

main app

- Home
- Analisis Exploratorio
- Analisis Estancia
- Analisis RandomForest
- Coocurrencias Diagnost...
- Regresion Logistica Dia...

EpiScope

EpiScope Envigado

Analítica Predictiva en
Salud Pública

Proyecto desarrollado



EpiScope Envigado

Analítica predictiva para la planeación hospitalaria y epidemiológica en Envigado

Selecciona una página en el menú lateral para explorar los análisis, modelos y resultados del proyecto.

main app

-  Home
-  Análisis Exploratorio
-  Análisis Estancia
-  Análisis RandomForest
-  Coocurrencias Diagnost...
-  Regresión Logística Dia...



EpiScope

 EpiScope Envigado

Analítica Predictiva en
Salud Pública

 Proyecto desarrollado



Home

EpiScope Envigado — Analítica predictiva para la planeación hospitalaria y epidemiológica en Envigado

Hoy, el sistema de salud de Envigado enfrenta un reto silencioso pero crítico: la dificultad para anticipar la demanda hospitalaria. Los picos de atención llegan sin aviso, los diagnósticos se dispersan en bases de datos extensas y las decisiones se toman mirando hacia atrás, no hacia adelante.

EpiScope Envigado nace para cambiar eso.

Es un proyecto de analítica predictiva que aprovecha los registros **RIPS 2023–2024** para detectar patrones en los diagnósticos **CIE-10** y anticipar tendencias. Con este modelo, los hospitales podrán prever picos de atención, optimizar su capacidad instalada y asignar recursos de manera más eficiente.



1. Contexto territorial y epidemiológico

El municipio de **Envigado** cuenta con una extensión de **51 km²** y una densidad poblacional estimada de **4.868,7 habitantes por km²** para el año 2024.

La población total proyectada asciende a **248.304 habitantes**, con una distribución de **54,1 % mujeres y 45,9 % hombres**.

De ellos, el **96,9 %** reside en zona urbana y el **3,1 %** en zona rural, según el **DANE** (Censo 2018, proyecciones 2024).

En 2023, el municipio contaba con **652 camas hospitalarias, 133 salas** (30 quirófanos) y **90 camillas**, reflejando una red hospitalaria sólida pero exigida por la alta demanda.

El análisis de morbilidad muestra que las **enfermedades no transmisibles (ENT)** representan la mayor proporción de consultas en todos los grupos etarios.

Las **condiciones transmisibles, nutricionales y las lesiones por causas externas** también tienen una participación importante, especialmente en la infancia y juventud.

 *Fuente: Análisis de Situación de Salud Participativo (ASIS) – Municipio de Envigado, 2024.*



2. Problema

El sistema de salud de Envigado enfrenta un desafío creciente:

- La **alta demanda hospitalaria y la capacidad instalada limitada** generan picos de atención imprevisibles.
- Las decisiones se basan principalmente en **datos históricos**, dificultando anticipar brotes o variaciones en la demanda.

- No existen **herramientas locales de analítica predictiva** que integren los RIPS 2023–2024 para generar alertas tempranas o estimaciones de morbilidad.
-



3. Solución: *EpiScope Envigado*

EpiScope Envigado es un proyecto de analítica avanzada que utiliza los **RIPS de hospitalización 2023–2024** para **identificar patrones diagnósticos (CIE-10)** y **anticipar tendencias de morbilidad**.

El modelo predictivo busca:

- Detectar cambios en los patrones de enfermedad.
 - Estimar la demanda futura por diagnóstico y especialidad médica.
 - Fortalecer la toma de decisiones en salud pública con base en datos reales.
-



4. Objetivos del proyecto

Objetivo general

Evaluar un modelo predictivo basado en analítica avanzada de los RIPS de hospitalización (2023–2024) del municipio de Envigado, para identificar patrones diagnósticos (CIE-10) y anticipar tendencias de morbilidad que fortalezcan la planeación epidemiológica y la gestión eficiente de recursos hospitalarios.

Objetivos específicos

- 1.Diseñar e implementar la infraestructura de datos del proyecto mediante un proceso ETL (Extracción, Transformación y Carga) que permita la creación y gestión eficiente de la base de datos, asegurando la integración adecuada de las fuentes de información.
 - 2.Realizar un Análisis Exploratorio de Datos (EDA) para caracterizar la población hospitalizada en Envigado durante el periodo 2023–2024, identificando tendencias de morbilidad, frecuencias de diagnóstico y variables relevantes para el modelado predictivo.
 - 3.Construir y validar un modelo predictivo basado en los códigos CIE-10, empleando técnicas de machine learning y algoritmos supervisados de clasificación, que permitan inferir relaciones entre diagnósticos y anticipar eventos de salud, contribuyendo a la toma de decisiones estratégicas en salud pública.
-



5. Valor e impacto esperado

El proyecto permitirá:

- Incrementar la **eficiencia operativa** mediante una mejor planeación de camas y servicios.
 - **Reducir** la congestión hospitalaria.
 - **Optimizar la inversión pública**, proyectando un ahorro en costos de atención.
 - Fortalecer la **capacidad institucional** de análisis de datos en salud pública local.
-



6. Equipo de trabajo

Equipo interdisciplinario de analítica y salud pública:

- Laura María Jaramillo Sánchez
 - Joshua Mateo Quiroz Márquez
 - Daniel Gil Arbeláez
 - Diego Eusse
 - Juan David Galego Ramírez
-

main app

Home

Analysis Exploratorio

Analysis Estancia

Analysis RandomForest

Coocurrencias Diagnósticas

Regresion Logistica Diagnóstica



EpiScope Envigado

Analítica Predictiva en
Salud Pública Proyecto desarrollado
por:

Análisis Exploratorio de los RIPS

Caracterización general de la base de datos

Registros totales: **45,147**

Descripción del dataset

Distribuciones básicas

Diagnósticos y causas externas



Panel de inspección del dataset

Descripción de columnas

Estadísticas descriptivas

Mostrar Primeras filas

Analisis Exploratorio

	fact_id	Cod_IPS	ID	Fecha_Ingreso	Fecha_Egreso	Duracion_Dias	via_ingreso_id	esta
0	1	52660212605	PAC00001	2023-09-04	2023-09-14	11	3	
1	2	52660212605	PAC00002	2023-02-05	2023-02-28	24	2	
2	3	52660212605	PAC00003	2023-09-30	2023-09-30	1	3	
3	4	52660212605	PAC00004	2023-08-31	2023-09-29	30	3	
4	5	52660212605	PAC00005	2023-08-26	2023-09-01	7	3	
5	6	52660212605	PAC00006	2023-10-16	2023-10-20	5	3	
6	7	52660212605	PAC00007	2023-10-22	2023-10-29	8	3	
7	8	52660212605	PAC00008	2023-08-31	2023-09-03	4	3	
8	9	52660212605	PAC00009	2023-06-29	2023-07-02	4	3	
9	10	52660212605	PAC00010	2023-07-31	2023-08-05	6	3	

main app

- Home
- Analysis Exploratorio
- Analysis Estancia
- Analysis RandomForest
- Coocurrencias Diagnost...
- Regresion Logistica Dia...



