

PRUEBA TÉCNICA PARA ANALÍTICO – GERENCIA DE SERVICIOS LOGÍSTICOS

Por: Alejandro Arango Giraldo

PUNTO 1

Pregunta 1

Consulta SQL:

```
query_p1 = ""
SELECT
    r.recurso_id,
    r.nombre_recurso,
    SUM(ps.cantidad_pedida) AS total_cantidad_pedida
FROM
    pedidos_sucursal ps
JOIN
    recursos r ON ps.recurso_id = r.recurso_id
WHERE
    ps.fecha_pedido >= DATE_ADD(CURDATE(), INTERVAL -3 MONTH)
GROUP BY
    r.recurso_id, r.nombre_recurso
ORDER BY
    total_cantidad_pedida DESC
LIMIT 5;
""
```

a.

Si observo que ciertos recursos son recurrentemente pedidos en grandes cantidades, recomendaría realizar un análisis exploratorio preliminar de los datos como primer acercamiento al problema.

En un principio, fijaría unas hipótesis iniciales que serían evaluadas estadísticamente para confirmarlas o rechazarlas. Para ello, caracterizaría las sucursales que representan la mayor demanda de estos recursos y los respectivos recursos identificados. Igualmente, evaluaría el comportamiento histórico de las solicitudes, utilizando técnicas de series de

tiempo para establecer si existe estacionalidad y reconocer tendencias históricas que puedan resultar relevantes.

Una de las hipótesis que plantearía sería que la mayor concentración de pedidos corresponde a tipos de sucursales grandes y en regiones de alta demanda (por alta demanda, me refiero a alto flujo de personal y clientes). Valiéndome de la tabla “sucursales”, determinaría la región y tipo de sucursal de las sucursales representativas utilizando las variables `region` y `tipo_sucursal` respectivamente. Esto permitiría realizar un análisis estadístico para confirmar o rechazar la hipótesis, pero, de igual forma, identificar casos anómalos y sesgos estadísticos. En el caso de que se confirme la hipótesis, los datos indicarán que las sucursales grandes en regiones de alta demanda deben ser priorizadas sobre las sucursales medianas o pequeñas en regiones de baja demanda. Sin embargo, el evaluar estos dos grupos bajo un mismo concepto resultaría en un sesgo a favor del primer grupo, lo que llevaría a pasar por alto sucursales del segundo grupo cuyo abastecimiento puede mostrarse prioritario en su contexto particular.

Por ejemplo, si la gran mayoría de sucursales grandes y medianas se encuentran concentradas en las regiones centrales, el abastecimiento de esta zona va a ser primordial según el análisis realizado. Al asumir una estrategia de abastecimiento que favorezca las regiones centrales sin considerar las regiones periféricas, aquellas sucursales cuya demanda, en el gran esquema de pedidos, sea insignificante, van a tener una prioridad reducida. No obstante, es importante tener en cuenta que estas sucursales pueden resultar ser las más difíciles de abastecer desde un punto de vista logístico (ya sea por las rutas de acceso, la distancia, o una estructura interna desorganizada que derive en solicitudes pequeñas, con un plazo corto entre ellas, y de carácter urgente cada una). Y, si a esto le adicionamos que la densidad de sucursales puede ser menor en estas zonas, algunas de estas pueden ser los únicos puntos de atención cercanos para suplir la demanda de la población local, dándoles un carácter crítico.

Con estas consideraciones en mente, pasaría a determinar las unidades de medida de los recursos más solicitados por región y tipo de sucursal, y, si cuento con la información, el monto monetario destinado a su compra y distribución. Ya que puede existir una discrepancia en las unidades de medida de los recursos, es fundamental llevar la demanda de estos a las mismas unidades, o por lo menos, ser consciente de esta diferencia. No es lo mismo comparar una solicitud de 10 unidades con una de 1 caja, la cual contiene 200 unidades. Asimismo, esta diferencia en unidades de medida se ve reflejada en el monto gastado por recurso. Si se solicitan 10 unidades de lápices A, cada uno a \$200, y 1 caja a \$100.000 que contiene 1000 unidades de lápices B, si únicamente se evalúa la tabla “pedidos_sucursal” y el campo “cantidad_pedida”, los datos pueden generar mal entendidos al identificar que hay mayor demanda de lápices A que de B, y que salen más rentables los primeros que los segundos, cuando esto no es cierto.

