

Planilla de dudas sobre variables del dataset

Equipo ML

26 de octubre de 2025

Resumen del dataset

Nombre del dataset	RECUIMA (Registro Cubano de Infarto Agudo de Miocardio)
Versión	v0.2 (post-limpieza inicial)
Fuente / Sistema origen	Registro hospitalario de pacientes con infarto agudo de miocardio
Periodo de cobertura	2016-2025
Población y unidad de análisis	Pacientes ingresados con diagnóstico de IAM;unidad: episodio de internación
Número de registros / variables	3,112 registros / 185 variables (después de limpieza inicial)
Fecha de extracción	02/04/2025
Objetivo analítico	Predicción de mortalidad intrahospitalaria en pacientes con IAM mediante modelos de machine learning
Restricciones legales / privacidad	Se eliminaron variables de identificación personal (nombres, números de identidad, números de contacto) para cumplir con protección de datos
Notas generales de calidad (duplicados, faltantes, codificaciones especiales)	Se observa presencia significativa de valores faltantes en múltiples variables. Existen duplicaciones de variables clave que requieren aclaración. El dataset parece ser resultado de la fusión de múltiples registros o fuentes. Se eliminaron variables redundantes identificadas en la limpieza inicial: anno , numero (identificador), unidad .

Nota importante sobre duplicaciones: Se han identificado las siguientes variables que aparecen duplicadas en el dataset: **presion_arterial_sistolica**, **presion_arterial_diastolica**, **asa**, **betabloqueadores**, **ieca**, **estatinas**, **clopidogrel**, **furosemida**, **nitratos**, **anticoagulantes**, **otros_diureticos**, **fecha_egreso** y **fecha_ingreso**. Se desconoce el motivo exacto de estas duplicaciones, aunque se presume que puede deberse a: (1) registro en diferentes momentos temporales (ingreso vs. egreso), (2) fusión de múltiples fuentes de datos, o (3) diferencias entre prescripción y administración real. Se requiere aclaración urgente sobre la interpretación correcta de estas columnas duplicadas.

Guía rápida

- Tipo: numérico, categórico, booleano, fecha/hora, texto libre, identificador. - Códigos especiales: por ejemplo, -1, 9, 99, 999 = “desconocido/no aplica”. - Estados: Pendiente, Enviado, Resuelto, Rechazado, En progreso.

1. Tabla maestra de variables

Cuadro 1: Tabla maestra de variables.

Variable	Descripción	Tipo	Unidad / Rango o Dominio	Faltantes y códigos	Reglas / Validación	Dudas principales
numero	Identificador único del paciente	numérico	enteros positivos	ninguno	único por paciente	Eliminada varias veces: variable redundante

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
anno	Año de registro	numérico	2016–2025	ninguno	–	Eliminada varias veces: variable redundante
unidad	Código de unidad hospitalaria	numérico	códigos específicos	presentes	–	Eliminada varias: variable redundante
fecha ingreso	Fecha de ingreso hospitalario	fecha	formato dd/mm/yyyy	presentes	fecha válida	DUPLICADA: requiere aclaración
fecha egreso	Fecha de egreso hospitalario	fecha	formato dd/mm/yyyy	presentes	>= fecha_ - ingreso	DUPLICADA: asociada a rein-gresos
numero identidad	Número de documento de identidad	texto	–	–	–	Eliminada: protección de datos
numero contacto	Número telefónico de contacto	texto	–	–	–	Eliminada: protección de datos
nombre	Nombre del paciente	texto	–	–	–	Eliminada: protección de datos
primer apellido	Primer apellido del paciente	texto	–	–	–	Eliminada: protección de datos
segundo apellido	Segundo apellido del paciente	texto	–	–	–	Eliminada: protección de datos
edad	Edad del paciente en años	numérico	0–120 años	escasos	>0, <120	Ninguna
sexo	Sexo del paciente	categorico	masculino, femenino	presentes	dominio cerrado	Requiere codificación binaria
color piel	Etnia o color de piel registrado	categorico	blanca, mestiza, negra	presentes	dominio cerrado	Requiere codificación numérica
peso	Peso corporal del paciente	numérico	kg, 20–200	presentes	>0	Ninguna
talla	Estatura del paciente	numérico	cm, 100–220	presentes	>0	Ninguna
imc	Índice de masa corporal calculado	numérico	kg/m ² , 10–60	presentes	peso/(talla/100)	Ninguna
provincia	Nombre de la provincia	categorico	nombres normalizados	presentes	–	Valores con caracteres especiales
municipio	Nombre del municipio	categorico	nombres normalizados	presentes	–	Valores con caracteres especiales
area_ - salud	Código de área de salud	categorico	códigos específicos	presentes	–	Valores con caracteres especiales
idprovincia	Identificador numérico de provincia	numérico	enteros positivos	ninguno	–	DUPLICADA: columna repetida

