Sacándole el Jugo a los Datos en el Cumplimiento de la Misión

Harvey Alférez, Ph.D.
Global Software Lab,
Facultad de Ingeniería y Tecnología,
Universidad de Montemorelos, México
www.harveyalferez.com
@harveyalferez

Agenda

1. Periodismo de datos
2. Descubriendo las necesidades de las personas con ciencia de datos
3. Comprendiendo a la feligresía con ciencia de datos
4. Conclusiones

Agenda

1. **Periodismo de datos**
2. Descubriendo las necesidades de las personas con ciencia de datos
3. Comprendiendo a la feligresía con ciencia de datos
4. Conclusiones

Periodismo de Datos

Recabar y **analizar grandes cantidades de datos** mediante software especializado y hacer comprensible la información a la audiencia a través de artículos, infografías, visualizaciones de datos o aplicaciones interactivas [1].

1. Cabra, M. (2013). Mar Cabra: «Hem d'acostumar les institucions a donar i als ciutadans a demanar». URL: http://lab.cccb.org/ca/mar-cabra-em-dacostumar-a-les-institucions-a-donar-i-als-ciutadans-a-demanar/

Big Data for Reaching a Big World

Andrews Research Conference

2. Alférez, G.H. (2015). Big Data for Reaching a Big World. Adventist Review, 192(11), 47-51

Análisis de datos masivos

para entender cómo la **cultura** percibe nuestras **creencias fundamentales**.

Datos Masivos y Nuestra Iglesia

- En este estudio, el análisis computacional de datos se basó en ***culturomics***.

- Recolección y análisis de datos masivos para el estudio de la cultura humana [3].

3. Michel, J.-B. et al. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331(6014):176–182.

Datos Masivos y Nuestra Iglesia

- El conjunto de datos utilizado en el experimento se puede descargar de:

  https://books.google.com/ngrams

  - Este conjunto de datos está compuesto por textos digitalizados que contienen cerca del **4% de todos los libros escritos entre 1800** y **2008** (5,195,769 libros).

  - **Libros** en **inglés** (361 mil millones de palabras) y en **español** (45 mil millones de palabras)

Datos Masivos y Nuestra Iglesia

- Este cuerpo **no puede ser leído por un humano** [4]:
  - Si se intenta leer solo las entradas en inglés del año 2000, a un ritmo razonable de 200 palabras / min, sin interrupciones para comer o dormir, se tomaría 80 años.

4. Michel, J.-B. et al. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331(6014):176–182.

Datos Masivos y Nuestra Iglesia

- El **Google Ngram Viewer** fue utilizado para visualizar los resultados (**https://books.google.com/ngrams**).

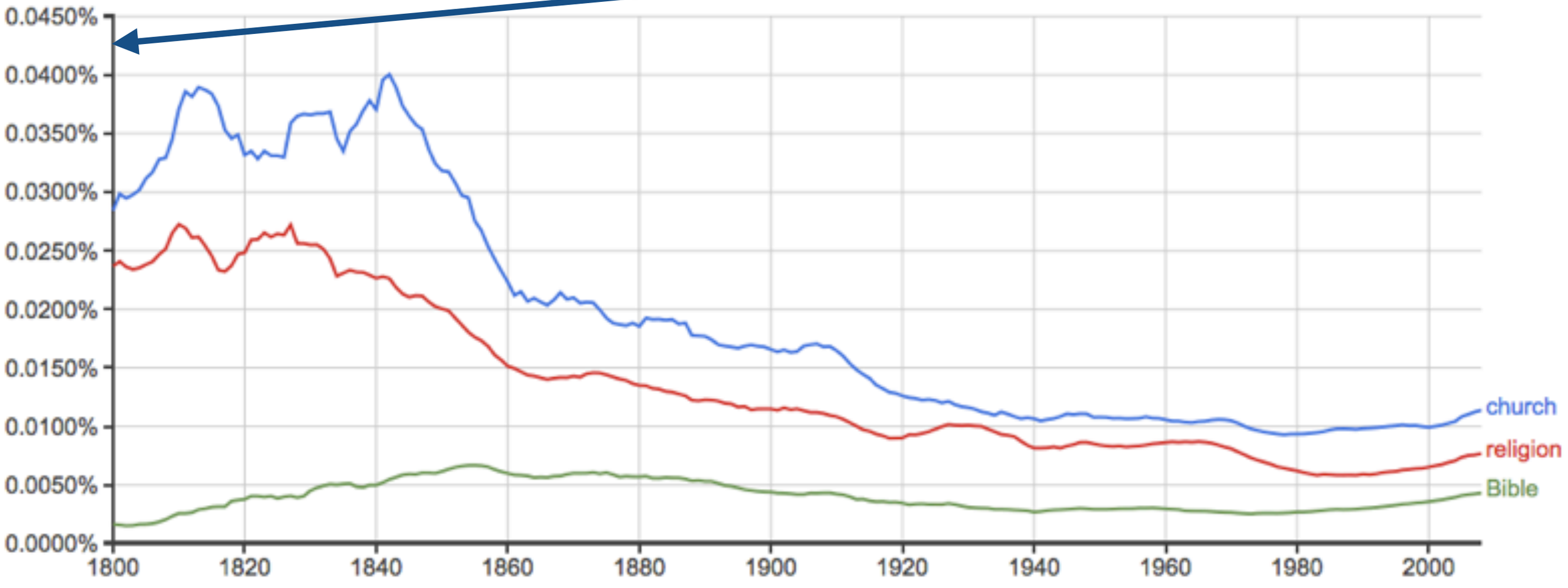
- Un **1-gram** es una cadena de caracteres ininterrumpida por espacios. Esto incluye palabras (“car”, “MICHIGAN”) pero también números (“3.14”) y errores de tipografía (“excesss”).

- Un **n-gram** es una secuencia de 1-grams, tal como la frase “stock market” (un 2-gram) y “the United States of America” (un 5-gram).

Google Books

Iglesia, Religión y Biblia

N-gram Frequency (Corpus of English Books)



Tendencias de Church, Religion y Bible

Years

Secularización

Tendencia de Secularization

Creación

Creationism vs. Theory of Evolution

El Sábado

Tendencia de Sabbath

La Naturaleza del Hombre

Tendencias de Pecado y Salvación

Segunda Venida de Cristo

Segunda venida de Cristo, Tendencias en Inglés vs. Español

Estilo de Vida Saludable

Interés Creciente en Estilo de Vida Saludable y Vegetarianismo

Living in the Age of Anxiety as Revealed in Google's Big Data

5. Alférez, G.H. (2017). Anxiety as revealed in Google's big data. URL: http://advindicate.com/articles/2017/11/23/living-in-the-age-of-anxiety-as-revealed-in-googles-big-data

