

PROYECTO FINAL DATA SCIENCE

ACCIDENTES DE TRÁFICO EN BRASIL
(2020-2022)

LOURDES OVELAR

Este análisis explora los datos sobre los accidentes de tráfico en Brasil, cubriendo desde 2020 al 2022. El foco está en descubrir los patrones y tendencias subyacentes que puedan esclarecer los factores contribuyentes a estos siniestros y apoyar en la formulación de estrategias más efectivas para incrementar la seguridad vial.

INTRODUCCIÓN

EL CONTEXTO Y EL PROBLEMA COMERCIAL

Necesidad de Medidas Preventivas: Enfoque en reducir accidentes viales en Brasil mediante análisis de datos detallados.

Preocupación y Soluciones en Seguridad Vial: "Aplicación de técnicas analíticas para comprender y mejorar la seguridad vial, basadas en patrones y correlaciones de datos

El proyecto se enfoca en analizar y visualizar datos de accidentes viales recopilados recientemente. El objetivo es identificar patrones para comprender mejor las circunstancias detrás de los accidentes y desarrollar visualizaciones interactivas para aumentar la conciencia sobre la seguridad vial.

OBJETIVO DEL PROYECTO

-
1. Horas Pico y Días de Semana: Impacto de las Condiciones Meteorológicas: La gravedad de los accidentes tiende a aumentar en condiciones meteorológicas con cielo despejado.
 2. Diferencias entre Autopistas y Calles Urbanas: Las autopistas muestran una mayor incidencia de accidentes graves en comparación con las calles urbanas.
 3. Patrones Geográficos de Accidentes: Se identifican áreas geográficas con alta concentración de ciertos tipos de accidentes, indicando la necesidad de medidas de seguridad específicas en esas zonas.

RECOMENDACIONES BASADAS EN INSIGHTS OBSERVADOS

HIPÓTESIS

1. La gravedad de los accidentes de tráfico es mayor en las autopistas que en las calles urbanas debido a la alta velocidad permitida en las autopistas.
2. Los accidentes de tráfico son más frecuentes durante los días laborables (de lunes a viernes) en comparación con los fines de semana, debido al aumento del tráfico durante las horas pico.
3. La gravedad de los accidentes es más alta durante las condiciones meteorológicas adversas, como lluvia intensa o neblina, debido a la reducción de la visibilidad y la disminución del agarre en la carretera.
4. Las zonas urbanas densamente pobladas tienen una mayor incidencia de accidentes de tráfico debido a la congestión del tráfico y a la interacción cercana entre vehículos y peatones.

PREGUNTAS DE INTERÉS

1. ¿Cuál es la tendencia en el número total de accidentes a lo largo de los años?
2. ¿Existe una estacionalidad en la ocurrencia de accidentes a lo largo de los meses o las estaciones del año?
3. ¿Cuáles son las condiciones meteorológicas más comunes en los accidentes y cómo afectan a la gravedad de los mismos "Mortos, Feridos_leves, Feridos_graves, llesos o Ignorados"?
4. ¿Qué días de la semana tienen la mayor cantidad de accidentes y cómo varía esta distribución?
5. ¿Cómo se distribuyen los tipos de carreteras en relación con los accidentes?
6. ¿Cuál es el mes y el día con mayor cantidad de accidentes a lo largo de los años analizados?
7. ¿Cómo varía el número de accidentes a lo largo de un solo día?
8. ¿Cuáles son las Unidades Federales con mayores índices de accidentes?

ENFOQUE METODOLÓGICO

Análisis detallado de datos de accidentes viales en Brasil de 2020 a mediados de 2022. Empleo de regresión lineal para comprender patrones y correlaciones entre variables como ubicación, hora del día, condiciones meteorológicas y tipos de carreteras.

Objetivo: aumentar la conciencia sobre seguridad vial y promover conductas más seguras. Oportunidad de aprendizaje y aplicación de soluciones prácticas en problemas reales.

