

ANALÍTICA PARA LOS NEGOCIOS

06327-ECO | Universidad Icesi | Periodo 2026-02

Caso Cóndor

Una fintech andina de dos lados

*Contexto del proyecto final: la empresa, sus dolores,
el ecosistema de datos y cómo elegir su ruta de análisis*

Empresa (ficticia)	Cóndor — plataforma de pagos y crédito
Rol de ustedes	Equipo de analistas contratado para un diagnóstico 360°
Rutas disponibles	Clasificación, regresión o segmentación
Fecha de corte	31 de marzo de 2026

Índice

1. La empresa: Cóndor	3
2. El encargo: un diagnóstico 360°	3
3. Los departamentos y sus dolores	3
4. El ecosistema de datos	5
4.1. dim_clientes — maestro de clientes	5
4.2. dim_comercios — maestro de comercios	6
4.3. fact_transacciones — histórico transaccional	6
4.4. fact_saldos_mensuales — panel mensual por cliente	6
4.5. creditos — solicitudes y desempeño de crédito	7
4.6. logs_interaccion — eventos de app y web	7
4.7. tickets_soporte — servicio al cliente	8
4.8. campanas_marketing — campañas y respuesta	8
4.9. Tablas de referencia	9
5. Dos formas de trabajar los datos	10
5.1. Ruta básica: la tabla analítica curada	10
5.2. Ruta avanzada: las tablas crudas	10
6. Cómo elegir su ruta de análisis	10
6.1. Ruta C — Clasificación (predecir una categoría)	11
6.2. Ruta R — Regresión (predecir un número)	11
6.3. Ruta S — Segmentación (agrupar sin objetivo)	11
7. Reglas de los datos	12
7.1. Fecha de corte y anti-fuga de información	12
7.2. Entrenamiento y prueba	12
7.3. Grano y unión de tablas	12
8. Carga de los datos en R	12

1. La empresa: Cóndor

Cóndor es una *fintech* andina con sede en Colombia y operación incipiente en Perú y Ecuador. Opera una **plataforma de pagos y crédito de dos lados**, es decir, tiene dos tipos de usuario que dependen el uno del otro:

- **Lado consumidor.** Una *wallet* (billetera) y app para guardar dinero, pagar con QR y tarjeta, transferir, recargar y pedir microcréditos.
- **Lado comercio.** Un datáfono o código QR para que los negocios acepten pagos, más **adelantos de capital de trabajo** basados en su flujo de ventas.

Cinco años de crecimiento explosivo dejaron a Cóndor con **datos abundantes pero desarticulados** entre departamentos: cada área registró su operación por separado y nadie ha mirado el negocio como un todo. Ese es, precisamente, el terreno donde ustedes van a trabajar.

Que la empresa sea “de dos lados” no es un detalle decorativo: significa que hay preguntas sobre **clientes** (¿quién va a caer en mora?, ¿a quién le ofrezco qué?) y preguntas sobre **comercios** (¿cuáles van a dejar de transaccionar?). Ambos lados viven en la misma base de datos.

2. El encargo: un diagnóstico 360°

La Gerencia General de Cóndor tiene una sospecha incómoda: **la empresa crece en usuarios pero pierde dinero por los flancos**, y no logra cuantificar dónde ni por qué. Por eso contrata a un equipo externo de analistas —ustedes— para hacer un diagnóstico y construir capacidades predictivas. Los síntomas que preocupan a la gerencia son:

- Las **pérdidas por fraude** transaccional crecen más rápido que el volumen de operaciones.
- La **cartera de microcréditos** muestra una mora preocupante.
- El **costo de adquisición** de clientes sube, pero muchos usuarios se vuelven inactivos pronto.
- Los **comercios se activan y luego dejan de transaccionar**, sin que nadie lo anticipe.
- **Finanzas no logra proyectar** ingresos con confianza para la próxima ronda de inversión.

Ustedes **no** tienen que resolver todos estos problemas. Cada grupo elige **un** frente —uno de los departamentos de la sección siguiente—, formula una pregunta de negocio propia y la responde con una de las tareas analíticas del curso.

3. Los departamentos y sus dolores

Cada departamento de Cóndor tiene un problema concreto que le duele y una o más rutas analíticas naturales para abordarlo. Este es el **menú**: elijan el frente que más les interese. La

tabla indica la ruta “natural”, pero varias preguntas admiten más de un enfoque.

