

Analítica para los negocios

06327-ECO

Semana 05: De la necesidad a la pregunta — y de la pregunta al producto

Eduard F. Martínez González, Ph.D.

Departamento de Economía, Universidad Icesi

1 de septiembre de 2026

Objetivo de la sesión

Hoy no abrimos RStudio: hoy practicamos **el oficio de pensar como analista**.

- Vamos a recorrer **dos proyectos reales de consultoría** — dirigidos desde el centro de investigación de la universidad — desde la necesidad vaga del cliente hasta el producto final entregado.
- En cada paso vamos a **reconocer los conceptos del podcast**: los tres criterios de una buena pregunta, el test de la pregunta, las siete etapas y los tipos de tarea analítica.
- Los dos casos arrancan en puntos **opuestos**: en uno *no existían los datos*; en el otro *sobraban*. Esa diferencia lo cambia todo.
- Al final, el método queda en sus manos: **leer el caso Cóndor, entender sus datos y proponer la pregunta de negocio de su grupo** — el insumo directo de la Entrega 1.

La idea que gobierna la sesión

Un proyecto de analítica no es exitoso porque el modelo acierte: es exitoso cuando **cambia una decisión**. Hoy lo veremos dos veces, con proyectos que sí ocurrieron.

¿Dónde estamos en el curso?

- **Semanas 1–4:** las herramientas. R y los scripts reproducibles, dplyr para transformar, ggplot2 para graficar. La semana pasada fuimos *del dato a la gráfica*.
- **Semana 5 (esta):** el mapa. El proceso analítico completo: de la pregunta a la decisión. La semana 4 entera era apenas *una* de las siete etapas.
- **Hoy en clase:** el mapa aplicado dos veces, con casos reales — y luego ustedes lo aplican por primera vez a su proyecto.
- **Semana 6:** EDA y calidad de datos — la etapa donde se va el 80% del tiempo.
- **Semana 7:** examen parcial 1 (unidades 1 a 5).

El proyecto final arranca hoy

La **Entrega 1** pide exactamente lo que vamos a practicar: la pregunta de negocio de su grupo sobre el caso Cóndor. Lo que salga de la clase de hoy es su materia prima.

Roadmap

- 1 El mapa que ya tienen (repaso en una lámina)
- 2 Caso A — Empresa P: predecir un mercado que no está en los datos
- 3 Caso B — Empresa M: vigilar una red que produce datos sin parar
- 4 El contraste: dos puntos de partida, dos productos
- 5 Su turno: el caso Cóndor

Roadmap

- 1 El mapa que ya tienen (repaso en una lámina)
- 2 Caso A — Empresa P: predecir un mercado que no está en los datos
- 3 Caso B — Empresa M: vigilar una red que produce datos sin parar
- 4 El contraste: dos puntos de partida, dos productos
- 5 Su turno: el caso Cóndor

Todo el podcast, en una lámina

Una buena pregunta de negocio

- 1 **Específica** — ¿qué, quién, cuándo?
- 2 **Accionable** — ¿qué decisión cambia?
- 3 **Medible** — ¿existen datos para responderla?

El test de la pregunta

Si te doy la respuesta ahora mismo, **¿sabes qué vas a hacer diferente mañana?** Si no: tienes un tema, no una pregunta.

Las siete etapas (y un ciclo)

pregunta → datos → limpieza → EDA → modelo → comunica → decide → (y de vuelta)

Las cinco tareas analíticas

Tarea	¿Target?	Output
Clasificación	sí	categoría
Predicción	sí	número
Segmentación	no	grupos
Anomalías	mixto	lista de raros
Optimización	—	decisión

Hoy

Dos recorridos completos por este mapa.
Su trabajo: ir marcando *en qué etapa estamos y qué tarea es.*

Roadmap

- 1 El mapa que ya tienen (repaso en una lámina)
- 2 Caso A — Empresa P: predecir un mercado que no está en los datos
- 3 Caso B — Empresa M: vigilar una red que produce datos sin parar
- 4 El contraste: dos puntos de partida, dos productos
- 5 Su turno: el caso Cóndor

Caso A · El contexto

Empresa P es una multinacional de **empaques de cartón**. Una parte importante de su negocio depende de un cliente muy concreto: **los exportadores de fruta** de América Latina — cada tonelada de fruta que sale del continente viaja en sus cajas.

