

BIG DATA & ANALYTICS

CASO CLARIN

_Hacia una empresa data driven

MAYO 2021



¿POR QUÉ SER UNA EMPRESA
DATA DRIVEN?

¿Qué es ser data driven?



WIKIPEDIA
The Free Encyclopedia

- Tomar decisiones en base a datos.
- Estructura analítica potente.
- Tener herramientas innovadoras de análisis.
- Rápida y óptima toma de decisiones, tanto a nivel departamental como general.
- Tecnología punta.
- Una **empresa Data Driven** cuenta con herramientas que tratan la información en **tiempo real**.

Hacia dónde vamos:



REAL TIME

Información en realtime sobre tráfico, muro paywall y altas.



PREDICIONES EN TIEMPO REAL

Predicción sobre el comportamiento futuro de las notas, según su performance.



RECOMENDADOR DE CONTENIDO

Sistema personalizado de recomendación, según perfil del lector, al pie de la nota.



COMPETENCIA

Monitoreo de notas de la competencia.



SEO

Semáforo personalizado con principales variables SEO.



REPORTES

Repositorio de información para Ipad, Desktop, con información y análisis de performance.



MODELOS

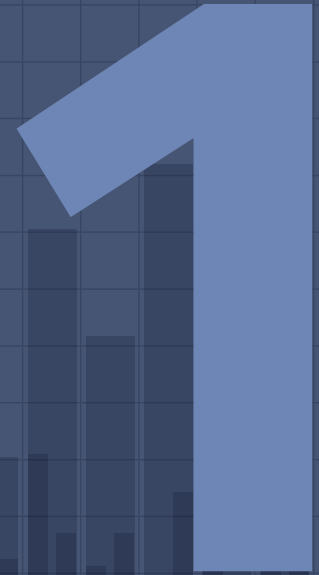
Machine Learning e Inteligencia Artificial para la generación de modelos predictivos, clusters, y nuevas herramientas.



PRICING

Optimización de precios en venta programática.

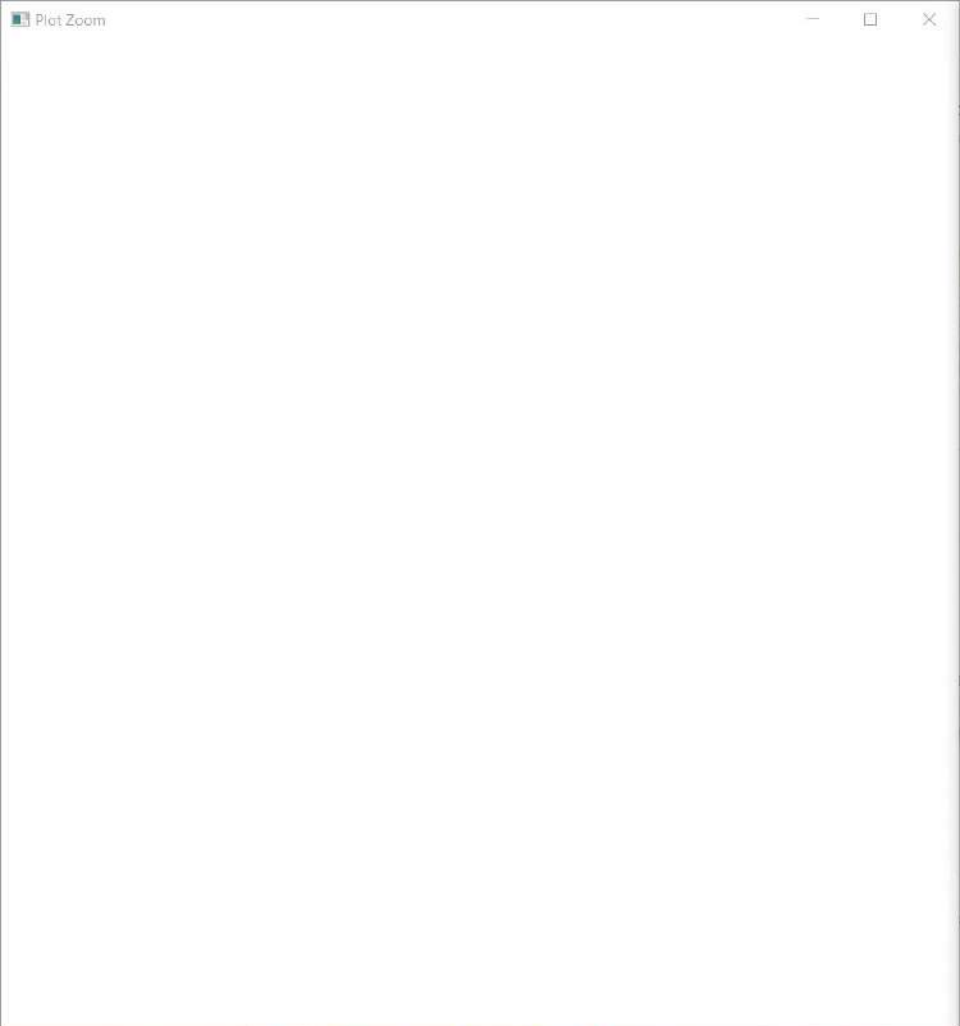
REAL TIME



¿Por qué Real Time?