Excavando en los Datos Masivos de Google

Búsquedas de “anxiety” en los Estados Unidos de Norteamérica
de 2004 hasta 2017

Google Trends

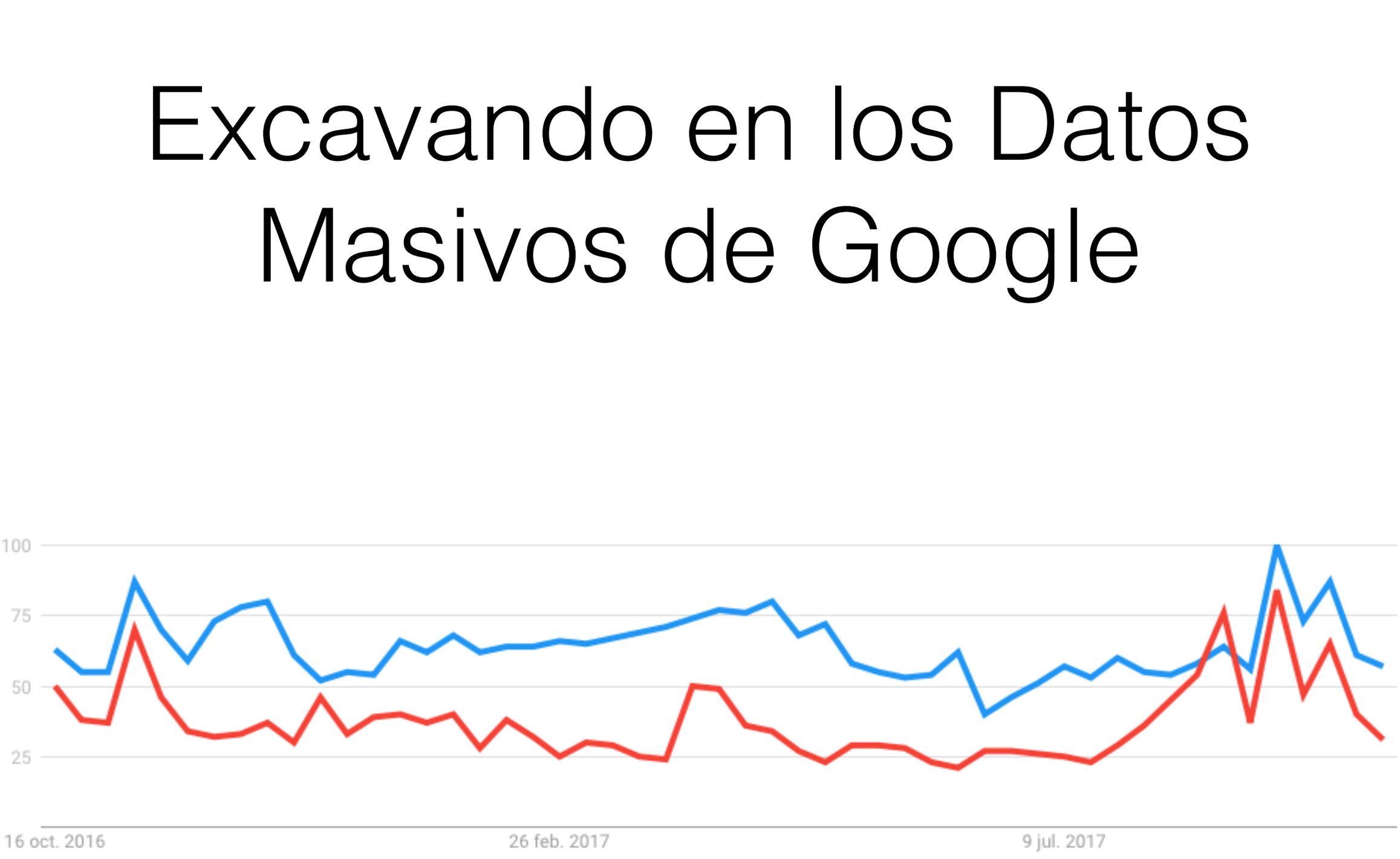
Excavando en los Datos Masivos de Google

Búsquedas de “uncertainty” en los Estados Unidos de Norteamérica en los últimos 5 años

Excavando en los Datos Masivos de Google

Búsquedas para “end of the world” en los Estados Unidos de Norteamérica de 2004 a 2017

Excavando en los Datos Masivos de Google



Búsquedas de “second coming” (en azul) y “Bible prophecy” (en rojo) en los Estados Unidos de Norteamérica de octubre 2016 a octubre 2017

Agenda

1. Periodismo de datos
2. **Descubriendo las necesidades de las personas con ciencia de datos**
3. Comprendiendo a la feligresía con ciencia de datos
4. Conclusiones

Understanding the Needs of People in Big Cities through Data Science

6. Alférez, G.H. (2016). Tweeting in New York City - Data Science Can Teach Us to Sympathize. Adventist Review, 193(2), 47-49

Las Ciudades Están Creciendo Rápidamente

- **66%** de la población mundial vivirá en zonas urbanas en **2050** [7].

- Hay más de **500** ciudades con una población de **1 millón o más de personas**. Sin embargo, estas ciudades tienen un promedio de **1 congregación adventista** por cada **89,000 personas** [8].

7. 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.

8. Oliver, A. "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.

“Cuando se trabaje en las **ciudades** como Dios quiere, el **resultado** será la puesta en **operación** de un **movimiento poderoso tal como nunca hemos presenciado hasta ahora**” [9].

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

“El método de Cristo solo dará verdadero éxito al llegar a la gente. El Salvador se mezcló con los hombres como alguien que deseaba su bien. Mostró su simpatía por ellos, ministró a sus **necesidades** y ganó su confianza. Luego les ordenó: "'Síganme.'” [10]

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

Utilización de **ciencia de datos** para entender las **necesidades** de la gente en la **ciudad de Nueva York**.

¿Qué ***datos*** usar para entender las necesidades de las personas en las grandes ciudades?

“Twitter es el archivo de pensamiento humano más grande que se puede buscar, que es público, y que alguna vez haya existido” [11] - *Chris Moody, Twitter’s vice president for data strategy*

11. Simonite, T., “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.

Entendiendo los Tweets

Análisis de sentimientos fue usado para descubrir las **necesidades** de las personas en sus tweets.

> El **estudio computacional** de **opiniones**, **sentimientos**, y **emociones** expresadas en un **texto** [12].

El análisis de sentimientos ha sido **satisfactoriamente** utilizado en la clasificación de sentimientos en tweets [13].

12. 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.

13. 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.

Entendiendo los Tweets

- **Tweets** están **clasificados**
  - como ***positivos*** cuando comunican un sentimiento positivo, tal como felicidad;
  - como ***negativos*** cuando un sentimiento negativo se adjunta a ellos (ej. tristeza);
  - y como ***neutro*** cuando no hay emociones implicadas.

Entendiendo los Tweets

Aprendizaje automático [14] fue utilize como una herramienta para diferenciar entre tweets con sentimientos *positivos*, *negativos*, y *neutros*.

Aprendizaje automático explora el estudio y la construcción de algoritmos que pueden usar los datos para **aprender** y **hacer predicciones**.

