

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

INTEGRANTE

- Roy Angel Choquehuanca Anconeysa

Introducción

- El análisis de registros electrónicos de salud (EHR) es clave en la investigación y monitoreo de pacientes críticos en UCI cardiovasculares.
- Los EHR contienen grandes volúmenes de datos heterogéneos, útiles para entender el estado de salud del paciente y apoyar decisiones clínicas.
- Uno de los retos principales es presentar esta información de forma comprensible y útil para los médicos.
- Los médicos requieren visualizaciones claras, rápidas e interactivas para extraer conclusiones relevantes.
- Entornos como Jupyter Notebooks permiten integrar código, resultados y visualizaciones, mejorando el análisis exploratorio clínico.
- A pesar de estas ventajas, todavía se requieren conocimientos técnicos avanzados, lo que limita la adopción de estas herramientas por personal no técnico.
- Muchas tareas (visualización, detección de anomalías, análisis de tendencias) aún son manuales, propensas a errores y rompen el flujo de trabajo clínico.

Introducción

- Han surgido herramientas que automatizan y simplifican el análisis exploratorio:
 - Lux: Sugiere visualizaciones automáticamente al mostrar un DataFrame.
 - Mage: Integra manipulaciones gráficas de datos con código.
 - AutoProfiler: Realiza perfilado continuo con resúmenes visuales e interactivos.
- Estas tecnologías nos podría permitir más otras herramientas:
 - Representar automáticamente tendencias clínicas (frecuencia cardíaca, presión arterial, etc.).
