Ejercicios:

1. Seleccionar los proveedores que han emitido órdenes de compra que solo poseen un ítem de producto tipo galletita (Galletitas Oreo, galletita opera, galletitas OBLEA, Galletita Sonrisa, etc.) o bien que hayan emitido órdenes de compras en el año 2019 y 2020 inclusive. **No se admitirán órdenes** de compras que no cumplan con la condición primera o **que hayan emitido ordenes solo en el 2019 o solo en el 2020**.

Emitió orden 2019 va en la tabla

Emitió orden 2020 tambien

1. Eliminar los registros de los proveedores que hayan tenido menos de 3 órdenes de compras en el año 2021.
2. Modificar el precio de venta de los productos a razón de 3 veces el precio de compra, solo para aquellos productos que estuvieron en alguna orden de compra cuya diferencia de días entre la "fecha entrega" y la "fecha de emisión" no supere los 3 días.
3. Insertar en una tabla (supuestamente ya creada), los registros provenientes de una consulta SQL que devuelva los 10 productos que hayan sido vendidos en el mes de noviembre de los años 2002, 2003 y 2004. La selección de estos productos debe estar ordenados según la ganancia real acumulada. En otras palabras, generar un ranking de los 10 productos que tuvieron mayor ganancia en noviembre de los años 2002, 203 y 2004, dejándolos almacenados en una tabla.
4. Actualizar el precio de venta al doble de los productos incluidos en órdenes de compra realizadas en abril.
5. Mostrar las descripciones que estén repetidas, y el precio de venta promedio de los mismos. Como resultado final se pide el producto y el precio de venta promedio únicamente. Esta es una relación de la tabla productos únicamente.
6. Borrar registro e historial de compra de los proveedores 514 y 831.
7. Crear una tabla con la estructura de cs\_productos, nombrarla como desee, e inserte los mismos registros que tiene la tabla cs\_productos. Luego actualizar el precio de costo en la nueva tabla con un aumento del 15% en productos con descripción de galletitas, un 5% para productos con descripción de chupetines y un 36% en barra de chocolates.
8. Crear una estructura procedural que recorra la tabla de productos y que por cada uno de los productos recorridos verifique la existencia del mismo en alguna orden de compra cuya fecha de emisión haya sido anterior al 01/01/2011. En caso de no existir el producto en alguna orden según la lógica antes mencionada, insertar los datos del producto en una tabla incluyendo los datos de código de producto, nombre y última fecha de orden de compra en la cual, incluido el producto, y a su vez disminuir el precio del producto al 90% del valor actual. Construir el bloque de código mediante sentencias PL/SQL de CURSORES, ASIGNACIONES, SELECT, IF, UPDATE, ETC.
10. ¿Qué diferencias hay entre clasificación y regresión? nombrar casos de uso para cada uno.
11. Una empresa de retail tiene problemas con productos que se vencen.

Para eso está interesada en saber cuántas ventas va a tener de determinados productos los próximos 6 meses.

¿Qué tipo de modelos se podrían usar? ¿Cómo medirías los resultados? que datos serán necesarios tener.

1. ¿Para qué se divide el dataset en train y test? ¿Cómo funciona cross-validation?
2. ¿Qué es un outlier, como se pueden detectar, porque es importante detectarlos?

RESOLUCIÓN EJERCICIO 10:

A) En base a lo estudiado en mi formación académica, específicamente, en la materia de Inteligencia Artificial, se desarrolla estos conceptos de clasificación y regresión. Para poder explicarlos, comenzaré a describir dos estrategias de aprendizaje dentro del aprendizaje automático: El aprendizaje supervisado y el aprendizaje no supervisado.

En Inteligencia Artificial hay dos estrategias de aprendizaje: supervisado y no supervisado. En el aprendizaje supervisado, al algoritmo le vamos a dar además de los ejemplos de entrada, también información adicional para que el algoritmo pueda encontrar la relación (el patrón) que le permita resolver el problema. Generalmente, esa información adicional puede ser una salida deseada o algo que le permita valorar si la salida que está generando es buena o mala. En el no supervisado, solo se dan datos de entrada, y a partir de esos ejemplos, el algoritmo va a tratar de determinar un modelo que los representa y de esa forma resolver un problema. La clasificación y regresión son dos tipos de algoritmos supervisados que buscan predecir variables discretas o continuas. La clasificación se usa para predecir variables categóricas, como clasificar si una imagen es de un perro o un gato, mientras que la regresión se usa para predecir variables continuas, como el precio de una casa en función de sus características (lo detallaré más abajo).

