



DATA SCIENCE II

Alumno: Yesid García

Comisión: 75690

“Clasificación Nutricional en la Industria Alimentaria”

CONTENIDO

- Introducción.
- Descripción del problema.
- Objetivo.
- Fuente.



- Preguntas de Interes
- Hipótesis.
- Conclusiones.
- Oportunidades e Insights Clave.



INTRODUCCIÓN

La clasificación precisa de productos alimenticios, basada en su perfil nutricional, es fundamental para la toma de decisiones estratégicas en la industria. Este enfoque permite una segmentación de mercado efectiva, un análisis competitivo profundo y el desarrollo de nuevas líneas de productos que respondan a las tendencias de salud y consumo.

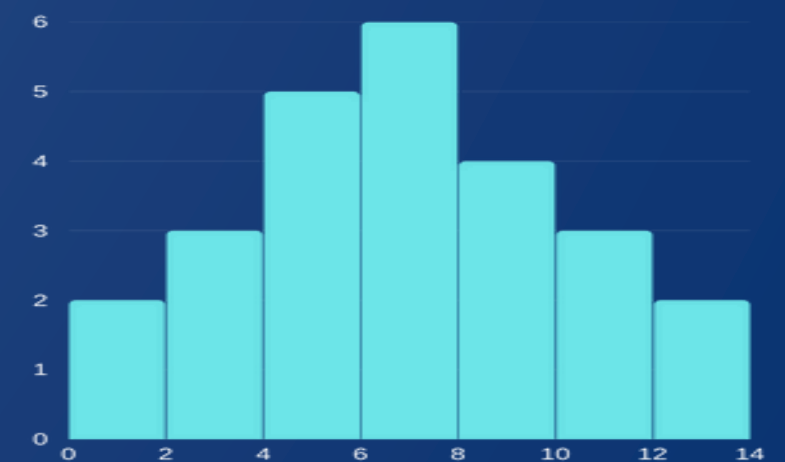
Aplicaciones Clave:

- Segmentación de Mercado: Identificar nichos y adaptar ofertas.
- Análisis Competitivo: Comparar productos y categorías.
- Desarrollo de Nuevas Líneas: Responder a tendencias de salud y consumo.



Nutrientes Clave:

- Energía (nutriments.energy100g)
- Grasas (nutriments.fat100g)
- Azúcares (nutriments.sugars100g)
- Sal (nutriments.salt100g)
- Proteínas (nutriments.proteins100g)
- Exceso de Grasas y Azúcares (a determinar)



DESCRIPCIÓN DEL PROBLEMA



El problema principal que aborda este proyecto, según la descripción proporcionada en el cuaderno de Colab, es la necesidad de clasificar de manera precisa productos alimenticios basándose únicamente en su perfil nutricional. Esto surge de la importancia de tener una clasificación correcta en la industria alimentaria para diversas aplicaciones, como:

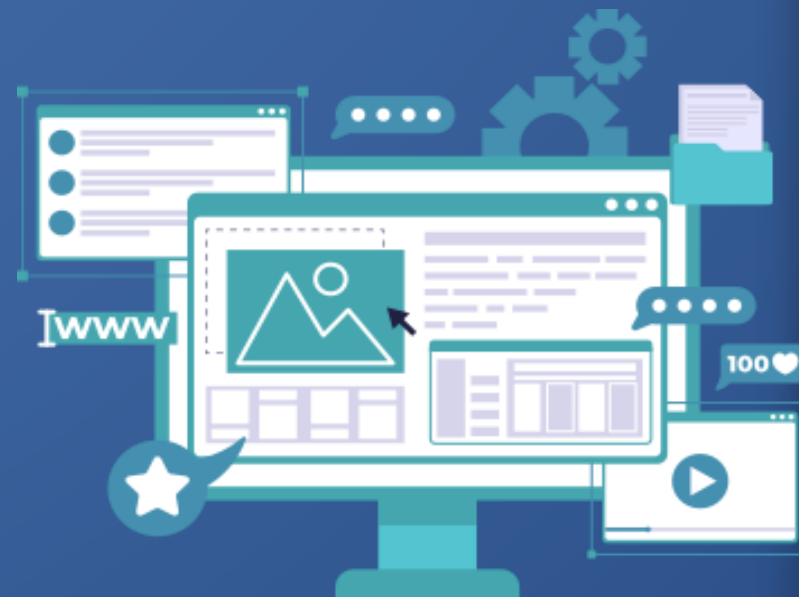
- **Segmentación de mercado.**
- **Análisis competitivo.**
- **Desarrollo de nuevos productos.**

El proyecto busca resolver este problema utilizando las variables nutricionales como base para predecir la categoría del producto, evaluando así la viabilidad de este enfoque para identificar patrones nutricionales típicos por segmento y evaluar oportunidades comerciales.

OBJETIVO PRINCIPAL: Predecir Categorías por Nutrientes

- 1. Identificar Perfiles:** Reconocer perfiles nutricionales típicos por segmento.
- 2. Evaluar Oportunidades:** Analizar oportunidades comerciales y patrones de consumo.
- 3. Facilitar Segmentación:** Mejorar la segmentación del mercado.

Este análisis busca predecir automáticamente la categoría de un producto alimenticio (snacks, bebidas, lácteos, etc.) utilizando únicamente sus valores nutricionales. Esto nos permitirá identificar perfiles nutricionales típicos por segmento, evaluar oportunidades comerciales y detectar patrones de consumo, facilitando la segmentación del mercado.