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
idmunicipio	Identificador numérico de municipio	numérico	enteros positivos	ninguno	–	Ninguna
idareasalud	Identificador numérico de área de salud	numérico	enteros positivos	presentes	–	Ninguna
atencion inicial	Tipo de atención inicial recibida	categorico	servicio, cuerpo, sala	presentes	–	Requiere documentación del significado
horario llegada	Horario de llegada al hospital	categorico	7am7pm, 7pm7am	presentes	–	Requiere codificación binaria
ecg_pre- vio	Electrocardiograma previo realizado	booleano	si, no	presentes	–	Requiere codificación binaria
ecg	Código de hallazgo electrocardiográfico	numérico	enteros (5–35)	presentes	–	Requiere tabla de correspondencia
llamada emergencias	Llamada al servicio de emergencias	categorico	si, no	presentes	–	Requiere codificación binaria
tiempo respuesta	Tiempo de respuesta de emergencias	numérico	probablemente minutos	presentes (solo si llamada=si)	≥ 0	Requiere confirmación de unidad
tiempo llegada	Tiempo de llegada al hospital	numérico	probablemente minutos	presentes (solo si llamada=si)	≥ 0	Requiere confirmación de unidad
primera asistencia medica	Tiempo hasta primera asistencia	numérico	probablemente minutos	presentes	≥ 0	Requiere confirmación de unidad
scacest	Síndrome coronario agudo con elevación ST	booleano	0, 1	presentes	–	Requiere explicación clínica detallada
scacest secundario	SCACEST secundario	booleano	0, 1	presentes	–	Requiere explicación clínica detallada
angina	Tipo de angina presentada	categorico	inestable, (otros)	presentes	–	Requiere dominio completo de valores
angina24h	Angina en las últimas 24 horas	booleano	0, 1	presentes	–	Requiere aclaración del período exacto
angina inestable	Presencia de angina inestable	booleano	0, 1	muchos	–	Posible variable de seguimiento
presentacion	Forma de presentación clínica	categorico	dolor_tipico, dolor_atipico, sincope, asintomatico, otros	presentes	–	Ninguna
depresion st	Depresión del segmento ST	booleano	si, no	presentes	–	Requiere explicación clínica y codificación

