

# Proyecto Ciencia de Datos Aplicada

Servicio de Trasplantes Hepáticos FSFB 2025-2

*Integrantes: Johana Rátiva, Juan Manuel Rivera, Lina Bejarano y Juan Diego Enriquez.*

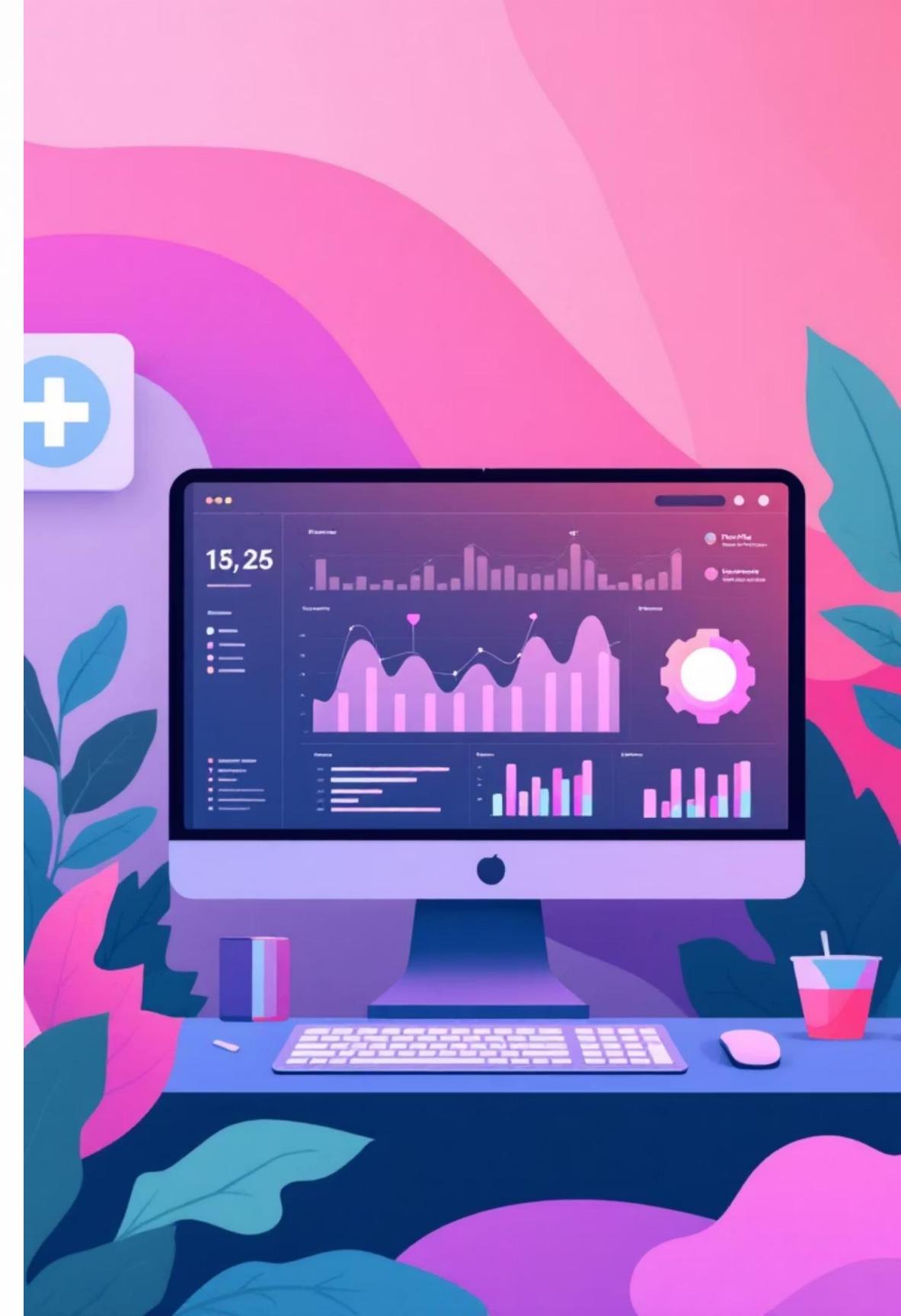
# Problemática y Contexto

## Situación Inicial

*La Fundación Santa Fe de Bogotá consolidó información dispersa de pacientes sometidos a trasplante hepático en una base de datos integral con variables operativas, clínicas, quirúrgicas y epidemiológicas.*

## Objetivos del Proyecto

- *Desarrollar panel interactivo en Power BI*
- *Facilitar reportes del año rural*
- *Apoyar decisiones clínicas*
- *Impulsar investigación mediante análisis de supervivencia*



# Métricas de Éxito (KPIs)



Crecimiento Anual

*Aumento sostenido del volumen anual de trasplantes mediante optimización de priorización*



