

ANÁLISIS COMPAS



Azahara Bravo

Daniel Álvarez

Maria Loza

08.10.2025

ÍNDICE

01

CONTEXTO Y OBJETIVO DEL PROYECTO

02

DATOS

03

EDA

04

RIESGO VS REALIDAD

05

ANÁLISIS PRELIMINAR
DE SESGOS

06

MODELO
ALTERNATIVO

07

DASHBOARD

08

CONCLUSIONES

09

RECOMENDACIONES

CONTEXTO

1

COMPAS es un algoritmo usado en EE.UU. para predecir el riesgo de reincidencia en acusados del sistema judicial.

QUÉ ES?

Un sistema que asigna puntuaciones de riesgo (1 a 10) y se emplea en decisiones judiciales como fianzas o libertad condicional

POLÉMICA

Investigaciones (ProPublica, 2016) mostraron **sesgos raciales**: más falsos positivos en afroamericanos y más falsos negativos en acusados blancos, generando un debate global sobre ética y justicia algorítmica.

IMPORTANCIA

Ejemplo clave de cómo los algoritmos pueden **reproducir desigualdades** en lugar de corregirlas, y de la necesidad de **modelos más justos y transparentes**.

DATOS

2

Origen de los datos

Fuente: **ProPublica (2016)**, Condado de Broward (Florida, EE.UU.).

11.757 acusados en dataset completo.

7.214 acusados en dataset reducido a **2 años de seguimiento**



Variables

- **Demográficas:** edad, rango de edad, género, etnia.
- **Judiciales:** grado de los delitos, fechas de arresto y custodia, antecedentes.
- **COMPAS:** puntuación de riesgo (1–10), categorías (Bajo, Medio, Alto)
- **Resultado real:** reincidencia a 2 años (sí/no)

Perfil de la muestra

- 81% hombres, 19% mujeres
- 51% African-American, 34% Caucasian, 9% Hispanic, minorías muy pequeñas (Asian 0,4%, Native American 0,2%)
- Mayoría de **25–45 años (57%)**

Limitaciones

- Cobertura solo 2013–2014 en Broward → baja generalización
- Datos de reincidencia = **nuevo arresto**, no necesariamente delito
- Alta proporción de nulos en variables de reincidencia violenta
- Posible sesgo estructural: refleja patrones de arresto, no tasas reales de criminalidad

DATOS

LIMITACIONES DEL DATASET DE PROPUBLICA

Otros autores y nuestro filtrado adicional

Dataset	Total casos	% No reincide	% Reincide
ProPublica (2016, two-year)	7.214	54,9%	45,1%
Barstein (2019, "corrected")	6.216	63,8%	36,2%
Nuestro filtrado (± 30, ≤ 730, $\leq 1/4/14$)	≈ 5.259	63,2%	36,8%



Dataset ProPublica:

Datos incoherentes: inflan artificialmente reincidencia.
El corte temporal es mayor para reincidentes.

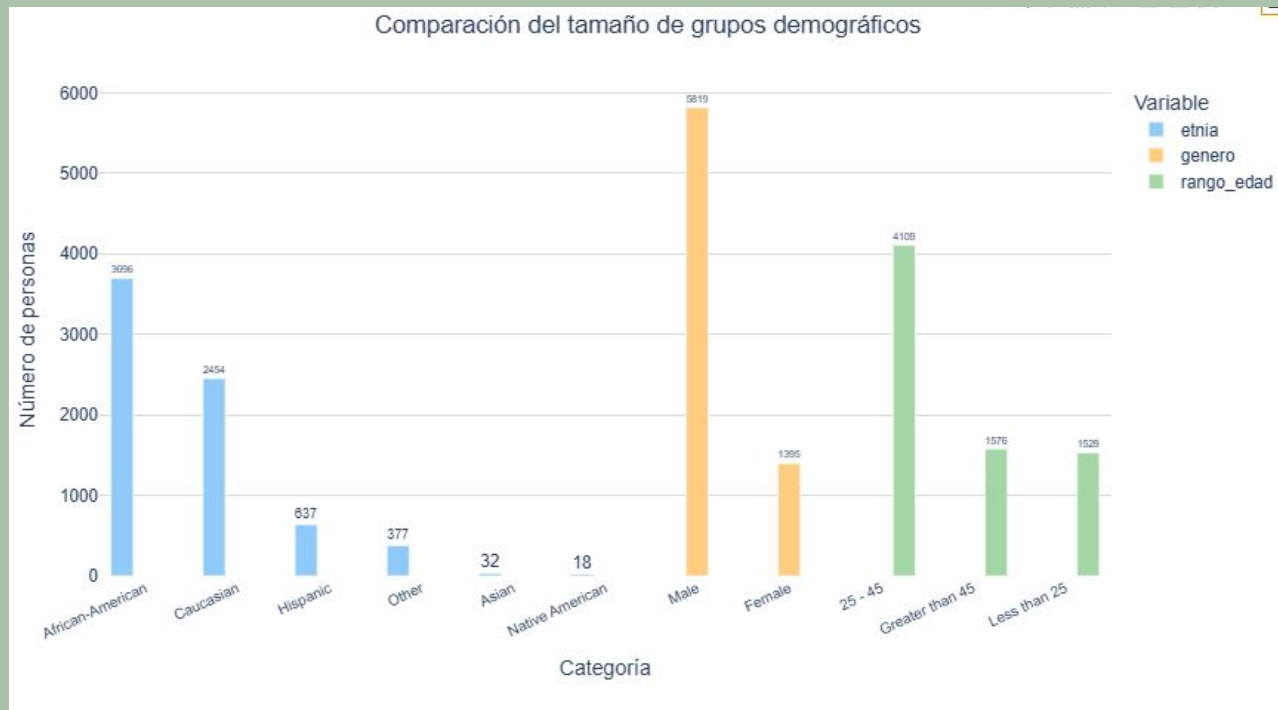
DECISIÓN SOBRE LOS DATOS

Seguimos utilizando el dataset de ProPublica por su valor comparativo y por ser el estándar internacional más usado en estudios sobre COMPAS.

