



Análisis del Catálogo de Sephora y Propuestas de Lanzamiento para Sephora Collection

TRABAJO FINAL: BOOTCAMP DE ALTO RENDIMIENTO EN DATA ANALYTICS

RAQUEL SÁNCHEZ CABEZA DE VACA

Índice

1.	Definición del Proyecto	2
1.1.	Descripción clara del caso de negocio	2
1.2.	Objetivos del proyecto y su impacto esperado	2
1.3.	Tecnologías y herramientas utilizadas	3
1.4.	Identificación y justificación de las fuentes de datos	3
2.	Proceso ETL	4
2.1.	Descripción del proceso	4
2.1.1.	Extracción	4
2.1.2.	Transformación	4
2.1.3.	Carga	4
2.2.	Esquema de la base de datos	4
2.3.	Problemas encontrados y soluciones	5
3.	Resultados del EDA	6
3.1.	Visión general del catálogo	6
3.1.1.	Posicionamiento de Sephora Collection en la tienda	6
3.1.2.	Éxito de Productos	7
3.1.3.	Análisis de Precios y Variaciones	7
3.1.4.	Outliers y Observaciones Relevantes	7
3.2.	Análisis Específico de Sephora Collection	8
3.3.	Limitaciones y Recomendaciones Futuras	8
3.4.	Análisis y Recomendaciones para Nuevos Lanzamientos de Sephora Collection	9
3.4.1.	Contexto general de la marca	9
3.4.2.	Lanzamiento Seguro: Brocha de Rostro	9
3.4.3.	Lanzamiento Arriesgado: Crema BB & CC	9
3.4.4.	Lanzamiento de Mejora: Sombras de Ojos	10
4.	Dashboard final	11
4.1.	Estética y diseño visual	11
4.2.	Estructura general del dashboard	11
5.	Next steps	13

1. Definición del Proyecto

Este proyecto se ha desarrollado en tres fases principales: una primera etapa de extracción y transformación de datos (**ETL**), seguida de un análisis exploratorio en profundidad (**EDA**), y culminando con el diseño de un **dashboard** interactivo. El punto de partida fue la definición del caso de negocio, al que le siguieron las tres fases mencionadas, estructuradas con un enfoque analítico y estratégico.

1.1. Descripción clara del caso de negocio

En este proyecto me he situado en el rol de analista de datos para Sephora, con el objetivo de apoyar la toma de decisiones estratégicas relacionadas con el **lanzamiento de nuevos productos de maquillaje bajo su propia marca, Sephora Collection**.

El enfoque inicial planteaba simplemente identificar una única oportunidad de lanzamiento, pero a medida que avanzó el análisis, se consolidó una estrategia más rica, basada en tres líneas diferenciadas de propuesta:

- Un lanzamiento seguro, basado en tendencias claras y patrones de éxito validados por datos.
- Un lanzamiento arriesgado, que apunta a innovar en un nicho aún poco explorado, pero con potencial de crecimiento.
- Y un lanzamiento de mejora, enfocado en cubrir carencias o reforzar categorías donde Sephora Collection ya compite, pero con margen de optimización.

1.2. Objetivos del proyecto y su impacto esperado

El objetivo principal ha sido definir, basándonos en datos reales, cuáles podrían ser los próximos lanzamientos de productos de maquillaje bajo la marca Sephora Collection.

Para alcanzar este propósito general, se desarrollaron una serie de objetivos técnicos y analíticos específicos:

- **Desarrollar un scraper automatizado** para extraer información detallada sobre los productos de maquillaje disponibles en la web de Sephora.
- **Diseñar e implementar una base de datos relacional en PostgreSQL**, con una estructura funcional que permitiera almacenar los datos de forma eficiente.
- **Automatizar el proceso de actualización de la base de datos**, asegurando que esta se mantuviera al día con nuevos productos y cambios en los ya existentes.
- **Conservar un histórico de datos clave**, especialmente en lo referente a precios y valoraciones, con el fin de identificar tendencias y cambios a lo largo del tiempo.
- **Llevar a cabo un análisis exploratorio exhaustivo con Python**, que permitiera comprender el comportamiento del mercado y detectar patrones relevantes: categorías con menor presencia de Sephora Collection, características comunes en productos bien valorados, precios de referencia por tipo de producto, etc.
- **Identificar oportunidades reales de lanzamiento**, basadas en datos objetivos, y estructuradas en las tres líneas estratégicas ya mencionadas.
- **Diseñar un dashboard interactivo en Power BI** para comunicar de forma visual, clara y accesible todos los hallazgos, conclusiones y recomendaciones del análisis.

