



Universidad
Francisco de
Vitoria

UFV Madrid

UNIVERSIDAD FRANCISCO DE VITORIA

ESCUELA POLITÉCNICA SUPERIOR

GRADO EN INGENIERÍA MATEMÁTICA

VISUALIZACIÓN DE DATOS

PLANIFICACIÓN DEL PROYECTO

Caso de Uso - **PUIG**

Miguel Poyatos García-Lorezana

Miguel Ángel Rojo Gala

29 de noviembre de 2024

ÍNDICE

- 1. Introducción 1
- 2. Metodología..... 2
- 3. Modelo De Datos 3
- 4. Plan 4
 - 4.a. Nivel Dificultad 1 4
 - Pregunta 1 4
 - Pregunta 2..... 4
 - Pregunta 3..... 4
 - Pregunta 4..... 5
 - Pregunta 5..... 5
 - 4.b. Nivel Dificultad 2 6
 - Pregunta 1 6
 - Pregunta 2..... 6
 - 4.c. Nivel Dificultad 3 6
 - Pregunta 1 6

INTRODUCCIÓN

La creciente relevancia de la inteligencia empresarial en la toma de decisiones ha impulsado a las empresas a incorporar herramientas avanzadas de análisis y visualización de datos. En este contexto, Puig, una multinacional española especializada en moda y perfumes, plantea un caso de uso para optimizar sus procesos de análisis de datos en el ámbito del testing de productos para la piel. Este proyecto va a evaluar y recopilar información sobre productos propios y de la competencia, considerando atributos sensoriales como brillo, suavidad o pegajosidad, además de analizar el impacto de los ingredientes y su aceptación en diferentes mercados.

El objetivo principal del informe consiste en desarrollar un plan de los dashboards interactivos que se van a desarrollar utilizando Power BI, que permitan visualizar los datos de manera efectiva para responder preguntas clave relacionadas con la calidad, composición y rendimiento de los productos. A través de este proyecto, se busca no solo responder a las preguntas iniciales planteadas por Puig, sino también proporcionar un marco analítico escalable y alineado con las mejores prácticas de Business Intelligence.

METODOLOGÍA

Para llevar a cabo este proyecto de visualización y análisis de datos para Puig, se aplica una metodología basada en cuatro pasos clave, adaptándola a las necesidades del cliente y al caso de uso.

Primero, **se define el mensaje y los objetivos**: el propósito principal es ofrecer una representación clara de los datos de testing de productos, destacando la calidad, composición y rendimiento de estos en comparación con los competidores. Esto incluye proporcionar resultados concluyentes sobre los atributos sensoriales más valorados (brillo, suavidad, pegajosidad) y el impacto de ingredientes específicos en diferentes mercados.

En segundo lugar, **se conocen los datos que apoyan los gráficos**: se recopilan métricas detalladas de productos propios y de la competencia, garantizando la calidad y relevancia de las fuentes. Se estructuran los datos de manera óptima para integrarlos eficientemente en Power BI, analizando patrones clave y relaciones entre atributos sensoriales y preferencias de mercado.

En tercer lugar, **se considera al público objetivo**, enfocado en los equipos de I+D y gerencia de Puig, quienes necesitan dashboards intuitivos y dinámicos que faciliten la toma de decisiones estratégicas. Diseñaremos gráficos interactivos que permitan explorar métricas clave de forma clara y eficiente, respondiendo a las preguntas planteadas.

Por último, **se identifican las limitantes**: el diseño de los dashboards se adapta al entorno de Power BI, por lo que las gráficas se adaptarán a lo que ofrece la herramienta. También la profundidad del desarrollo del proyecto está muy limitada por el tiempo, debido a que tenemos de plazo hasta el 16 de enero, y además, se presentará en 20 minutos.

Además, se sigue estrictamente las especificaciones visuales de Puig, utilizando su paleta de colores oficial, tipografías corporativas y un enfoque estético alineado con su imagen de marca. De este modo, se logran representaciones visuales profesionales que no solo cumplen con los objetivos analíticos, sino que también refuerzan su identidad corporativa en cada presentación. Este enfoque integral asegura que el proyecto esté alineado con los objetivos estratégicos de Puig y maximiza el valor que la inteligencia empresarial aporta a sus procesos.