El problema que la desvela: el cambio climático está moviendo el mapa frutícola mundial. Zonas que hoy producen dejarán de hacerlo; otras que hoy no pesan, despegarán. Y las inversiones de la empresa (plantas, capacidad, relaciones comerciales) se deciden *años antes*.

Caso A · El contexto

Empresa P es una multinacional de **empaques de cartón**. Una parte importante de su negocio depende de un cliente muy concreto: **los exportadores de fruta** de América Latina — cada tonelada de fruta que sale del continente viaja en sus cajas.

El problema que la desvela: el cambio climático está moviendo el mapa frutícola mundial. Zonas que hoy producen dejarán de hacerlo; otras que hoy no pesan, despegarán. Y las inversiones de la empresa (plantas, capacidad, relaciones comerciales) se deciden *años antes*.

La necesidad, tal como llegó

“Queremos entender hacia dónde debemos orientar nuestro trabajo futuro en este mercado.”

¿Qué tiene de vaga? No dice *qué* medir, ni *dónde*, ni *a qué horizonte*, ni *qué decisión* está esperando la respuesta. Es una inquietud legítima — pero todavía no se puede trabajar. **Convertir esto en una pregunta es la etapa 1 del proceso.**

Caso A · La pregunta de negocio

De la necesidad vaga... a la pregunta que dispara el proyecto

*“¿Cómo evolucionarán las **exportaciones de fruta** (toneladas), **por país y por tipo de fruta**, hacia **2045** bajo distintos **escenarios de temperatura** — y dónde crecerá o caerá la demanda potencial de empaques?”*

Caso A · La pregunta de negocio

De la necesidad vaga... a la pregunta que dispara el proyecto

*“¿Cómo evolucionarán las **exportaciones de fruta (toneladas)**, **por país y por tipo de fruta**, hacia **2045** bajo distintos **escenarios de temperatura** — y **dónde** crecerá o caerá la demanda potencial de empaques?”*

¿Cumple los tres criterios?

- **Específica** — qué: exportaciones en toneladas; quién: 14 países productores y competidores, fruta por fruta; cuándo: horizonte 2025–2045.
- **Accionable** — prioriza *dónde* enfocar la estrategia comercial y qué mercados vigilar. Test de la pregunta: con la respuesta en la mano, mañana la agenda cambia.
- **Medible** — *casi*: hay historial de exportaciones y hay escenarios climáticos... pero **nadie los ha juntado**. Ahí está todo el proyecto.

Caso A · La pregunta de negocio

De la necesidad vaga... a la pregunta que dispara el proyecto

*“¿Cómo evolucionarán las **exportaciones de fruta (toneladas)**, **por país y por tipo de fruta**, **hacia 2045** bajo distintos **escenarios de temperatura** — y **dónde crecerá o caerá la demanda potencial de empaques?**”*

¿Cumple los tres criterios?

- **Específica** — qué: exportaciones en toneladas; quién: 14 países productores y competidores, fruta por fruta; cuándo: horizonte 2025–2045.
- **Accionable** — prioriza *dónde* enfocar la estrategia comercial y qué mercados vigilar. Test de la pregunta: con la respuesta en la mano, mañana la agenda cambia.
- **Medible** — *casi*: hay historial de exportaciones y hay escenarios climáticos... pero **nadie los ha juntado**. Ahí está todo el proyecto.

¿Qué tarea analítica es?

Output = **número** (toneladas futuras) + historial con la respuesta correcta ⇒ **predicción, supervisada**.

