



ENTREGABLE FINAL

H&M Business Intelligence & Data Science

Autores

María Lourdes Linares Barrera
Pablo Reina Jiménez

ÍNDICE DE CONTENIDOS

0 _____
Introducción

1 _____
Branding

2 _____
Dashboards

3 _____
Asistencia a
marketing

4 _____
Comentarios
finales

Pablo Reina Jiménez



M. Lourdes Linares Barrera



¿QUIÉNES SOMOS?

Identificador:

H&M BI. NLP, visualización y predicción

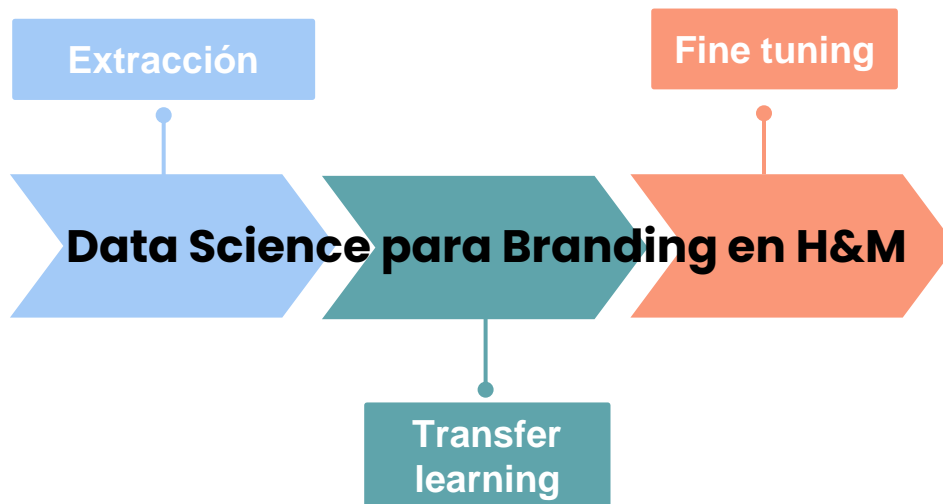
Título:

**Business Intelligence y Data Science para la
empresa de ropa H&M**



RubricaSoft

BRANDING



Objetivo

¿Cuál es la opinión de nuestra empresa?

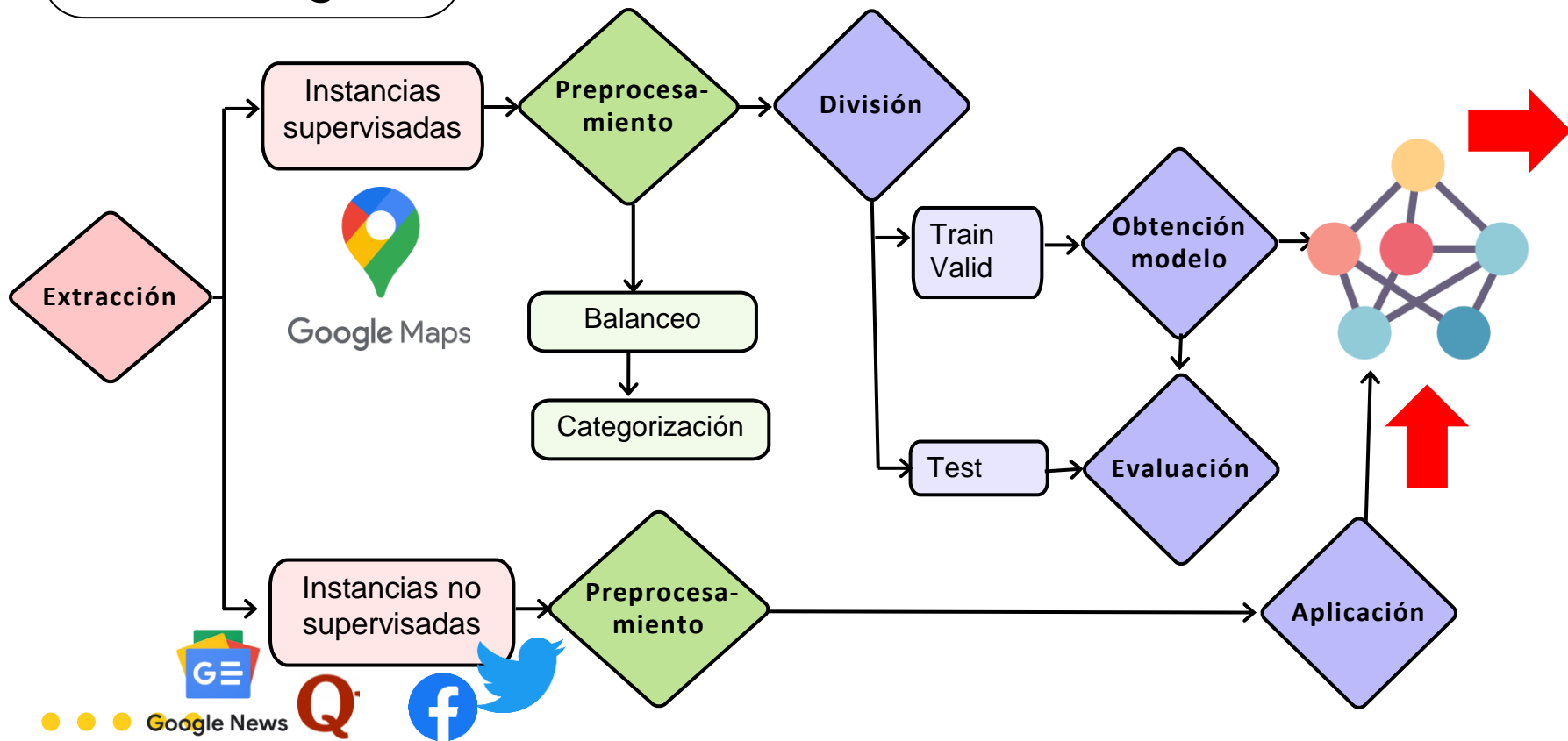


En qué mejorar

Dónde mejorar



Metodología



Preprocesamiento

Scraping



Preprocesamiento



Eliminación de reseñas vacías

Limpieza de reseñas

Idioma (inglés)



