

Visualizando la Vida: Exploración Intuitiva de Datos Clínicos en UCI Cardiovasculares

Roy Angel Choquehuanca Anconeyra

Introducción

En los últimos años, el análisis de registros electrónicos de salud (EHR) se ha vuelto una herramienta clave en la investigación y monitoreo de pacientes críticos, especialmente en Unidades de Cuidados Intensivos (UCI) cardiovasculares. Estos registros contienen grandes volúmenes de datos heterogéneos que, si se analizan correctamente, pueden ofrecer una mejor comprensión del estado de los pacientes y apoyar la toma de decisiones clínicas. Sin embargo, uno de los desafíos principales es presentar esta información de forma comprensible y útil para los médicos, quienes requieren visualizaciones claras, rápidas e interactivas para extraer conclusiones relevantes.

Con la llegada de entornos computacionales interactivos como los notebooks (por ejemplo, Jupyter), se ha abierto la posibilidad de integrar código, resultados y visualizaciones en un mismo espacio de trabajo [1]. Estas herramientas han transformado el análisis exploratorio de datos clínicos, permitiendo mayor flexibilidad, trazabilidad y replicabilidad. Sin embargo, siguen existiendo barreras, como la necesidad de conocimientos técnicos avanzados para generar visualizaciones útiles, lo cual puede limitar su adopción por parte del personal médico o de investigadores sin formación en ciencia de datos.

Tareas críticas como la creación de visualizaciones, la identificación de anomalías en los datos clínicos o la exploración de tendencias en signos vitales y resultados de laboratorio continúan siendo procesos manuales, que interrumpen el flujo de trabajo y pueden generar errores o pérdida de información relevante [1][3]. Para atender estos retos, se han desarrollado herramientas como Lux, que genera automáticamente visualizaciones sugeridas al mostrar un dataframe, facilitando así la detección de patrones sin necesidad de escribir comandos complejos [1]. Asimismo, herramientas como Mage integran manipulaciones gráficas de datos y modelos con código, reduciendo la fricción cognitiva [2], y AutoProfiler (Dead or Alive) proporciona un perfilado de datos continuo con resúmenes visuales interactivos que permiten identificar errores y comportamientos inesperados en tiempo real [3].

Estas tecnologías surgen de la necesidad de asistir a usuarios durante la fase exploratoria de los datos clínicos, etapa crucial en la comprensión del estado de salud de los pacientes. Aplicadas a UCI cardiovasculares, estas herramientas pueden representar de manera automática tendencias de presión arterial, frecuencia cardíaca, oxigenación, historial de procedimientos o medicación, permitiendo a los médicos visualizar la evolución de los pacientes y tomar decisiones más informadas con rapidez.

El impacto de estas herramientas recae en su capacidad para aumentar la productividad, mejorar la calidad del análisis de datos clínicos y facilitar la accesibilidad para usuarios con diferentes niveles técnicos. Automatizar parcialmente tareas como la generación de gráficos, la validación de calidad de los datos y el resumen continuo se convierte en una estrategia crucial para fortalecer el monitoreo, predicción y prevención de reingresos a la UCI en pacientes cardiovasculares [1][2][3].

Visualizar datos clínicos complejos y multidimensionales puede ser un reto para los profesionales de salud, especialmente si se carece de una plataforma intuitiva y automatizada. El análisis exploratorio de datos clínicos toma una gran parte del tiempo de los científicos de datos, representando hasta el 50% del esfuerzo en proyectos de salud, lo cual se traduce en demoras para proveer insights accionables. Además, los notebooks actuales no brindan un soporte interactivo o automatizado que facilite la generación de visualizaciones médicas, validaciones de calidad de los datos clínicos ni una integración fluida entre gráficos y código, lo que entorpece los flujos de trabajo, en particular para el personal clínico que necesita retroalimentación visual inmediata para evaluar condiciones críticas.

Herramientas como Lux [1] presentan limitaciones al trabajar con grandes volúmenes de datos de EHR, ya que su sistema de sugerencias visuales puede resultar poco intuitivo o engorroso, al depender de una desconexión entre la manipulación de los datos y la generación automática de gráficos útiles. Mage [2], por su parte, carece de interacción bidireccional entre código y gráficos; las modificaciones en la GUI no se reflejan automáticamente en el código, dificultando la trazabilidad y replicabilidad del análisis. Esto es especialmente problemático cuando se quiere generar dashboards clínicos dinámicos que respondan a los cambios del paciente en tiempo real.