- **Disponibilización** de la información inmediata.
 - **Personalización** de lo que queremos ver.
 - **Ahorro** en costos para mediano y largo plazo.
 - Aplicación de **modelos predictivos** en RT
 - **Recomendaciones** de contenido en RT.
-
- Prepara el terreno para lo que viene...

FUNCIONAMIENTO EN VIVO



Google

google.com/webhp?hl=es&sa=X&ved=0ahLKEw...

Aplicaciones IMPALE Color Wheel ViernesDulce Otros marcadores Lista de lectura

Gmail Imágenes

Google

Buscar con Google Me siento con suerte

Ofrecido por Google en: English Español (Latinoamérica)

Argentina

Sobre Google Publicidad Negocios Cómo funciona la Búsqueda

Privacidad Condiciones Preferencias

PRINCIPALES HITOS

1. REPORTING

- Tablero performance general.
- Tablero de altas.

2. MODELOS PREDICTIVOS

- Predicción performance de la nota.
- Predicción de Choques.
- Análisis de comportamiento.

3. PERSONALIZACIÓN

- Editable según las preferencias del negocio.

CARACTERISTICAS

- **Escalable**, soporta un incremento de más del *500%*.
- **Personalizable**, en la creación de métricas.
- **Plug & Play**, puede ser implementado en cualquier sitio.
- **Flexible**, Cloud Ready.
- **Confiable**, permite guardar datos históricos.

Recomendador de contenido



2

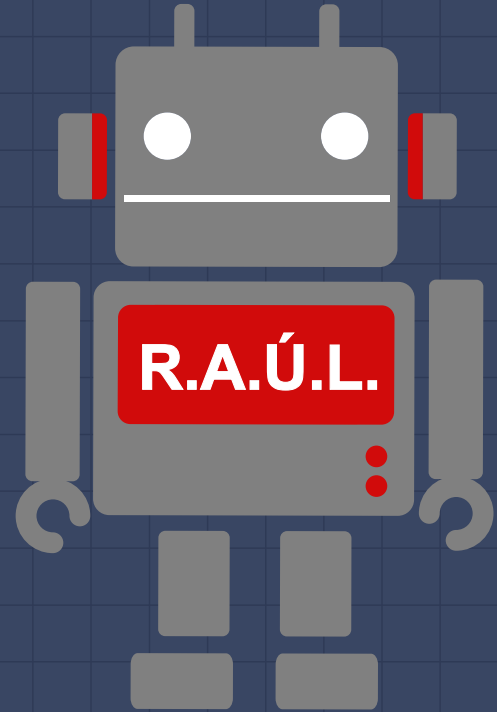
R.A.U.L.

Recomendador

Automático

Últimas

Lecturas



Machine Learning + Inteligencia Artificial

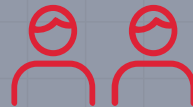
¿Qué recomienda?



Último
artículo



Historia de
navegación



Usuario
similar



Nota más
leída

NAVEGACIÓN

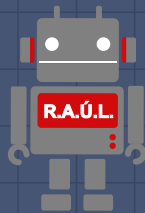
FILTRO COLABORATIVO

Performance del recomendador



- **Recomendación de contenido propio + contenido publicitario.**
- **Sin posibilidad de generar reglas de recomendación.**
- **Pocas métricas de performance.**

Vs



- **Recomendación de contenido 100% propio.**
- **Recomendación personalizada para cada usuario.**
- **Gran apertura de datos para monitoreo.**

Performance del recomendador

TABOOLA vs R.A.U.L

0,81%

CTR

Taboola

VS

2,13%

CTR



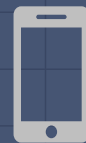
* Taboola: Información widget recomendación Septiembre 2020.

* Raul: Información recomendación en Suscriptores Abril 2021.