EpiScope

EpiScope Envigado

Analítica Predictiva en
Salud Pública

Proyecto desarrollado

🔍 Análisis Exploratorio de los RIPS

Caracterización general de la base de datos

Registros totales: **45,147**

Descripción del dataset

Distribuciones básicas

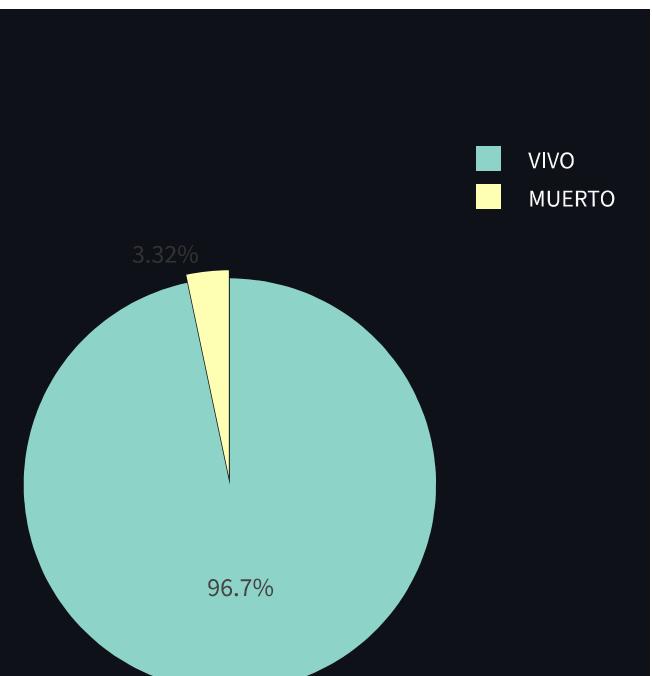
Diagnósticos y causas externas

Distribuciones principales

Distribución de la vía de ingreso

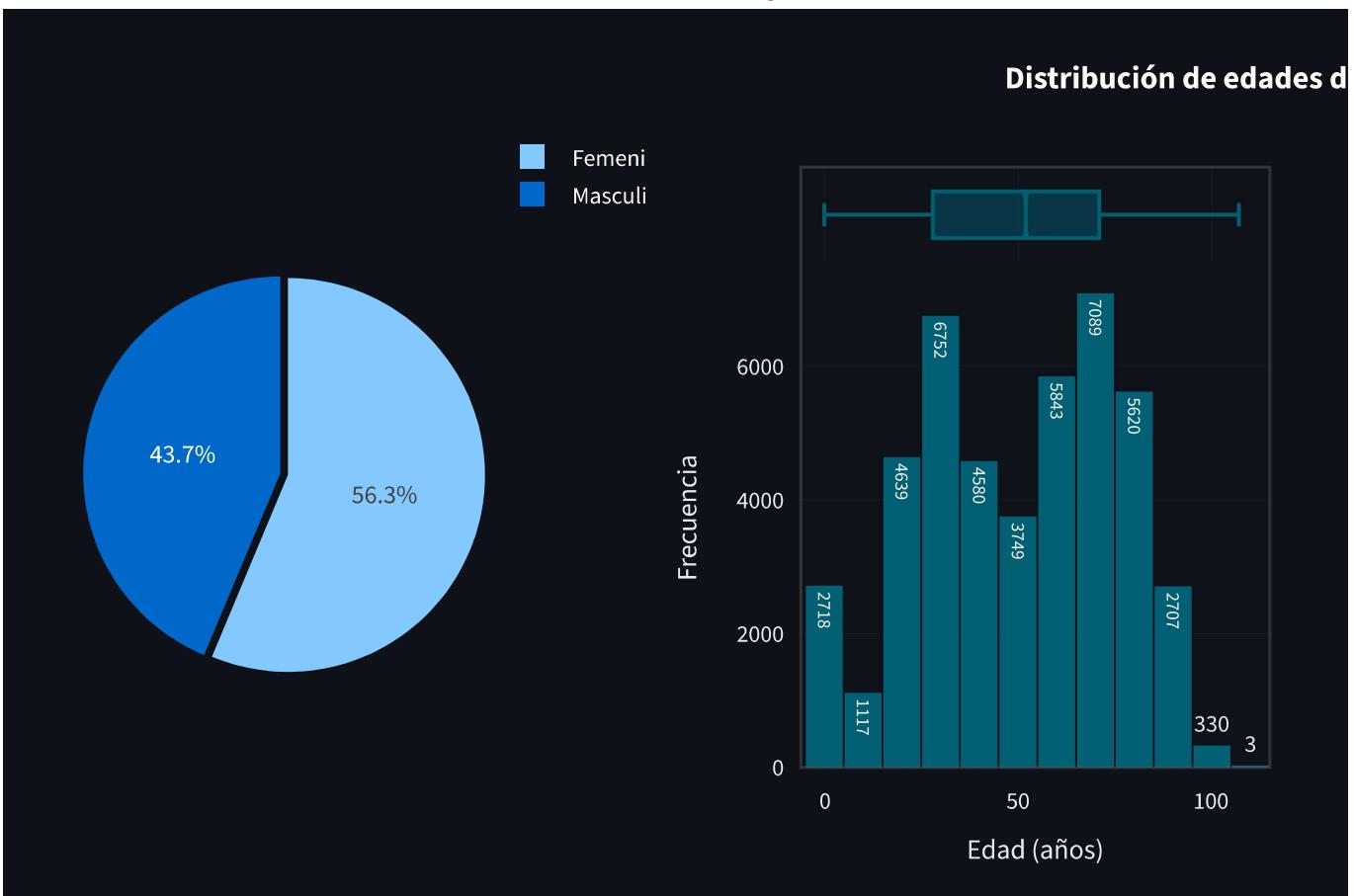


Estado de salida

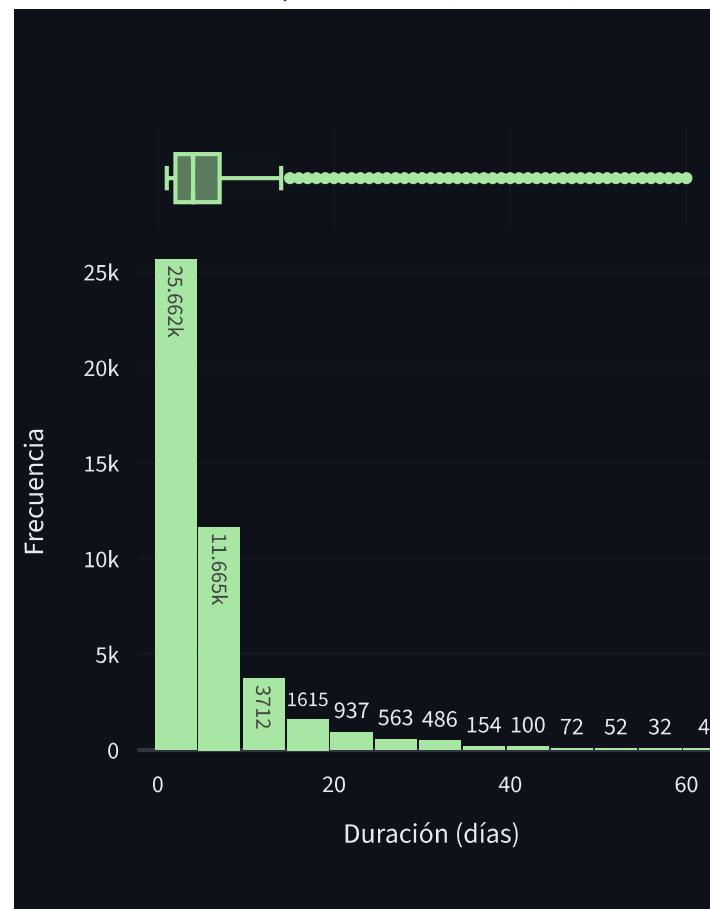




👤 Distribución por sexo



⌚ Duración de hospitalización (≤ 60 días)



main app

- Home
- Analysis Exploratorio
- Analysis Estancia
- Analysis RandomForest
- Coocurrencias Diagnósticas
- Regresion Logistica Diagnóstica



EpiScope

EpiScope Enviado

Analítica Predictiva en Salud Pública

Proyecto desarrollado por:

Análisis Exploratorio de los RIPS

Caracterización general de la base de datos

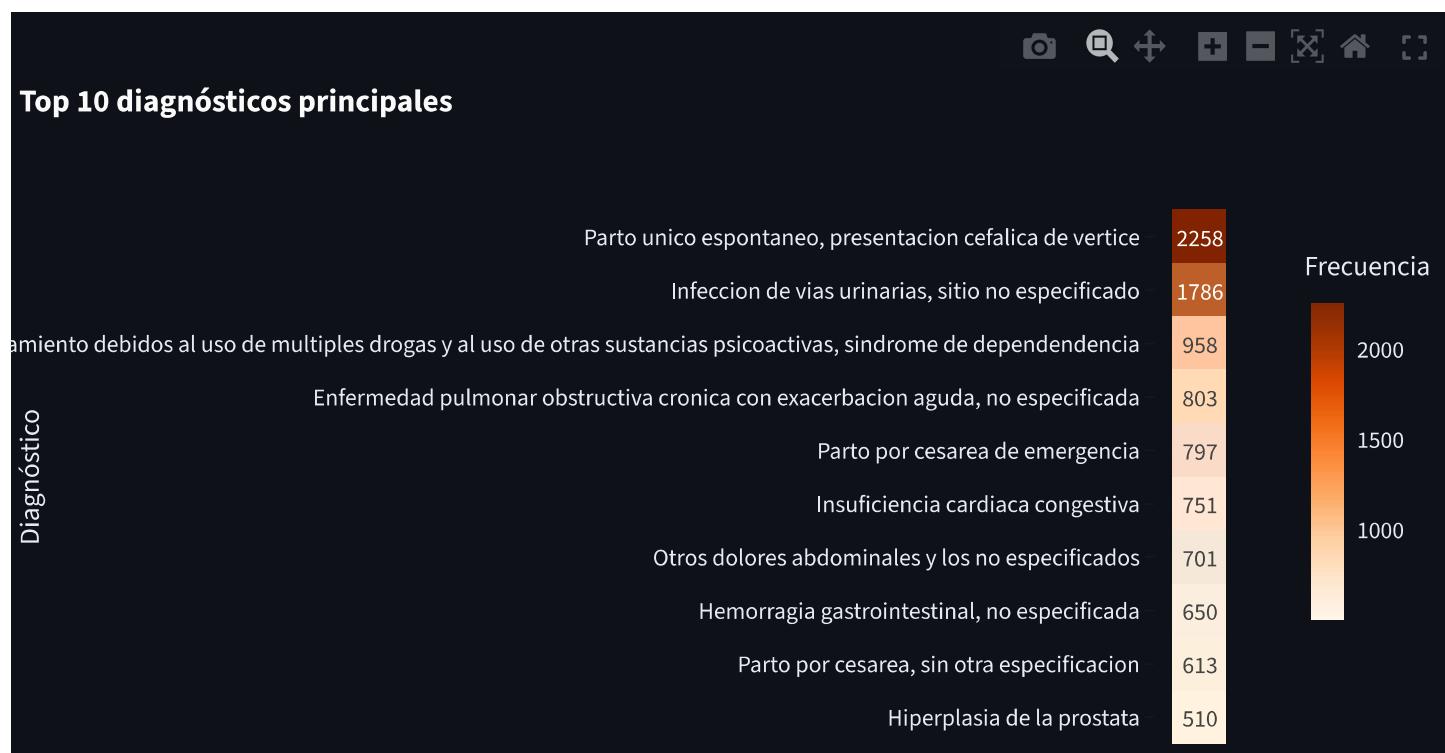
Registros totales: **45,147**

Descripción del dataset

Distribuciones básicas

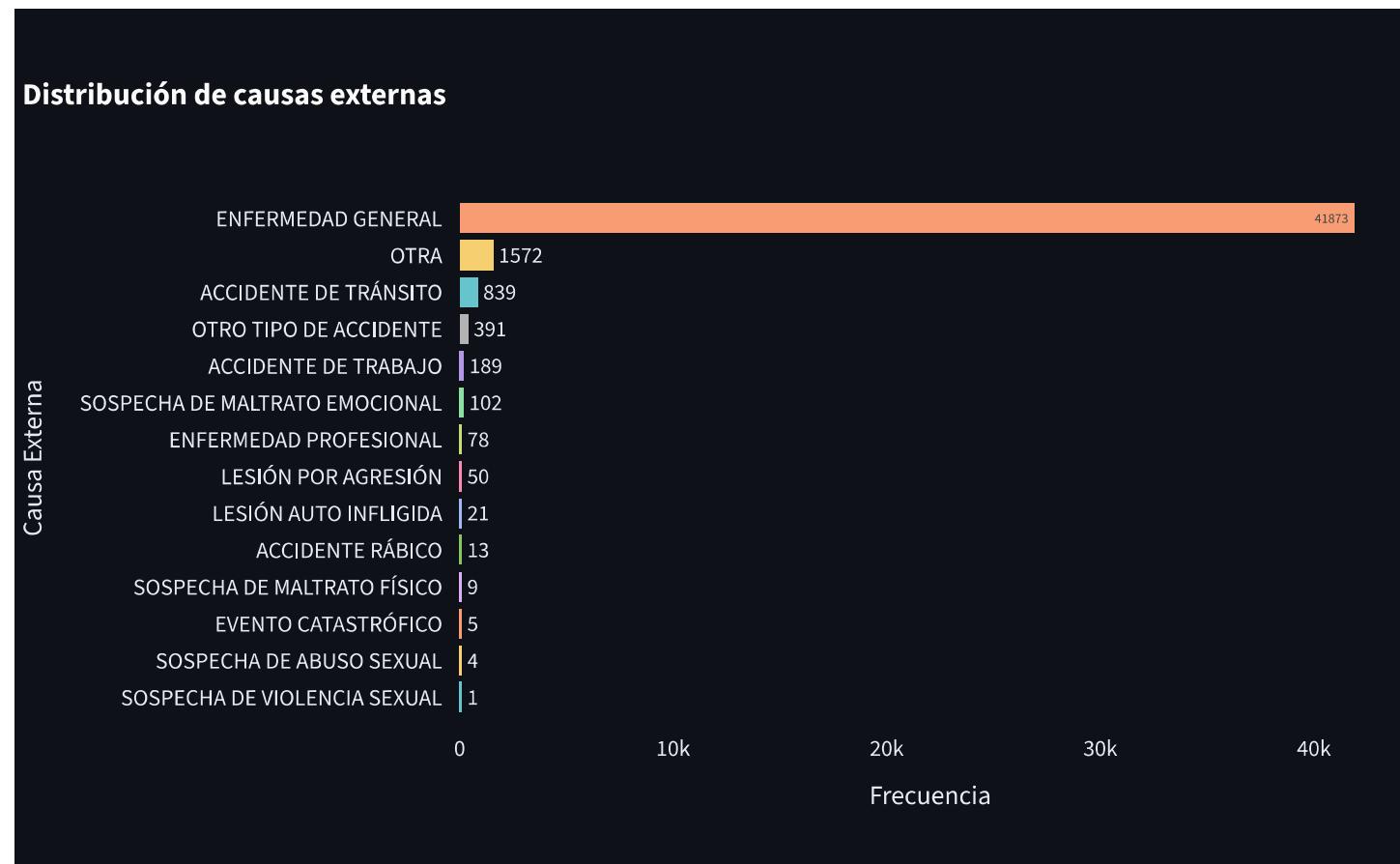
Diagnósticos y causas externas

Top 10 diagnósticos principales



Frecuencia

⚠ Distribución por causa externa



main app

 Home

 Analysis Exploratorio

 **Analysis Estancia**

 Analysis RandomForest

 Coocurrencias Diagnósticas

 Regresion Logistica Dia...

 EpiScope

 EpiScope Envigado

Analítica Predictiva en
Salud Pública

 Proyecto desarrollado
por:

Análisis de la Duración de la Estancia Hospitalaria

Explora paso a paso el proceso de análisis y modelado de la duración de la estancia hospitalaria. Cada sección puede expandirse con un *checkbox* para facilitar la exploración. Se incluyen explicaciones para interpretar los resultados estadísticos y gráficos.

 Cargar y visualizar datos

 Seleccionar y explorar variables

 Preparar datos para el modelo

Preparación de variables

Variables finales para el modelo:

X: 20 columnas | y: 45141 registros

	const	EDAD_ANIOS	SEXO_M	Via_Ingreso_Desc_NACIDO EN LA INSTITUCION	Via_Ingreso_Desc_REMITII
0	1	29	0		0
1	1	22	1		0
2	1	22	0		0
3	1	23	1		0
4	1	22	1		0

Interpretación:

- Se crean **variables dummy** para variables categóricas, necesarias para regresión lineal.
- `drop_first=True` evita la multicolinealidad perfecta.
- Se revisa que no haya valores nulos ni inconsistencias en X o y.

  Ajustar modelo de regresión OLS

Modelo de Regresión Lineal (OLS)



Resumen del modelo

	Métrica	Valor
0	R ²	0.0500
1	R ² ajustado	0.0490
2	F-Statistic	123.9520
3	Prob (F-stat)	0.0000

Métrica	Valor
4 Observaciones	45,141.0000
5 Condición numérica	11,836.2000

Interpretación estadística del resumen:

- **R²:** porcentaje de la variabilidad explicada por el modelo.
- **R² ajustado:** ajusta R² por número de variables; útil para comparar modelos con diferente cantidad de predictores.
- **F-Statistic y Prob (F-stat):** evalúa si el modelo completo es significativo.
- **Condición numérica:** alerta sobre posibles problemas de multicolinealidad alta.