1.3. Tecnologías y herramientas utilizadas

Este proyecto ha sido desarrollado en Python 3.13.0 y utiliza diversas herramientas para llevar a cabo el scraping, procesamiento, análisis y visualización de los datos. A continuación, se detallan los componentes clave:

Tecnologías y herramientas:

- **Python:** Lenguaje principal para la extracción, transformación, análisis y conexión a la base de datos.
- **PostgreSQL:** Base de datos relacional utilizada para almacenar los datos estructurados.
- **DBeaver:** Cliente visual para la administración de la base de datos.
- **Power BI:** Herramienta de visualización utilizada para crear el dashboard final.

Librerías principales:

- **Web Scraping:** selenium, webdriver-manager
- **Procesamiento de datos:** pandas, numpy
- **Análisis y visualización:** plotly
- **Base de datos:** psycopg2, python-dotenv
- **Otras utilidades:** datetime, os, time, re, json

1.4. Identificación y justificación de las fuentes de datos

En este caso no ha sido necesario buscar datasets externos porque la fuente de datos es clara: la propia **página web de Sephora**. Desde allí se ha podido acceder a la información de todos los productos de maquillaje, incluyendo datos básicos como nombre, marca o precio, y otros más específicos como el formato del producto, la cobertura, el tipo de piel para el que está recomendado, etc.

Gracias a que la web es accesible y estructurada, se ha podido scrapear sin problemas y se puede automatizar para mantener la base de datos actualizada. Esto además permite tener un histórico de cambios que resulta clave para detectar variaciones de precio, evolución de las valoraciones o aparición de nuevos productos, lo cual enriquece mucho más el análisis final.

2. Proceso ETL

Se ha partido de la obtención de los datos directamente desde la página web, pasando por una mínima transformación necesaria para mantener la calidad del dato, hasta su almacenamiento en una base de datos relacional que ha permitido su posterior análisis.

2.1. Descripción del proceso

2.1.1. Extracción

La fase de extracción se ha llevado a cabo mediante técnicas de **web scraping** utilizando **Python**. El proceso principal consiste en: Obtener las URLs de todos los productos de maquillaje disponibles en la web de Sephora España e ir iterando por cada URL individual para extraer información como: nombre del producto, marca, precio, número de valoraciones, valoración media, descripción, número de variaciones del producto. Además, se ha implementado un sistema para extraer la información de los filtros (como cobertura, texturas, acabados, formulaciones, etc.) mediante una lógica alternativa: se accede a las URLs correspondientes a cada subcategoría dentro de cada filtro (por ejemplo, “cobertura ligera”, “cobertura media”, etc.), se obtienen los productos que aparecen en cada uno, y se comparan las URLs con las previamente extraídas. Esta lógica permite determinar de forma indirecta a qué categorías pertenece cada producto.

El proceso incluye también operaciones como scroll automático en la página, manejo de excepciones y verificación de estructuras HTML, lo cual ha requerido un trabajo técnico importante dada la variabilidad del contenido web.

2.1.2. Transformación

En este caso, la **transformación** ha sido **mínima** y se ha incorporado directamente **durante la fase de extracción**. Las únicas transformaciones realizadas han sido: conversión y limpieza del tipo de dato de ciertas columnas (por ejemplo, convertir valores de precio a float, valoraciones a int, etc.), y normalización de valores textuales en lo posible.

No se han imputado nulos ni realizado transformaciones adicionales ya que, tras análisis, los valores ausentes eran representaciones correctas del dato (por ejemplo, ausencia de cierto filtro en un producto).

2.1.3. Carga

La carga se ha diseñado cuidadosamente para permitir un proceso de **actualización incremental**. Se ha implementado una lógica que: en primer lugar, verifica si un producto ya existe en la base de datos (identificado de forma única mediante la URL del producto). Si el producto no existe, se inserta junto con sus atributos asociados, incluyendo sus relaciones con marcas, categorías, subcategorías y filtros (a través de tablas intermedias). Si el producto ya existe, se registra únicamente una nueva entrada en la tabla histórico con los valores que puedan variar en el tiempo (precio, número de valoraciones, etc.).

De esta forma, se podrá realizar un análisis temporal de la evolución de cada producto cuando se disponga de un número considerable de scrapeos.