Eliminación de emojis

CLEAN-TEXT

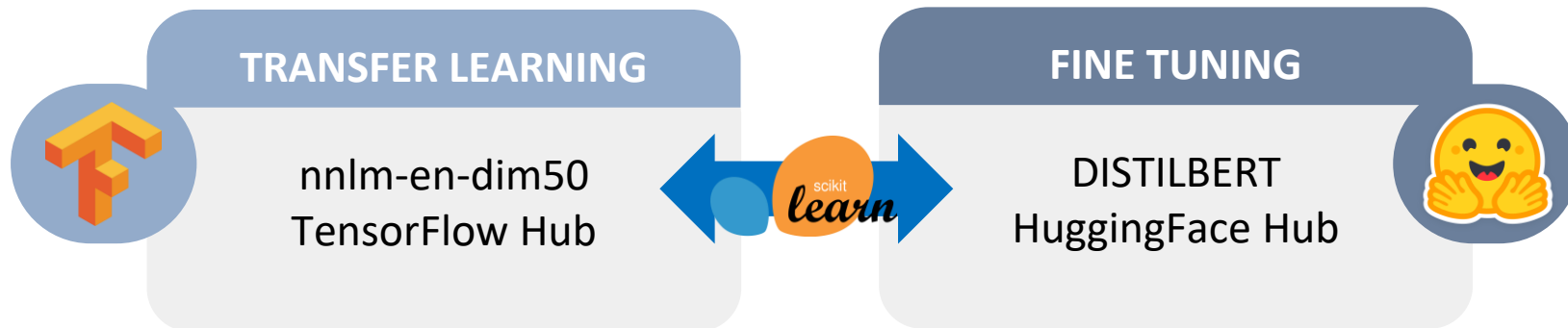
Filtro por subjetividad

VADER



TextBlob

Modelos

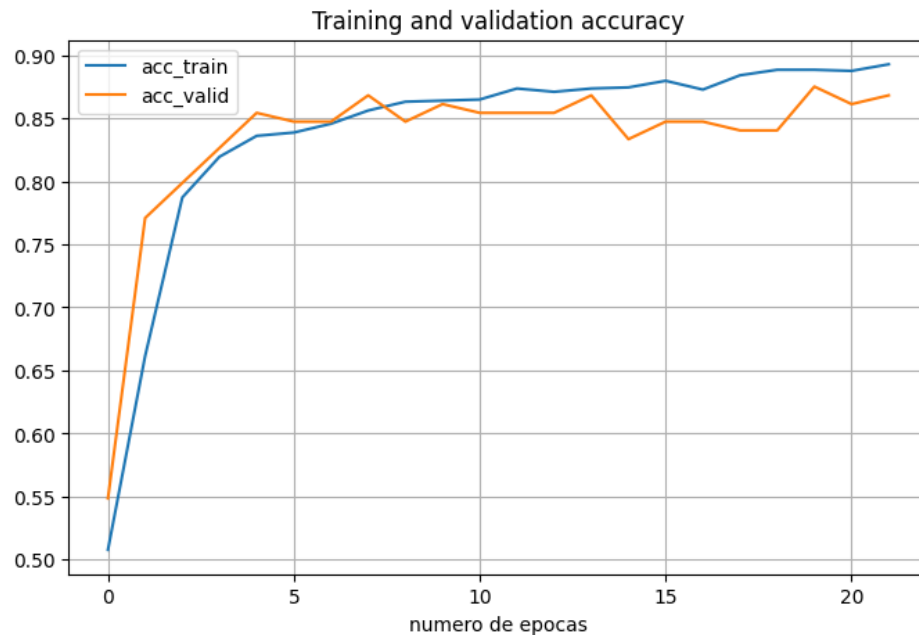


Modelos

TRANSFER LEARNING




nnlm-en-dim50
TensorFlow Hub

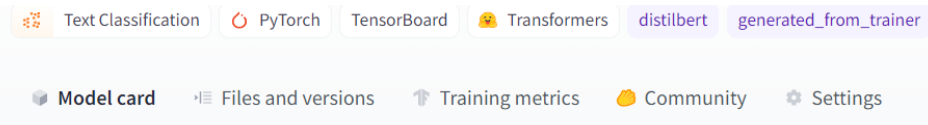


➔ **86%**

Modelos

FINE TUNING

 **DISTILBERT**
HuggingFace Hub



finetuning-sentiment-model-ie

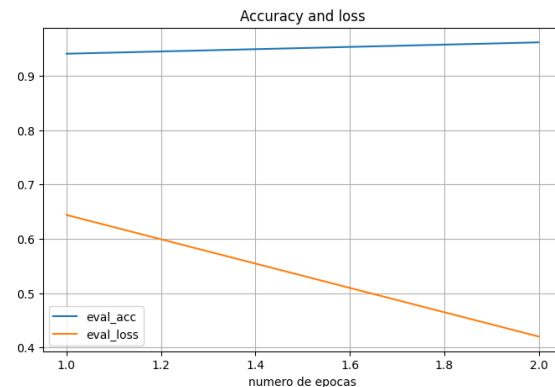
This model is a fine-tuned version of [distilbert-base-uncased](#) on google_reviews dataset. It achieves the following results on the evaluation set:

- Loss: 0.15
- Accuracy: 0.95
- F1: 0.95

Model description

More information needed


Intended uses & limitations

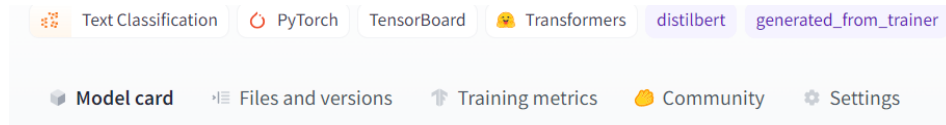


 **95%**

Modelos

FINE TUNING

 **DISTILBERT**
HuggingFace Hub



finetuning-sentiment-model-ie

This model is a fine-tuned version of [distilbert-base-uncased](#) on google_reviews dataset. It achieves the following results on the evaluation set:

- Loss: 0.15
- Accuracy: 0.95
- F1: 0.95

Model description

More information needed

Intended uses & limitations

 **95%**





Resultados

Text Classification PyTorch TensorBoard Transformers **distilbert** generated_from_trainer License: apache-2.0

Model card Files and versions Training metrics Community Settings

Train Deploy Use in Transformers

Edit model card

Downloads last month

6

Ejemplo de funcionamiento



Hosted inference API

Text Classification

Examples

this store is so cool its set out in a great way that you can see everything

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: cached

LABEL_1

0.967

LABEL_0

0.033

JSON Output

Maximize

finetuning-sentiment-model-ie

This model is a fine-tuned version of distilbert-base-uncased on google_reviews dataset. It achieves the following results on the evaluation set:

- Loss: 0.15
- Accuracy: 0.95
- F1: 0.95

Métricas de eficiencia

Model description

More information needed

Intended uses & limitations

SCAN
ME!