CTR por plataforma



2,16%
CTR DESKTOP



2,07%
CTR MOBILE



1,98%
CTR HOMBRES



2,50%
CTR MUJERES

CTR por rango etario

Navegación Desktop



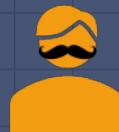
1,42%

CTR MILLENIALS



1,68%

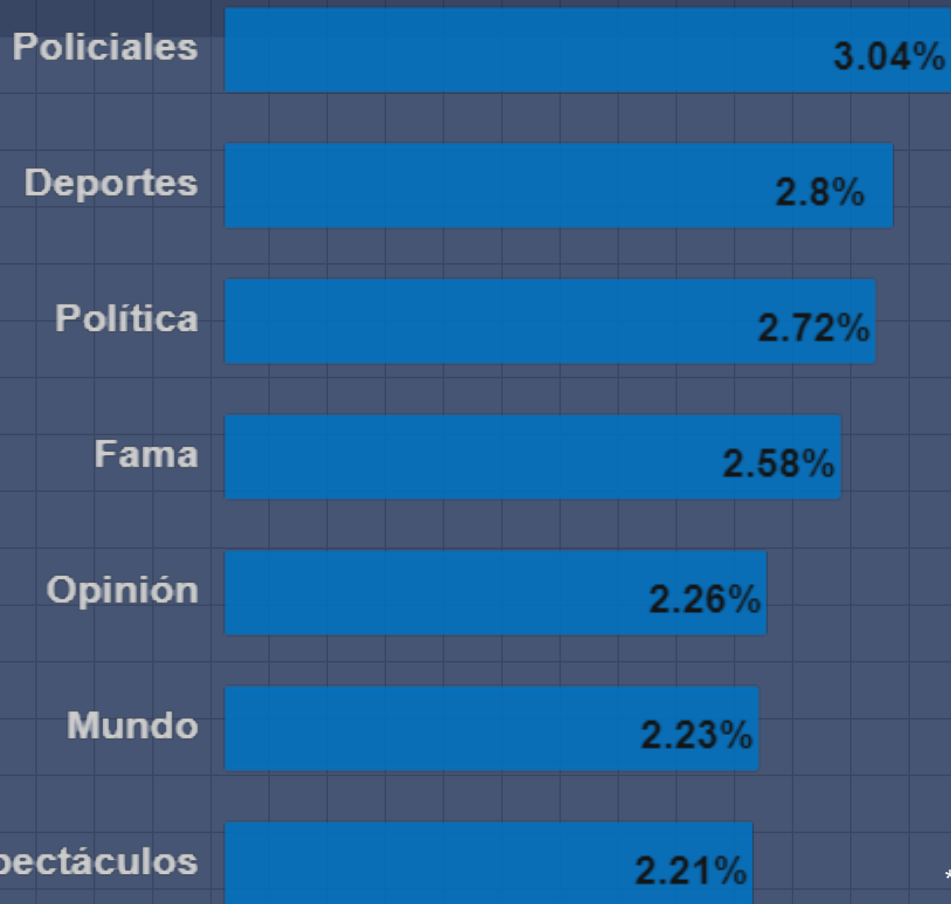
CTR GENERACIÓN X



2,34%

CTR BOOMERS

* Raul: Información recomendación en Suscriptores Abril 2021.



Top secciones con mayor CTR

Navegación Desktop

* Raul: Información recomendación en Suscriptores Abril 2021.

Competencia




3

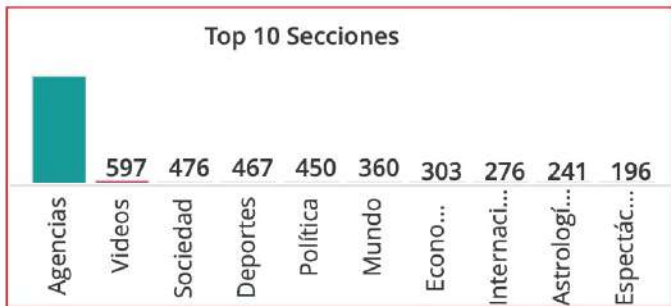
¿QUÉ ES UN SCRAPER?

INTRO:

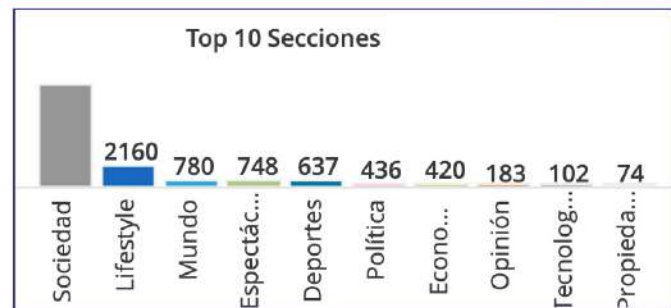
Es un **bot** que cada una hora **baja información** de los sitios editoriales elegidos.