EDA

3

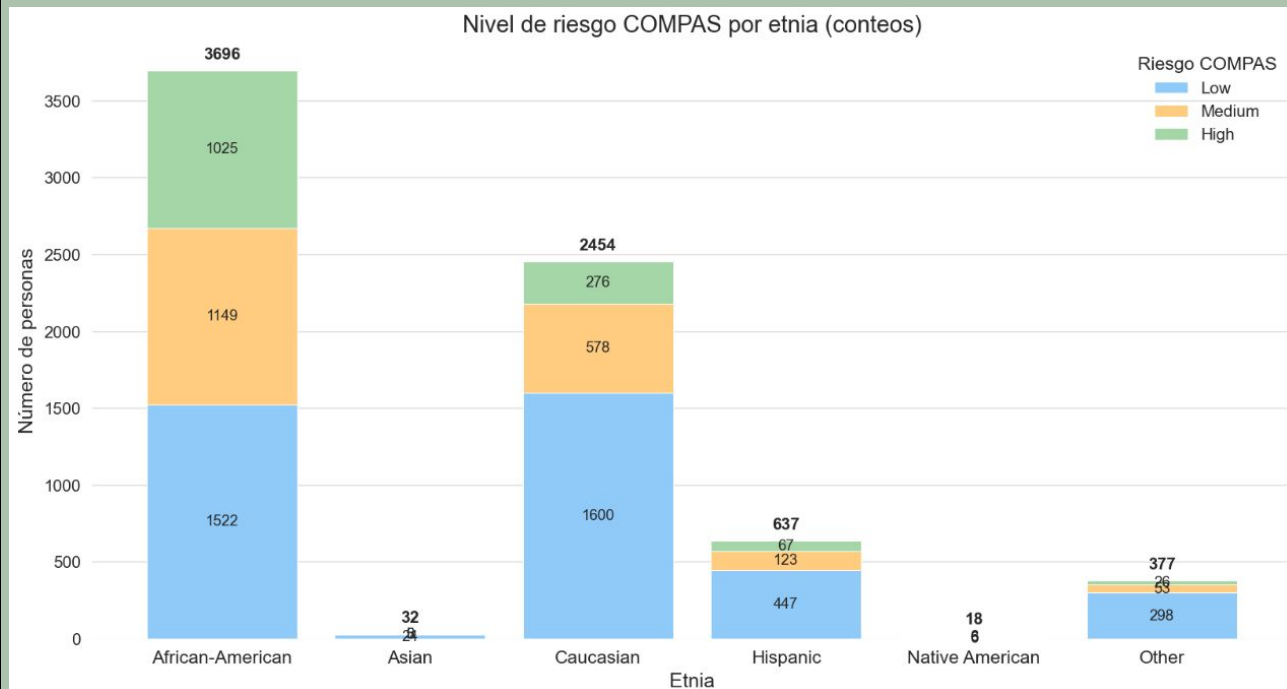
Distribuciones demográficas



El dataset no es equilibrado, hay fuerte subrepresentación de mujeres y minorías → esto condiciona cualquier modelo.

- Género (81% hombres vs 19% mujeres).
- Etnia (African-American 51% / Caucasian 34% / resto minoritario).
- Edad (mayoría 25–45 años).

Resultados COMPAS por etnia



“COMPAS puntúa sistemáticamente más alto a afroamericanos lo que apunta a un posible sesgo potencial.”

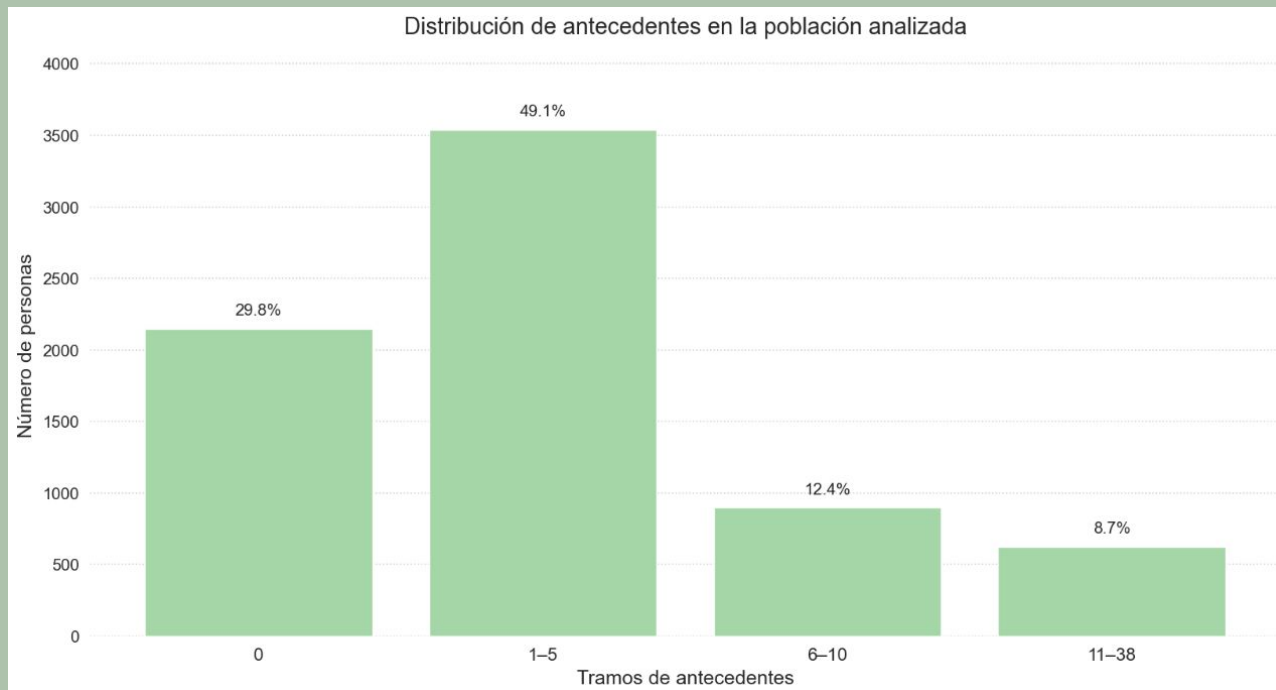
Puntuaciones medias:

African-American: 5.37
Caucasian: 3.74

Hombres: 4.59 vs
Mujeres: 4.17

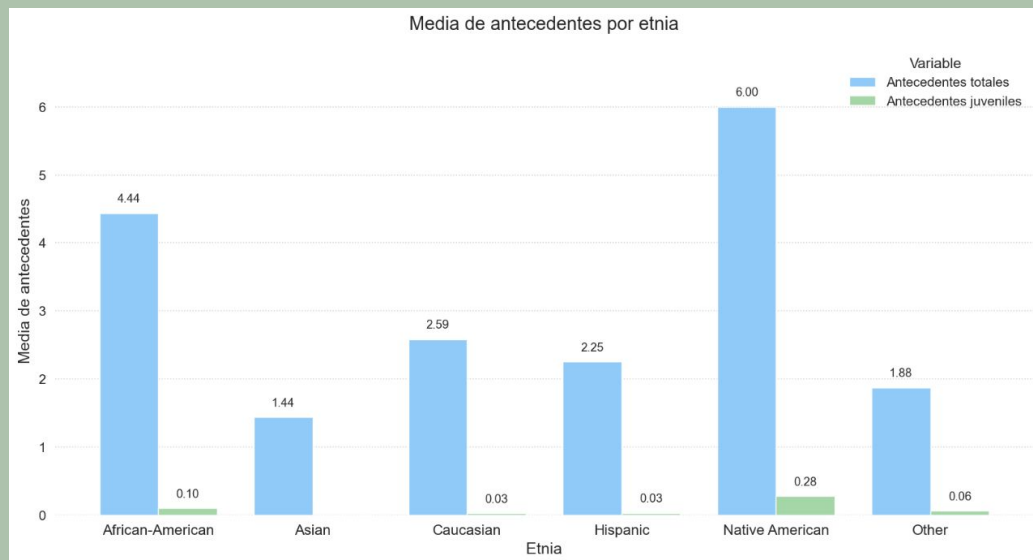
Jóvenes <25 años: 5.91
vs >45 años: 2.92

Comportamiento delictivo de la población estudiada



“La población analizada combina un gran grupo con poca experiencia delictiva y un pequeño núcleo de reincidencia persistente, aspecto que puede influir en la calibración del algoritmo COMPAS.”

Antecedentes por etnia y género



Género	Media de antecedentes totales	Media de antecedentes juveniles
Hombre	3.76	0.08
Mujer	2.28	0.01

Conclusiones iniciales del EDA

Representación

El EDA revela **desequilibrios demográficos marcados** en el dataset, con una fuerte sobrerrepresentación de hombres y personas African-American.

Tendencia

Las **puntuaciones de riesgo y los antecedentes** muestran una tendencia consistente: los grupos con más historial delictivo reciben puntuaciones más altas.

Reproducir desigualdades existentes

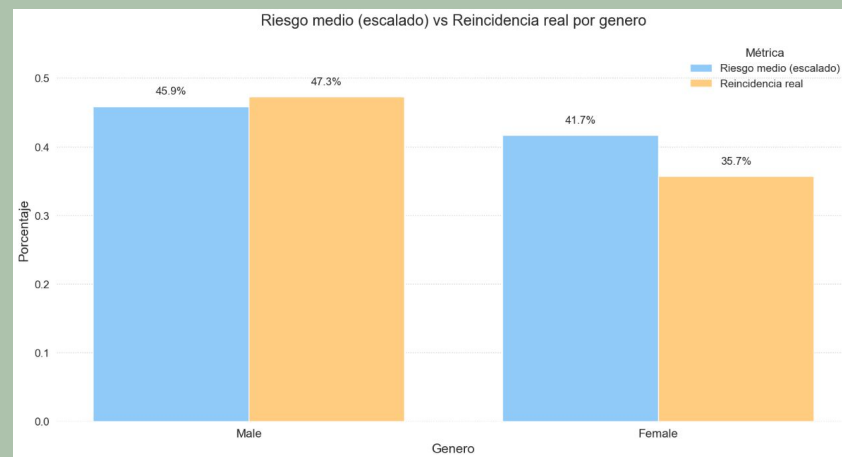
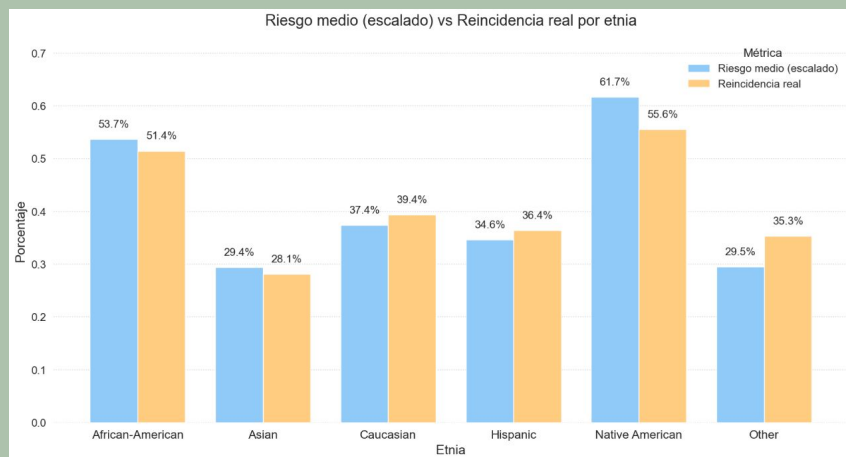
Esto no demuestra sesgo por sí mismo, pero sí sugiere que **las condiciones de partida de los datos pueden predisponer al algoritmo** a reproducir desigualdades existentes.