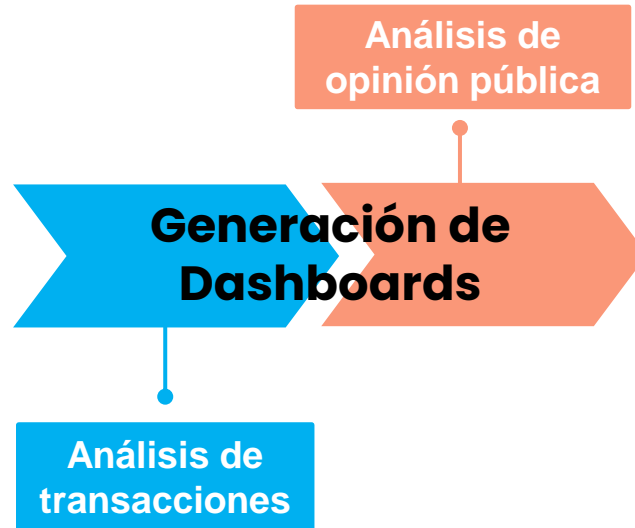




Modelos

site	review	label	score
facebook	great selection and great service	1	0,97160404920578
facebook	how on earth do you get in touch with an actual store! the "customer service" line is far from customer service with an endless loop and voice	0	0,981667280197144
facebook	5 hours later and 2 email i still haven progressed. i also rang the 1800 number, pressed 3 to speak to a customer service officer. i was promptl	0	0,972457706928253
facebook	the dress i bought shrunk in a very unusual way and i have made 7 phone calls to date. i only getting message, ' . . .busy call back later.' or i an	0	0,982921183109283
facebook	this store is so cool its set out in a great way that you can see everything	1	0,967416524887085
facebook	great prices and fun style	1	0,973173201084137
facebook	The clothes are well attended, and a lot of products of very good quality as well as good prices.	1	0,973054766654968
facebook	Terrible service. I'm coming out of a fracture and I can't stand for a long time! The supervisor Alex sent me to form in a great line! what a lack	0	0,979790568351746
facebook	I'm the one who buys shoes today, but you said... they didn't accept my debit card, but they already charged me. and I asked the bank that th	0	0,98026043176651
facebook	nice selection. great prices. found some good deals.	1	0,975844621658325
facebook	good service clean and tidy	1	0,970619261264801
facebook	all current in fashion, affordable prices, diversity of sizes and above all clothing ordered by size.	1	0,939377248287201

DASHBOARDS



**SCAN
ME!**



Organización del dashboard

1. Análisis de transacciones

Análisis de ventas

Informe de ventas global

Informe de ventas anual

Análisis de productos

Análisis de clientes

2. Análisis de opinión pública

Análisis de reseñas (Google Reviews)

Análisis de sentimientos (predicción)

Redes sociales (Twitter, Facebook)

Foros (Quora)

Periódicos (Google News)

WordClouds

1. Dashboard de transacciones

1. Análisis de transacciones

Análisis de ventas

Informe de ventas global

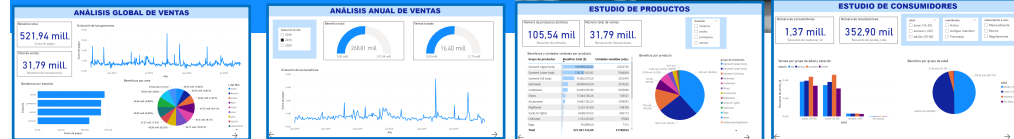
Informe de ventas anual

Análisis de productos

Análisis de clientes

01

ANÁLISIS DE VENTAS



ANÁLISIS GLOBAL DE VENTAS

Beneficio total

521,94 mill.

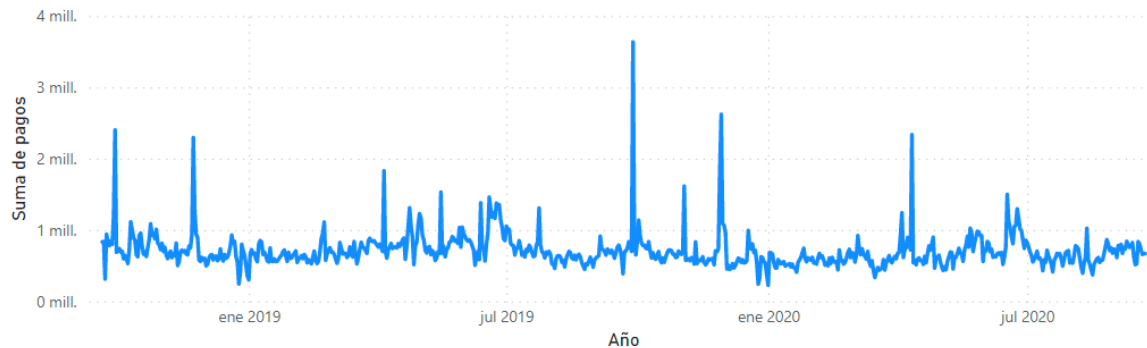
Suma de pagos

Total de ventas

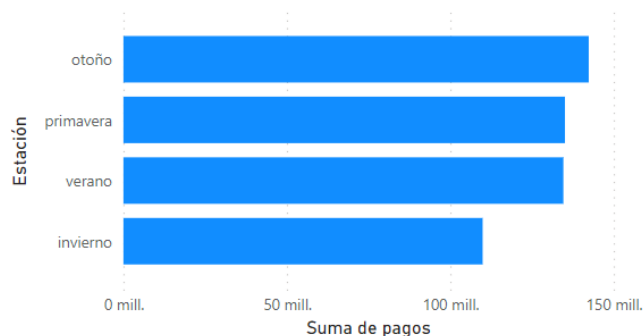
31,79 mill.