2.2. Esquema de la base de datos

La base de datos relacional diseñada sigue una **estructura** altamente **normalizada**. Se destacan la tabla de productos (tabla principal que contiene los productos). Las tablas auxiliares como marcas, categorías y subcategorías con claves foráneas en productos. La tabla de histórico que permite guardar distintas extracciones temporales de cada producto para seguimiento de cambios. Y, por

último, las tablas de filtros, donde para cada filtro (acabado, textura, cobertura, formulación, etc.) se ha creado una tabla de categorías y otra tabla intermedia que vincula productos con sus correspondientes atributos múltiples. Esto permite manejar relaciones muchos a muchos.

El uso de claves foráneas y la política de ON DELETE CASCADE permite mantener la **integridad referencial**.

2.3. Problemas encontrados y soluciones

Durante el desarrollo del proyecto, se han enfrentado diversos desafíos técnicos. A continuación, se detallan algunos de ellos:

- **Desafíos propios del scraping:** Implementación del scroll automático y manejo dinámico del contenido web, dificultades para localizar ciertos datos dentro del HTML debido a clases variables o estructura inconsistente, ajustes necesarios para detectar pequeñas excepciones en la estructura HTML (por ejemplo, precios en clases distintas para algunos productos).
- **Problemas descartados:** Se intentó extraer la información de los ingredientes de cada producto, pero resultó inviable debido a la enorme variabilidad en el formato de presentación (comas, bullet points, texto entre ingredientes, etc.). Se concluyó que requeriría un tratamiento de NLP o reglas muy complejas para ser abordado adecuadamente.
- **Problemas con identificadores:** Inicialmente se usó el nombre del producto como identificador, pero se detectaron productos con nombres idénticos. Se corrigió utilizando la URL del producto como identificador único.
- **Errores en la carga del histórico:** Se detectó que, tras el primer scraping completo, aparecían 7 duplicados en la tabla histórico, lo cual no debería ocurrir en la primera ejecución. El error no ha podido ser identificado aún, por lo que se deja como pendiente para una próxima revisión.

3. Resultados del EDA

3.1. Visión general del catálogo

	Sephora	Sephora Collection
Número productos	1822	198
Número marcas	79	-
Número categorías	13	11 (sin presencia en Papeles matificantes ni Pestañas postizas).
Número subcategorías	48	40 (ausente en varias relacionadas con cuidados y paletas específicas).
Precio mediano	34,99	13,99
Número mediano valoraciones	42,5	26
Valoración mediana	4,4	4,4
Número mediano variaciones	2	0
Correlación (precio / éxito)	0,05	0,31

*Éxito: Productos con una valoración mediana mayor o igual que 4,5 y un número mediano de valoraciones mayor o igual a 50

Estos límites se definieron a partir del análisis de la distribución general de los datos, teniendo en cuenta que:

- Una valoración de 4.5 supera la mediana (4.4) y se aproxima al cuartil superior sin restringirse a outliers.
- Un mínimo de 50 valoraciones garantiza cierta base de usuarios, superando levemente la mediana (~42) sin volverse demasiado restrictivo.

Este enfoque busca equilibrar representatividad y calidad, capturando productos con buena aceptación sin limitarse a los casos más extremos.

Además, en el análisis se optó por utilizar la mediana como medida de tendencia central, en lugar de la media, debido a la presencia de valores atípicos que podrían distorsionar los resultados. La mediana proporciona una visión más robusta y precisa del comportamiento típico de los productos, lo cual es clave para tomar decisiones más informadas en contextos con distribuciones asimétricas o dispersas.

3.1.1. Posicionamiento de Sephora Collection en la tienda

Sephora Collection es la marca con **mayor presencia** en el catálogo, contando con un total de 198 productos, lo que representa el 10,87% del total disponible. Esta marca es la segunda más **económica**, con un precio mediano de 13,99 €. Su dominio se evidencia en categorías clave como estuches de maquillaje, labios, ojos, así como pinceles y accesorios. Dentro de estas categorías, lidera en varias subcategorías importantes, incluyendo brochas de rostro, productos de contouring y máscara de pestañas. En general, los productos de Sephora Collection presentan una **escasa variedad de variaciones por producto**, con una mediana de cero, excepto en las categorías de labios, rostro y cejas, donde la variedad es algo mayor.

En general, **el precio no muestra una relación significativa con el éxito** de los productos en Sephora, ya que la correlación observada es prácticamente nula. Esto sugiere que el precio, por sí solo, no es un factor determinante del éxito.