El EDA establece así la base para analizar en profundidad si COMPAS **exagera o corrige** esas diferencias en las fases siguientes del proyecto.

RIESGO VS REALIDAD

4

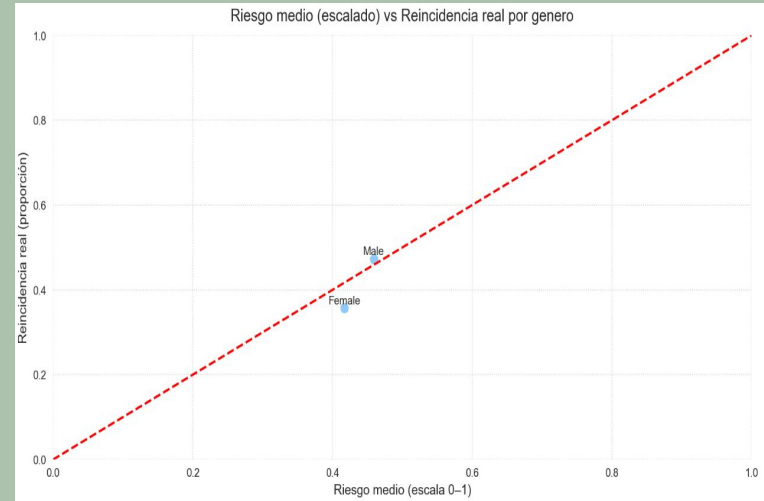
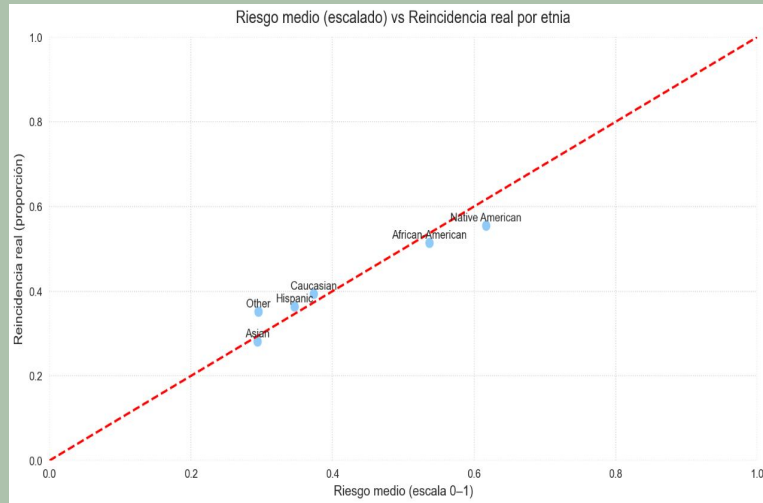
Riesgo Medio vs Reincidencia real



¿Predice COMPAS igual para todos los grupos?

Correlación de Spearman

Los grupos con mayor riesgo predicho realmente reinciden más?



$$\rho = 0,35$$

COMPAS mantiene el orden general –a mayor score, mayor reincidencia–, pero no con exactitud: el ajuste no es perfecto, lo que indica que el sistema ‘ordena’ los grupos correctamente, pero sin calibrarlos de forma precisa.

Conclusiones Riesgo vs Realidad

Spearman correlación monótona perfecta

En general son valores cercanos, pero con diferencias sistemáticas en algunos grupos:



Sobreestimación de riesgo

African-American
Mujeres

Subestimación de riesgo

Caucasian

Posibles implicaciones éticas

Cuando un modelo como COMPAS tiende a sobreestimar sistemáticamente el riesgo en ciertos colectivos, en este caso, en personas afroamericanas lo que hace en la práctica es reforzar estereotipos preexistentes.

ANÁLISIS PRELIMINAR DE SESGOS


5

ANÁLISIS PRELIMINAR DE SESGOS

ESTADÍSTICA DESCRIPTIVA


GRUPO	N	MEDIA	MEDIANA	REINCIDENCIA_REAL	INTERPRETACIÓN
AFRICAN AMERICAN	3696	5.37	7	51.4	Sobreestima el riesgo
CAUCASIA N	2454	3.74	5	39.4	Subestima el riesgo
MALE	5819	4.59	4	47.3	Mejor calibración
FEMALE	1395	4.17	4	35.7	Sobreestima el riesgo /Mismo riesgo asignado pese a menor reincidencia

COHORTES

 Cohorte F_0 - tablas por ETNIA, RANGO DE EDAD y GÉNERO

	etnia	n	media_score	sd_score	tasa_reincidencia
0	African-American	526	4.36	2.59	35.9
2	Caucasian	406	3.36	2.39	30.0

	genero	n	media_score	sd_score	tasa_reincidencia
1	Male	903	3.69	2.55	33.8
0	Female	243	3.77	2.43	27.2

 Cohorte F_11-38 - tablas por ETNIA, RANGO DE EDAD y GÉNERO

	etnia	n	media_score	sd_score	tasa_reincidencia
0	African-American	321	7.80	2.04	75.4
1	Caucasian	70	6.69	2.51	74.3
	genero	n	media_score	sd_score	tasa_reincidencia
1	Male	386	7.49	2.25	75.6
0	Female	36	8.22	1.82	66.7

MÉTRICAS

Esenciales para la evaluación del modelo COMPAS y de la Regresión logística.

ACCURACY

Mide el porcentaje total de aciertos.

