Caso práctico II (Bloque II — IA Agéntica) — Construcción de material multimedia mediante IA agéntica

Asignatura: Deep Learning para Business Analytics — Comillas (ICADE). **Profesor:** Eduardo C. Garrido-Merchán · `ecgarrido@comillas.edu`. **Curso:** 2026–2027. **Bloque:** II — IA Agéntica (Temas A y B, semanas 8–13). **Modalidad: entrega en grupo + test individual.**

1. Qué se construye

Cada grupo elige **uno** de los 66 casos del repositorio de dossieres de la asignatura, con su dossier de negocio y su `datos.csv` de un sector real:

https://eduardogarrido90.github.io/docencia/ml1/casos/index.html · contraseña `DataICADE`

Sobre él construye, **dirigiendo a un agente**, lo que una consultora entrega a un cliente: un **dashboard web interactivo** con el análisis de esos datos y unas **slides** de ocho minutos que venden el producto. Las dos cosas, en una tarde.

Tres exigencias. Toda cifra sale del CSV; lo que añadáis de negocio (un precio, un margen, un ROI) va etiquetado como estimación, porque inventar está permitido y disfrazar lo inventado de medido no. La web y las slides comparten paleta, tipografía, marca y gráficos. Y el acabado visual puntúa, al contrario que en el Caso I. No hay memoria escrita que entregar, pero sí hay que mirar la sesión mientras ocurre: qué harness, qué modelo, abierto o cerrado, qué herramientas, cuánto contexto y cuánto dinero. De eso pregunta el test del §4, y lo que separa a quien encarga esa tarde de quien sabe venderla es la factura.

2. Los cuatro bloques analíticos del dashboard

Dentro del dashboard va el análisis entero que sabéis hacer, con sus controles. El agente escribe el código; vosotros decidís qué análisis tiene sentido sobre estos datos y leéis cada salida.

Bloque	Controles	Salida obligatoria
Limpieza	—	Recuadro visible: categorías colapsadas, centinelas traducidos a ausencia, imposibles retirados, filas y valores perdidos finales.
1. Descriptivo univariante	Selector de variable, interruptor crudo/limpio, escala lineal o logarítmica	Numérica: histograma, diagrama de caja, y n, media, mediana, desviación, CV, cuartiles, asimetría y atípicos por 1,5·RIQ. Categórica: barras y frecuencias.
2. Descriptivo bivariante	Selectores de X, de Y y de variable de color	Num-num: nube, recta, r de Pearson y ρ de Spearman. Cat-num: cajas por grupo y medias. Cat-cat: contingencia, χ^2 y V de Cramér. Más la matriz de correlaciones.
3. Inferencia	Selector de contraste (Welch, ANOVA, χ^2, correlación nula), de variables y de α; casillas de regresores	Estadístico, grados de libertad, p, **intervalo de confianza** y **tamaño del efecto**, con la conclusión en español. Regresión: coeficientes con error típico, t, p e IC95, R^2, F, **VIF** y residuos.
4. ML explicable	Selector de modelo (**tres** más una referencia ingenua, misma partición y semilla escrita) y de métrica	Métrica **sobre el conjunto de prueba**. Global: importancia por permutación y media de $\lvert\phi_i\rvert$ de Shapley. Local: descomposición de Shapley de una predicción, que sume a la predicción.

La regla del cero servidor. Una página autocontenida no ejecuta `pandas` ni `scikit-learn`. Todo lo seleccionable se calcula **antes**, en Python, y viaja dentro del HTML como JSON: `panel.py` limpia, calcula la rejilla entera de opciones y vuelca `panel.json`; `index.html` elige qué enseñar. Decidir de antemano esa rejilla es la conversación que vais a tener con el agente.

El panel de referencia: `casos/ejemplo_panel/`. Los cuatro bloques resueltos sobre el caso `delivery`, 9 200 pedidos de una plataforma de reparto: cinco filas entraron multiplicadas por mil y retirarlas baja la desviación típica del importe de 513,78 € a 15,55 €; y la tesis de la casa de que el suscriptor gasta más por pedido se cae en el contraste de Welch (0,28 € de diferencia, IC95 de -0,43 € a 0,99 €, $p = 0{,}433$) y en la regresión (+0,33 €, $p = 0{,}14$, $R^2 = 0{,}616$); el *gradient boosting* predice con RMSE 9,03 € frente a los 15,23 € de la media, y la explicación local de un pedido parte de 31,63 € y le suma 20,73 € por sus cinco comensales. Su `README.md` lo cuenta bloque a bloque, y **el dashboard entero montado sobre esa rejilla está publicado en la web de la asignatura**: ahí se ve qué se espera, controles incluidos. Si elegís `delivery`, el listón es ir más allá.