Reducción Fast Track

*Disminuir tiempo entre cirugía y hospitalización/alta mediante detección temprana*



Mortalidad Postoperatoria

*Reducir mortalidad anticipando patrones de riesgo y apoyando decisiones oportunas*



Reportes Clínicos

*Automatizar consultas y agilizar lectura de casos críticos*



Productividad Científica

*Facilitar generación de estudios y publicaciones de alto valor*

# Solución Propuesta

Dashboard Power BI con 4 Vistas Principales

01

Panorama General

*Reportes operativos y descripción poblacional con filtros por grupo etario, sexo, año y variables demográficas*



02

Pre-trasplante

*Tiempo en lista de espera, etiologías, Child-Pugh, MELD score y antecedentes para evaluar riesgos*

03

Intra-quirúrgico

*Cirujanos, tiempo de isquemia, antibióticos, profilaxis antitrombótica y días de hospitalización*

04

Post-trasplante

*Rechazo agudo/crónico, gráficas Kaplan-Meier, falla cardíaca, arritmias e infarto POP*

# Panorama General de Trasplantes



Year

Todas

Month

Todas

Sexo Receptor

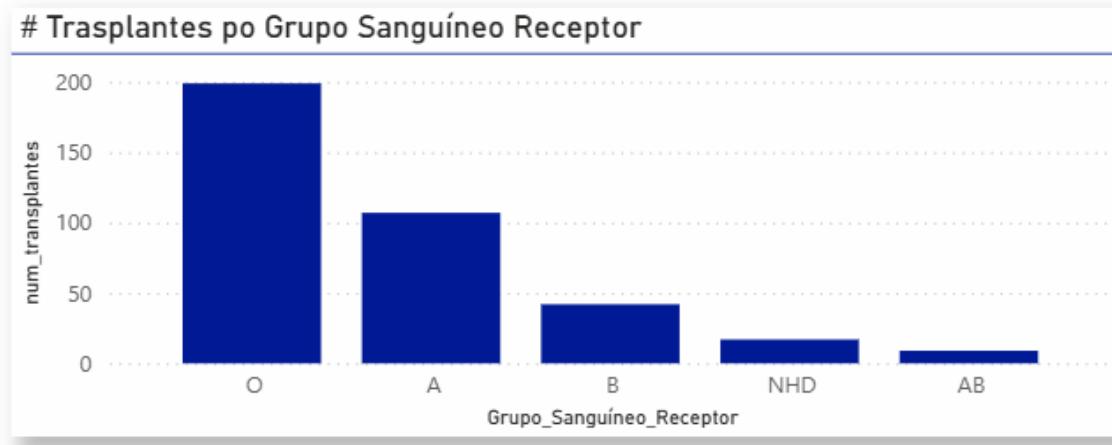
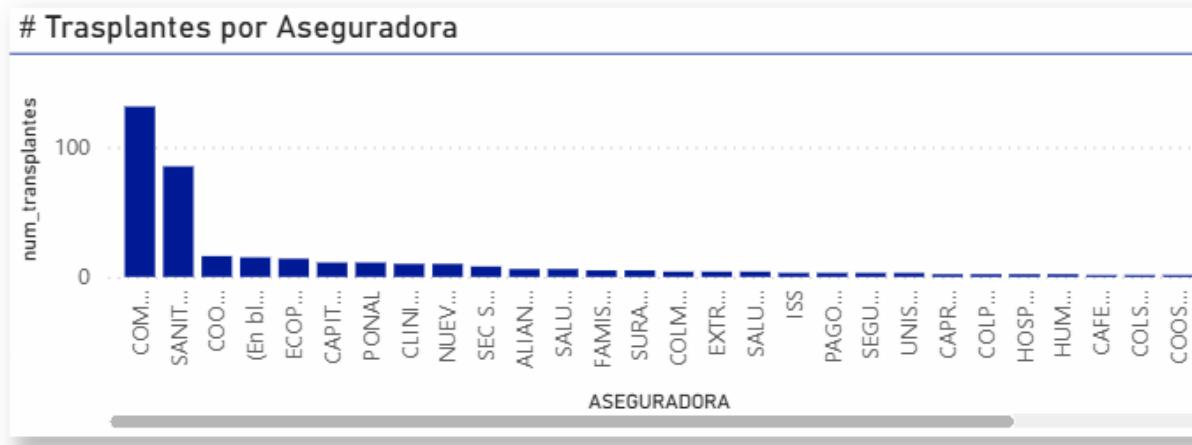
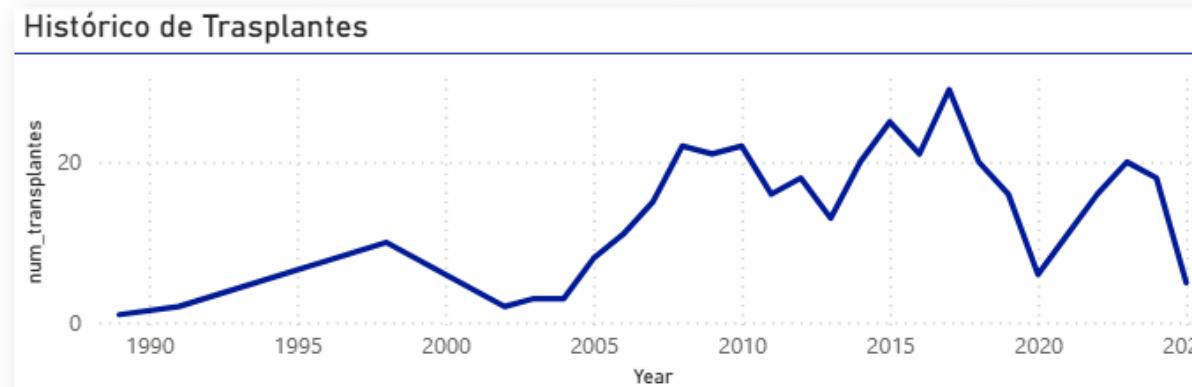
1

