

Filtrado y Agregación de un Catálogo Online

Práctica de el módulo de Bases de Datos No SQL

Laura Rodríguez Roperó



Filtrado y Agregación de un Catálogo Online

Práctica de el módulo de Bases de Datos No SQL

por

Laura Rodríguez Roperó

07/11/2024

Índice

1. Introducción y contexto	1
2. Desarrollo de ejercicios	2
2.1 Inserción y Actualización de Datos	2
2.2 Proyección y Filtrado	4
2.3 Pipeline de Agregación	5
3. Conclusión	12

Profesor: Álvaro Bravo
Facultad: Facultad de Estudios Estadísticos
Universidad: Universidad Complutense de Madrid
Máster: Big Data, Data Science e Inteligencia Artificial



MongoDB



NoSQLBooster

1

Introducción y contexto

Actualmente, el análisis de datos provenientes de productos de moda es un área de interés creciente tanto para investigadores como para profesionales de la industria. Con el auge de las plataformas de comercio electrónico, las marcas y los minoristas recopilan grandes volúmenes de información sobre sus productos, que pueden incluir desde características físicas (como el tipo de prenda o el color) hasta descripciones detalladas y datos temporales (año de lanzamiento, temporada, entre otros). Estos datos se utilizan no solo para la gestión de inventarios y la optimización de ventas, sino también para identificar patrones de consumo, detectar tendencias de moda y mejorar la experiencia del cliente. Mi motivación para elegir este conjunto de datos es mi gran interés por el mundo de la moda y el diseño. Como soy una chica que acaba de llegar de Extremadura, me inspira la idea de que algún día pueda crear y patentar mi propia marca de moda.

Este trabajo se centra en el análisis de un **dataset de productos de moda**, que consta de **44,446 registros** de productos, cada uno con una serie de atributos clave. Entre estos atributos se incluyen:

- **ID del producto:** Identificador único para cada artículo.
- **Género:** Clasificación del producto por género (por ejemplo, "Men", "Women").
- **Categoría principal (masterCategory):** Tipo general del producto (como "Apparel" o "Accessories").
- **Subcategoría (subCategory):** Subdivisión más específica del tipo de producto (por ejemplo, "Shirts", "Watches", "Shoes").
- **Color base (baseColour):** Color principal del producto.
- **Temporada (season):** Época del año en la que el producto está destinado a ser usado.
- **Año (year):** Año en el que el producto fue lanzado o añadido al catálogo.
- **Uso (usage):** Contexto en el que el producto se utiliza, como "Casual" o "Formal".
- **Nombre descriptivo del producto (productDisplayName):** Nombre que describe el producto para su presentación al cliente.
- **URL de la imagen (image_url):** Enlace a la imagen del producto, que permite visualizarlo.

El objetivo de este trabajo es realizar un análisis exhaustivo de este dataset, se utilizará **MongoDB** como base de datos y **MongoDB Compass** junto con **NoSQLBooster** como herramientas de cliente para la ejecución de consultas y agregaciones. Además, se emplearán herramientas como **Python** para la generación de visualizaciones.

MongoDB es una base de datos NoSQL que permite manejar grandes volúmenes de datos estructurados y no estructurados de manera eficiente, lo que la convierte en una opción ideal para datasets heterogéneos como el que se maneja en este proyecto.

2

Desarrollo de ejercicios

A continuación, presento el desarrollo de los ejercicios de inserción, actualización, proyección, filtrado y agregación de datos que se han realizado sobre el dataset de productos elegido. Este análisis incluye varios scripts que permiten modificar y analizar la base de datos.

2.1. Inserción y Actualización de Datos

El dataset original se ha modificado para incluir las URLs de las imágenes asociadas a cada producto. Esta modificación es clave para que la información de las imágenes esté correctamente integrada en el análisis, y podamos realizar visualizaciones y consultas posteriores.