Sin embargo, al analizar específicamente la línea Sephora Collection, se encuentra una correlación positiva ($r = 0.31$) entre precio y éxito. Aunque esta relación sigue siendo débil, podría indicar que, dentro de una marca conocida por su enfoque económico, los productos ligeramente más caros son percibidos como de mayor calidad o valor diferencial, lo que puede influir positivamente en su recepción.

Aun así, dado que la correlación es moderada y no muy fuerte, se puede concluir que el precio no es el principal factor de éxito, y que otros elementos como la calidad percibida, la reputación, las reseñas o el marketing probablemente tienen un mayor peso en la valoración final por parte de los consumidores.

3.1.2. Éxito de Productos

Los productos considerados exitosos representan el **21,57%** del total del catálogo. En términos estratégicos, marcas como Natasha Denona, Huda Beauty y Anastasia Beverly Hills destacan por su alta eficiencia y volumen de productos exitosos. Por otro lado, marcas nicho como Glow Recipe y Pai muestran una alta eficiencia, aunque con un volumen reducido. Sephora Collection cuenta con un **índice de éxito interno del 16%**, lo que refleja una alta relevancia dada su gran volumen.

Analizando por categorías, la sección de rostro se destaca como la de mayor volumen y eficiencia interna, alcanzando un 24,24%, por lo que es considerada una categoría clave. Las categorías de labios y ojos presentan un alto volumen de productos, aunque su eficiencia es moderada, con índices de éxito del 20,12% y 16,94% respectivamente, evidenciando una mayor dispersión en calidad. Las paletas de maquillaje, aunque con menor volumen, resultan ser las más eficientes proporcionalmente, con un 40,82%. Asimismo, pinceles y accesorios, junto con la categoría de cejas, combinan un buen volumen con una eficiencia destacada, posicionándose como categorías nicho efectivas.

En cuanto a subcategorías, colorete y paletas de sombras destacan por su alto volumen y eficiencia, ambas superiores al 38%. Las brochas de rostro y sombras de ojos tienen una participación considerable, con eficiencias moderadas alrededor del 29-30%. Las cremas BB & CC y las paletas de rostro presentan altísimos índices de eficiencia, superiores al 45%, aunque con catálogos más pequeños. Finalmente, los estuches de pinceles muestran un buen rendimiento, con un 36,36% de eficiencia, validando la relevancia de estos nichos específicos.

3.1.3. Análisis de Precios y Variaciones

Las categorías con los precios medianos más altos corresponden a paletas de maquillaje, rostro y ojos, mientras que las más económicas incluyen rizadores de pestañas, productos para uñas y sacapuntas. Las subcategorías con precios más elevados son paletas de rostro, sombras y productos de contouring, mientras que las más accesibles son esmaltes de uñas, quitaesmaltes y limpiadores de pinceles. En este contexto, Sephora Collection se destaca por mantener una **política de precios competitivos** y un **bajo número de variaciones por producto**, lo que simplifica su oferta y potencia su accesibilidad.

3.1.4. Outliers y Observaciones Relevantes

Los productos con los precios más altos se concentran principalmente en las marcas Sisley y en bases de maquillaje, mientras que La Mer también presenta una mediana alta en precios generales. Las marcas Benefit y la categoría de máscara de pestañas cuentan con los productos que acumulan

mayor número de valoraciones, destacando también algunos productos de Anastasia Beverly Hills y NARS. Por otro lado, existen 232 productos sin valoraciones, con una presencia notable de artículos de Dior y barras de labios. En cuanto a las valoraciones bajas, Sephora Collection aparece frecuentemente, aunque esto puede atribuirse a su amplio volumen dentro del catálogo. Finalmente, las variaciones más destacadas se observan en productos de Rem Beauty y en bases de maquillaje, que presentan un mayor número de opciones para el consumidor.

3.2. Análisis Específico de Sephora Collection

En cuanto al **éxito interno** de Sephora Collection, la marca cuenta con 32 productos considerados exitosos, lo que representa el 8,14% del total de productos de Sephora y un 16,16% dentro de la propia colección Sephora Collection. Dentro de las diferentes categorías, los pinceles y accesorios se destacan como el principal motor de éxito, alcanzando una eficiencia interna del 40%. Asimismo, la categoría de desmaquillantes, aunque pequeña en volumen, muestra una alta eficiencia, posicionándose como un segmento muy relevante. En contraste, la categoría de labios presenta un menor éxito relativo, lo que podría señalar áreas de mejora o necesidad de optimización en la oferta.