Número de transacciones

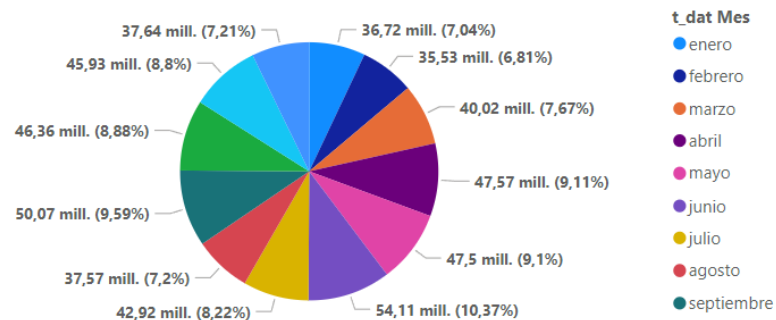
Evolución de las ganancias



Beneficios por estación



Beneficios por mes

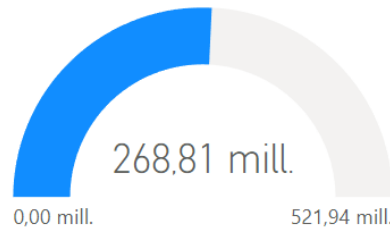


ANÁLISIS ANUAL DE VENTAS

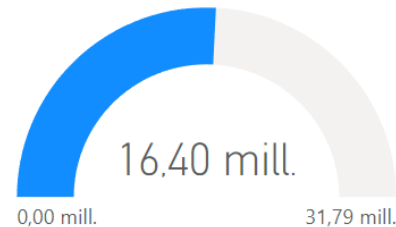
Selección de año

☐ 2018☒ 2019☐ 2020

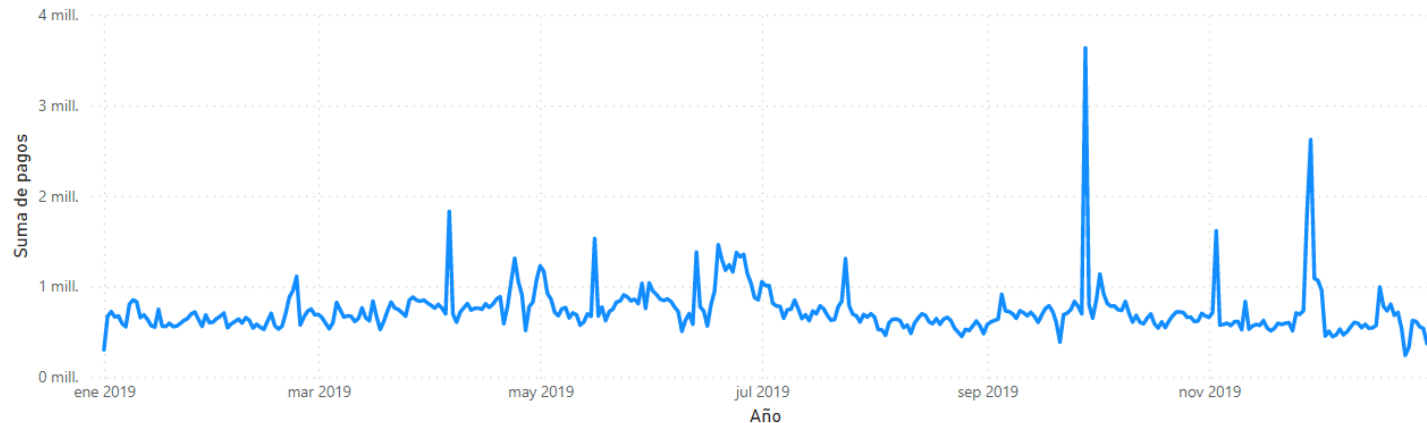
Beneficio total



Ventas totales



Evolución de los beneficios



ESTUDIO DE PRODUCTOS

Número de productos distintos

105,54 mil

Recuento de artículos

Número total de ventas

31,79 mill.

Recuento de transacciones

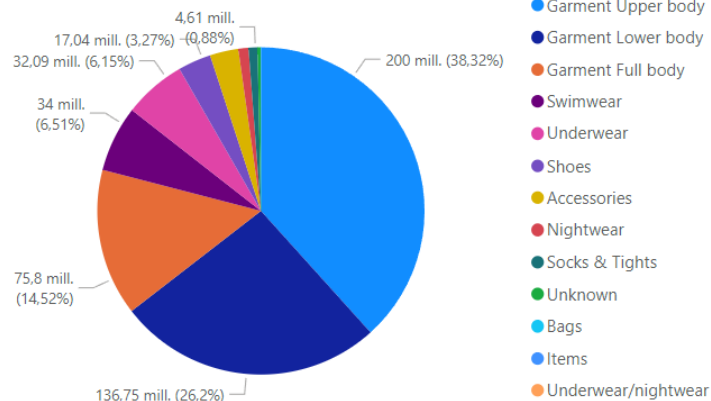
Estación

☐ invierno☐ otoño☐ primavera☐ verano

Beneficios y unidades vendidas por producto

Grupo de productos	Beneficio total (\$)	Unidades vendidas (uds.)
Garment Upper body	199.999.255,63	12552755
Garment Lower body	136.747.025,82	7046054
Garment Full body	75.802.575,59	3552470
Swimwear	34.000.618,24	2579222
Underwear	32.093.555,90	2565858
Shoes	17.044.130,26	745521
Accessories	14.687.355,23	1599593
Nightwear	5.223.161,08	348180
Socks & Tights	4.608.570,55	685712
Unknown	1.533.253,09	97040
Bags	143.889,50	7313
Total	521.941.124,69	31788324

Beneficio por producto





ESTUDIO DE CONSUMIDORES

Número de consumidores

1,37 mill.

