

PROYECTO SAFECLAIM

IGRUPO# 2

INTEGRANTES:

- Aldahir Ramirez
- Ginger Samaniego
- Roberto Prado
- Melina Mendez



INTRODUCCION

- Comprender patrones y comportamientos en la gestión de reclamaciones de seguros.
- Detectar variables clave relacionadas con el fraude.
- Identificar patrones atípicos (temporales y demográficos).



DATASET

DATASET Y VARIABLES ANALIZADAS

El dataset contiene información completa de reclamaciones de seguros, incluyendo datos personales del asegurado, características del vehículo, detalles del reclamo, historial del cliente y fechas clave del proceso.

TRANSFORMACIONES REALIZADAS

Se identificaron valores nulos, columnas numéricas, outliers. También se detectó un fuerte desbalance en la variable objetivo, lo que influye en cualquier análisis predictivo futuro.

ESTADO DEL DATASET



VALORES NULOS

- AccidentArea: 2.09%
- MaritalStatus: 1.39%

OUTLIERS Y ERRORES

- Edades fuera de rango (incluye registros con edad 0)
- Deducibles con valores atípicos o poco realistas

DESBALANCE DE LA VARIABLE OBJETIVO

- 94% casos sin fraude
- 6% casos con fraude

ANALISIS DE PATRONES TEMPORALES DE FRAUDE

OBJETIVO

Identificar patrones temporales asociados a comportamientos fraudulentos mediante el análisis de variables temporales existentes y la creación de nuevas variables derivadas. Este análisis integrará múltiples dimensiones temporales con características demográficas y de contexto del siniestro.



VARIABLES TEMPORALES

Variables temporales generadas

- Meses entre accidente y reclamo (Months_Accident_to_Claim)
- Velocidad del reclamo (Claim_Speed_Category)
- Trimestre del accidente y del reclamo
- Tipo de día (laboral o fin de semana)

Identificar si el tiempo entre el accidente y el reclamo muestra patrones o comportamientos inusuales.

==== DISTRIBUCION DE VELOCIDAD DE RECLAMACION ===

Claim_Speed_Category	
Mismo Mes	11442
1 Mes Despues	3211
2-3 Meses	535
4-6 Meses	155
7-12 Meses	76

LIMPIEZA DE DATOS TEMPORALES

VALORES NULOS MÍNIMOS

Solo un registro tenía valores no asignados en MonthClaimed y DayOfWeekClaimed.

MAPEOS VÁLIDOS:

casi todos los meses y días se convirtieron correctamente a formato numérico.

REGISTROS TARDÍOS

Menos del 0.5% reporta el reclamo más de 6 meses después del accidente (casos poco comunes).

