

Somos una empresa de datos.
Construimos soluciones centradas
en el cliente, la tecnología
y la información

Somos

El **Dato** como **Centro** del negocio

Data
Centric

Pioneros en soluciones **Data Driven** para objetivos de marketing y para la toma de decisiones.

El más completo **DataLake** de información veraz de empresas y consumidores

DATOS

APIs accesibles de **Data As A Service**, Tecnologías **Big Data** y Plataforma de gestión de campañas

TECNOLOGÍA

Un equipo de más de 60 profesionales entre Data Scientists, Informáticos y Asesores de Marketing

SERVICIO



Aprendemos de los datos para:



Mejorar el performance de los procesos de negocio



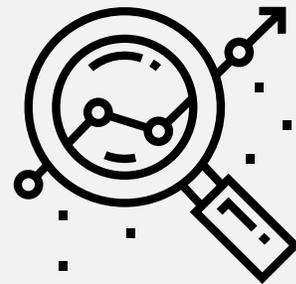
Optimizar las inversiones en acciones comerciales y de marketing



Generamos Insights de tu mercado a través de Data Analytics



Te ayudamos a
entender



DATA CRAWLING & SCRAPING

OBJETIVO: Identificación de tiendas online y creación de un data set de tiendas ecommerce para prospección comercial

Keywords Objetivo

Creación del Scoring

► Para la identificación de tiendas online en las URL's de empresas en España analizadas, se han definido **keywords habituales en los procesos de compra** online y se han ponderado en base al potencial de estar asociados a un eCommerce; a mayor puntuación (scoring) mayor probabilidad de afinidad, es decir, mayor probabilidad de que dicha URL sea realmente un ecommerce

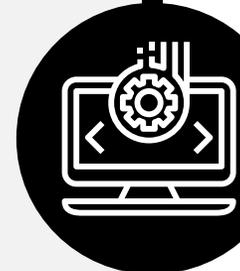
Keyword	Peso
▪ Agregar al carro	100
▪ carrito de la compra	100
▪ añadir a la cesta	100
▪ tu cesta	65
▪ Formas de pago	35
▪ cesta de la compra	100
▪ compra online	35
▪ gastos de envío	60
▪ tienda online	100
▪ mastercard visa paypal	60
▪



Web Crawling

Descarga de url's

- ▶ Descarga de todas la URL's españolas que contengan, al menos, **una de las keywords** anteriormente definidas; de esta forma se extraerán todas la posibles tiendas de ecommerce y las keywords por las que han sido seleccionadas: 1.241.441 URL's brutas
- ▶ **Segmentación:** Selección de aquellas URL's que contengan tres (3) o mas keywords diferentes: 123.103 URL's netas, incorporando la ponderación realizada inicialmente



All_online_shops				
Palabras	url			
gastos de envío	http://www.flamashop.es/			
gastos de envío	http://www.fcsoluciones.es/			
gastos de envío	http://www.entretodos-comparting.es/			
gastos de envío	http://www.eglobalmoto.es/			
gastos de envío	http://www.e-toy.es/			
gastos de envío	http://www.edel-optics.es/			
gastos de envío	http://www.educmed.es/sec/revista.php			
gastos de envío	http://www.galeonvigo.es	www.vivamondo.es	mi cuenta	0,4
gastos de envío	http://www.tentutienda.es	www.vivapets.es	mi cuenta	0,4
gastos de envío	http://www.tarrinasperso.es	www.vivaq.es	mi cuenta	0,4
gastos de envío	http://www.safetycloud.es	www.vivaracha.es	mi cuenta	0,4
gastos de envío	http://www.sanico.es/	www.vivecorriendo.es	mi cuenta	0,4
gastos de envío	http://www.retratos-a-la.es	www.viveencornella.es	mi cuenta	0,4
gastos de envío	http://www.lacasadelzur.es	www.viveenviladecans.es	mi cuenta	0,4
gastos de envío	http://www.relojate.es/	www.vivendo.es	mi cuenta	0,4
gastos de envío	http://www.relojes-mare.es	www.viveourense.es	mi cuenta	0,4
gastos de envío	http://www.roansasl.es/	www.viverosbarber.com	mi cuenta	0,4
gastos de envío	http://www.comprasmil.es	www.viverslaroda.es	mi cuenta	0,4
gastos de envío	http://www.tentutienda.es	www.vivit.es	mi cuenta	0,4
gastos de envío	http://www.sacaba.es/in	www.vivoenmovimiento.es	mi cuenta	0,4
		www.vivonium.es	mi cuenta	0,4
		www.vizuals.es	mi cuenta	0,4
		www.vlaigo.es	mi cuenta	0,4
		www.vmarco.es	mi cuenta	0,4
		www.vnetlibrerias.es	mi cuenta	0,4
		www.vocesenlared.com	mi cuenta	0,4
		www.vodafoneencasa.es	mi cuenta	0,4

