



Práctica Final

Grupo 05

ALVARO LEONARDO SEIJAS
ROI SÁNCHEZ SERNA
ADRIAN ANDREA RUBIO ORTEGA
ALEJANDRO SOLANO RODRIGUEZ
JACOBO JAVIER GALACHE GOMEZ-COALLA
JAIME REDONDO LOPEZ



Introducción

1. Organización
2. Acotación del análisis
3. Primer análisis
4. Análisis de rating
 - a. Outliers
 - b. Reviews y rating
 - c. Controles realizados y rating
 - d. Rating según atributos
5. Predicción de nuevos negocios
6. Análisis de los usuarios: Influencers
7. Conclusiones



Organización

- Decidimos el alcance del análisis
 - En qué sección de los datos nos vamos a centrar
- Dividimos la tarea
 - Compilar las tablas de usuarios y reviews
 - Buscar relaciones entre datos
 - Representar los elementos de forma gráfica
 - Probar hipótesis de modelos predictivos
- Conclusiones
 - Juntamos el trabajo realizado
 - Compilamos los resultados



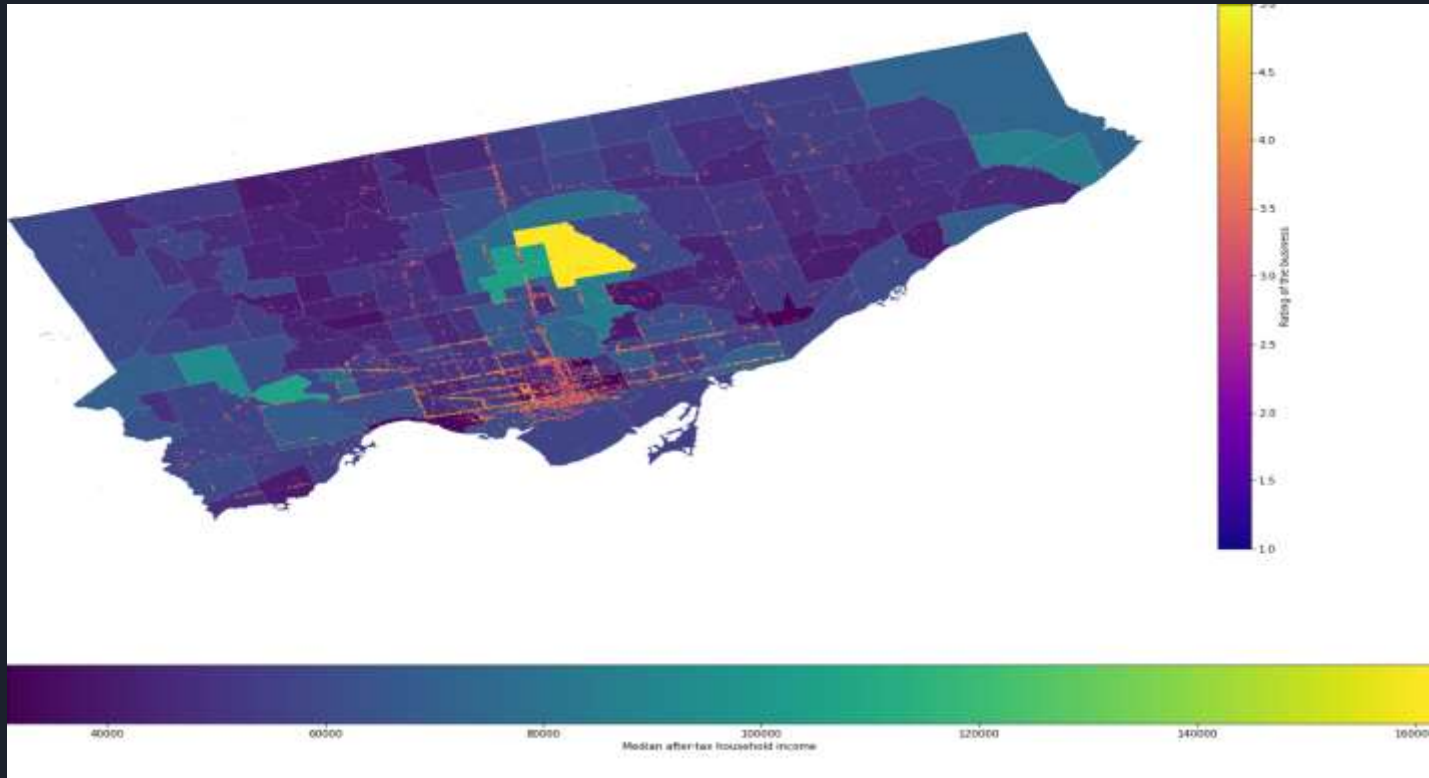
Acotación del análisis

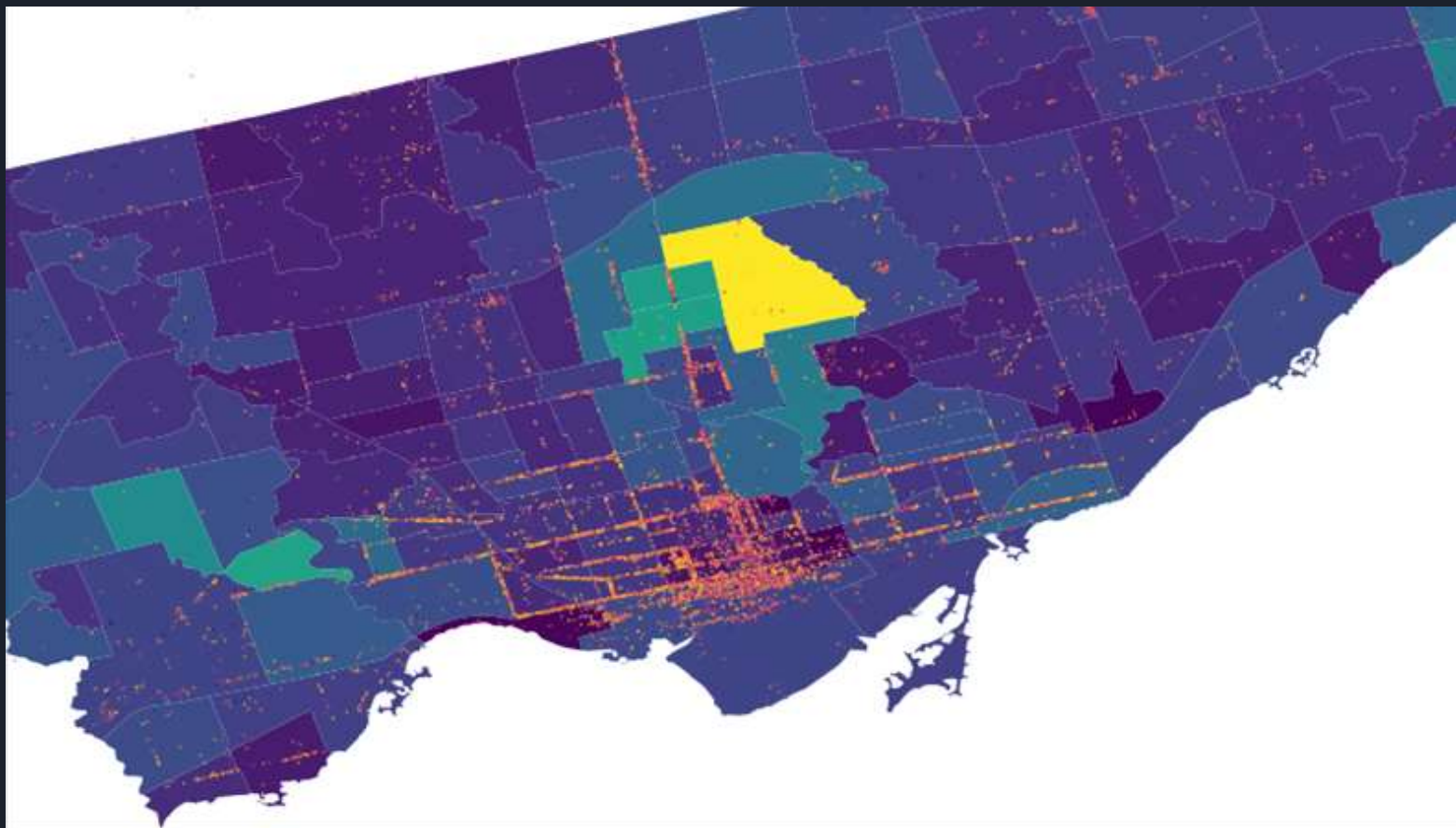
- Hemos decidido centrarnos en Toronto
- Nuestro objetivo principal es **analizar la puntuación de los negocios.**