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
depression ondat	Depresión de onda T	categorico	(valores diversos)	presentes	–	Requiere expli- cación clínica completa
supradesni- vel	Supradesnivel del segmento ST	numérico	enteros	muchos	–	Requiere expli- cación clínica y unidad
infradesni- vel	Infradesnivel del segmento ST	numérico	enteros	muchos	–	Requiere expli- cación clínica y unidad
v1	Derivación precor- dial V1	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v2	Derivación precor- dial V2	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v3	Derivación precor- dial V3	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v4	Derivación precor- dial V4	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v5	Derivación precor- dial V5	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v6	Derivación precor- dial V6	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v7	Derivación precor- dial V7	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v8	Derivación precor- dial V8	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v9	Derivación precor- dial V9	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
d1	Derivación de miembro D1	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
d2	Derivación de miembro D2	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
d3	Derivación de miembro D3	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
avl	Derivación aumen- tada AVL	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
avf	Derivación aumen- tada AVF	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
avr	Derivación aumen- tada AVR	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v3r	Derivación precor- dial derecha V3R	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
v4r	Derivación precor- dial derecha V4R	booleano	0, 1	muchos	–	Requiere explica- ción: ¿alteración presente?
presion_ arterial_ sistolica	Presión arterial sistólica	numérico	mmHg, 60–250	escasos	>0	DUPLICADA: requiere aclara- ción
presion_ arterial_ diastolica	Presión arterial diastólica	numérico	mmHg, 40–150	escasos	>0	DUPLICADA: requiere aclara- ción
frecuencia_ cardiaca	Frecuencia cardí- aca	numérico	latidos/min, 30– 200	escasos	>0	Ninguna
shock	Presencia de shock	booleano	0, 1	presentes	–	Requiere tipo de shock (cardiogé- nico, etc.)
indice_ mkillip	Clasificación Killip modificada	categorico	I, II, III, IV (ro- manos)	presentes	–	Requiere explica- ción de diferen- cia con Killip
indice_ killip	Clasificación Killip	categorico	I, II, III, IV (ro- manos)	presentes	–	Requiere explica- ción de diferen- cia con Killip-M
ingresos_ anteriores	Ingresos previos por IAM	booleano	0, 1	presentes	–	Requiere codifi- cación binaria
diabetes_ mellitus	Diabetes mellitus previa	booleano	0, 1	ninguno	–	Ninguna
insulina	Tratamiento con insulina	numérico	(valores diversos)	muchos	–	Requiere acla- ración: ¿dosis, tipo, booleano?
insuficiencia cardiaca congestiva	Insuficiencia car- díaca congestiva previa	booleano	0, 1	ninguno	–	Ninguna
insuficiencia cardiaca	Insuficiencia car- díaca (evento)	booleano	0, 1	muchos	–	Diferencia con ICC como ante- cedente
hipertension arterial	Hipertensión arte- rial previa	booleano	0, 1	ninguno	–	Ninguna
hiperlipo- proteinemia	Dislipidemia pre- via	booleano	0, 1	ninguno	–	Ninguna

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
enfermedad arterias coronarias	Enfermedad ar- terial coronaria previa	booleano	0, 1	ninguno	–	Ninguna
infarto_- miocar- dio_agudo	Infarto de miocar- dio previo	booleano	0, 1	ninguno	–	Ninguna
fibrilacion_ auricular	Fibrilación auricu- lar previa	booleano	0, 1	ninguno	–	Ninguna
intervencion_ coronaria percutanea	ICP previa	booleano	0, 1	ninguno	–	Ninguna
cabg	Cirugía de revascu- larización corona- ria previa (CABG)	booleano	0, 1	ninguno	–	Requiere aclara- ción del acróni- mo CABG
enfermedad venosa periferica	Enfermedad veno- sa periférica	booleano	0, 1	ninguno	–	Ninguna
insuficiencia renal_cro- nica	Insuficiencia renal crónica	booleano	0, 1	ninguno	–	Ninguna
dialisis	Tratamiento con diálisis	booleano	0, 1	ninguno	–	Ninguna
enfermedad cerebro_- vascular	Enfermedad cere- brovascular previa	booleano	0, 1	ninguno	–	Ninguna
anemia	Anemia previa o actual	booleano	0, 1	ninguno	–	Ninguna
epoc	Enfermedad pul- monar obstructiva crónica	booleano	0, 1	ninguno	–	Ninguna
otros	Otros antecedentes	booleano	0, 1	presentes	–	Requiere espe- cificación de contenido
otras	Otras complicacio- nes o condiciones	booleano	0, 1	muchos	–	Requiere espe- cificación de contenido
tabaquismo	Antecedente de tabaquismo	booleano	0, 1	ninguno	–	Ninguna
tipo_taba- quismo	Tipo de tabaquis- mo	categorico	activo, noactivo	presentes	–	Requiere codifi- cación binaria
annos_- fumando	Años de consumo de tabaco	numérico	años, 0–80	presentes	>=0	Ninguna
annos_- sin_fumar	Años sin fumar (ex-fumadores)	numérico	años, 0–80	muchos	>=0	Ninguna
asa	Ácido acetilsalicí- lico (aspirina)	numérico	(valores diversos)	presentes	–	DUPLICADA: ¿dosis o boo- leano?
betabloquea- dores	Beta bloqueadores	numérico	(valores diversos)	presentes	–	DUPLICADA: requiere escala o codificación