 - Mejorar la calidad del análisis de datos clínicos.
 - Aumentar la productividad del personal médico y técnico.
 - Facilitar la toma de decisiones médicas rápidas e informadas.
- Su impacto puede ser especialmente relevante en:
 - Monitoreo, predicción y prevención de reingresos a UCI en pacientes cardiovasculares.

[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>

Problema

- Visualizar datos clínicos complejos y multidimensionales
- El análisis exploratorio de datos clínicos consume hasta el 50% del tiempo de los científicos de datos en proyectos de salud.
- Esto entorpece el flujo de trabajo clínico, especialmente para médicos que requieren retroalimentación visual inmediata para evaluar condiciones críticas.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId      891 non-null int64
Survived          891 non-null int64
Pclass           891 non-null int64
Name              891 non-null object
Sex              891 non-null object
Age              714 non-null float64
SibSp            891 non-null int64
Parch            891 non-null int64
Ticket           891 non-null object
Fare             891 non-null float64
Cabin            204 non-null object
Embarked         889 non-null object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.6+ KB
```




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

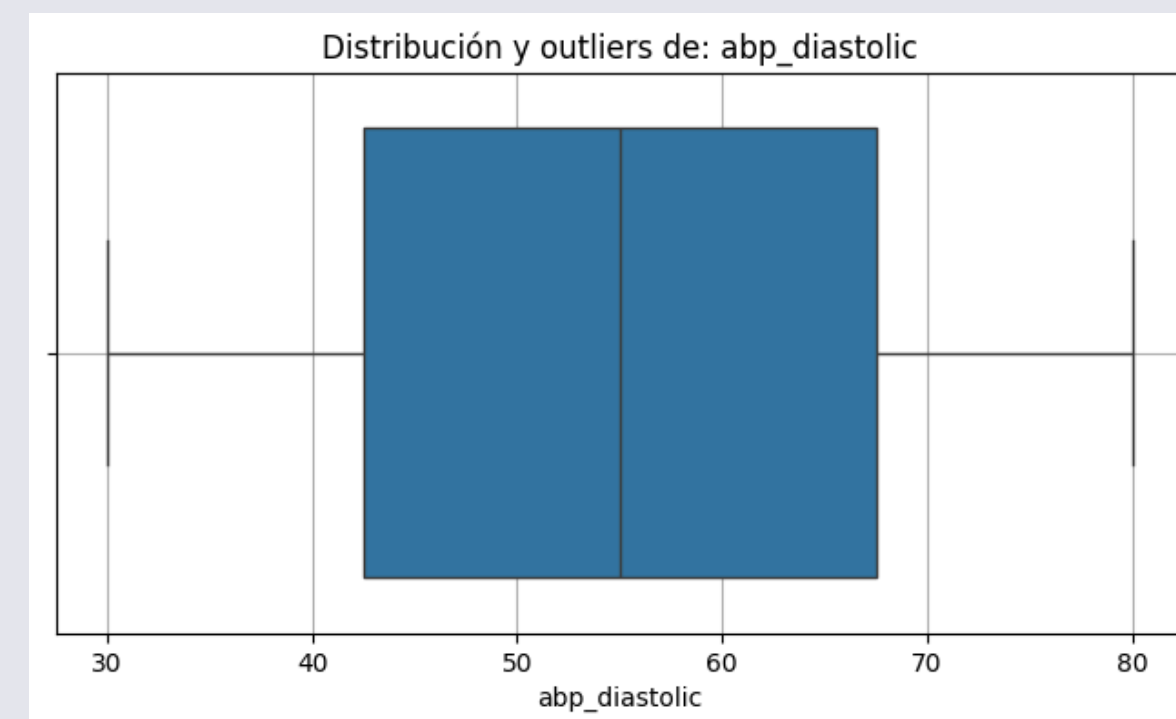
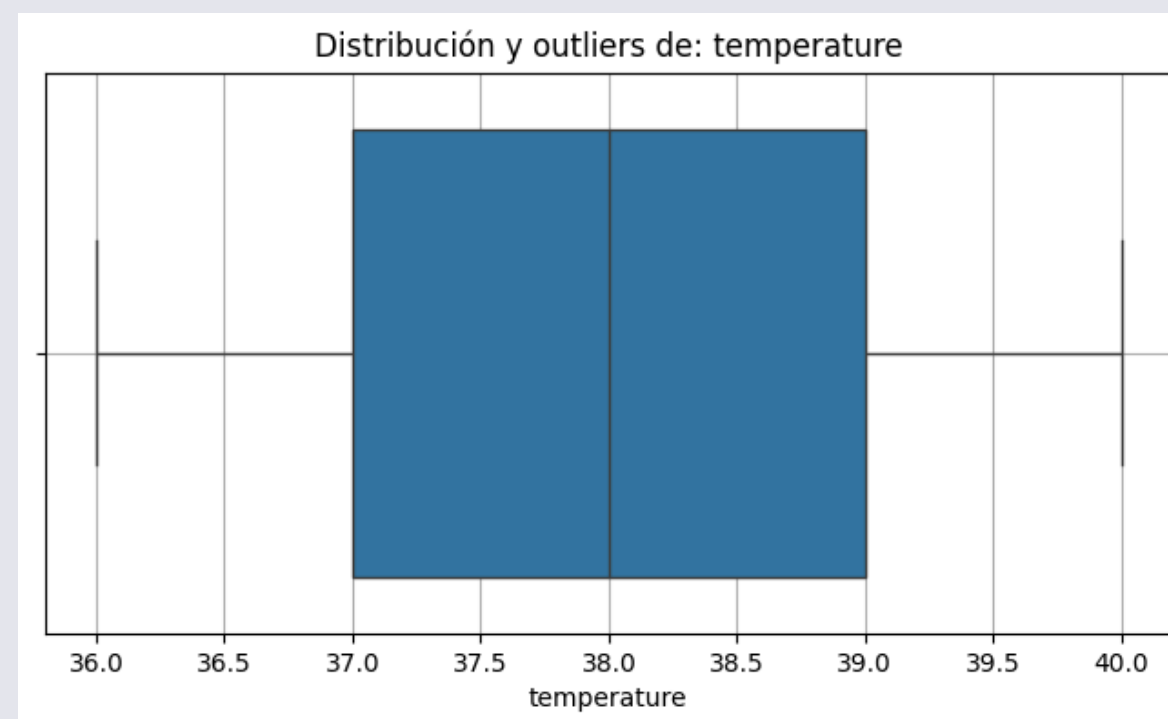
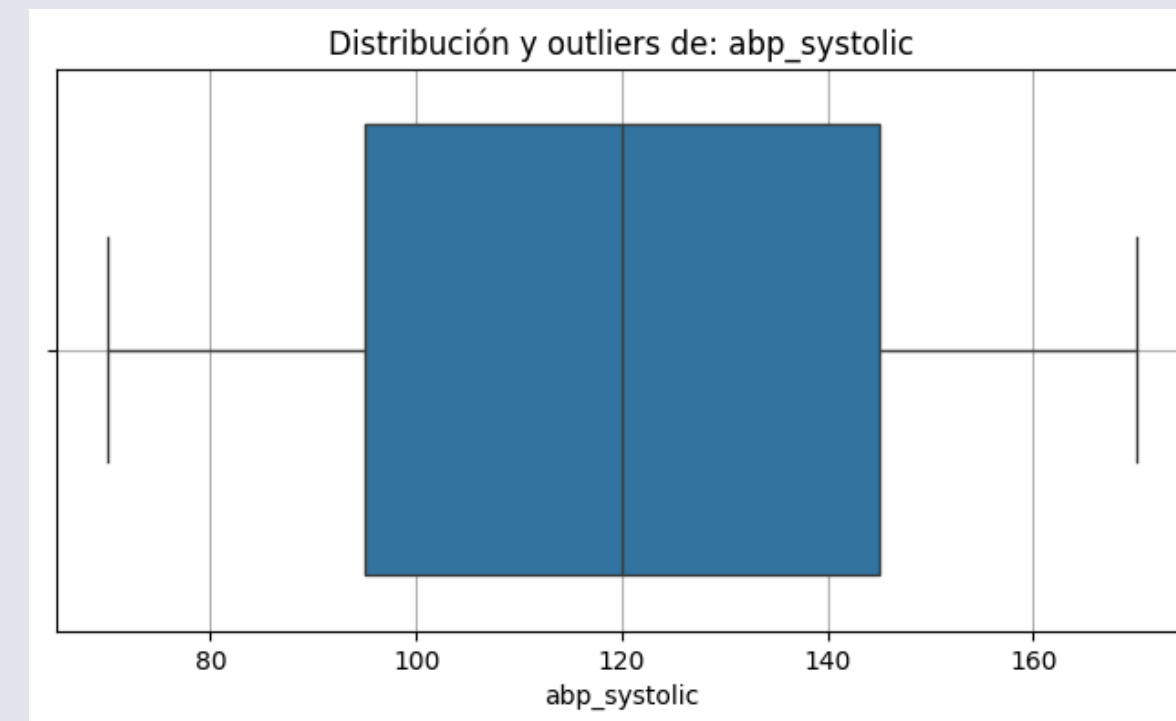
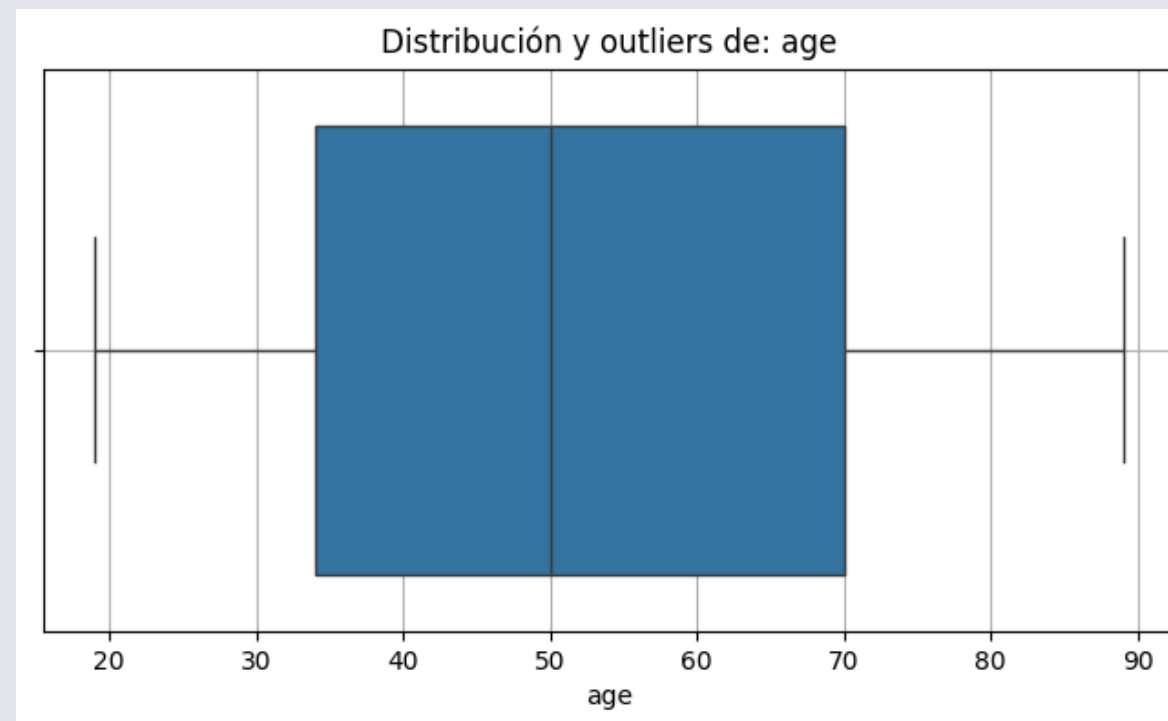


Descripción del dataset

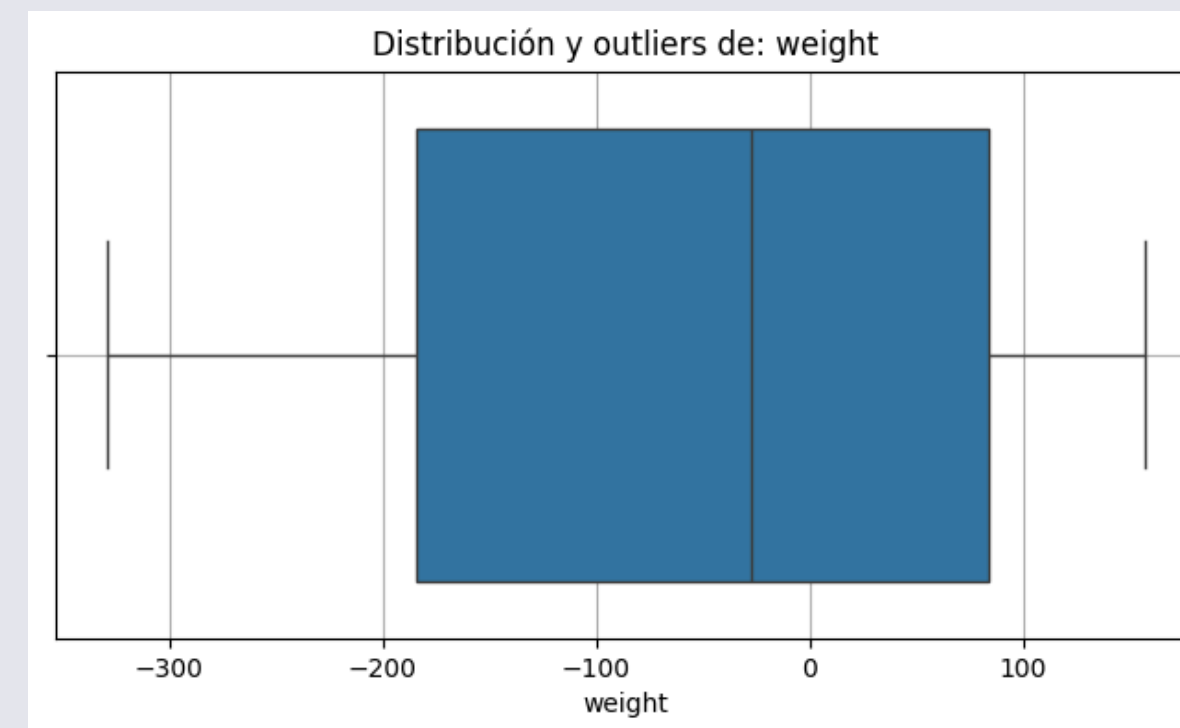