Caso A · El problema: la respuesta no estaba en ningún dataset

No existía ninguna base de “exportaciones frutícolas condicionadas al clima”.
Hubo que construirla desde fuentes abiertas — la etapa 2 fue el corazón del proyecto:

El target (Y)

Exportaciones de fruta en **toneladas**, por país–fruta–año: series históricas de **FAO**.

Las explicativas (X)

Temperatura histórica y **proyectada** por país: portal climático del **Banco Mundial** (CCKP), bajo **tres escenarios** de emisiones (SSP1-1.9 optimista · SSP2-4.5 moderado · SSP5-8.5 intensivo).

El resultado: un **panel país–fruta–tiempo** armonizado — con rezagos, tendencias y transformaciones construidas a mano.

Caso A · El problema: la respuesta no estaba en ningún dataset

No existía ninguna base de “exportaciones frutícolas condicionadas al clima”. **Hubo que construirla** desde fuentes abiertas — la etapa 2 fue el corazón del proyecto:

El target (Y)

Exportaciones de fruta en **toneladas**, por país–fruta–año: series históricas de **FAO**.

Las explicativas (X)

Temperatura histórica y **proyectada** por país: portal climático del **Banco Mundial** (CCKP), bajo **tres escenarios** de emisiones (SSP1-1.9 optimista · SSP2-4.5 moderado · SSP5-8.5 intensivo).

El resultado: un **panel país–fruta–tiempo** armonizado — con rezagos, tendencias y transformaciones construidas a mano.

La regla 80/20, en vivo

La mayor parte del esfuerzo se fue en **encontrar, integrar y validar** datos de fuentes distintas. El modelo vino después — y fue lo más rápido.

Caso A · El análisis

- 1 **Modelos supervisados** entrenados sobre el panel: Random Forest, XGBoost y una red neuronal — tres algoritmos compitiendo por capturar la relación clima–exportaciones (que no tiene por qué ser lineal).
- 2 **Validación fuera de muestra con enfoque temporal:** se entrena con el pasado y se evalúa contra años que el modelo *no vio*, respetando el orden del tiempo. Métricas de error de predicción (MAE, RMSE, MAPE) por fruta y país.
- 3 **Selección del mejor modelo** por desempeño y estabilidad — no por sofisticación.
- 4 **Proyección:** al modelo ganador se le inyectan las trayectorias de temperatura de cada escenario SSP, y produce las exportaciones esperadas hacia 2045.

Caso A · El análisis

- 1 **Modelos supervisados** entrenados sobre el panel: Random Forest, XGBoost y una red neuronal — tres algoritmos compitiendo por capturar la relación clima–exportaciones (que no tiene por qué ser lineal).
- 2 **Validación fuera de muestra con enfoque temporal**: se entrena con el pasado y se evalúa contra años que el modelo *no vio*, respetando el orden del tiempo. Métricas de error de predicción (MAE, RMSE, MAPE) por fruta y país.
- 3 **Selección del mejor modelo** por desempeño y estabilidad — no por sofisticación.
- 4 **Proyección**: al modelo ganador se le inyectan las trayectorias de temperatura de cada escenario SSP, y produce las exportaciones esperadas hacia 2045.

Conecta con la teoría

“No se evalúa un modelo con los datos con los que se entrenó”: aquí el examen de verdad es predecir años que el modelo no vio. (Volvemos a esto en la semana 10.)

Caso A · El producto final

Un **informe prospectivo ejecutivo**: resumen gerencial, escenarios, y un cuadro que prioriza las combinaciones país–fruta con mayor expansión y mayor contracción a 2045. Una muestra real (escenario moderado SSP2-4.5, variación 2045 vs. promedio 2020–2024):

País	Fruta	Variación a 2045	Particip. mercado 2045
Colombia	limones	+248%	3,5%
Estados Unidos	frambuesas	+108%	60,5%
México	aguacates	+50%	49,2%
Guatemala	melones	−43%	20,4%
Brasil	piña	−87%	0,1%

Caso A · El producto final

Un **informe prospectivo ejecutivo**: resumen gerencial, escenarios, y un cuadro que prioriza las combinaciones país–fruta con mayor expansión y mayor contracción a 2045. Una muestra real (escenario moderado SSP2-4.5, variación 2045 vs. promedio 2020–2024):