Objetivos

- Evaluar el uso de AutoProfiler como herramienta complementaria a los notebooks computacionales para facilitar la exploración visual de registros electrónicos de salud en pacientes cardiovasculares en UCI.
- Proporcionar visualizaciones automáticas y recomendadas de signos vitales, tratamientos y evolución clínica sin requerir instrucciones explícitas del usuario.
- Permitir a los médicos e investigadores alternar entre interacciones gráficas y programación en notebooks, incrementando la adaptabilidad del análisis clínico.
- Implementar mecanismos de perfilado continuo de datos clínicos que generen resúmenes visuales interactivos en tiempo real, permitiendo detectar patrones críticos o anomalías que puedan indicar riesgo de reingreso o complicaciones.

Descripción del dataset

Aún no se tiene la columna de salida, pero se determinará tras entrenar el modelo con esta data. Será multiclase, ya que al ser binario sólo hay dos opciones, se podría devolver el porcentaje para si un paciente necesita ser readmitido o no. Si un valor domina demasiado (por ejemplo, 95% "no readmitido"), según si queremos que un modelo:

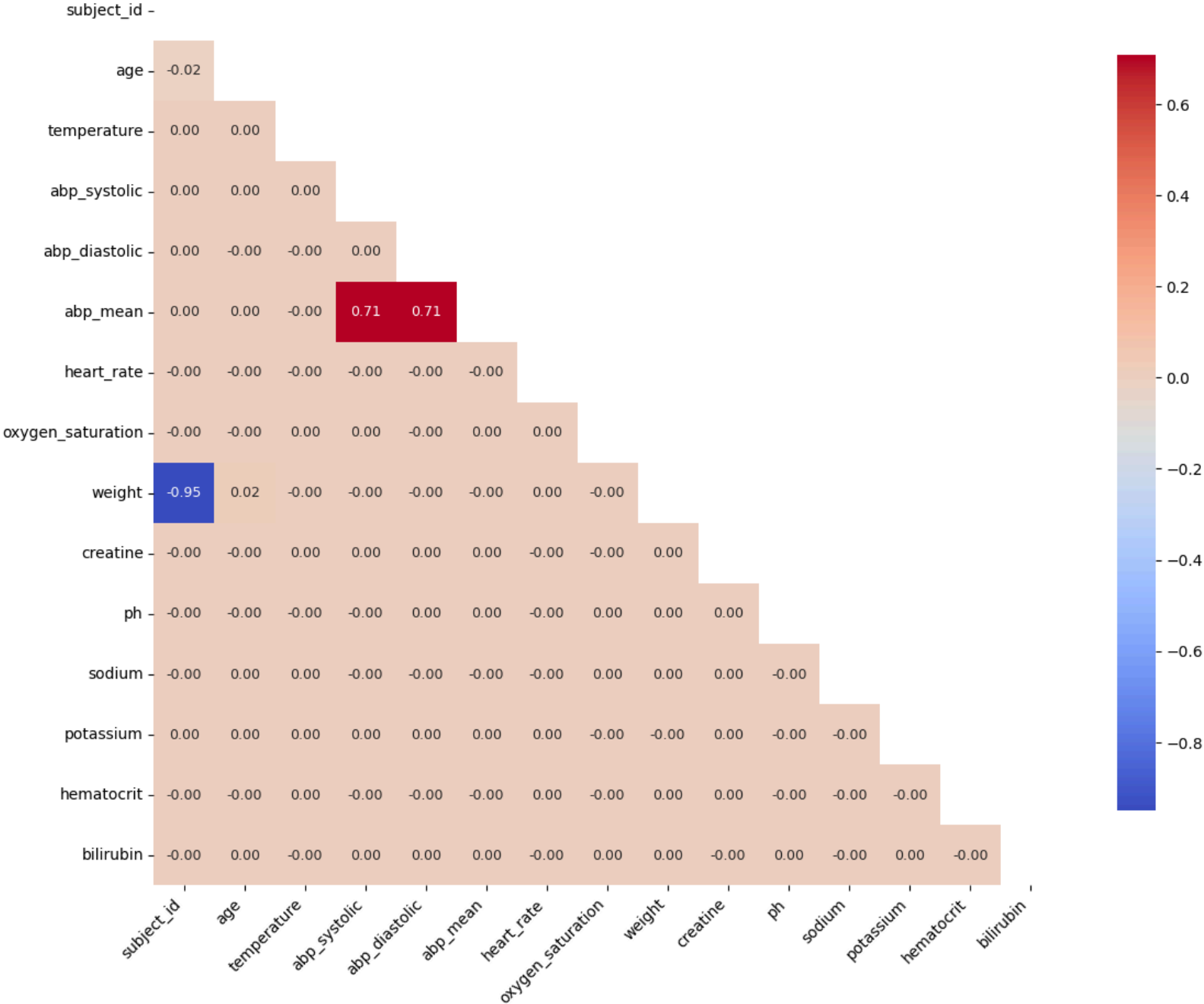
- Si el paciente será readmitido a UCI.
- Si un paciente está estable o crítico.
- Si sobrevivirá 30 días después de salir de UCI.
- Si el tratamiento fue exitoso o fallido.