Primer análisis: bussines (campos numéricos)

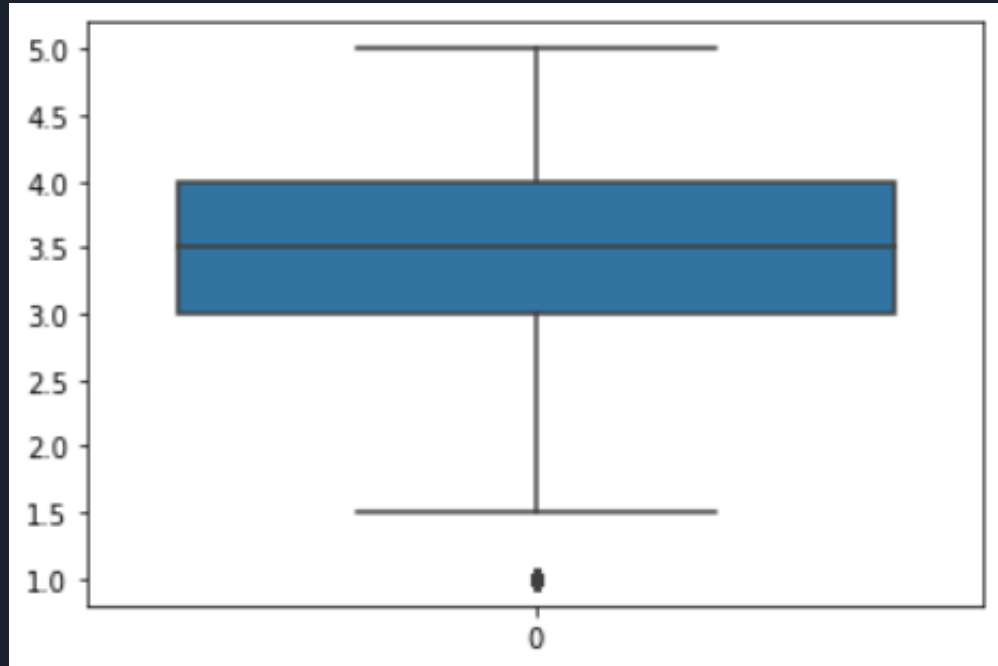
| | Unnamed: 0 | Unnamed: 0.1 | lat | long | rating | num_reviews | open |
|--------------|--------------|---------------|--------------|--------------|--------------|--------------|--------------|
| count | 18904.000000 | 18904.000000 | 18904.000000 | 18904.000000 | 18904.000000 | 18904.000000 | 18904.000000 |
| mean | 16732.029994 | 96958.203502 | 43.678689 | -79.395003 | 3.444430 | 27.022958 | 0.757988 |
| std | 9667.104717 | 55714.722559 | 0.043223 | 0.059808 | 0.934474 | 59.525537 | 0.428313 |
| min | 3.000000 | 19.000000 | 43.560365 | -79.713930 | 1.000000 | 3.000000 | 0.000000 |
| 25% | 8370.750000 | 48813.750000 | 43.650500 | -79.418793 | 3.000000 | 4.000000 | 1.000000 |
| 50% | 16686.500000 | 96804.000000 | 43.664216 | -79.394142 | 3.500000 | 9.000000 | 1.000000 |
| 75% | 25144.250000 | 145526.750000 | 43.688899 | -79.375570 | 4.000000 | 24.000000 | 1.000000 |
| max | 33409.000000 | 192592.000000 | 43.881942 | -79.019777 | 5.000000 | 2121.000000 | 1.000000 |