Como último punto del análisis exploratorio, emplearía visualizaciones, estadísticas descriptivas, y pruebas estadísticas para valorar las series de tiempo de los recursos de interés, y, de ser necesario, entrar al detalle de los tipos de sucursales y regiones. De esta manera, se puede determinar si los datos tienen patrones predecibles o si son dominados por ruido, si existen ciclos estacionales o influencias externas, si hay comportamientos atípicos relevantes, y que tipo de modelo puede ser empleado para pronosticar el comportamiento de los datos.

Al evaluar y compartir estos hallazgos con los stakeholders, haría las siguientes propuestas, las cuales estarían sujetas a los resultados previos y a las capacidades de gestión, almacenamiento y distribución del banco:

1. Optimización de Inventarios

- **Mantenimiento de Stock Mínimo por Demanda:** Definir un umbral mínimo de inventario para los recursos más solicitados, considerando factores como la demanda por tipo y región de sucursal, y ciclos estacionales que determinen el comportamiento de esta demanda. Esto asegura que las sucursales no queden desabastecidas y evita costos asociados con pedidos urgentes al mantener un “stock de emergencia” disponible.
- **Predicción de la Demanda:** Considerando los resultados del análisis histórico, se podrían crear modelos de predicción basados en Machine Learning para anticipar el consumo futuro de estos recursos y optimizar el stock disponible. Dependiendo de las características de las series temporales, y a que nivel de detalle existen los datos, se decidiría que modelo utilizar (SARIMAX, ARIMA, Prophet, etc).

2. Abastecimiento Centralizado

- **Consolidación de Pedidos:** En lugar de que cada sucursal haga pedidos individuales, la sede administrativa del banco podría centralizar las solicitudes por medio de flujos en aplicativos comunes para todas las sucursales, y encargarse del abastecimiento posterior al procesamiento de la solicitud (negociación y pedido al proveedor, pago, recepción, almacenamiento, y distribución de los recursos). De esta manera, se reducen costos, se negocian mejores precios con los proveedores, y se reducen las posibilidades de error en la gestión por parte de las sucursales.
- **Cadenas de Suministro Regionales:** Establecer centros de distribución regionales que puedan servir a varias sucursales cercanas, optimizando tiempos y costos de transporte. Como se mencionó en el ejemplo de la influencia de las regiones en la distribución, se deben optimizar las rutas y distancias entre el origen y destino de los recursos.

3. Optimización de Recursos

- **Alternativas Sustitutas:** Si se identifican recursos costosos o difíciles de conseguir, en colaboración con los negociadores que conciertan los contratos con los proveedores, se pueden ponderar otras alternativas que puedan cumplir la misma función con menor costo y mayor facilidad de adquisición. Pero si el precio y disponibilidad de los recursos se considera adecuado, es importante diversificar los proveedores para prevenir inconvenientes que se deriven del monopolio de un proveedor sobre el suministro de un recurso.

4. Análisis de Frecuencia y Temporadas

- **Frecuencia de Pedidos y Ciclos Temporales:** A partir del análisis histórico, identificar si los recursos se piden en grandes cantidades debido a una baja frecuencia de pedidos. Si es así, recomendar pedidos más regulares para evitar acumulaciones o excesos (siempre teniendo en cuenta las capacidades de almacenamiento y el presupuesto establecido para la sucursal). Si hay estacionalidad en la demanda, coordinar con las sucursales para planificar los pedidos en períodos de alta demanda y provisionarlos progresivamente.

5. Implementación de Sistemas Inteligentes

- **Automatización de Pedidos:** Diseñar un sistema automático que realice pedidos cuando los niveles de inventario de los recursos más demandados bajen del umbral mínimo establecido. Utilizando tecnologías como RFID o códigos de barras para monitorear en tiempo real los niveles de inventario y su consumo, se puede tener mayor control sobre el inventario y evitar pedidos urgentes.

6. Ineficiencias en las Sucursales

- **Revisar Procesos Operativos:** Realizar una campaña de capacitación con las sucursales que presenten comportamientos ineficientes en las solicitudes para replantear su manejo y gestión de los recursos.

Es importante tener en cuenta que estas propuestas van dirigidas a abordar la alta demanda de los recursos más solicitados. No obstante, también son aplicables en caso de presentarse problemas de abastecimiento para todos los demás recursos.

b.