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Rango o Dominio	Faltantes y códigos	Reglas / Validación	Dudas principales
clopidogrel	Clopidogrel	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
heparina	Heparina	categorico	si, no	presentes	–	Requiere codificación binaria
estatinas	Estatinas (hipolipemiantes)	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
furosemida	Furosemida (diurético de asa)	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
nitratos	Nitratos	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
anticoagulantes	Anticoagulantes	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
anticalcicos	Antagonistas del calcio	categorico	si, no	presentes	–	Requiere codificación binaria
anticalcico	Antagonista del calcio (¿duplicado?)	booleano	0, 1	presentes	–	Posible duplicación de anticalcicos
ieca	Inhibidores de ECA	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
otros__diureticos	Otros diuréticos	categorico	si, no	presentes	–	DUPLICADA: requiere codificación binaria
estreptoquinasa recombinante	Administración de estreptoquinasa	categorico	si, no	presentes	–	Requiere codificación binaria
tiempo__puerta__aguja	Tiempo puerta-aguja	numérico	probablemente minutos	presentes	>=0	Requiere confirmación de unidad y definición
tiempo__isquemia	Tiempo de isquemia	numérico	probablemente minutos	presentes	>=0	Requiere confirmación de unidad
escala__grace	Puntaje de riesgo GRACE	numérico	0–372	presentes	–	Requiere validación de rango y significado
reperfusion	Tipo de reperfusión realizada	categorico	no, parcial, total, otro	presentes	–	Requiere codificación y explicación detallada
coronariografia	Coronariografía realizada	categorico	no, si, otro, centro	presentes	–	Requiere codificación y explicación
lugar__trombolisis	Lugar de trombolisis	categorico	ucie, sala, servicio, otro	presentes	–	Candidata a eliminación según análisis

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
ergometria	Prueba de esfuerzo (ergometría)	categorico	positiva, negativa, no	muchos	–	Requiere domi- nio completo y codificación
avc	Asistencia ventri- cular o complica- ción	booleano	0, 1	ninguno	–	Requiere aclara- ción del acróni- mo
mpt	Marcapaso tempo- ral	booleano	0, 1	ninguno	–	Requiere con- firmación del acrónimo
vam	Ventilación asisti- da mecánica	booleano	0, 1	ninguno	–	Requiere con- firmación del acrónimo
mpp	Marcapaso perma- nente	booleano	0, 1	ninguno	–	Requiere con- firmación del acrónimo
aminas	Uso de aminas vasoactivas	booleano	0, 1	ninguno	–	Ninguna
balon	Balón de contra- pulsación intraaór- tico	booleano	0, 1	ninguno	–	Ninguna
proxima	Próxima consulta o evento	numérico	enteros	muchos	–	Requiere ex- plicación del significado
fecha_- consulta	Fecha de consulta de seguimiento	fecha	formato dd/mm/yyyy	muchos	fecha >= fe- cha_egreso	Ninguna
resultado	Resultado del se- guimiento	categorico	vivo_sin, vivo_- con, noevaluado, alta, fallecido	muchos	–	Requiere explica- ción de catego- rías
fecha_- defuncion	Fecha de defunción	fecha	formato dd/mm/yyyy	muchos	>= fecha_- ingreso	Ninguna
motivo	Motivo de reingre- so	categorico	scacest (principal- mente)	muchos	–	Asociada a fe- cha_ingreso duplicada
fecha_- realizacion	Fecha de realiza- ción de procedi- miento	fecha	formato dd/mm/yyyy	muchos	–	Requiere espe- cificación de procedimiento
razones_- documen- tadas	Razones documen- tadas	booleano	0, 1	muchos	–	Requiere contex- to: ¿razones de qué?
riesgo_- beneficio	Evaluación riesgo- beneficio	booleano	0, 1	muchos	–	Requiere contex- to específico
anti_agre- gacion_- plaqueta- ria	Antiagregación plaquetaria	booleano	0, 1	muchos	–	Requiere aclara- ción del contex- to
proteccion_ embolica	Protección embóli- ca	booleano	0, 1	muchos	–	Requiere aclara- ción del contex- to
funcion_- renal	Evaluación de fun- ción renal	booleano	0, 1	muchos	–	Requiere aclara- ción del contex- to