Recuento de customer_id

Número de localizaciones

352,90 mil

Recuento de postal_code

edad

- ☐ joven (16-30)
- ☐ anciano (+60)
- ☐ adulto (30-60)

membresía

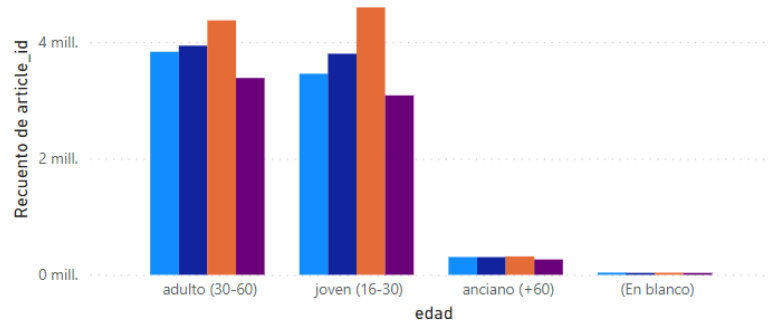
- ☐ Activo
- ☐ Antiguo miembro
- ☐ Precreada

suscripción a noti...

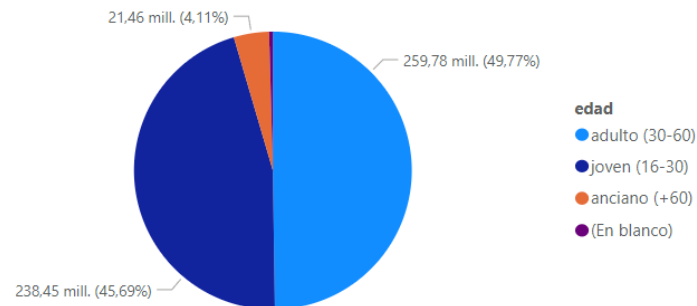
- ☐ Mensualmente
- ☐ Nunca
- ☐ Regularmente

Ventas por grupo de edad y estación

season ● fall ● spring ● summer ● winter



Beneficio por grupo de edad



2. Dashboard de branding

2. Análisis de opinión pública

Análisis de reseñas (Google Reviews)

Análisis de sentimientos (predicción)

Redes sociales (Twitter, Facebook)

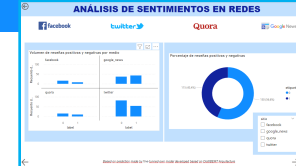
Foros (Quora)

Periódicos (Google News)

WordClouds

02

OPINIÓN PÚBLICA DE LA MARCA



ANÁLISIS GOOGLE REVIEWS

Valoración media por ciudad

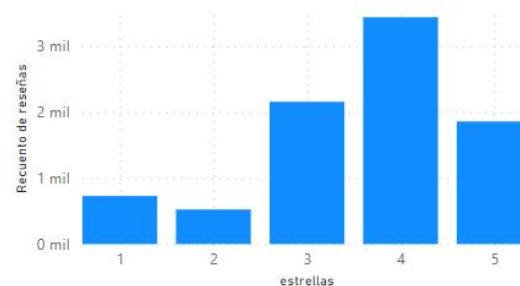
ciudad ● Boston ● Vancouver ● Toronto ● London ● Quebec ● Lakeland