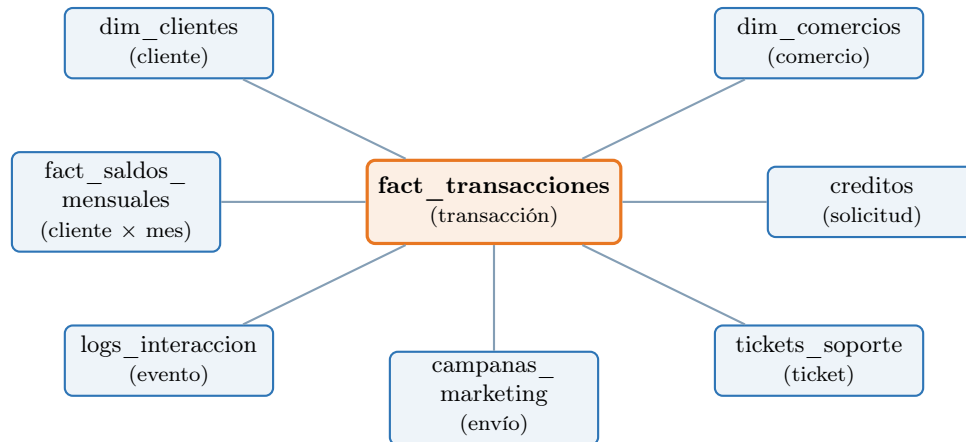
Departamento	Problema que le duele	Ruta(s) natural(es)
Riesgo de Crédito	Anticipar qué solicitudes de crédito van a caer en mora.	Clasificación / Regresión
Fraude y Seguridad	Detectar transacciones fraudulentas casi en tiempo real.	Clasificación
Growth / Marketing	A quién adquirir, a quién retener y a quién ofrecerle qué.	Segmentación / Clasificación
Merchant Success (Operaciones)	Predecir y frenar la deserción de comercios.	Clasificación / Segmentación
Finanzas y Planeación	Proyectar ingresos, gasto y valor de vida del cliente (CLV).	Regresión
Servicio al Cliente	Anticipar la carga de tickets y su tiempo de resolución.	Regresión / Clasificación

Idea clave

El foco del proyecto **no es escribir código perfecto**. Es plantear una buena pregunta, elegir la tarea analítica adecuada, ejecutar el análisis con las herramientas del curso e **interpretar** los resultados para que alguien sin conocimientos técnicos pueda entender y actuar.

4. El ecosistema de datos

Los datos de Cóndor están organizados como un **modelo tipo estrella**: una tabla de hechos masiva en el centro (**fact_transacciones**, con millones de filas) rodeada de tablas maestras (*dimensiones*, una fila por cliente o por comercio) y de tablas satélite que registran otros procesos del negocio (créditos, campañas, soporte, uso de la app).



Todas las tablas se conectan por **cliente_id** (y las compras, además, por **comercio_id**).

Las variables marcadas con ★ **objetivo** son **variables objetivo**: lo que un modelo intentaría predecir. No están en todas las tablas y no todas las rutas usan las mismas.

4.1 dim_clientes — maestro de clientes

Grano: una fila por cliente. Quién es cada usuario de la wallet.

Variable	Tipo	Descripción
cliente_id	id	Identificador único del cliente (llave).
fecha_registro	fecha	Cuándo se registró en la plataforma.
edad	entero	Edad en años.
genero	categorica	F / M / Otro.
ciudad, departamento, pais	categoricas	Ubicación declarada del domicilio.
lat_domicilio, lon_domicilio	continuas	Coordenadas aproximadas del domicilio.
canal_adquisicion	categorica	Cómo llegó: orgánico, ads_meta, ads_google, referido, influencer.
ingreso_declarado	continua	Ingreso mensual declarado (COP).
ocupacion	categorica	empleado / independiente / estudiante / pensionado.
kyc_nivel	ordinal	Nivel de verificación de identidad (nivel_1 a nivel_3).
antiguedad_meses	entero	Meses desde el registro hasta la fecha de corte.

4.2 dim_comercios — maestro de comercios

Grano: una fila por comercio. Los negocios que aceptan pagos con Cóndor.

