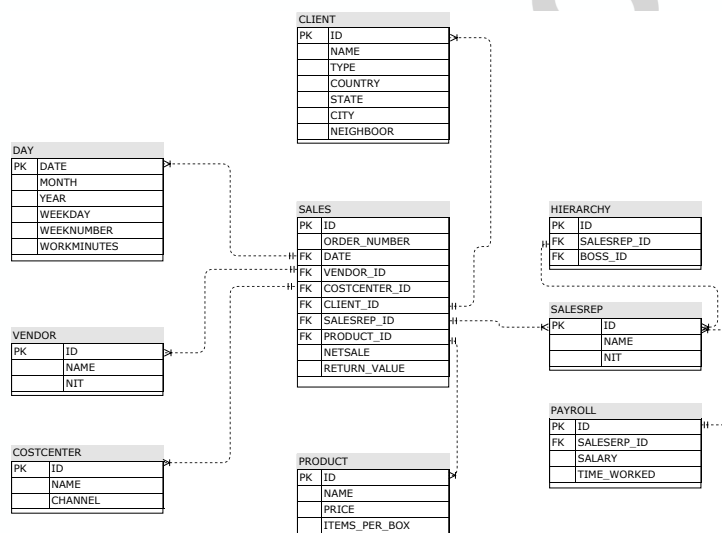


**Prueba técnica del proceso de selección**  
**Cargo: Científico de Datos (Data Scientist)**  
**Gerencia de Tecnología – Ingeniería y Analítica de Datos**  
**Altipal - Octubre de 2021**

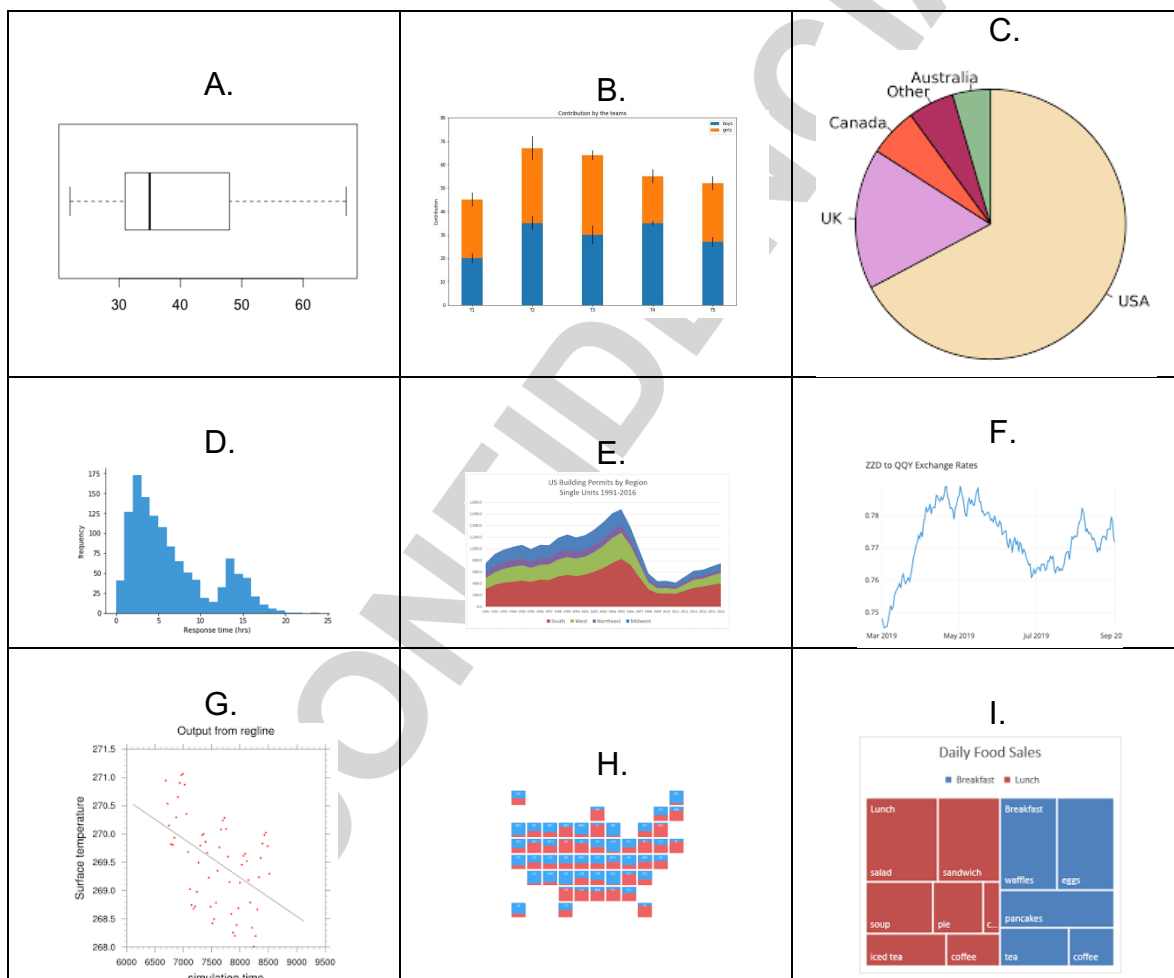
1. Indique y defina las funciones de agregación disponibles en SQL estándar
2. Defina los siguientes conceptos:
  - a. ETL
  - b. Data warehouse
  - c. Data lake
  - d. Data mart

