

L U M B R E



El flujo de exploración

Grado: 4

Materia: Análisis Exploratorio de Datos · Análisis de Datos

11/10/2026

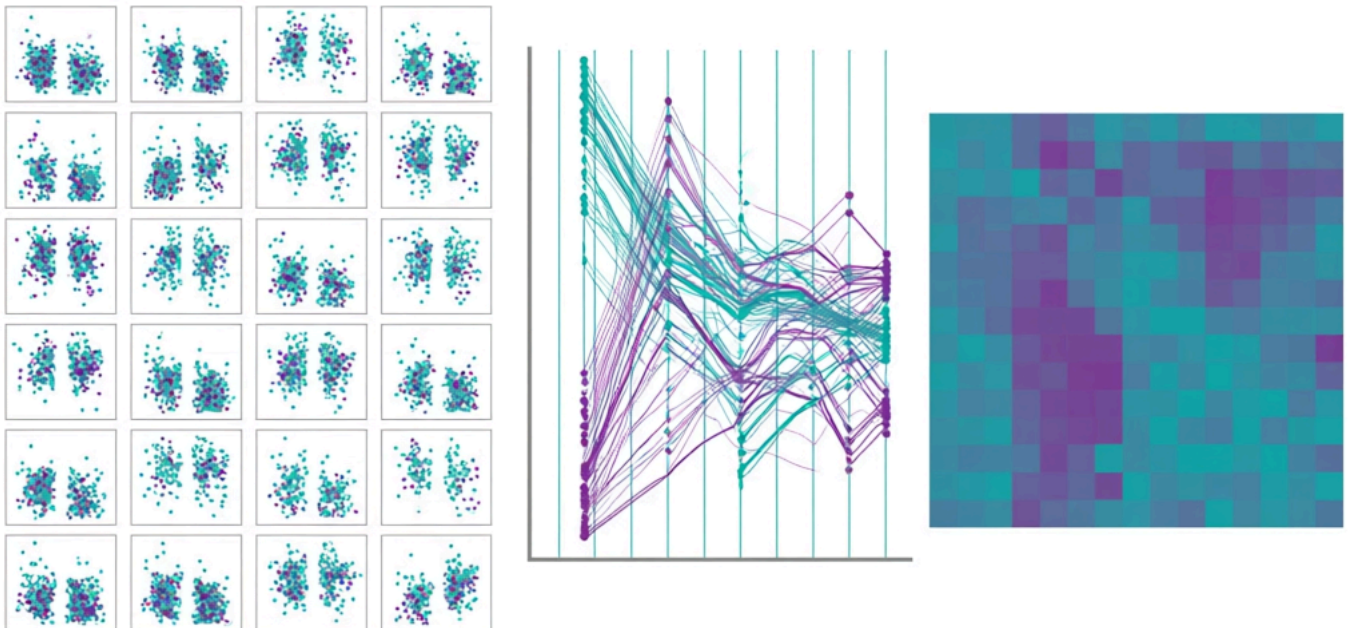
Antes de empezar

Antes de modelar hay que conocer. La exploración es la fase en que se pregunta a los datos qué forma tienen, dónde mienten y qué historia cuentan, con la mente abierta y sin hipótesis rígidas que la conviertan en una mera confirmación de lo que ya se creía.

QUÉ APRENDERÁS

- Describir el flujo ordenado del análisis exploratorio, desde la definición de la pregunta hasta el primer vistazo a los datos, y su papel como fase de descubrimiento frente a la confirmación.

El flujo de exploración



Análisis exploratorio de datos: visualización multivariante y detección de patrones.

Mapa conceptual

- **EDA**

- Univariante
 - Histogramas
 - Boxplots
 - Atipicos
- Bivariante
 - Scatter
 - Correlacion
 - Confusion
- Multivariante
 - Pairs plot
 - Parallel coord
- Modelo
 - Hipotesis
 - Seleccion

El mapa mental interactivo es interactivo en la versión online.



Conocer antes de modelar

Explorar es mirar antes de creer. Es la diferencia entre dejar que los datos te digan algo y forzarlos a decir lo que querías oír. Un buen analista dedica más tiempo a mirar que a ajustar.

Explorar frente a confirmar

El análisis se divide en dos espíritus. La exploración busca patrones sin hipótesis previas, abre preguntas, se permite sorprenderse. La confirmación pone a prueba hipótesis concretas con rigor. Confundirlas produce el vicio de «buscar una aguja» y luego anunciar haberla hallado por azar.

Pregunta antes que método

La exploración no es un paseo aleatorio: empieza con una pregunta de negocio o curiosidad que oriente la mirada sin cerrarla. Sin pregunta se vaga; con demasiada rigidez se deja de explorar y se empieza a confirmar. El equilibrio es el oficio.