Edad

0 85

Aseguradora

Todas



# Información sobre Pre trasplante

Year	Month	Sexo Receptor	Edad	Aseguradora
Todas	Todas	Todas	0	85

Panorama de información

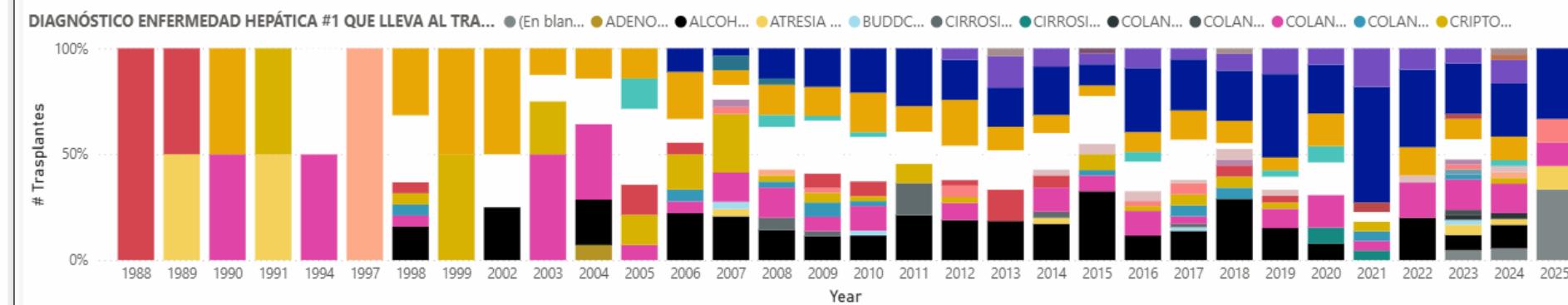
<b>Pre trasplante</b>	Post trasplante	Intra Quirúrgico	Donante
-----------------------	-----------------	------------------	---------

- 1) Seleccione un panorama de información del cuál desea ver información  
 2) Seleccione una categoría de información.

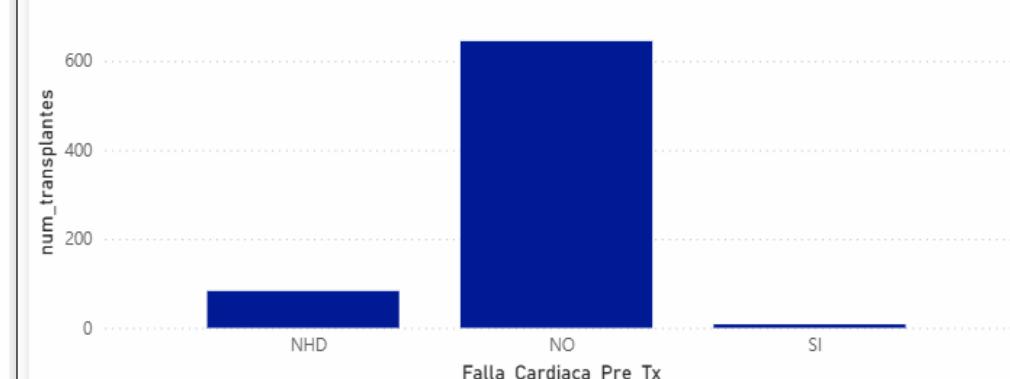
Selección una categoría de la cual deseas ver información