Variable	Tipo	Descripción
comercio_id	id	Identificador único del comercio (llave).
fecha_alta	fecha	Cuándo se afilió a la plataforma.
mcc	categoría	Código de categoría de comercio (ver dim_mcc).
rubro	categoría	Rubro legible (supermercado, restaurante, ...).
tamano	ordinal	micro / pequeño / mediano.
ciudad, departamento, pais	categorías	Ubicación del comercio.
lat, lon	continuas	Coordenadas del comercio.
tipo_terminal	categoría	QR / datáfono / link de pago.
ticket_promedio	continua	Valor promedio de una compra en ese comercio (COP).

4.3 fact_transacciones — histórico transaccional

Grano: una fila por transacción (la tabla masiva, ~2 millones de filas). El corazón del negocio.

Variable	Tipo	Descripción
transaccion_id	id	Identificador único de la transacción.
cliente_id	id (FK)	Cliente que realiza la operación.
comercio_id	id (FK)	Comercio receptor (vacío si no es compra).
timestamp	fecha-hora	Momento exacto de la transacción.
monto	continua	Valor de la operación (COP).
tipo	categoría	compra / transferencia / retiro / recarga / pago_servicio.
canal	categoría	app_qr / tarjeta / pse / link.
mcc	categoría	Categoría del comercio (vacío si no es compra).
moneda	categoría	COP / USD / PEN.
es_internacional	binaria	1 si la operación es internacional.
lat_tx, lon_tx	continuas	Coordenadas donde ocurrió la transacción.
dispositivo_id	id (FK)	Dispositivo desde el que se operó.
estado	categoría	aprobada / rechazada.
etiqueta_fraude	binaria ★ objetivo	1 si la transacción fue fraudulenta.

4.4 fact_saldos_mensuales — panel mensual por cliente

Grano: una fila por cliente y mes. La “foto” financiera de cada cliente, mes a mes. Es la materia prima natural para construir un objetivo de **regresión** (por ejemplo, el gasto del próximo trimestre).

Variable	Tipo	Descripción
cliente_id	id (FK)	Cliente.
mes	fecha	Mes del panel (primer día del mes).
saldo_promedio	continua	Saldo promedio del mes (COP).
saldo_fin_mes	continua	Saldo al cierre del mes (COP).
gasto_total_mes	continua	Gasto total del mes (COP).
ingreso_total_mes	continua	Ingresos recibidos en el mes (COP).
num_tx_mes	entero	Número de transacciones del mes.
productos_activos	entero	Productos de Cóndor activos ese mes.

4.5 credits — solicitudes y desempeño de crédito

Grano: una fila por solicitud de crédito. No todos los clientes piden crédito; algunos piden varias veces.

Variable	Tipo	Descripción
credito_id	id	Identificador de la solicitud.
cliente_id	id (FK)	Cliente solicitante.
fecha_solicitud	fecha	Cuándo se solicitó el crédito.
monto_aprobado	continua	Monto aprobado (COP).
plazo_meses	entero	Plazo del crédito en meses.
tasa	continua	Tasa de interés mensual.
score_buro	continua	Puntaje de buró al momento de solicitar (300–850).
destino	categorica	consumo / educación / negocio / salud / otros.
default_90d	binaria ★ objetivo	1 si el cliente cayó en mora de 90+ días.
dias_mora	entero	Días de mora acumulados (0 si no hubo default).
monto_recuperado	continua ★ objetivo	Monto recuperado de la cartera en mora (COP).

Regla de oro (anti-fuga de información)

En **credits** hay dos variables que **solo se conocen después** de otorgar el crédito: **dias_mora** y **monto_recuperado**. Si su objetivo es predecir **default_90d**, **no** pueden usar esas dos como predictores: al momento de decidir el préstamo todavía no existen. Usarlas sería hacer trampa con información del futuro (*data leakage*).

4.6 logs_interaccion — eventos de app y web

Grano: una fila por evento. Alto volumen: cada login, cada pantalla abierta, cada sesión. Sirve para medir **engagement** (qué tan activo está un cliente).

Variable	Tipo	Descripción
cliente_id	id (FK)	Cliente.
timestamp	fecha-hora	Momento del evento.
tipo_evento	categorica	login / push_open / consulta_saldo / pantalla_pago / pantalla_credito / pago_completado.
pantalla	categorica	Pantalla de la app donde ocurrió el evento.
duracion_sesion	continua	Duración de la sesión en segundos.
app_version	categorica	Versión de la app.
os	categorica	Android / iOS.

4.7 tickets_soporte — servicio al cliente

Grano: una fila por ticket de soporte. La variable **tiempo_resolucion_horas** es un objetivo natural de **regresión**.