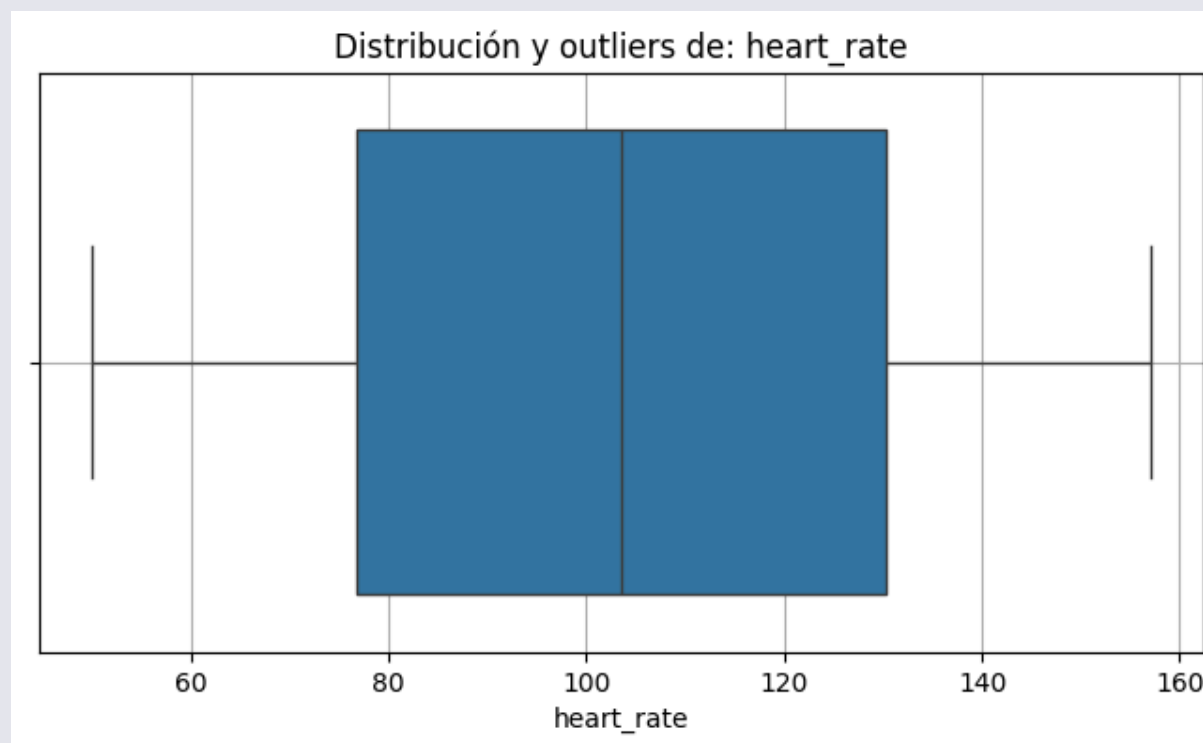
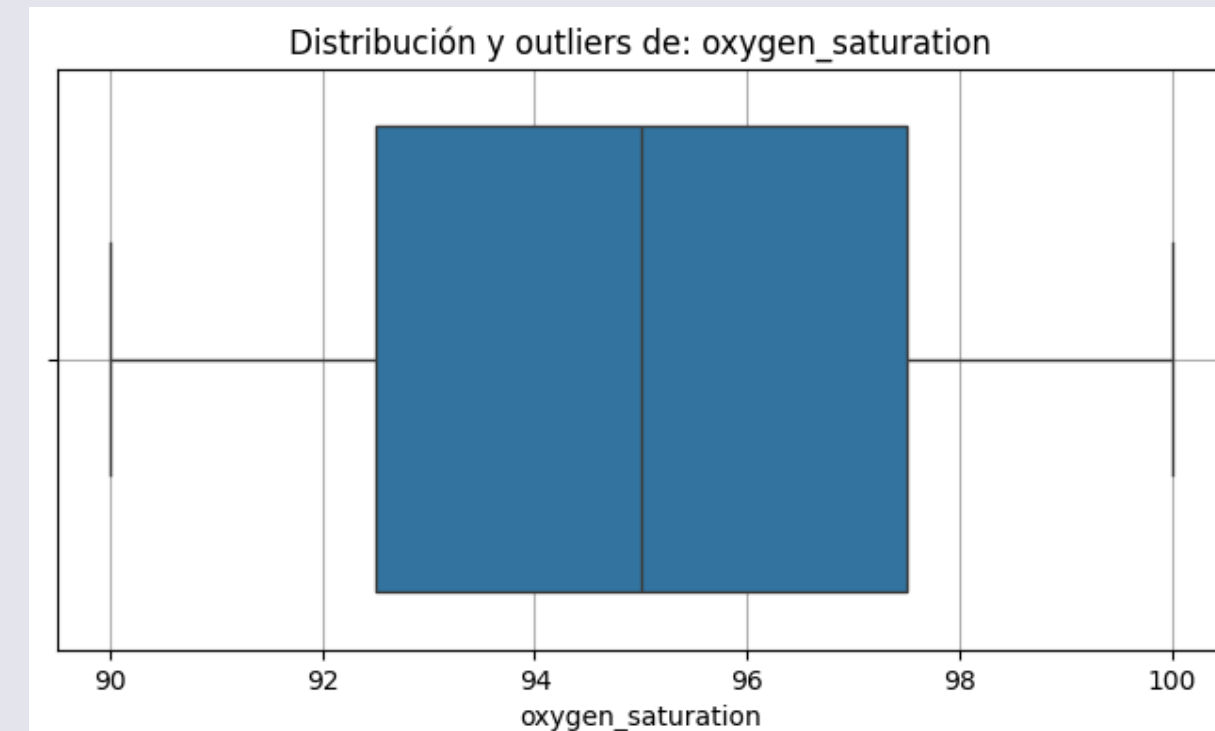
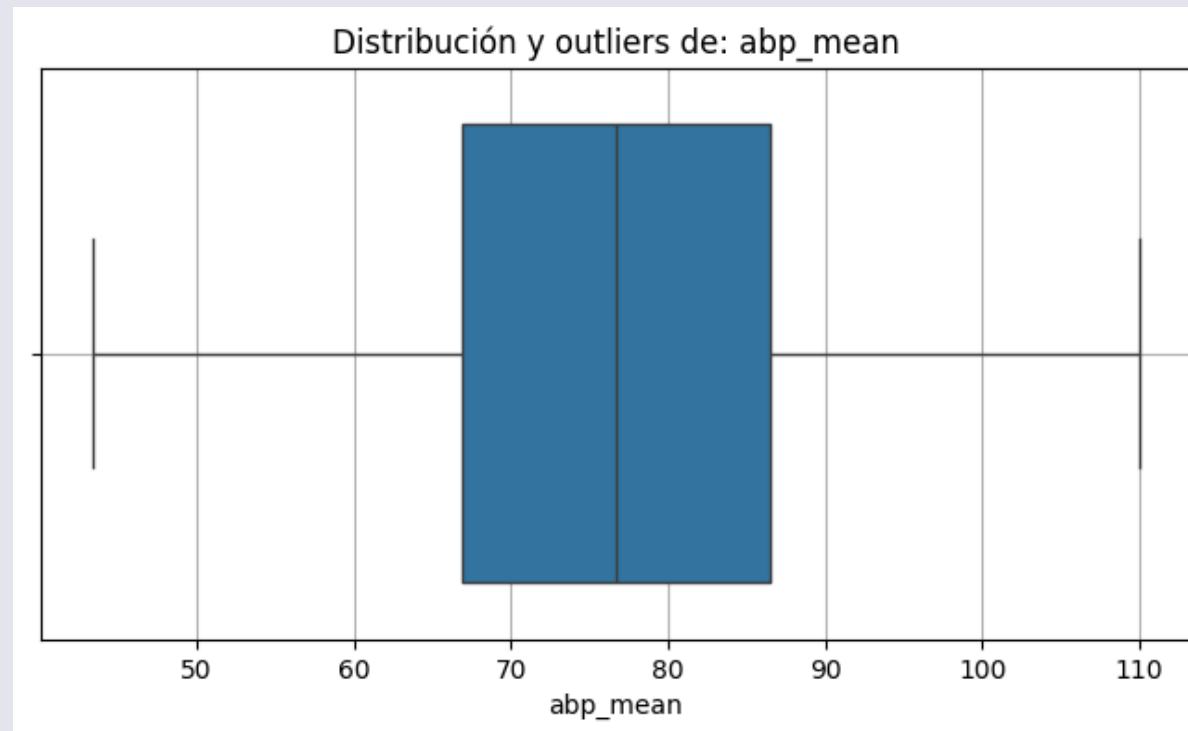
Columna	Descripción	Tipo	Naturaleza	Límites	Unidad de medida	Porcentaje de datos faltantes
subject_id	Identificador único y anonimizado del paciente. Permite rastrear registros individuales a lo largo del tiempo sin revelar su identidad.	int / str	Discreto, categórico	1000 – 1499	*	0%
date	Fecha en la que se registraron los signos vitales o exámenes clínicos. Formato: YYYY-MM-DD.	fecha	Discreto temporal	2001-03-31 – 2002-03-31	*	0%
time	Hora del día en que se tomó la muestra o se midió el dato clínico.	hora	Discreto temporal	00:00:00 – 23:59:59	*	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)	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%