DECISIÓN TOMADA

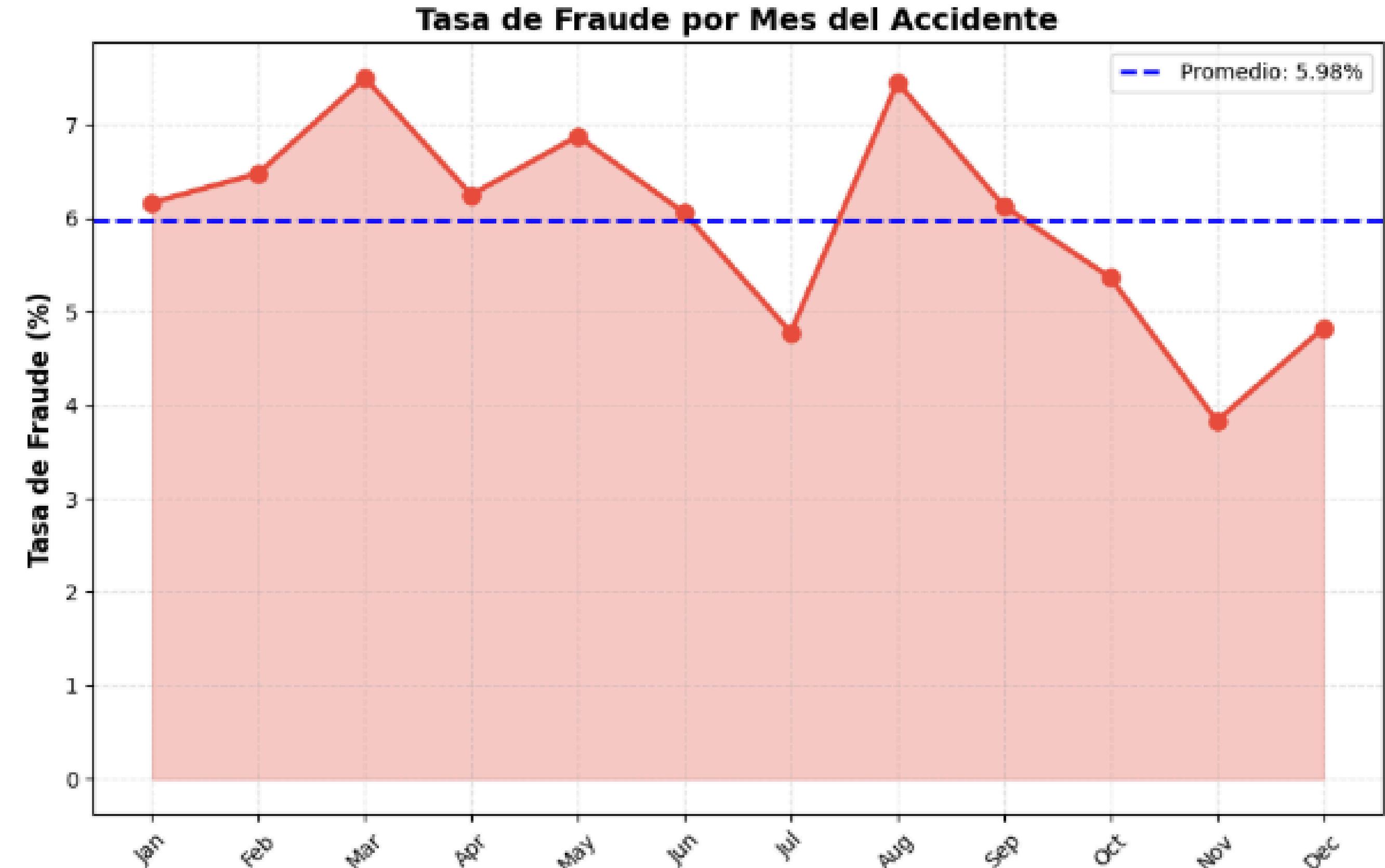
Se eliminó únicamente 1 registro, ya que su información temporal estaba incompleta.

LIMITACIONES

Las variables temporales contenían rangos amplios en lugar de valores numéricos.

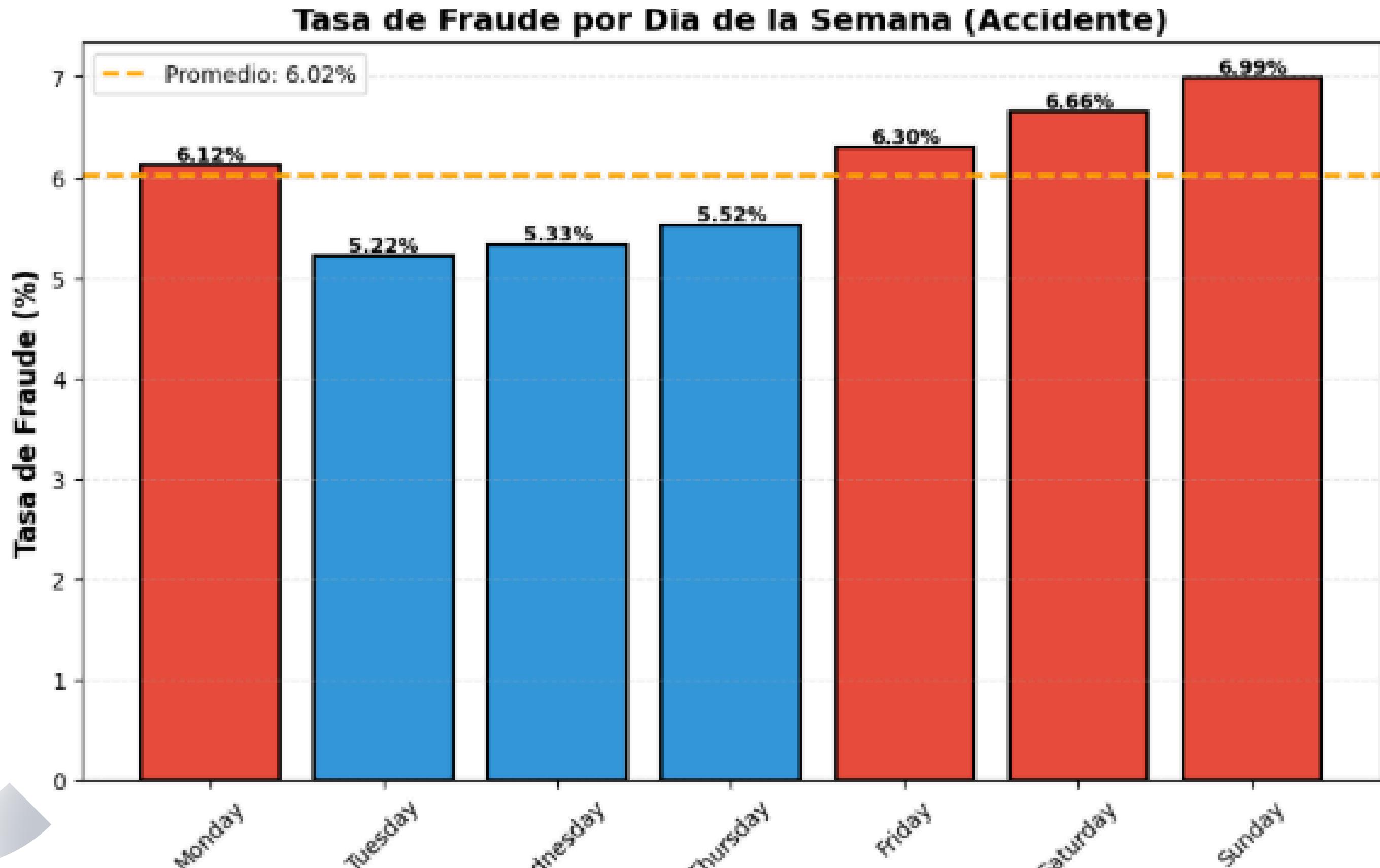
PREGUNTA 1: ¿HAY MESES CON MAYOR FRAUDE?