Los siguientes son factores internos y externos que podrían explicar esta demanda en el último trimestre:

Factores internos:

- **Tipo de sucursal:** Sucursales grandes o con más empleados suelen requerir más recursos. Sin embargo, sucursales pequeñas podrían tener alta demanda para recursos clave propios de una operación específica que solo se lleve a cabo allí.

- **Fluctuación operativa:** Sucursales con meses de actividad inusual, ya sea por eventos internos, auditorias, campañas, o normativas especiales. Por ejemplo, en el cierre de año, donde los empleados de áreas contables deben cumplir con horas extra para cerrar la contabilidad, recursos como el café, papel, papel higiénico, entre otros, pueden aumentar su demanda por la permanencia prolongada de las personas en las sucursales.
- **Procesos ineficientes:** Procesos internos que hagan mal uso de los recursos, lo que, en consecuencia, aumenta la demanda de estos. O que de por sí, hagan uso de recursos completamente innecesarios si se emplearan tecnologías de vanguardia más eficientes.
- **Errores humanos:** El personal de las sucursales puede cometer errores en el formato de solicitud, resultando en una demanda superior a la real de un recurso.
- **Problemas con el sistema de centralización:** Pueden presentarse problemas técnicos en el flujo centralizador, derivando en datos erróneos o desactualizados.
- **Provisiones:** En los cierres contables de diciembre se solicita a los usuarios del banco (personal de sucursales y sede administrativa) que provisionen los recursos de enero en diciembre, aumentando considerablemente las solicitudes en este mes. Esta estacionalidad se ve representada en un pico en la demanda de recursos. De igual forma, puede que llegue el final de año, y algunas sucursales no hayan gastado la totalidad de su presupuesto, lo que las lleva a hacer compras de última hora bajo la suposición de que si no lo hacen, van a perder este dinero al cambiar de año.
- **Inventario desactualizado:** En la gestión interna del inventario de cada sucursal, se le puede pasar por alto al personal encargado hacer la actualización de la base de datos que indica la cantidad de un recurso dentro de la sucursal. En este caso, el campo “cantidad_disponible” de la tabla “inventario_actual” está desactualizado porque el personal responsable no ha realizado la gestión (esto se ve reflejado en el campo “fecha_actualizacion”). Estos datos erróneos llevan a los encargados de las solicitudes a pensar que hay suficientes unidades de un recurso cuando en realidad es lo contrario. Una vez corregida esta cifra, aumentan las solicitudes para un recurso que está a punto de agotarse y se requiere urgentemente.
- **Casos particulares:** Al hacer el análisis considerando el campo “fecha_pedido”, se pueden estar dejando por fuera solicitudes que se hicieron antes del trimestre actual pero que todavía no han sido gestionadas. Esto se puede deber a varios factores:
 - Acumulación de pedidos de meses anteriores resultado de ineficiencias en la distribución. Un retroceso en la distribución puede llevar a que la entrega de un recurso tarde más, causando que las sucursales, en su afán por recibirlo, hagan más solicitudes.
 - Demoras en la gestión interna de las solicitudes. Ya sea por saturación de capacidades dentro del área encargada de centralizar y procesar las solicitudes. O en las áreas encargadas de la negociación y vinculación de

proveedores cuando no existe un proveedor vinculado que pueda suplir esa demanda.

Factores externos:

- **Región de la sucursal:** Sucursales ubicadas en regiones de alta clientela, y, por ende, mayor concentración de empleados, pueden presentar demandas elevadas. Sucursales en regiones de baja clientela, pero retiradas de los centros de distribución y con rutas de difícil acceso pueden requerir periodos más largos para gastar sus recursos, pero una vez se acaban, el reabastecimiento puede ser complicado y tardarse más tiempo. Esto puede derivar en la acumulación de solicitudes para recursos urgentes. Asimismo, las propias condiciones climáticas de la región pueden influir en la demanda. Por ejemplo, en regiones frías, la demanda de café permanece elevada de manera uniforme a lo largo del año. Pero en regiones más templado, la demanda de café aumenta en el invierno y se reduce en el verano.
- **Estacionalidad y eventos especiales:** Festividades locales pueden llevar al aumento del flujo de clientes en las sucursales. Por ejemplo, en diciembre es de esperarse que más personas realicen más transacciones y requieran de la ayuda del personal de sucursales. Esto finalmente representa un aumento en la demanda de ciertos recursos.
- **Condiciones macroeconómicas:** Frente al prospecto del aumento de precios en los recursos (ya sea por inflación, escasez de recursos, recesión, o regulaciones a nivel nacional), las sucursales pueden aumentar sus solicitudes con la esperanza de ahorrar dinero.
- **Normativas regulatorias:** Cambios en las normativas locales pueden exigir un mayor uso de ciertos recursos para cumplir con los requerimientos.
- **Novedad:** Si alguno de los proveedores lanza un nuevo artículo de menor precio y mayor calidad, la novedad de este puede llevar a un repentino aumento en la demanda del recurso.
- **Adopción tecnológica:** La adopción de nuevas tecnologías puede aumentar el consumo de ciertos recursos que sostengan esta implementación.
- **Perfil del cliente:** En ciertas regiones, el perfil del cliente cambia. Por ejemplo, en las zonas rurales, donde la adopción tecnológica por parte de los campesinos es baja, habrá una mayor demanda de papel que de recursos digitales.
- **Crecimiento poblacional:** El propio crecimiento en la población de una región está relacionado en cierta medida con el aumento de los clientes, y, en consecuencia, de los recursos necesarios para satisfacerlos.

Pregunta 2

Consulta SQL

```
query_p2 = ""
SELECT
    ps.pedido_id,
    ps.sucursal_id,
    s.region,
    s.tipo_sucursal,
    ps.recurso_id,
    r.nombre_recurso,
    ps.cantidad_pedida,
    ps.cantidad_entregada,
    (ps.cantidad_pedida - ps.cantidad_entregada) AS diferencia
FROM
    pedidos_sucursal ps
JOIN
    sucursales s ON ps.sucursal_id = s.sucursal_id
JOIN
    recursos r ON ps.recurso_id = r.recurso_id
WHERE
    ps.fecha_pedido >= DATE_ADD(CURDATE(), INTERVAL -6 MONTH)
    AND ps.cantidad_entregada < ps.cantidad_pedida
ORDER BY
    diferencia DESC;
""
```

a.

Las propuestas sugeridas en la pregunta 1.a también engloban esta pregunta. Sin embargo, la medida que considero más pertinente y que puede tener el mayor impacto positivo en el abastecimiento sería una solución tecnológica que integre la digitalización del inventario, automatización y centralización de pedidos, y la predicción de la demanda.

Inicialmente, se debe equipar a las sucursales con herramientas tecnológicas (códigos de barras, escáneres o RFID) que les permitan hacer seguimiento en tiempo real de los movimientos del inventario (entradas, salidas y consumo). Con esta información actualizada, se pueden crear dashboards para el control y visualización del inventario a nivel administrativo dentro de las sucursales, e implementar dentro de estos paneles recordatorios para la actualización de la información (campos “cantidad_disponible” y “fecha_actualizacion” de la tabla “inventario_actual”).

Una vez digitalizado el inventario, se puede implementar un sistema de solicitudes automáticas personalizado para los ciclos de consumo de cada sucursal. Tomando en cuenta variables como el comportamiento histórico, el historial de recursos pedidos,

inventario actual, región y tipo de sucursal, tiempos de entrega, disponibilidad de inventario, cantidad de recursos mínima, capacidad de almacenamiento, entre otros, se configuran reglas para la generación automática de solicitudes, optimizando el inventario de la sucursal y previniendo la escasez de recursos. Estas solicitudes serían centralizadas por medio de un flujo de un aplicativo común (SAP, Bizagi, etc.) para que en la sede administrativa se encarguen de su gestión y posterior distribución.

Finalmente, en la sede administrativa se construiría un modelo analítico para la predicción de la demanda considerando el historial de solicitudes de cada sucursal. De esta manera, el ente centralizador se puede anticipar a los pedidos de meses posteriores y separar un presupuesto más preciso, gestionar negociaciones y vinculaciones (establecer cláusulas a largo plazo con los proveedores para prevenir contratiempos o cambios abruptos en los precios), adelantar pedidos, y preparar la estrategia de distribución.