Descripción del dataset

Columna	Descripción	Tipo	Naturaleza	Límites	Unidad de medida	% de datos faltantes
subject_id	Identificador único y anonimizado del paciente. Permite rastrear registros individuales a lo largo del tiempo sin revelar su identidad.	int / str	Discreto, categórico	1000 – 1499	*	0%
date	Fecha en la que se registraron los signos vitales o exámenes clínicos. Formato: YYYY-MM-DD.	fecha	Discreto temporal (2,191 fechas únicas)	2001-03-31 – 2007-03-29	*	0%
time	Hora del día en que se tomó la muestra o se midió el dato clínico.	hora	Discreto temporal (1,440 valores únicos)	00:00:00 – 23:59:00	*	0%
age	Edad del paciente al momento del registro.	int	Discreto (puede tratarse como continuo)	19 – 89	años	0%
gender	Sexo biológico del paciente.	str / categórico	Nominal (2 valores únicos: 'M' y 'F')	M, F	*	0%
temperature	Temperatura corporal del paciente. Indicador de infecciones o respuesta inflamatoria.	float	Continuo	36.0 – 40.0	°C	0%
abp_systolic	Presión arterial sistólica	float	Continuo	70.0 – 170.0	mmHg	0%
abp_diastolic	Presión arterial diastólica	float	Continuo	30.0 – 80.0	mmHg	0%
abp_mean	Presión arterial media, muy importante en UCI para evaluar perfusión.	float	Continuo	43.4 – 110.0	mmHg	0%

heart_rate	Frecuencia cardíaca en latidos por minuto.	float	Continuo	50.0 – 157.0	bpm (latidos por minuto)	0%
oxygen_saturation	Saturación de oxígeno en sangre	float	Continuo	90.0 – 100.0	%	0%
weight	Peso del paciente	float	Continuo (con posible error de signos)	-329.0 – 157.0	kg	0%
creatine	Nivel de creatinina en sangre	float	Continuo	0.40 – 2.60	mg/dL o $\mu\text{mol/L}$	0%
ph	Medida del pH sanguíneo. El valor normal está entre 7.35 y 7.45	float	Continuo	6.8 – 7.7	adimensional	0%
sodium	Concentración de sodio en sangre	float	Continuo	117.0 – 166.0	mEq/L	0%
potassium	Nivel de potasio en sangre	float	Continuo	2.0 – 8.8	mEq/L	0%
hematocrit	Porcentaje de volumen de glóbulos rojos en la sangre	float	Continuo	8.9 – 53.3	%	0%
bilirubin	Nivel de bilirrubina en sangre	float	Continuo	0.1 – 45.0	mg/dL o $\mu\text{mol/L}$	0%

Mapa de calor de correlación entre variables numéricas



Preguntas

Hipótesis 1: Existe una relación significativa entre la edad del paciente y su frecuencia cardíaca (heart_rate)

A medida que las personas envejecen, el sistema cardiovascular sufre cambios que podrían modificar la frecuencia cardíaca basal. En pacientes críticos, esta relación puede ser más evidente por el estado de salud comprometido.