PLANTEAMIENTO DE LA EXPLORACIÓN DE DATOS: ACCIDENTES DE TRÁFICO EN BRASIL (2020-2022)

1. Los datos abarcan accidentes de tráfico en Brasil desde 2020 hasta 2022 e incluyen variables como fecha y hora del accidente, ubicación geográfica, condiciones meteorológicas, tipo de carretera, número de víctimas y gravedad del accidente.
2. El análisis univariable incluirá estadísticas descriptivas para variables numéricas como el número de víctimas por accidente, y examinará la distribución de tipos de carreteras para entender las proporciones de autopistas y calles urbanas, entre otros

PLANTEAMIENTO DE LA EXPLORACIÓN DE DATOS: ACCIDENTES DE TRÁFICO EN BRASIL (2020-2022)

1. El análisis bivariable investigará la relación entre la gravedad del accidente y la hora del día, así como las correlaciones entre condiciones meteorológicas y el número de accidentes para identificar patrones climáticos asociados con un aumento en los accidentes.
2. También se realizará una exploración temporal mediante un gráfico de línea para visualizar la tendencia anual de accidentes. En cuanto a la visualización de datos, se utilizarán gráficos de barras y torta para mostrar la frecuencia de diferentes tipos de accidentes

1 Id_acidente
2 Año_acidente
3 Dia_semana: Los días de semanas de los accidentes
4 Horario: Horario de ocurrencia de cada accidente
5 Unidad_federal: División por estados y regiones
6 Br: El el número de las rutas o autopistas
7 Km: El kilómetro en dónde ocurren los accidentes
8 Municipio: Municio de ocurrencia de los accidentes
9 Causa_acidente
10 Tipo_acidente: Motivo de los accidentes
11 Classificacao_acidente: Clasificación de los accidentes de acuerdo al estado de las víctimas
12 Fase_dia: Horario del día del accidente
13 Sentido_via: Dirección del carril
14 Condicao_metereologica: La condición climática de los accidentes
15 Tipo_pista: Tipos de pistas de las rutas
16 Tracado_via: Disposición de las vías
17 Uso_solo: Muestra si los carriles son manos únicas o dobles
18 Pessoas: Cantidad de persnas afectadas
19 Mortos: Cantidad personas fallecidas
20 Feridos_leves: Números de personas con heridas leves
21 Feridos_graves: Números de personas con heridas graves
22 Ilesos: Números de personas hilésos
23 Ignorados: Números de personas sin registro de estado
24 Feridos: Números de personas con heridas
25 Veiculos: Cantidad de vehículos involucrados
26 Latitude
27 Longitude
28 Regional: Región en la cual se predujo el accidente
29 Delegacia: Delegación de las denuncias
30 uop: Id de la delegación

CONTEXTO ANALÍTICO

VISUALIZACION DE DATOS

c:\Users\lourd\OneDrive\Escritorio\TP\Tp-Ovelar.ipynb > df (191348, 31)