Pregunta 3

Consulta SQL

```
query_p3 = ""
SELECT
    ia.sucursal_id,
    s.region,
    s.tipo_sucursal,
    ia.recurso_id,
    r.nombre_recurso,
    ia.cantidad_disponible,
    ia.fecha_actualizacion
FROM
    inventario_actual ia
JOIN
    recursos r ON ia.recurso_id = r.recurso_id
JOIN
    sucursales s ON ia.sucursal_id = s.sucursal_id
WHERE
    ia.cantidad_disponible < 5
ORDER BY
    ia.cantidad_disponible ASC;
""
```

a.

Para realizar una adecuada gestión de casos de inventario bajo, primero evaluaría la criticidad del caso puntual. Como he mencionado previamente, existen sucursales que, por su tamaño, demanda y región, son críticas y su abastecimiento debe ser prioritario.

Igualmente, se pueden presentar casos donde una sucursal tenga 1 o 0 unidades de cierto recurso crucial.

Una vez identificados los casos críticos, se definiría una estrategia de suministro adecuada, considerando los recursos, tiempos de entrega de los proveedores, rutas optimizadas, entre otros aspectos, y despacharía este inventario de la manera más ágil posible. Si alguno de estos casos tiene una urgencia mayor, ya sea por el tipo de recurso requerido, o porque la entrega está programada para una fecha posterior a que se agote, recurriría a la sucursal más cercana con suficiente inventario para reabastecer el caso crítico y mitigar el riesgo. Una vez gestionados, pasaría a los demás casos.

Tras abarcar este problema de inventario reducido, reevaluaría la estrategia correctiva implementada, identificando que esta no es óptima y representa dificultades logísticas y costos adicionales. La mejor manera de gestionar estos casos es asumiendo una estrategia preventiva y predictiva por medio de soluciones tecnológicas y análisis de datos, como la planteada en la anterior pregunta. Aunque la implementación de esta medida constituye retos e inversiones considerables, los beneficios a largo plazo supondrían retribuciones aún mayores.

b.

Considero que una política de reabastecimiento automatizada, complementada con una solución tecnológica robusta, tendría vastos beneficios para el banco. Algunas de las ventajas que encuentro derivadas de esta implementación son:

- **Mejora de la eficiencia operativa:** Se reducen los tiempos de espera al realizar solicitudes inmediatas a partir de datos en tiempo real. Estas solicitudes requerirían una intervención humana mínima, reduciendo al máximo los errores como pedidos duplicados, omisiones o cálculos errados. Esta liberación de capacidad da cabida a un mayor enfoque en tareas estratégicas que mejoren el servicio al cliente y la gestión de la sucursal.
- **Optimización de costos:** El monitoreo automático del inventario impide que se pasen por alto recursos que están a punto de agotarse, y, en consecuencia, se minimizan las compras de emergencia, cuyo costo suele ser mayor. Asimismo, las cifras certeras que arrojaría este sistema posibilitarían mejores negociaciones con los proveedores y rutas de transporte optimizadas.
- **Mejor experiencia del cliente:** Al impedir que las sucursales se queden sin recursos esenciales, se evitan interrupciones en las operaciones, pero, sobre todo, genera un efecto mariposa que beneficia al cliente. Si los empleados cuentan con todos los recursos necesarios para desempeñar sus labores correctamente, su ánimo se va a ver impulsado por esta abundancia. Este buen ánimo resulta en un mejor trato del cliente, el cual, a su vez, va a encontrar una sucursal que proyecta profesionalismo y donde no falta nada de lo que este pueda requerir.

- **Escalabilidad:** Una política y tecnología robusta se pueden adaptar al incesante crecimiento de la red de sucursales nacionales sin sobrecargar las capacidades del personal administrativo. El establecimiento de una nueva sucursal, aunque supone falta de datos para alimentar el modelo predictivo, no es una limitante para el sistema propuesto, ya que este, alimentado de datos de sucursales similares en la misma región o regiones semejantes, puede hacer pronósticos certeros hasta que haya suficiente información para personalizar el servicio.
- **Toma de decisiones basada en datos:** Con toda esta información a la mano, la sede administrativa puede tener una visión general del panorama de abastecimiento nacional, y tomar decisiones estratégicas que con una vista más limitada no sería prudente tomar.
- **Impacto medioambiental:** Al optimizar la gestión de inventario, las rutas, y las negociaciones, se reduce el desperdicio de recursos y la generación de residuos contaminantes ya sea directa o indirectamente.