Coeficientes del modelo

	Variable	Coeficiente	Error Std	t-Valor	p-Valor
3	Via_Ingreso_Desc_NACIDO EN LA INSTITUCION	9.9576	0.3698	26.9276	00000000000000000000000000000000
4	Via_Ingreso_Desc_REMITIDO	4.8151	0.1856	25.9395	00000000000000000000000000000000
2	SEXO_M	1.5507	0.0739	20.9831	00000000000000000000000000000000
5	Via_Ingreso_Desc_URGENCIAS	3.0417	0.1818	16.731	0.0000000000000000
1	EDAD_ANIOS	0.018	0.0015	12.0614	
19	Estado_Salida_Desc_VIVO	-1.296	0.2031	-6.3801	
10	Causa_Externa_Desc_EVENTO CATASTRÓFICO	16.8215	3.449	4.8772	
6	Causa_Externa_Desc_ACCIDENTE DE TRÁNSITO	2.8442	0.6129	4.6403	
11	Causa_Externa_Desc_LESIÓN AUTO INFILGIDA	4.4656	1.7511	2.5503	
0	const	1.5803	0.6259	2.5248	

Interpretación de coeficientes:

- Un coeficiente positivo indica que al aumentar la variable, la duración de estancia tiende a aumentar.
 - Un p-valor < 0.05 indica significancia estadística.
 - El error estándar indica precisión de la estimación.



Multicolinealidad (VIF)

	Variable	VIF
0	const	305.2155
8	Causa_Externa_Desc_ENFERMEDAD GENERAL	16.1926
13	Causa_Externa_Desc_OTRA	9.0697
5	Via_Ingreso_Desc_URGENCIAS	6.2024
4	Via_Ingreso_Desc_REMITIDO	6.1054
6	Causa_Externa_Desc_ACCIDENTE DE TRÁNSITO	5.3391
14	Causa_Externa_Desc_OTRO TIPO DE ACCIDENTE	3.0502
16	Causa_Externa_Desc_SOSPECHA DE MALTRATO EMOCIONAL	1.5378
9	Causa_Externa_Desc_ENFERMEDAD PROFESIONAL	1.412
3	Via_Ingreso_Desc_NACIDO EN LA INSTITUCION	1.3581
	,	,

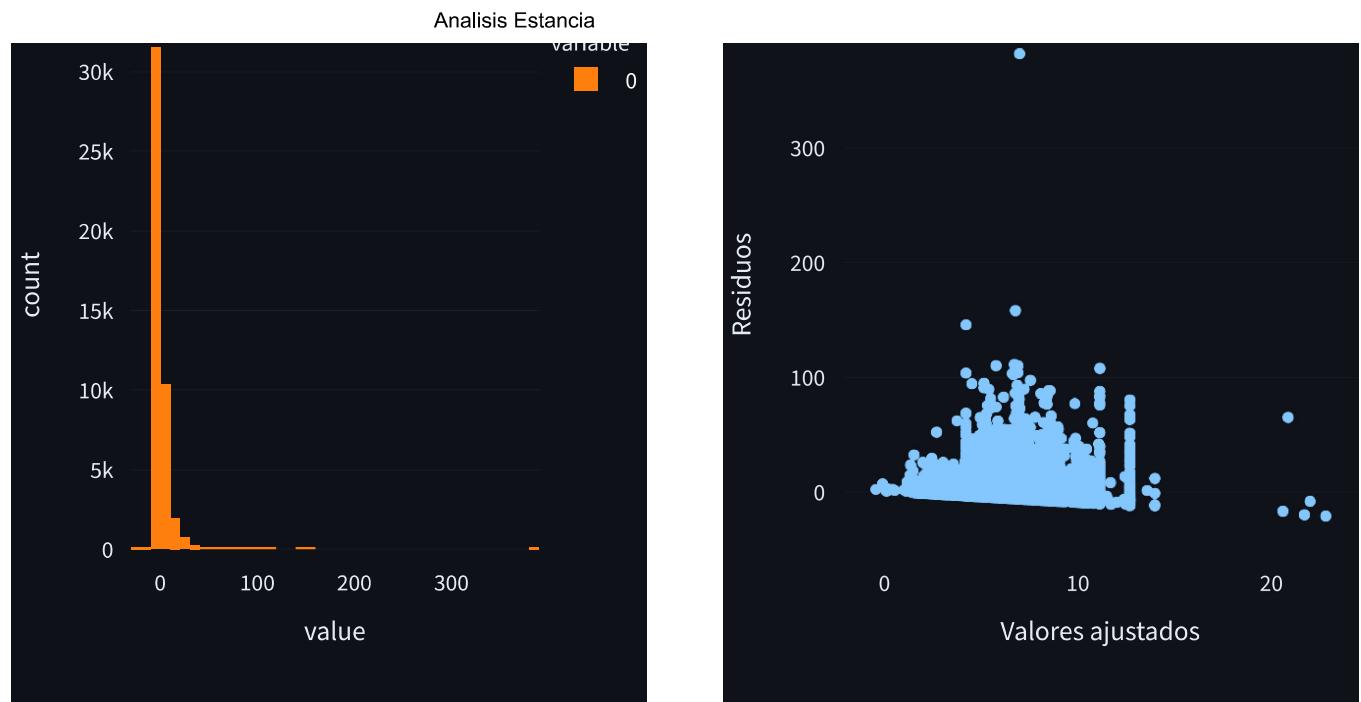
Interpretación de VIF:

- VIF > 5-10 indica posible multicolinealidad alta.
- Variables muy correlacionadas pueden inflar errores estándar y afectar interpretación de coeficientes.



Análisis de residuos





Interpretación de residuos:

- Los residuos deberían seguir distribución aproximadamente normal (histograma simétrico).
- Residuos vs valores ajustados permite verificar **homocedasticidad** (varianza constante).
- Patrones claros o funil en el scatter sugieren problemas en el modelo.

Predicción interactiva

Estimación de duración de estancia

Ingrasa los valores o selecciones los parámetros clínicos para estimar la duración de la estancia hospitalaria.

Edad (años)

40

Sexo

F

Vía de ingreso

- Seleccionne -

Estado de salida

- Seleccionne -

M

Causa externa

- Seleccione -



Calcular duración estimada

Interpretación de predicciones:

- El valor central es la duración estimada para un paciente con estas características.
- El **intervalo de confianza 95%** indica el rango donde se espera que caiga la duración real del paciente el 95% de las veces.

Este modelo usa regresión lineal (OLS) para estimar la duración de estancia hospitalaria según variables clínicas y demográficas.

Incluye análisis de residuos, multicolinealidad y predicciones con intervalo de confianza para facilitar interpretación estadística.

main app



Analysis Exploratorio

Analysis Estancia

Analysis RandomForest

Coocurrencias Diagnost...

Regresion Logistica Dia...



EpiScope Envigado

Analítica Predictiva en
Salud PúblicaProyecto desarrollado
por:

Análisis de la Duración de la Estancia Hospitalaria con Random Forest

Explora un modelo basado en **Random Forest** para estimar la duración de la estancia hospitalaria.

El enfoque no asume linealidad ni normalidad, y puede capturar relaciones no lineales entre variables.