Correlación Pearson entre edad y ritmo cardíaco: -0.00, p-valor: 0.2164

Coefficiente de correlación: -0.00

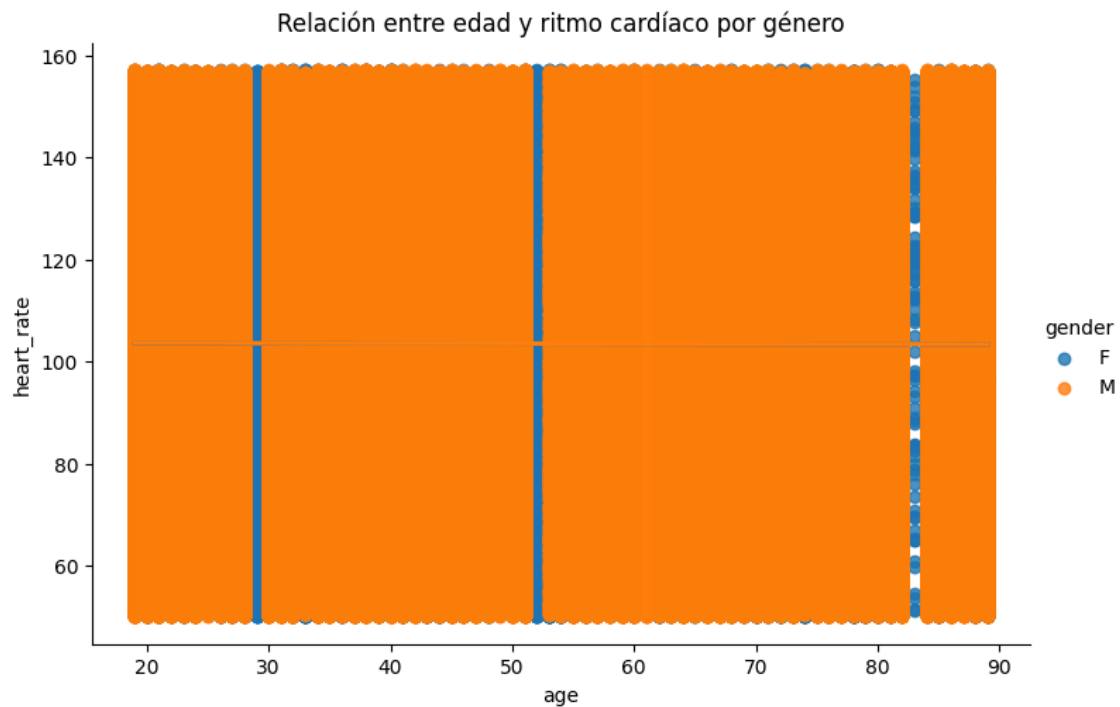
- Este valor está muy cerca de 0, lo que indica que no existe una relación lineal aparente entre la edad y el ritmo cardíaco en este conjunto de datos.
- El signo negativo sugiere una posible tendencia decreciente, pero al ser tan cercano a 0, no es significativo.

P-valor: 0.2164

- Este valor representa la probabilidad de obtener una correlación igual o más extrema que -0.00 si no hubiera relación real en la población (hipótesis nula).
- Como $0.2164 > 0.05$, no se puede rechazar la hipótesis nula, es decir, no hay evidencia estadísticamente significativa de que la edad influya en el ritmo cardíaco en esta muestra.

¿La relación entre la edad y la frecuencia cardíaca varía según el género del paciente (gender)?

El género puede influir en la fisiología cardíaca. Por ejemplo, se ha observado que las mujeres tienden a tener frecuencias cardíacas ligeramente mayores en condiciones normales. Este análisis permitiría descubrir si esta diferencia persiste en pacientes con enfermedades cardíacas en UCI.



Hipótesis 2: Niveles anormales de creatinina (creatine) están asociados con alteraciones en la presión arterial media (abp_mean)

La creatinina es un marcador de función renal. Una función renal deficiente puede afectar la regulación de la presión arterial, especialmente en pacientes críticos.