- Cantidad de notas publicadas.
 - Contenido.
 - Horario de publicación.
- 

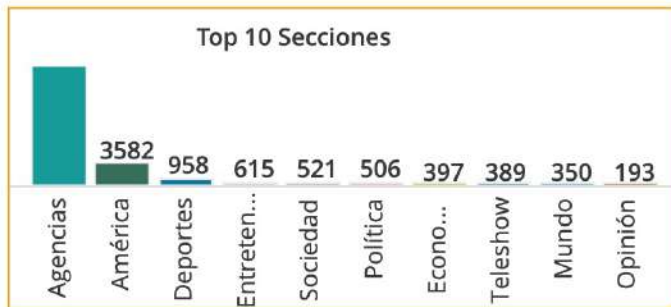
A



B



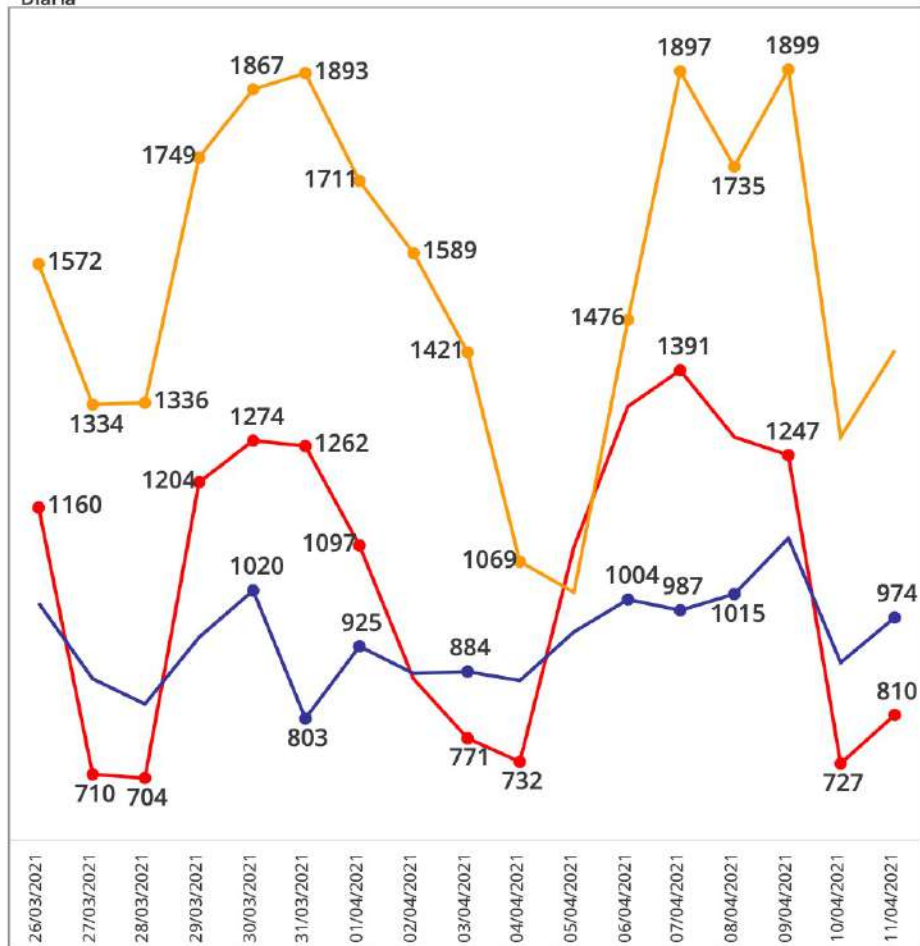
C



Evolución Diaria

Desde: 26/03/2021

Hasta: 11/04/2021



A

Indica	Cant Notas																							
	Dia	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
do.	38	51	49	50	35	94	109	79	96	126	197	156	171	172	184	157	165	179	196	164	127	131	130	94
lu.	83	73	73	50	75	90	200	208	243	286	283	324	330	293	321	233	239	210	233	197	177	126	143	105
ma.	76	67	64	53	88	120	210	214	241	319	329	372	345	358	357	269	284	261	307	197	206	154	185	130
mi.	75	86	60	74	92	130	246	226	226	323	353	335	375	332	327	355	271	257	273	245	191	160	200	113
ju.	123	108	65	84	97	125	270	234	305	307	369	459	386	421	436	366	394	328	346	293	289	216	231	148
vi.	122	69	75	83	123	121	259	270	314	303	320	424	348	373	345	335	296	259	293	246	190	132	162	89
sá.	81	62	48	52	48	102	115	92	104	138	153	160	147	189	182	197	160	157	175	140	137	96	129	54

B

Indica	Cant Notas																							
	Dia	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
do.	136	69	188	295	668	1064	269	105	20	32	31	39	47	29	34	43	28	48	41	42	28	20	15	91
lu.	70	34	142	288	503	804	206	146	33	58	68	84	79	66	64	86	85	113	90	70	59	43	14	106
ma.	57	43	151	265	565	896	235	124	29	63	68	99	119	81	103	115	115	159	95	89	65	56	35	101
mi.	60	43	144	183	383	629	163	100	26	113	625	119	83	87	79	77	75	106	80	75	53	31	17	65
ju.	71	44	246	340	724	1128	309	142	50	122	121	101	124	124	117	150	165	141	138	155	85	60	29	147
vi.	85	80	176	336	712	1149	306	148	37	75	120	122	98	74	195	122	144	132	121	118	76	59	38	142
sá.	181	45	112	257	677	1067	277	137	10	44	35	34	63	47	111	48	46	57	64	40	23	11	16	110

C

Indica	Cant Notas																							
	Dia	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
do.	356	162	418	117	91	96	326	440	150	307	187	197	293	240	203	244	200	339	251	191	232	42	59	218
lu.	203	214	353	100	141	142	287	328	222	344	341	264	262	299	338	270	268	315	348	234	224	43	87	155
ma.	188	114	334	131	147	113	182	347	215	428	328	322	366	419	333	401	316	453	383	361	280	72	81	275
mi.	266	232	485	225	219	183	330	513	352	438	400	310	392	465	346	435	353	378	335	363	304	50	69	232
ju.	325	213	625	245	242	249	321	650	328	510	444	452	441	575	484	460	484	482	489	403	352	59	113	227
vi.	304	205	649	228	287	234	382	588	357	504	418	500	375	575	331	429	434	509	420	371	293	59	143	245
sá.	215	277	459	128	84	111	226	437	163	331	273	210	295	319	225	313	253	335	273	203	173	36	97	133

SEO



4

- ✓ **Palabras clave** en título y primer párrafo.
- ✓ **Subtítulos.**
- ✓ **Cantidad de caracteres** en **título.**
- ✓ **Cantidad de palabras** en el **cuerpo.**
- ✓ **Cantidad de caracteres** en **párrafos.**
- ✓ **Cantidad de caracteres** en la **bajada.**
- ✓ **Cantidad de caracteres** en la **volanta.**
- ✓ **Cantidad de epígrafes.**
- ✓ **Cantidad de párrafos** en **negrita.**
- ✓ **Cantidad de contenido multimedia.**
- ✓ **Cantidad de links.**
- ✓ **Firma** del autor.



DASHBOARD CATEGORÍA SEO

Actualizado: 2021-03-02-11:33:42

Categoría SEO:

(All)

Autor:

(All)

Content id:

Buscar Content id

Información de las **notas**
(fecha publicación, título, sección).