Cargar y visualizar datos

Seleccionar y explorar variables

Preparar datos para el modelo

Entrenar modelo Random Forest

Entrenamiento y evaluación

Número de árboles

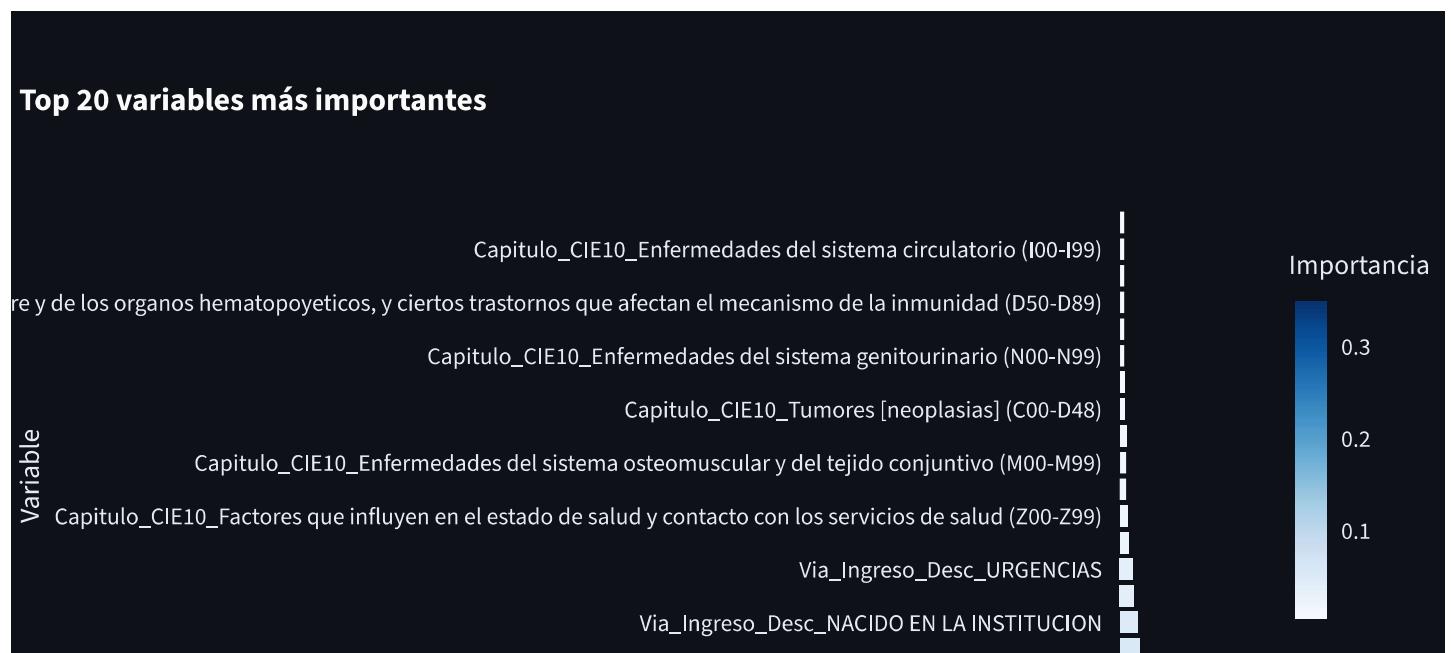
200

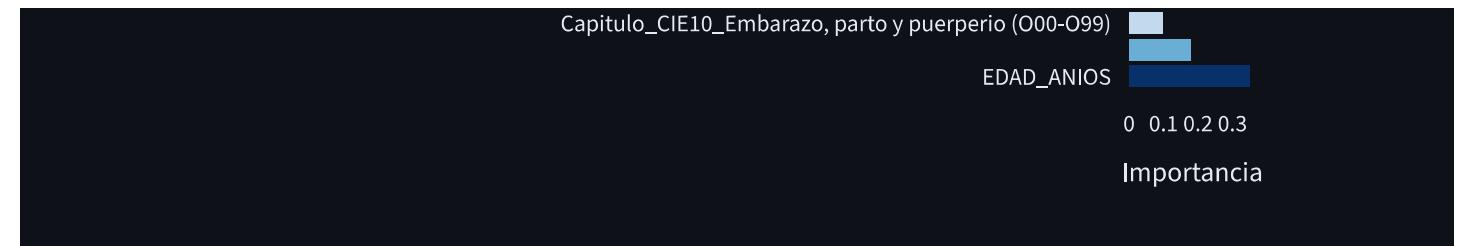
Profundidad máxima del
árbol
10

	Métrica	Valor
0	R ²	0.1340
1	RMSE	7.2590
2	MAE	4.1300

Interpretación:

- **R²**: proporción de variabilidad explicada por el modelo (1 es perfecto).
- **RMSE**: error cuadrático medio, penaliza errores grandes.
- **MAE**: error absoluto medio, mide desviación promedio.





Predicción interactiva

Predicción con Random Forest

Ingrera los valores o selecciones los parámetros clínicos para estimar la duración de la estancia hospitalaria.

Edad (años)

40

—

Sexo

F
 M

Vía de ingreso

NACIDO EN... ▾

Estado de salida

MUERTO ▾

Causa externa

ENFERMEDAD PROFESIONAL

Capítulo CIE10

Embarazo, parto y puerperio (O00-O...

Calcular duración estimada

Duración estimada de estancia:

3.27 días

Este modelo usa **Random Forest**, un método de ensamble no lineal que captura interacciones complejas entre variables.

No requiere supuestos de normalidad ni homocedasticidad, y puede mejorar la precisión predictiva respecto a modelos lineales.

main app



Analisis Exploratorio



Analisis RandomForest



Regresion Logistica Dia...

**EpiScope Envigado**Analítica Predictiva en
Salud Pública**Proyecto desarrollado
por:****Equipo EpiScope Envigado**

Análisis Coocurrencias Diagnósticas

Análisis de Coocurrencias Significativas entre Diagnósticos

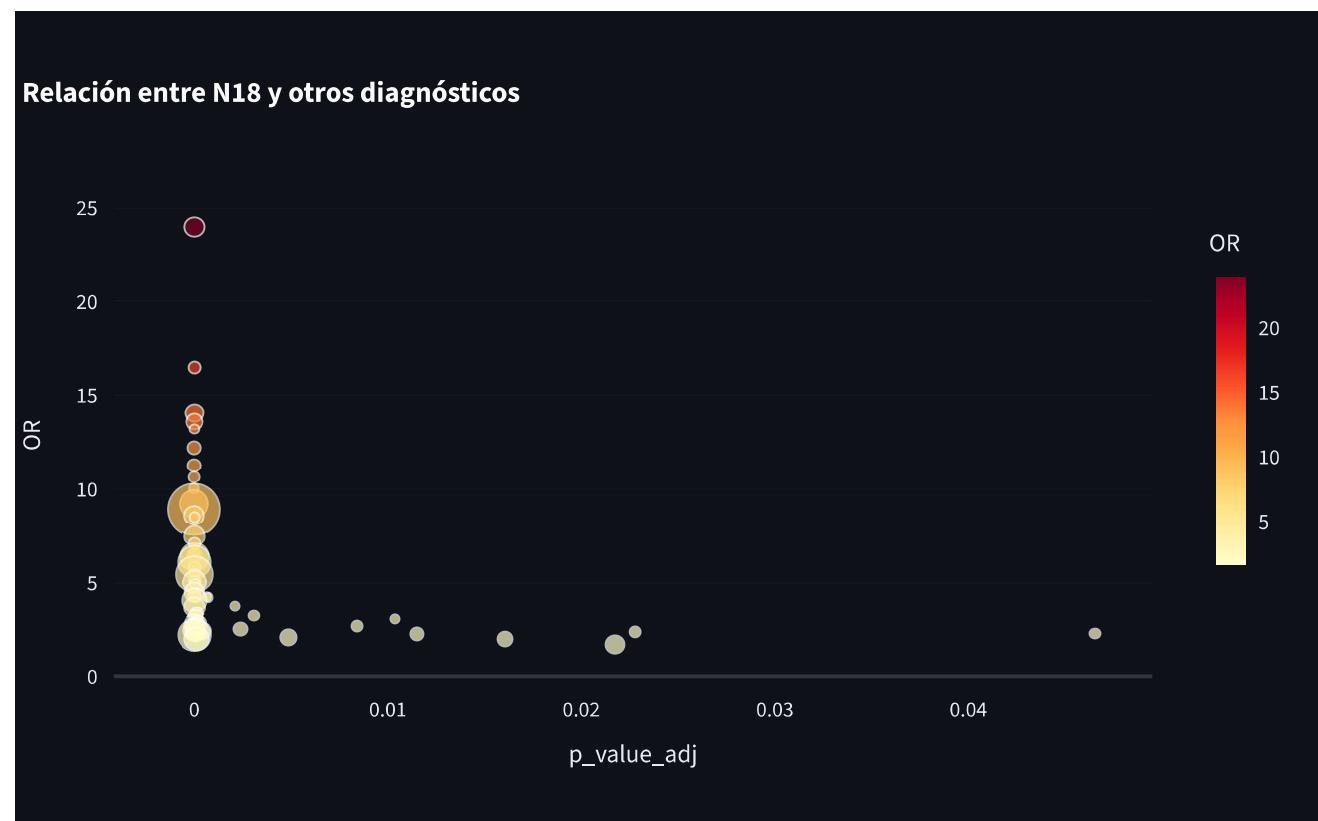
Selecciona diagnóstico:

N18 — Enfermedad Renal Cronica



53 asociaciones con N18 — Enfermedad Renal Cronica

	Dx1	Desc1	Dx2	Desc2
1379	J81	Edema Pulmonar	N18	Enfermedad Renal Cronica
1364	E16	Otros Trastornos De La Secrecion Interna D	N18	Enfermedad Renal Cronica
1381	K65	Peritonitis	N18	Enfermedad Renal Cronica
1387	A49	Infeccion Bacteriana De Sitio No Especifica	N18	Enfermedad Renal Cronica
1348	I13	Enfermedad Cardiorrenal Hipertensiva	N18	Enfermedad Renal Cronica
1366	I77	Otros Trastornos Arteriales O Arteriolares	N18	Enfermedad Renal Cronica
1969	N18	Enfermedad Renal Cronica	Y84	Otros Procedimientos Medicos Como La Causa De Re
1365	D63	Anemia En Enfermedades Cronicas Clasifica	N18	Enfermedad Renal Cronica
1353	I51	Complicaciones Y Descripciones Mal Definid	N18	Enfermedad Renal Cronica
1391	E10	Diabetes Mellitus Insulinodependiente	N18	Enfermedad Renal Cronica