País	Fruta	Variación a 2045	Particip. mercado 2045
Colombia	limones	+248%	3,5%
Estados Unidos	frambuesas	+108%	60,5%
México	aguacates	+50%	49,2%
Guatemala	melones	−43%	20,4%
Brasil	piña	−87%	0,1%

Miren dos filas: Colombia–limones y Brasil–piña tienen las variaciones **más espectaculares** de la tabla... y las participaciones de mercado **más pequeñas**. ¿Entonces qué se recomienda? → *siguiente lámina*.

Caso A · La lectura de negocio (la cuarta pata)

El modelo produce números; **el analista produce prioridades**. La clave: la *doble lectura* — variación proyectada $\times$ participación de mercado.

- **Colombia, limones (+248%, part. 3,5%)** → *oportunidad emergente*: crece muchísimo, pero desde una base pequeña. Se vigila; no redefine la estrategia.
- **EE.UU., frambuesas (+108%, part. 60,5%)** → *foco material*: crecimiento alto y dominio del mercado. Aquí sí se mueve capacidad.
- **Guatemala, melones (−43%, part. 20,4%)** → *riesgo prioritario*: cae fuerte en un mercado donde el país pesa.
- **Brasil, piña (−87%, part. 0,1%)** → espectacular... e *irrelevante* para la decisión.

Caso A · La lectura de negocio (la cuarta pata)

El modelo produce números; **el analista produce prioridades**. La clave: la *doble lectura* — variación proyectada $\times$ participación de mercado.

- **Colombia, limones (+248%, part. 3,5%)** → *oportunidad emergente*: crece muchísimo, pero desde una base pequeña. Se vigila; no redefine la estrategia.
- **EE.UU., frambuesas (+108%, part. 60,5%)** → *foco material*: crecimiento alto y dominio del mercado. Aquí sí se mueve capacidad.
- **Guatemala, melones (−43%, part. 20,4%)** → *riesgo prioritario*: cae fuerte en un mercado donde el país pesa.
- **Brasil, piña (−87%, part. 0,1%)** → espectacular... e *irrelevante* para la decisión.

El error del analista junior

Presentar la tabla ordenada por variación y detenerse ahí. El +248% *sin* el 3,5% al lado manda la estrategia al lugar equivocado. **Esa columna extra es el enfoque de negocio.**

Caso A · Del producto a la decisión

Lo que la empresa hizo con el informe:

- Priorizar **dónde** enfocar la estrategia comercial de los próximos años (focos de expansión con peso de mercado).
- Construir una **lista de vigilancia** de riesgos: mercados relevantes con contracción proyectada.
- Cruzar los focos con **riesgos no climáticos** (logística, concentración productiva) antes de decidir.

El entregable se lee **una vez** y alimenta decisiones de **largo plazo**. Es un producto tipo *informe*.

Ficha del proyecto

- Duración: **3 semanas**.
- Entrega: **única** (informe + nota metodológica).
- Etapas dominantes: **1–3** (pregunta y construcción de datos).
- Tarea: **predicción** supervisada.
- Roles protagonistas: *business analyst* + *data engineer*.

Roadmap

- 1 El mapa que ya tienen (repaso en una lámina)
- 2 Caso A — Empresa P: predecir un mercado que no está en los datos
- 3 Caso B — Empresa M: vigilar una red que produce datos sin parar
- 4 El contraste: dos puntos de partida, dos productos
- 5 Su turno: el caso Cóndor

Caso B · El contexto

Empresa M es una empresa de **servicios públicos domiciliarios**: atiende $\sim 1,4$ millones de usuarios residenciales y opera una red con ~ 300 **estaciones telemedidas** que registran el consumo **hora a hora**.

Cómo se vigila esa red hoy: un profesional llega a las 7:00 a.m., saca el reporte del día anterior de los grandes consumidores y lo compara contra un día típico. La regla vigente obliga a revisar en detalle **toda variación mayor al $\pm 20\%$** . Estación por estación. En hoja de cálculo.