Respecto a **precios y variaciones**, los productos exitosos tienen un precio mediano de 18,49 €. Se observa que el número de variaciones por producto no guarda una correlación significativa con el éxito, lo que indica que una mayor variedad no siempre se traduce en mejor desempeño. Entre las categorías, los estuches de maquillaje son los más caros, mientras que los sacapuntas ocupan el extremo opuesto con precios más bajos. En cuanto a las subcategorías, las bases de maquillaje y lápices de ojos son las que presentan mayor variabilidad en precios y opciones.

En términos de **outliers**, el producto más caro dentro de Sephora Collection es un estuche de pinceles que se encuentra fuera de la tendencia general de precios. Por otra parte, los productos con más valoraciones tienden a ser eyeliners y lápices para cejas, los cuales cuentan además con buenas valoraciones, reflejando una buena aceptación por parte de los consumidores. Finalmente, se identifican pocos productos sin valoraciones o con valoraciones negativas, y no se aprecian patrones claros que expliquen estas situaciones.

3.3. Limitaciones y Recomendaciones Futuras

En este análisis, se ha utilizado el **precio bruto** sin realizar ajustes en función del volumen del producto, lo que representa una limitación importante. No se ha considerado el precio por unidad (€/ml o €/g), factor que podría sesgar algunos resultados y conclusiones. Por ello, se recomienda incluir esta variable en futuros análisis para obtener una visión más precisa y poder identificar de manera más fiable los outliers reales en cuanto a precios.

Otro aspecto a destacar son las **incoherencias presentes en algunos atributos específicos del catálogo**. Se han detectado problemas en la categorización y en los filtros disponibles en la web, lo que dificulta realizar análisis fiables sobre formatos, acabados, texturas y otros detalles relevantes de los productos. Aunque estos campos no fueron empleados en los análisis agregados actuales, podrían explorarse de manera puntual para obtener insights más detallados y específicos.

Finalmente, se optó por utilizar la mediana en lugar de la media para el tratamiento de variables numéricas debido a la **presencia de valores atípicos (outliers)**. Este enfoque permite ofrecer una visión más representativa y robusta del comportamiento general de los datos, evitando que valores extremos distorsionen las conclusiones.

3.4. Análisis y Recomendaciones para Nuevos Lanzamientos de Sephora Collection

3.4.1. Contexto general de la marca

Sephora Collection se posiciona como la **marca más representativa** dentro del catálogo de Sephora en términos de número de productos, abarcando un 10,87% del total disponible. Aunque aún no ha logrado cubrir 8 de las 48 subcategorías existentes, domina claramente en 13 de ellas, mostrando un fuerte liderazgo en diversidad y presencia.

Esta marca se distingue por su **estrategia de precios competitivos**, situándose entre las opciones más económicas dentro de Sephora. Además, goza de una **valoración generalmente positiva** por parte de los consumidores. En concreto, el 8,14% de los productos exitosos dentro de Sephora pertenecen a Sephora Collection, cifra que solo es superada ligeramente por Benefit Cosmetics con un 8,91%.

Sephora Collection es percibida como una marca que ofrece una excelente relación calidad-precio, atrayendo principalmente a consumidores que buscan un rendimiento funcional sólido sin incurrir en el coste de productos de lujo innecesarios.

Por lo tanto, para mantener y potenciar el éxito en futuros lanzamientos, Sephora Collection debe seguir reforzando su propuesta de valor basada en tres pilares fundamentales: **ofrecer precios competitivos, garantizar una buena relación calidad-precio, y mantener su atractivo para un público que prioriza el rendimiento práctico por encima del lujo.**

3.4.2. Lanzamiento Seguro: Brocha de Rostro

Sephora Collection cuenta con un liderazgo sólido en la subcategoría de brochas de rostro, con un total de 25 productos que representan el 12,63% del catálogo interno de la marca. Destaca que 12 de los 32 productos exitosos de Sephora Collection (un 37,5%) son brochas, lo que subraya la alta tasa de éxito de esta línea, donde cerca del 48% de las brochas disponibles son consideradas exitosas.

En el panorama global de Sephora, esta subcategoría ocupa el segundo lugar en número de productos exitosos, con un 7,63%, y de las 30 brochas exitosas en Sephora, 12 pertenecen a Sephora Collection. Además, el modelo de brochas de rostro dentro de la marca es sencillo, con poca variabilidad en diseño (mediana de variaciones igual a cero).