DIAGNÓSTICO ENFERMEDAD HEPÁTICA #1 QUE LLEVA AL TRASPLANTE	MELD_Score	Hipertensión_Arterial_Pre_Tx	LDL_Pre_Tx	Colesterol_TotalPre_Tx
Tiempo_En_Lista_Meses	Falla_Cardiaca_Pre_Tx	Antecedente_De_Tabaquismo	HDL_Pre_Tx	
Child_Pugh_Letra	Diabetes_Mellitus_Pre_Tx	Antecedente_De_Alcoholismo	Triglicéridos_Pre_Tx	

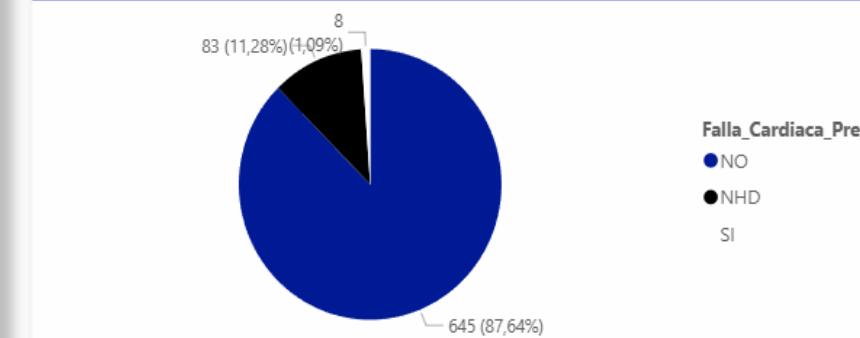
Por favor seleccione una opción de panorama de información y una de sus categorías



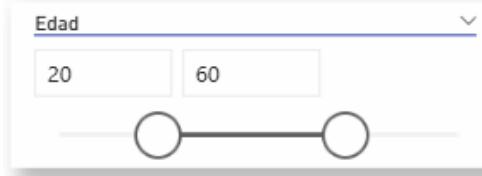
Por favor seleccione una opción de panorama de información y una de sus categorías



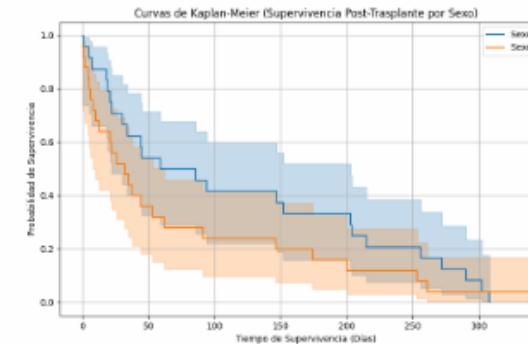
Por favor seleccione una opción de panorama de información y una de sus categorías