Selección

Suma de las diferentes ponderaciones

- Selección: Suma de las diferentes ponderaciones asignadas a cada keyword, obteniendo una puntuación total para cada URL concreta; solo se seleccionan aquellas con una puntuación total superior a 0,3, obteniendo finalmente **56.911 URL's únicas**

host	Score_host
www.donostifotomaton.es	0,7
www.donostik.com	0,4
www.donpanchito.es	0,3
www.donpapel.es	1,6
www.donpet.es	5,8
www.donplacer.com	0,4
www.donquijotedelared.es	0,3
www.donreloj.com	5,35
www.donsaturnino.es	0,3
www.donsofaycia.es	3,5
www.dontbelieve.es	0,3
www.dontpanic.es	0,3
www.dontstopme.es	3,4
www.dontworryspain.es	0,3
www.donum.es	0,3
www.donvaquero.es	6,1
www.donyagreta.com	0,3
www.doodigital.es	0,7
www.doodoo.es	0,35
www.doombox.es	0,3
www.doonde.es	1
www.doordefender.es	4,15
www.doparana.es	1,3
www.dopra.es	2,9
www.doriangray.es	0,3



Scraping

Extracción de datos

- Web Scraping: Extracción de los datos de contacto de cada URL (Razón Social, Teléfono, CIF, Dirección postal, para el proceso de matching sobre la base de datos de negocios de Datacentric y, de esta forma, realizar el enriquecimiento con variables financieras: Actividad, Ventas, Empleados, etc.:

nie 20.235

tiendas ecommerce con datos de contacto



WEB	nif	rzs	direccion	cp	localidad
www.mallkart.es	BB6943677	SHOP CART SL	PLAZA REPUBLICA DEL ECUADOR, 7 1 A	28016	MADRID
www.teniscap.es	BB6944872	TENISCAP RC SL	PLAZA DE LA ENONA, 2	28100	ALDOBENDAS
www.kulubaa.es	BB6946795	KULUBAA FASHION SL	CALLE LANGAROTE, 9 2 C	28705	SAN SEBASTIAN DE LOS REYES
www.multimaradegal.es	BB6946734	MULTIMARADA LEGAL SLP	PARCO CASTELLANA, 114 B	28048	MADRID
www.gro-natura.es	BB6949059	DISTRIBUCIONES TICHE NATURAL SL	CALLE ECHegarai, 8	28014	MADRID
www.laaya.es	BB6959996	LAPTE SOLUTIONS SL	AVENIDA FRANCISCO JAVIER SÁENZ DE OJEDA, 220 B B	28055	MADRID
www.typr.es	BB6966439	FRANQUICIAS AGTOYS SL	CALLE DE MANUEL JIMÉNEZ, 28 1 DOHA	28701	SAN SEBASTIAN DE LOS REYES
www.morecon.es	BB6968409	MORECON SERVICIOS PARA EMPRESAS SL	CALLE DEL SOL, 2 2 B 5	28523	LA PARTIDA - SANTA MONICA
www.cristalar.es	BB6979517	CRISTALIA VIDRO 2000 SL	CALLE CANTABRIA, 59	28979	SERRANILLOS DEL VALLE
www.acostar-garshop.es	BB6982949	JO & JO TRADING SL	CALLE ALFREDO NOBEL, 16 PTO 1 2 4	28922	ALCORCON
www.itechnologymadrid.com	BB6984085	ITECHNOLOGY MADRID SL	CALLE TARRAGONA, 32 3 07RO	28045	MADRID
www.bricceramica.es	BB6984881	BRICO CERAMICA Y BADO SL	AVENIDA DEL JARAFA, 18	28823	COLLADA
www.ingoste-madrid.es	BB6985553	INGOSTE INVERSIONES SL	CALLE LOPE DE RUEDA, 11 B 02A	28009	MADRID
www.atenaya.es	BB6986758	ATENAYS TECHNOLOGIES SL	CALLE VELAZQUEZ, 15 1	28001	MADRID
www.americaomotor.es	BB6987589	MULTIMARCA AMERICA MOTOR SL	CALLE CORAZON DE MARIA, 11 LOCAL	28002	MADRID

Gracias

MADRID

Tel. +34 91 382 20 00

C/ Nuria, 57. 28034

datacentricmad@datacentric.es

BARCELONA

Tel. +34 93 467 48 00

C/ Valencia, 243-245. 08007

datacentricbcn@datacentric.es

LISBOA

Tel. +351 935 746 448

Edifício Infante. Av. D. João II, 35

1990-083

info@datacentric.pt

