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].

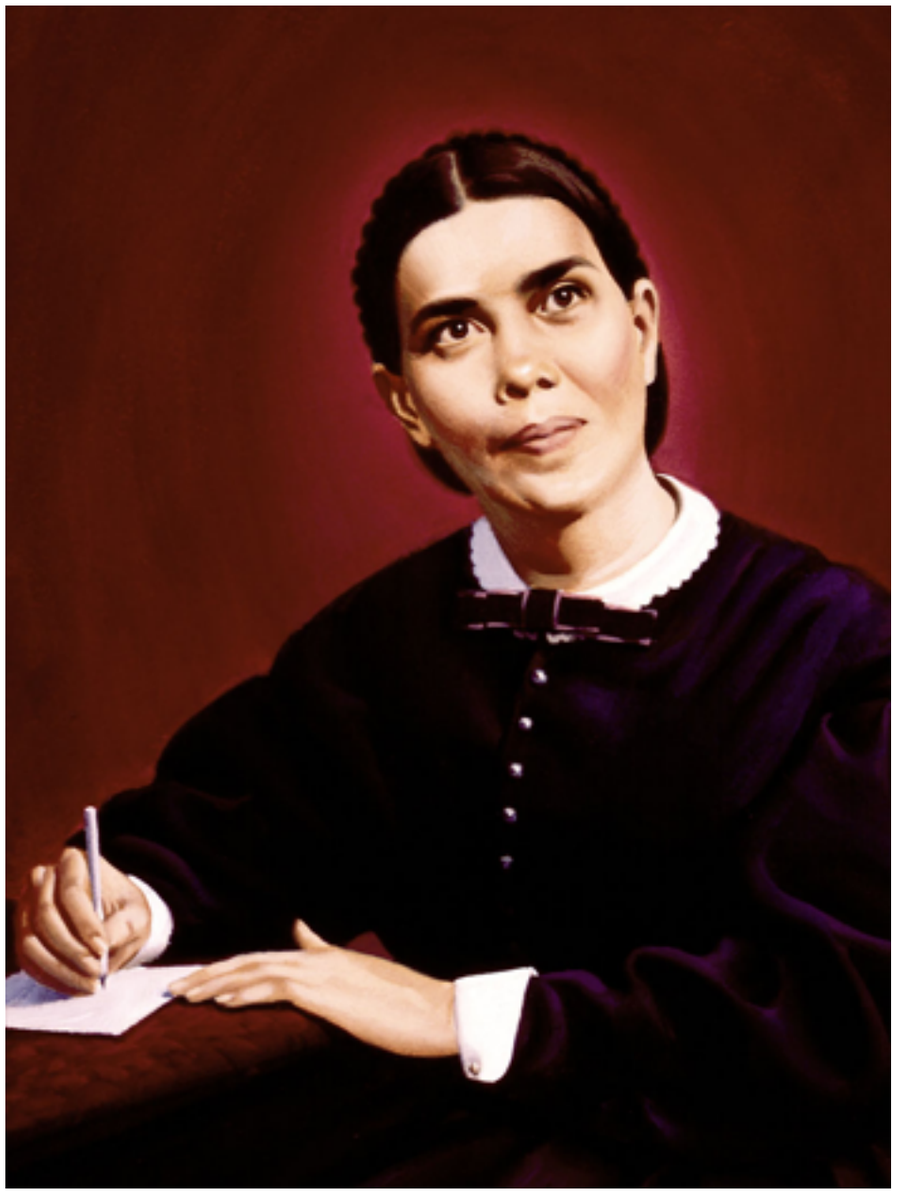
- There are more than **500** cities with a population of **1 million or more people**. However, these cities have an average of **1 Adventist congregation** for every **89,000 people**! [2].

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**.

