

---

# Cencosud Desafío Data Science

24 de Octubre del 2022

La siguiente prueba para el cargo de Data Scientist consiste en evaluar sus capacidades en el trabajo relacionado con series de tiempo, particularmente, series de tiempo de ventas de artículos. Para ello se facilita el archivo *datos\_ventas.csv* el cual tiene las siguientes columnas

- **fecha:** En formato YYYY-MM-DD, corresponde a la fecha que se realizó la transacción.
- **producto\_id:** Identificador del producto que se vendió.
- **venta:** Cantidad vendida del producto. Notar que en caso de ser un decimal, el producto en cuestión es a granel (Ej : Frutas).
- **rubro\_id:** Código correspondiente al rubro al que pertenece el producto.

Los datos entregados corresponden a ventas realizadas en una única tienda Jumbo en el período que comprende desde el 2021-01-02 hasta el 2022-01-09. El **objetivo final** de la evaluación es realizar una predicción **diaria** de las ventas de la semana del 17 al 23 de Enero del 2022, es decir, un valor para cada día de la semana y para cada producto (Notar el desfase de una semana entre los datos y el período deseado). Los resultados numéricos concretos de dicha predicción no son relevantes para la evaluación, si no más bien el desarrollo y su dominio que quedará reflejado en el trabajo.

Para la entrega se espera un notebook de jupyter que incluya el desarrollo del trabajo. Se valora el código limpio y el buen uso de celdas markdown, además se valora igualmente el manejo en Python que se despliegue en dicho notebook (Sobretudo como lenguaje orientado a objetos).

La entrega es libre, sin embargo se sugiere seguir la siguiente estructura, junto con responder las preguntas adjuntas (No es necesario responderlas todas):

1. **Análisis exploratorio de datos** (¿Qué rubro se lleva la mayor venta? ¿Qué series de tiempo se ven problemáticas a priori? Utilice gráficos adecuados.)
2. **Limpieza y preparación de datos** (¿Qué series podemos dejar afuera y bajo que criterio? Asegurarse de completar las fechas faltantes puede ser fundamental para el modelo. ¿Qué hacer si una serie contiene outliers?)
3. **Modelamiento** (¿Qué modelo propone para solucionar el problema? ¿Cómo sugiere entrenarlo? Es importante defender la propuesta, incluyendo los pros y los contra. De necesitar alguna transformación adicional a los datos, por necesidad del modelo, asegúrese de incluirlo en su análisis.)
4. **Entrega, evaluación de resultados y conclusiones.** (Presente el modelo ya entrenado con su predicción de ventas. ¿Puede estimar el error de su predicción? ¿Qué métrica o métricas considera adecuadas para evaluar este resultado? Tenga en consideración la posible heterogeneidad de las series. ¿Qué otros datos consideraría útiles para mejorar el modelo? ¿Cómo diseñaría una librería que implemente su solución? Los datos corresponden a una única tienda, ¿Cómo podríamos escalar el trabajo anterior para más tiendas?)