Análisis visual: mapa de Toronto





Análisis de la puntuación de los negocios





Outliers

- Los outliers están formados por 71 negocios, de los cuales su puntuación es 1
- Entre ellos destacan 3 McDonald's (de 47 que hay en total)
- Las categorías más repetidas son:

Movers, Home Services

3

Car Rental, Hotels & Travel

2

Banks & Credit Unions, Financial Services

2

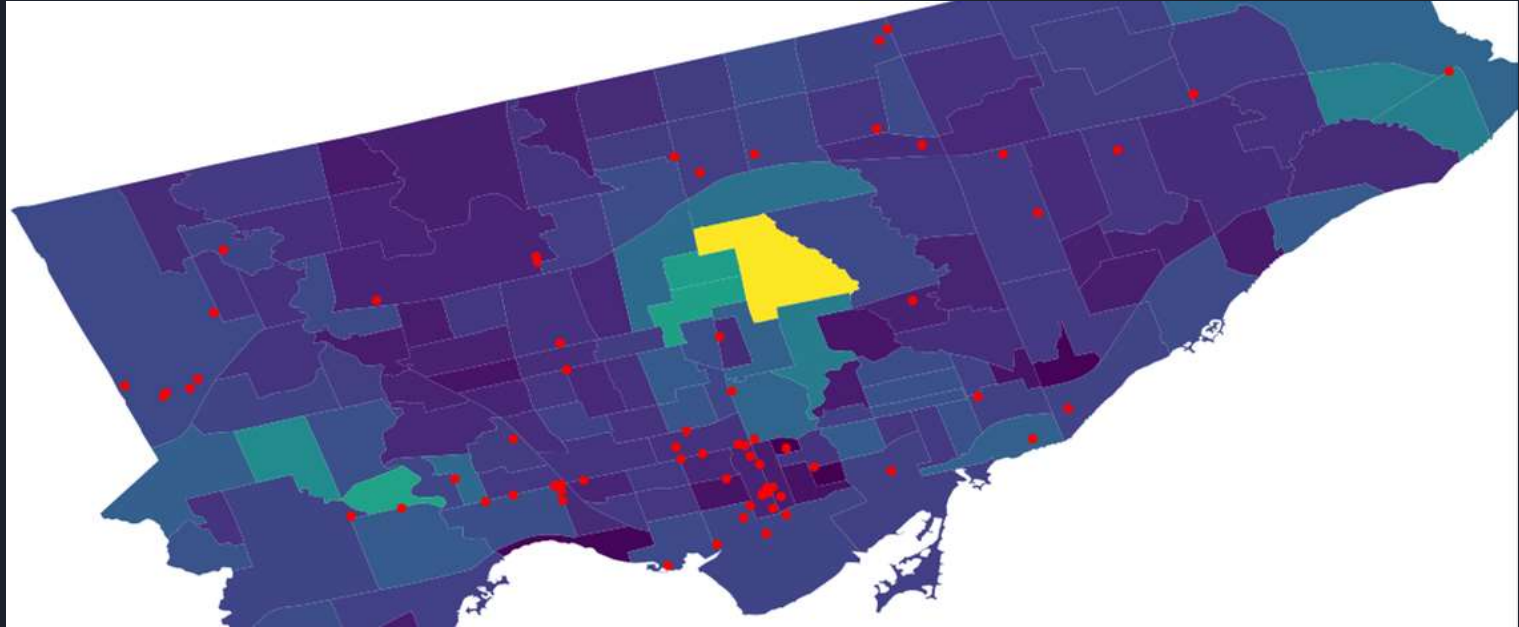
Apartments, Real Estate, Home Services

2

Event Planning & Services, Hotels, Hotels & Travel

2

Posición de los outliers en el mapa