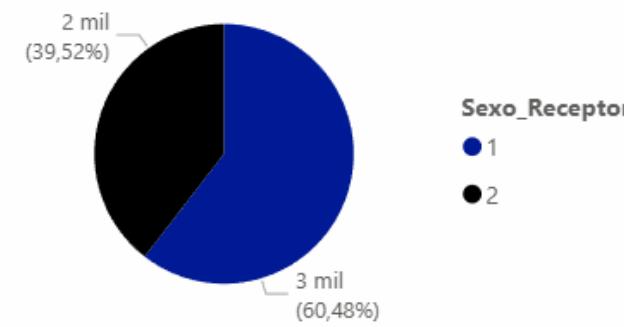
# Análisis de Supervivencia Poblacional



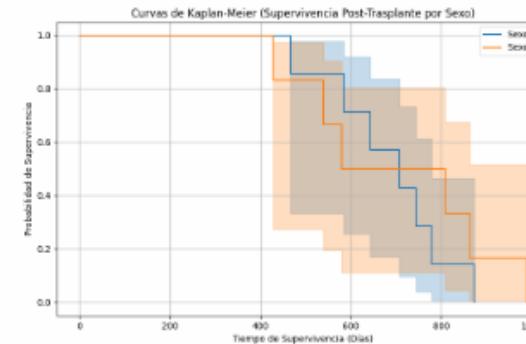
Supervivencia a 1 año



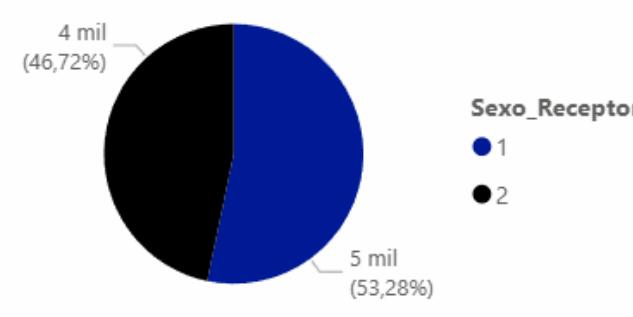
% Muerte al año por Genero



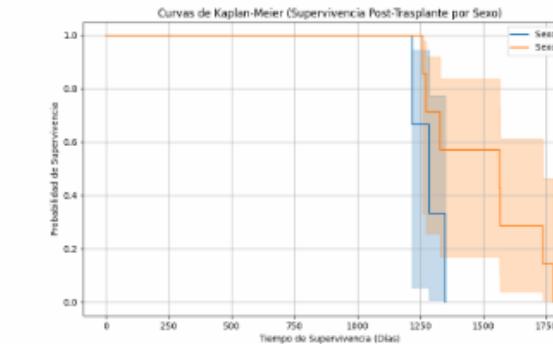
Supervivencia a 3 años



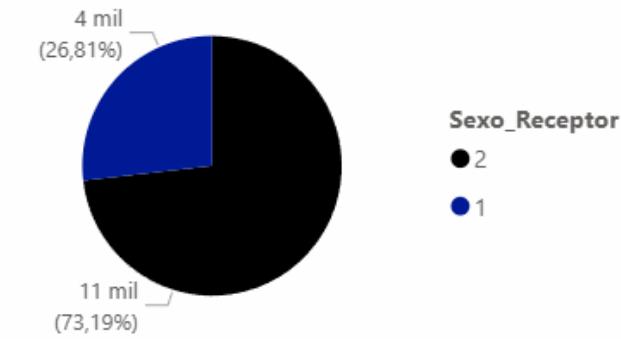
% Muerte a los dos años por Genero



Supervivencia a 5 años



% Muerte a los tres años por Genero



# Enfoque Analítico

Pregunta Central

*¿Qué factores clínicos y demográficos se asocian con mayor riesgo de complicación postoperatoria?*

Metodología

*Enfoque mixto combinando técnicas estadísticas, visualización y machine learning supervisado con 4 algoritmos:*

- *Regresión Logística*
- *Random Forest*
- *Árboles de Decisión*
- *AdaBoost*

*Métrica principal: F1-score ponderado para equilibrar precisión y recall en contexto con desbalance de clases.*