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
volumen_- contraste	Volumen de con- traste utilizado	booleano	0, 1	muchos	–	Requiere aclara- ción: ¿adecua- do/inadecuado?
prescripcion- optima	Prescripción ópti- ma al egreso	booleano	0, 1	muchos	–	Requiere defini- ción de "óptima"
rehabilitacion	Rehabilitación cardíaca	booleano	0, 1	muchos	–	DUPLICADA: aparece dos ve- ces
participacion- registro	Participación en registro	booleano	0, 1	muchos	–	Requiere aclara- ción del propósi- to
coronariografia- medico	Coronariografías por médico	booleano	0, 1	muchos	–	Requiere aclara- ción del significa- do
coronariografia- centro	Coronariografías por centro	booleano	0, 1	muchos	–	Requiere aclara- ción del significa- do
idresultado	Identificador de resultado	numérico	enteros	muchos	–	Requiere explica- ción del sistema de ID
arteria	Arteria afectada	categorico	cod, cx, ada, otros	muchos	–	Requiere tabla completa de códigos
localizacion	Localización de lesión	categorico	proximal, media, distal, sin_lesio- nes	muchos	–	Ninguna
estenosis	Grado de estenosis	numérico	porcentaje, 0–100	muchos	0–100	Ninguna
abordaje	Tipo de abordaje	categorico	stent_farmaco, stent_metalico, ninguno	muchos	–	Requiere tabla completa de valores
fecha_- egreso	Fecha de egreso hospitalario	fecha	formato dd/mm/yyyy	presentes	>= fecha_- ingreso	DUPLICADA: requiere aclara- ción
estado_- vital	Estado vital al egreso	categorico	vivo, fallecido	presentes	–	Requiere codifi- cación binaria
otra_- institucion	Traslado a otra institución	categorico	si, no	muchos	–	Requiere codifi- cación binaria
estadia_- ucie	Estadía en UCIE	numérico	días	presentes	>=0	Ninguna (uni- dad confirmada: días)
estadia_- uci	Estadía en UCI	numérico	días	presentes	>=0	Ninguna (uni- dad confirmada: días)
estadia_- intrahospi- talaria	Estadía intrahospi- talaria total	numérico	días	presentes	>=0	Ninguna (uni- dad confirmada: días)
consejeria_- antitaba- quica	Consejería antita- báquica brindada	booleano	0, 1	presentes	–	Ninguna

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
consejeria	Consejería general brindada	booleano	0, 1	presentes	–	Requiere espe- cificación del tipo
dieta	Consejería dietéti- ca brindada	booleano	0, 1	presentes	–	Ninguna
complicaciones	Complicaciones durante hospitali- zación	categorico	valores diversos	muchos	–	Requiere tabla de códigos com- pleta
observaciones	Observaciones clínicas libres	texto libre	texto único por caso	muchos	–	Ninguna
colesterol	Colesterol total	numérico	mg/dL o mmol/L	muchos	>0	Requiere con- firmación de unidad
creatinina	Creatinina sérica	numérico	mg/dL o μ mol/L	muchos	>0	Requiere con- firmación de unidad
filtrado_- glomerular	Tasa de filtrado glomerular estima- do	numérico	mL/min/1.73m ²	muchos	>0	Requiere con- firmación de unidad
trigliceridos	Triglicéridos	numérico	mg/dL o mmol/L	muchos	>0	Requiere con- firmación de unidad
glicemia	Glucemia	numérico	mg/dL o mmol/L	muchos	>0	Requiere con- firmación de unidad
leuco	Leucocitos	numérico	células/mm ³ o $\times 10^3/\mu$ L	muchos	>0	Requiere con- firmación de unidad
hb	Hemoglobina	numérico	g/dL	muchos	>0	Requiere con- firmación de unidad
ck	Creatina quinasa	numérico	U/L	muchos	>0	Requiere con- firmación de unidad
ckmb	Creatina quinasa MB	numérico	U/L o ng/mL	muchos	>0	Requiere con- firmación de unidad
acd	Arteria coronaria derecha	booleano	0, 1	presentes	–	Requiere con- firmación del acrónimo
ada	Arteria descenden- te anterior	booleano	0, 1	presentes	–	Requiere con- firmación del acrónimo
acx	Arteria circunfleja	booleano	0, 1	presentes	–	Requiere con- firmación del acrónimo
fraccion_- eyeccion	Fracción de eyec- ción del VI	numérico	porcentaje, 0–100	muchos	0–100	Ninguna
ddvi	Diámetro diastóli- co del VI	numérico	mm	muchos	>0	Requiere con- firmación de unidad