SEO

Semáforo SEO
(puntuación ítem por ítem).

Versiones

9 - CREADA - 04/05/2021 10:09:14

Privada

NO

NOTA

RELACIONES

VERSIÓN

MILENIUM

WEB

SEMÁFORO SEO

xiGy2qUhN

Score: 72.43 / 100

Tier 1

* Keywords en título y primer párrafo

* Cantidad caracteres título

* Cantidad caracteres bajada

* Cantidad palabras en cuerpo

Tier 2

* Cantidad subtítulos H2

* Cantidad caracteres volanta

* Cantidad párrafos negrita

* Cantidad de fotos multimedia

Tier 3

* Cantidad de links

* Cantidad caracteres en párrafos

* Cantidad de epígrafes

* Firma autor

Observaciones:

Cantidad máxima de caracteres en párrafos: 400 | Número de párrafos que no lo cumplen: 2.

Sin palabras claves detectadas en el primer párrafo

(keywords sugeridas: marcos, españa, ", uruguay, matecaps).

Cantidad mínima de links totales: 3 | Número actual: 1 + Mirá También: 0.

Reportes



5

Autores

BIG DATA
ANÁLISIS Y VISUALIZACIÓN

Performance de Autores

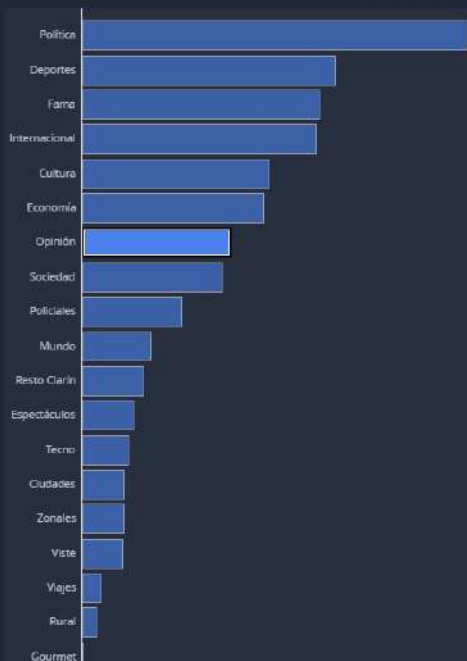
Autores








Autores y Notas

Notas

#

%



Autor	Notas	PVs/Notas	Br./Notas	Reg./Notas	Susc./Notas	Conv./Notas
 Alejandro Borensztein	1	106.801	54.458	27.363	24.184	33,00
 Julio Barbaro	1	35.078	26.948	6.688	5.735	1,00
 Ricardo Kirschbaum	1	9.244	5.847	3.729	3.413	1,00
 Eduardo van der Kooy	1	23.183	11.979	6.944	6.163	1,00
 Rubén Torres	1	16	13	4	4	0,00
 Félix V. Lonigro	1	489	322	31	21	0,00
 Mariano Turd	1	1.105	804	98	71	0,00



Ventas Olé Digital al 30/4/2021

Cartera Activa

47.587

Altas Mes Actual

12.250

Altas del Día

48

Cartera activa por plan

Plan	Cartera	% Total
Digital Olé	2.533	5,32%
Digital Olé + 365	504	1,06%
Corporat.	44.550	93,62%
Total	47.587	100,00%

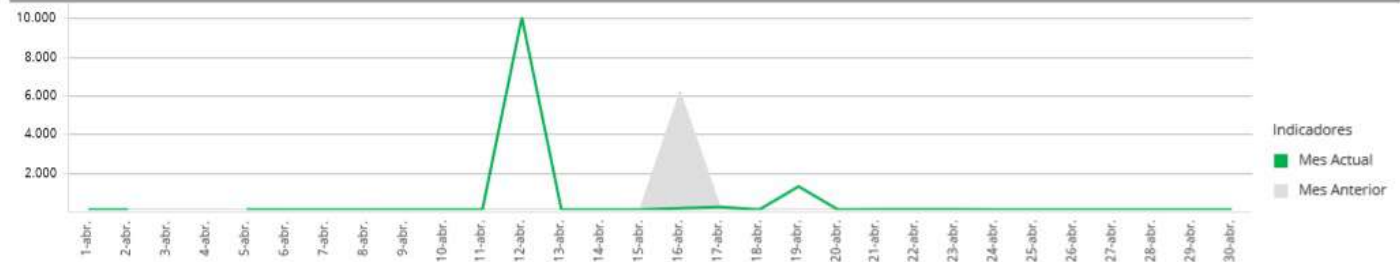
Variación contra mes anterior

Plan	Altas	% Vs. Mes Ant.
Digital Olé	415	43,60%
Digital Olé + 365	79	36,21%
Corporat.	11.756	86,54%
Total	12.250	84,24%

Variación contra día anterior

Plan	Alta	% Vs. Día Ant.
Digital Olé	41	(12,77%)
Digital Olé + 365	7	(22,22%)
Corporat.	0	(100,00%)
Total	48	(14,29%)

Altas por día



Dashboards
Ole Digital

Dashboard Choques y Altas

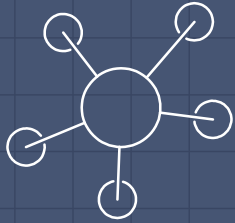


Modelos Predictivos



6

MODELOS PREDICTIVOS



CLUSTERS:

- Propensión *Paywall*.
- Cierre de notas.
- Acciones puntuales.