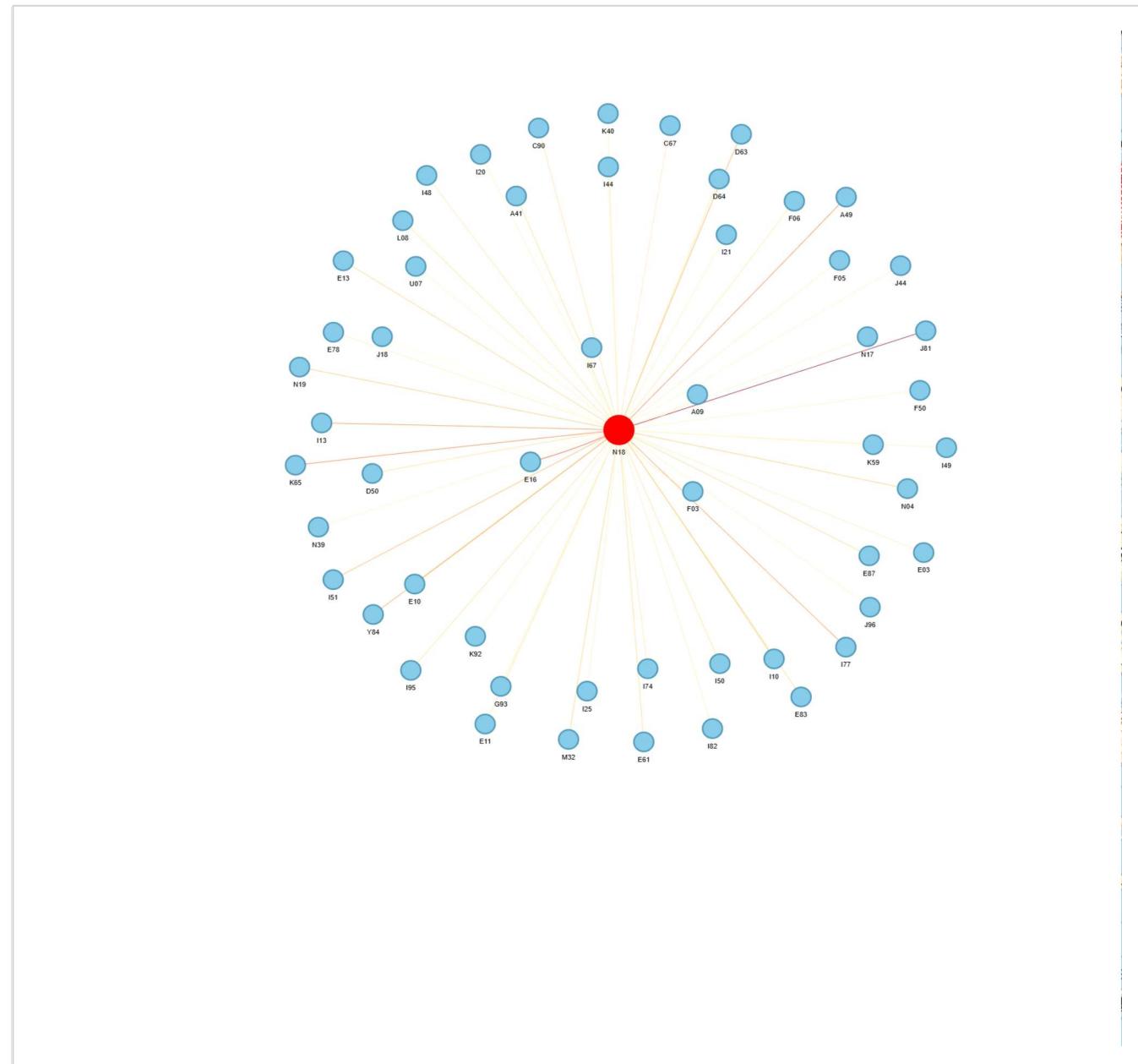


🌐 Visualización de red

Umbral mínimo de coocurrencias para incluir en la red

5

Generar red interactiva



main app



Analysis Exploratorio



Analysis RandomForest



Regresion Logistica Dia...



EpiScope Enviado

Analítica Predictiva en
Salud PúblicaProyecto desarrollado
por:

🧠 Regresión Logística: Probabilidad de Dx2 dado Dx1

Este módulo permite explorar la **probabilidad de que un paciente con un diagnóstico (Dx1) también presente otro diagnóstico (Dx2)****, considerando la edad y el sexo mediante **regresión logística**.

✓ Dataset cargado con 35807 pacientes

🔍 Selección de diagnósticos

Seleccione Dx1

E10 - Diabetes Mellitus Insulinodependiente

Dx1 seleccionado: E10 - Diabetes Mellitus Insulinodependiente

Pacientes con E10: 447

Seleccione uno o más Dx2

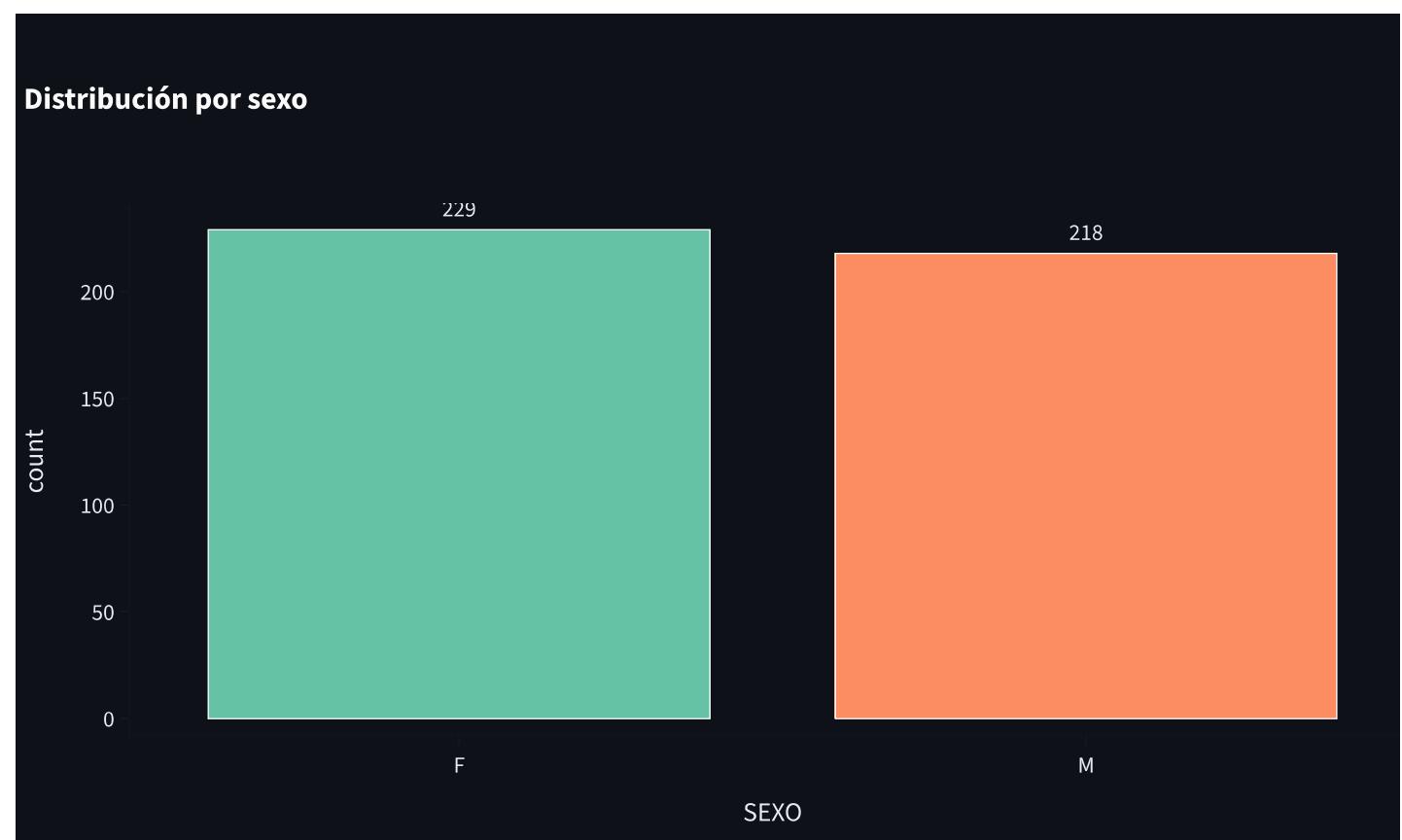
A41 - Otras sepsi... × I50 - Insuficienci... × N18 - Enfermeda... ×