El siguiente script (Actualizar_dataset.js) actualiza los productos existentes añadiendo el campo `image_url` a partir de una colección separada de imágenes. Se conecta a MongoDB y itera sobre cada imagen para asociarla al producto correspondiente.

```
1 const { MongoClient } = require('mongodb');
2
3 async function updateProductsWithImages() {
4   const uri = "mongodb://localhost:27017";
5   const client = new MongoClient(uri, { useNewUrlParser: true, useUnifiedTopology: true });
6
7   try {
8     await client.connect();
9     const db = client.db('fashion_products');
10    const productosCollection = db.collection('styles');
11    const imagenesCollection = db.collection('images');
12
13    const imagenes = await imagenesCollection.find({}).toArray();
14
15    for (const imagen of imagenes) {
16      const productId = parseInt(imagen.filename.split('.')[0]);
17
18      await productosCollection.updateOne(
19        { id: productId },
20        { $set: { image_url: imagen.link } }
21      );
22    }
23    console.log("Actualización de productos con imágenes completada.");
24  } catch (error) {
25    console.error("Error al actualizar productos:", error);
26  } finally {
27    await client.close();
28  }
29 }
30
31 updateProductsWithImages();
```

Ejercicios sobre inserción y borrado de productos:

```
1 // Insertar un producto
2 db.styles.insertOne({
3   "id": 60000,
4   "gender": "Women",
5   "masterCategory": "Accessories",
6   "subCategory": "Hats",
7   "articleType": "Hats",
8   "baseColour": "Black",
9   "season": "Spring",
10  "year": 2024,
11  "usage": "Formal",
12  "productDisplayName": "Beret",
13  "image_url": "http://assets.myntassets.com/v1/images/style/properties/60000_images.jpg"
14 });
15
16 // Borrar el producto que acabamos de añadir
17 db.styles.deleteOne({
18   "id": 60000,
19   "gender": "Women",
20   "masterCategory": "Accessories",
21   "subCategory": "Hats",
22   "articleType": "Hats",
23   "baseColour": "Black",
24   "season": "Spring",
25   "year": 2024,
26   "usage": "Formal",
27   "productDisplayName": "Beret",
28   "image_url": "http://assets.myntassets.com/v1/images/style/properties/60000_images.jpg"
29 });
```

Ejercicios sobre actualización de productos:

```
1 // Actualizar la categoría de los relojes para hombre a 'Luxury Accessories'
2 db.styles.updateMany(
3   { "subCategory": "Watches", "gender": "Men" },
4   { $set: { "masterCategory": "Luxury Accessories" } }
5 );
6
7 // Volver a dejar la categoría como estaba
8 db.styles.updateMany(
9   { "subCategory": "Watches", "gender": "Men" },
10  { $set: { "masterCategory": "Accessories" } }
11 );
12
13 // Contar productos que no tengan url para comprobar que los hemos actualizado todos
14   correctamente, y de no existir imagen para algún producto escribir 'Undefined' (aunque
15   sabemos que todos lo tienen)
16 db.styles.find({ "image_url": { $exists: false } }).count();
17 db.styles.updateMany(
18   { $or: [ { "image_url": "" }, { "image_url": null } ] },
19   { $set: { "image_url": "Undefined" } }
20 );
```

2.2. Ejercicios sobre Proyección y Filtrado








Mostrar sólo los productos con categoría 'Apparel' enseñando únicamente los campos `image_url` y `productDisplayName`

```
1 db.styles.find(
2   { "masterCategory": "Apparel" },
3   { "productDisplayName": 1, "image_url": 1, "_id": 0 }
4 );
```

styles 0.025 s 21.400 Docs		
	productDisplayName	image_url
1	Turtle Check Men Navy Blue Shirt	http://assets.myntassets.com/v1/images/style/properties/7a5b82d1372a7a5c6de67ae7a314fd91
2	Peter England Men Party Blue Jeans	http://assets.myntassets.com/v1/images/style/properties/4850873d0c417e6480a26059f83aac29
3	Manchester United Men Solid Black Track Pants	http://assets.myntassets.com/v1/images/style/properties/8153dc35d9a5420eeb93922067137db6
4	Puma Men Grey T-shirt	http://assets.myntassets.com/v1/images/style/properties/Puma-Men-Grey-T-shirt_32668f8a6145
5	Inkfruit Mens Chain Reaction T-shirt	http://assets.myntassets.com/v1/images/style/properties/9c1b19682ecf926c296f520d5d6e0972_