- Meses como marzo, mayo y agosto superan el promedio general (~6%), por lo que se consideran los meses con mayor incidencia de fraude.



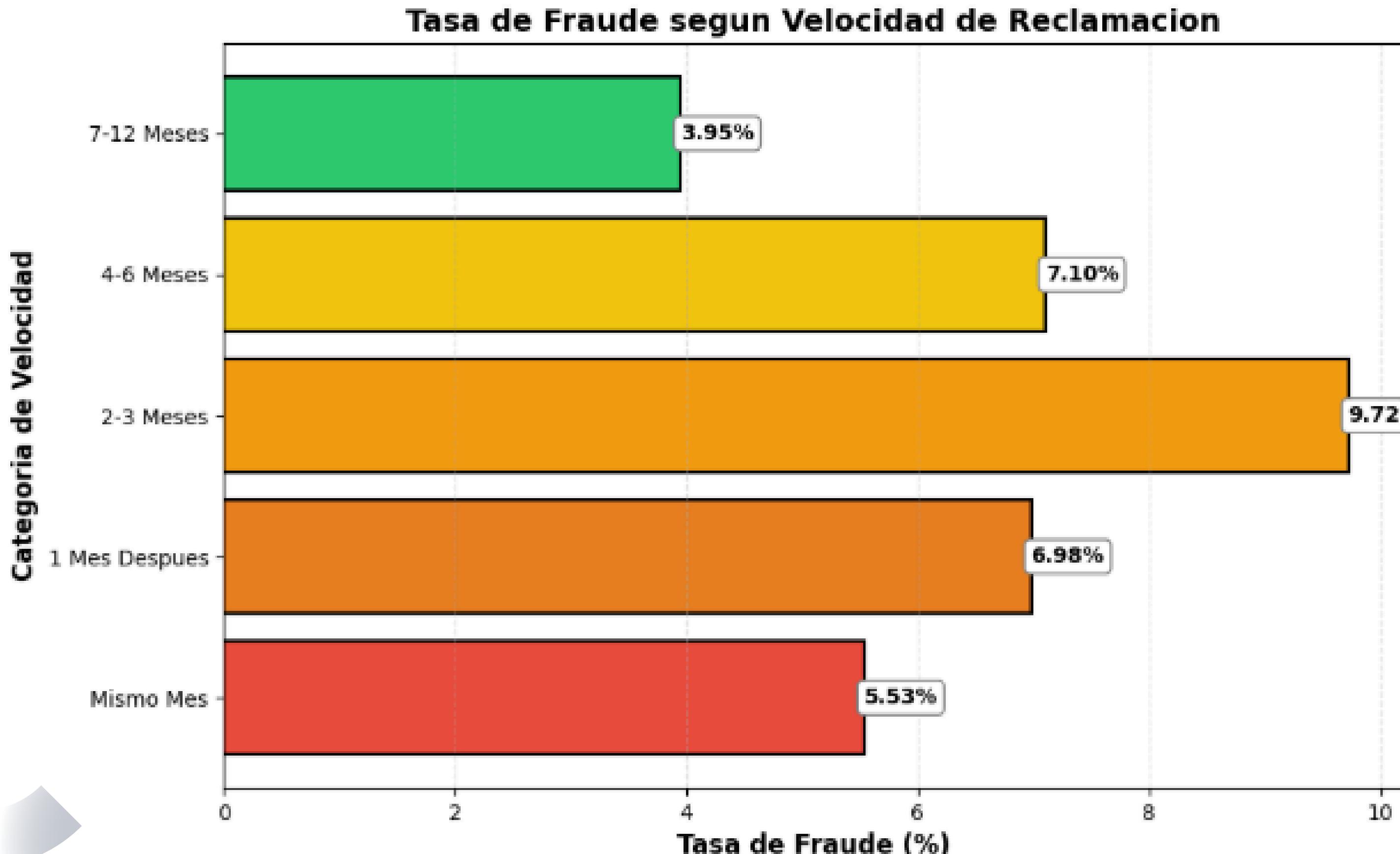
PREGUNTA 2: ¿HAY DÍAS DE LA SEMANA CON MAYOR PROPORCIÓN DE CASOS FRAUDULENTOS?