Caso B · El contexto

Empresa M es una empresa de **servicios públicos domiciliarios**: atiende $\sim 1,4$ millones de usuarios residenciales y opera una red con ~ 300 **estaciones telemedidas** que registran el consumo **hora a hora**.

Cómo se vigila esa red hoy: un profesional llega a las 7:00 a.m., saca el reporte del día anterior de los grandes consumidores y lo compara contra un día típico. La regla vigente obliga a revisar en detalle **toda variación mayor al $\pm 20\%$** . Estación por estación. En hoja de cálculo.

La necesidad, tal como llegó

"Necesitamos detectar consumos atípicos a tiempo — sin ahogarnos revisando estación por estación."

El número que le duele al negocio

Mantener las **pérdidas** — la diferencia entre lo medido a la entrada de la red y lo facturado a clientes — **por debajo del 1%**. Cada anomalía no detectada es pérdida, fraude o falla que nadie vio.

A diferencia de la Empresa P, aquí **los datos sobran**: millones de lecturas horarias. El problema es convertirlas en *decisión diaria*.

Caso B · La pregunta de negocio

De la necesidad vaga... a la pregunta operativa

*“¿Qué estaciones se desviaron **ayer** de su consumo **esperado** — y en qué horas — para priorizar la revisión de **hoy**?”*

Caso B · La pregunta de negocio

De la necesidad vaga... a la pregunta operativa

*“¿Qué estaciones se desviaron **ayer** de su consumo **esperado** — y en qué horas — para priorizar la revisión de **hoy**?”*

- **Específica** — qué estación, qué horas, qué día. Nada de “consumos raros en general”.
- **Accionable** — la respuesta es la lista de trabajo del turno de la mañana.
Test de la pregunta: aprobado por diseño.
- **Medible** — un año de lecturas horarias ya almacenadas.

(El mismo proyecto tiene una **pregunta gemela** para la fase residencial: ¿qué usuarios muestran patrones atípicos compatibles con fraude? — misma lógica, otra escala.)

Caso B · La pregunta de negocio

De la necesidad vaga... a la pregunta operativa

*“¿Qué estaciones se desviaron **ayer** de su consumo **esperado** — y en qué horas — para priorizar la revisión de **hoy**?”*

- **Específica** — qué estación, qué horas, qué día. Nada de “consumos raros en general”.
- **Accionable** — la respuesta es la lista de trabajo del turno de la mañana.
Test de la pregunta: aprobado por diseño.
- **Medible** — un año de lecturas horarias ya almacenadas.

(El mismo proyecto tiene una **pregunta gemela** para la fase residencial: ¿qué usuarios muestran patrones atípicos compatibles con fraude? — misma lógica, otra escala.)

¿Qué tarea analítica es?

“¿Qué es **raro** aquí?” ⇒ **detección de anomalías**. La palabra clave es *esperado*: sin un **baseline**, “anómalo” no significa nada.

Caso B · Los datos ya existían — el reto era otro

La materia prima entregada: **2.577.432 lecturas horarias** de **317 estaciones**, durante 12 meses. Nadie tuvo que construir el dato... pero la EDA (etapas 3–4) igual se ganó su fama:

- **Histórico desigual:** solo 273 estaciones traen el año completo; el resto entró a mitad de año o reporta a saltos.
- **Medidores congelados:** 95 estaciones repiten *exactamente* el mismo valor por más de 24 horas seguidas — el récord: **744 horas (31 días)** marcando lo mismo. Un medidor trabado congela su propio histórico: **ningún método de rangos lo detecta** (se caza por *racha*, no por rango).
- **Concentración:** **88 estaciones acumulan el 80% del volumen** del año. Una alerta no vale lo mismo en todas: ahí nace la priorización.