GEOMARKETING:

- Sucursales más cercanas para cada usuario de Clarín 365.

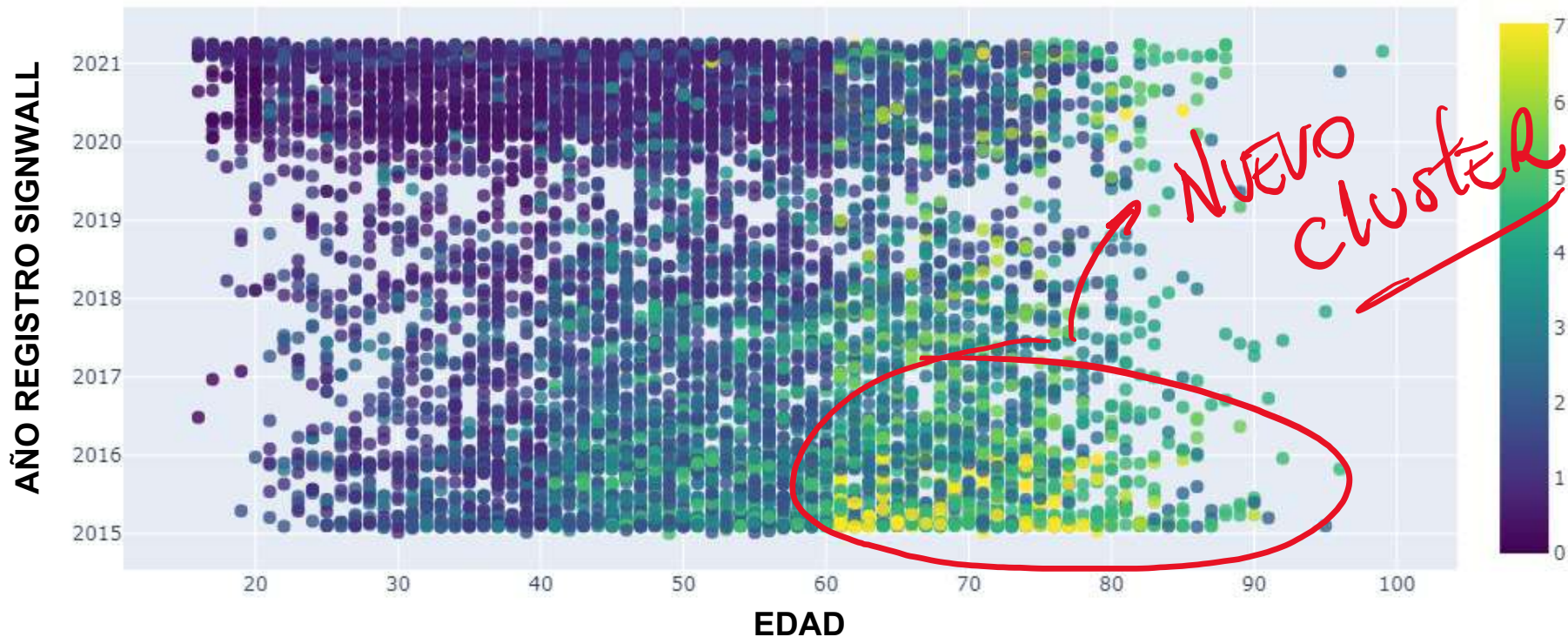


CHURN:

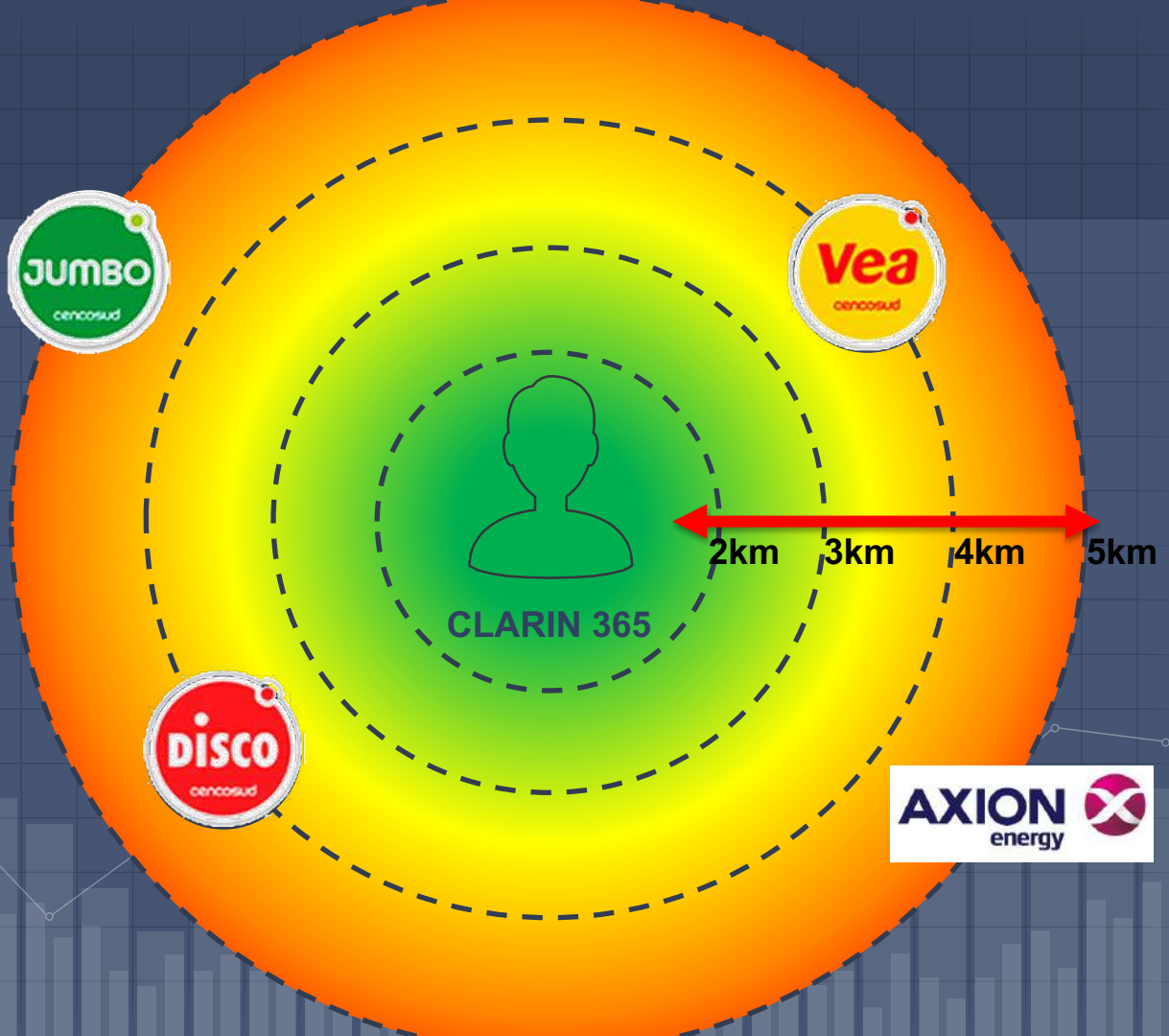
- Suscripciones digitales.

CLUSTERS – OPTIMIZAR BLOQUEOS PAYWALL

%CR por Edad y Antigüedad



GEO-MARKETING



CHURN

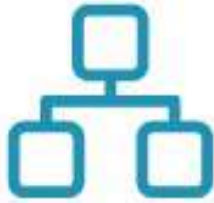
CHURN RATE: EL PROCESO CONSTA DE 4 PARTES



DATOS



VARIABLES



MODELO



CLUSTER

**BASE
FEBRERO 2021**

30% con más prob. de baja
Incluimos al **60%** de las bajas reales.

Potencial...

- Reforzar analítica de datos.
- Hypercards en tiempo real.
- Home Analytics.
- Conversión Rate Optimization
- Periodismo de datos.
- Nuevas tecnologías.



Potencial...

- Reforzar analítica de datos.
- Hypercards en tiempo real.
- Home Analytics.
- Conversión Rate Optimization.
- Periodismo de datos.
- Nuevas tecnologías.

SER DATA DRIVEN



Potencial...

- Reforzar analítica de datos.
- **Hypercards en tiempo real.**
- Home Analytics.
- Conversión Rate Optimization.
- Periodismo de datos.
- Nuevas tecnologías.



The screenshot shows a news article on the Clarín website. The page header includes the Clarín logo and the section 'Política'. The main headline reads: 'Tenía 71 años Murió Eugenio Zanmarini, superintendente de Servicios de Salud'. Below the headline, a sub-headline states: 'Era un hombre cercano al ex ministro Ginés González García. Fue hallado muerto en su domicilio.' The image below the text shows two men in front of a screen displaying the logo for 'Salud SUPERINTENDENCIA DE SERVICIOS DE SALUD'.

Potencial...

- Reforzar analítica de datos.
- Hypercards en tiempo real.
- Home Analytics.
- Conversión Rate Optimization.
- Periodismo de datos.
- Nuevas tecnologías.

The screenshot displays a news website dashboard with a dark purple header. The header includes the site name 'LA VACACION EN ARGENTINA', an 'ENTRAR' button, and several statistics: 'Días distribuidos' (9,326,408), 'Cae 1 dólar' (6,926,650, 15.12%), 'Cae 2 dólares' (938,563, 2.05%), 'Mejores' (60.08%), 'Hombres' (39.71%), and 'Otros' (0.21%). Below the header, there are several article cards. The main article is 'Adelanto exclusivo Muerte de Maradona: concluyen que el tratamiento que recibió fue "deficiente y temerario"'. Other articles include 'Presión a los precios El Gobierno sube 9% la luz y le pide la renuncia al subsecretario de Energía', 'Sin clases presenciales La Provincia autorizó una suba del 12.5% en las cuotas de los colegios privados', 'Tenía 21 años Confirmaron que un joven argentino murió en la estampida de Israel', 'Refuerzo excepcional de 15 mil pesos', 'Una amiga tradición Cómo es la festividad de Lag BaOmer donde murió el joven argentino', 'Habla a las 17 Nuevas restricciones en la Ciudad: cuáles serán los anuncios de Larreta', 'Pocos cambios Una por una, todas las restricciones que regirán hasta el 21 de mayo', and 'A los 83 años Murió la ex vedette y actriz Libertad Leblanc, un ícono del cine argentino'. Each article card features a thumbnail image, a title, a brief description, and a percentage indicator.