Continúa en la siguiente página

Continuación de la Tabla maestra de variables (encabezado repetido)

Variable	Descripción	Tipo	Unidad / Ran- go o Dominio	Faltantes y códigos	Reglas / Va- lidación	Dudas princi- pales
dsvi	Diámetro sistólico del VI	numérico	mm	muchos	>0	Requiere con- firmación de unidad
tiv	Grosor del tabique interventricular	numérico	mm	muchos	>0	Requiere con- firmación de unidad
pp	Grosor de pared posterior	numérico	mm	muchos	>0	Requiere con- firmación de unidad
ud	Parámetro ecocardiográfico (desconocido)	numérico	(unidad desconocida)	muchos	>0	Requiere aclaración completa
fs	Fracción de acortamiento	numérico	porcentaje, 0–100	muchos	0–100	Requiere confirmación
tapse	TAPSE (función ventricular derecha)	numérico	mm	muchos	>0	Requiere con- firmación de unidad
ea	Relación E/A mitral (función diastólica)	numérico	ratio	muchos	>0	Requiere confirmación
ee	Relación E/e' (función diastólica)	numérico	ratio	muchos	>0	Requiere confirmación
pat	Parámetro ecocardiográfico (desconocido)	numérico	(unidad desconocida)	muchos	>0	Requiere aclaración completa
insao	Insuficiencia aórtica	booleano	0, 1	muchos	–	Requiere aclaración del acrónimo
estao	Estenosis aórtica	booleano	0, 1	muchos	–	Requiere aclaración del acrónimo
insmit	Insuficiencia mitral	booleano	0, 1	muchos	–	Requiere aclaración del acrónimo
estmit	Estenosis mitral	booleano	0, 1	muchos	–	Requiere aclaración del acrónimo
mortality_inhospital	Mortalidad intrahospitalaria	booleano	0, 1	ninguno	–	Variable objetivo del modelo