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 termed 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

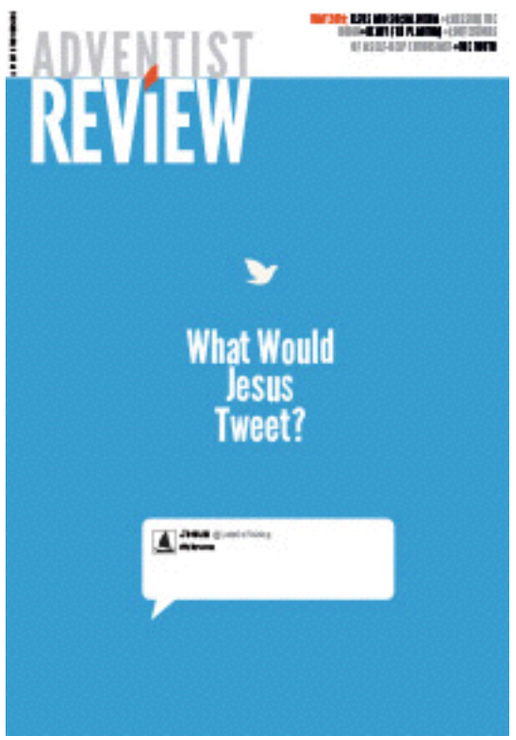
My Way Towards Research on Data Science

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

Andrews Research Conference

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. Indurkhya, & 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
2. Classifies
+ and -
3. Stores
python

Positive Tweet about Vegetarian Food

- Positive
- her*
- 2015/10/02 02:08:16
- I want to be vegetarian. I really do. @arrogantswine @ 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

