

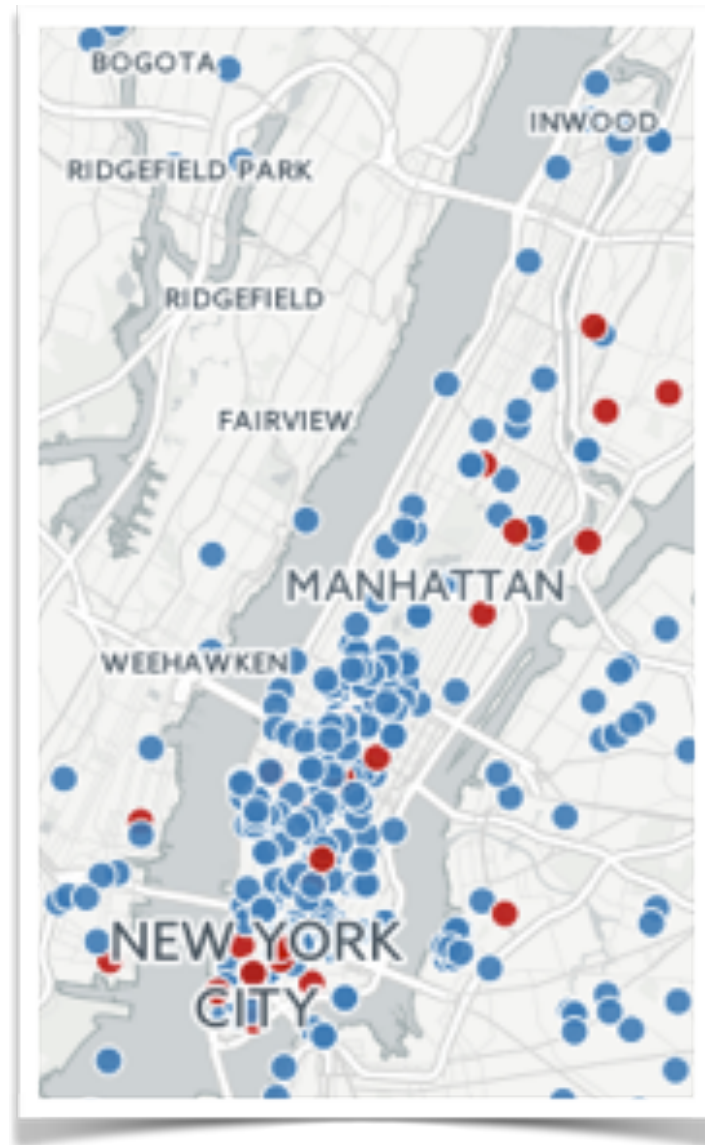


Figure 4. Positive [blue] and negative [red] tweets about family in Manhattan.

Data science has the potential to help us understand the **needs of people in big cities** in an **unprecedented way**.



What's Next?

- Other cities, other datasets: London, Mumbai, Buenos Aires, and Mexico City
- Build an easy-to-use tool that pastors, evangelists, and church leaders will be able to use to understand the needs of people in their cities

What's Next?

- Other areas: Health (Universidad de Montemorelos):
 - Find hidden patterns in thousands of dental records. School of Dentistry
 - Diagnosis of glaucoma by means of machine learning. Ophthalmological Clinic
 - Discover hidden reasons of maternal mortality in Mexico. School of Medicine

1. Let God grant us grace and
bless us;
shine on us,

2
**known on earth,
salvation becomes known
among all the nations.**

3 Let the people thank you, God!
Let all the people thank you!

Psalms 67:1-3 (CEB)



Understanding the Needs of People in Big Cities through Data Science

Harvey Alférez, Ph.D.

Global Software Lab

School of Engineering and Technology

Universidad de Montemorelos, Mexico

www.harveyalferez.com