Correlación Pearson entre creatinina y presión arterial media: 0.00, p-valor: 0.0386

Coefficiente de correlación: 0.00

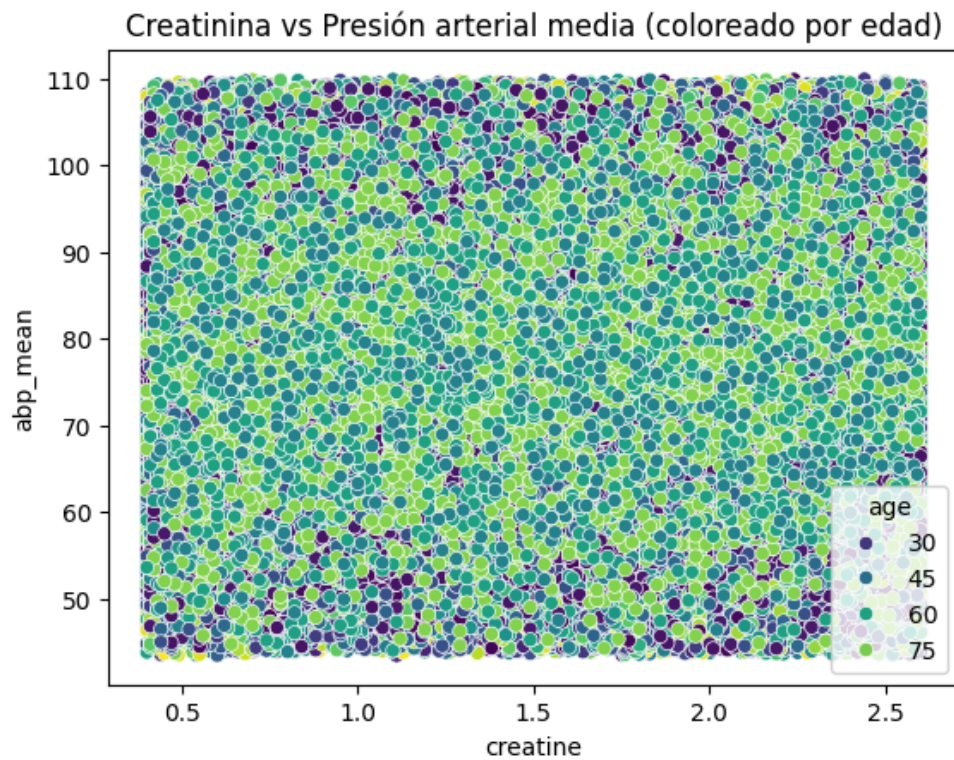
- Al igual que antes, un valor cercano a cero indica que no hay relación lineal fuerte o clara entre la creatinina y la presión arterial media.
- Esto no significa que no haya relación alguna, solo que la relación no es lineal o es muy débil.

P-valor: 0.0386

- Aquí el p-valor es menor a 0.05, lo que significa que sí se puede rechazar la hipótesis nula al 5% de significancia.
- En términos estadísticos: aunque la relación es muy débil (casi nula), el resultado es significativo debido al tamaño de la muestra o la distribución de los datos.

¿La relación entre creatinina y presión arterial media es moderada por la edad del paciente?

En adultos mayores, la función renal suele deteriorarse con la edad, lo que podría amplificar los efectos de la disfunción renal sobre la presión arterial. Explorar esta relación permite identificar si los efectos renales sobre la hemodinámica son más marcados en ciertas edades.



Referencias

- [1] Lee, D., Moritz, D., Wang, K., Ko, A. J., & Howe, B. (2021). Lux: Always-on Visualization Recommendations for Exploratory Dataframe Workflows. arXiv:2105.00121. <https://arxiv.org/abs/2105.00121>
- [2] Head, A., Hohman, F., Barik, T., Drucker, S. M., & DeLine, R. (2020). Mage: Fluid Moves Between Code and Graphical Work in Computational Notebooks. arXiv:2009.10643. <https://arxiv.org/abs/2009.10643>
- [3] Zhang, H., Song, X., Hohman, F., & Liu, Z. (2023). Dead or Alive: Continuous Data Profiling for Interactive Data Science. arXiv:2308.03964. <https://arxiv.org/abs/2308.03964>