El 6.99% de los fraudes por accidente se concentran los días Domingos



PREGUNTA 3: ¿CUÁL ES LA DIFERENCIA TEMPORAL ENTRE EL ACCIDENTE Y LA RECLAMACIÓN EN CASOS DE FRAUDE?

- Los casos fraudulentos reclaman 55% más tarde que los legítimos.
- Reclamar en el mismo mes tiene la tasa de fraude más baja (5.37%).



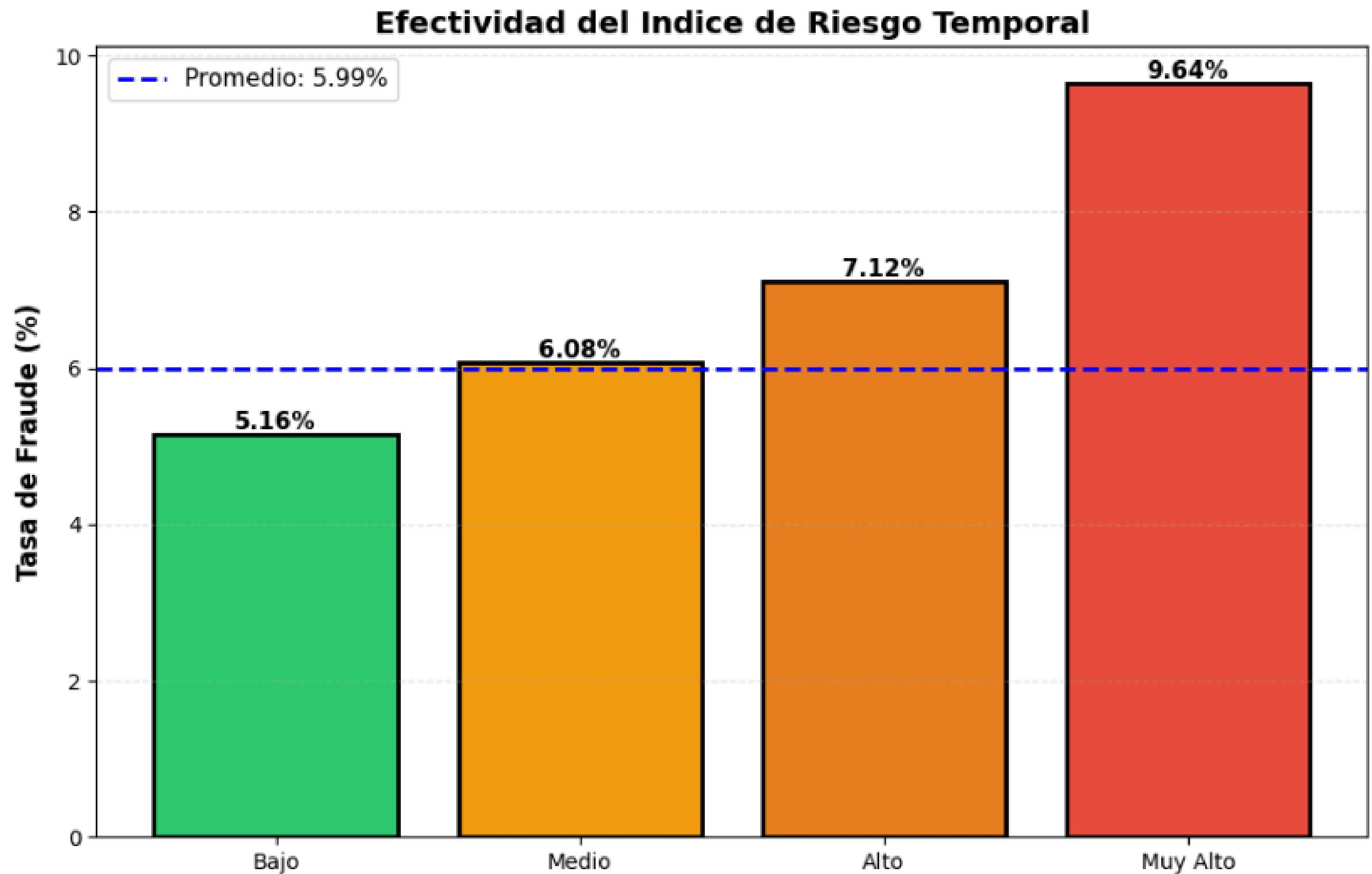
PREGUNTA 4: ¿LOS PATRONES TEMPORALES IDENTIFICAN BANDERAS ROJAS PARA DETECCION TEMPRANA?