Ordenar por año de forma descendente y limitar la salida a las 50 primeras filas

```
1 db.styles.find({}).sort({ year: -1}).limit(50);
```

styles 0.025 s 50 Docs																
	_id *	id	gender	masterCategory	subCatego	usage	articleType	baseColor	season	year	productDisplayName	image_url				
1	 67161ccda	55243	Women	Personal Care	Makeup	Casual	Foundation ar	Skin	Summer	2019	Lakme Perfecting Natural F	http://assets.myntassets.o				
2	 67161ccda	55238	Women	Personal Care	Makeup	Casual	Foundation ar	Skin	Summer	2019	Lakme Invisible Finish 04 F	http://assets.myntassets.o				
3	 67161ccda	55986	Women	Personal Care	Lips	Casual	Lipstick	Magenta	Summer	2019	ColorBar Velvet Matte Addi	http://assets.myntassets.o				
4	 67161ccda	55136	Women	Personal Care	Nails	Casual	Nail Polish	Purple	Summer	2019	Lakme Truewear Color Cru	http://assets.myntassets.o				
5	 67161ccda	56010	Women	Personal Care	Lips	Casual	Lipstick	Red	Summer	2019	Colorbar Velvet Matte Hot F	http://assets.myntassets.o				
6	 67161ccda	55270	Women	Personal Care	Makeup	Casual	Kajal and Eye	Black	Summer	2019	Lakme Insta-Liner	http://assets.myntassets.o				
7	 67161ccda	55440	Women	Personal Care	Makeup	Casual	Concealer	Beige	Summer	2019	Lakme Flawless White Tint	http://assets.myntassets.o				








Ordenar por año de forma descendente y limitar la salida a las 50 segundas filas







```
1 db.styles.find({}).sort({ year: -1}).skip(50).limit(50);
```

styles

0.036 s

50 Docs



	_id	id	gender	masterCate	subCategory	articleType	baseColour	season	year	usage	productDisplayName	image_url
2	 67161ccca	2390	Men	Footwear	Shoes	Casual Shoes	Black	Winter	2018	Casual	Red Tape Men's Semi Blac	http://assets.myntassets.co
3	 67161ccca	28318	Men	Footwear	Shoes	Casual Shoes	Brown	Winter	2018	Casual	Red Tape Men Brown Shoe	http://assets.myntassets.co
4	 67161ccca	39255	Men	Footwear	Shoes	Formal Shoes	Black	Winter	2018	Formal	Red Chief Men Black Form	http://assets.myntassets.co
5	 67161ccca	2397	Men	Footwear	Shoes	Casual Shoes	Brown	Winter	2018	Casual	Red Tape Men's Brown Cas	http://assets.myntassets.co
6	 67161ccca	39290	Men	Footwear	Shoes	Casual Shoes	Brown	Winter	2018	Casual	Red Chief Men Brown Sho	http://assets.myntassets.co
7	 67161ccca	50668	Men	Apparel	Loungewear	Lounge Pants	Navy Blue	Fall	2018	Sports	Chromosome Men Navy Bli	http://assets.myntassets.co

Descartar productos que sean de años mayores de 2015 y menores de 2014 (\$nor)

```
1 db.styles.find({ $nor: [ { "year" : { $lt : 2014 } }, { "year" : { $gt : 2015 } } ] });
```

styles

0.026 s

3017 Docs

	_id	id	gender	masterCategory	subCategory	articleType	baseColour	season	year	usage	productDisplayName	image_url
2		67161ccca39988	Men	Footwear	Shoes	Casual Shoes	White	Winter	2014	Casual	Gas Men Europa White Sho	http://assets.m
3		67161ccca2886	Women	Footwear	Shoes	Flats	Brown	Winter	2015	Casual	Catwalk Women Leather B	http://assets.m
4		67161ccca2872	Women	Footwear	Shoes	Heels	Brown	Winter	2015	Casual	Catwalk Women Leather F	http://assets.m
5		67161ccca8746	Unisex	Footwear	Flip Flops	Flip Flops	Navy Blue	Summer	2015	Casual	Crocs Kids Navy Blue Clog	http://assets.m
6		67161ccca11518	Women	Footwear	Shoes	Heels	Grey	Winter	2015	Casual	Catwalk Women Gun Metal	http://assets.m
7		67161ccca30854	Women	Apparel	Blouses	Blouses	Red	Winter	2015	Casual	Salon Art Women Blouse	http://assets.m