Todos estos beneficios aportar a la consecución del objetivo primordial del banco: Promover el desarrollo sostenible para lograr el bienestar de todos.

Claramente, un proyecto de esta magnitud requiere de una elevada inversión inicial y su éxito depende de la adopción tecnológica, la calidad de los datos, la resistencia al cambio del personal de sucursal, y la integración a tecnologías existentes. No obstante, estos riesgos pueden mitigarse con un enfoque estratégico y pruebas rigurosas antes de la implementación completa.

Pregunta 4

a.

El objetivo del modelo sería predecir la cantidad de cada recurso que cada sucursal necesitará en los próximos meses. Tomando en consideración la información disponible y en la manera que se presenta, se trata de un problema de predicción basado en series temporales con múltiples variables, con la tabla “pedidos_sucursal” como insumo primario. Ya que el consumo de recursos depende de factores históricos, estacionales, y específicos del negocio, se tomarían los datos desde 2021 hasta el presente para entrenar modelos que puedan identificar las tendencias temporales en el consumo de recursos por sucursal. Si no existen datos hasta el 2021, desde el 2023 es suficiente para tener pronósticos suficientemente precisos. Pero, si por el contrario, hay datos que van hasta años previos a 2021, estos se omitirían para evitar alimentar el modelo con los datos atípicos producto de la pandemia y datos que ya no son relevantes después de tantos años.

Antes de determinar el modelo ideal para estimar el consumo futuro de recursos por sucursal, establecería los recursos computacionales disponibles y la urgencia del proyecto. Con estos dos factores en mente, y después de un análisis exploratorio exhaustivo,

exploraría modelos de series temporales clásicos como el ARIMA, SARIMA y SARIMA, modelos de Gradient Boosting como el XGBoost o de RNN (Redes Neuronales Recurrentes) como LSTM o GRU. Sin embargo, con la experiencia que obtuve desarrollando un proyecto similar en la sección gestión de proveedores para la predicción del gasto de áreas usuarias para diferentes cuentas, me enfocaría en los modelos de series temporales, especialmente en modelo híbridos como el SARIMAX o el Prophet.

El SARIMAX maneja estacionalidad y tendencias, tiene una buena flexibilidad en la modelación de ruido por sus componentes AR (autoregresivo) y MA (media móvil), tiene un mayor grado de personalización en sus parámetros de entrada, admite variables exógenas que, en este caso en particular, pueden resultar fundamentales para mejorar las predicciones (días festivos, cierres contables, festividades, cambios en políticas nacionales, condiciones macroeconómicas, etc.), y finalmente, es óptimo cuando existen 2 o más años de datos históricos.

El Prophet es más fácil de desplegar que el SARIMAX, pero también maneja estacionalidad y tendencias, permite incorporar eventos específicos que puedan tener un impacto en la serie, es robusto frente a escenarios complejos, falta de datos, y outliers, y funciona bien con series temporales cortas, y claramente, mejor con series largas.

Determinar cual es mejor ya requiere de un análisis detallado de las métricas de evaluación y de cómo se adaptan al problema particular. Por ello, desplegaría ambos modelos y los enfrentaría uno contra el otro.

b.

Las variables principales del modelo serían: sucursal_id, recurso_id, fecha_pedido, y cantidad_pedido. Después de agrupar por sucursal_id y recurso_id, obtendría las series temporales de todos los recursos para todas las sucursales. La variable “fecha_pedido” representaría el componente temporal y “cantidad_pedido” es la variable objetivo que se busca predecir. Dependiendo de si la predicción se requiere a nivel de día, semana, mes, o trimestre, se agruparía la variable objetivo para reflejar esta cantidad. Ya que “fecha_pedido” está en formato “date”, lo más probable es que la frecuencia de los pedidos esté a nivel de día. En el caso que las predicciones se requieran diarias, para los días que no haya solicitudes, a estos se les asignaría una “cantidad_pedido” igual a 0.

Adicional a estas variables, tanto el SARIMAX como el Prophet admitirían variables referentes a eventos especiales, lo que permitiría integrar días festivos, festividades, periodos de implementación de normativas especiales, cierres contables, accidentes o disturbios de orden público, etc.