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

Aprendiendo de los Tweets

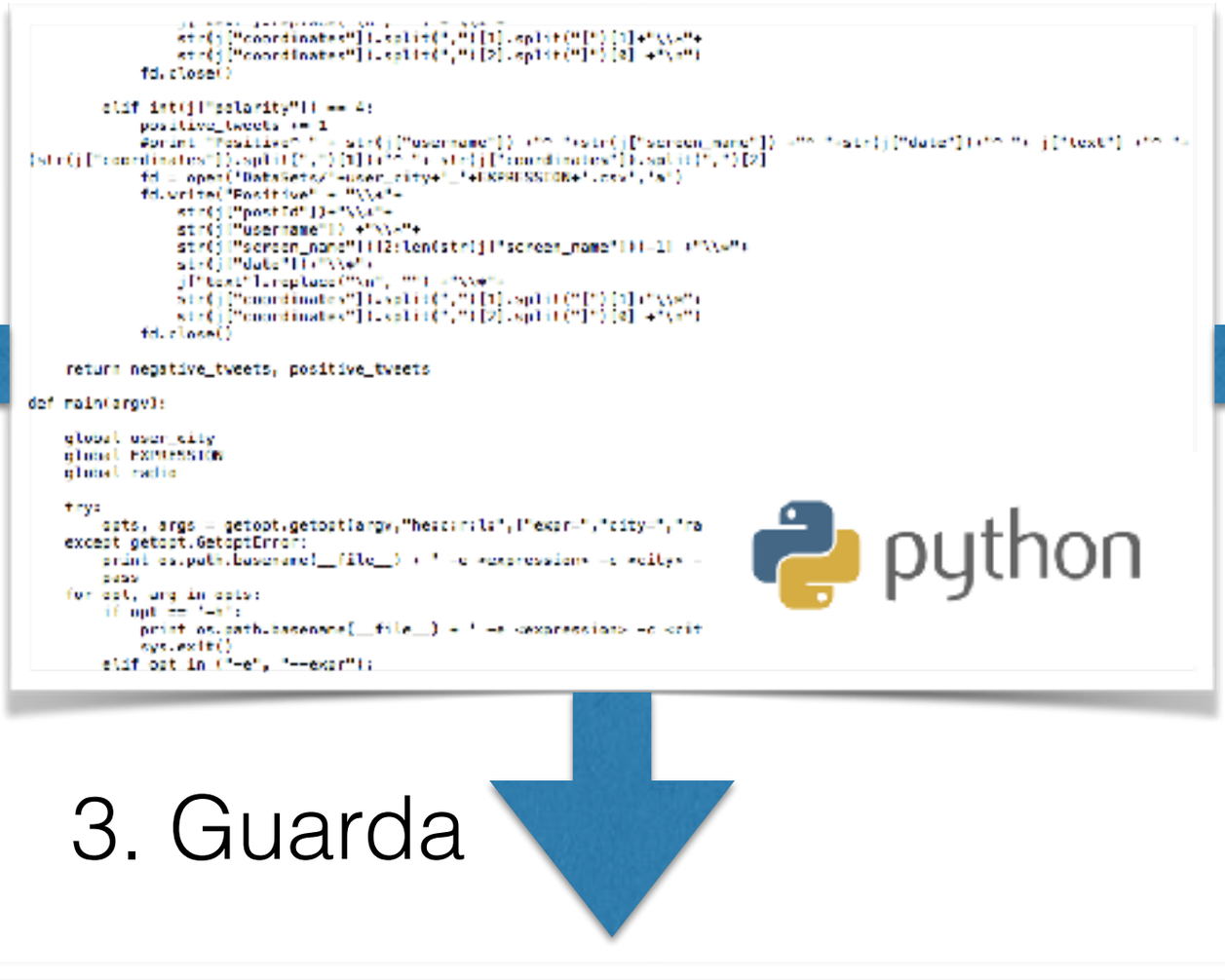
Durante un período de seis semanas (del 22 de septiembre al 3 de noviembre de 2015), se recolectaron 2.084 tweets de la ciudad de Nueva York, de los cuales 1.633 tenían sentimientos positivos y 451 sentimientos negativos. Los tweets con sentimientos neutros no se utilizaron.

Aprendiendo de los Tweets

30 palabras:

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. Recoger



2. Clasifica

+ and -

3. Guarda

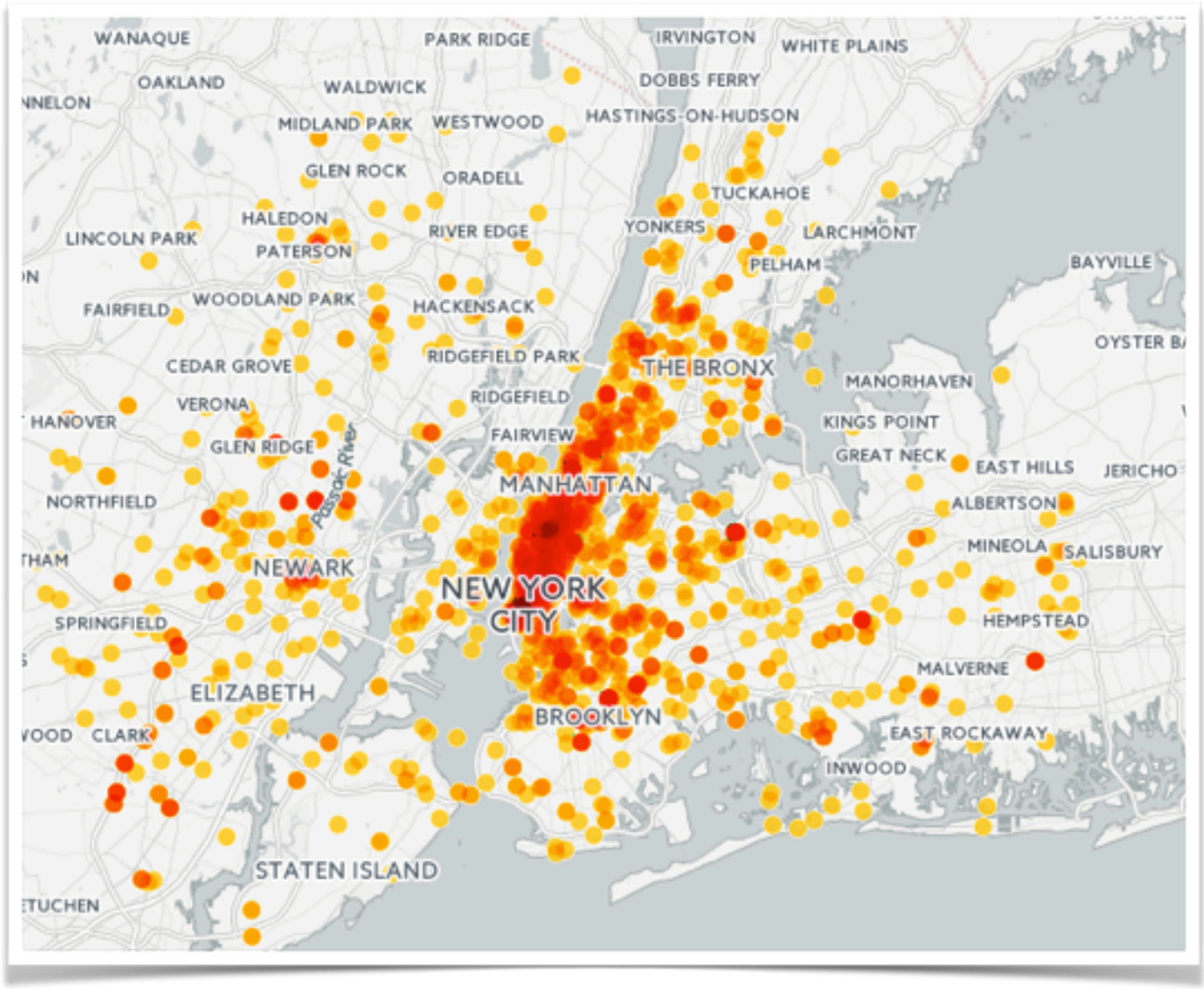
Tweet Positivo acerca de Comida Vegetariana

- 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)

Tweet Negativo acerca de Familia

- 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)