Objetivos Específicos del Análisis

- 1. Identificar Perfiles Nutricionales Característicos:**
¿Qué nutrientes definen a cada tipo de producto?
Análisis de tendencias centrales y distribución de nutrientes por categoría.
- 2. Desarrollar y Evaluar un Modelo de Clasificación Automática:**
¿Se puede predecir la categoría con un modelo automático usando solo los nutrientes?
Selección de características, algoritmo, entrenamiento y evaluación de precisión.
- 3. Comparar Perfiles Nutricionales entre Categorías:**
¿Qué categorías tienen un mejor o peor perfil nutricional general? Visualizaciones comparativas de nutrientes y cálculo de puntuaciones.
- 4. Analizar Concentración de Azúcares y Grasas:**
¿Qué tipos de productos tienden a concentrar más azúcares o grasas?





FUENTE DE DATOS



Origen: Los datos provienen de la API pública de **Open Food Facts**, una base de datos colaborativa con más de un millón de productos.



LINK:

https://www.google.com/url?q=https%3A%2F%2Fworld.openfoodfacts.org%2Fcgi%2Fsearch.pl%3Faction%3Dprocess%26json%3Dtrue%26page_size%3D100

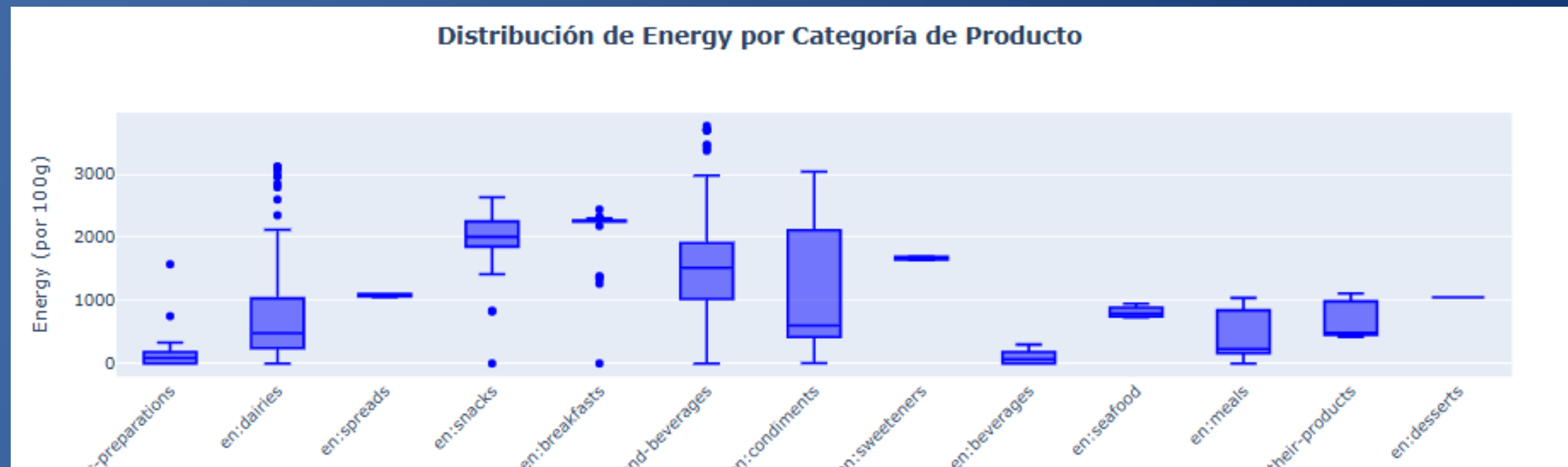


Variables Extraídas: Categorías de productos c(ategoriestags), Nutrientes por 100g: energía, grasas, azúcares, proteínas, sal, fibra, Marca (para análisis secundarios).

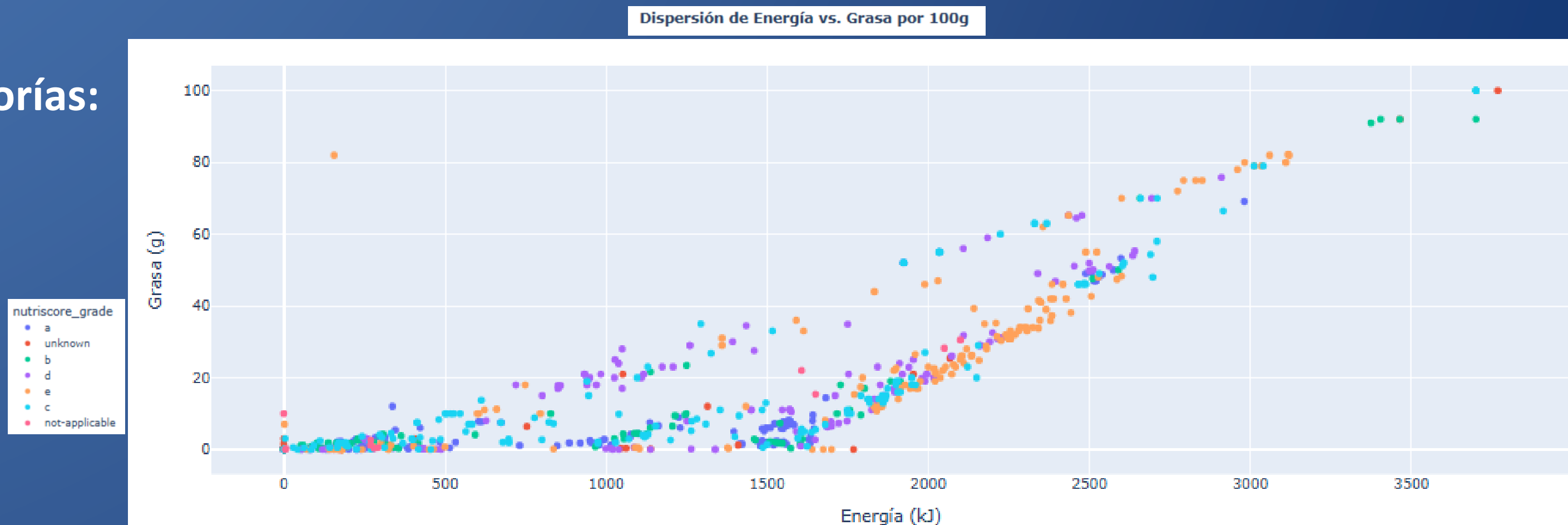
Tamaño de la Fuente: 713 filas x 23 columnas.