Para el modelo SARIMAX se podrían agregar otras variables exógenas como complemento a los eventos especiales. Algunas de las variables que considero serían relevantes y pueden ser extraídas de las tablas son: cantidad_disponible (cantidad del recurso disponible en la

fecha del pedido si es a nivel de día), fecha_actualizacion (también solo aplica si es a nivel de día), region, y tipo_sucursal. Variables que no se encuentran en la tabla, pero pueden resultar relevantes, también deben ser consideradas al punto de no saturar el modelo con ruido o causar sobre ajuste, pero si aportar características que mejoren la predicción. Algunas de estas variables podrían ser: número de empleados por sucursal (puede ser redundante con “tipo_sucursal” dependiendo de como se determinan las categorías de “grande”, “mediana”, y “pequeña”), clientes atendidos en la fecha, presupuesto mensual o anual asignado, capacidad de almacenamiento, horas laborales o de atención de la sucursal, distancia al centro de abastecimiento (asumiendo que mientras más lejos estoy, mayor cantidad voy a pedir en una sola solicitud por el tiempo de envío), tipo de clima (puede afectar en cierta medida los patrones de consumo)

Factores específicos del negocio como cierres contables, provisiones presupuestarias, auditorias, o campañas internas podrían influir en la demanda de recursos, en adición a los demás factores internos y externos previamente formulados.

C.

Para evaluar el modelo identificaría primero cuantos recursos computacionales tengo a mi disposición, si puedo ejecutar el entrenamiento en la nube o paralelizándolo. Con base en esto seleccionaría una de estas dos opciones:

- **Train-Test Split Temporal:** Si cuento con pocos recursos, o el conjunto de datos es muy grande, seleccionaría esta opción. Esta implica dividir los datos de entrenamiento y prueba considerando la naturaleza temporal de estos. Por ejemplo, para el entrenamiento, tomaría los primeros 2 años en adelante hasta una fecha de 2 meses previos a la última fecha disponible. Estos últimos 2 meses serán el conjunto de prueba.
- **Rolling Window:** Si cuento con los recursos y el tiempo, y la data histórica es suficiente, aplicaría un enfoque de validación cruzada para evaluar el modelo en múltiples horizontes de predicción, dividiendo la serie en un número determinado de subconjuntos.

Claramente, las evaluaciones de las predicciones dependerían del horizonte de interés. Si es a largo plazo (meses o años), o a corto plazo (días o semanas), esto lo determina la estrategia de distribución que se esté implementando.

Las métricas de rendimiento que usaría serían: MAE, RMSE, MAPE y R^2 . El MAE, MAPE y R^2 serían suficientes si la serie tiene valores estables y sin outliers, pero en el caso de que se presenten casos atípicos, el MAPE, que penaliza los errores grandes, sería una métrica importante para considerar. Adicionalmente, todas estas métricas las compararía con un modelo base (baseline) como una media móvil o el último valor de la serie, para asegurarme de que el modelo aporta valor.

d.

Inicialmente, realizaría un diagnóstico de los resultados, identificando aquellos recursos y sucursales con baja precisión. Analizando los resultados, buscaría si existen patrones en los errores y establecería una lista de posibles causas. Aunque solo son suposiciones, estimo que las fuentes de imprecisión están relacionadas en cierta proporción con algunos de estos motivos:

- **Datos insuficientes o irregulares:** Para ciertos recursos o sucursales puede haber pocos datos históricos o brechas significativas en los pedidos. Por ejemplo, un recurso recientemente introducido al inventario de una sucursal no tendrá suficiente historial para capturar patrones claros. O una nueva sucursal que tan solo lleva unos pocos meses en funcionamiento, y sus patrones de consumo no se han establecido. De igual forma, una sucursal pudo haber cerrado por un tiempo, o un recurso discontinuado por algún problema con el proveedor, pero después se reincorporó, generando datos irregulares.
- **Variabilidad alta o comportamiento atípico:** Algunos recursos o sucursales pueden tener demandas altamente volátiles, estar influenciados por eventos inesperados, o tener patrones estacionales complejos que el modelo no logra capturar.
- **Falta de consideración de factores externos:** Puede que en la selección de variables exógenas se hayan omitido algunas que puedan tener un gran impacto en la predicción.
- **Problemas con el modelo:** Puede que el modelo seleccionado no sea el adecuado para el tipo de datos, o que los parámetros utilizados no sean los óptimos para el problema.