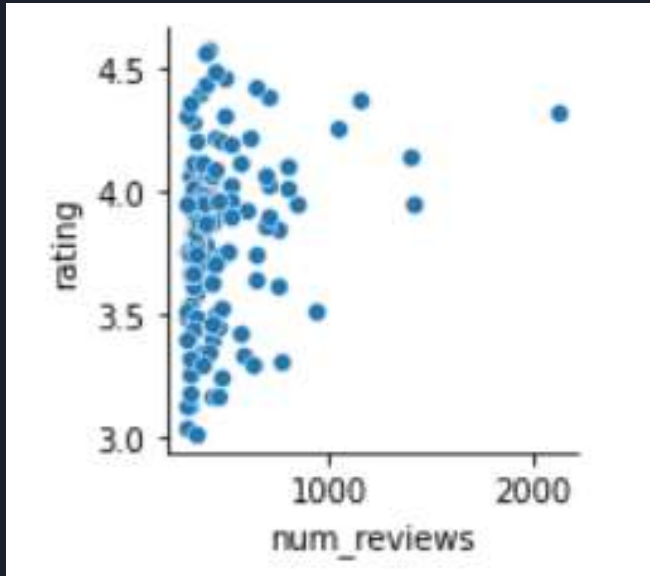




Número de reviews y rating

- ¿Cuántas más reviews tiene un negocio, mayor es su la puntuación de sus opiniones?
- Analizar el impacto de las reviews en las ventas

Número de reviews y rating



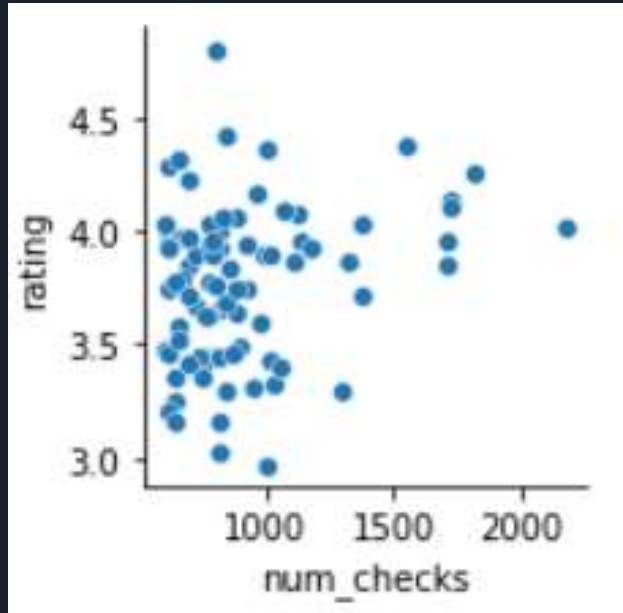
- Acumulación de datos incluso después de filtrar con mínimo número de reviews
- Se aprecia algo de correlación



Controles realizados y rating

- Relación entre el número de controles y puntuaciones del negocio
- Tiempo desde el primer hasta el último control