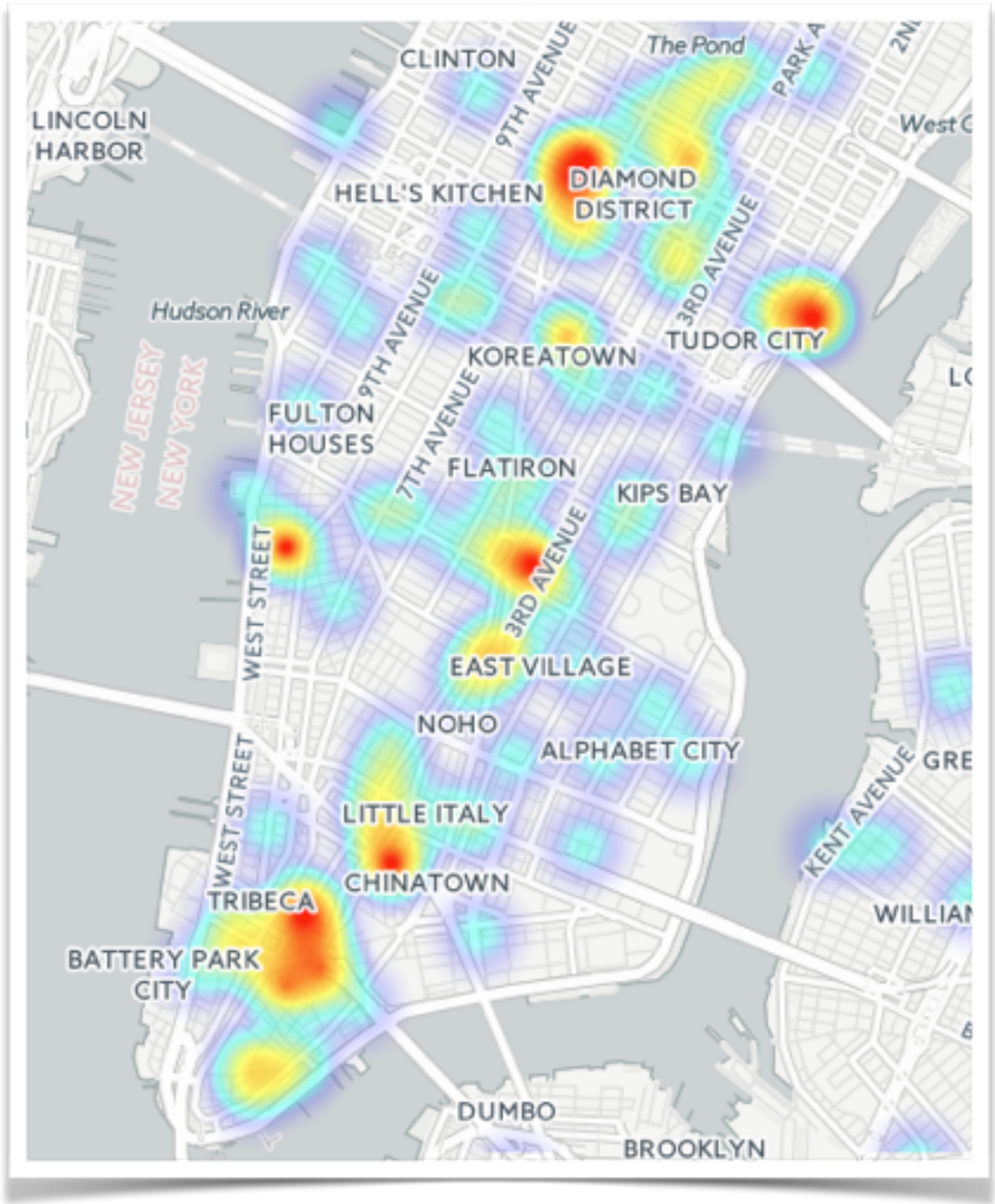
Aprendiendo de los Tweets



Intensidad de los tweets en la ciudad de New York

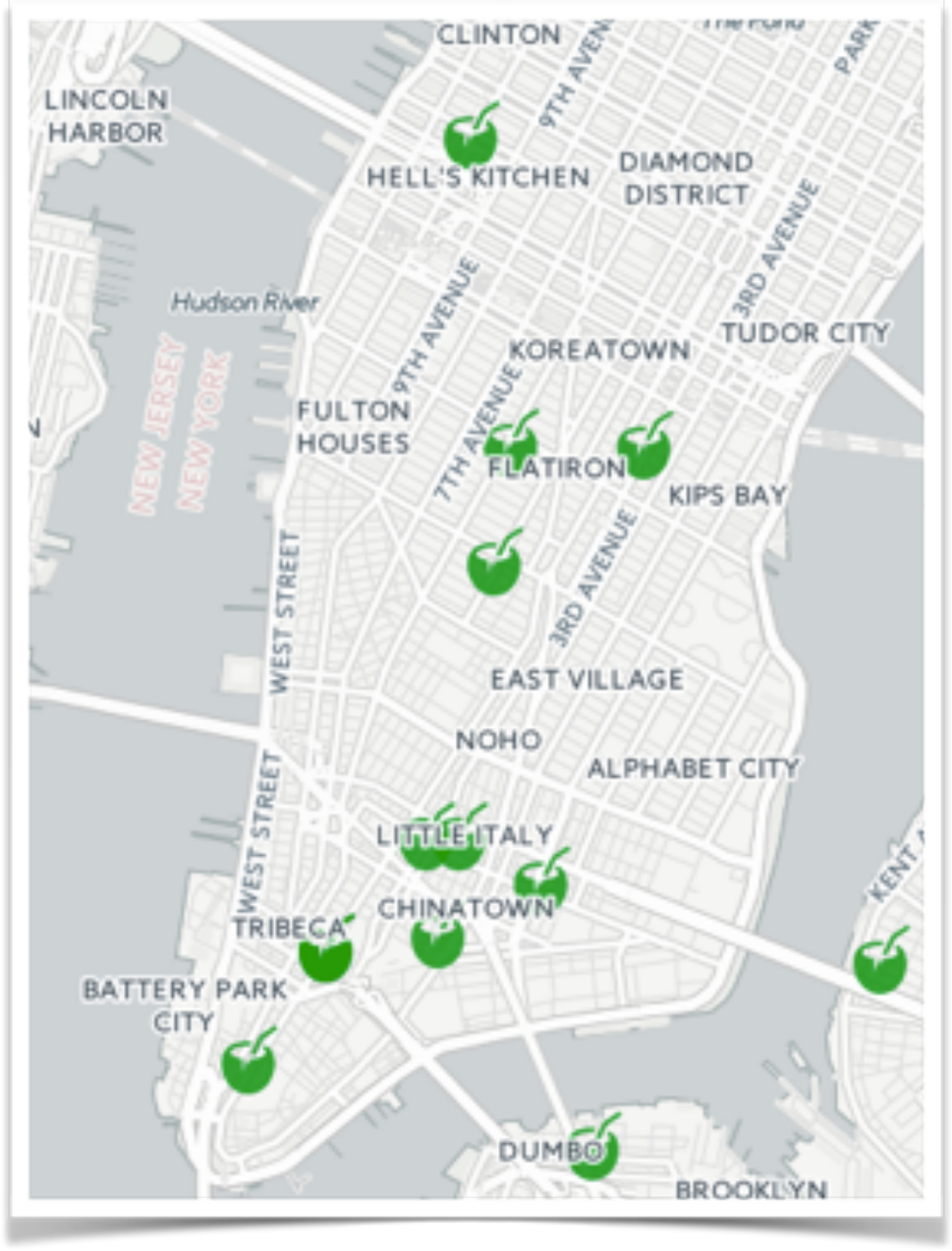
CARTO

Aprendiendo de los Tweets



Áreas con tweets negativos en Manhattan

Aprendiendo de los Tweets



Tweets positivos acerca de comida vegetariana en Manhattan

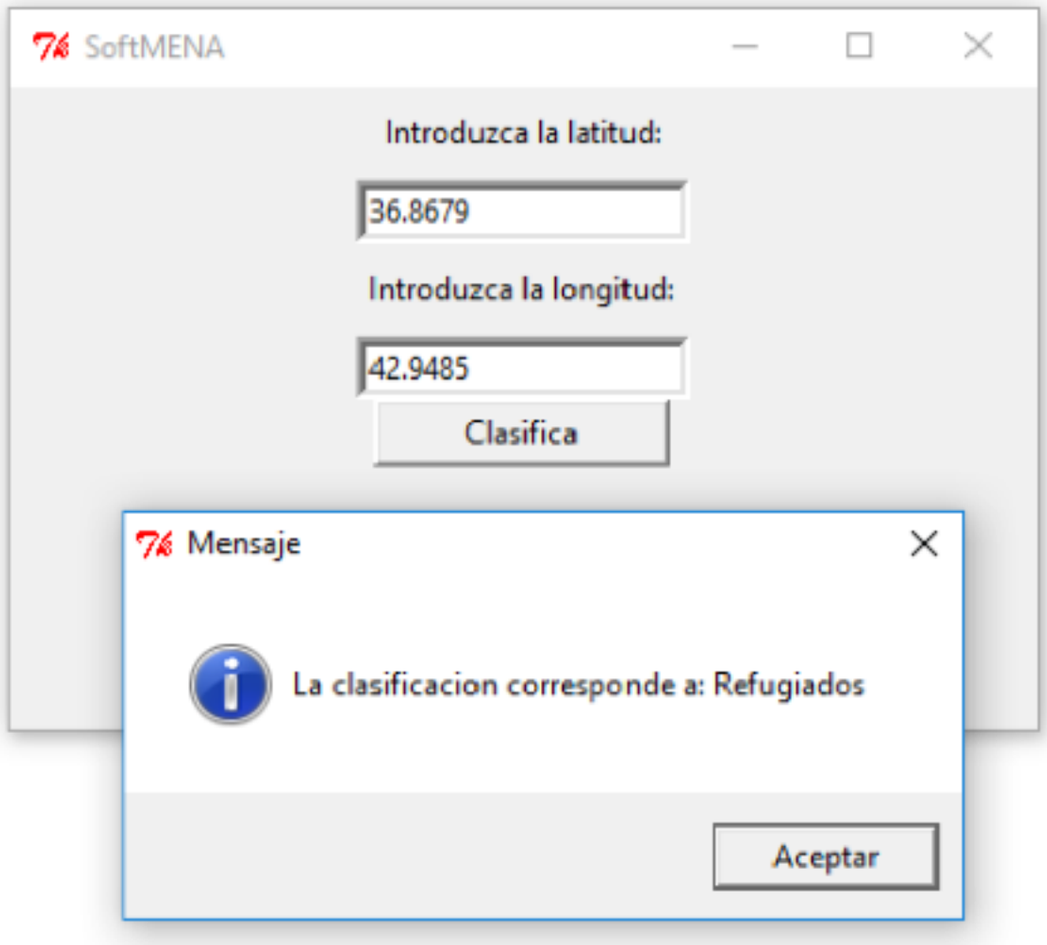
The GDELT Project

100 idiomas

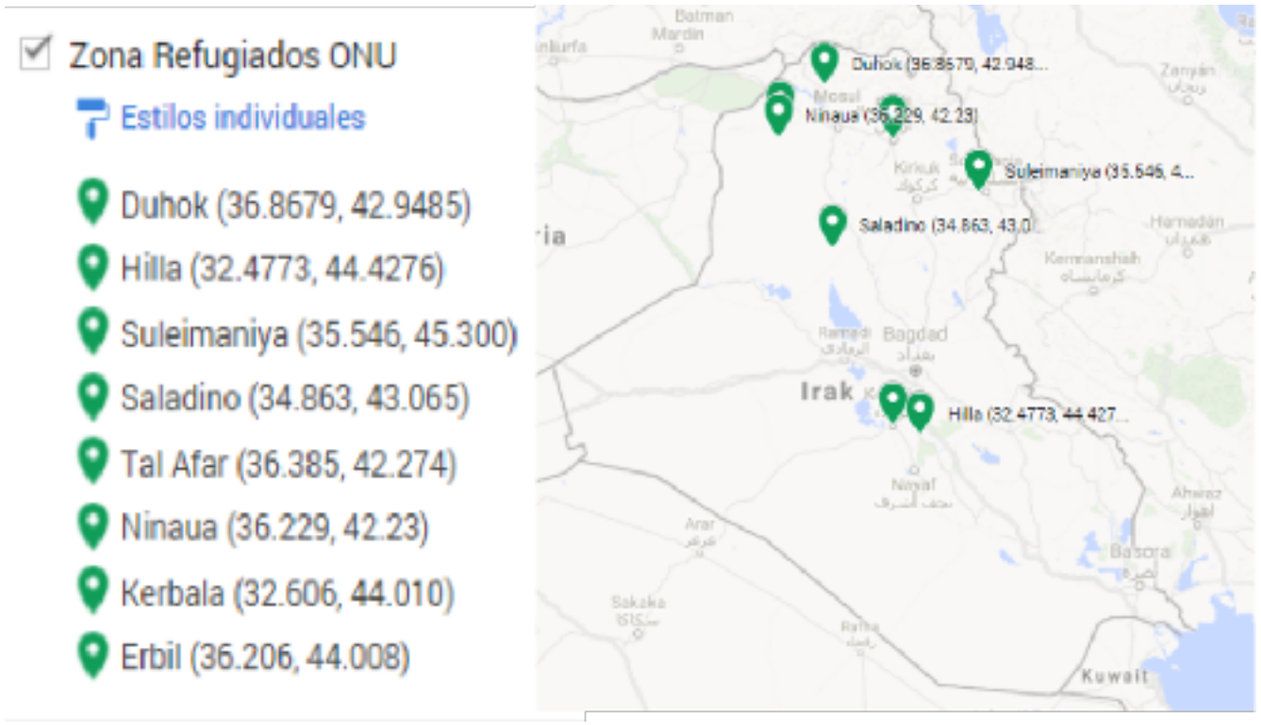
Software para Descubrir las Necesidades de las Personas en la Ventana 10/40 Usando Ciencia de Datos. Caso de Estudio: Middle East and North Africa Union

AfricaNews, Agence France Presse, Associated Press, Associated Press Online, Associated Press Worldstream, BBC Monitoring, Christian Science Monitor, Facts on File, Foreign Broadcast Information Service, The New York Times, United Press International and The Washington Post

Software para Descubrir las Necesidades de las Personas en la Ventana 10/40 Usando Ciencia de Datos



Clasificación Duhok, Iraq



Zonas de refugiados con latitud y longitud (adaptado de UNHigh Commissioner for Refugees, 2017)