MODELO DE DATOS

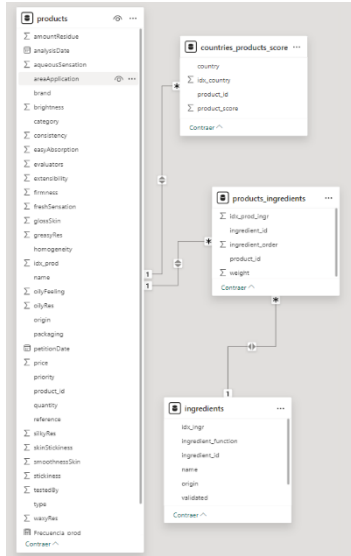
En el desarrollo de este proyecto, el procesado de datos ha sido una etapa fundamental para garantizar la calidad y coherencia de la información disponible antes de proceder al análisis. Durante esta fase, se abordaron varios aspectos clave. En primer lugar, se identificaron y corrigieron valores erróneos presentes en los datos. Para resolverlo, se estudió cuál era la naturaleza del error que aparecía, en concreto, aparecía un error en la tabla *products* en el registro 1277 como se muestra en la *Figura 1*. Al abrir el archivo como blog de notas se descubrió que contenía información desconocida, por lo que se concluyó que lo óptimo era eliminar esas filas en el Power BI, perdiendo posible información.

Figura 1: Error detectado

1276;64ccbf20c171551f708c4f09;HEXANEDIOL	Active, in use	1,2;competitor;True;
1277;64ccbf20c171551f708c4f0a;"HEXANEDIOL	HEXANODIOL	1,100000
SM001995	MYRISTYL MYRISTATE	Active, in use MYRISTYL MYRISTATE MIRISTATO DE MIRISTILO 1,000000
SM001404	POTASSIUM CETYL PHOSPHATE	Active, in use POTASSIUM CETYL PHOSPHATE CETIL FOSFATO DE POTASIO 0,500000
2999999999	Carga inicial	Active, in use 0,500000
SM001609	CAPRYLYL GLYCOL	Active, in use CAPRYLYL GLYCOL CAPRILIL GLICOL 0,350000
SM001280	CARBOMER	Active, in use CARBOMER CARBOMER 0,300000
SM001256	ACRYLATES/C10"	competitor;True;
1278;64ccc10fc171551f708c50e7; Aceite de camelia	ecologico (C.OLEIFERA);xr;True;	
1279;64ccc10fc171551f708c50e8;NONAPEPTIDE-1;inci;True;Active		

Otro aspecto importante fue la corrección de los números decimales. El formato de los números en los csv contenía los números indicando la parte decimal con puntos, por lo que se procesó en cambiando las columnas numéricas los puntos por comas para adaptarlo al formato de Power BI. Se verificó la unicidad de las claves principales de las tablas, como los identificadores de productos e ingredientes, asegurando que no existieran duplicados y que cada registro fuese único. Además, se escalaron los datos numéricos de 0 a 10, dejándolos todos en la misma escala para ser más fácilmente comparables.

Figura 2: Visualización del modelo



El modelo de datos se estructuró siguiendo el esquema de la *Figura 2*, en el cual la tabla de *products_ingredients* actúa como el eje central, conectándose con las tablas de productos e ingredientes. La relación de la tabla principal con las otras dos es de “1:N”, porque en las otras dos tablas está la clave primaria que es única que identifica los productos/ingredientes, pero en la tabla principal aparecen repetidas veces, es decir, un número N.

Asimismo, la tabla de productos está vinculada con la tabla de puntuaciones por país, lo que permite identificar la aceptación de los productos en diferentes mercados. Esta relación también es de “1:N” por lo mismo explicado, la tabla de productos tiene la clave primaria, pero la tabla de países contiene repetidas veces la clave que identifica a los productos.

Hay que destacar que, como se ve en la gráfica, todas las relaciones son bidireccionales para que al filtrar en las gráficas y tablas afecte en ambos sentidos. Además, en el siguiente apartado para responder a algunas preguntas se adaptará un postprocesado con código DAX, utilizando tablas auxiliares y medidas rápidas que permitan hacer gráficas más avanzadas y dinámicas.

PLAN

NIVEL DIFICULTAD 1

▪ PREGUNTA 1

Se quiere visualizar los atributos más relevantes de cada producto (origen, marca, ...) pudiendo filtrar los productos por cada uno de estos campos, pero también por ingredientes para poder saber que productos contienen un ingrediente concreto.