Controles realizados y rating



- Filtrado por tiempo abierto
- Negocios establecidos

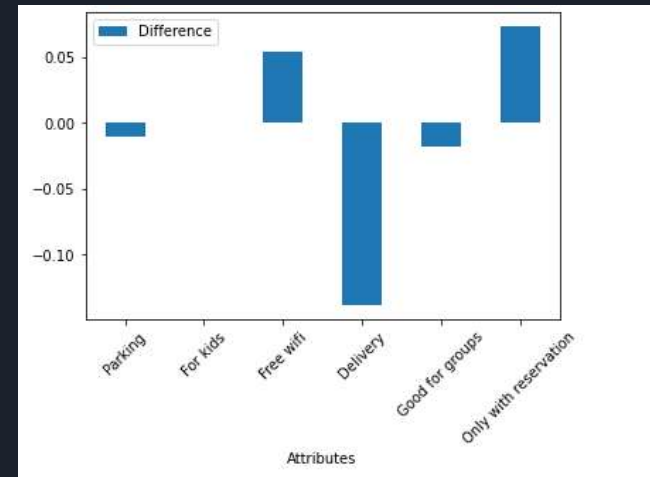
Rating sin filtrar = 3.441944

Rating filtrando = 3.773964

Rating según atributos

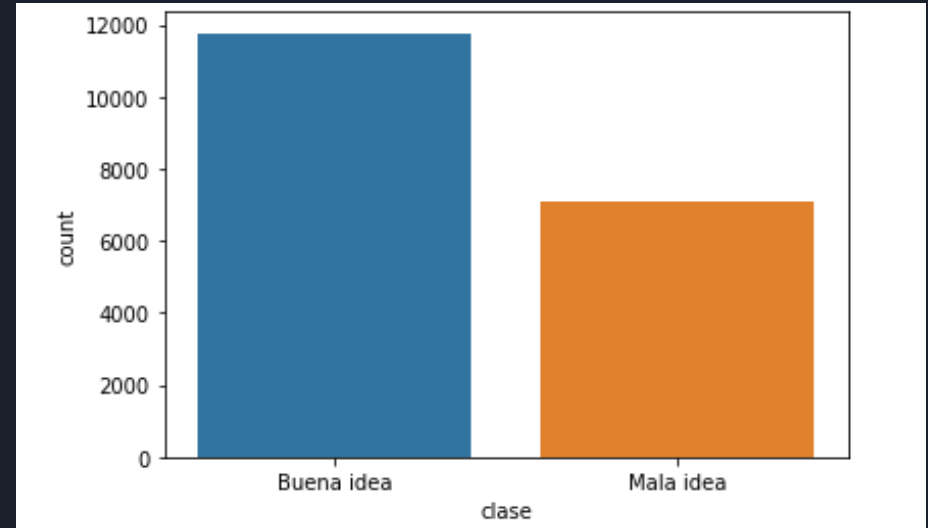
- Hemos seleccionado varios atributos para analizar cómo cambia el rating medio de los negocios según posean o no el atributo
- De esta forma un empresario puede modificar su negocio para aumentar la satisfacción de sus clientes.

| | Attribute | Yes | No | Difference |
|---|-----------------------|----------|----------|------------|
| 0 | Parking | 3.519880 | 3.530760 | -0.010880 |
| 1 | Good for kids | 3.504792 | 3.504792 | 0.000000 |
| 2 | Free wifi | 3.513061 | 3.458878 | 0.054183 |
| 3 | Delivery | 3.373378 | 3.512042 | -0.138663 |
| 4 | Good for Groups | 3.415595 | 3.433912 | -0.018317 |
| 5 | Only with reservation | 3.694185 | 3.620148 | 0.074037 |



Predicción para nuevos comercios en Toronto

- Técnica utilizada: TEXT MINING
- Datos de entrada: Categorías de los comercios existentes en Toronto
- Clase:
 - Buena idea: $\text{rating} \geq 3.5$
 - Mala idea: $\text{rating} < 3.5$





Resultados obtenidos

| ALGORITMO | ACCURACY | MATRIZ DE CONFUSIÓN |
|---------------------|----------|-------------------------|
| Regresión Logística | 0.672053 | [2036 329] [909 501] |
| Multinomial NB | 0.657219 | [1985 380] [914 496] |
| Passive-Aggressive | 0.615629 | [1698 667] [784 626] |
| Árbol de decisión | 0.627550 | [1783 582] [824 586] |
| RandomForest | 0.649007 | [1902 463] [862 548] |



Análisis de los resultados

PRINCIPALES PROBLEMAS

- Escaso volumen de datos para la ciudad de Toronto
- El Text Mining necesita un gran volumen de datos
- Clases no balanceadas

Análisis de los usuarios: Influencers





Influencers: Proceso

- Disponemos de un csv con ids de negocios y su localización
- Filtramos y obtenemos los negocios de Toronto
- Obtenemos las reviews correspondientes a los negocios
- Obtenemos los usuarios que realizaron dichas reviews