PREGUNTAS DE INTERÉS Y PATRONES NUTRICIONALES

1. **Nutrientes por Categoría:** Los perfiles nutricionales son distintivos y pueden usarse para clasificación.



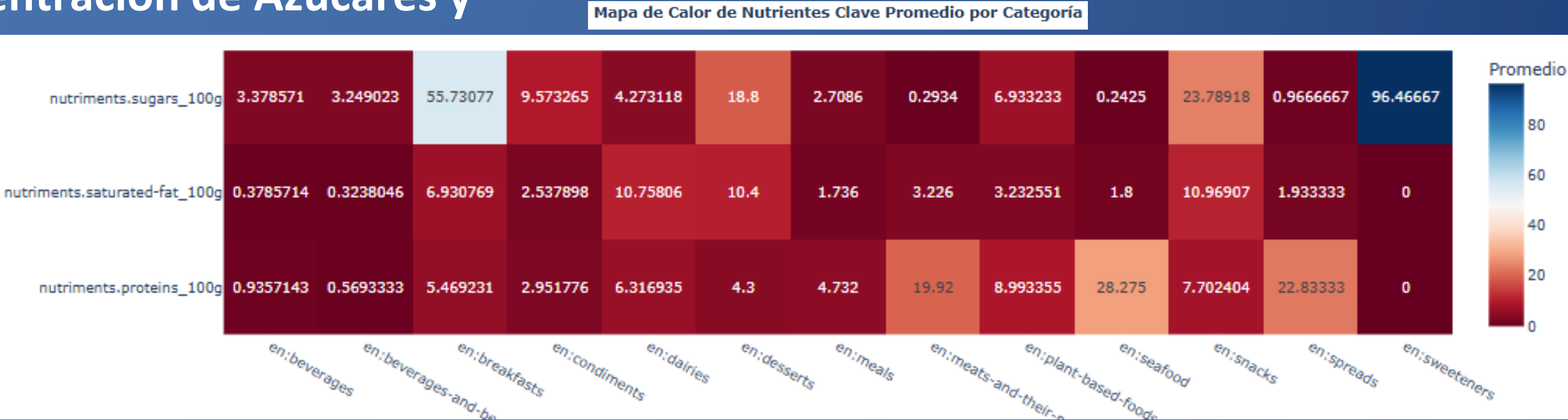
2. **Relación de Categorías:** Las mejor relación encontrada es la siguientes.