		index	Id_aci...	Año_a...	Dia_se...	Horario	Unida...	Br	Km	Munic...	Causa...	Tipo_a...	Classif...	Fase_dia	Sentid...	Condi...	Tipo_...	Tracad...	Uso_s...	Pessc	
		364	365	262321	2020-01-...	domingo	11:20:00	MG	251	354	FRUTA D...	Velocida...	Tombam...	Com Víti...	Pleno dia	Decresce...	Sol	Simples	Não Infor...	Não	2
		365	366	262333	2020-01-...	domingo	13:20:00	SC	101	264,4	PAULO L...	Falta de ...	Colisão l...	Com Víti...	Pleno dia	Decresce...	Sol	Dupla	Reta	Não	4
		366	367	262341	2020-01-...	domingo	02:30:00	MG	381	720,7	CARMO ...	Velocida...	Tombam...	Com Víti...	Pleno dia	Decresce...	Nublado	Dupla	Curva	Não	1
		367	368	262343	2020-01-...	domingo	15:00:00	RJ	116	225,6	PIRAI	Pista Esc...	Colisão tr...	Com Víti...	Pleno dia	Decresce...	Chuva	Dupla	Curva	Não	4
		368	369	262344	2020-01-...	domingo	13:50:00	RJ	40	111	DUQUE ...	Falta de ...	Colisão l...	Com Víti...	Pleno dia	Decresce...	Nublado	Múltipla	Não Infor...	Sim	21
		369	370	262348	2020-01-...	domingo	15:00:00	MG	381	500	BETIM	Defeito ...	Incêndio	Com Víti...	Pleno dia	Decresce...	Céu Claro	Dupla	Reta	Não	1
		370	371	262349	2020-01-...	domingo	13:00:00	MT	70	507	CUIABA	Ultrapass...	Colisão fr...	Com Víti...	Pleno dia	Crescente	Céu Claro	Simples	Reta	Não	3
		371	372	262358	2020-01-...	domingo	16:00:00	ES	262	58	MARECH...	Condutor...	Colisão fr...	Com Víti...	Pleno dia	Crescente	Sol	Simples	Curva	Não	6
		372	373	262368	2020-01-...	domingo	17:00:00	RN	304	305	PARNAM...	Desobedi...	Colisão tr...	Com Víti...	Anoitecer	Crescente	Céu Claro	Dupla	Reta	Não	3
		373	374	262373	2020-01-...	domingo	16:20:00	MT	364	474	JANGADA	Defeito ...	Colisão c...	Com Víti...	Pleno dia	Decresce...	Sol	Simples	Não Infor...	Não	4
		374	375	262380	2020-01-...	domingo	17:00:00	MG	381	293	ANTONI...	Velocida...	Colisão fr...	Com Víti...	Pleno dia	Decresce...	Céu Claro	Simples	Reta	Não	2
		375	376	262385	2020-01-...	domingo	16:50:00	SC	101	156	PORTO B...	Carga ex...	Colisão c...	Sem Víti...	Pleno dia	Decresce...	Sol	Simples	Reta	Não	1
		376	377	262400	2020-01-...	domingo	19:00:00	PA	155	338	MARABA	Desobedi...	Colisão fr...	Com Víti...	Anoitecer	Decresce...	Chuva	Simples	Reta	Não	20
		377	378	262405	2020-01-...	domingo	09:00:00	MA	222	541	BOM JES...	Velocida...	Colisão fr...	Com Víti...	Pleno dia	Decresce...	Sol	Simples	Curva	Não	2
		378	379	262408	2020-01-...	domingo	14:55:00	PR	476	48	ADRIAN...	Desobedi...	Colisão l...	Com Víti...	Pleno dia	Crescente	Nublado	Simples	Curva	Não	3
		379	380	262417	2020-01-...	domingo	17:15:00	MG	40	822,4	SIMAO P...	Condutor...	Colisão tr...	Com Víti...	Pleno dia	Decresce...	Nublado	Dupla	Reta	Não	4
		380	381	262418	2020-01-...	domingo	16:30:00	MA	316	378	BACABAL	Desobedi...	Colisão fr...	Com Víti...	Pleno dia	Crescente	Céu Claro	Simples	Reta	Sim	2
		381	382	262424	2020-01-...	domingo	17:00:00	ES	262	29	DOMING...	Falta de ...	Colisão c...	Sem Víti...	Plena No...	Decresce...	Nublado	Dupla	Curva	Não	1
		382	383	262428	2020-01-...	domingo	20:55:00	SE	101	47	JAPARAT...	Desobedi...	Colisão tr...	Com Víti...	Plena No...	Decresce...	Céu Claro	Dupla	Reta	Não	6
		383	384	262437	2020-01-...	domingo	19:30:00	RO	364	441	JARU	Desobedi...	Colisão fr...	Com Víti...	Plena No...	Decresce...	Céu Claro	Simples	Reta	Não	2
		384	385	262440	2020-01-...	domingo	19:20:00	TO	230	29	NAZARE	Desobedi...	Colisão fr...	Com Víti...	Plena No...	Crescente	Céu Claro	Simples	Reta	Não	4
		385	386	262452	2020-01-...	domingo	19:40:00	PE	316	331	FLORESTA	Ingestão ...	Atropela...	Com Víti...	Plena No...	Decresce...	Céu Claro	Simples	Reta	Não	3
		386	387	262456	2020-01-...	domingo	22:30:00	MG	381	477,2	CONTAG...	Falta de ...	Colisão tr...	Com Víti...	Plena No...	Crescente	Céu Claro	Dupla	Não Infor...	Sim	2
		387	388	262467	2020-01-...	segunda-...	00:10:00	PA	155	336	MARABA	Ingestão ...	Saída de ...	Sem Víti...	Plena No...	Crescente	Céu Claro	Simples	Reta	Não	3

PROCESAMIENTO DE DATOS

Exploración valores vacíos

```
valores_nulos = df.isnull().sum()
columnas_con_nulos = valores_nulos[valores_nulos == 0]
columnas_con_nulos
```

[53]

```
... Id_accidente      0
    Año_accidente     0
    Dia_semana        0
    Horario           0
    Unidad_federal    0
    Municipio         0
    Causa_accidente   0
    Tipo_accidente    0
    Classificacao_accidente 0
    Fase_dia          0
    Sentido_via       0
    Condicao_meteorologica 0
    Tipo_pista        0
    Tracado_via       0
    Uso_solo          0
    Pessoas           0
    Mortos            0
    Feridos leves     0
```

Remoción de duplicados

```
[270] df.duplicated().any()
...   True

[271] df.drop_duplicates(inplace=True)
      df.duplicated().any()
...   False

[272] df['Dia_semana'] = df['Dia_semana'].str.title()
      df['Municipio'] = df['Municipio'].str.title()
```