2.3. Ejercicios sobre Pipeline de Agregación

Mostrar los 10 tipos de producto que tienen más variedad



```
1 db.styles.aggregate([
2   { $group: { _id: "$articleType", totalProductos: { $sum: 1 } } },
3   { $sort: { totalProductos: -1 } },
4   { $limit: 10 }
5 ]);
```

	_id *	totalProductos
1	Tshirts	7070
2	Shirts	3217
3	Casual Shoes	2846
4	Watches	2542
5	Sports Shoes	2036
6	Kurtas	1844
7	Tops	1762
8	Handbags	1759
9	Heels	1323
10	Sunglasses	1073

Mostrar las 15 combinaciones baseColour-articleType predominantes y las menos comunes

```
1 db.styles.aggregate([
2   { $group: { _id: { subCategory: "$articleType", baseColour: "$baseColour" }, totalProductos
3     : { $sum: 1 } } },
4   { $sort: { totalProductos: -1 } },
5   { $limit: 15 }
6 ]);
7 db.styles.aggregate([
8   { $group: { _id: { subCategory: "$articleType", baseColour: "$baseColour" }, totalProductos
9     : { $sum: 1 } } },
10  { $sort: { totalProductos: 1 } },
11  { $limit: 15 }
12 ]);
```

	styles	0.043 s	15 Docs
	_id		totalProductos
	subCategory	baseColour	
1	Tshirts	White	1119
2	Tshirts	Black	1049
3	Tshirts	Blue	1046
4	Watches	Black	987
5	Sports Shoes	White	898
6	Casual Shoes	Black	875
7	Shirts	Blue	785
8	Tshirts	Grey	628
9	Tshirts	Red	602
10	Tshirts	Green	591
11	Watches	White	579
12	Sports Shoes	Black	554
13	Shirts	White	526
14	Casual Shoes	Brown	511
15	Jeans	Blue	446

	styles	 0.043 s	15 Docs
	_id		
	subCategory	baseColour	totalProductos
1	Jewellery Set	Turquoise Blue	1
2	Kajal and Eyeliner	Lavender	1
3	Lounge Shorts	Charcoal	1
4	Shorts	Off White	1
5	Sports Shoes	Khaki	1
6	Perfume and Body Mist	Olive	1
7	Earrings	Olive	1
8	Free Gifts	Steel	1
9	Capris	Brown	1
10	Lip Gloss	Gold	1
11	Jeans	Brown	1
12	Blazers	Charcoal	1
13	Travel Accessory	Red	1
14	Shapewear	Pink	1
15	Face Wash and Cleanser	Red	1

Mostrar el año u años con más y con menos productos

```

1 var fase1 = { $group: { "_id": "$year", "cuenta": { $sum: 1 } } };
2 var fase2 = { $group: { "_id": "$cuenta", "años": { $push: "$_id" } } };
3 var sortMax = { $sort: { "_id": -1 } }; // Para encontrar el máximo número de productos
4 var sortMin = { $sort: { "_id": 1 } }; // Para encontrar el mínimo número de productos
5 var limit = { $limit: 1 };
6 var unwind = { $unwind: "$años" };
7
8 var pipelineMax = [ fase1, fase2, sortMax, limit, unwind ];
9 var pipelineMin = [ fase1, fase2, sortMin, limit, unwind ];
10
11 db.styles.aggregate([
12   {
13     $facet: {
14       "maxYear": pipelineMax,
15       "minYear": pipelineMin
16     }
17   }
18 ]);

```

Nos sale como año 'null' porque hay un producto que no le viene asignado ningún año, entonces vamos a añadir un match al principio para evitar esto

```

1 var matchValidYear = { $match: { year: { $ne: null } } };
2 var fase1 = { $group: { "_id": "$year", "cuenta": { $sum: 1 } } };
3 var fase2 = { $group: { "_id": "$cuenta", "años": { $push: "$_id" } } };
4 var sortMax = { $sort: { "_id": -1 } };
5 var sortMin = { $sort: { "_id": 1 } };
6 var limit = { $limit: 1 };
7 var unwind = { $unwind: "$años" };
8
9 var pipelineMax = [ matchValidYear, fase1, fase2, sortMax, limit, unwind ];
10 var pipelineMin = [ matchValidYear, fase1, fase2, sortMin, limit, unwind ];
11
12 db.styles.aggregate([
13   {
14     $facet: {
15       "maxYear": pipelineMax,
16       "minYear": pipelineMin
17     }
18   }
19 ]);