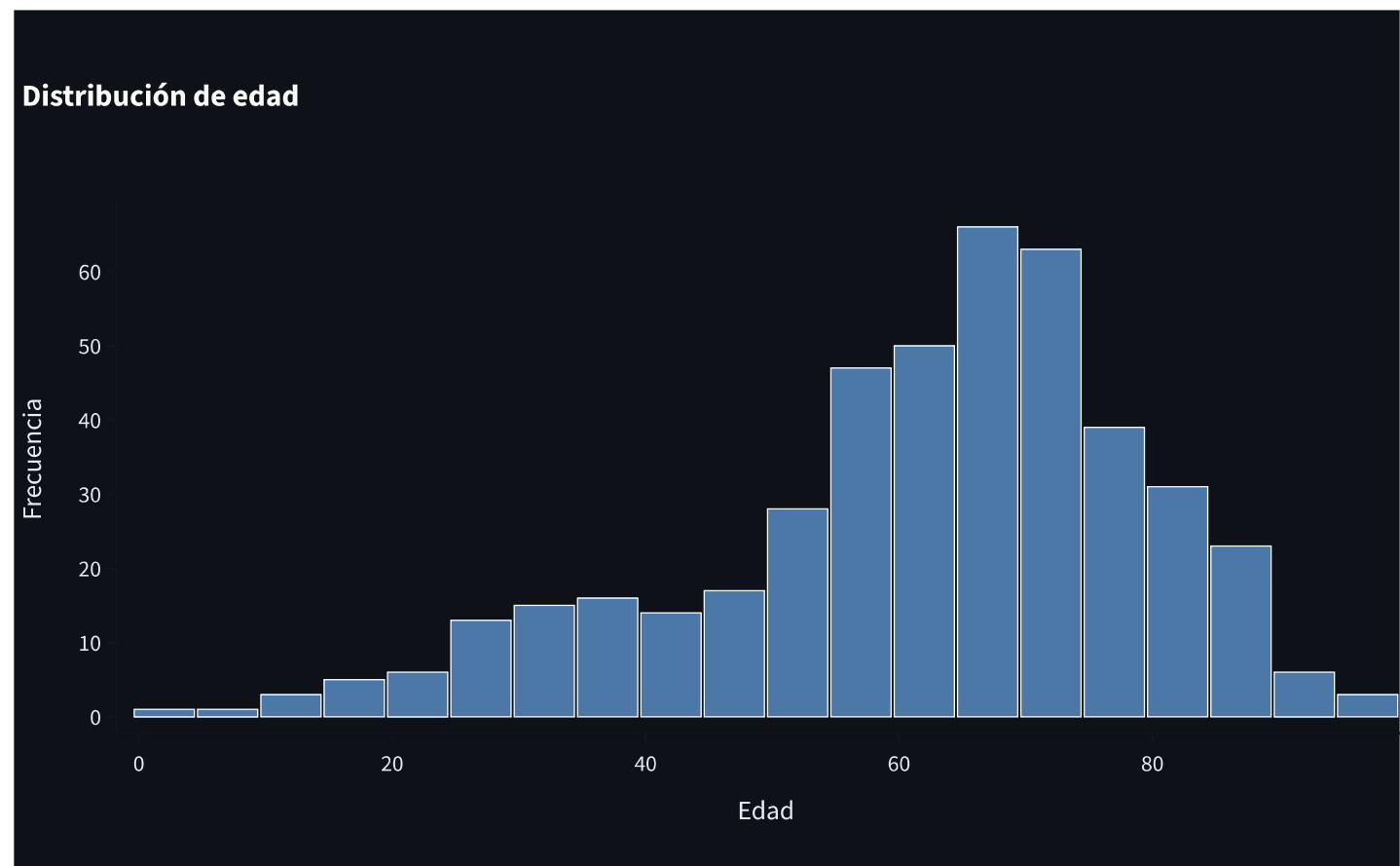


Dx2 seleccionados:

- A41 - Otras sepsis
 - I50 - Insuficiencia Cardiaca
 - N18 - Enfermedad Renal Cronica
-

Distribución de edad y sexo para pacientes con E10





Regresión logística: Probabilidad de Dx2 dado Dx1

La regresión logística estima cómo cambia la **probabilidad de presentar Dx2** según la **edad** y el **sexo**, entre pacientes que ya tienen Dx1.

📌 Modelo para A41 - Otras sepsis

▼ Ver resumen estadístico completo

Logit Regression Results

=

Dep. Variable: Dx2_presente No. Observations: 447

Model: Logit Df Residuals: 444

Method: MLE Df Model: 2

Date: Thu, 13 Nov 2025 Pseudo R-squ.: 0.001160

Time: 14:56:37 Log-Likelihood: -158.50

converged: True LL-Null: -158.68

Covariance Type: nonrobust LLR p-value: 0.8319

=

coef	std err	z	P> z	[0.025	0.975]
------	---------	---	------	--------	--------

const	-1.8447	0.563	-3.274	0.001	-2.949	-0.740
-------	---------	-------	--------	-------	--------	--------

EDAD_ANIOS	-0.0020	0.008	-0.239	0.811	-0.018	0.014
------------	---------	-------	--------	-------	--------	-------

M	-0.1720	0.300	-0.574	0.566	-0.759	0.415
---	---------	-------	--------	-------	--------	-------

=

✳️ Interpretación general

- **OR > 1:** la variable aumenta la probabilidad de Dx2.

- **OR < 1:** la variable disminuye la probabilidad.
- **p < 0.05:** asociación estadísticamente significativa.

	Variable	OR	IC95_Lower	IC95_Upper	p-value
0	const	0.158000	0.052000	0.477000	0.001000
1	EDAD_ANIOS	0.998000	0.982000	1.015000	0.811000
2	M	0.842000	0.468000	1.515000	0.566000

Predicción interactiva para A41

Edad del paciente

Sexo del paciente

F
▼

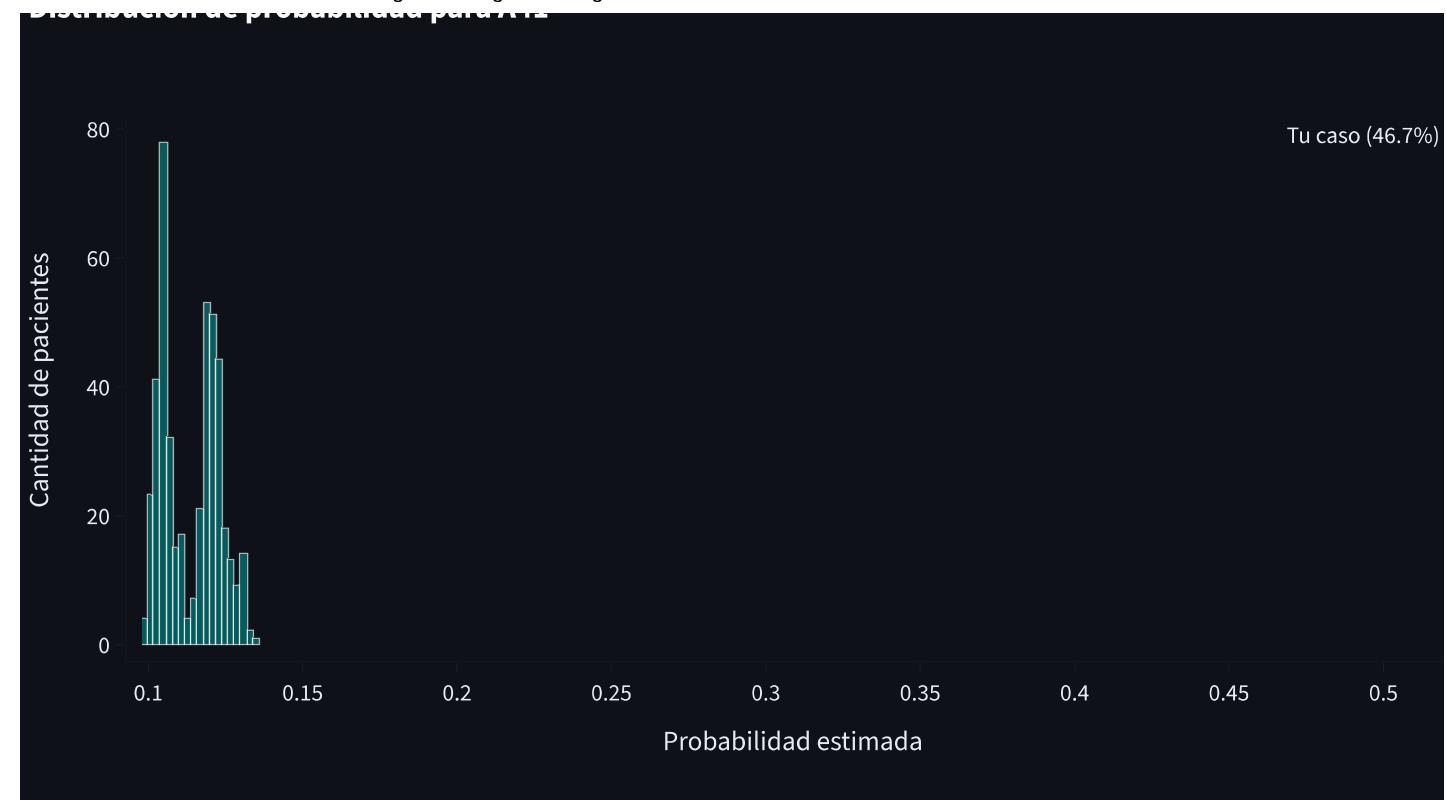
👉 Probabilidad estimada de tener **A41** dado **E10**