El rango de precios va desde 11 € hasta 34,99 €, destacando que las brochas exitosas tienen un precio mediano superior (22,5 €) en comparación con las que no alcanzan éxito (12,99 €), lo que indica una disposición de los consumidores a pagar más por calidad dentro de la gama económica.

Riesgos y mitigación: Aunque la subcategoría está bastante saturada con 25 productos, y la competencia es intensa tanto interna como externamente, se recomienda apostar por una brocha con una diferenciación clara, ya sea mediante doble funcionalidad, una edición especial, colaboraciones artísticas o un packaging sostenible. Se sugiere posicionar el precio en el rango alto (22–28 €) para aumentar el valor percibido sin perder la identidad económica de la marca.

3.4.3. Lanzamiento Arriesgado: Crema BB & CC

En esta subcategoría, Sephora Collection no tiene presencia actualmente, lo que representa una oportunidad potencial para incursionar. Las cremas BB & CC muestran una alta tasa de éxito en Sephora, alcanzando el 50%, y esta categoría está bien valorada con una media de 4,55 sobre 5.

La competencia está liderada por marcas como Erborian, cuyos productos tienen un precio mediano más elevado, alrededor de 30 €. Los productos exitosos en esta categoría presentan menos variaciones y un precio mediano de 32 €, mientras que los que no tienen éxito tienden a tener más variaciones y un precio más alto (39 €), lo que sugiere que el precio es un factor crítico, más que la cantidad de variaciones.

Estrategia recomendada: Se aconseja lanzar un único producto sin variaciones para validar la respuesta del mercado, con un precio competitivo situado entre 24 y 27 €, por debajo del rango de los productos exitosos, para mantener un posicionamiento accesible. La fórmula debe ser sencilla, universal y apta para todo tipo de piel, acompañada de una comunicación clara y educativa para diferenciarse de las marcas líderes.

Riesgos y mitigación: La falta de reconocimiento de Sephora Collection en esta categoría, la fuerte competencia consolidada y la complejidad en la percepción del producto son riesgos importantes. Para mitigarlos, es fundamental mantener una comunicación transparente, una fórmula sencilla y hacer un seguimiento cercano del feedback para ajustar futuras expansiones.

3.4.4. Lanzamiento de Mejora: Sombras de Ojos

Actualmente, Sephora Collection cuenta con 17 sombras de ojos en su catálogo, de las cuales 4 son consideradas exitosas (23,53%). Aunque es la marca con más sombras dentro de Sephora, la mayoría de los éxitos en esta subcategoría corresponden a otras marcas (19 de 23). Los productos no exitosos de Sephora Collection muestran menor cantidad de valoraciones y peores puntuaciones en comparación con la competencia.

El precio accesible de estas sombras (mediana 13,99 €) no ha garantizado el éxito, y se observa que una mayor cantidad de variaciones no asegura mejores resultados, pues los productos exitosos tienen menos variaciones. Además, las sombras exitosas presentan un precio ligeramente superior (14,99 € frente a 12 € en los no exitosos).