```

styles	0.104 s	1 Doc
1	{	
2	"maxYear" : [
3	{	
4	"_id" : 16290,	
5	"años" : 2012	
6	}	
7],	
8	"minYear" : [
9	{	
10	"_id" : 1,	
11	"años" : null	
12	}	
13]	
14	}	

styles	0.163 s	1 Doc
1	{	
2	"maxYear" : [
3	{	
4	"_id" : 16290,	
5	"años" : 2012	
6	}	
7],	
8	"minYear" : [
9	{	
10	"_id" : 2,	
11	"años" : 2007	
12	}	
13]	
14	}	

Agrupar por temporada y contar cuántos productos se lanzaron en cada una, ordenando de forma descendente

```
1 db.styles.aggregate( [
2   { $group: { "_id": "$season", "cuenta": { $sum: 1 } } },
3   { $sort: { "cuenta" : -1 } } ] );
```

	styles	0.036 s	5 Docs
	_id *	cuenta	
1	Summer	21476	
2	Fall	11445	
3	Winter	8519	
4	Spring	2985	
5		21	

Mostrar los 5 tipos de artículo de categoría Footwear de los que hay más productos

```
1 var fasefilter={ $match: { "masterCategory": "Footwear" } }
2 var fasegroup={ $group: { "_id": "$articleType", "cuenta": { $sum: 1 } } }
3 var fasesort={ $sort: { "cuenta": -1 } }
4 var faselimit={ $limit: 5 }
5 var etapas=[ fasefilter, fasegroup, fasesort, faselimit ]
6 db.styles.aggregate(etapas);
```

	styles	0.036 s	5 Docs
	_id *	cuenta	
1	Casual Shoes	2846	
2	Sports Shoes	2036	
3	Heels	1323	
4	Flip Flops	916	
5	Sandals	897	

Mostramos las subcategorías junto con el número de productos que tiene cada subcategoría y la información del producto más reciente en esa subcategoría.

```
1 db.styles.aggregate([
2   { $sort: { "year": -1 } },
3   { $group: {
4     _id: "$subCategory",
5     totalProductos: { $sum: 1 },
6     mostRecentProduct: { $first: "$$ROOT" }
7   } },
8   { $sort: { "_id": 1 } },
9   { $project: {
10    _id: 1,
11    totalProductos: 1,
12    "mostRecentProduct.productDisplayName": 1,
13    "mostRecentProduct.image_url": 1,
14    "mostRecentProduct.year": 1
15  } }
16 ]);
```

Quiero apuntar que al no haber mes, día ni hora de publicación van a existir muchos productos de cada subcategoría publicados en el último año de publicación de esta misma, cogeremos simbólicamente el primero de ellos.

styles 0.094 s 45 Docs					
	_id *	totalProducts	year	productDisplayName	mostRecentProduct image_url
34	Saree	427	2018	Xoxo Women Maroon Pure Georgette Solid Ready-to	http://assets.myntassets.com/assets/ima
35	Sandal	963	2018	Red Chief Men Khaki Sandals	http://assets.myntassets.com/v1/images/
36	Loungewear and N	470	2018	Jockey Women Charcoal Grey Lounge Pants	http://assets.myntassets.com/assets/ima
37	Innerwear	1808	2018	Amante Women Pack of 2 Bikini Briefs PFFL03	http://assets.myntassets.com/v1/image/st
38	Flip Flops	915	2018	Red Tape Men's Black Flip Flop	http://assets.myntassets.com/v1/images/
39	Bottomwear	2694	2018	Jockey Women Grey Melange Knit Capris 1300	http://assets.myntassets.com/assets/ima
40	Bags	3055	2018	Wildcraft Unisex Grey & Green Rucksack	http://assets.myntassets.com/v1/images/
41	Nails	329	2019	Lakme Truewear Color Crush 04 Nail Polish	http://assets.myntassets.com/assets/ima
42	Makeup	307	2019	Lakme Nine to Five Flawless Makeup Shell Foundatic	http://assets.myntassets.com/assets/ima
43	Lips	527	2019	Colorbar Velvet Matte Hot Hot Hot Red Lipstick 01R	http://assets.myntassets.com/assets/ima
44	Eyes	43	2019	Colorbar Women Lash Illusion Mascara Duo DM001	http://assets.myntassets.com/assets/ima
45	Bath and Body	12	2019	Colorbar Acetone Free Nail Enamel Remover	http://assets.myntassets.com/assets/ima