# Procesamiento de Datos



## Datos Iniciales

*736 registros con 285 variables clínicas recibidas el 25 de septiembre de 2025*



## Selección

*114 variables iniciales basadas en análisis de interés, completitud y variabilidad*



## Limpieza

*Estandarización de NaN, corrección de tipos, depuración de rangos clínicos (MELD ≤40, Child-Pugh)*



## Validación

*121 variables finales validadas con stakeholder para modelado*

# Hallazgos y Resultados



## Perfil Demográfico

Mayoría de pacientes entre 60-74 años. Distribución equitativa por sexo (masculino 50,8%, femenino 49,2%).  
EPS Sanitas predominante.



## Temporalidad

Pico de trasplantes en 2017. Reducción notable en 2020 por COVID-19. Mediana de espera: 0 años, aunque algunos casos esperan 4+ años.



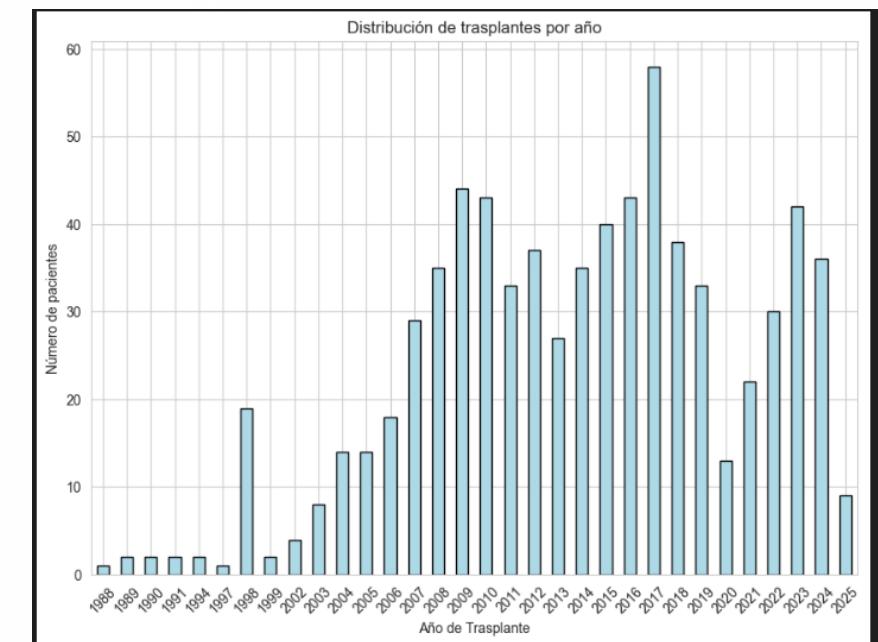
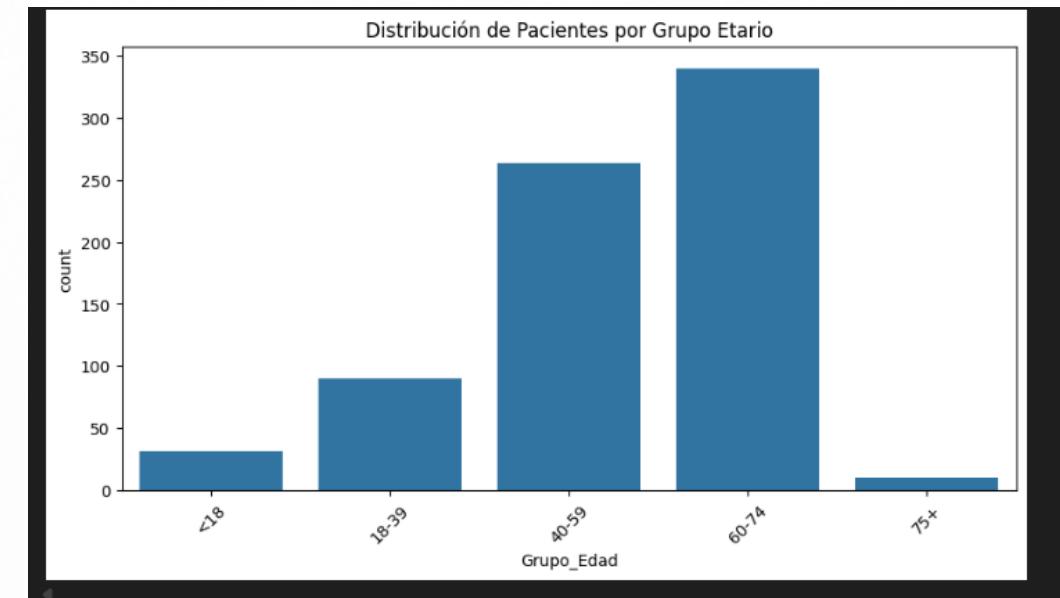
## Etiología Principal