Responda las siguientes preguntas usando el siguiente modelo de datos:



3. Construya una consulta que contenga los campos:
  - a. ORDER\_NUMBER
  - b. DATE
  - c. VENDOR\_NAME
  - d. CLIENTE\_NAME
  - e. CLIENT\_TYPE
  - f. CLIENT\_STATE
  - g. CLIENT\_CITY
  - h. SALESREP\_ID
  - i. PRODUCT\_ID
  - j. SALESREP\_BOSS\_ID
  - k. NETSALE
  - l. Unidades vendidas
  - m. Número de cajas/BOXES vendidas
  - n. Total de venta por ORDER\_NUMBER

4. Realice una consulta para mostrar:
  - Nombre y ID del vendedor concatenado
  - Antigüedad
  - ID del jefe
  - Salario
  - Ventas totales (neto – devoluciones) durante los últimos 12 meses
  - Número de clientes atendidos con ventas
5. Indique para que se utilizan los siguientes gráficos y qué información se puede presentar en ellos:



6. Desarrolle un modelo para la competencia de Kaggle Predict Future Sales (<https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data>).  
 Reporte de forma autocontenida su exploración de los datos, el modelo seleccionado justificando su selección y los resultados obtenidos usando las

métricas de la competencia. Determine si su modelo tiene un ajuste adecuado, si la calidad de los datos es suficiente y si variables adicionales pueden ayudar a mejorar su predicción.

7. Analizar la satisfacción de los clientes y los motivos de sus opiniones son tareas fundamentales en el posicionamiento de productos en el mercado. A diferencia de un modelo predictivo, este análisis corresponde a una tarea de investigación para identificar variables relevantes y efectos que impacten significativamente la percepción de los compradores. Usando la información de Olist (<https://www.kaggle.com/olistbr/brazilian-ecommerce>) construya un modelo para entender qué variables (dentro del control de la compañía) son las más relevantes para lograr la satisfacción del cliente. Utilice un modelo de sentiment analysis para determinar el estado de ánimo de los compradores.

8. Proponga de forma teórica soluciones a los siguientes problemas de negocio. Suponga que existen todas las áreas de negocio necesarias para definir reglas y aplicabilidad, y que se tienen todos los conjuntos de datos necesarios.

a. Actualmente existe un alto volumen de pedidos que son cancelados en las siguientes 48 horas después de la solicitud del cliente. Estos pedidos implican desgaste operativo en alistamiento, entrega y muchas veces una recogida. Proponga un sistema para reducir el desgaste operativo.

b. Los clientes suelen hacer varios pedidos en el transcurso de un ciclo de visitas de venta (15 días), esto implica alistar múltiples envíos de mercancía con camiones separados que van parcialmente llenos y que visitan el mismo punto de venta. Proponga un sistema de recomendación para sugerirle al asesor de venta que ofrezca otros productos de la canasta al cliente con el fin de consolidar los envíos y reducir el desperdicio en transporte.

c. En la compañía se tiene un altísimo número de clasificaciones de clientes debido a la caracterización de los establecimientos de comercio según su tamaño, portafolio de ventas y modelo de operación (On-trade, Off-trade). Sin embargo, esta clasificación es actualmente realizada por los asesores de venta al momento de la creación del cliente y por agilidad o desconocimiento suele seleccionarse una categoría incorrecta. Diseñe una solución para realizar verificación de la calidad de la información *a posteriori* para confirmar la categoría vigente y la categoría sugerida según su perfil de compras y demás valores propios del establecimiento.