Nota: Correlación entre nutriments.energy_100g y nutriments.fat_100g: 0.803202

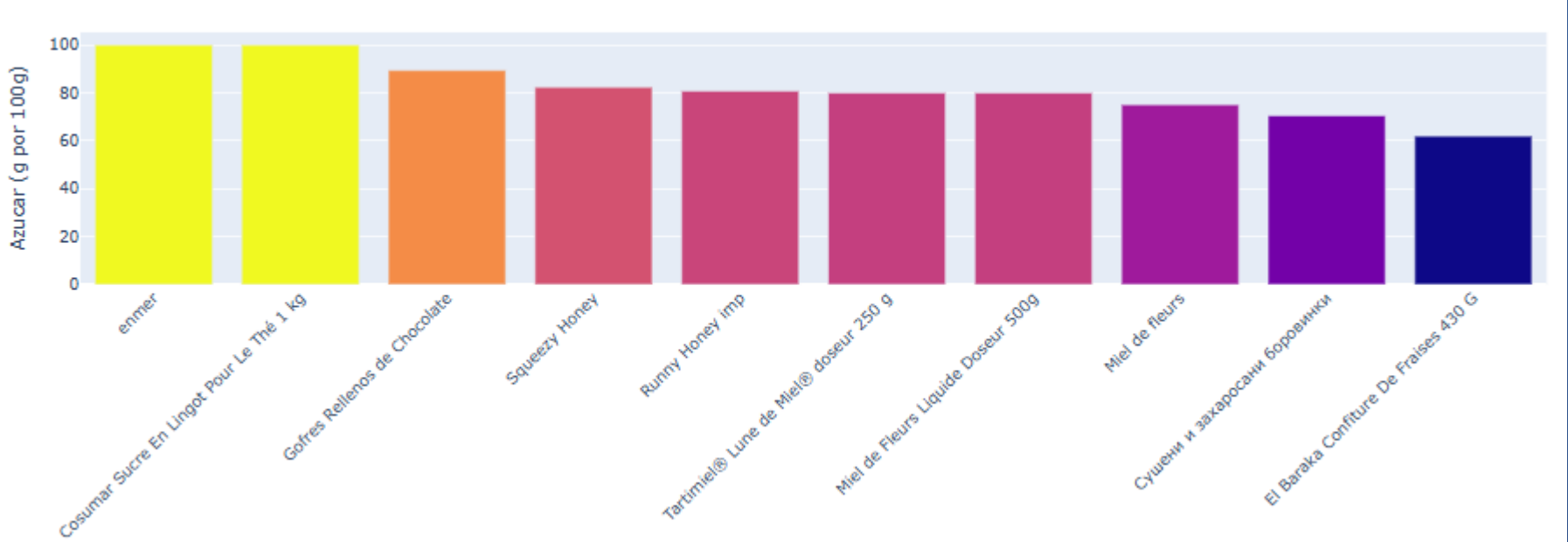
PREGUNTAS DE INTERÉS Y PATRONES NUTRICIONALES

3. Concentración de Azúcares y Grasas:

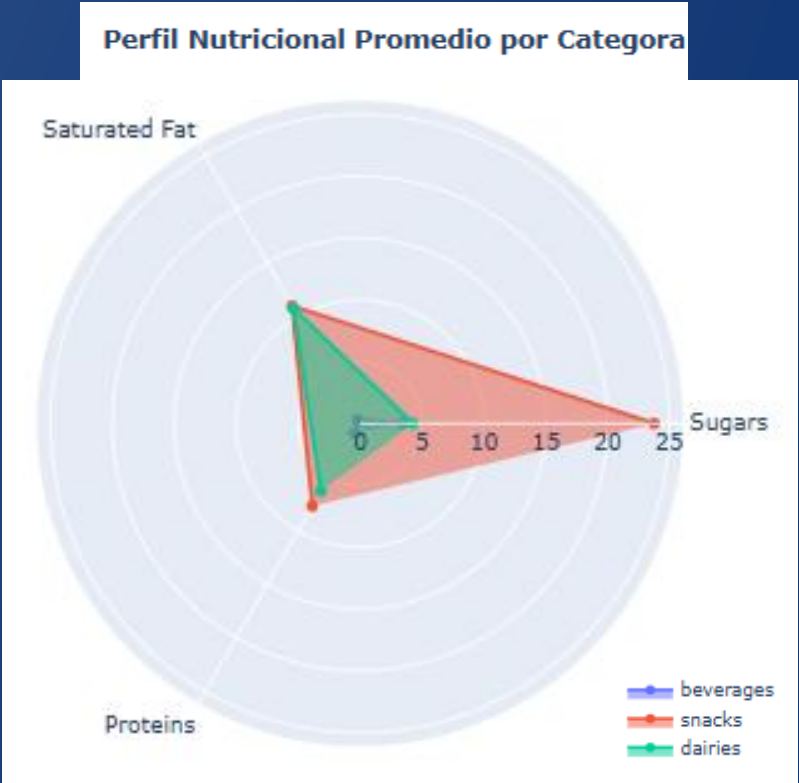


4. Productos con mayor concentración de Azúcares:

Top 10 Productos con Mayor Concentración de Azucres (con Exceso de Azucres)



5. Oportunidades para innovar en snacks o postres más saludables: Reducción de Azúcares Añadidos y Grasas Saturadas, Incremento del Contenido de Fibra y Proteínas, Uso de Ingredientes Naturales y Menos Procesados.



HIPÓTESIS PLANTEADAS

Hipotesis 6: Todos los productos con Nutri Score D y E poseen alergen.