NASH (esteatohepatitis no alcohólica) representa el 21,1%. Mayoría sin hepatocarcinoma previo al trasplante.



## Factores de Riesgo

Relación significativa entre complicaciones y: Soporte Vasopresor PeriTx, Noradrenalina PeriTx y número de complicaciones simultáneas.



# Preparación para Modelado



## Imputación Variable Objetivo

*Creación de "¿Complicación?" a partir de tipo de complicaciones, infección, arritmia, falla cardíaca e infarto*

*Proporción inicial: Con complicación: 53% , Sin complicación: 47%*

*Proporción después de imputación: Con complicación 57% , Sin complicación 43%*



## Selección Inicial

*70 características numéricas, booleanas y ordinales sin procesos adicionales de limpieza*



## Procesamiento Categórico

*OneHotEncoder y MultiLabelBinarizer aplicados, resultando en 173 columnas*

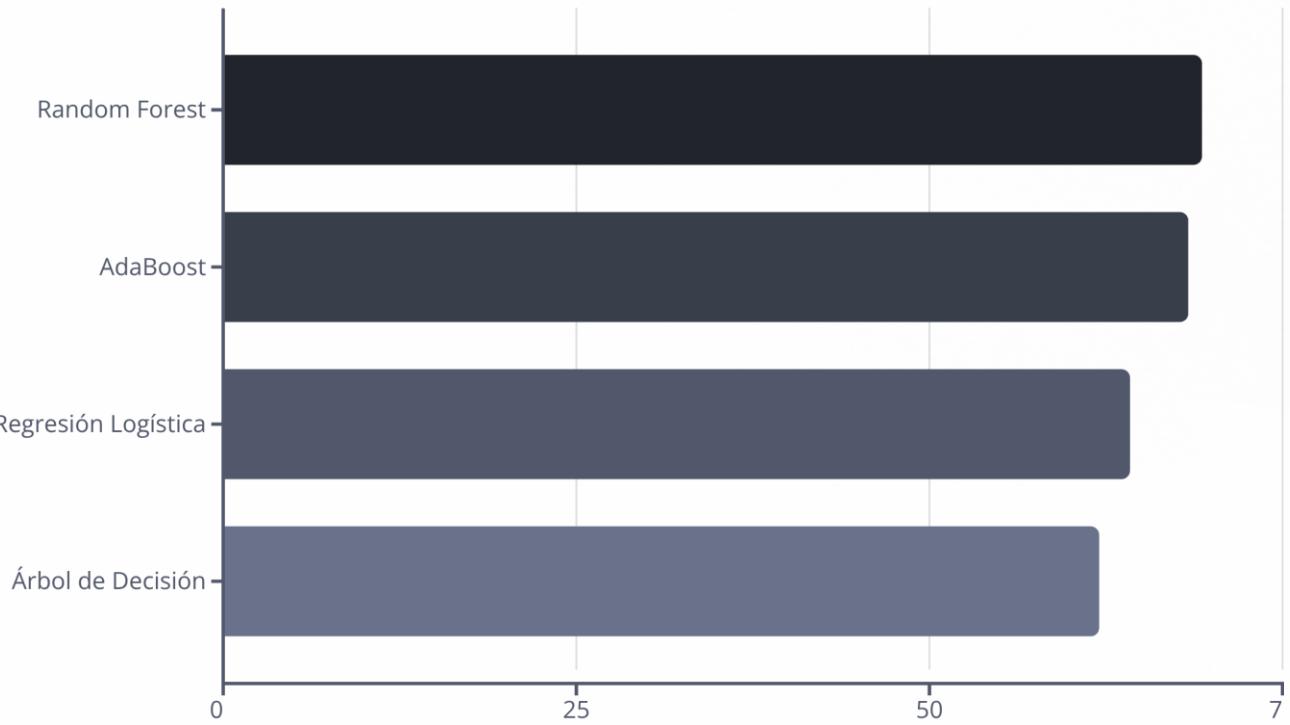


## Filtrado Final

*Umbral mínimo 16% no nulidad, eliminación de filas nulas: 466 filas × 143 columnas*

- ☐ División de datos: 70% entrenamiento (326 registros) y 30% validación (140 registros) con estratificación