Score IRT			
Factor Temporal	Condición (Regla)	Puntuación (Peso)	Justificación (Basada en el Análisis)
1. Mes del Accidente	Ene-May: Month_Num ∈[1,2,3,4,5]	3	Estos meses mostraron las tasas de fraude más altas en
	Nov-Dic: Month_Num ∈[11,12]	1	Riesgo medio.
2. Trimestre del Accidente	Q1 o Q2: Quarter_Accident ∈ ['Q1 (Ene-Mar)', 'Q2 (Abr-	2	Refuerza el patrón de alto riesgo en el primer
3. Día de la Semana	Domingo: DayOfWeek == 'Sunday'	2	El Domingo tuvo la tasa de fraude más alta.
	Sábado: DayOfWeek == 'Saturday'	1	Riesgo medio.
4. Velocidad de Reclamación	Reclamo Tardío (Muy Sospechoso):	2	Se identificó que los casos de fraude
	Reclamo Tardío (Sospechoso):	1	
5. Reclamación en Diferente Mes	Diferente Mes: Same_Month == 0	1	Complementa el Factor 4, penalizando la no

Nivel de Riesgo	Rango de Puntos
Bajo	0-2 puntos
Medio	3-5 puntos
Alto	6-7 puntos
Muy Alto	8-10 puntos

PREGUNTA 4: ¿LOS PATRONES TEMPORALES IDENTIFICAN BANDERAS ROJAS PARA DETECCION TEMPRANA?