Una vez identificado el o los motivos de bajo rendimiento, estos se deben abordar de manera individual y transversal.

Para los datos insuficientes o irregulares, primero verificaría que no haya otras fuentes de información que me puedan suministrar estos datos completos. En el caso de los datos irregulares, evaluaría el modelo imputando los datos faltantes con base a los datos previos y posteriores de la brecha. Si no resulta conveniente este enfoque, si la brecha es muy grande, eliminaría los datos previos a esta y solo trabajaría con los datos posteriores. Si son brechas pequeñas, considero que no tendrían un impacto muy significativo en las predicciones (ya que puede deberse a estacionalidad). Para los datos insuficientes (recursos o sucursales nuevas), entrenaría un modelo de clustering considerando diversas variables propias de los recursos y las sucursales, para que, de esta manera, se puedan simular datos históricos con cierto grado de fidelidad a la realidad para recursos y sucursales nuevas basándose en otros grupos similares.

Para variabilidad alta y comportamiento atípico, probaría inicialmente normalizando los datos y aplicándoles alguna transformación de ser necesario (aunque SARIMAX y Prophet ya hacen las transformaciones internamente), y evaluaría los outliers según el modelo usado (SARIMAX es sensible a estos, mientras que Prophet no). Si esto no funciona, segmentaría los recursos o sucursales que presentan este comportamiento y utilizaría modelos más robustos o un híbrido de varios.

Para la falta de consideración de variables externas, simplemente me remitiría a las variables planteadas en el ítem **b** y obtendría información histórica sobre estas para alimentarlas al modelo como variables exógenas.

Para problemas con el modelo, emplearía métodos iterativos para la selección de hiperparámetros como Grid Search o Random Search, o, en el caso de estar usando SARIMAX, cambiaría al modelo AutoARIMA, el cual hace esta optimización de hiperparámetros internamente. Si esto no da resultado, evaluaría otros modelos como XGBoost o LSTM.

e.

Asumiendo que el sistema de solicitudes ya se encuentra centralizado, y que esta información es instantáneamente almacenada en la LZ, empleando el orquestador se programaría la descarga de los insumos relevantes utilizando queries de SQL. Posteriormente, se diseñaría un pipeline automatizado que incluya tanto la descarga como el preprocesamiento y limpieza de los datos. Se calendarizaría este subproceso y se emplearían estos insumos para reentrenar el modelo cada cierto periodo de tiempo (dependiendo del flujo y disponibilidad de los datos).

Dependiendo de la estrategia de distribución activa, se programaría la generación de pronósticos cada día, semana o mes. Estos resultados serían presentados en un tablero de PowerBI para la toma de decisiones, y, de ser posible, tras las debidas verificaciones y pruebas, conectarse directamente con el sistema de pedidos para que las solicitudes se hagan automáticamente con base a los pronósticos (no sería del todo automático, pues se requieren validaciones humanas para evitar predicciones erróneas que supongan un gasto de capital muy elevado). Adicional al modelo, se pueden agregar alertas y notificaciones tanto para las sucursales como para el área administrativa sobre tiempo estimado antes de que se acabe un recurso crítico, cantidad de inventario disponible, última fecha de actualización del inventario, fechas estimadas de entrega posteriores a que se agote un recurso, etc.

Una vez se tenga este sistema bien estructurado y formalizado, se pueden realizar pruebas piloto con un subconjunto de sucursales con características variadas para exponer el modelo a casos variados. Durante el periodo de prueba (3-6 meses) se compararían las predicciones con los consumos reales, identificando problemas de precisión, y reajustando y retroalimentando el modelo. Una vez se aprueben los resultados de la prueba piloto y se

hagan las correcciones necesarias, se crearían dashboards que funcionen como reportes en tiempo real, tanto para los encargados de la distribución en la sede administrativa, como los dirigentes y stakeholders de cada sucursal. A medida que se validen los resultados y la solución se vuelva más robusta, se llevaría a cabo una expansión gradual a las demás sucursales.

En todo momento, el modelo sería monitoreado, reentrenado, y se le haría el mantenimiento pertinente. A la vez, se capacitaría al personal de las sucursales y de la sede administrativa para interpretar los dashboards y tomar decisiones sobre la información allí desplegada. Para la toma de estas decisiones se tendría que definir un protocolo y documentación clara para prevenir contratiempos y mejorar la eficiencia.