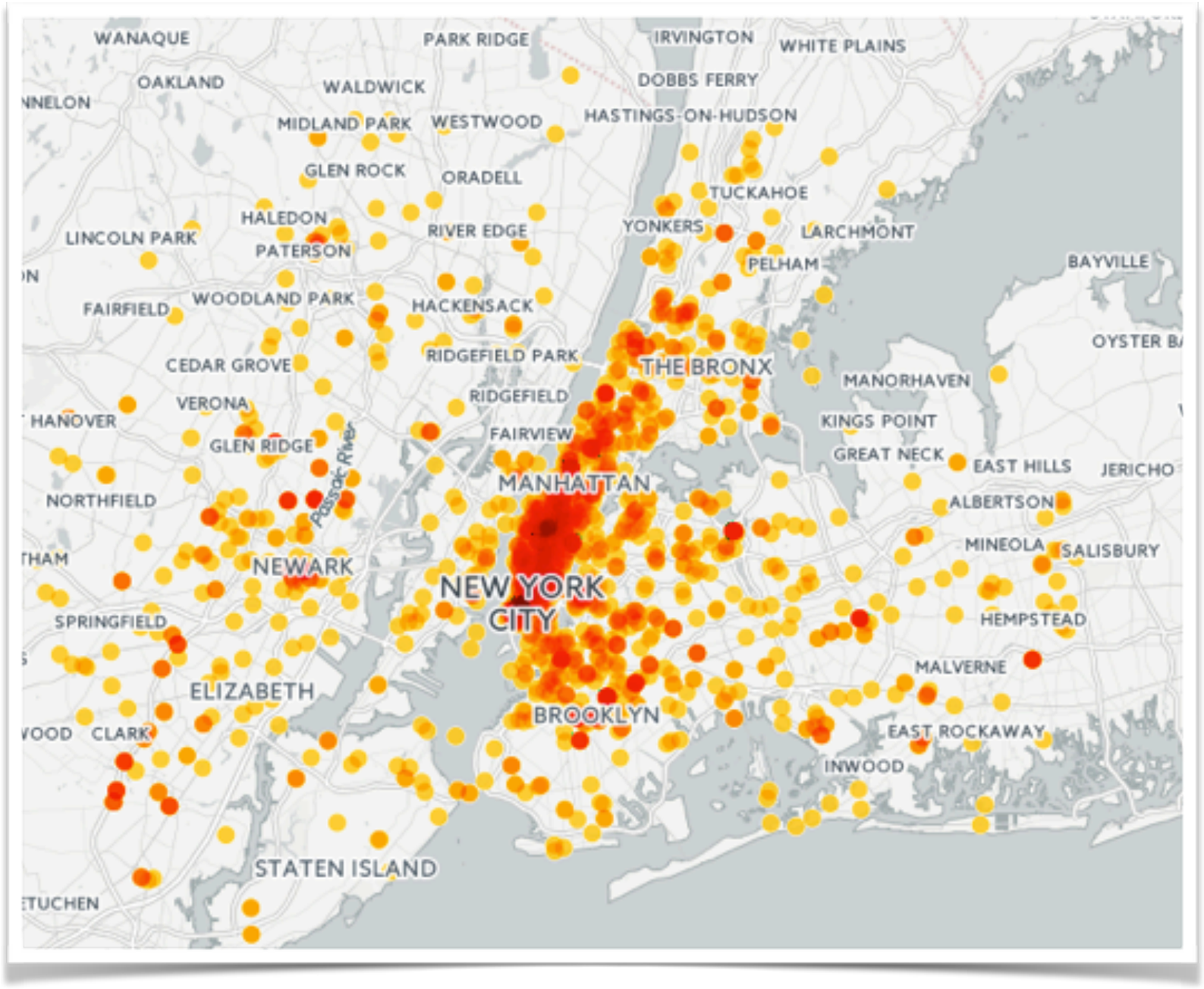


Figure 1. Intensity of tweets in New York City

Listening Closely to the Birds

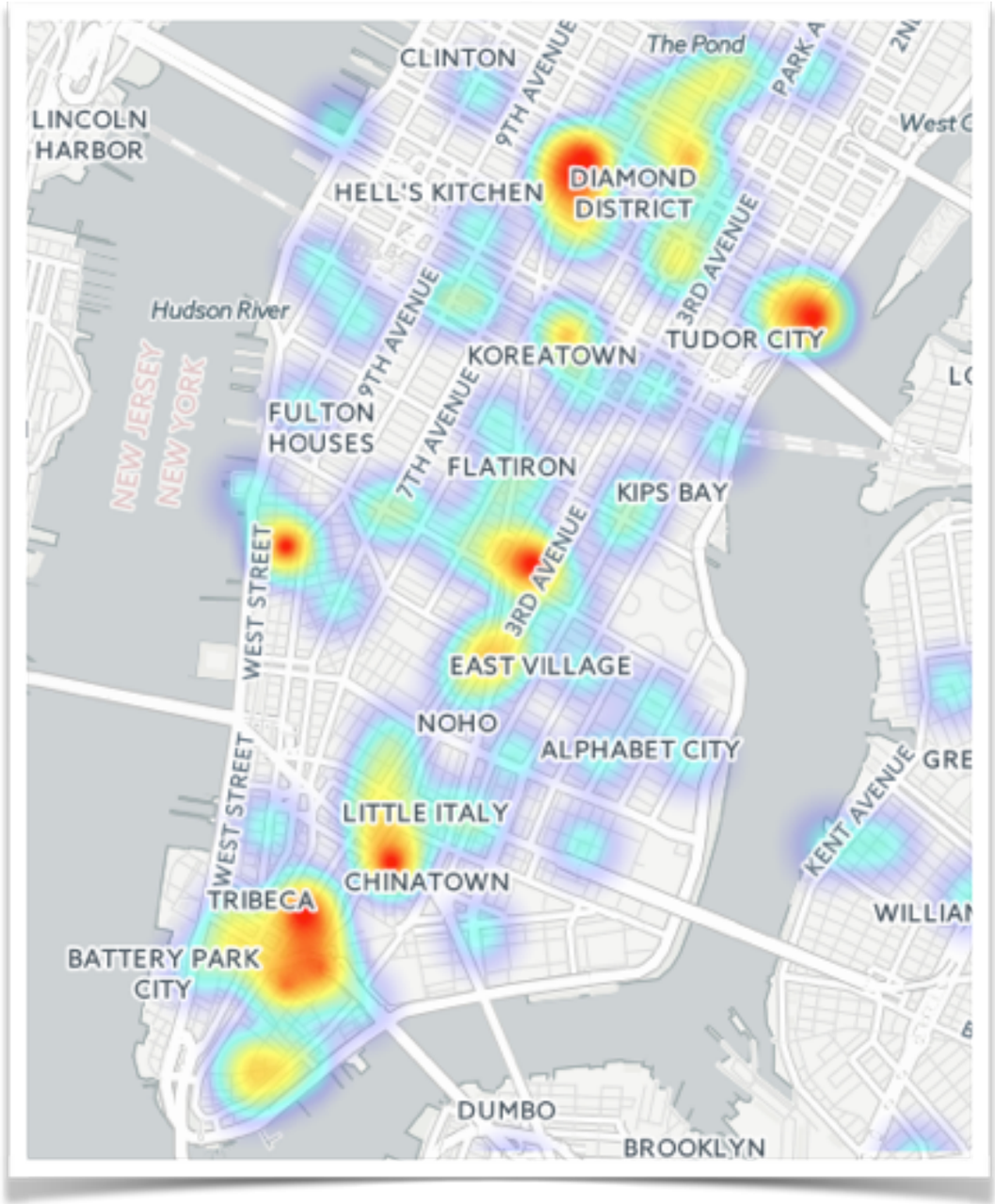


Figure 2. Areas with negative tweets in Manhattan