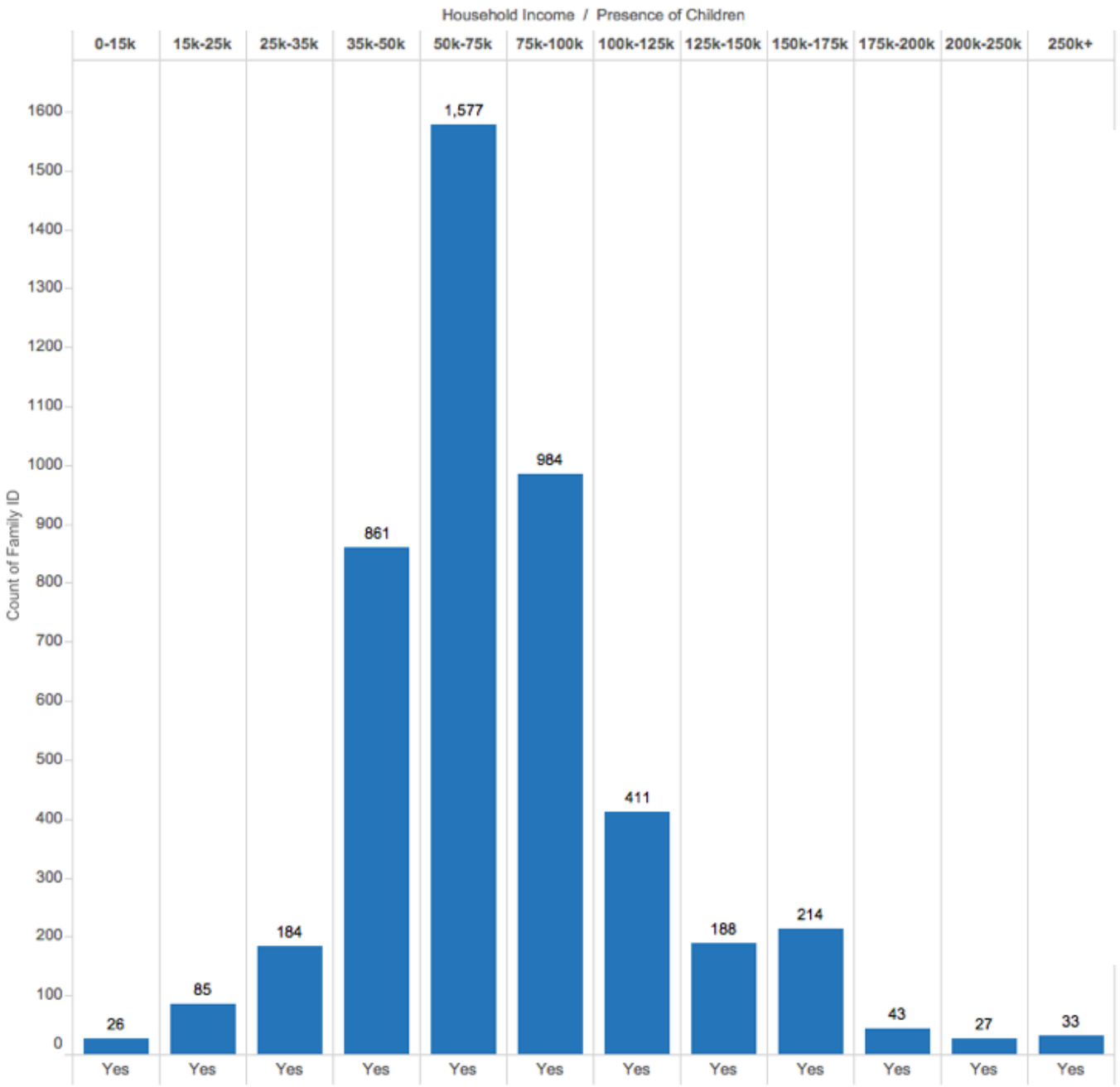
Google BigQuery

scikit learn

Agenda

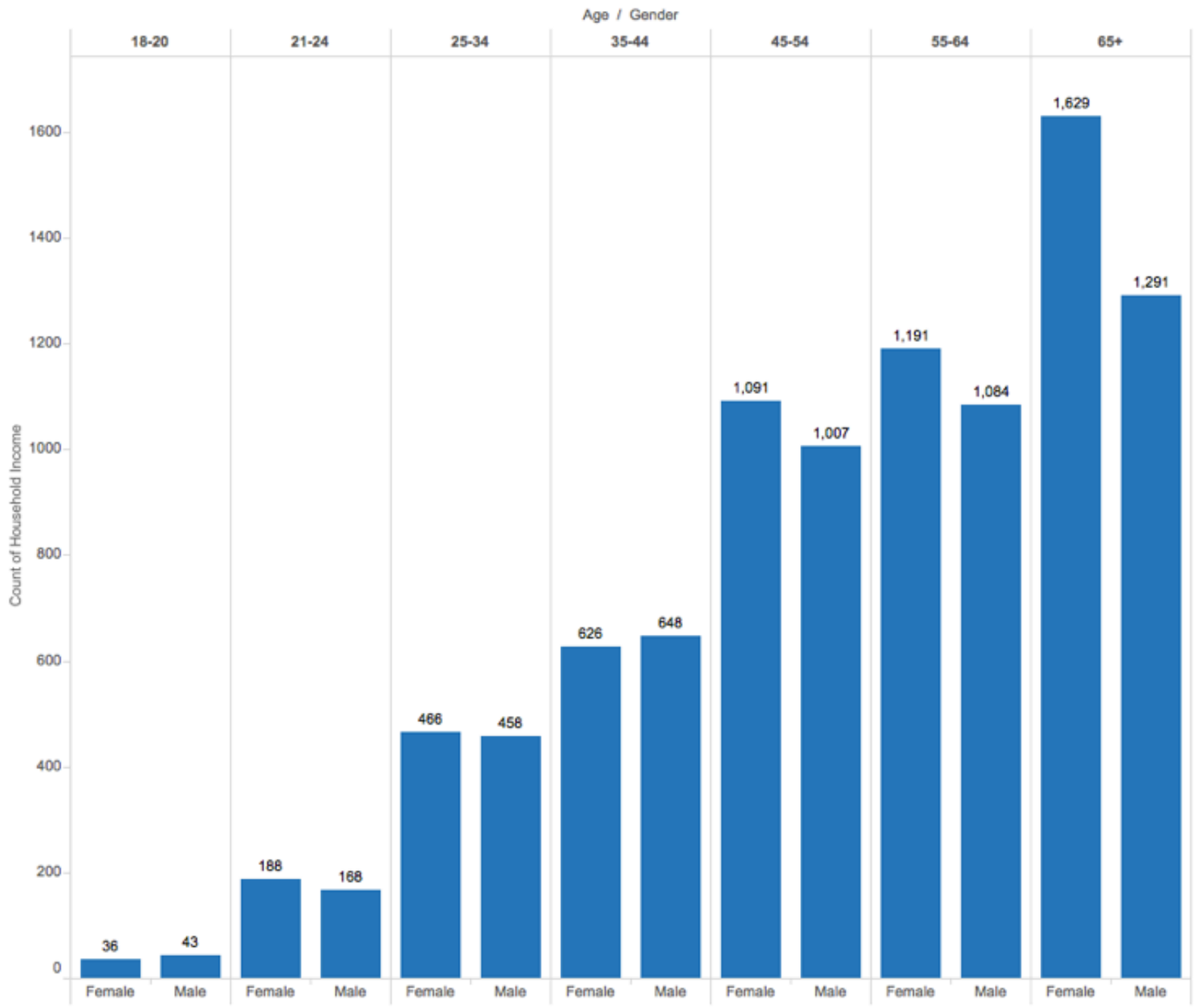
1. Periodismo de datos
2. Descubriendo las necesidades de las personas con ciencia de datos
3. **Comprendiendo a la feligresía con ciencia de datos**
4. Conclusiones