Caso B · Los datos ya existían — el reto era otro

La materia prima entregada: **2.577.432 lecturas horarias** de **317 estaciones**, durante 12 meses. Nadie tuvo que construir el dato... pero la EDA (etapas 3–4) igual se ganó su fama:

- **Histórico desigual:** solo 273 estaciones traen el año completo; el resto entró a mitad de año o reporta a saltos.
- **Medidores congelados:** 95 estaciones repiten *exactamente* el mismo valor por más de 24 horas seguidas — el récord: **744 horas (31 días)** marcando lo mismo. Un medidor trabado congela su propio histórico: **ningún método de rangos lo detecta** (se caza por *racha*, no por rango).
- **Concentración:** **88 estaciones acumulan el 80% del volumen** del año. Una alerta no vale lo mismo en todas: ahí nace la priorización.

Conecta con la teoría

Esto es “conocer los datos antes de modelar”: los tres hallazgos **cambiaron el diseño** del sistema. La semana 6 es esto, con sus propios datos.

Caso B · El análisis: definir “lo esperado” es el corazón

- 1 **El punto de partida del cliente:** un piloto interno que marcaba alerta cuando el consumo horario salía de límites estadísticos calculados sobre el histórico de cada estación. Se **replicó y validó** antes de proponer nada — respeto por el trabajo previo del cliente.
- 2 **La mejora central: elegir mejor el baseline.** ¿Comparar cada estación contra... su propio histórico? ¿Contra las estaciones *que se le parecen*? Para eso la red se segmentó en **4 grupos** según la *forma* de su curva de 24 horas — no su nivel. (Segmentación, no supervisada, al servicio de la detección de anomalías: las tareas se combinan.)
- 3 **La alerta:** qué tan lejos quedó cada hora de lo esperado, en unidades de desviación ($|z|$): **amarillo** $\geq 1,5$ · **naranja** $\geq 1,75$ · **rojo** ≥ 2 .
- 4 **Los umbrales se definieron CON el equipo de operaciones** — reunión semanal. Las etapas 1 y 6 nunca desaparecen: sin ellas, el sistema muere en un cajón.

Caso B · El producto final no es un informe: es un sistema

Una **aplicación de monitoreo** autocontenida y documentada, que el equipo del cliente puede abrir cualquier mañana:

Presentación del proyecto

El encargo es automatizar la detección de consumos atípicos en la red de estaciones telemedidas de Empresa M y priorizar la revisión diaria: que el analista llegue a su turno y ya tenga la lista de qué revisar. Esta aplicación reúne el trabajo de la primera fase en tres partes.

Esta primera pestaña describe la materia prima —un año de lecturas horarias— y cómo se comportan las estaciones: cuánta información hay por estación, dónde se concentra el consumo, con qué ritmo consume cada tipo de estación y qué límites tiene la base. Cierra con la segmentación de la red en cuatro grupos, que es la que permite comparar cada estación contra las que se le parecen. La segunda pestaña reproduce el tablero del cliente —el piloto Merlin, con límites por rango intercuartil— para tenerlo como punto de partida. La tercera presenta el tablero propuesto: la misma revisión diaria, pero con la línea base por grupo, el tratamiento de ceros y el diagnóstico de cada caso.

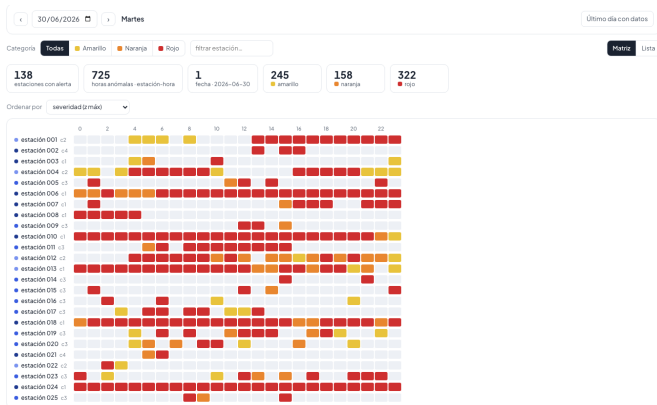
Hoy la validación es estación por estación

Un profesional llega a las 7:00 a.m., saca el reporte del día anterior de los grandes consumidores y lo compara contra un día típico del mismo día de la semana. La regla de negocio vigente obliga a revisar en detalle toda variación mayor al $\pm 20\%$. Con ~300 estaciones y datos horarios, esa validación una a una es el cuello de botella.