Variable	Tipo	Descripción
ticket_id	id	Identificador del ticket.
cliente_id	id (FK)	Cliente que abrió el ticket.
fecha	fecha	Fecha de apertura.
tema	categorica	problema_pago / consulta_credito / fraude_reclamo / actualizacion_datos / otro.
canal	categorica	chat / email / teléfono.
tiempo_resolucion_horas	continua ★ objetivo	Horas hasta resolver el ticket.
csat	ordinal	Satisfacción del cliente (1–5).
reabierto	binaria	1 si el ticket se reabrió.

4.8 campanas_marketing — campañas y respuesta

Grano: una fila por envío de campaña a un cliente. Registra a quién se le envió qué oferta y si respondió.

Variable	Tipo	Descripción
campana_id	id	Identificador del envío.
cliente_id	id (FK)	Cliente que recibió la campaña.
fecha_envio	fecha	Cuándo se envió.
canal	categorica	push / email / sms.
tipo_oferta	categorica	cashback / descuento / upgrade_credito / genérico.
costo	continua	Costo del envío para Cóndor (COP).
convertido	binaria ★ objetivo	1 si el cliente respondió a la oferta.

4.9 Tablas de referencia

Tres tablas pequeñas de apoyo: **dim_geografia** (ciudades con coordenadas), **dim_mcc** (códigos de categoría de comercio y su rubro) y **dim_tiempo** (calendario día a día de la ventana de datos). Sirven para enriquecer o traducir códigos; no son obligatorias.

5. Dos formas de trabajar los datos

Cóndor les entrega **el ecosistema completo** (todas las tablas anteriores) y, además, una **tabla analítica ya construida a nivel cliente**. Ustedes eligen por dónde entrar según su nivel de comodidad con R.

5.1 Ruta básica: la tabla analítica curada

`tabla_analitica_clientes` tiene **una fila por cliente** y reúne, ya calculadas, las variables más útiles de varias tablas: perfil demográfico, comportamiento transaccional (recencia, frecuencia y monto —**RFM**—), uso de la app, resumen de soporte, resumen de crédito y respuesta a campañas. Con esta sola tabla un grupo puede plantear una clasificación (por ejemplo, `default_90d` o `convertido`), una regresión o una segmentación RFM **sin tener que hacer joins**.

Grupo de variables	Contenido
Demografía	<code>edad</code> , <code>genero</code> , <code>ciudad</code> , <code>departamento</code> , <code>pais</code> , <code>canal_adquisicion</code> , <code>ingreso_declarado</code> , <code>ocupacion</code> , <code>kyc_nivel</code> , <code>antiguedad_meses</code> .
Comportamiento (RFM)	<code>recencia_dias</code> , <code>frecuencia_tx</code> , <code>monto_total</code> , <code>monto_promedio</code> , <code>ticket_mediano</code> , <code>pct_tx_internacional</code> , <code>num_categorias_mcc</code> , <code>canal_preferido</code> .
Saldos y app	<code>saldo_promedio_ult3m</code> , <code>gasto_promedio_mensual</code> , <code>num_sesiones_ult30d</code> , <code>duracion_sesion_promedio</code> , <code>os_principal</code> .
Soporte	<code>num_tickets</code> , <code>csat_promedio</code> .
Crédito	<code>tiene_credito</code> , <code>score_buro</code> , <code>default_90d</code> .
Campañas	<code>num_campanas_recibidas</code> , <code>convertido</code> .

Idea clave

Algunas columnas de esta tabla tienen valores faltantes **con sentido**: `score_buro` y `default_90d` están vacías para quien nunca pidió crédito; `convertido` está vacía para quien no recibió campaña; `csat_promedio` para quien no abrió tickets. Diagnosticar y decidir qué hacer con esos faltantes es parte de la Entrega 2.

5.2 Ruta avanzada: las tablas crudas

Los grupos que quieran hacer su propio *feature engineering* pueden partir de las tablas crudas y construir sus variables desde cero: unir `fact_transacciones` con `dim_clientes`, agregar `fact_saldos_mensuales` por cliente, cruzar `logs_interaccion` para medir actividad, etc. Esta ruta es **más exigente** pero se reconoce en la rúbrica del script (“pipeline correcto”).

6. Cómo elegir su ruta de análisis

La pregunta de negocio manda: primero decidan **qué** quieren responder, y de ahí se desprende la tarea analítica y las tablas que necesitan. Aquí hay un ejemplo por cada ruta; **no** son las únicas

preguntas posibles.