Upbeat and Downbeat

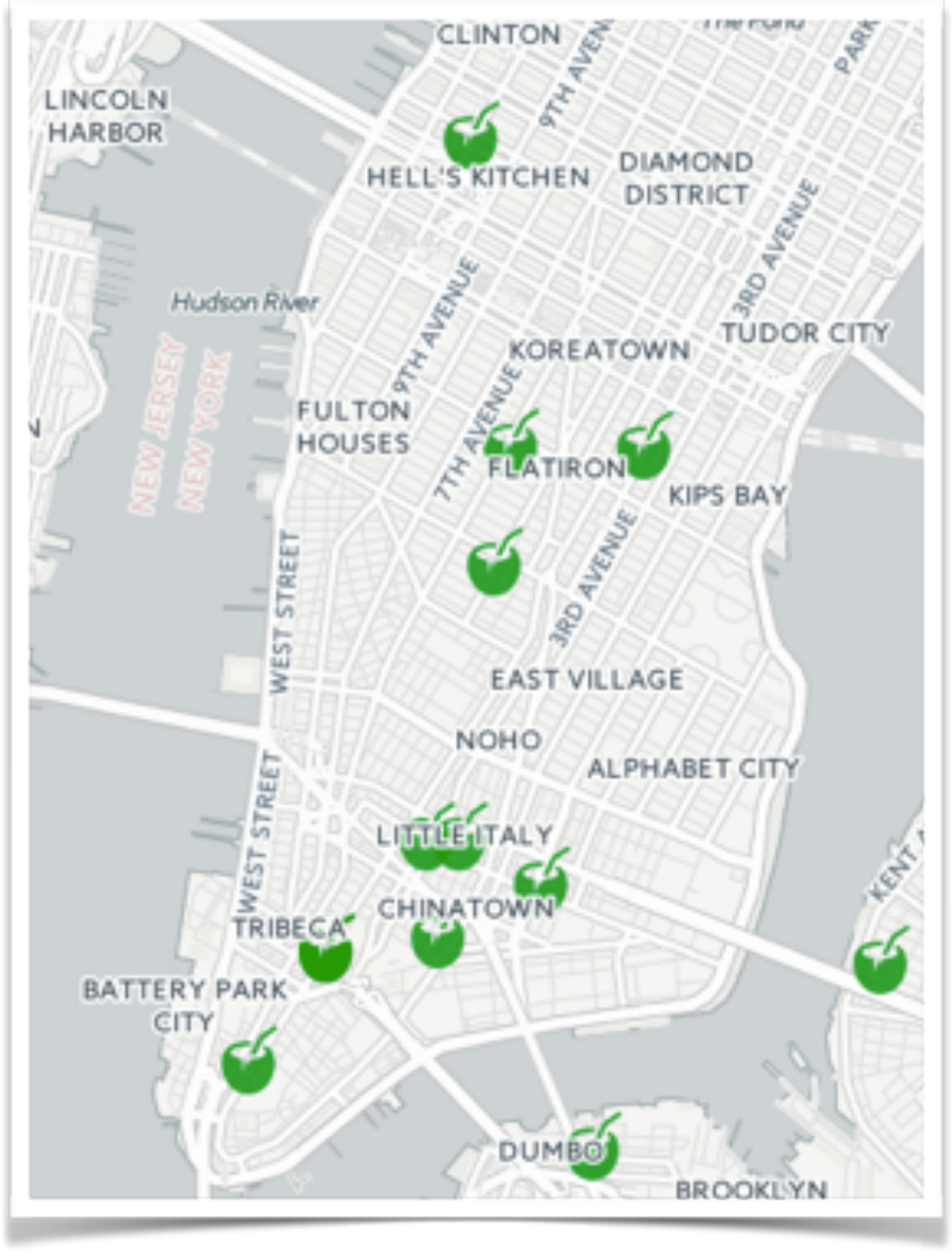


Figure 3. Positive tweets about vegetarian food in Manhattan

Upbeat and Downbeat

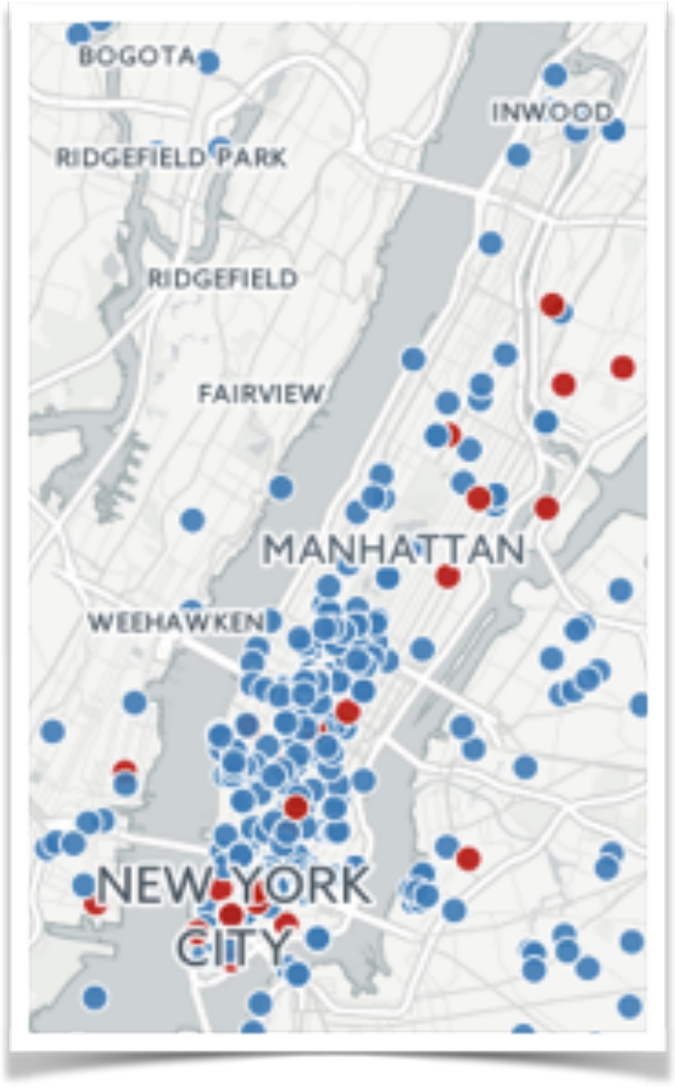


Figure 4. Positive [blue] and negative [red] tweets about family in Manhattan.

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!

Psalm 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