En resumen, la principal diferencia entre la clasificación y la regresión es que la clasificación se usa para predecir variables discretas o categóricas, mientras que la regresión se usa para predecir variables continuas.

Otro ejemplo de clasificación: Si se tienen datos de diferentes tipos de verduras. En este caso, se tendrían datos de entrenamiento que contienen imágenes de diferentes tipos de verduras, como tomates, zanahorias, zapallos, lechugas,etc, y se quiere predecir a que categoría pertenece una nueva verdura de acuerdo con las características como tamaño, color, etc. Se estaría realizando una nueva tarea de clasificación, en donde la nueva verdura se asignaría a una categoría ya antes predefinida como tomate, zanahoria, zapallo o lechuga, según la característica que tenga.

Ejemplo de regresión: Si se quiere predecir el precio de una casa en función de sus características, como el tamaño, la ubicación, el número de habitaciones, etc., se podría utilizar un modelo de regresión. A partir de un conjunto de datos de entrenamiento que contiene información sobre casas previamente vendidas y sus características, se podría construir un modelo que pueda predecir el precio de una nueva casa en función de sus características.

B) Se pueden utilizar regresiones lineales o temporales para predecir las ventas futuras de productos, dependiendo si se cree que las ventas tienen una tendencia o un patrón estacional. La regresión lineal se ajustaría a los datos históricos y buscaría una relación lineal entre las variables relevantes, como el tiempo y las ventas. Por otro lado, la regresión temporal puede ser útil para modelar las tendencias y patrones estacionales que se repiten en los datos a lo largo del tiempo.

Para medir los resultados del modelo de predicción, se pueden utilizar diferentes métricas, dependiendo del objetivo de la empresa y las características del modelo, pero podrían ser por ejemplo el error de predicción utilizando el error cuadrático medio (MSE) o el error absoluto medio (MAE) o el coeficiente de determinación (R²) para evaluar el ajuste del modelo de datos.

Es importante tener datos históricos de ventas, inventario, fechas de vencimiento y precios para entrenar y ajustar los modelos de predicción. También se podría utilizar un modelo de redes neuronales artificiales, pero cada modelo tiene sus propias fortalezas y debilidades y es importante elegir el modelo adecuado según el contexto de la empresa.

C) En el aprendizaje automático, el dataset se divide en dos conjuntos, el conjunto de entrenamiento y el conjunto de prueba, para evaluar el rendimiento de un modelo antes de su implementación en el mundo real. La finalidad de esta división es entrenar el modelo con el conjunto de entrenamiento y comprobar su capacidad para generalizar en el conjunto de prueba, que contiene datos que el modelo no ha visto antes. Es importante dividir el dataset de esta manera para evaluar el rendimiento del modelo de forma objetiva y precisa.

**Cross-Validation:**

Es una técnica que ayuda a evaluar el rendimiento de un modelo de aprendizaje automático donde se divide el conjunto de datos en varios grupos más pequeños y se utiliza uno como conjunto de prueba y los demás como conjunto de entrenamiento.

En cada ronda, se entrena el modelo en los subconjuntos de entrenamiento y se evalúa en el subconjunto de prueba, repitiendo este proceso varias veces, con diferentes subconjuntos de prueba en cada ronda, hasta que se hayan utilizado todos los subconjuntos como conjuntos de prueba al menos una vez.

Al final, se promedian todos los resultados obtenidos en cada grupo de validación cruzada para obtener una estimación más precisa y generalizable del rendimiento del modelo en datos que nunca ha visto antes.

D) Es un punto de dato que es muy diferente, o se aleja significativamente de la distribución general de los demás valores en un conjunto de datos. Pueden ser detectados usando gráficos o pruebas estadísticas y es importante detectarlos porque pueden afectar los resultados de un análisis o modelo de aprendizaje automático. Si no se detectan y se usan en un modelo, pueden hacer que éste no funcione bien en situaciones reales. Por lo tanto, es importante identificarlos y tratarlos antes de usarlos en un análisis o modelo.