Para visualizar los atributos más relevantes de cada producto, se diseñará un dashboard que permita explorar de manera clara y detallada la información disponible. Este dashboard mostrará en una **tabla** los atributos clave de los productos, como la marca, el tipo y la zona de aplicación entre otros. En esta tabla, **cada fila representará un producto**.

Además, esta será una **tabla interactiva** en la que se podrán implementar **filtros avanzados**, que permitirán a los usuarios seleccionar productos en función de atributos específicos. Por ejemplo, será posible filtrar por la marca o el tipo de producto, así como buscar directamente aquellos que contengan un ingrediente concreto. Los filtros estarán configurados para interactuar con la tabla, actualizando automáticamente los gráficos y tablas en función de los criterios seleccionados.

▪ PREGUNTA 2

Se quiere visualizar la distribución de cada dimensión categórica pudiendo elegir dinámicamente la dimensión representada (sugerencia: usa un parámetro).

Para visualizar la distribución de las dimensiones categóricas de forma dinámica, se diseñarán **dos gráficos para cada dimensión: un gráfico de tarta y un gráfico de barras**, de tal manera que permita seleccionar entre diferentes dimensiones. En el **gráfico de tarta**, habrá una categoría "Otros" que agrupará aquellas categorías menos relevantes cuando existan más de 5 para visualizar el reparto con respecto al total. El objetivo de este gráfico es **comparar cada categoría con respecto al total**. En cambio, el objetivo del gráfico de barras busca **comparar la frecuencia absoluta de todas las categorías, entrando en más detalle** (más categorías).

La implementación se realizará utilizando un parámetro dinámico en Power BI. Este parámetro incluirá las dimensiones categóricas disponibles en los datos, permitiendo al usuario seleccionar una de ellas desde un menú desplegable. Para conectar este parámetro con los gráficos, se creará una medida en DAX que determine qué dimensión debe mostrarse según la selección realizada. Esta medida actualizará automáticamente los datos y las etiquetas de los gráficos, ajustando la visualización en tiempo real.

▪ PREGUNTA 3

Se quiere visualizar el perfil sensorial de uno o varios productos. Bonus (nivel 3): hay visuals que no vienen por defecto, pero los puedes descargar gratuitamente. Prueba varios.

Para visualizar el perfil sensorial de uno o varios productos, se propone el uso de un **gráfico de radar**, también conocido como gráfico de telaraña, que permite representar de forma clara y comparativa los valores de los atributos sensoriales. Estos atributos, como consistencia, brillo, frescura y suavidad, están medidos en una escala, de 0 a 10. Este tipo de gráfico es especialmente útil para comparar fortalezas y debilidades de manera intuitiva.

Una vez procesados los datos, se utiliza una visualización de radar, que no está disponible de forma predeterminada en Power BI pero se ha descargado gratuitamente. El gráfico se configura para mostrar cada atributo sensorial como un eje radial, mientras que los valores de los productos seleccionados se representan como puntos distribuidos a lo largo de cada eje. Esta configuración permite comparar múltiples productos en la misma visualización.

Adicionalmente, se añaden filtros interactivos que permiten elegir los productos que se quieren visualizar en el gráfico.

▪ PREGUNTA 4

¿Cuáles son los ingredientes más comunes para los 20 productos más aceptados en cada país?

Para identificar los ingredientes más comunes en los 20 productos más aceptados en cada país, se realizará un análisis basado en las puntuaciones por país y los ingredientes asociados a cada producto. El enfoque se centrará en seleccionar los productos mejor valorados en cada país y luego analizar qué ingredientes son los más frecuentes en estos productos.

La implementación comienza con la filtración de la tabla de puntuaciones por país, donde se seleccionan los 20 productos con las mejores puntuaciones en cada país. Una vez filtrados los productos, se utilizará la tabla intermedia de productos-ingredientes para identificar los ingredientes presentes en cada uno de estos productos. Luego, se contará la frecuencia de cada ingrediente en este grupo de productos seleccionados.

Para la visualización, se utilizará un **gráfico de barras** que muestre los ingredientes más comunes, ordenados por su frecuencia de aparición en los 20 productos más aceptados. Además, se incluirán filtros para que los usuarios puedan seleccionar un país específico y observar cómo varían los ingredientes según la región.

▪ PREGUNTA 5

¿Cuál es el ingrediente conservante más utilizado por categoría (dinámico) de producto?

Para determinar cuál es el ingrediente conservante más utilizado por categoría de producto, se procederá a clasificar los ingredientes según su función (por ejemplo, conservante, emulsionante, etc.) y a contar su frecuencia de uso dentro de cada categoría de producto. Este enfoque permitirá identificar cuáles son los conservantes más comunes en cada tipo de producto y analizar su distribución de manera dinámica.

