Logotipo

Descripción generada automáticamente





# **Reporte de Comprensión del Negocio**

## **Proyecto: Predicción de Stock para Productos In and Out en Supermercados**

**Responsables:** Dra. Laura Hervert Escobar y Dr. Raúl Valente Velarde  
**Objetivo:** Predecir el nivel óptimo de stock de productos estacionales y reducir los costos de inversión y almacenamiento.

## **1. Determinación de los Objetivos del Negocio**

La industria retail se caracteriza por vender productos en pequeñas cantidades a múltiples clientes. En particular, este proyecto se centra en la gestión de inventario de productos **"In and Out"**, también llamados **productos de temporada o de eventos**. Estos productos experimentan fluctuaciones drásticas en la demanda debido a factores como estaciones del año, festividades o eventos específicos.

### Objetivos específicos del negocio:

1. **Optimización del inventario**: Minimizar el exceso de stock para evitar costos de almacenamiento elevados y pérdidas por caducidad o baja demanda.
2. **Reducción de costos**: Minimizar la inversión en productos que tendrán un rápido declive en demanda.
3. **Planeación de precios**: Ajustar los precios dinámicamente en función de la demanda prevista.
4. **Maximización de ventas**: Garantizar la disponibilidad de productos estacionales en los momentos clave para maximizar ingresos.

Los criterios de éxito de este proyecto serán evaluados por la **Dra. Laura Hervert Escobar y el Dr. Raúl Valente Velarde**, quienes supervisarán la alineación del modelo con los objetivos comerciales.

## **2. Valoración de la Situación**

Antes de definir los objetivos de minería de datos, es esencial evaluar la situación actual del negocio y los factores que pueden afectar el éxito del proyecto.

### Recursos Disponibles

* **Datos históricos de ventas** de productos estacionales.
* **Información de proveedores** y tiempos de entrega.
* **Conjunto de datos** con **400,000 registros en una base y 150,000 en otra**.
* **Variables clave** en los datos:
  + ID del producto
  + Ventas mensuales
  + Precio de venta
  + Ventas del mes anterior
  + Año
* **Herramientas tecnológicas**: Bases de datos, sistemas de gestión de inventario, plataformas de análisis de datos.
* **Experiencia de expertos** en retail y gestión de inventarios.

### Fuentes de Referencia

* [Inventory Planner: Seasonal Inventory](https://www.inventory-planner.com/seasonal-inventory/)
* [Tada Now: Inventory Management Best Practices](https://www.tadanow.com/blog/inventory-management-best-practices-to-cater-to-seasonal-demand)
* [Shopify: Seasonal Inventory](https://www.shopify.com/my/retail/seasonal-inventory)

## **4. Preparación de los Datos**

### 4.1. Limpieza de Datos

* Manejo de valores nulos.
* Identificación de outliers en ventas.
* Homogeneización de formatos.

### 4.2. Transformación de Variables

* Normalización/Estandarización de ventas y precios.
* Creación de nuevas variables si es necesario (ej. tasa de crecimiento de ventas).

### 4.3. Segmentación Inicial con Clustering

* Aplicar PCA para reducción de dimensionalidad.
* Aplicar Clustering (K-Means o DBSCAN) para agrupar productos según comportamiento.

## **5. Metodología de Aplicada**

El proyecto se apoya en modelos de **clustering** para segmentar los productos In and Out, con un enfoque en dos dimensiones clave:

### 1. Clustering basado en Precio:

* Agrupa productos con etiquetas de precio similares.
* Considera el margen de contribución de cada producto.
* La información obtenida de esta agrupación se enlaza con el clustering de tiendas.

### 2. Clustering basado en Tiendas:

* Segmenta las tiendas según sus patrones de venta de productos In and Out.
* Toma en cuenta factores como formato de la tienda y nivel socioeconómico, volumen de ventas.
* Se realiza por separado del clustering de precios, ya que cada análisis se usa en diferentes módulos del sistema de información.

Para este análisis, se ha implementado un enfoque de **clustering aglomerativo** con el método de enlace **WARD**, que minimiza la varianza dentro de cada clúster y es similar en objetivo a la función del algoritmo k-means.

Los resultados iniciales indican que:

* Se han analizado **163,017 productos In and Out** pertenecientes a **1,751 subcategorías**.
* Se han identificado agrupaciones óptimas con un coeficiente de **silhouette** promedio superior a **0.7** en la mayoría de los casos.

## **5. Elaboración del Plan de Proyecto**

Para alcanzar estos objetivos, se estructurará el proyecto en las siguientes fases:

1. **Comprensión del Negocio** (fase actual)
2. **Comprensión de los Datos**
   * Revisión de bases de datos disponibles
   * Análisis de calidad de datos
3. **Preparación de los Datos**
   * Limpieza y transformación de datos
   * Selección de variables relevantes
4. **Modelado**
   * Aplicación de modelos de predicción de series temporales y machine learning
   * Validación de modelos
5. **Evaluación**
   * Comparación de resultados con datos reales
   * Ajustes en el modelo según métricas de desempeño
6. **Despliegue**
   * Implementación del modelo en el sistema de gestión de inventarios
   * Monitoreo y ajuste continuo