Edad: 65 años · Sexo: F

- ◆ Probabilidad: **0.467 (46.7%)**
- ◆ Intervalo de confianza 95%: **23.2% – 71.9%**

💡 Interpretación: Según el modelo, la probabilidad de presentar Otras sepsis dado que el paciente tiene Diabetes Mellitus Insulinodependiente es **alta**.

Distribución de probabilidad para A41



Modelo para I50 - Insuficiencia Cardiaca

▼ Ver resumen estadístico completo

Logit Regression Results

=====

=

Dep. Variable: Dx2_presente No. Observations: 447

Model: Logit Df Residuals: 444

```

Method: MLE Df Model: 2
Date: Thu, 13 Nov 2025 Pseudo R-squ.: 0.02698
Time: 14:56:37 Log-Likelihood: -135.53
converged: True LL-Null: -139.29
Covariance Type: nonrobust LLR p-value: 0.02333
=====
=
      coef  std err      z   P>|z|  [0.025    0.975]
-----
const    -3.7942   0.781  -4.860   0.000   -5.324   -2.264
EDAD_ANIOS  0.0256   0.011   2.350   0.019    0.004    0.047
M       -0.2941   0.334  -0.881   0.378   -0.948    0.360
=====
=

```

Interpretación general

- **OR > 1:** la variable aumenta la probabilidad de Dx2.
- **OR < 1:** la variable disminuye la probabilidad.
- **p < 0.05:** asociación estadísticamente significativa.

	Variable	OR	IC95_Lower	IC95_Upper	p-value
0	const	0.023000	0.005000	0.104000	0.000000
1	EDAD_ANIOS	1.026000	1.004000	1.048000	0.019000
2	M	0.745000	0.388000	1.433000	0.378000



Predicción interactiva para I50

Edad del paciente

65

Sexo del paciente

F

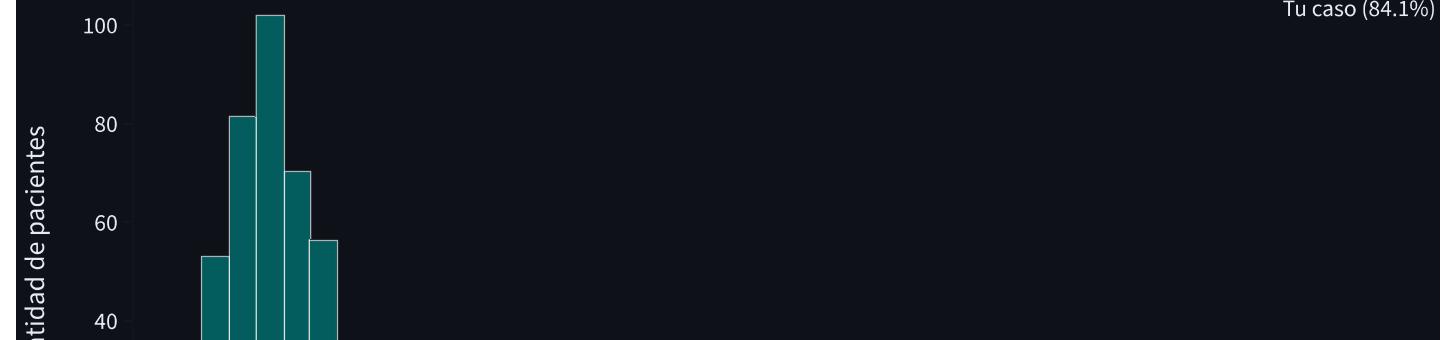
👉 Probabilidad estimada de tener **I50** dado **E10**

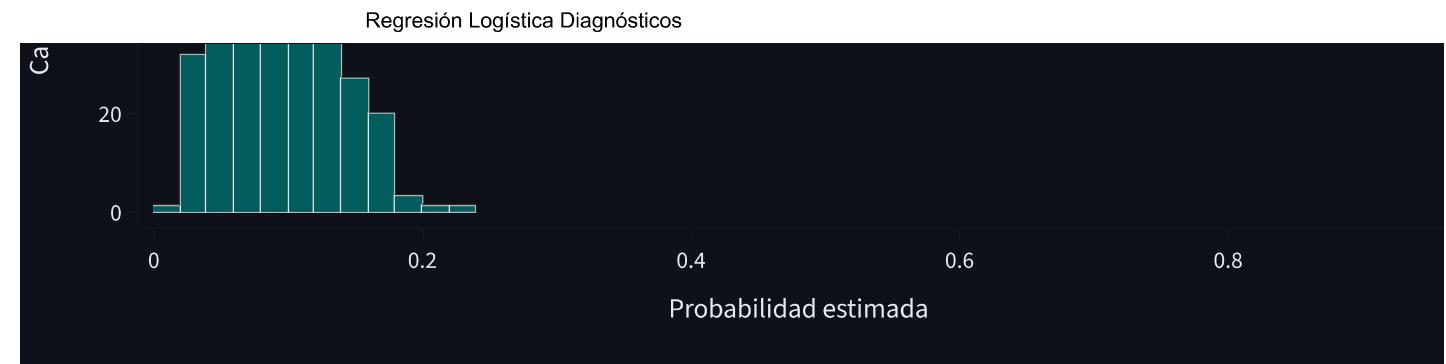
Edad: 65 años · Sexo: F

- ◆ Probabilidad: 0.841 (84.1%)
- ◆ Intervalo de confianza 95%: 56.9% – 95.5%

💡 Interpretación: Según el modelo, la probabilidad de presentar Insuficiencia Cardiaca dado que el paciente tiene Diabetes Mellitus Insulinodependiente es **muy alta**.

Distribución de probabilidad para I50





📌 Modelo para N18 - Enfermedad Renal Cronica

▼ Ver resumen estadístico completo

Logit Regression Results

```
=====
=
Dep. Variable: Dx2_presente No. Observations: 447
Model: Logit Df Residuals: 444
Method: MLE Df Model: 2
Date: Thu, 13 Nov 2025 Pseudo R-squ.: 0.01830
Time: 14:56:37 Log-Likelihood: -151.71
converged: True LL-Null: -154.54
Covariance Type: nonrobust LLR p-value: 0.05916
=====
=
      coef  std err      z   P>|z|  [0.025  0.975]
```

Regresión Logística Diagnósticos

```

const      -3.0417   0.653   -4.655   0.000   -4.322   -1.761
EDAD_ANIOS  0.0090   0.009    0.976   0.329   -0.009    0.027
M          0.6923   0.315   2.199   0.028   0.075   1.309
=====
=

```

Interpretación general

- **OR > 1:** la variable aumenta la probabilidad de Dx2.
- **OR < 1:** la variable disminuye la probabilidad.
- **p < 0.05:** asociación estadísticamente significativa.

	Variable	OR	IC95_Lower	IC95_Upper	p-value
0	const	0.048000	0.013000	0.172000	0.000000
1	EDAD_ANIOS	1.009000	0.991000	1.027000	0.329000
2	M	1.998000	1.078000	3.703000	0.028000

Predicción interactiva para N18

Edad del paciente

65

Sexo del paciente

F