Agrupar los productos por género y color base, mostrando los 5 colores más populares para cada género.

```
1 db.styles.aggregate([
2   { $group: {
3     _id: { gender: "$gender", baseColour: "$baseColour" },
4     totalProductos: { $sum: 1 }
5   }
6 },
7 { $sort: { totalProductos: -1 } },
8 { $group: {
9   _id: "$_id.gender",
10  topColours: { $push: { colour: "$_id.baseColour", count: "$totalProductos" } }
11 }
12 },
13 { $project: { _id: 0, gender: "$_id", topColours: { $slice: ["$topColours", 5] } } }
14 ]);
```

```
1 /* 1 */
2 {
3   "gender": "Boys",
4   "topColours": [
5     {
6       "colour": "Blue",
7       "count": 159
8     },
9     {
10      "colour": "White",
11      "count": 157
12    },
13    {
14      "colour": "Red",
15      "count": 167
16    },
17    {
18      "colour": "Yellow",
19      "count": 90
20    },
21    {
22      "colour": "Black",
23      "count": 83
24    }
25  ],
26 },
27
28 /* 2 */
29 {
30   "gender": "Girls",
31   "topColours": [
32     {
33       "colour": "Pink",
34       "count": 167
35     },
36     {
37       "colour": "White",
38       "count": 109
39     },
40     {
41       "colour": "Blue",
42       "count": 182
43     },
44     {
45       "colour": "Green",
46       "count": 43
47     },
48     {
49       "colour": "Black",
50       "count": 41
51     }
52   ],
53 },
54
55 /* 3 */
56 {
57   "gender": "Men",
58   "topColours": [
59     {
60       "colour": "Black",
61       "count": 5885
62     },
63     {
64       "colour": "White",
65       "count": 3193
66     },
67     {
68       "colour": "Blue",
69       "count": 2688
70     },
71     {
72       "colour": "Brown",
73       "count": 1951
74     },
75     {
76       "colour": "Grey",
77       "count": 1810
78     }
79   ],
80 },
81
82 /* 4 */
83 {
84   "gender": "Unisex",
85   "topColours": [
86     {
87       "colour": "Black",
88       "count": 774
89     },
90     {
91       "colour": "Blue",
92       "count": 263
93     },
94     {
95       "colour": "Red",
96       "count": 192
97     },
98     {
99       "colour": "Grey",
100      "count": 143
101    },
102    {
103      "colour": "White",
104      "count": 142
105    }
106  ],
107 },
108
109 /* 5 */
110 {
111   "gender": "Women",
112   "topColours": [
113     {
114       "colour": "Black",
115       "count": 2949
116     },
117     {
118       "colour": "White",
119       "count": 2629
120     },
121     {
122       "colour": "Blue",
123       "count": 1718
124     },
125     {
126       "colour": "Pink",
127       "count": 1488
128     },
129     {
130       "colour": "Brown",
131       "count": 1408
132     }
133   ],
134 }
```

Agrupar los productos por año y mostrar el número total de productos lanzados por cada año.

```
1 db.styles.aggregate([
2   { $group: {
3     _id: "$year",
4     totalProductos: { $sum: 1 }
5   }
6 },
7 { $sort: { _id: 1 } },
8 { $project: { year: "$_id", totalProductos: 1, _id: 0 } }
9 ]);
```

Mostrar cómo se distribuyen los productos lanzados por año y estación

```
1 db.styles.aggregate([
2   { $group: {
3     _id: { year: "$year", season: "$season" },
4     totalProductos: { $sum: 1 }
5   }
6 },
7 { $sort: { "_id.year": 1, "_id.season": 1 } },
8 { $project: {
9   year: "$_id.year",
10  season: "$_id.season",
11  totalProductos: 1,
12  _id: 0
13 }
14 }
15 ]);
```

	totalProductos	year
1	1	null
2	2	2007
3	7	2008
4	20	2009
5	846	2010
6	13689	2011
7	16290	2012
8	1213	2013
9	236	2014
10	2780	2015
11	6007	2016
12	2917	2017
13	405	2018
14	33	2019

	totalProductos	year	season
2	2	2007	Spring
3	1	2008	Fall
4	3	2008	Spring
5	2	2008	Summer
6	1	2008	Winter
7	19	2009	Fall
8	1	2009	Summer
9	640	2010	Fall
10	3	2010	Spring
11	133	2010	Summer
12	70	2010	Winter