PRECISION

Del total de predicciones como reincidentes, cuantos lo fueron de verdad.

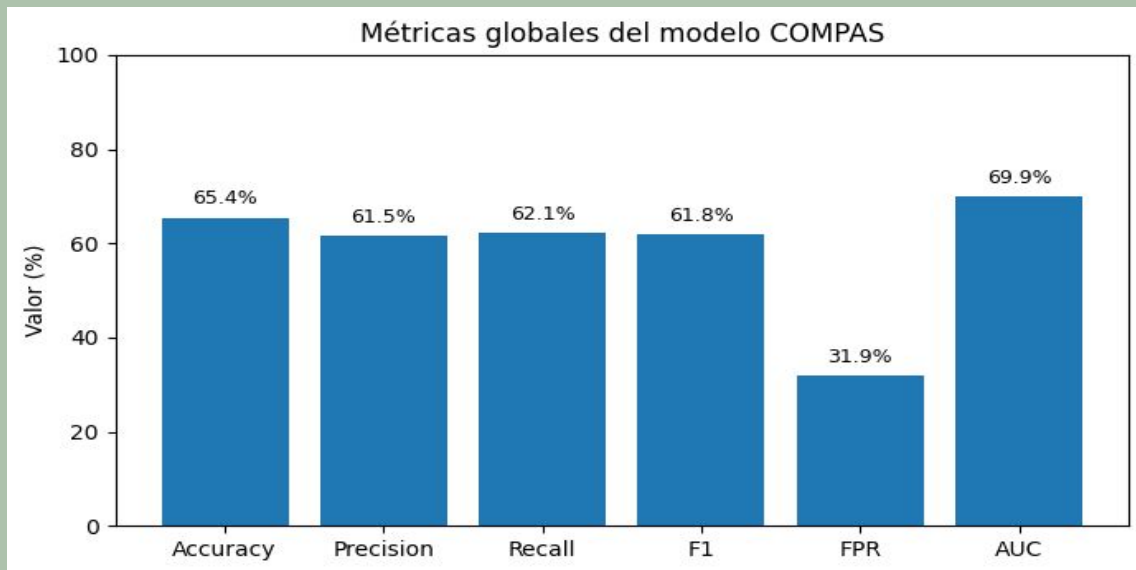
RECALL

De los individuos que realmente reincidentieron, a cuántos el modelo logró detectar?

FPR

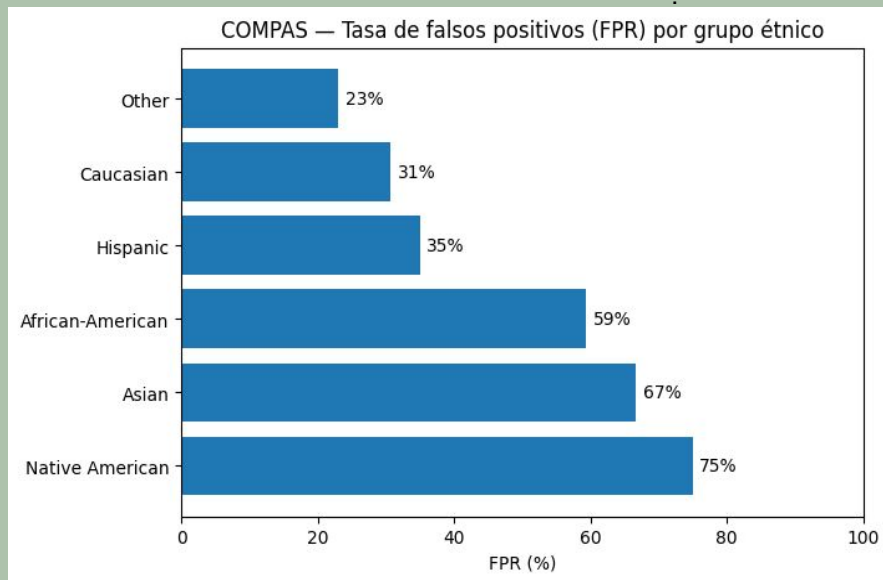
Tasa de falsos positivos. Cuántos de los individuos fueron clasificados como alto riesgo y en realidad no reincidentieron.

Rendimiento global del modelo COMPAS



El desempeño global es **moderado**, con valores equilibrados entre precisión, recall y F1, pero una **tasa de falsos positivos alta (31.9 %)** indica que el modelo **tiende a sobreestimar el riesgo** en parte de la población no reincidente.

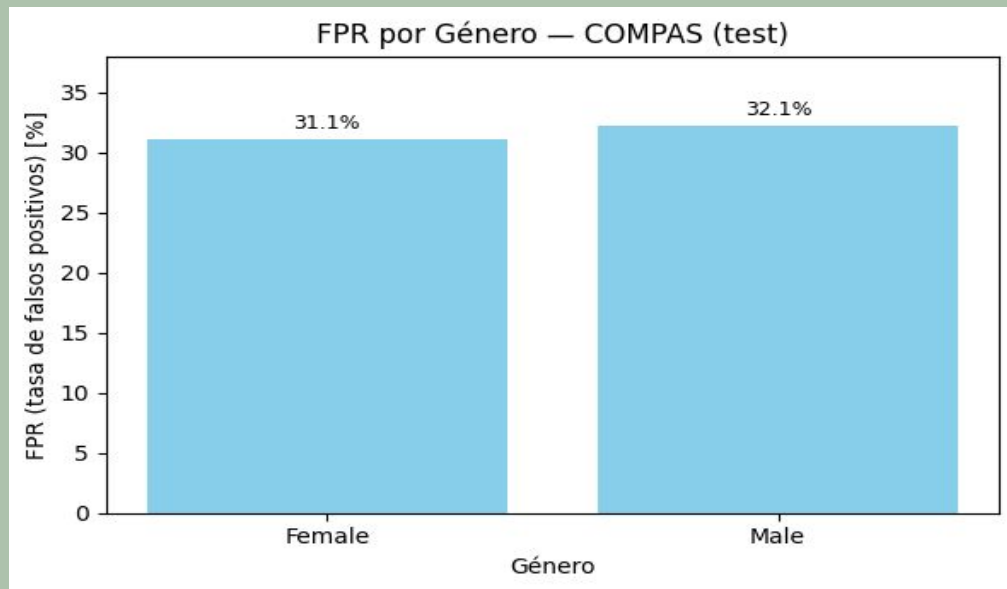
Desglose por etnia de la tasa de falsos positivos (FPR)



La tasa de falsos positivos para afroamericanos (59 %) es casi el doble que la de caucásicos (31 %), lo que refleja un **sesgo evidente en COMPAS** que perjudica a ciertos grupos étnicos.

