



# Predicción de demanda

## Prueba técnica

En Sumz, como consultora analítica, nos enfrentamos al desafío de desarrollar modelos predictivos para clientes de diferentes sectores económicos. Para esta prueba trabajamos con un cliente del sector retail que necesita un modelo de estimación de demanda para su comercio. La disponibilidad de los productos es fundamental para ofrecer un excelente servicio al cliente, lo que implica considerar dos factores críticos: evitar la falta de stock y no excederse en el abastecimiento de unidades. Así, nuestro objetivo es construir una proyección de demanda precisa que permita optimizar el inventario.

Por lo tanto, te desafiamos a crear el mejor modelo de estimación de unidades vendidas para cada uno de los artículos y para cada uno de los días del año. Este modelo permitirá planear la frecuencia y el volumen de las compras necesarias para garantizar la disponibilidad de los productos requeridos sin incurrir en la rotura de stock o en el exceso de aprovisionamiento.

Nos gustaría ver tu enfoque en el desarrollo de este modelo y cómo abordarías los desafíos específicos de este caso. ¡Buena suerte, esperamos ver tus resultados!

## Aviso de confidencialidad

Este documento y toda la información contenida en él, incluidos los enunciados de la prueba y las bases de datos, son de carácter estrictamente confidencial y están destinados únicamente al uso del destinatario específico para el propósito de esta evaluación. Se prohíbe expresamente la divulgación, distribución o reproducción de este documento y cualquier información relacionada en cualquier medio o formato, ya sea electrónico, impreso o de cualquier otro tipo, sin el consentimiento previo y por escrito de la organización. Al participar en esta prueba de selección, usted se compromete a mantener en la más estricta confidencialidad todo el contenido y la información proporcionados, reconociendo que el incumplimiento de esta obligación puede resultar en acciones legales en su contra.

## Descripción

El objetivo de esta prueba es evaluar su capacidad para analizar y desarrollar modelos analíticos en un tiempo limitado. Se espera que no le dedique más de 10 horas para completar la prueba, lo que incluye el tiempo para documentar su proceso. Puede utilizar cualquier herramienta de su preferencia (Python, R, Spark, SAS Guide/Miner, SPSS, etc.) y recursos en línea, pero no se permite consultar directamente con otras personas a través de ningún medio.



Tiene libertad para realizar supuestos que considere necesarios, por favor enúncielos explícitamente en la solución. No es obligatorio utilizar todos los datos o variables proporcionados ya que esto depende de la estrategia que se adopte para abordar el problema. Tenga en cuenta que no existe una única solución.

## El dataset

El dataset integra la información de las ventas y atributos de producto para las fechas comprendidas entre 01/01/2020 y el 30/11/2022.

### Ficheros

- El dataset "catalogo\_productos.csv" contiene las características de todos los productos, como su tamaño, categoría, proveedor, entre otros. La clave principal en este dataset es el "id\_producto"
- El dataset "demanda.csv" se utilizará para construir un modelo predictivo que estime las ventas diarias por producto. Los datos abarcan desde el 01/01/2020 hasta el 30/11/2022. Este dataset proporciona información sobre la demanda diaria por fecha y por "id\_producto"
- El dataset "demanda\_test.csv" se empleará para aplicar el modelo construido y evaluar su precisión fuera de muestra. Los datos cubren el período del 01/12/2022 al 30/2/2023. Este dataset contiene dos columnas: fecha y "id\_producto"

### Variables

- date: momento del tiempo en el que se produce el evento.
- Id\_producto: número identificador del artículo.
- Categoría: Corresponde a la categoría del artículo
- Sub\_categoria: Toma valores de sub categorías del campo categoría
- Tamaño: toma los siguientes valores (pequeño, mediano, grande)
- Premium: Variable dummy. 0 no es premium y 1 es premium
- Marca\_exclusiva: Variable dummy. 0 no es exclusiva y 1 es exclusiva
- Estacional: variable dummy que identifica si el producto se vende todo el año o no. Vale 1 para productos que solo se venden en noviembre, diciembre y enero, y 0 si se venden todo el año.
- Nit\_proveedor: código de identificación del proveedor
- demanda: variable a predecir. Nos indica las unidades vendidas para cada día y cada artículo.

### Formato y escritura

Los datasets con formato csv tienen como estructura:



- Nombres de campo: Incluidos en la cabecera.
- Separador: ","
- Codificación: UTF-8.

## Notas

El 2 de julio de 2021 entró en operación una tienda de la competencia a pocos metros. El negocio notó una disminución de las ventas a partir de ese día.

## Preguntas a responder

- ¿Cuál fué el impacto de la apertura del competidor? Por favor cuantifíquelo
- ¿Las ventas tienen una tendencia creciente o decreciente? Excluya el impacto de la apertura del competidor del análisis. Se quiere entender si en el futuro se espera que las ventas crezcan, decrezcan o se mantendrán estables.

## Estos son los entregables que se esperan de la resolución de la prueba:

- Documento explicativo: Se espera un documento detallado que explique el proceso seguido para resolver la prueba, incluyendo todas las etapas que se hayan seguido para la solución de la misma. Si alguna fase del desarrollo analítico no se incluye en este documento, se asumirá que no se hizo.
- Código documentado: Se espera que el candidato entregue el código documentado que respalden el ejercicio analítico. Se puede usar cualquier formato para la entrega, pero se debe explicar cómo reproducir el análisis en el documento explicativo. Dado que el ejercicio es corto, se invita a entregar un único notebook incluyendo el código y el análisis (documento explicativo).
- Archivo CSV: Se debe entregar un archivo CSV con nombre "resultado\_prueba.csv" que contenga las predicciones del modelo sobre el dataset "demanda\_test.csv". Debe contener las columnas: date, id\_producto y demanda

La calidad del modelo generado y la técnica utilizada para su creación serán evaluadas y valoradas. Se compararán los valores predichos por el modelo con la demanda esperada de acuerdo a la estrategia de generación de los datos. La métrica utilizada será el RMSE.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Siendo:

- "n" el número de casos



- “ $y_i$ ” el valor real
- “ $\hat{y}$ ” el valor estimado

## Metodología de calificación

Nuestra metodología de calificación evalúa de manera integral las habilidades de los participantes en la prueba a través de cuatro etapas clave: Análisis Exploratorio de Datos, Modelamiento y Presentación de Resultados. Buscamos identificar candidatos con sólidos conocimientos de analítica y capacidad para resolver problemas de negocio para enfrentar los desafíos en el entorno empresarial actual. Valoramos tanto la competencia técnica como las habilidades de comunicación y presentación.

Detalles de los pesos:

- Análisis Descriptivo (30%)
  - Análisis exploratorio (10%)
  - Respuesta a las 2 preguntas mencionadas arriba (20%)
- Modelamiento (20%)
  - Creación y selección de atributos (4%)
  - Imputación de datos (2%)
  - Partición train/test (4%)
  - Selección de hiper parámetros (5%)
  - Evaluación del modelo (5%)
    - Visualizaciones de la demanda real y la predicha
    - Conclusiones
- Presentación de Resultados (10%)
  - Documentación clara y organizada (5%)
  - Comunicación efectiva de los resultados y hallazgos (5%)
- Error en dataset fuera de muestra (demanda\_test.csv) **(40%)**

¡Mucha suerte!