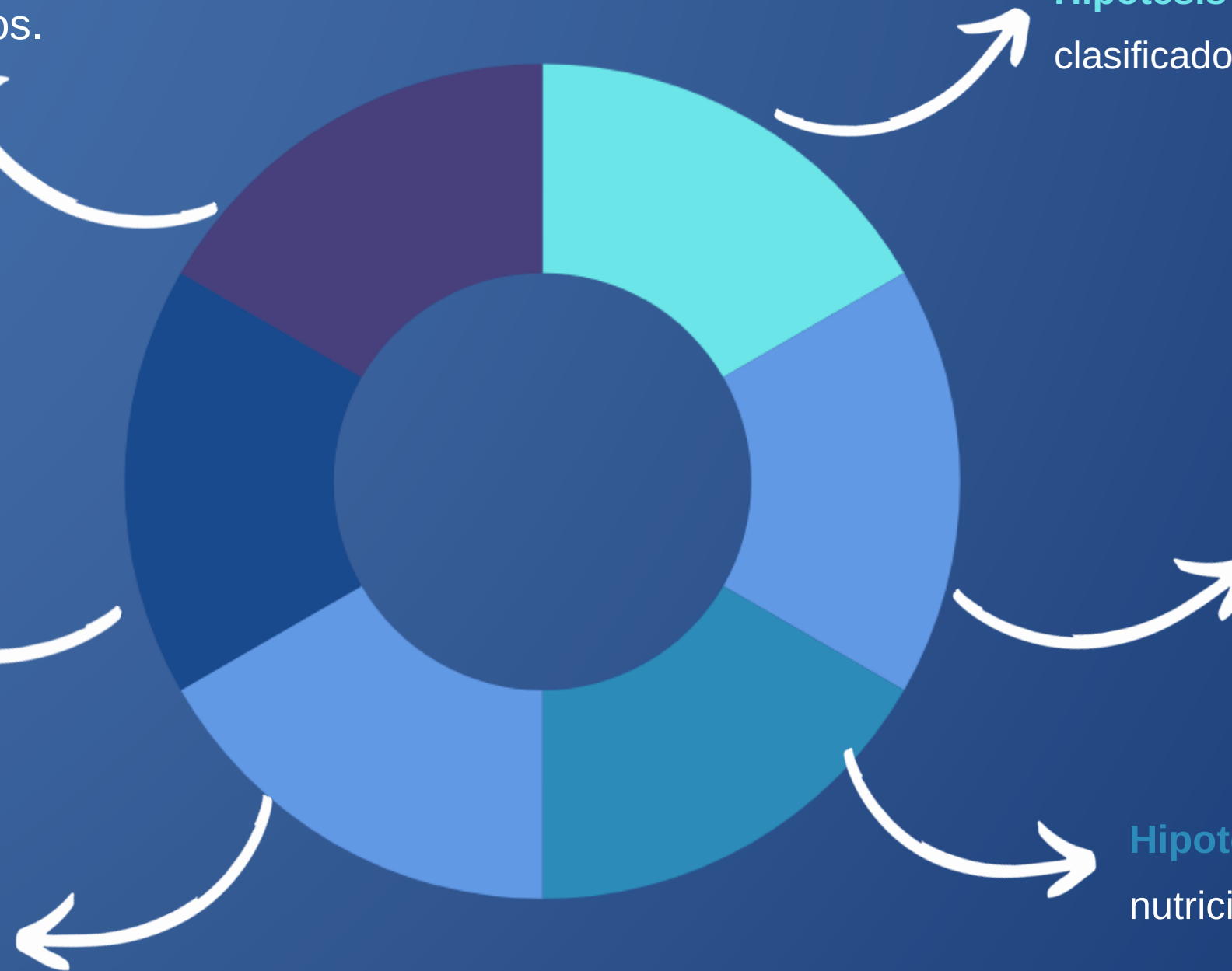
Hipotesis 1: Los productos pueden ser correctamente clasificados en categorías usando su perfil nutricional.

Hipotesis 2: Las bebidas tienen menor contenido de proteínas y grasas que los snacks o postres.

Hipotesis 3: Las categorías presentan perfiles nutricionales distintos y predecibles.

Hipotesis 4: Algunas marcas como Sidi Ali presentan más oportunidades de mejora nutricional azucares que otras marcas.

Hipotesis 5: Los productos más saludables como Snacks y bebidas tienden a concentrarse en ciertas categorías específicas y tienen un Nova Group por debajo de 3.



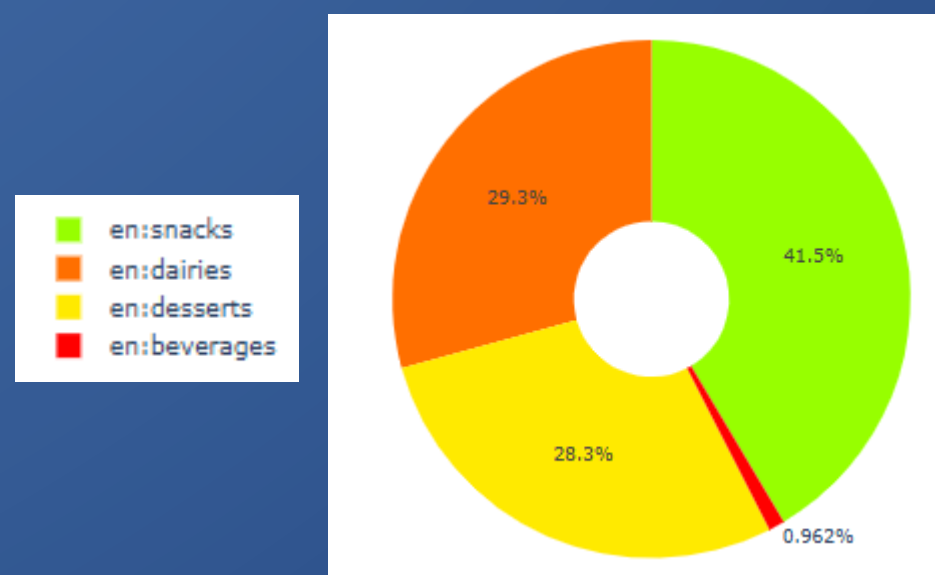
HIPÓTESIS 1: Los productos pueden ser correctamente clasificados en categorías usando su perfil nutricional.

Verdadero, El análisis exploratorio confirma una fuerte asociación entre el perfil nutricional y la categoría de un producto. Las diferencias en nutrientes entre categorías sugieren que cada tipo de producto tiene un "perfil nutricional característico".continuación de la siguiente visualización podremos tener la siguiente conclusión:

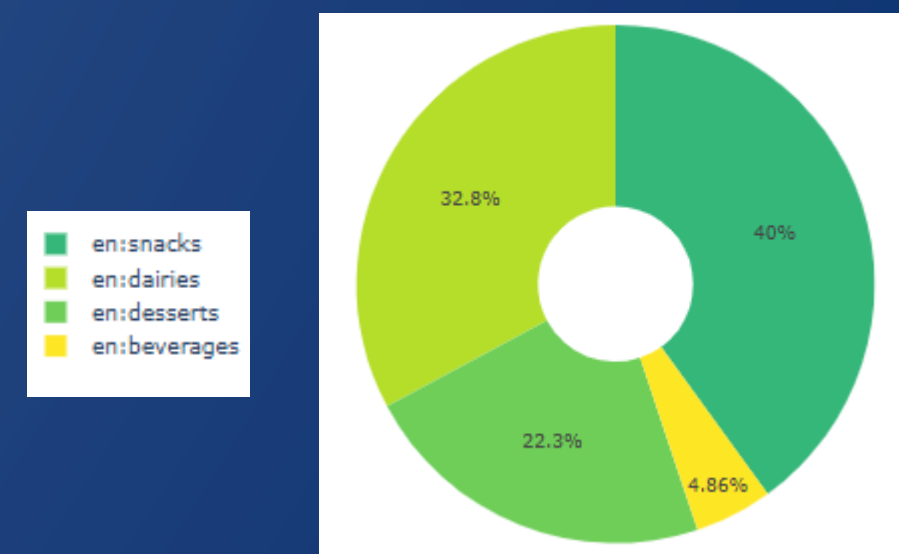


HIPÓTESIS 2: Las bebidas tienen menor contenido de proteínas y grasas que los snacks o postres.

Verdadero, puede verse esa diferenciación y resultado:



Distribución Promedio de Grasa por Categoría (Bebidas, Snacks, Postres)



Distribución Promedio de Proteínas por Categoría (Bebidas, Snacks, Postres)

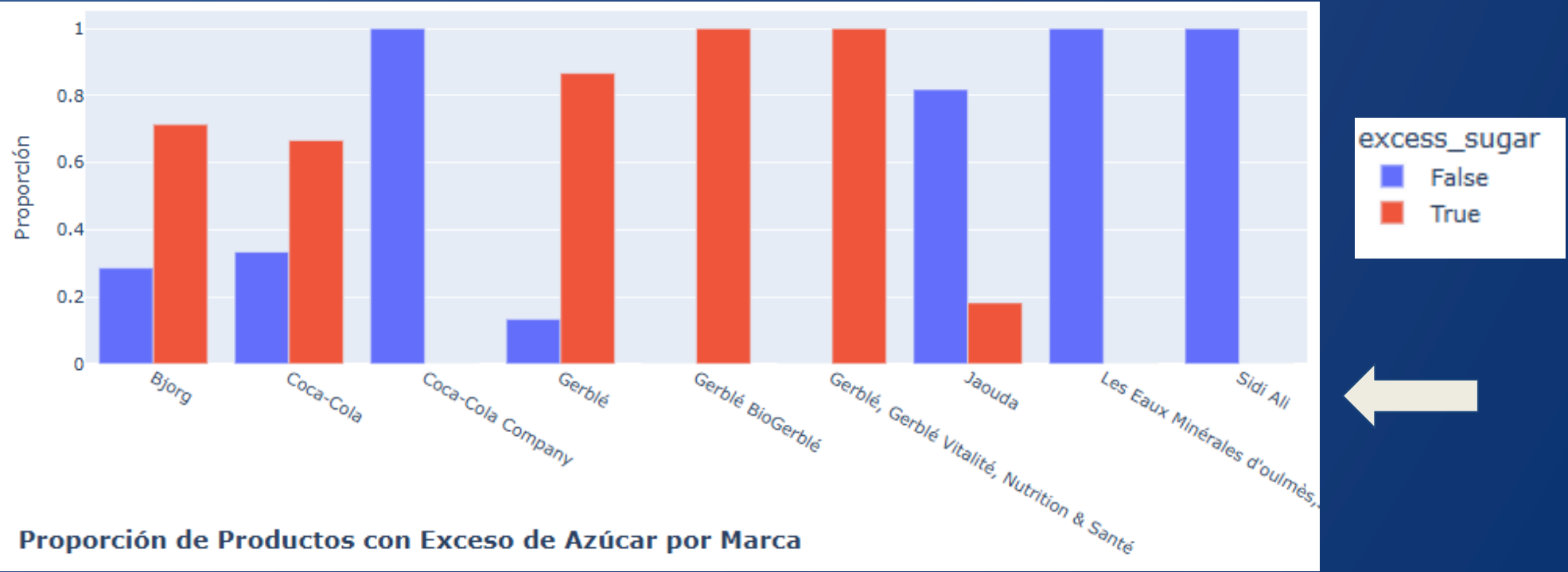
HIPÓTESIS 3: Las categorías presentan perfiles nutricionales distintos y predecibles.

Verdadero, los hallazgos en el EDA apoyan fuertemente este resultado de NUTRIENTES CLAVE POR CATEGORÍA:.