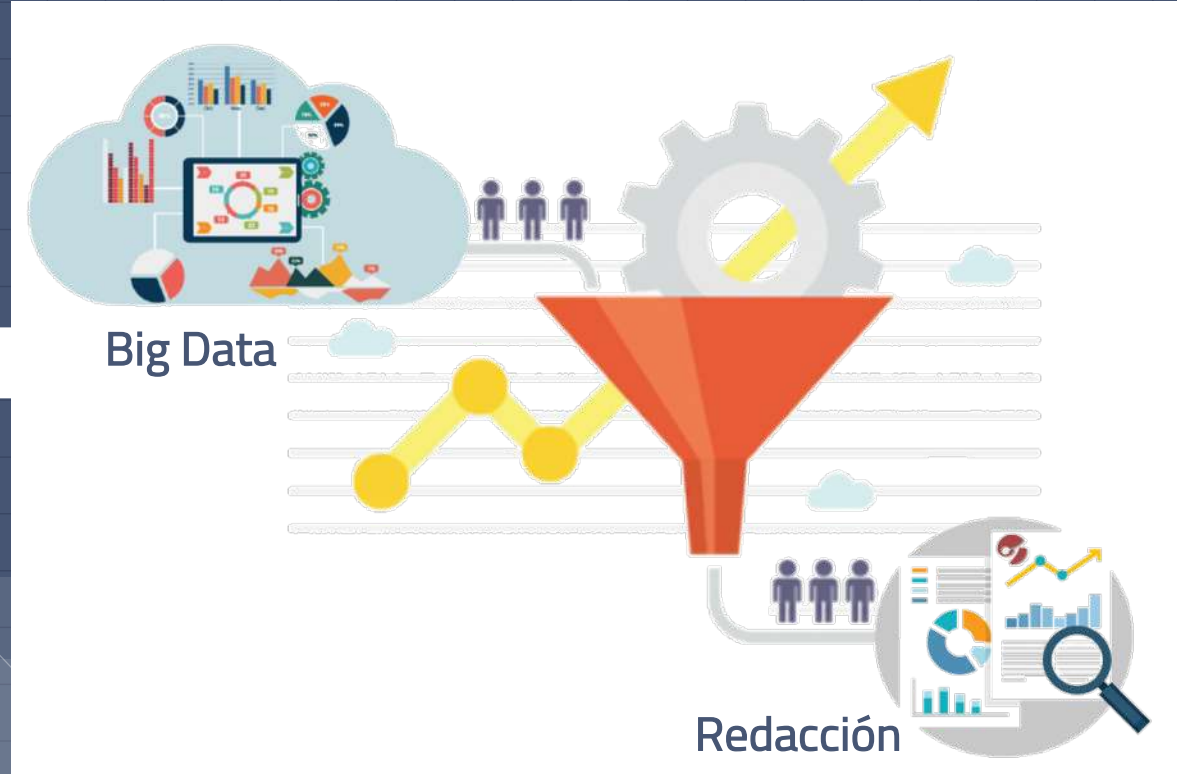
Potencial...

- Reforzar analítica de datos.
- Hypercards en tiempo real.
- Home Analytics.
- **Conversión Rate Optimization.**
- Periodismo de datos.
- Nuevas tecnologías.



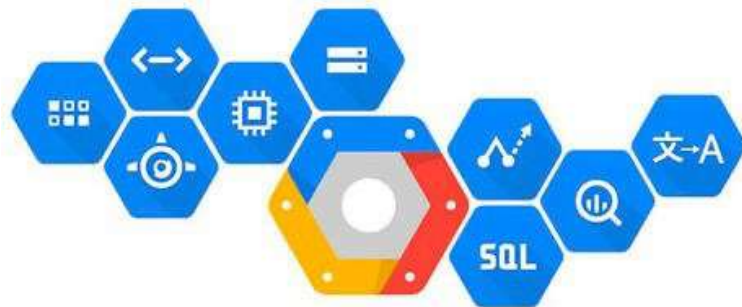
Potencial...

- Reforzar analítica de datos.
- Hypercards en tiempo real.
- Home Analytics.
- Conversión Rate Optimization.
- Periodismo de datos.
- Nuevas tecnologías.



Potencial...

- Reforzar analítica de datos.
- Hypercards en tiempo real.
- Home Analytics.
- Conversión Rate Optimization.
- Periodismo de datos.
- Nuevas tecnologías.



Google Cloud Platform

- Mejores procesos iterativos.
- Mayor integración de datos.
- Mas velocidad de procesamiento.
- Simplicidad y herramientas worldclass.

Propensión Paywall Scrapper
Reporting Performance GCP
RAUL Recomendador AGEA Competencia Grafos
Equipo Usuarios
Datos Tráfico Repositorios Modelado Choques
Tableros Real Time
Reportes Home Analytics Data Driven Geomarketing Servicios
Big Data Automatización Analytics Herramientas Tecnología
SEO Contenido Predicciones Optimización Semáforo
Churn Información Experiencia de Usuario Forecasting Pricing
Altas Clustering Modelos Predictivos Clarin 365 Perfil
Navegación Periodismo de Datos Hypercards
CRO Auto-Bot Macine Learning AnálisisMIRTA

Bigdata