Columna	Descripción	Tipo	Naturaleza	Límites	Unidad de medida	Porcentaje de datos faltantes
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 posición)	-329.0 – 157.0	kg	0%
creatine	Nivel de creatinina en sangre	float	Continuo	0.40 – 2.60	mg/dL o μmol/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 μmol/L	0%

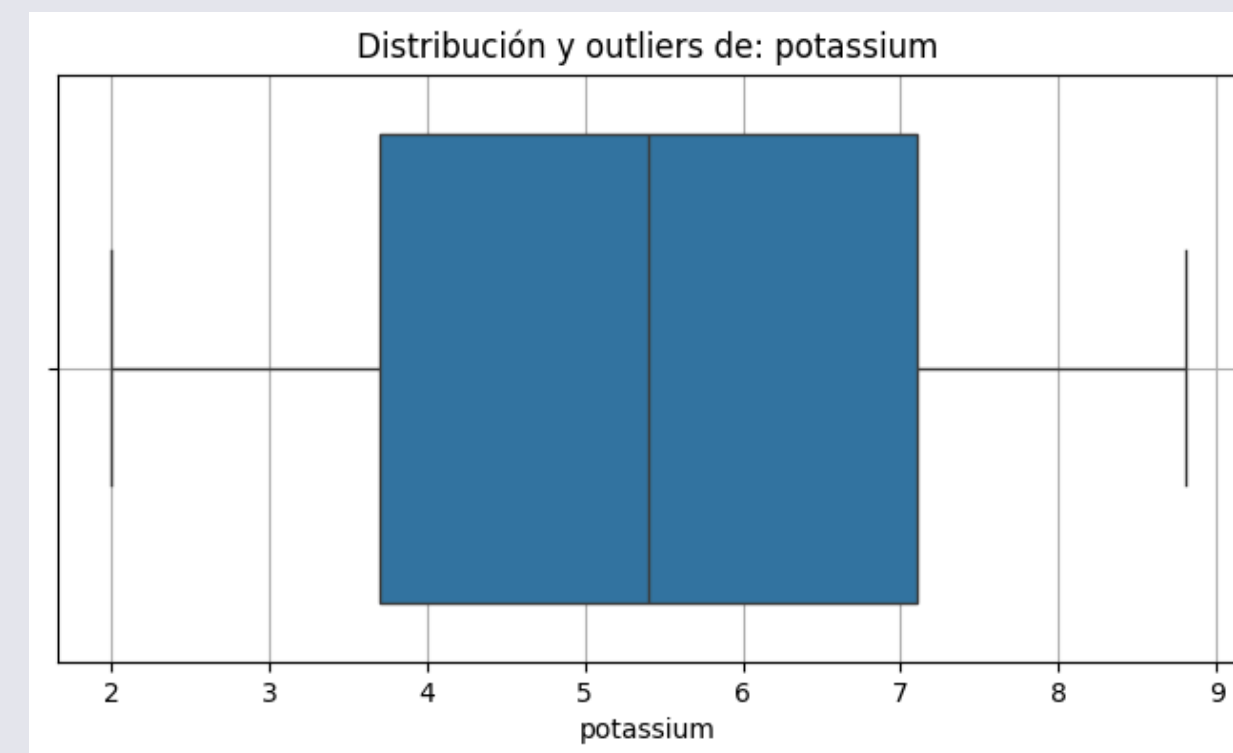
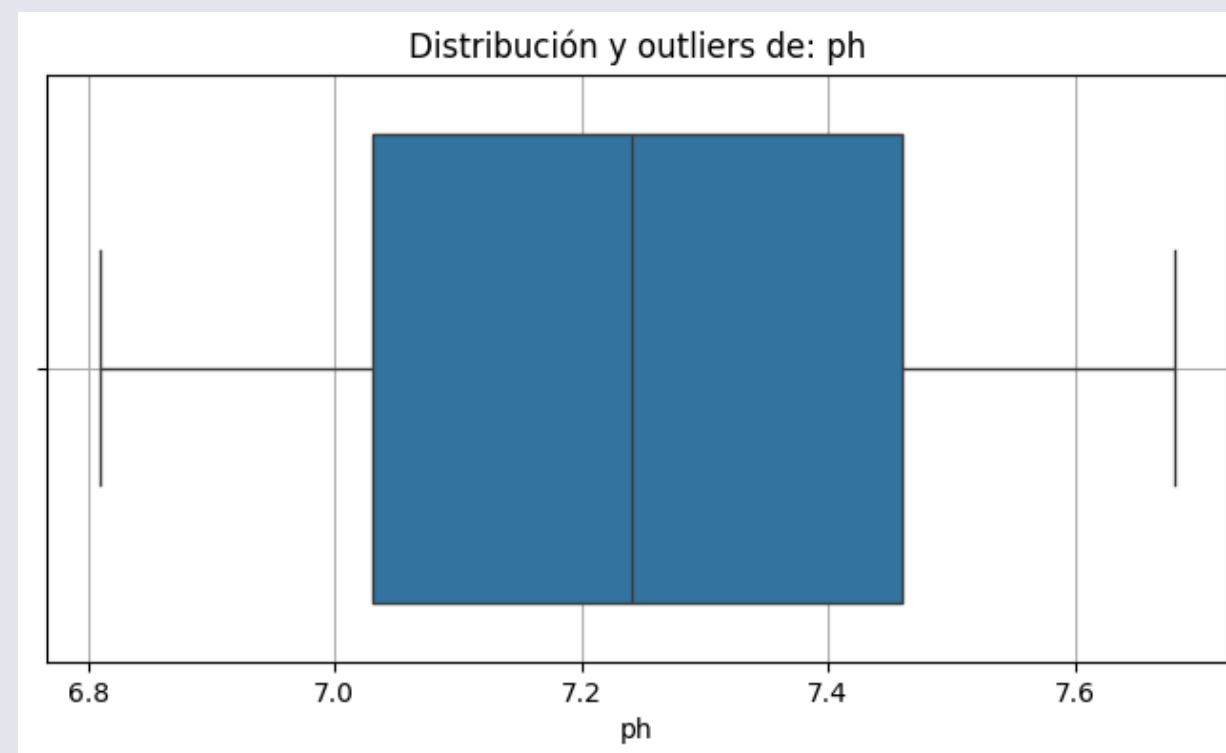
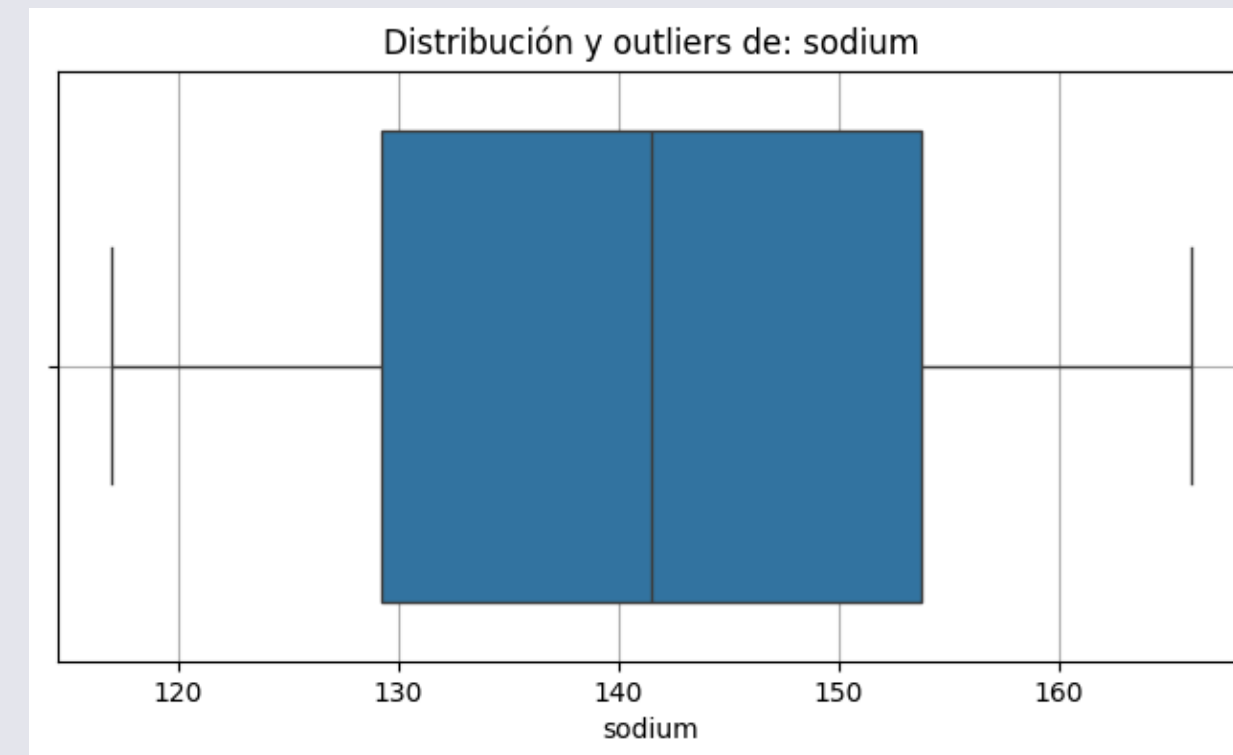
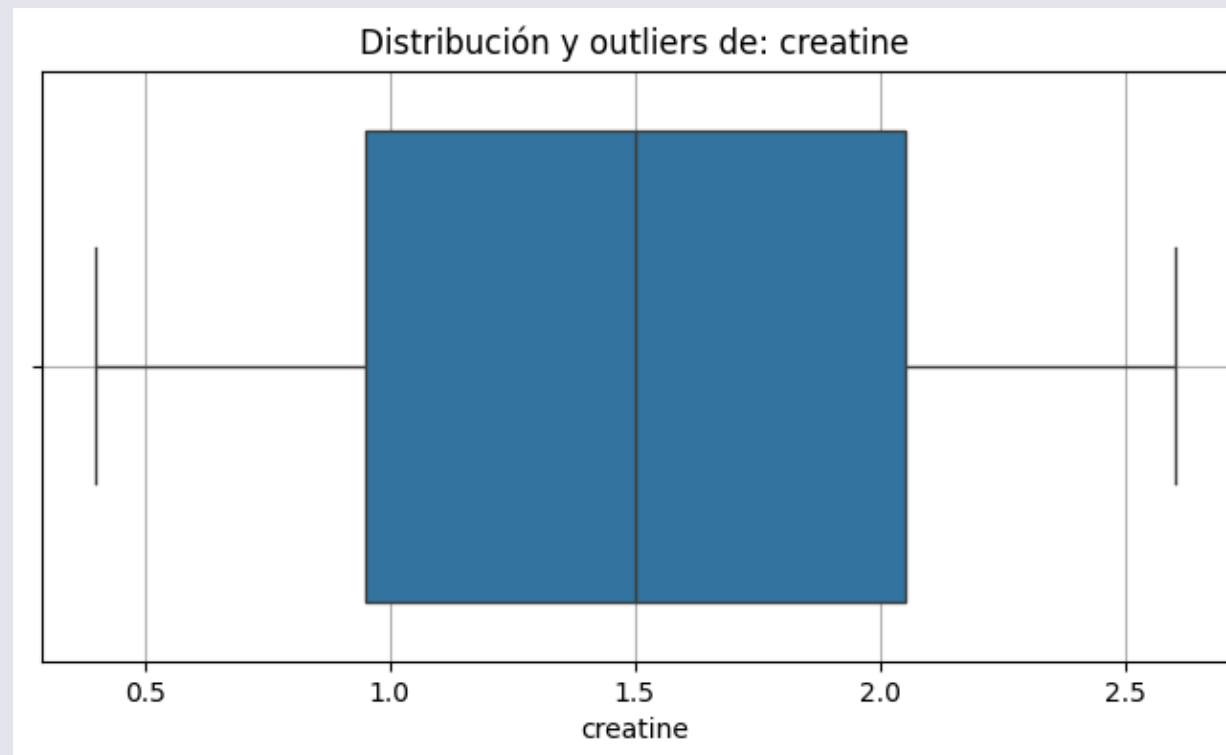
Análisis de outliers



Análisis de outliers



Análisis de outliers

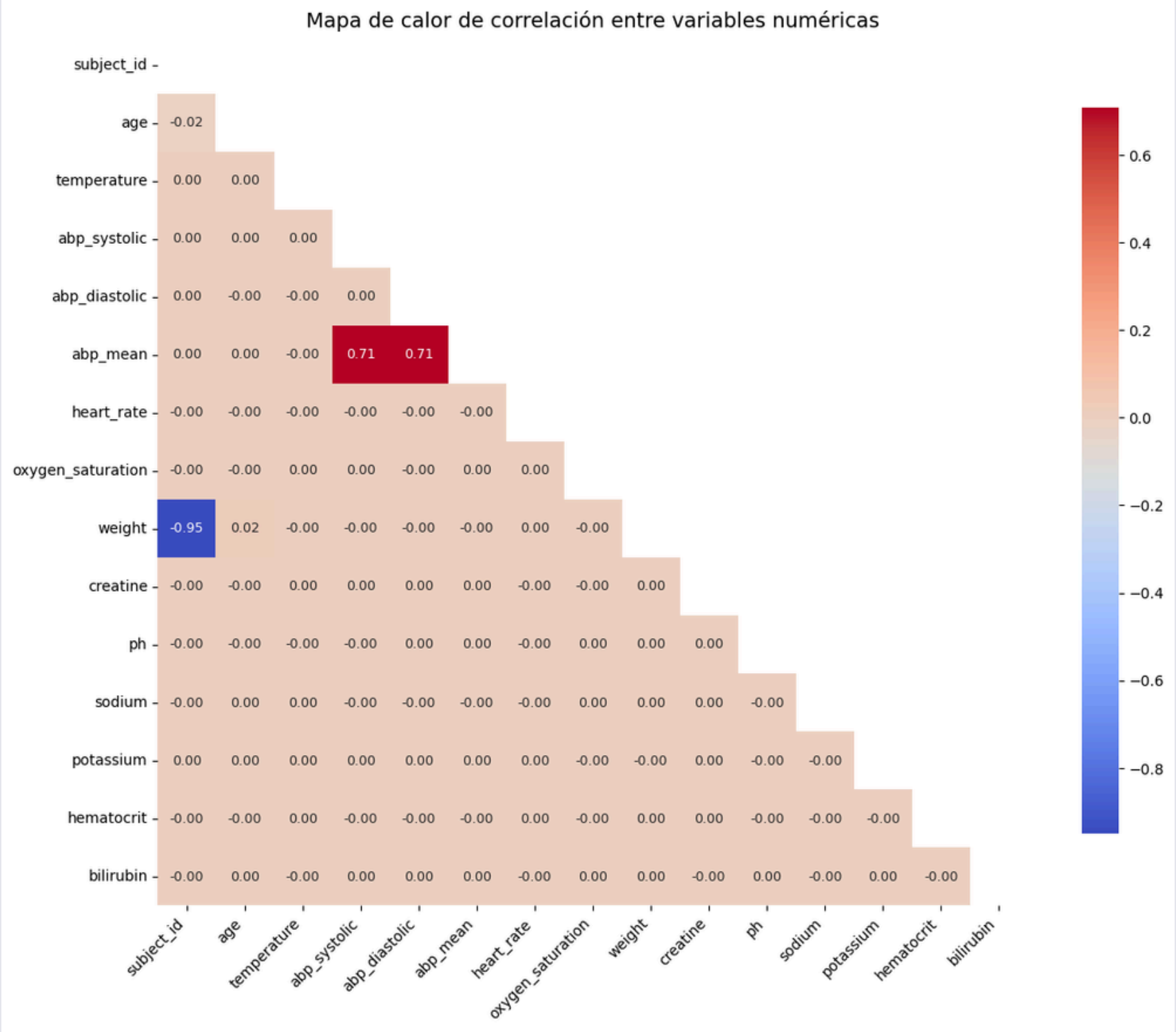


Two overlapping blue squares with rounded corners are positioned in the top-left corner of the slide. The top square is slightly offset to the right and up from the bottom square.