Objetivo de largo plazo de Empresa M: mantener las pérdidas — la diferencia entre el gas medido en los city gates y el facturado a clientes — por debajo del 1%.

Cuatro vistas: la materia prima y sus límites · la réplica del piloto del cliente · las desviaciones hora a hora · y las **alertas del día** con su diagnóstico.

Caso B · La revisión de la mañana, resuelta



El analista elige la fecha y el sistema responde la pregunta de negocio **tal cual se formuló**: ese martes, **138 estaciones** con alerta y **725 horas–estación** anómalas, ordenadas por severidad, con la hora exacta del problema — y un clic abre el diagnóstico del caso (volumen, temperatura, presión). **El analista de las 7:00 a.m. llega y la lista ya está hecha.**

Caso B · Del producto a la decisión

La decisión que cambia cada día:

- **A quién revisar hoy** — de ~ 300 estaciones a una lista corta priorizada por severidad y por peso en el volumen.
- Cada caso revisado **alimenta el ciclo**: ¿era fraude, falla del medidor o un cambio real de rutina? Esa respuesta refina los umbrales del mes siguiente.
- Es la etapa 7 girando **cada 24 horas**: la decisión genera datos nuevos y una nueva pregunta.

Ficha del proyecto

- Duración: **3 meses**, dos fases en paralelo.
- Equipo: dos científicas de datos + un consultor senior; **reunión semanal** con operaciones.
- Entrega: **evolutiva** (victorias tempranas → sistema).
- Tarea: **anomalías** (+ segmentación de apoyo).
- Roles protagonistas: *data engineer* + *data analyst*.

Roadmap

- 1 El mapa que ya tienen (repaso en una lámina)
- 2 Caso A — Empresa P: predecir un mercado que no está en los datos
- 3 Caso B — Empresa M: vigilar una red que produce datos sin parar
- 4 El contraste: dos puntos de partida, dos productos
- 5 Su turno: el caso Cóndor

Mismo mapa, recorridos opuestos

	Caso A · Empresa P	Caso B · Empresa M
Necesidad	“¿hacia dónde va nuestro mercado?”	“detectar consumos raros a tiempo”
Pregunta	prospectiva, horizonte 2045	operativa, horizonte <i>mañana</i>
Tarea analítica	predicción (supervisada)	anomalías (+ segmentación)
Datos al arrancar	no existían : se construyó el panel desde fuentes abiertas	sobran : 2,6 millones de lecturas propias
El esfuerzo se fue en...	integrar y armonizar fuentes (etapas 2–3)	definir “lo esperado” y depurar (etapas 3–5)
Producto	informe que se lee una vez	sistema que corre todos los días
Decisión	estratégica, de largo plazo	operativa, del turno de hoy
Ritmo del ciclo	se revisará en años	gira cada 24 horas

La lección

No todos los proyectos de Business Analytics empiezan en el mismo punto ni entregan lo mismo. **La pregunta determina los datos, la tarea, el producto y el ritmo.** Por eso se formula primero.

Lo que los dos casos comparten

- ❶ **La pregunta se escribió antes que el código.** En ambos, la parte más difícil no fue técnica: fue convertir una frase vaga en una pregunta que pasara el test.
- ❷ **Un baseline como vara de medir.** El promedio histórico 2020–2024 en un caso; el consumo esperado por estación en el otro. Sin punto de comparación, ningún número significa nada.
- ❸ **Validación honesta.** Fuera de muestra y contra el trabajo previo del cliente — nunca contra los mismos datos que entrenaron el modelo.
- ❹ **El entregable habla el idioma de quien decide.** Un cuadro de prioridades gerenciales; un semáforo con la lista del día. Ninguno entrega “el modelo”.
- ❺ **Ninguno empezó por el algoritmo.** El modelo apareció en la etapa 5 de 7 — y en ambos casos fue lo *menos* sofisticado del proyecto.