2. Registro de dudas y resoluciones

ID	Variable	Duda / Observación	Evidencia / Contexto	Impacto	Prioridad	Responsable (consulta)	Estado	Fechas (sol./cierre)
001	Variables duplicadas (grupo)	Se identifican 12 variables que aparecen duplicadas en el dataset. Se requiere aclaración sobre el significado de cada columna duplicada y cuál utilizar para el análisis.	Variables afectadas: presion_arterial_sistolica, presion_arterial_diastolica, asa, betabloqueadores, ieca, estatinas, clopidogrel, furosemda, nitratos, anticoagulantes, otros_diureticos, fecha_ingreso. Posibles hipótesis: (a) medición en momentos diferentes (ingreso vs egreso), (b) fusión incorrecta de datasets, (c) prescripción vs administración real	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
002	asa	¿La variable ASA representa dosis en mg, días de tratamiento, o simplemente presencia/ausencia?	Valores observados son numéricos diversos (rango 0–24). No se identifica patrón claro de dosificación estándar	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
003	betabloqueadores	¿Qué escala o codificación se utiliza para betabloqueadores?	Valores numéricos diversos (rango 0–36). No se identifica si es dosis, días, o código categórico	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
004	ecg	Solicitar tabla de correspondencia completa para códigos ECG	Valores observados: 5, 10, 12, 15, 20, 25, 30, 35. Sin documentación disponible sobre significado clínico	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
005	scacest, scacest_secundario	Requiere explicación clínica detallada de la diferencia entre SCA-CEST primario y secundario	Ambas variables booleanas (0/1), pero no se comprende la distinción clínica ni criterios diagnósticos	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
006	indice_killip, indice_mkkillip	¿Cuál es la diferencia entre clasificación Killip y Killip modificada? ¿Cuándo se utiliza cada una?	Ambas utilizan números romanos I–IV. En algunos registros difieren, en otros coinciden	Medio	Alta	Cardiólogo	Pendiente	2025-10-25 /
007	tiempo_puerta_aguja	Confirmar definición clínica exacta y unidad de medida	Asumimos minutos desde llegada hospitalaria hasta inicio de trombolisis, pero requiere confirmación oficial	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
008	tiempo_isquemia	Confirmar definición clínica exacta y unidad de medida	Asumimos minutos desde inicio de síntomas hasta reperusión, pero requiere confirmación oficial	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
009	tiempo_respuesta, tiempo_llegada	Confirmar unidades de medida y definiciones exactas de cada variable	Valores presentes solo cuando llamada_emergencias=si. Asumimos minutos pero sin documentación	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
010	escala_grace	Validar rango de valores observados contra rango teórico esperado (0–372)	Valores parecen consistentes pero requiere validación clínica del cálculo	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
011	Derivaciones ECG (V1–V9, D1–D3, AVL, AVF, AVR, V3R, V4R)	¿Estas variables indican presencia de alteración en cada derivación? ¿Qué tipo de alteración (supradesnivel, infradesnivel, onda Q)?	Variables booleanas (0/1) con muchos faltantes. Requiere documentación de criterios diagnósticos utilizados	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /

ID	Variable	Duda / Observación	Evidencia / Contexto	Impacto	Prioridad	Responsable (consulta)	Estado	Fechas (sol./cierre)
012	avc, mpt, vam, mpp	Confirmar significado exacto de acrónimos	Asumimos: AVC=asistencia ventricular, MPT=marcapaso temporal, VAM=ventilación mecánica, MPP=marcapaso permanente	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
013	cabg	Confirmar acrónimo CABG (¿Coronary Artery Bypass Graft?)	Asumimos cirugía de revascularización coronaria pero requiere confirmación	Bajo	Baja	Cardiólogo	Pendiente	2025-10-25 /
014	reperfusion	Explicar diferencia entre categorías y criterios de clasificación	Valores: no, parcial, total, otro. Requiere definición de criterios clínicos utilizados	Alto	Alta	Cardiólogo	Pendiente	2025-10-25 /
015	coronario-grafía	Explicar diferencia entre valores observados	Valores: no, si, otro, centro. Especialmente aclarar significado de "centro" vs "otro"	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
016	Grupo variables de calidad	Grupo de variables con muchos faltantes que requieren contexto clínico	Variables: razones_documento, riesgo_beneficio, anti_agregacion_plaquetaria, proteccion_embolica, funcion_renal, volumen_contraste, prescripcion_optima. ¿Son indicadores de calidad o checklist?	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
017	Laboratorio (grupo)	Confirmar unidades de medida para todas las variables de laboratorio	Variables: colesterol, creatinina, filtrado_glomerular, trigliceridos, glicemia, leuco, hb, ck, ckmb. Especificar unidades (mg/dL, mmol/L, etc.)	Alto	Alta	Laboratorio	Pendiente	2025-10-25 /
018	Ecocardiografía (grupo)	Confirmar significado de acrónimos: ud, pat, insao, estao, insmit, estmit	Asumimos insuficiencias y estenosis valvulares. ud y pat sin identificar	Medio	Media	Cardiología	Pendiente	2025-10-25 /
019	arteria	Solicitar tabla completa de códigos de arterias coronarias	Códigos: cd, cx, ada. Asumimos: CD=coronaria derecha, CX=circunfleja, ADA=descendente anterior	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
020	abordaje	Solicitar tabla completa de tipos de abordaje/intervención coronaria	Valores: stent_farmaco, stent_metalico, ninguno. ¿Existen otros valores posibles?	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
021	resultado	Explicar categorías de seguimiento y su significado clínico	Valores: vivo_sin, vivo_con, noevaluado, alta, fallecido. Requiere explicación de "vivo_sin" vs "vivo_con" (¿con/sin complicaciones?)	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
022	angina24h	¿Se refiere a angina en 24h previas al ingreso o durante primeras 24h de hospitalización?	Variable booleana con contexto temporal ambiguo que afecta interpretación clínica	Bajo	Baja	Cardiólogo	Pendiente	2025-10-25 /
023	provincia, municipio, area_salud	Normalizar caracteres especiales en nombres geográficos	Caracteres especiales mal codificados: Sancti Spiritus, Cabaiguán, Camagüey, Manatí, Güines	Bajo	Baja	Admin. datos	Pendiente	2025-10-25 /