ciudad

- ☐ Boston
- ☐ Lakeland
- ☐ London
- ☐ Quebec
- ☐ Toronto
- ☐ Vancouver

Recuento de valoraciones

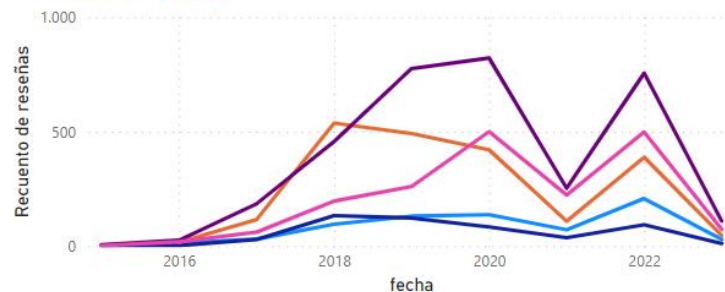


Evolución de valoración media



Evolución por valoración

estrellas ● 1 ● 2 ● 3 ● 4 ● 5

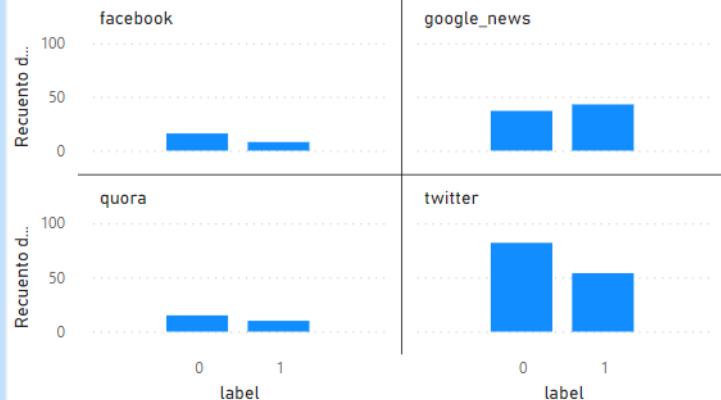




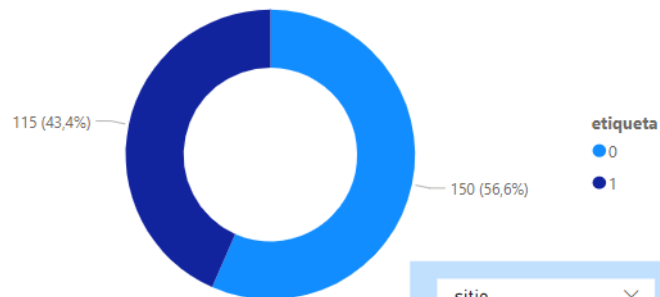
ANÁLISIS DE SENTIMIENTOS EN REDES



Volumen de reseñas positivas y negativas por medio



Porcentaje de reseñas positivas y negativas



- sitio
- ☐ facebook
 - ☐ google_news
 - ☐ quora
 - ☐ twitter



BRANDING ANÁLISIS DAFO

Fortalezas y oportunidades



Debilidades y amenazas



sitio

- ☐ facebook
- ☐ google_news
- ☐ google_reviews
- ☐ quora
- ☐ twitter



ASISTENCIA A MARKETING





Introducción

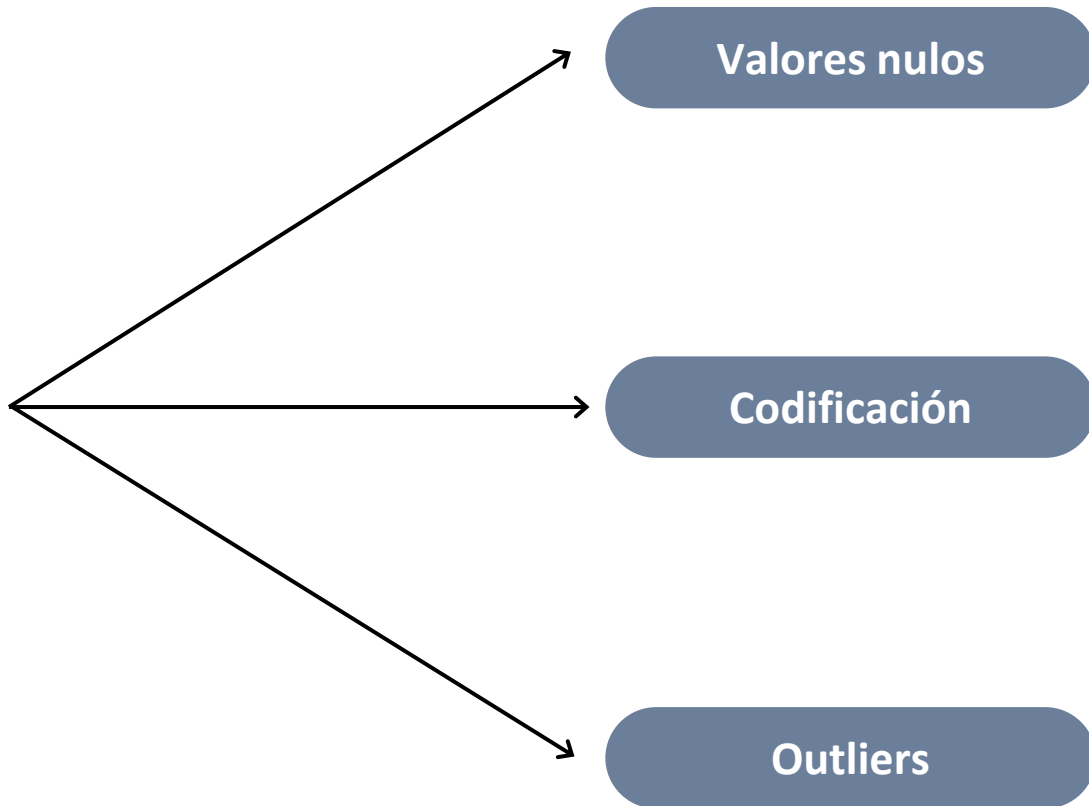
Branding

Análisis de
ventas

Asistencia a
marketing

Comentarios
finales

Preprocesamiento





Introducción

Branding

Análisis de
ventas

Asistencia a
marketing

Comentarios
finales

EDA

Visualizaciones preliminares



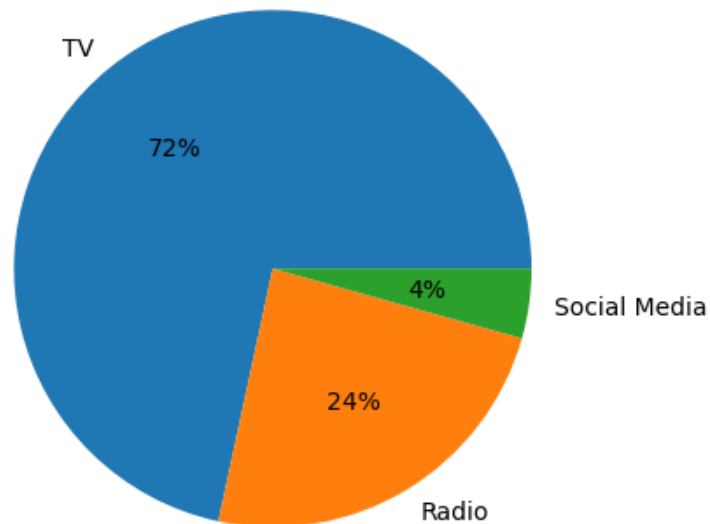
Presupuesto por medio

Densidad de ventas

Correlaciones

Más...

Budget per media source



EDA

Visualizaciones preliminares



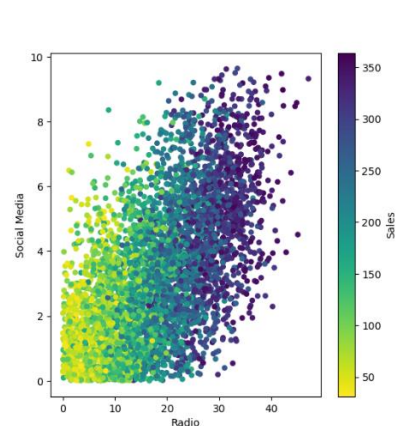
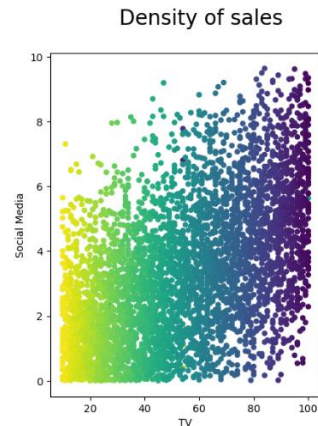
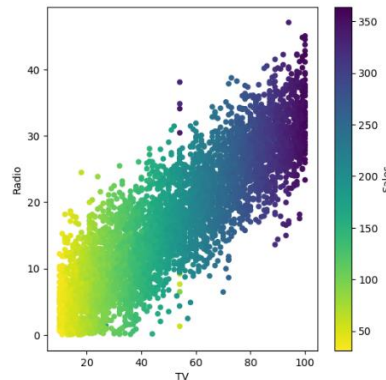
Presupuesto por medio