Mostrar el número de productos por género y las subcategorías más representativas de cada uno

```
1 db.styles.aggregate([
2   { $match: { gender: { $in: ["Men", "Women"] } } },
3   { $group: {
4     _id: { gender: "$gender", subCategory: "$subCategory" },
5     totalProductos: { $sum: 1 }
6   }
7 },
8 { $sort: { totalProductos: -1 } },
9 { $group: {
10  _id: "$_id.gender",
11  subCategories: {
12    $push: {
13      subCategory: "$_id.subCategory",
14      totalProductos: "$totalProductos"
15    }
16  },
17  totalProductos: { $sum: "$totalProductos" }
18 }
19 ]);
```

```

18   }
19 },
20 { $project: {
21   _id: 1,
22   totalProductos: 1,
23   firstSubCategory: { $arrayElemAt: ["$subCategories.subCategory", 0] },
24   secondSubCategory: { $arrayElemAt: ["$subCategories.subCategory", 1] },
25   thirdSubCategory: { $arrayElemAt: ["$subCategories.subCategory", 2] },
26 }
27 },
28 { $addFields: {
29   description: {
30     $concat: [
31       "En la categoría ", "$_id",
32       ", la subcategoría más popular es ", "$firstSubCategory",
33       ", seguida de ", "$secondSubCategory",
34       " y ", "$thirdSubCategory", "."
35     ]
36   }
37 }
38 },
39 { $project: { _id: 1, totalProductos: 1, description: 1 } }
40 ]);

```

	styles	0.045 s	2 Docs
	_id *	totalProductos	description
1	Women	18632	En la categoría Women, la subcategoría más popular es Topwear, seguida de Shoes y Bags.
2	Men	22165	En la categoría Men, la subcategoría más popular es Topwear, seguida de Shoes y Watches.

Mostrar los productos de la colección Primavera-Verano del año 2008 para destacar las imágenes de estos mismos.

```

1 db.styles.aggregate([
2   { $match: { season: { $in: ["Spring", "Summer"] } , year: 2008 } },
3   { $project: {
4     productDisplayName: 1,
5     image_url: 1,
6     season: 1
7   }}
8 ]);

```

styles

0.042 s

5 Docs

	<div><div>_id *</div><div></div></div>	<div><div>season</div><div></div></div>	<div><div>productDisplayName</div><div></div></div>	<div><div>image_url</div><div></div></div>
1	<div><div></div><div>67161ccea0ef8f406584946f</div></div>	Summer	Raymond Men Black & B	http://assets.myntassets.com/v1/images/style/property
2	<div><div></div><div>67161ccea0ef8f406584bc00</div></div>	Spring	U.S. Polo Assn. Kids Boy	http://assets.myntassets.com/v1/images/style/property
3	<div><div></div><div>67161ccfa0ef8f406584d474</div></div>	Summer	Reebok Men Realflex Op	http://assets.myntassets.com/v1/images/style/property
4	<div><div></div><div>67161ccfa0ef8f406584d609</div></div>	Spring	test dispName	http://assets.myntassets.com/v1/images/style/property
5	<div><div></div><div>67161ccfa0ef8f406584db85</div></div>	Spring	ADIDAS Men White Sukc	http://assets.myntassets.com/v1/images/style/property