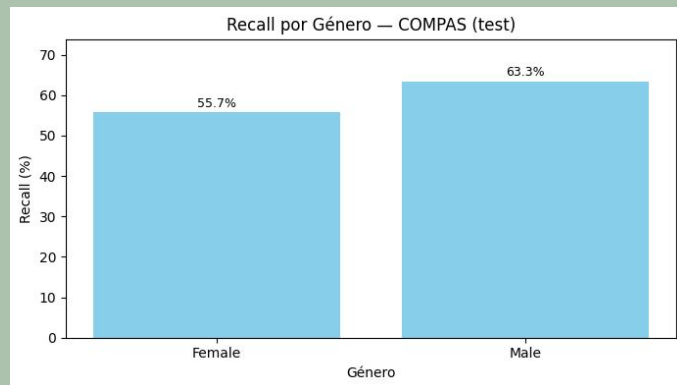
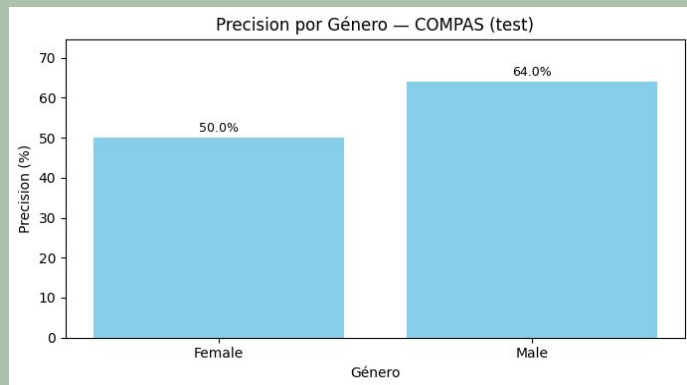
Por otro lado, los valores elevados en **nativos americanos y asiáticos** deben interpretarse con cautela, ya que provienen de **muestras muy pequeñas** dentro del conjunto de datos.

Tasa de Falsos Positivos por Género



- La **tasa de falsos positivos (FPR)** es muy similar entre hombres y mujeres, lo que podría dar la **impresión de equidad** en el modelo COMPAS. Sin embargo, esta igualdad aparente no garantiza una ausencia de sesgo, ya que el modelo podría estar reproduciendo desigualdades en otras métricas o dimensiones más sutiles del desempeño.

Precision y Recall por Género



Pero observando diferentes dimensiones del modelo y analizando diferentes métricas, se puede observar un **recall** más alto y una **precisión** más alta para los hombres, lo que indica que el modelo **identifica mejor a los reincidentes masculinos** y **acierta con mayor frecuencia** cuando predice su reincidencia.

MODELO ALTERNATIVO

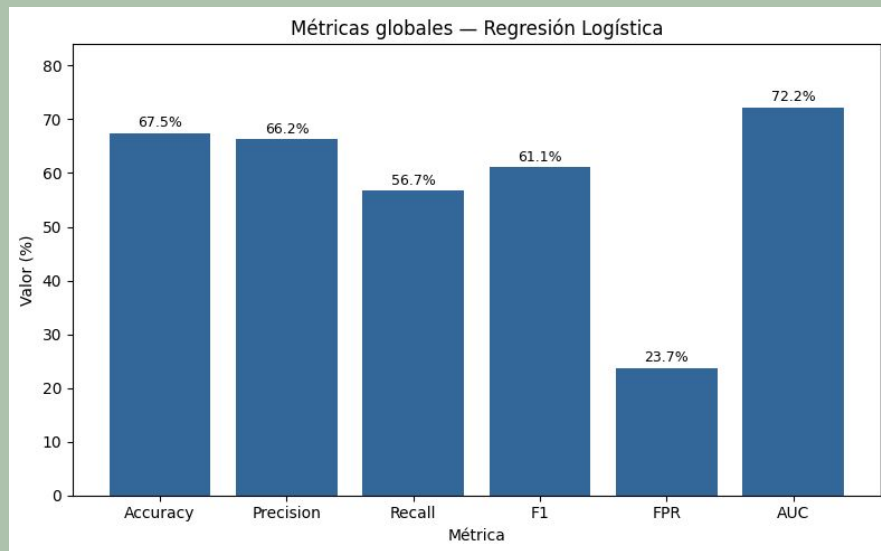
6



“Construcción del modelo de regresión logística para reincidencia”

- El modelo se entrenó con datos históricos de personas condenadas, utilizando únicamente variables relevantes al riesgo de reincidencia (por ejemplo: número de antecedentes, tipo de delito, duración de la condena, tiempo en libertad, etc.).
- **No se utilizaron variables protegidas** como género, etnia, raza, religión, orientación sexual o nacionalidad, para evitar que el modelo incorpore sesgos sistemáticos.
- Resultado: un modelo explicable, defendible, más equitativo y listo para comparar resultados globales y por grupos.