Para la implementación en Power BI se preparará haciendo transformaciones de la tabla de ingredientes en Power Query para preparar la tabla (añadiendo columnas o con tablas auxiliares), donde se identifican los conservantes y se categoriza cada ingrediente según su función. Esta transformación facilitará la creación de medidas específicas para contar el uso de conservantes en productos de cada categoría.

Una vez que los datos estén preparados, se creará una medida en DAX que cuente cuántas veces se utiliza cada conservante en los productos de cada categoría. Esta medida permitirá actualizar los resultados de forma dinámica al aplicar filtros y seleccionar diferentes categorías de productos.

Para la visualización, se diseñará un **gráfico de barras agrupadas** o una **matriz interactiva**. En el gráfico de barras, las categorías de productos aparecerán en el eje Y, mientras que los

conservantes más comunes se ubicarán en el eje X. Alternativamente, en la matriz interactiva, las categorías se mostrarán en las filas y los conservantes en las columnas, lo que permitirá comparar rápidamente la frecuencia de uso de cada conservante dentro de las diferentes categorías.

NIVEL DIFICULTAD 2

▪ PREGUNTA 1

Se quiere visualizar el perfil sensorial medio de los productos favoritos (top 10) de cada país. Bonus (nivel 3): que el top sea flexible y que también se pueda definir el top por nota (ejemplo: nota mayor que 9).

Para visualizar el perfil sensorial medio de los productos favoritos (top 10) de cada país, se utilizará **un gráfico de araña** que se utilizó en una pregunta anterior, sirviéndonos de las mismas medidas.

La implementación para este caso comienza con la identificación de los productos favoritos en cada país. Esto se realiza filtrando la tabla de puntuaciones para seleccionar los productos con las mejores notas en cada región, definiendo un top 10 flexible que permita ajustar el número de productos o establecer un umbral mínimo de puntuación. A continuación, se calculan los valores promedio de los atributos sensoriales para los productos seleccionados, agrupados por país.

▪ PREGUNTA 2

¿Cuáles son los ingredientes clasificados por tipo de ingrediente que utiliza la competencia y Puig no?

Para identificar los ingredientes clasificados por tipo que son utilizados por la competencia, pero no por Puig, se realizará un análisis previo a la gráfica que compare las tablas de ingredientes asociadas a los productos de ambas partes. Esto permitirá determinar y guardar qué ingredientes específicos están presentes en los productos de la competencia y ausentes en los de Puig de una forma que se pueda representar.

El primer paso en la implementación consiste en transformar los datos en Power Query. Se creará una columna que clasifique los ingredientes por tipo (conservantes, emulsionantes, fragancias, etc.) y se identificará si cada ingrediente pertenece a productos de Puig, de la competencia o de ambos. Una vez procesados los datos, se filtrará la nueva tabla para mostrar únicamente los ingredientes utilizados exclusivamente por la competencia.

En Power BI, se diseñará una visualización interactiva que presente esta información. Un **gráfico de barras** agrupadas permitirá mostrar los tipos de ingredientes en el eje Y y el número de ingredientes únicos de la competencia en el eje X. Alternativamente, el dashboard se apoyará de una **tabla o matriz interactiva** para listar los ingredientes específicos, categorizados por tipo, con indicadores claros de exclusividad.

NIVEL DIFICULTAD 3

▪ PREGUNTA 1

Visualizar en que productos se encuentran una combinación de dos ingredientes (sugerencia: crea medidas y úsalas en los filtros).

Para visualizar en qué productos se encuentran simultáneamente dos ingredientes seleccionados, se diseñará una **tabla interactiva** que muestre los productos que cumplen esta condición. Esta tabla incluirá atributos clave de cada producto, como la marca, el tipo y la zona de aplicación, proporcionando contexto adicional. Esto permitirá analizar de manera precisa y dinámica la combinación de ingredientes en los productos.

Para la implementación se creará una medida en DAX que evalúe si un producto contiene ambos ingredientes seleccionados. Esta medida devolverá un valor booleano (TRUE o FALSE) dependiendo de si el producto incluye los dos ingredientes especificados. Para facilitar la selección de ingredientes, se configurarán slicers o listas desplegadas en el dashboard, permitiendo al usuario elegir dinámicamente las combinaciones de interés.