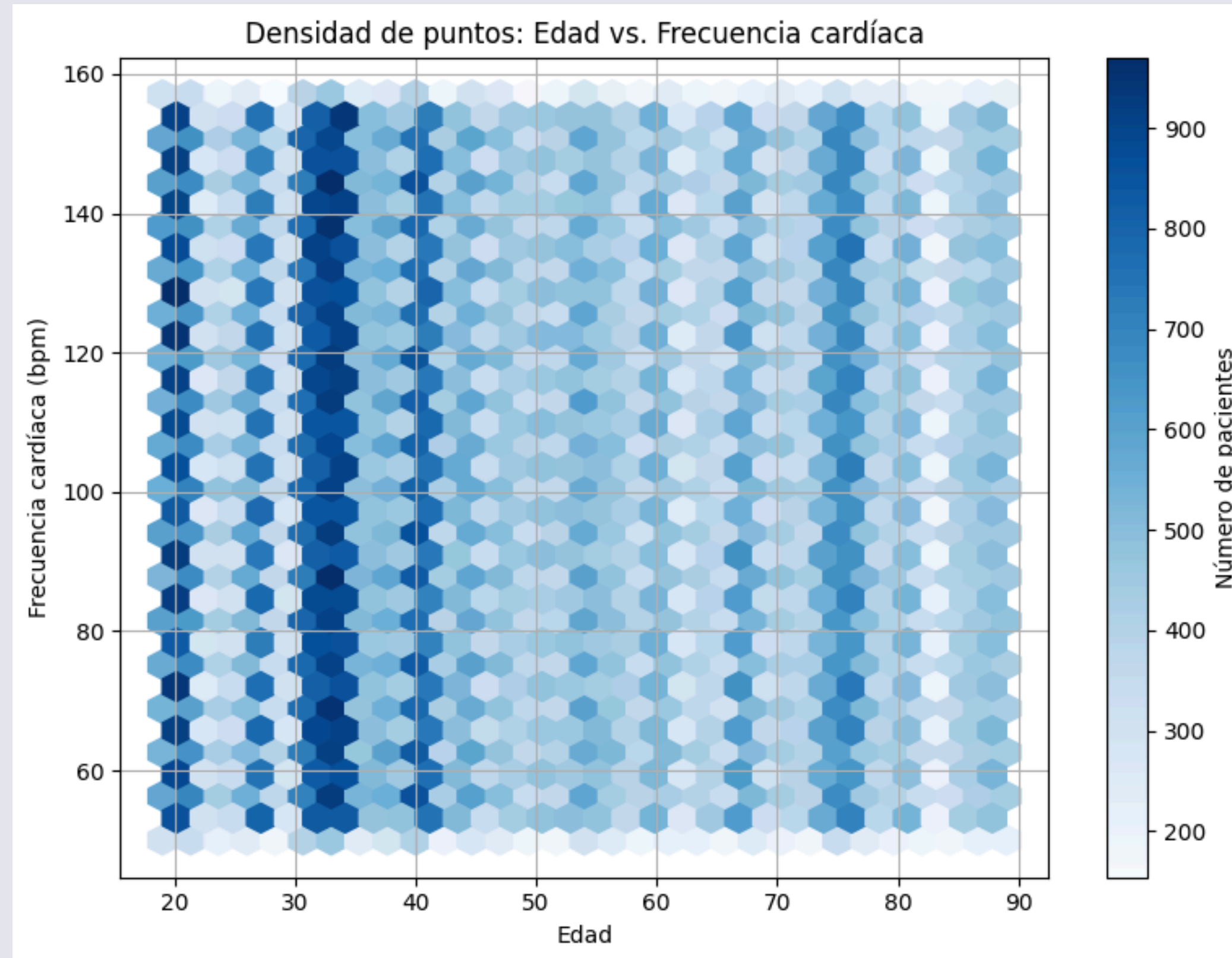
Análisis de outliers

- No hay Outliers

Relación de datos

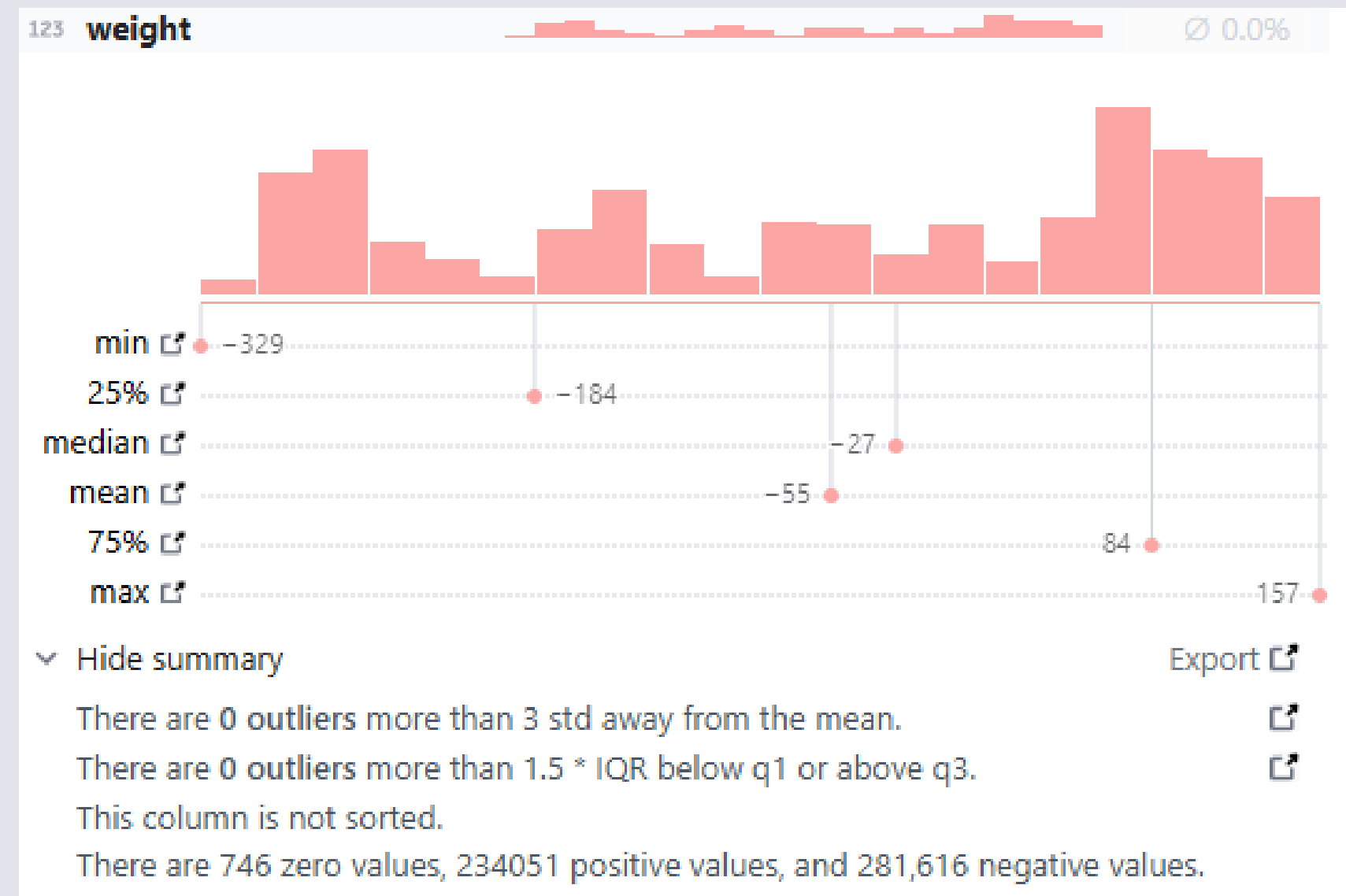


Relación de datos



Problemas encontrados en la data

- La columna weight contiene valores negativos





Preguntas



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



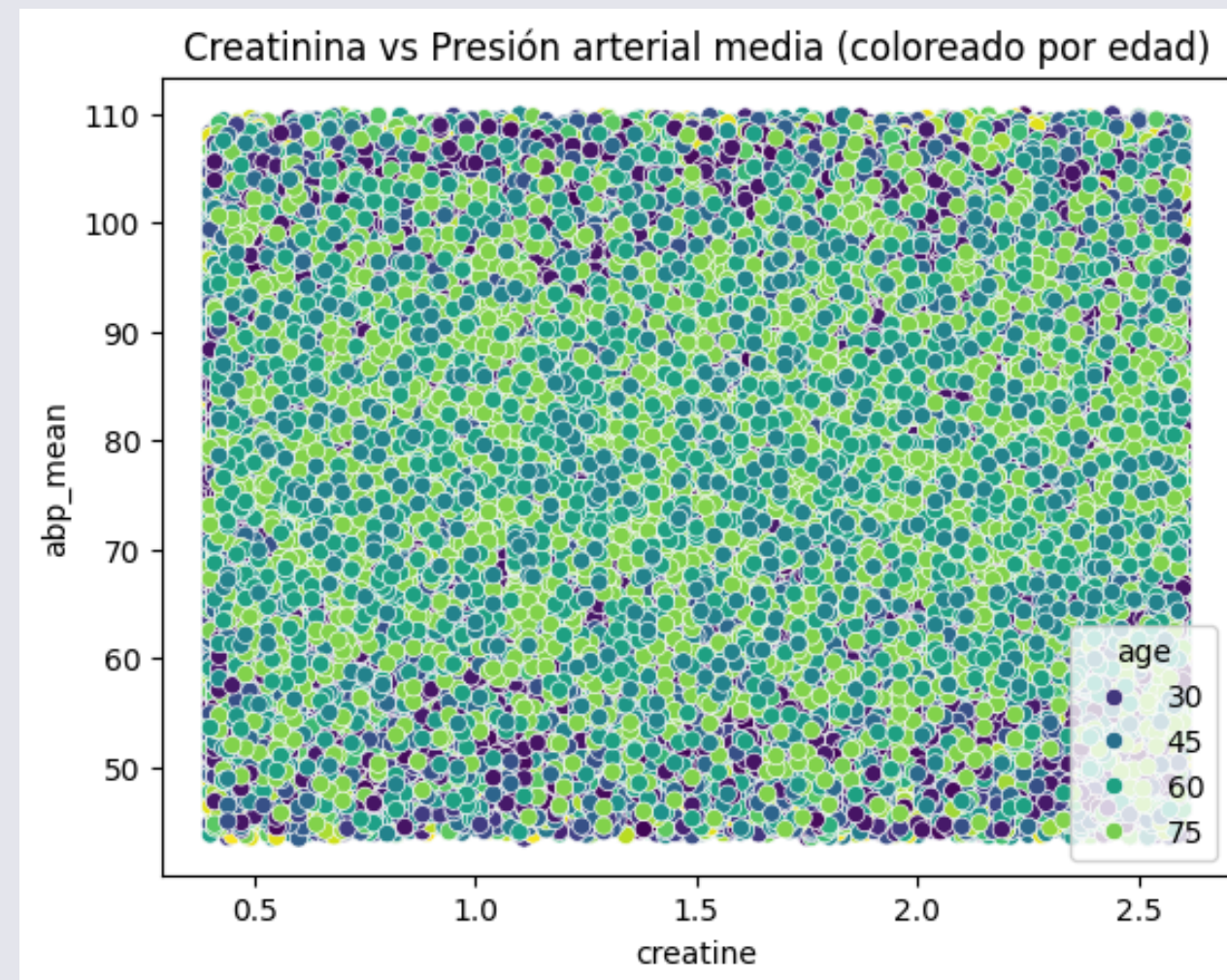
Niveles anormales de creatinina (creatine) están asociados con alteraciones en la presión arterial media (abp_mean)

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

¿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.





GRACIAS

Automatización del Análisis Exploratorio en Ciencia de Datos con AutoProfiler

INTEGRANTES

- Roy Angel Choquehuanca Anconeyra