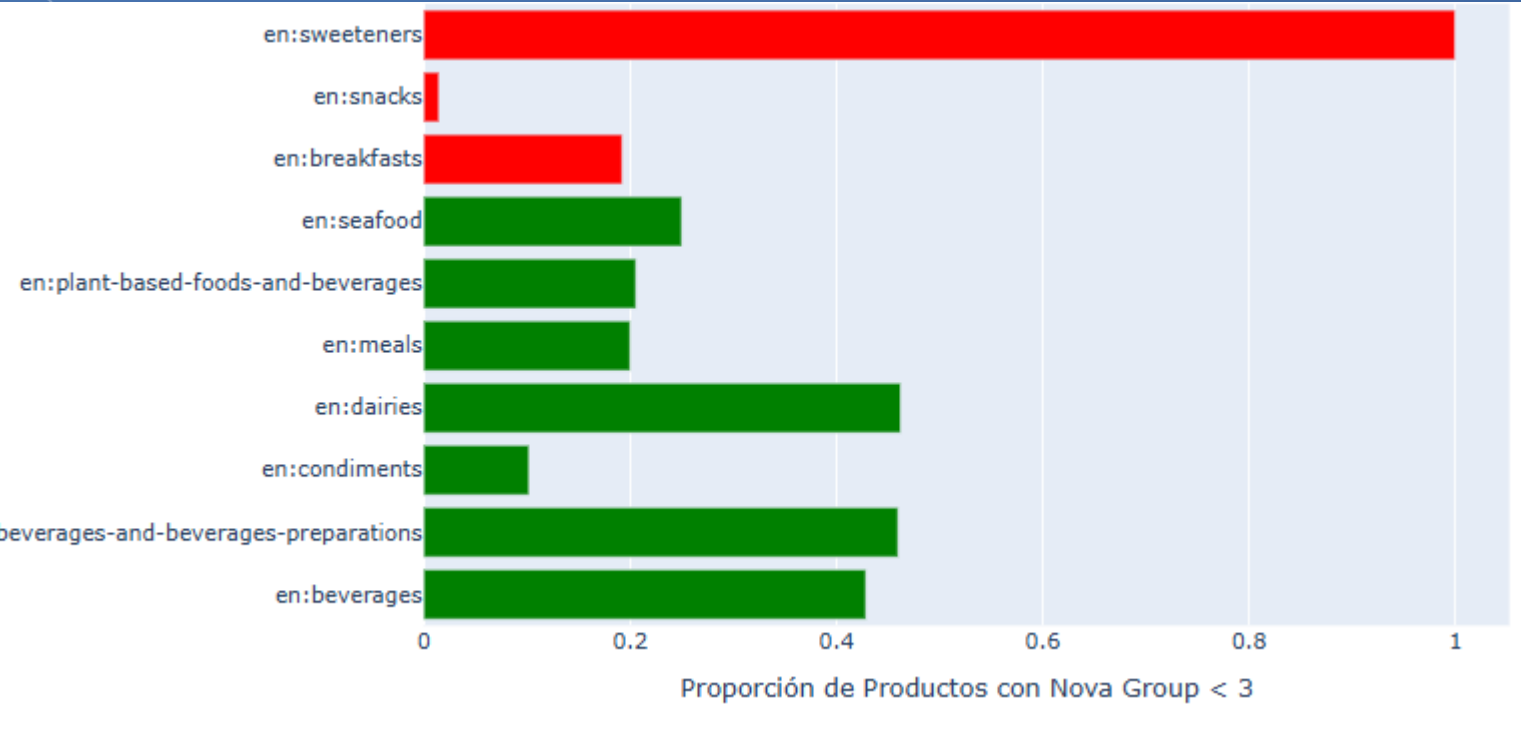
HIPÓTESIS 4: Sidi Ali presenta más oportunidades de mejora nutricional en azúcares que otras marcas.

Falso, Falso, Sidi Ali de hecho se encuentra entre las marcas con menores contenidos de azúcar y sin excesos de azúcares por normativa establecida por debajo de otras marcas que deberían tener prioridad en estas mejoras como Gerbié, Coca-Cola ó Bjorg.