Cargar y mirar

El primer contacto con un conjunto de datos debe ser casi físico: tamaño, tipos de cada columna, primeras y últimas filas, recuento de vacíos. Este inventario humilde detecta el noventa por ciento de las sorpresas desagradables antes de que se conviertan en conclusiones erróneas.

Resumen automático

Estadísticos descriptivos de una ojeada —media, mediana, desviación, mínimos, máximos, cuantiles por columna— ofrecen un esqueleto numérico. Ayudan, pero son traicioneros sin contexto: un máximo imposible delata un error que ninguna media revela. El resumen abre puertas, no las cierra.

Tipos y semántica

Un campo declarado como texto puede esconder fechas; uno numérico puede ser en realidad una categoría codificada (un código postal no se promedia). Verificar que el tipo de dato concuerda con el significado es el primer acto de higiene analítica.

La forma de mirar

- De lo general a lo particular: panorama global y luego zoom sobre lo raro.
- De lo univariante a lo multivariante: una variable, dos, muchas.
- De lo esperado a lo inesperado: confirmar lo obvio para poder detectar lo anómalo.
- Con gráfica y número juntos: cada resumen numérico gana con su contrapartida visual.

Distribución antes que resumen

La media resume pero oculta. Dos conjuntos con igual media pueden tener distribuciones radicalmente distintas: bimodales, sesgadas, con colas. Mirar la forma completa de la distribución es más honesto que fiarse de un único número.

Iterar, no linealizar

La exploración es un ciclo de miradas sucesivas, no una tubería recta. Cada gráfico sugiere la siguiente pregunta, cada anomalía sugiere un filtro, cada filtro revela nueva estructura. El orden emerge durante el proceso.

Documentar el rastro

Un buen explorador deja huella: qué miró, qué descartó y por qué. Sin bitácora, los resultados no son reproducibles y el propio analista olvida los callejones ya probados. La exploración es ciencia, no memoria dependiente.

Sesgo de confirmación

El mayor enemigo del explorador es su propia expectativa: tender a ver lo que se espera y pasar por alto lo que lo contradice. Combatirlo exige buscar deliberadamente evidencia contraria y preguntarse qué haría fallar la hipótesis favorita.

La curiosidad disciplinada

Explorar bien combina dos virtudes en tensión: curiosidad para seguir el raro y disciplina para no contarse cuentos. La primera genera hipótesis; la segunda las desecha sin piedad antes de que echen raíces.

Anomalías como oro

El punto que rompe el patrón, la categoría con un solo caso, el valor en el año 1900: son molestias que suelen señalar errores, pero también las ventanas a hallazgos. Descartar lo raro sin mirarlo es tirar la parte más informativa del dato.

Del caos al orden

Al terminar la exploración, un conjunto de datos bruto debe haberse vuelto comprensible: con sus variables caracterizadas, sus errores acotados, sus relaciones intuidas. Esa comprensión es el suelo firme sobre el que se puede empezar a modelar.

Errores frecuentes

- Saltar directo al modelo sin mirar los datos, confiando en que el algoritmo «se las arreglará».
- Quedarse en los resúmenes numéricos sin graficar nunca la distribución.
- Explorar hasta encontrar lo que se quería encontrar y detenerse ahí satisfecho.

1. La diferencia esencial entre análisis exploratorio y confirmatorio es que el exploratorio:

- ☐ Usa más datos
- ☐ Busca patrones sin hipótesis fijas, admitiendo sorprenderse
- ☐ Siempre es gráfico
- ☐ No requiere rigor

2. Los resúmenes numéricos de una ojeada bastan para conocer bien un conjunto de datos.

- ☐ Verdadero
- ☐ Falso

Explorar en equipo

Diferentes miradas detectan distintas anomalías. Revisar juntos los primeros gráficos multiplica las preguntas y reduce el riesgo de que el sesgo de uno solo guíe toda la exploración. La diversidad es un control de calidad cognitivo.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Exploración



4. Ordena las etapas del primer contacto con datos:

Elementos:

- abrir y ver tamaño y aspecto
- revisar tipos, vacíos, dominio de cada columna
- caracterizar variable a variable
- explorar cómo se relacionan

Cargar	. . .
Inventario	. . .
Univariante	. . .
Relaciones	. . .

Gapminder <https://www.gapminder.org/> Exploración visual de datos del mundo que despierta curiosidad.

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=M7tX2MJGEY>

5. Recibes una tabla nueva de clientes. ¿Qué tres mirarías primero y qué tipo de problema esperarías encontrar en cada una?

Referencia curricular

Grado:	4
Materia:	Análisis Exploratorio de Datos · Análisis de Datos

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-anex-001-el-flujo-de-exploracion>

Material educativo generado desde Lumbre · lumbre.lat