Densidad de ventas

Correlaciones

Más...

seaborn





Introducción

Branding

Análisis de
ventas

Asistencia a
marketing

Comentarios
finales

EDA

Visualizaciones preliminares

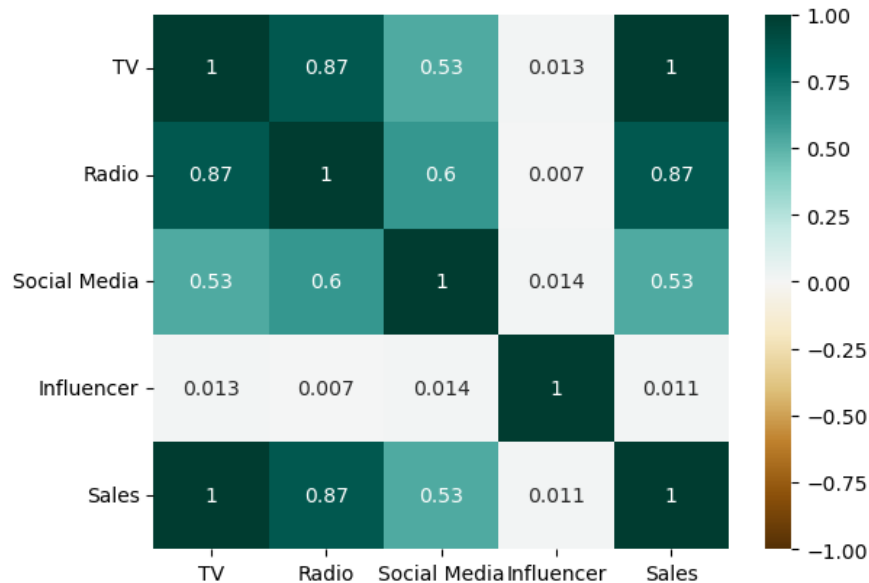


Presupuesto por medio

Densidad de ventas

Correlaciones

Más...



Modelos

REGRESIÓN LINEAL

Modelo de predicción
Explicabilidad

RANDOM FOREST REGRESSOR

Modelo de predicción
Importancia de los
atributos

scikit
learn



Introducción

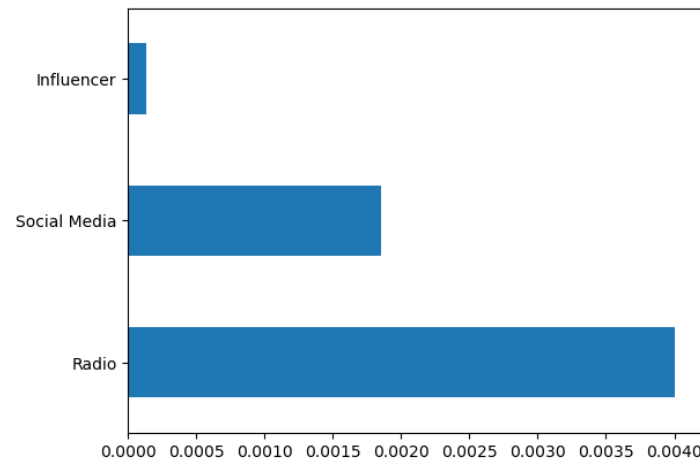
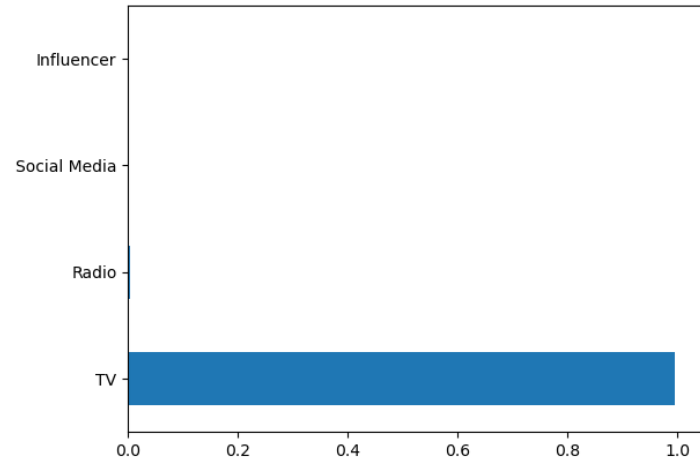
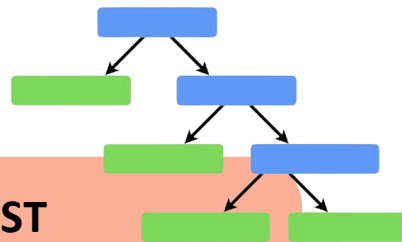
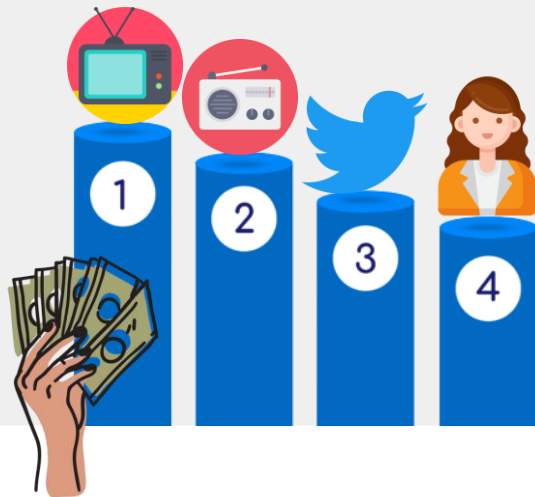
Branding

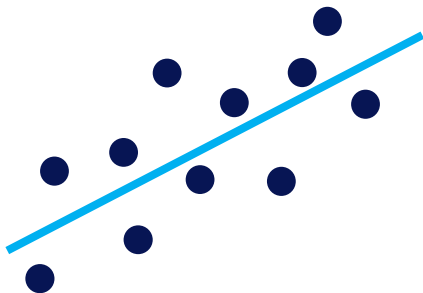
Análisis de
ventas

Asistencia a
marketing

Comentarios
finales

RANDOM FOREST





MODELO DE REGRESIÓN LINEAL

$$\text{sales} = 0.58 + 3.51 * \text{TV} + 0.11 * \text{Radio} + 0.06 * \text{SocialMedia} + -0.10 * \text{Influencer}$$