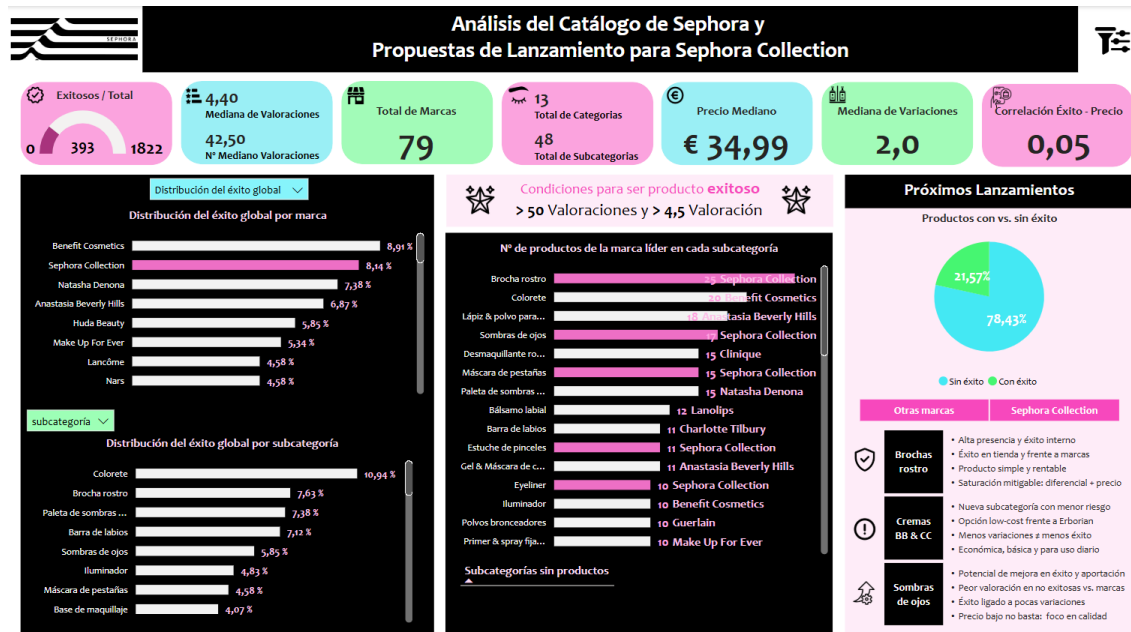
Oportunidad: El principal competidor en esta subcategoría es Natasha Denona, con un precio mediano considerablemente más alto (73,99 €). Las paletas de sombras tienen una tasa de éxito superior (39,19%) en comparación con las sombras individuales (30,26%), mientras que Sephora Collection tiene una presencia limitada en paletas (solo tres, ninguna exitosa).

Recomendaciones para el lanzamiento de mejora: Se recomienda apostar por una paleta compacta, sin variaciones en formato, con un surtido pequeño de tonos (entre 3 y 5), pensada para ser versátil tanto para el uso diario como para ocasiones especiales. Se sugiere un ligero incremento del precio (15–17 €) para reflejar una mejora en calidad, junto con mejoras en la fórmula que aumenten la pigmentación, duración y facilidad de aplicación. El diseño y el packaging deben ser atractivos y funcionales para transmitir una percepción de calidad.

Finalmente, es fundamental implementar una estrategia de marketing digital que incluya activaciones con influencers para incrementar la visibilidad y fomentar reseñas positivas.

4. Dashboard final

Este dashboard es el resultado visual y funcional del análisis exploratorio de datos (EDA) realizado sobre el catálogo de productos de Sephora, y constituye una herramienta clave para la toma de decisiones estratégicas relacionadas con el portafolio actual y las propuestas de lanzamiento para la línea Sephora Collection.



4.1. Estética y diseño visual

La estética del dashboard ha sido cuidadosamente alineada con la **identidad visual de Sephora**. Se han utilizado las rayas negras y blancas, elementos icónicos del packaging de la marca, así como tonos rosas, habituales en el entorno visual de sus tiendas. Además, se han incorporado detalles en azul y verde, inspirados en la iluminación y ambientación de los espacios físicos de Sephora. Estos recursos cromáticos no solo refuerzan la coherencia con la marca, sino que también contribuyen a estructurar visualmente las distintas secciones del panel y guiar la navegación del usuario de forma intuitiva.

4.2. Estructura general del dashboard

El dashboard se divide en **dos secciones principales**, diferenciadas por el color de fondo:

- **Sección de análisis general (fondo negro)**

Esta primera parte del dashboard presenta un análisis amplio del catálogo actual, abarcando **todas las marcas, categorías y subcategorías** de productos. Para ello se han utilizado principalmente **gráficos de barras**, por ser los más adecuados para comparar magnitudes entre distintos elementos de forma clara y directa.

En esta sección se muestran **7 indicadores clave (KPIs)**, algunos de ellos con doble valor representado, seleccionados estratégicamente por su relevancia dentro del análisis del portafolio: la proporción de productos exitosos sobre el total, el número mediano de valoraciones y la valoración mediana, el precio mediano, el número mediano de variaciones, el total de categorías/subcategorías disponibles, así como la correlación entre el precio y el éxito del producto. En conjunto, estas métricas permiten obtener una fotografía precisa y detallada del rendimiento y composición del portafolio actual.

Dado que uno de los objetivos principales de esta fase del análisis es **posicionar a Sephora Collection frente al resto de marcas** de la tienda, las barras correspondientes a Sephora Collection se han representado en color rosa, en contraste con el resto de las marcas, que aparecen en blanco. Esta diferenciación visual facilita su localización durante la presentación y permite apreciar su desempeño de forma inmediata. No obstante, el dashboard ha sido diseñado para ofrecer flexibilidad total de exploración, de modo que cualquier usuario pueda examinar en profundidad otras marcas o categorías si así lo desea.