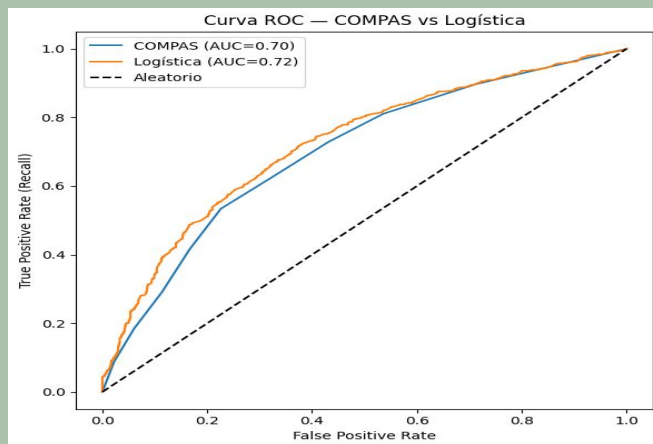
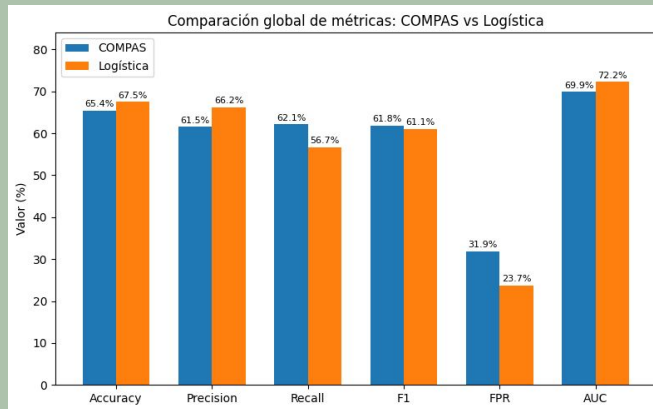
Desempeño de la Regresión Logística



Interpretable y reproducible, el modelo de **regresión logística** ofrece un desempeño global sólido sin recurrir a variables protegidas. Muestra una **precisión ligeramente superior** a COMPAS y una **tasa de falsos positivos más baja**, lo que implica menos clasificaciones erróneas de personas no reincidentes.

Además, al ser un modelo **transparente y auditable**, sirve como **referencia base** para comparar resultados globales y por grupos, incluyendo la curva que relaciona la **tasa de verdaderos positivos con la de falsos positivos**.

Comparación visual de desempeño global y Curva ROC (Receiver Operating Characteristic)



- En el gráfico se observa la comparación de las métricas globales de desempeño entre **COMPAS** y la **regresión logística**, mostrando valores muy similares en la mayoría de las métricas, con diferencias visibles en **Recall**, **FPR** y **AUC**.

- En la curva se observa el rendimiento de **COMPAS** y de la **regresión logística** al distinguir entre personas reincidentes y no reincidentes, mostrando trayectorias muy próximas y **áreas bajo la curva (AUC)** similares, con una **ligera ventaja para la regresión logística**.

COMPAS vs Regresión Logística — Desempeño y Curva ROC

COMPAS vs Regresión Logística — Desempeño y Curva ROC

- Ambos modelos muestran un **desempeño global similar** en Accuracy, Precisión y F1.
- La **regresión logística**, además de ser **transparente**, logra una **tasa de falsos positivos mucho menor**, mientras que **COMPAS obtiene un Recall ligeramente mayor**, detectando más reincidentes reales.
- En conjunto, la logística mantiene resultados comparables a COMPAS, pero con **menos errores injustos** y mayor interpretabilidad.
- En la **curva ROC**, ambas áreas bajo la curva (AUC) son cercanas (**0.70 vs 0.72**), mostrando un rendimiento parecido.
- El AUC más alto de la regresión logística indica que **distingue mejor entre quienes reinciden y quienes no**, reflejando una **ligera ventaja en capacidad predictiva**.

Cierre: Reflexiones sobre el modelo de regresión logística

- El modelo combina **transparencia** y ética con resultados comparables a sistemas opacos como COMPAS. Al excluir variables protegidas, favorece decisiones más justas, evitando discriminación directa. Su enfoque probabilístico permite una mejor valoración del riesgo y un control más fino sobre errores.
- No obstante, puede reproducir **sesgos implícitos** presentes en los datos (como vigilancia desigual o condenas históricas) y depende fuertemente de la calidad del conjunto de datos. Por eso debe usarse como **herramienta de apoyo**, no como sustituto del juicio humano, y debe someterse a auditorías regulares de equidad.
- **Conclusión:** *“Sin importar el número o la complejidad de los modelos, si los datos raíz están incompletos o sesgados, los modelos seguirán replicando desigualdades favoreciendo determinados grupos.”*

DASHBOARD

7



Dashboard

Diseño basado en los colores representativos de la justicia: blanco, azul, gris. Hemos añadido rojo para contextualizarlos en EEUU.

Logos: como origen de los datos trabajados por Propublica añadimos su logo y el escudo norteamericano de su bandera.

Analisis de sesgos en Compas



PROPUBLICA

genero

etnia

rango_edad

Numero de casos
7.214

Puntuación media
4,51

Reincidencia Real
45,07 %

Acierto COMPAS
65,37 %

PRECISION
0,61

RECALL
0,63

Tasa Falsos Negativos
0,37

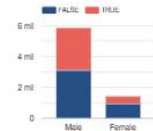
Tasa Falsos Positivos
0,32



Resultado Compas

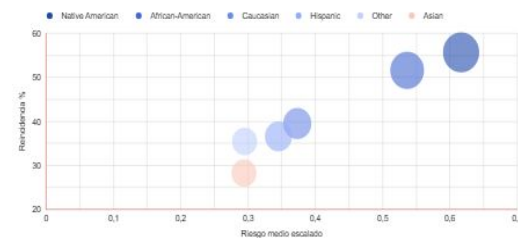


Reincidencia

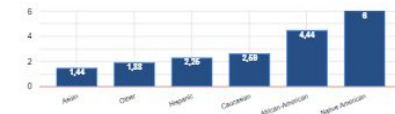


Reincidencia por género

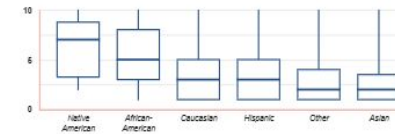
Distribución del Riesgo Medio vs Reincidencia Real