Análisis de Datos para Washington Conference



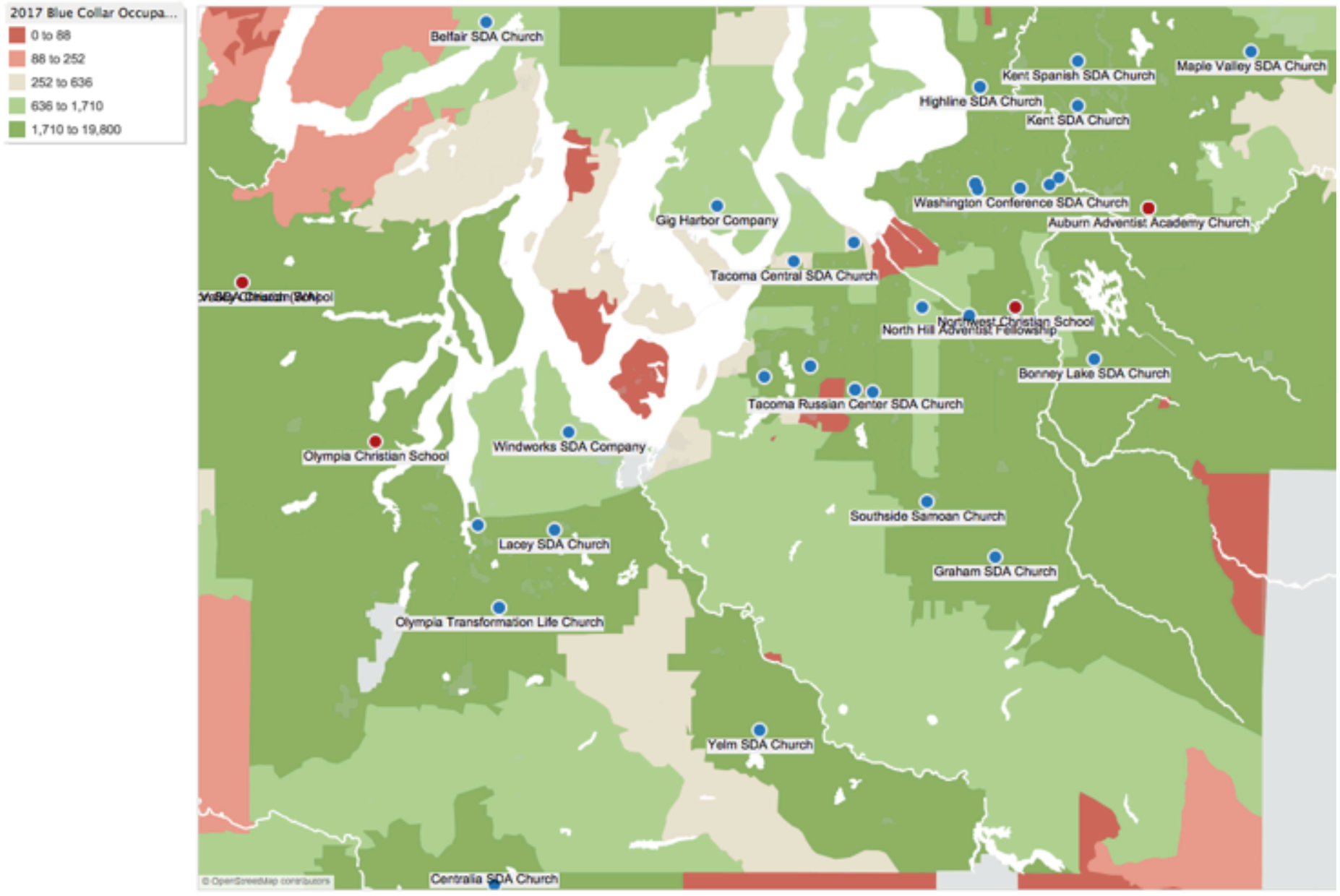
Ingresos familiares de familias con hijos en el hogar

Análisis de Datos para Washington Conference



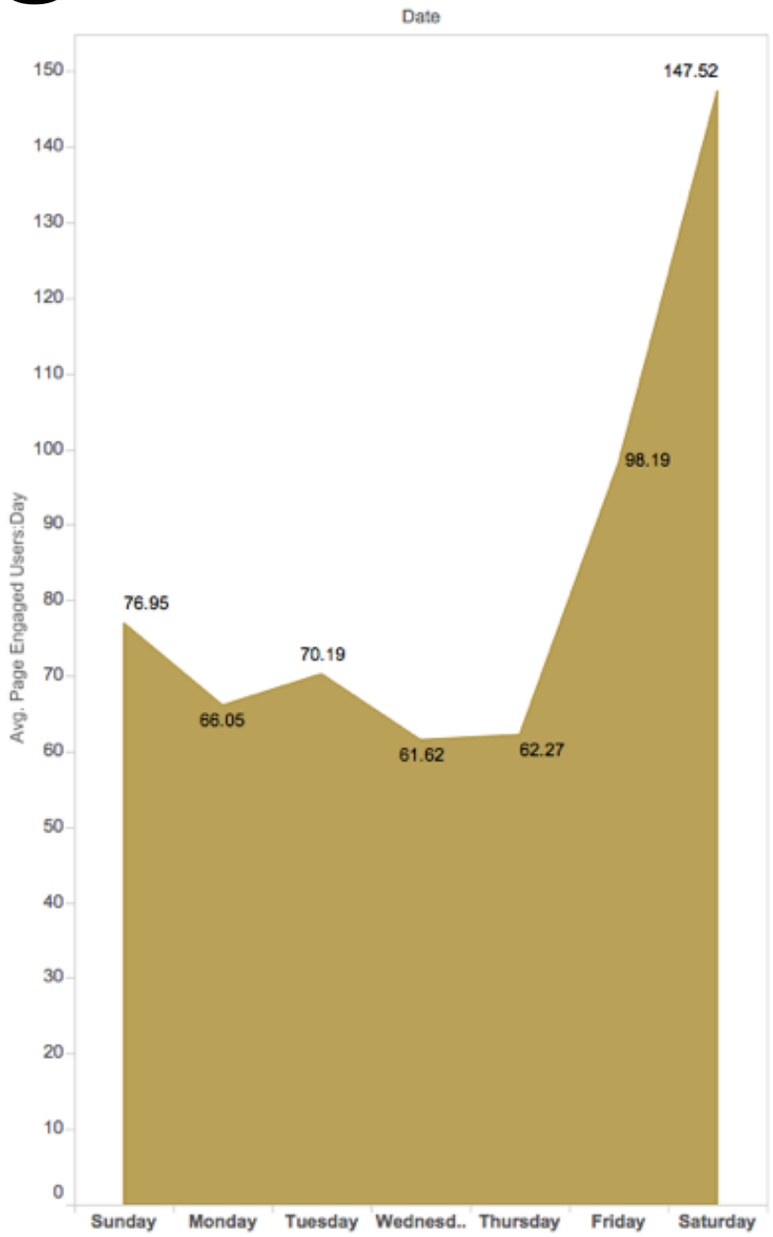
Ingresos por género y edad

Análisis de Datos para Washington Conference



Mapa de muestra con la geolocalización de iglesias (en azul) y escuelas (en rojo) y datos del censo relacionados con la ocupación manual (colores de fondo)