Adjunto por último aquí el código de python (Print_images.py) del que he hecho uso para imprimir las imágenes de la última consulta.

```
1 from PIL import Image
2 import requests
3 from io import BytesIO
4 import matplotlib.pyplot as plt
5
6 # Lista de URLs de las imágenes
7 urls = [
8     "http://assets.myntassets.com/v1/images/style/properties/Raymond-Men-Raymond-Leather-
9     Belts-White-Belts_f942698ad78d2e13a4007024e6c73f3c_images.jpg",
10    "http://assets.myntassets.com/v1/images/style/properties/US-Polo-Assn-Kids-Boys-Orange-T-
11    shirt_a7c1b583f1c04c05f9bd6ffca5358241_images.jpg",
12    "http://assets.myntassets.com/v1/images/style/properties/Reebok-Men-REALFLEX-OPTIMAL-
13    White-Sports-Shoes_6e49515f918bc0fb66df906799386112_images.jpg",
14    "http://assets.myntassets.com/v1/images/style/properties/Reid---Taylor-Men-Blue-Striped-
15    Shirt_b49d226c8a40fc2049dfb6280149e4ff_images.jpg",
16    "http://assets.myntassets.com/v1/images/style/properties/Adidas-Men-SUKOI-White-Sports-
17    Shoes_2d184f92b877bb5f06e2ba7c725590f7_images.jpg"
18 ]
19
20 # Descargar y mostrar cada imagen
21 images = []
22 for url in urls:
23     response = requests.get(url)
24     img = Image.open(BytesIO(response.content))
25     images.append(img)
26
27 # Crear una cuadrícula de 1x5 para mostrar las imágenes
28 fig, axes = plt.subplots(1, 5, figsize=(15, 5))
29 for ax, img in zip(axes, images):
30     ax.imshow(img)
31     ax.axis('off') # Quitar los ejes para una mejor visualización
32
33 plt.show()
```

3

Conclusiones

En este análisis del catálogo de productos de moda, se observó que las categorías de productos con mayor diversidad incluyen tipos como "Tshirts", "Shirts", "Casual Shoes" y "Watches", lo que sugiere que estos artículos son clave en la estrategia de la marca debido a su alta variedad. Al estudiar las combinaciones de color y tipo de artículo, se identificaron combinaciones predominantes como "White Tshirts", "Black Tshirts", y "Black Watches", destacando preferencias claras de color y tipo en la oferta. Mientras que en las combinaciones menos frecuentes podemos observar que se trata de colores y de tipos de artículo menos básicos, como "Turquoise Blue Jewellery Set", "Lavender Kajal and Eyeliner", y "Charcoal Lounge Shorts". En cuanto a la estacionalidad, cabe destacar la predominancia en cuanto a la publicación de artículos de las colecciones Verano y Primavera frente a Invierno y Otoño.

El análisis de los años con mayor y menor número de lanzamientos reveló que los años con más productos fueron 2011 y 2012, mientras que tanto años anteriores como posteriores, tuvieron un menor volumen de lanzamientos. Además, la consulta que agrupó productos por género y color base mostró que para las categorías "Men" y "Women", que son las más pobladas, los colores predominantes incluían el negro, el blanco, el azul y el marrón, reflejando una muy poco representativa segmentación de color basada en preferencias de género. Por último, las subcategorías más representativas para estos mismos géneros ("Men" y "Women") fueron "Topwear", "Shoes" y "Watches" para hombres, y "Topwear", "Shoes" y "Bags" para mujeres, lo que subraya las prioridades de inventario y la oferta enfocada en los productos más buscados por cada segmento.

Para enriquecer el análisis del catálogo de productos de moda y abordar cuestiones más profundas y personalizadas, propongo la inclusión de varios campos adicionales en el dataset. Uno de los campos que sugiero es el precio del producto, que permitiría explorar cómo el costo influye en la popularidad, identificar rangos de precios y evaluar la accesibilidad o exclusividad de ciertos artículos. Además, un campo de fecha de añadido posibilitaría un análisis temporal detallado (por día, mes o trimestre) y facilitaría la detección de patrones de lanzamiento en fechas clave, como festividades o temporadas altas. Incluir tamaño también sería útil, ya que permitiría identificar preferencias de talla en cada categoría, ajustar inventarios según demanda y analizar la rotación de productos con mayor precisión.

Para profundizar en la percepción del cliente, sugiero agregar una valoración promedio de usuarios junto con el número de reseñas asociadas daría una visión integral de la satisfacción del cliente y popularidad del producto, permitiendo segmentar por productos de alta o baja valoración. Asimismo, incluir un nivel de inventario que permitiría monitorear las existencias en tiempo real, identificar productos de alta demanda y planificar mejor los reabastecimientos. En el ámbito digital, un campo de popularidad en redes sociales ofrecería una perspectiva sobre la presencia en redes de cada producto, permitiendo correlacionar la popularidad en estas plataformas con las ventas y destacando productos con potencial viral.