## Resultados del Modelado

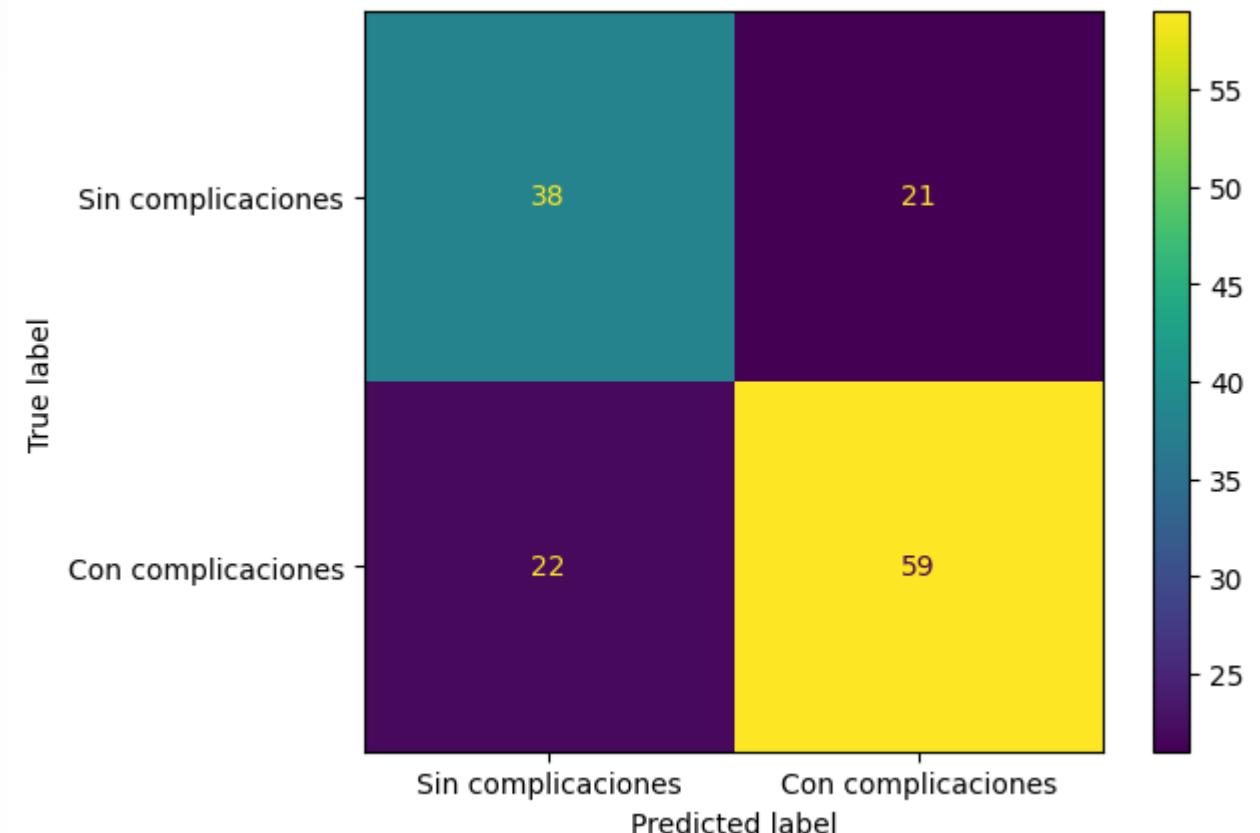


Mejor Modelo: Random Forest

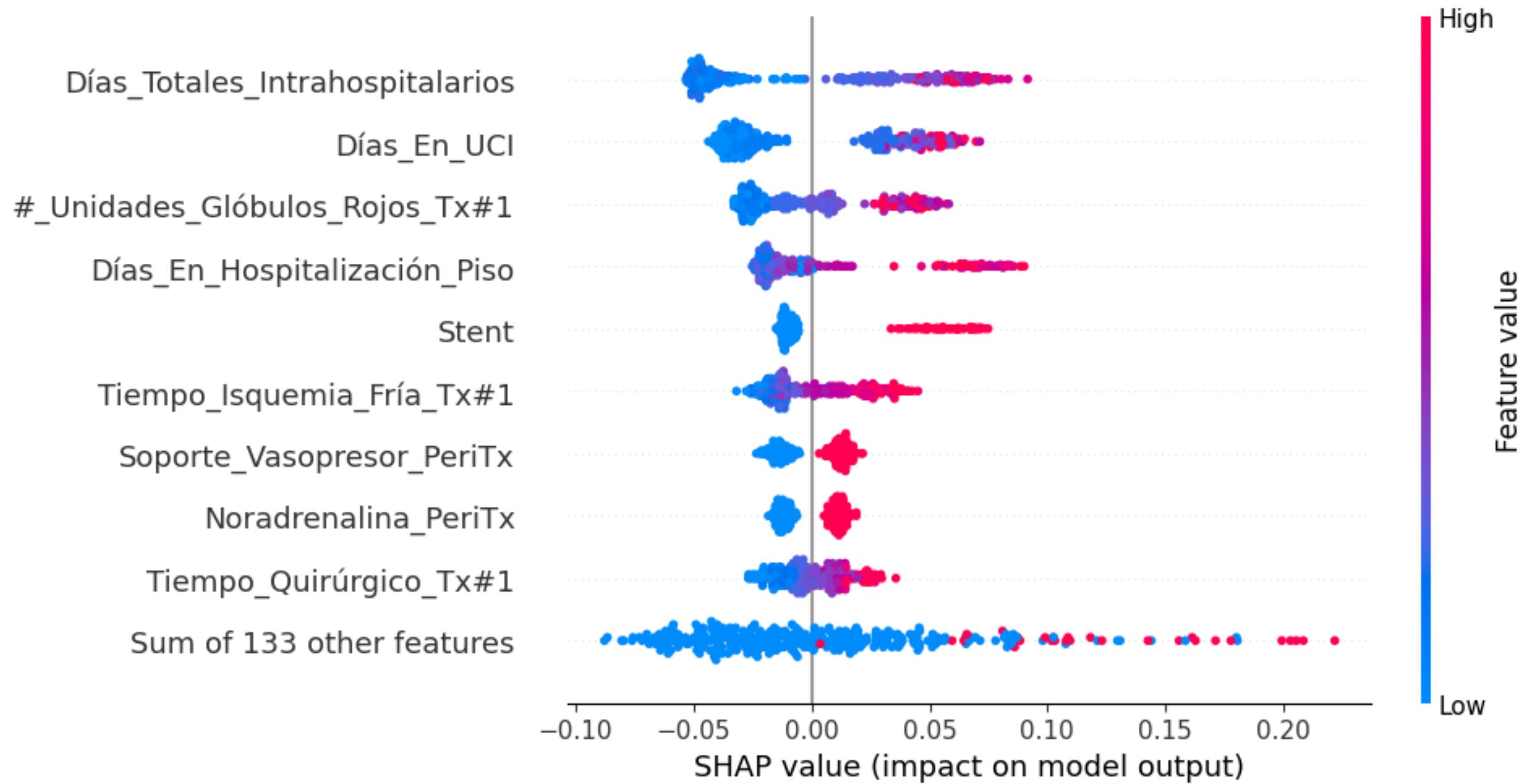
**F1-Score ponderado de 69,3% con equilibrio óptimo entre sensibilidad y precisión.**

Desempeño Destacado

- *Recall: 73% (59 de 81 casos positivos)*
- *Precisión: 74% (3 de 4 predicciones correctas)*
- *Accuracy: 69,3%*



¿Cómo esta funcionando el modelo?



## Impacto Esperado

### Logros del Proyecto

*Power BI implementado con 5 secciones, diccionario rediseñado y modelo Random Forest con 69,3% F1-score para predicción de complicaciones.*

### Impacto Esperado

- *Reducir tasa de mortalidad*
- *Reducir tiempo de hospitalización*
- *Potenciar artículos científicos*

### Desafíos Identificados

*Calidad de datos, valores faltantes e inconsistencias afectaron desarrollo. Se requiere pipeline mejorado y software especializado.*

### Próximos Pasos

*Validación con datos 2025, implementación progresiva tipo Canary Deployment y mejora continua del sistema de captura de datos.*