Análisis de Datos para Washington Conference



Usuarios activos en la página en Facebook de la conferencia (enero 12 a junio 6, 2017)

Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013

Dr. Germán H. Alférez, *Universidad de Montemorelos,* Erón Zebadúa, *Inter-Oceanic Mexican Union Conference,* and Enoc Cruz, *Universidad Linda Vista*

Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos

Abstract—Removing individuals from membership in the Seventh-day Adventist Church is the ultimate discipline that the church can administer. Our contribution is to present how we have applied state-of-the-art data science techniques to identify the segments of individuals who have been baptized from 2005 to 2013 and also been removed from membership in the same period of time at the Inter-Oceanic Mexican Union Conference. The dataset that was analyzed is composed of 14,388 records of members who have been removed. The results can guide further church decisions to prevent membership lost, specially among youth and among people who are baptized after evangelistic campaigns. Our data-science approach could be easily extrapolated to other divisions and conferences.

15. Alférez, G.H., Zebadúa, E., & Cruz, E. (2016). Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013. Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos. URL: http://www.harveyalferez.com/publications/TechnicalReport_June_23_2016_GSL_UM.pdf

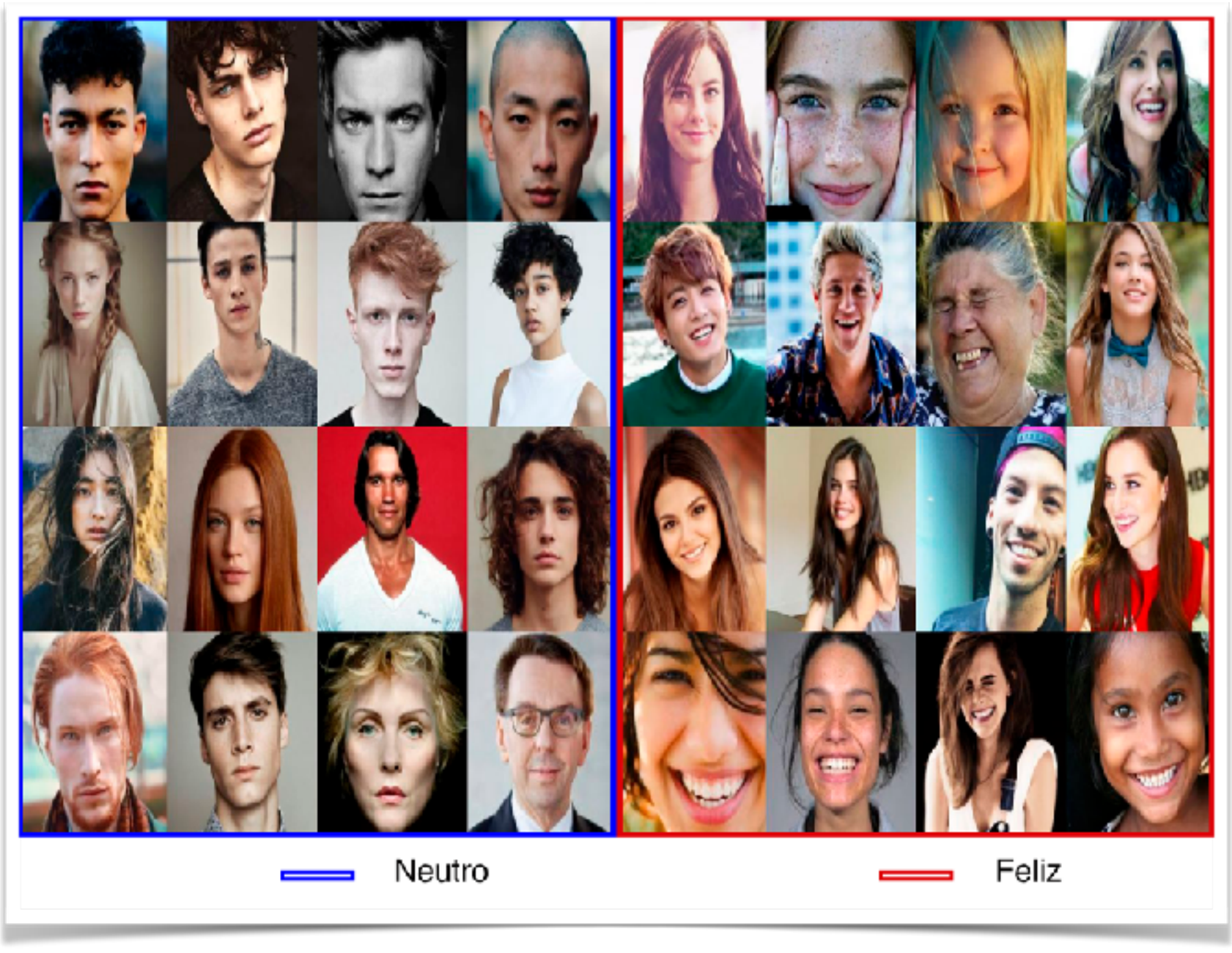
1. La mayoría de los individuos que apostatan son jóvenes y duran alrededor de 3 años en la iglesia.

2. El porcentaje de retención de jóvenes (15 a 33 años de edad) con antecedentes adventistas es muy similar al segmento de jóvenes sin ningún trasfondo religioso.

3. Las personas que ingresan a la iglesia después de una campaña evangelística tienden a abandonar la iglesia en un porcentaje más alto que las personas que toman cursos bíblicos o que han sido invitados por amigos.

15. Alférez, G.H., Zebadúa, E., & Cruz, E. (2016). Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013. Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos. URL: http://www.harveyalferez.com/publications/TechnicalReport_June_23_2016_GSL_UM.pdf

Aprendizaje Profundo para Descubrir las Emociones de la Audiencia al Ver Videos



Aprendizaje Profundo para Descubrir las Emociones de la Audiencia al Ver Videos

Agenda

1. Periodismo de datos
2. Descubriendo las necesidades de las personas con ciencia de datos
3. Comprendiendo a la feligresía con ciencia de datos
4. **Conclusiones**

La ciencia de datos tiene el potencial de ayudarnos a entender las necesidades y las tendencias de una forma sin precedentes.

Áreas de Interés

- **Evangelismo:** Entendiendo tu comunidad, etc.
- **Tesorería de la Iglesia**: Análisis de diezmos y ofrendas, perfiles de donantes, etc.
- **Educación:** Perfiles de estudiantes, predicción de retención, etc.
- **Medios de Comunicación:** Crear perfiles de la audiencia en TV/radio/Internet (i.e., personas), etc.
- **Análisis de Geolocalización:** Crear mapas relacionados con escuelas y estudiantes, iglesias y miembros, grupos pequeños, información demográfica, etc.

Sacándole el Jugo a los Datos en el Cumplimiento de la Misión

Harvey Alférez, Ph.D.
Global Software Lab,
Facultad de Ingeniería y Tecnología,
Universidad de Montemorelos, México
www.harveyalferez.com