Introducción

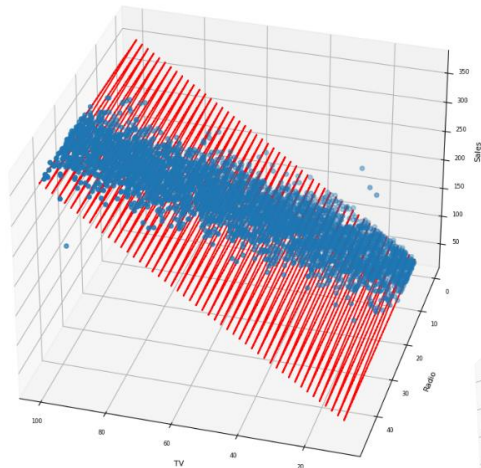
Branding

Análisis de
ventas

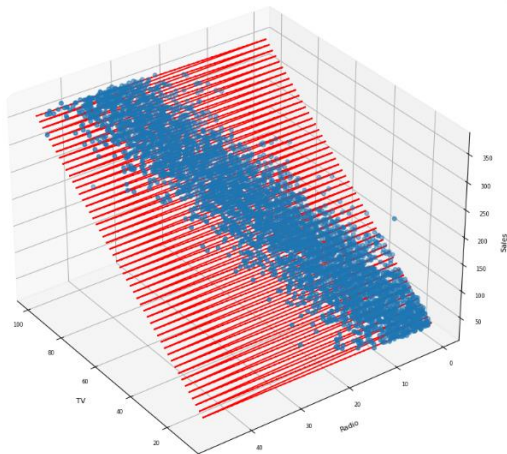
Asistencia a
marketing

Comentarios
finales

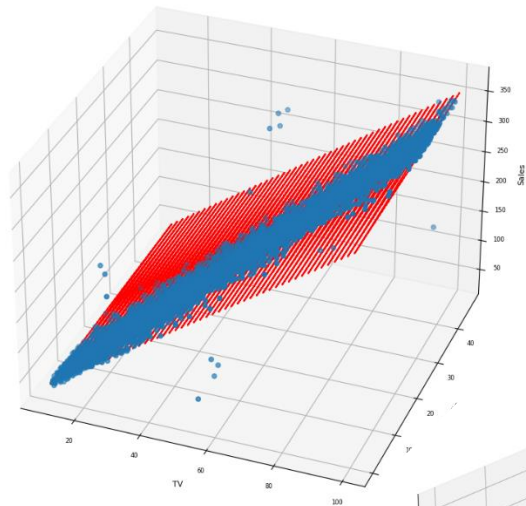
Proyeccion TV-Radio vs Sales



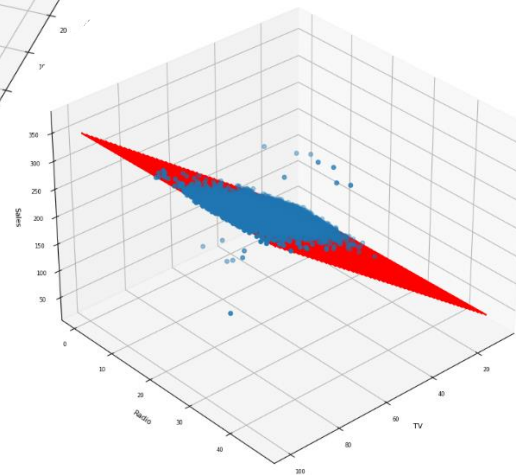
Proyeccion TV-Radio vs Sales



Proyeccion TV-Radio vs Sales



Proyeccion TV-Radio vs Sales





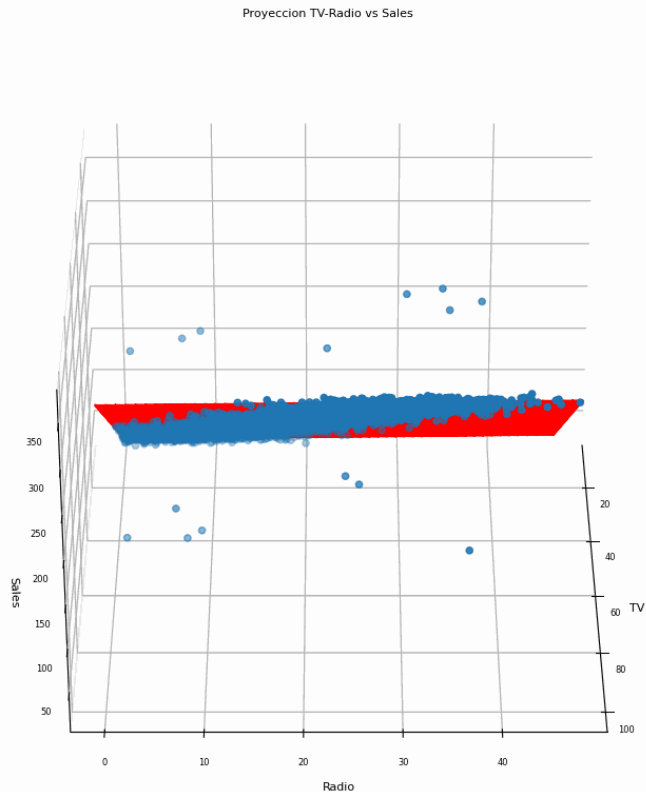
Introducción

Branding

Análisis de
ventas

Asistencia a
marketing

Comentarios
finales



Rendimiento

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^N \left| \frac{y - \hat{y}}{y} \right|$$

donde:

y – variable respuesta

\hat{y} – valor predicho y

\bar{y} – valor medio de y

	Regresión lineal	Random Forest
MAPE	1,966149	1,238034
R ²	0,994919	0,0993878

MUCHAS GRACIAS POR SU ATENCIÓN

Identificador:

H&M BI. NLP, visualización y predicción

Título:

**Business Intelligence y Data Science para la
empresa de ropa H&M**