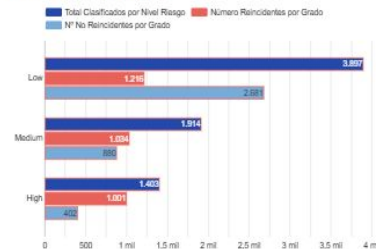
Media de Antecedentes



Distribución de la Puntuación COMPAS por etnia



Puntuación vs Reincidencia



Dashboard

Dashboard de dos páginas divididas para mostrar insights del EDA y del estudio de los sesgos y otra para visualizar los resultados comparativos de los modelos.

Evaluación de modelos

COMPAS vs. Regresión Logística



PROPUBLICA

Modelo

Métrica

Grupo

COMPAS
54,07 %

Logística
54,07 %

Métricas por grupo

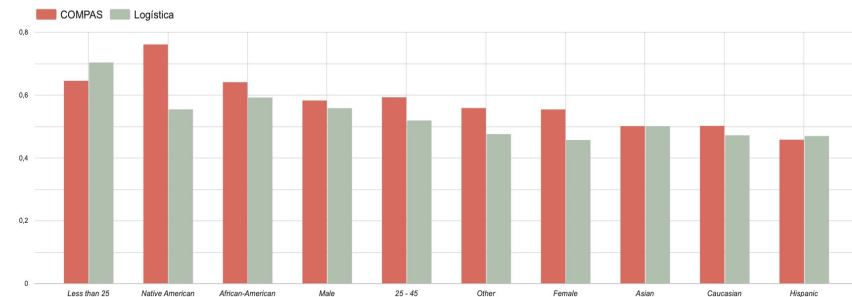


Tabla dinámica COMPAS/ Logística

Grupo	Métrica / Modelo / Valor			
	COMPAS	Precision Logística	COMPAS	Accuracy Logística
Native American	100 %	66,67 %	80 %	80 %
Less than 25	61,26 %	63,82 %	61,74 %	61,74 %
African-American	67,72 %	62,64 %	67,52 %	64,31 %
Male	66,56 %	64 %	66,41 %	65,72 %
25 - 45	72,29 %	63,08 %	67,88 %	66,06 %
Other	96,24 %	73,91 %	75 %	67,31 %

Tabla dinámica COMPAS/Logística (Solo FPR)

Grupo	Métrica / Modelo / Valor	
	COMPAS	FPR Logística
Less than 25	53,61 %	70,48 %
African-American	44,16 %	33,12 %
Male	32,14 %	26,59 %
25 - 45	32,5 %	16,43 %
Female	31,06 %	14,47 %
Hispanic	19,79 %	20,83 %

CONCLUSIONES

8

CONCLUSIONES

Sesgos y desigualdades

El análisis reveló sesgos tanto étnicos como de género.

En particular, COMPAS tiende a asignar puntuaciones más altas a hombres y a minorías raciales, y sobreestima el riesgo en mujeres, lo que sugiere que la herramienta amplifica desigualdades preexistentes en lugar de corregirlas.

Estos grupos sufren más falsos positivos, siendo clasificados como de “alto riesgo” mayormente.

Falta de calibración y riesgos éticos

Aunque COMPAS tiene cierta capacidad predictiva...


...no está calibrado de forma equitativa entre subpoblaciones.



Su aplicación sin un análisis ético y contextual puede reforzar desigualdades judiciales y poner en riesgo principios fundamentales de justicia e imparcialidad.

Limitaciones del proyecto

Cobertura restringida: solo datos de 2013-2014 (Broward, Florida).

 **Reincidencia = arresto:** puede sobreestimar la reincidencia real.

Desbalance demográfico: baja representación de mujeres y minorías.

Variables limitadas: sin factores socioeconómicos o educativos.

Modelo simplificado (regresión logística): enfoque didáctico, no optimizado para producción.

RECOMENDACIONES

9

RECOMENDACIONES SISTEMA JUDICIAL EE.UU

Implicaciones éticas y legales

Delegar decisiones judiciales en algoritmos sin supervisión humana ni auditoría plantea riesgos éticos y legales.

♦ Recomendamos:

- Supervisión humana obligatoria.
- Transparencia y auditorías públicas.
- Derecho de apelación algorítmica (*caso Wisconsin vs Loomis*).

Reducción de sesgos y mejora de los datos

Sin datos justos, no hay algoritmos justos.

Los sesgos de COMPAS reflejan desigualdades en los procesos de detención.

- Ampliar y equilibrar las bases de datos (más diversos y actualizados).
- Incorporar variables socioeconómicas y educativas.
- Homogeneizar criterios metodológicos (± 30 días, 2 años reales).
- Definir qué entendemos por “equidad” antes de medirla.
- Promover equipos de desarrollo diversos y formados en ética.

Algoritmos transparentes y auditable

La justicia algorítmica requiere control humano.

Modelos simples como la regresión logística demuestran que la transparencia puede coexistir con la eficacia.

- Mantener modelos de referencia como benchmarks.
- Revisiones periódicas de sesgo y precisión mediante dashboards públicos.
- Formación en alfabetización algorítmica para jueces y fiscales.
- Protocolos de transparencia y rendición de cuentas institucional.

“Este proyecto nos ha desbordado, retado y
enseñado.

Nos ha mostrado lo mucho que ignorábamos, lo
que somos capaces de aprender y lo que aún nos
queda por recorrer.”

Aza, Dani y María