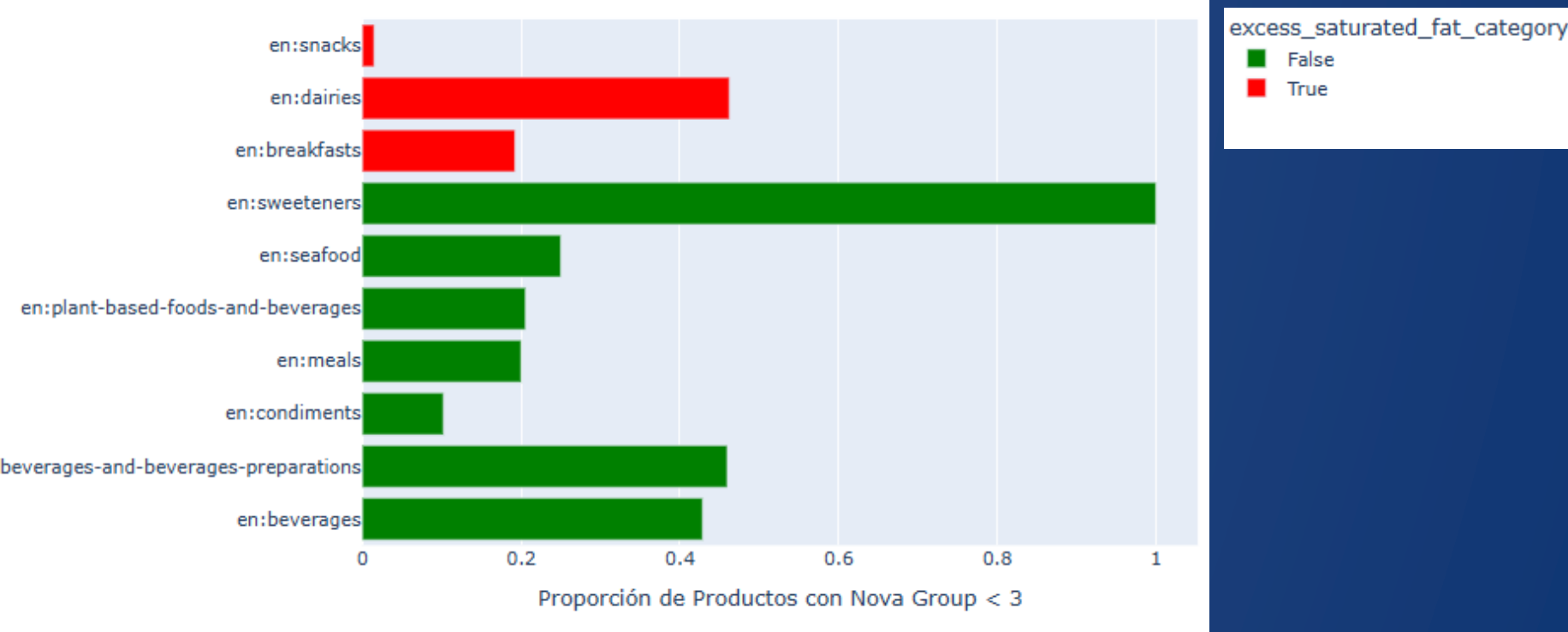


HIPÓTESIS 5: Los productos más saludables como Snacks y bebidas tienden a concentrarse en ciertas categorías específicas y tienen un Nova Group por debajo de 3.

Falso, Al agruparse, no se cumple esta condición en ningún contexto.



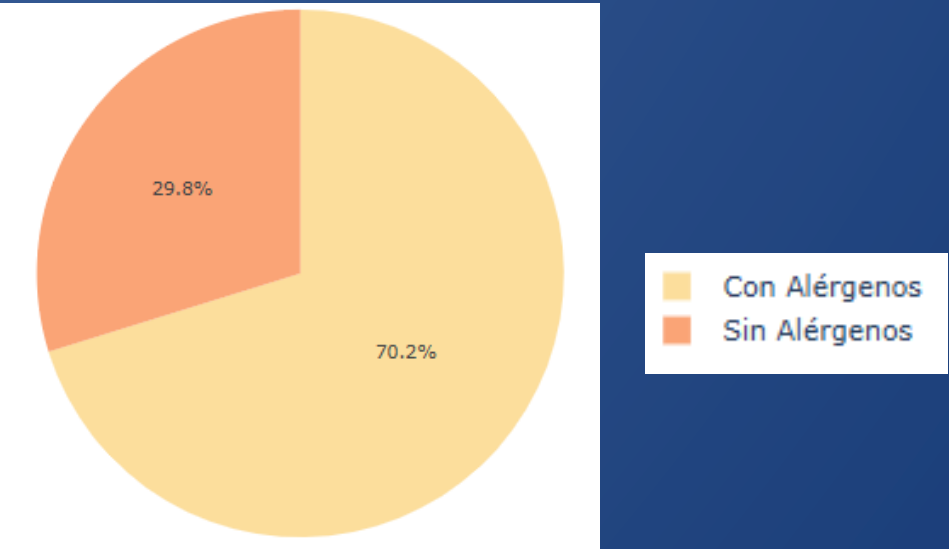
Proporción de Categorías con Nova Group < 3 Exceso de Azúcares



Proporción de Categorías con Nova Group < 3 Exceso de G. Sat

HIPÓTESIS 6: Todos los productos con Nutri Score D y E poseen alérgenos

Falso, Apenas el 70.2% de los productos con Nutri Score D y E tienen alérgenos, lo que aún está lejos de crear una tendencia real hacia esta Hipótesis con el catálogo actual de productos:



Proporción de Productos con Nutri Score D o E y Alérgenos

CONCLUSIONES FINALES



- **Clasificación por Perfil Nutricional:** Confirmamos que los productos tienen perfiles nutricionales únicos y predecibles. Nuestro modelo de clasificación automática, usando sólo datos nutricionales, puede predecir con éxito la categoría de un producto (validando H1).
- **Diferencias Nutricionales entre Categorías:** Se observaron diferencias significativas en los perfiles nutricionales entre categorías. Por ejemplo, las bebidas tienen menos proteínas y grasas que los snacks o postres (validando H2 y H3).
- **Concentración de Azúcares y Grasas:** Identificamos las categorías con mayores concentraciones de azúcares y grasas, información crucial para la salud pública y patrones de consumo.
- **Impacto de la Normativa Argentina:** Un número considerable de productos excede los límites de azúcares y grasas saturadas según la Ley de Etiquetado Frontal de Argentina.
- **Calidad de los Datos:** Los datos de la API de Open Food Facts son adecuados para este análisis.

INSIGHTS CLAVE



1. **Innovación Nutricional:** Hay una clara oportunidad para desarrollar versiones más saludables de productos, especialmente en categorías con altos niveles de azúcares y grasas como snacks y postres. Esto responde a la creciente demanda de opciones más sanas.
1. **Diversificación Estratégica:** Las categorías con perfiles nutricionales menos favorables son ideales para expandirse hacia líneas "light", "reducidas en" o "funcionales", alineándose con las preferencias del consumidor.
1. **Análisis por Marca:** Investigar la relación entre marcas y perfiles nutricionales podría revelar qué marcas tienen productos menos saludables o con más sellos de advertencia. Esto es útil para mejorar productos y analizar a la competencia.
1. **Futuras Investigaciones:** Los datos actuales son una base sólida para explorar hipótesis adicionales, como la conexión entre Nutri-Score, Nova Group y la presencia de alérgenos, aunque esto requiere un análisis más profundo.



THANK YOU