Además, se han incorporado controles interactivos avanzados, como desplegados que permiten modificar dinámicamente **el eje X y el eje Y** de ciertos gráficos. Esto brinda la posibilidad de combinar diferentes métricas (como el total de productos, la distribución del éxito global, la tasa de éxito por grupo o el precio mediano) y analizarlas a distintos niveles (por categoría, subcategoría o marca).

Finalmente, en la esquina superior derecha se encuentran **tres filtros ocultos** que se despliegan al hacer clic sobre ellos, y que permiten filtrar la información por marca, categoría o subcategoría. Esta funcionalidad está orientada a quienes deseen realizar un análisis más detallado o centrarse en segmentos específicos del catálogo.

- **Sección de lanzamientos propuestos (fondo rosa)**

La segunda parte del dashboard, identificada visualmente con un fondo rosa, está centrada en las **propuestas de nuevos lanzamientos para la línea Sephora Collection**. Esta sección recoge de forma visual y sintética las principales conclusiones del análisis exploratorio, materializadas en tres propuestas concretas de producto.

Cada propuesta va acompañada no solo de la recomendación de lanzamiento, sino también de una justificación estratégica respaldada por datos, donde se destacan oportunidades detectadas, huecos de mercado y patrones relevantes observados en el catálogo actual. La sección está pensada para ser interactiva, de manera que el usuario puede explorar las justificaciones a través de filtros, botones y parámetros que reflejan de forma dinámica los resultados del análisis.

5. Next steps

Aunque a lo largo del informe ya se han mencionado algunas oportunidades de mejora y ampliación, este apartado recoge de manera estructurada todos los posibles próximos pasos que permitirían continuar desarrollando y refinando el análisis presentado.

En primer lugar, una de las principales limitaciones del análisis actual ha sido el uso del precio bruto sin **tener en cuenta el volumen del producto**. La ausencia del precio por unidad (€/ml o €/g) puede haber condicionado ciertas conclusiones y dificultado la identificación precisa de productos anómalos en cuanto a posicionamiento de precio. Incluir esta variable en futuros análisis permitiría una comparación más justa entre productos y un análisis más robusto de precios extremos.

Además, se detectaron ciertas inconsistencias en algunos de los **atributos más específicos** de los productos, como en los formatos, acabados o texturas, debido a limitaciones en la estructura de datos de la web original. Aunque estos campos no se incluyeron en el análisis principal por su baja fiabilidad, sí podrían **explorarse de forma puntual o mejorar su tratamiento** con técnicas de limpieza más avanzadas. En la misma línea, también se intentó sin éxito **extraer los ingredientes** de los productos, pero la extrema variabilidad en el formato de presentación hizo inviable su procesamiento sin aplicar técnicas de NLP o reglas complejas. No obstante, este tipo de información supondría un valor añadido importante en análisis futuros.

Desde el punto de vista técnico, persiste un error detectado tras la primera carga del histórico, donde se observaron 7 duplicados a pesar de que no deberían existir en esa primera ejecución. La causa concreta no ha sido localizada, por lo que se propone **revisar el proceso de inserción en futuras actualizaciones**. Asimismo, la base de datos se ha dejado montada y lista para seguir almacenando datos a lo largo del tiempo, permitiendo así **futuros análisis temporales**, que no se han llevado a cabo en esta versión por falta de volumen histórico suficiente.

En cuanto al dashboard final, podría considerarse la **inclusión de una sección específica dedicada al análisis de outliers**. También sería interesante explorar visualizaciones centradas en la evolución temporal de productos si se dispone de datos suficientes en el futuro.

Finalmente, a nivel de organización y buenas prácticas, tanto el notebook de análisis exploratorio (EDA) como los scripts de la ETL podrían haberse **documentado** de forma más exhaustiva. En el caso del EDA, habría sido recomendable modularizar las funciones, trasladarlas a un directorio src/ y automatizar la generación de resultados clave, por ejemplo, exportándolos en PDF. En la parte de la ETL, aunque el código está estructurado en scripts .py, ha faltado añadir docstrings explicativos en las funciones, algo que se podría mejorar para facilitar la reutilización del código en futuros desarrollos. También, **mejorar el archivo README** para que funcione como guía clara y estructurada que permita a cualquier persona replicar el proyecto de forma autónoma.