Eliminación de registros con valor nulo

```
[3] df.isnull().any().sum()
0

[4] df.dropna(inplace=True)
    df.isnull().any().any()
```

PROCESAMIENTO DE DATOS

Función para eliminar valores atípicos a más de 5 desviaciones estándar de la media.

```
def outliers(df, columnas):  
    indices_incomuns = []  
    todos_indices = np.zeros(len(df), dtype=bool)  
  
    for columna in columnas:  
        df_coluna = np.array(df[columna])  
  
        desvio = df_coluna.std()  
        media = df_coluna.mean()  
  
        indice_incomum = df_coluna > (media + 5 * desvio)  
  
        indices_incomuns.append(indice_incomum)  
  
    for i in range(len(indices_incomuns)):  
        todos_indices = np.logical_or(todos_indices, indices_incomuns[i])  
  
    return todos_indices
```

```
# Removendo os outliers das colunas Pessoas e Veículos.  
indices_outliers = outliers(df, ['Pessoas', 'Veiculos'])  
df_sin_anio = df[~indices_outliers]
```

```
df.describe()
```

	Año_accidente	Br	Pessoas	Mortos	Feridos_leves	Feridos_graves	Ilesos	Ignorados	Veiculos	km
count	191338	191338.000000	191338.000000	191338.000000	191338.000000	191338.000000	191338.000000	191338.000000	191338.000000	191338.000000
mean	2021-07-06 18:11:12.142491648	211.711453	2.354493	0.083747	0.849194	0.274812	0.986208	0.160533	1.647164	259.599213
min	2020-01-01 00:00:00	10.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
25%	2020-10-12 00:00:00	101.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	78.800000
50%	2021-07-08 00:00:00	158.000000	2.000000	0.000000	1.000000	0.000000	1.000000	0.000000	2.000000	192.800000
75%	2022-04-06 00:00:00	324.000000	3.000000	0.000000	1.000000	0.000000	1.000000	0.000000	2.000000	406.700000
max	2022-12-31 00:00:00	495.000000	75.000000	19.000000	50.000000	31.000000	73.000000	54.000000	23.000000	1454.500000
std	NaN	130.916179	1.890825	0.338424	1.038329	0.609714	1.479499	0.471636	0.723807	226.246341

VISUALIZACION DE GRÁFICOS

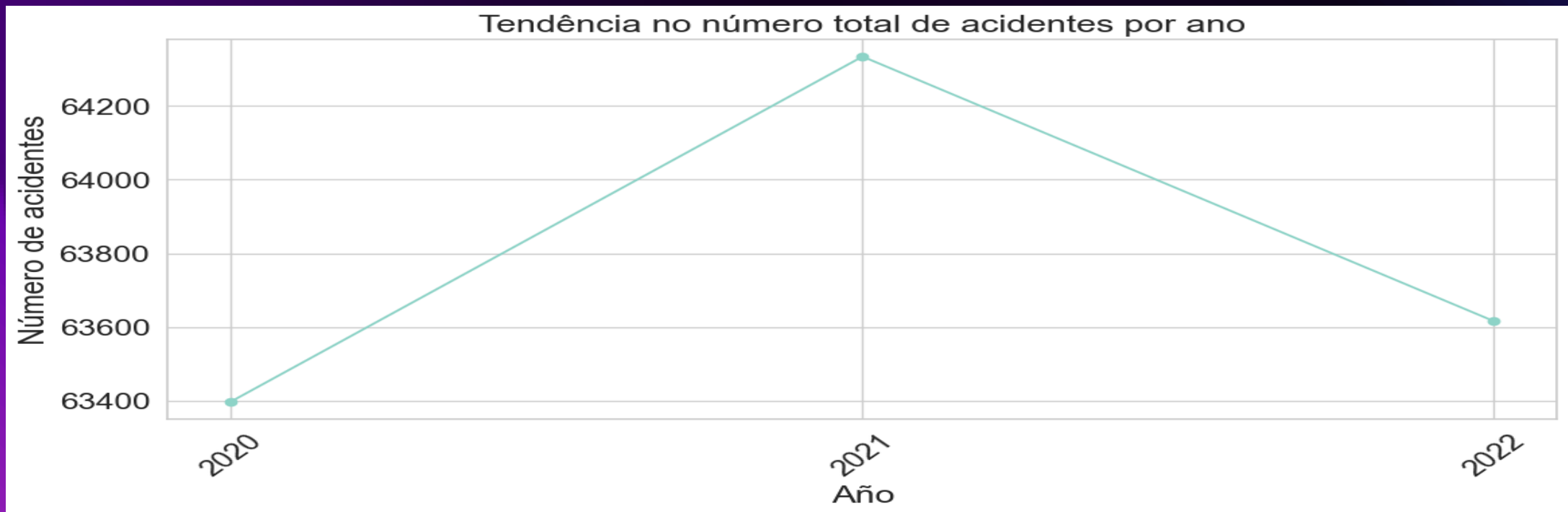
Tendencia en el número total de accidentes a lo largo de los años

En 2020, el gráfico comienza con aproximadamente 63,400 accidentes.

