**Desafío Cargo Data Scientist - Carozzi**

**Antecedentes**

El objetivo es resolver un problema simulado utilizando datos disponibles, con el objetivo de medir capacidades en el contexto de la resolución de un problema y presentación de resultados en el proceso de postulación al cargo de Data scientist en Carozzi. Se espera que logres resolver el desafío utilizando los datos que se te van a compartir, complementándolo con otros datasets disponibles en la web si consideras necesario y realizando las actividades que se detallaran a continuación

Consideraciones: Para resolver este desafío, se deben tener presente los siguientes puntos:

* En base a tu experiencia debes plantear la metodología y forma de resolver el problema considerando las mejores prácticas
* Se debe estructurar una presentación de máximo 10 slides, con la presentación de los principales resultados del caso orientados a un comité mixto entre negocio/técnico. La presentación la puedes complementar con visualizaciones en herramientas de visualización que consideres complementarias.
* La presentación debe durar máximo 20 minutos, 20 minutos serán para preguntas y respuestas
* El código que respalde el análisis debe ser compartido para revisión posterior a presentación de resultados
* Se reciben entregables hasta el 4/9/2025 a las 17:00 hrs
* Puedes contactar vía mail/whatsapp/teléfono al contacto definido para este caso para resolver dudas:
  + **Contacto**: Felipe Contreras ([felipe.contreras@carozzi.cl](mailto:felipe.contreras@carozzi.cl) - +56945052321)

**Caso: Proyección de ventas por Store**

1. **Antecedentes dataset**

Data.csv = Corresponden a las ventas de salas de supermercados en Ecuador y tiene las siguientes columnas:



Stores.csv = Corresponde a información complementaria de las salas de supermercado y tiene las siguientes columnas:



Oil.csv= Corresponde al precio del petróleo de manera diaria y contiene las siguientes columnas:



1. **Objetivo del caso**

* Predecir el volumen de ventas de manera semanal, para cada store\_nbr, en un horizonte de medio semestre (26 semanas), desde la última observación presente en el dataset. (considera el periodo de entrenamiento/pruebas y la metodología de evaluación que estimes necesarios)
* Evaluar cómo se ven impactadas las ventas por factores como, precio del petróleo, promociones y otras variables relevantes que puedas crear o integrar y aporten valor al modelo

1. **Estructura para considerar en el desarrollo del caso** Para que el comité pueda reproducir y auditar tu trabajo, todo el desarrollo debe venir empaquetado en un único .zip y, a la vez, versionado en un repositorio Git (público o privado) cuyo enlace nos compartirás.
2. **Temas específicos para resolver: Se espera que el desarrollador sea capaz de responder mediante la resolución del caso al menos a las siguientes preguntas durante el desarrollo de su presentación del caso:**
   1. Monto a predecir: ¿Cuánto se estima que será el monto a vender, en los próximos 6 meses desde la última observación en los datos? Considere agregar semanalmente la serie y calcule además indicadores como % crecimiento proyectado en las ventas de este periodo y compárelo contra los crecimientos en los mismos periodos, pero en años anteriores
   2. Análisis del modelo: métricas de calidad, predicción, insights relevantes
   3. Análisis de Tendencias: ¿Qué patrones estacionales o tendencias observaste en los datos de ventas a lo largo del tiempo?
   4. Relación entre Promociones y Ventas: ¿Cómo influyen las promociones en el volumen de ventas? ¿Es este impacto constante a lo largo de todas las categorías de productos o varía en función de la categoría y la tienda?
   5. Precio del Petróleo y Comportamiento del Consumidor: ¿Existe alguna correlación significativa entre el precio del petróleo y las ventas? Si es así, ¿cómo podrías justificar este hallazgo en términos de comportamiento del consumidor o costo operativo?
   6. Recomendaciones para el Negocio: Con base en tus hallazgos, ¿qué estrategias recomendarías para optimizar las ventas durante eventos especiales o en tiendas que muestran una mayor sensibilidad a variables externas?