3. Qué se entrega

Una carpeta comprimida `grupoNN_caso_2.zip`, por debajo de 20 MB, con exactamente esto.

#	Fichero	Qué es
1	`capturas/`	6 a 10 capturas del dashboard en PNG legible: vista general, los cuatro bloques del §2 y dos estados de interacción. Numeradas, con un `README.md` de una línea cada una.
2	`panel/`	El script de cálculo y el JSON de resultados. Es lo que hace comprobable que toda cifra sale del CSV.
3	`slides/slides.pdf`	8–10 diapositivas del mismo producto, con el fuente al lado.

Tres cosas y nada más: ni prompts, ni memoria, ni informe. El dashboard tiene que existir y funcionar: una sola página autocontenida, sin CDN ni conexión, y en clase se os puede pedir que la abráis y mováis un control.

4. El test individual

Los primeros 25 minutos de la clase siguiente a la entrega, a solas, ordenadores cerrados, sin apuntes y sin IA. **20 preguntas tipo test**, cuatro opciones, una correcta: acierto +0,5, error −0,125, en blanco 0, sobre 10 puntos. Va de la **teoría del Tema A** entera, que es lo que acabáis de aplicar: del bucle y las piezas que distinguen un agente de un chatbot al punto de equilibrio de un modelo abierto, pasando por el harness, las skills, las barandillas, las arquitecturas, los modos, el contexto y la fórmula del coste. Cuatro de las veinte preguntan por **vuestra propia sesión**: qué harness usasteis, si el modelo era abierto o cerrado, qué permiso le disteis a la herramienta de escritura y cuánto os costó. No hay que entregar nada de eso, pero sí haberlo mirado con `/cost` y `/context` mientras construíais: como no hay memoria que contrastar, esas cuatro se corrigen contra lo que conteste la mayoría de vuestro grupo, y quien no tocó el caso se delata solo. Se estudia con el deck del Tema A y el simulacro con soluciones que se publica con este enunciado; ninguna de sus preguntas sale en el test real.

5. Rúbrica

Componente	Peso	Qué se valora
Los cuatro bloques del §2	35 %	Limpieza declarada (5), univariante (6), bivariante (7), inferencia con IC y tamaño del efecto (9), ML con explicación global y local (8).
Acabado e interacción	15 %	Controles que funcionan, gráficos con ejes y unidades, identidad visual propia, lectura en móvil, página autocontenida.
Slides	10 %	Ocho minutos defendibles: problema, hallazgo, producto y precio, con la marca de la web.
Test individual del §4	40 %	La nota sobre 10 del test, tal cual.

Calcular no es analizar. Cada bloque lleva al lado una o dos frases que digan qué significa ese número para el cliente: una salida correcta que nadie lee se lleva la mitad de su parte y una lectura correcta con el gráfico feo se lleva la parte entera. **Suspende el caso**, con independencia del acabado visual, que una cifra no se pueda reproducir desde el CSV con el script entregado o que el dashboard no abra o sus controles no hagan nada.

Bonus, hasta +1 punto, sólo si lo obligatorio está completo: un simulador «qué pasaría si» sobre la rejilla precomputada (+0,5); intervalos por bootstrap con el número de remuestreos declarado (+0,25); comportamiento del modelo por subgrupos, con la disparidad medida (+0,25).

6. Fechas, política de IA y ayudas

Publicación en la semana 11 y entrega en Moodle al final de la 13, antes del Bloque III; el test, en la clase inmediatamente siguiente; notas una semana después. Los grupos son los del proyecto final, y como el test es individual el grupo no reparte esa nota. La IA está permitida y fomentada sin restricción para construir, que es de lo que va el caso, y prohibida en el test (regla C del Día 01): la responsabilidad de lo que el agente escriba es vuestra. Para dudas, el foro primero y el profesor después; `casos/ejemplo_panel/` resuelve las del §2 y `casos/demo_caso_2/` enseña cómo se mide una sesión, que es lo que pregunta el test. Y si en vez de dictárselo todo a un agente preferís repartir el encargo, en la web hay **una arquitectura posible de siete agentes** con su flujo y sus órdenes listas para copiar: orientación, no obligación.