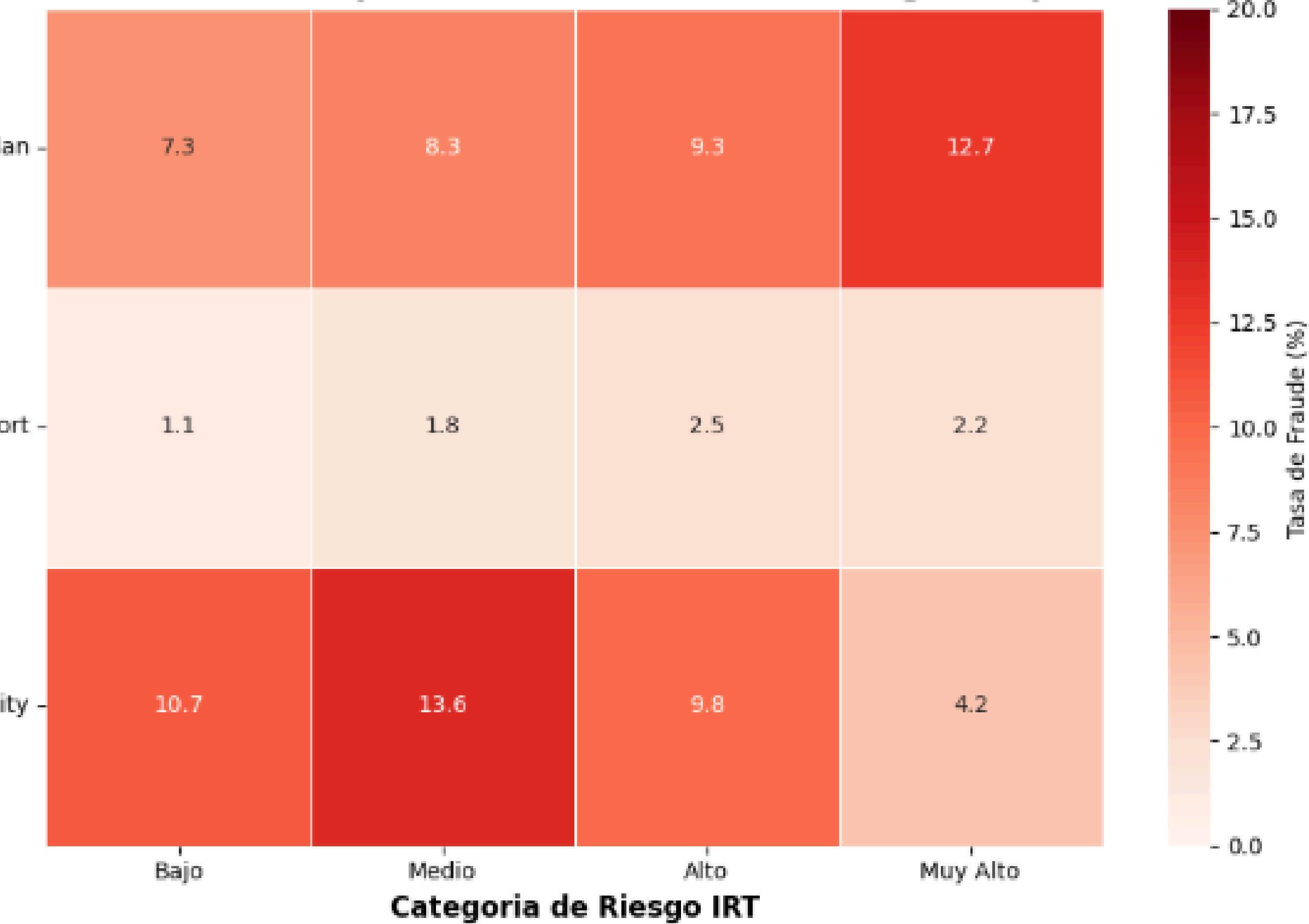
- Riesgo Bajo: 5.16%, por debajo del promedio (5.99%).
- Riesgo Medio y Alto ya se ubican por encima del promedio.
- Riesgo Muy Alto: 9.64%, el nivel más crítico.



PREGUNTA 5: ¿EL TIPO DE VEHÍCULO TAMBIÉN INFUYE EN EL IRT?

- Nivel de IRT Utility 23% en la categoría Medio y Alto
- Nivel de IRT Sedan 22% en la categoría Alto y Muy Alto

Tasa de Fraude: Tipo de Vehículo x Nivel de Riesgo Temporal



• CONCLUSIONES

- El análisis temporal permitió identificar patrones claros asociados al fraude, mostrando que ciertos días, secuencias y tiempos inesperados elevan significativamente el riesgo.
- El IRT sintetiza estas dimensiones temporales y distingue con precisión los casos más sospechosos, permitiendo separar reclamos de bajo riesgo de aquellos que requieren revisión inmediata.
- Integrar variables temporales con características del siniestro y del vehículo mejora la detección temprana sin necesidad de modelos complejos.

• RECOMENDACIONES

- Usar el IRT para priorizar los reclamos con mayor riesgo.
- Automatizar la revisión de casos de bajo riesgo.
- Reforzar la investigación en reclamos con patrones temporales sospechosos.
- Capacitar al equipo en detección temprana basada en tiempos y secuencias.
- Mantener un monitoreo continuo de los patrones temporales para ajustar estrategias.