Roadmap

- 1 El mapa que ya tienen (repaso en una lámina)
- 2 Caso A — Empresa P: predecir un mercado que no está en los datos
- 3 Caso B — Empresa M: vigilar una red que produce datos sin parar
- 4 El contraste: dos puntos de partida, dos productos
- 5 Su turno: el caso Cóndor

El encargo de hoy (en grupos de proyecto)

Ustedes acaban de ver dos veces el recorrido completo. Ahora arranca el suyo, con el **caso Cóndor** (el documento del caso y su diccionario de datos están en la página del curso):

- 1 **Lean el caso como analistas**, no como estudiantes: ¿qué le duele a esta fintech? ¿qué decisiones está tomando a ciegas?
- 2 **Inventaríen los datos**: una base principal y tres anexos. Para cada tabla que les interese: ¿qué es *una fila*? ¿qué periodo cubre? ¿qué variable podría ser un *target*?
- 3 **Formulen UNA pregunta de negocio** que pase el test — y clasifiquen qué tarea analítica es.

El encargo de hoy (en grupos de proyecto)

Ustedes acaban de ver dos veces el recorrido completo. Ahora arranca el suyo, con el **caso Cóndor** (el documento del caso y su diccionario de datos están en la página del curso):

- 1 **Lean el caso como analistas**, no como estudiantes: ¿qué le duele a esta fintech? ¿qué decisiones está tomando a ciegas?
- 2 **Inventarían los datos**: una base principal y tres anexos. Para cada tabla que les interese: ¿qué es *una fila*? ¿qué periodo cubre? ¿qué variable podría ser un *target*?
- 3 **Formulen UNA pregunta de negocio** que pase el test — y clasifiquen qué tarea analítica es.

El entregable de la sesión: la ficha de su grupo

(1) La pregunta (específica, accionable, medible) · (2) la decisión que cambia con la respuesta · (3) las tablas/variables que la responden (incluido el target, si aplica) · (4) la tarea analítica y por qué.

Esta ficha es el insumo directo de la Entrega 1 del proyecto final

Antes de entregar: pásenle el detector

Chequeo final para su pregunta — el mismo que usamos con las empresas P y M:

- ¿Dice **qué, quién y cuándo**? (“mejorar la mora” no; “¿qué clientes de crédito entrarán en mora mayor a 30 días el próximo trimestre?” sí.)
- Si les doy la respuesta ahora mismo, ¿**saben qué haría Cóndor diferente mañana?** Si no hay decisión esperando, tienen un tema — no una pregunta.
- ¿**Existen las columnas** para responderla en las cuatro tablas — o al menos se pueden construir? (Recuerden la Empresa P: “no está la columna” a veces significa “hay que construirla”.)
- ¿Qué **tarea** es? ¿Hay target histórico o no? Eso ya define la mitad de su semestre.

Antes de entregar: pásenle el detector

Chequeo final para su pregunta — el mismo que usamos con las empresas P y M:

- ¿Dice **qué, quién y cuándo**? (“mejorar la mora” no; “¿qué clientes de crédito entrarán en mora mayor a 30 días el próximo trimestre?” sí.)
- Si les doy la respuesta ahora mismo, ¿**saben qué haría Cóndor diferente mañana?** Si no hay decisión esperando, tienen un tema — no una pregunta.
- ¿**Existen las columnas** para responderla en las cuatro tablas — o al menos se pueden construir? (Recuerden la Empresa P: “no está la columna” a veces significa “hay que construirla”.)
- ¿Qué **tarea** es? ¿Hay target histórico o no? Eso ya define la mitad de su semestre.

Para cerrar

El 87% de los proyectos de datos muere — y casi siempre muere **en el arranque**, no en el modelo. El arranque del suyo es hoy. Háganlo como profesionales.