6.1 Ruta C — Clasificación (predecir una categoría)

El objetivo es una variable **binaria**. Ejemplos servidos y listos:

- **Fraude:** predecir `etiqueta_fraude` en `fact_transacciones` usando monto, canal, hora, si es internacional, etc.
- **Mora de crédito:** predecir `default_90d` en `creditos` con `score_buro`, `monto_aprobado`, `plazo_meses`, ingreso del cliente...
- **Conversión de campaña:** predecir `convertido` en `campanas_marketing` con el tipo de oferta, el canal y qué tan activo es el cliente.

6.2 Ruta R — Regresión (predecir un número)

El objetivo es una variable **continua**. Ejemplos:

- **Gasto / CLV futuro:** agregar `fact_saldos_mensuales` por cliente y predecir su gasto del próximo trimestre. (*Aquí ustedes construyen el objetivo a partir del panel.*)
- **Tiempo de resolución:** predecir `tiempo_resolucion_horas` en `tickets_soporte`.
- **Recuperación de cartera:** predecir `monto_recuperado` entre los créditos en mora.

6.3 Ruta S — Segmentación (agrupar sin objetivo)

No hay variable objetivo: la meta es **descubrir grupos**. Ejemplos:

- **Segmentar clientes** por un perfil RFM + comportamiento (recencia, frecuencia, monto, mezcla de categorías y canales, uso de la app) construido desde `fact_transacciones` y `logs_interaccion`, o directo desde la tabla analítica.
- **Segmentar comercios** por su rubro, tamaño, ticket promedio y actividad.

Un segmento solo sirve si se puede **describir** y **accionar**: “clientes jóvenes de alto gasto y baja antigüedad” es útil porque sugiere una acción; “grupo 3” no dice nada. La segmentación termina en un **perfilamiento** con nombre y una recomendación para cada grupo.

7. Reglas de los datos

7.1 Fecha de corte y anti-fuga de información

Todos los datos de comportamiento están medidos **hasta el 31 de marzo de 2026** (la fecha de corte). Los objetivos se refieren a lo que pasó **en o después** de esa fecha. Esta separación imita la realidad: cuando la empresa toma una decisión, solo conoce el pasado.

Regla de oro (anti-fuga de información)

No usen como predictor ninguna variable que, en la práctica, **solo se conoce después** de que ocurre el evento que quieren predecir. Ejemplos: `dias_mora` y `monto_recuperado` para predecir `default_90d`; el `estado` final para predecir el fraude de una transacción que aún no se aprueba. Cuando una variable “predice demasiado bien”, sospechen fuga de información.

7.2 Entrenamiento y prueba

Como se vio en el curso, para saber si un modelo **generaliza** deben separar los datos en **entrenamiento** y **prueba**, entrenar solo con la parte de entrenamiento y medir el desempeño en la de prueba, comparándolo siempre contra un **baseline** (la clase mayoritaria en clasificación, o la media/mediana en regresión). Los objetivos de este caso están servidos con esa disciplina en mente.

7.3 Grano y unión de tablas

Antes de unir dos tablas, verifiquen su **grano**: `dim_clientes` tiene una fila por cliente, pero `fact_transacciones` tiene muchas por cliente. Para llevar transacciones al nivel de cliente hay que **agregar** primero (contar, sumar, promediar) y luego unir. Unir sin agregar duplica filas y distorsiona el análisis.

8. Carga de los datos en R

Todas las tablas están publicadas como archivos `.csv`. Se cargan con `read.csv()` apuntando a la URL correspondiente. Por ejemplo:

```
url_base <- "https://eduard-martinez.github.io/databases/ba/condor/"
clientes <- read.csv(paste0(url_base, "dim_clientes.csv"))
creditos <- read.csv(paste0(url_base, "creditos.csv"))
analitica <- read.csv(paste0(url_base, "tabla_analitica_clientes.csv"))
```

Idea clave

`fact_transacciones.csv` es **muy grande** (~2 millones de filas). Si solo necesitan un subconjunto (por ejemplo, un rango de fechas o un tipo de transacción), conviene filtrarlo temprano. Para muchas preguntas, la `tabla_analitica_clientes` o las tablas satélite (`creditos`, `campanas_marketing`, `tickets_soporte`) son suficientes y mucho más livianas.

Cóndor es una empresa ficticia creada con fines pedagógicos. Cualquier parecido con una empresa real es coincidencia.