Influencers: Resultado

- Total de **109060** usuarios que han realizado una review en Toronto
- Top 500 con más seguidores
- Top 500 con más reviews
- Obtenemos la intersección de los dos conjuntos
- Resultado: 43 usuarios con más de 800 reviews y 500 seguidores cada uno

Influencers: Resultado

| user_id | name | num_reviews | user_since | useful | fun | cool | expert |
|------------------------|----------|-------------|------------------|--------|-------|-------|--|
| Hi10sGSZNxQH3NLyWSZ1oA | Fox | 6407 | 26/05/2009 11:33 | 89418 | 86122 | 82128 | 2014,2015,2016,2017,2018 |
| CxDOIDnH8gp9KXzpBHJYXw | Jennifer | 6314 | 09/11/2009 20:44 | 20336 | 8823 | 12995 | 2009,2010,2011,2012,2013,2014,2015,2016,2017,2018 |
| bLbSNkLggFnqwNNzzq-ljw | Stefany | 4321 | 29/06/2011 14:40 | 44667 | 22790 | 30846 | 2012,2013,2014,2015,2016,2017,2018 |
| U4INQZOPSUaj8hMjLIZ3KA | Michael | 4308 | 31/01/2008 2:55 | 22757 | 10209 | 16249 | 2008,2010,2011,2012,2013,2014,2015 |
| CQUdH80m48xnxUkx-X5NAw | David | 4027 | 29/12/2008 21:03 | 20314 | 8669 | 14142 | 2009,2010,2011,2012,2013,2014 |
| GHoG4X4FY8D8L563zzPX5w | Hannah | 3956 | 22/08/2012 10:40 | 28051 | 15649 | 21930 | 2013,2014,2015,2016,2017,2018 |
| WWnhqRnWWjutMqh-2SzEuQ | Steph | 3379 | 09/07/2008 6:30 | 21901 | 15788 | 17440 | 2009,2010,2011,2012,2013,2014,2015,2016,2017,2018 |
| gjhzKWsqClrpEd9pevbKZw | Miriam | 3148 | 21/09/2005 16:54 | 19224 | 8003 | 15438 | 2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018 |
| ITa3vh5ERI90G_WP4SmGUQ | Peter | 2939 | 06/05/2008 1:30 | 41172 | 45625 | 31778 | 2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018 |
| UsXqCXRZwSCSw0AT7y1uBg | Candice | 2816 | 25/01/2008 19:32 | 47055 | 39914 | 38957 | 2009,2010,2011,2012,2013,2014,2015,2016,2017,2018 |
| 5MCBLBxr10NLUKZ4AboAMg | Colleen | 2770 | 07/09/2006 19:27 | 33675 | 21647 | 24589 | 2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018 |
| w-w-k-QXosIKQ8HQVwU6IQ | Anthony | 2741 | 13/03/2009 21:18 | 36969 | 25191 | 31736 | 2009,2010,2011,2012,2013,2014,2015,2016,2017 |
| QnlrewrXxFnssw_gG1gF-Q | Sarah | 2638 | 10/07/2010 18:54 | 37788 | 20241 | 32330 | 2010,2011,2012,2013,2014,2015,2016,2017,2018 |



Influencers

- ¿Cómo pueden usarse para promocionar un negocio?
- ¿Qué alcance tienen?
- ¿Qué resultados se pueden esperar?



Conclusiones finales

- El posicionamiento es clave para su visibilidad y afecta en todos los aspectos
 - Mejor en zonas más pobladas, calles más concurridas
 - Ciertas calles son mejores que otras
- Las reviews son la clave
 - Conseguir que los usuarios realicen reviews sobre tu negocio es clave para mejorar tu standing
 - El número de reviews es un predictor de mayor rating para el negocio
- La longevidad es un marcador de calidad
- El tipo de negocio determina la viabilidad



Final

¿Alguna pregunta?