ID	Variable	Duda / Observación	Evidencia / Contexto	Impacto	Prioridad	Responsable (consulta)	Estado	Fechas (sol./cierre)
024	insulina	¿Variable representa dosis, tipo de insulina, duración de tratamiento, o uso binario?	Valores numéricos diversos sin patrón identificable. Dificulta su uso en modelado	Medio	Media	Cardiólogo	Pendiente	2025-10-25 /
025	lugar trombolisis	¿Es necesaria esta variable para el objetivo analítico del proyecto?	Valores: ucie, sala, servicio. Considerada candidata a eliminación según relevancia clínica	Bajo	Baja	Equipo ML	Pendiente	2025-10-25 /

Acciones siguientes y próximos pasos

Consultas prioritarias

- Duplicación de variables (Alta prioridad):** Aclarar el significado y uso correcto de las 12 variables que aparecen duplicadas en el dataset. Esta es la duda más crítica que afecta la calidad del análisis.
- Escalas y codificaciones de medicamentos:** Definir las escalas utilizadas para asa, betabloqueadores, insulina y confirmar codificaciones de otros medicamentos.
- Variables electrocardiográficas:** Proporcionar tabla de correspondencia para códigos ECG y documentar criterios para derivaciones (V1–V9, D1–D3, AVL, AVF, AVR, V3R, V4R).
- Variables de tiempo:** Confirmar definiciones clínicas exactas y unidades de medida para tiempo_puerta_aguja, tiempo_isquemia, tiempo_respuesta, tiempo_llegada.
- Clasificaciones clínicas:** Explicar diferencias entre SCACEST primario/secundario, Killip/Killip modificado, y categorías de perfusión.
- Acrónimos y términos técnicos:** Confirmar significado de acrónimos no documentados (avc, mpt, vam, mpp, cabg, acd, ada, acx, insao, estao, insmit, estmit, ud, pat).
- Tablas de códigos:** Solicitar tablas completas para arteria, abordaje, resultado, complicaciones.
- Especificar unidades de medida para todas las variables de laboratorio (colesterol, creatinina, filtrado glomerular, triglicéridos, glicemia, leucocitos, hemoglobina, CK, CK-MB).
- Confirmar rangos de referencia y límites de detección de los equipos utilizados.

Tareas del equipo de machine learning

- Mantener actualizado el Registro de dudas y resoluciones con estados y fechas.
- Documentar todas las decisiones de preprocesamiento tomadas en ausencia de aclaraciones.
- Priorizar variables según impacto en el objetivo analítico (predicción de mortalidad intrahospitalaria).
- Preparar pipeline de preprocesamiento flexible que permita incorporar aclaraciones posteriores.
- Normalizar caracteres especiales en nombres geográficos.

Nota sobre el proceso de limpieza inicial

Este dataset es resultado de una primera etapa de limpieza en la que se eliminaron:

- Variables de identificación personal (nombre, primer_apellido, segundo_apellido, numero_identidad, numero_contacto) para cumplir con protección de datos.

- Variables redundantes identificadas en análisis preliminar (anno, numero como identificador, unidad).

El dataset original parece ser resultado de la fusión de múltiples fuentes o registros hospitalarios, lo cual explicaría algunas de las duplicaciones observadas. Se requiere confirmación de esta hipótesis por parte del equipo responsable de la recolección de datos.