Luego, en 2021, hay un pico significativo que muestra un aumento hasta alrededor de 64,200 accidentes.

Para 2022, se observa una disminución hasta 63,600 accidentes.

El pico en 2021 puede sugerir varios factores, como cambios en las condiciones de la carretera, modificaciones en las leyes de tráfico, o incluso factores externos como el clima.



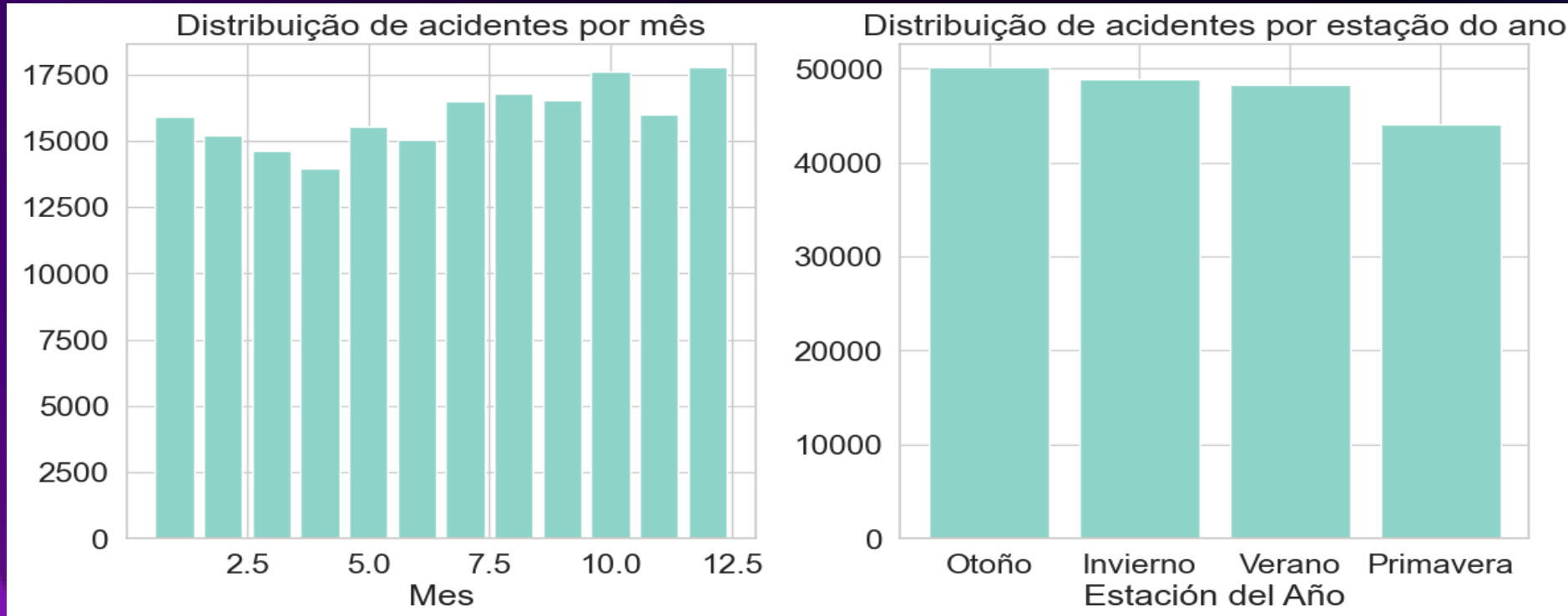
VISUALIZACION DE GRÁFICOS

Estacionalidad en la ocurrencia de accidentes a lo largo de los meses o las estaciones del año

El análisis examina la distribución de accidentes a lo largo de los meses y las estaciones del año mediante gráficos de barras y estadísticas resumidas. No se observa una tendencia clara ni picos significativos que sugieran una estacionalidad fuerte en los datos mensuales.

Verano: Algo más de 40,000 accidentes

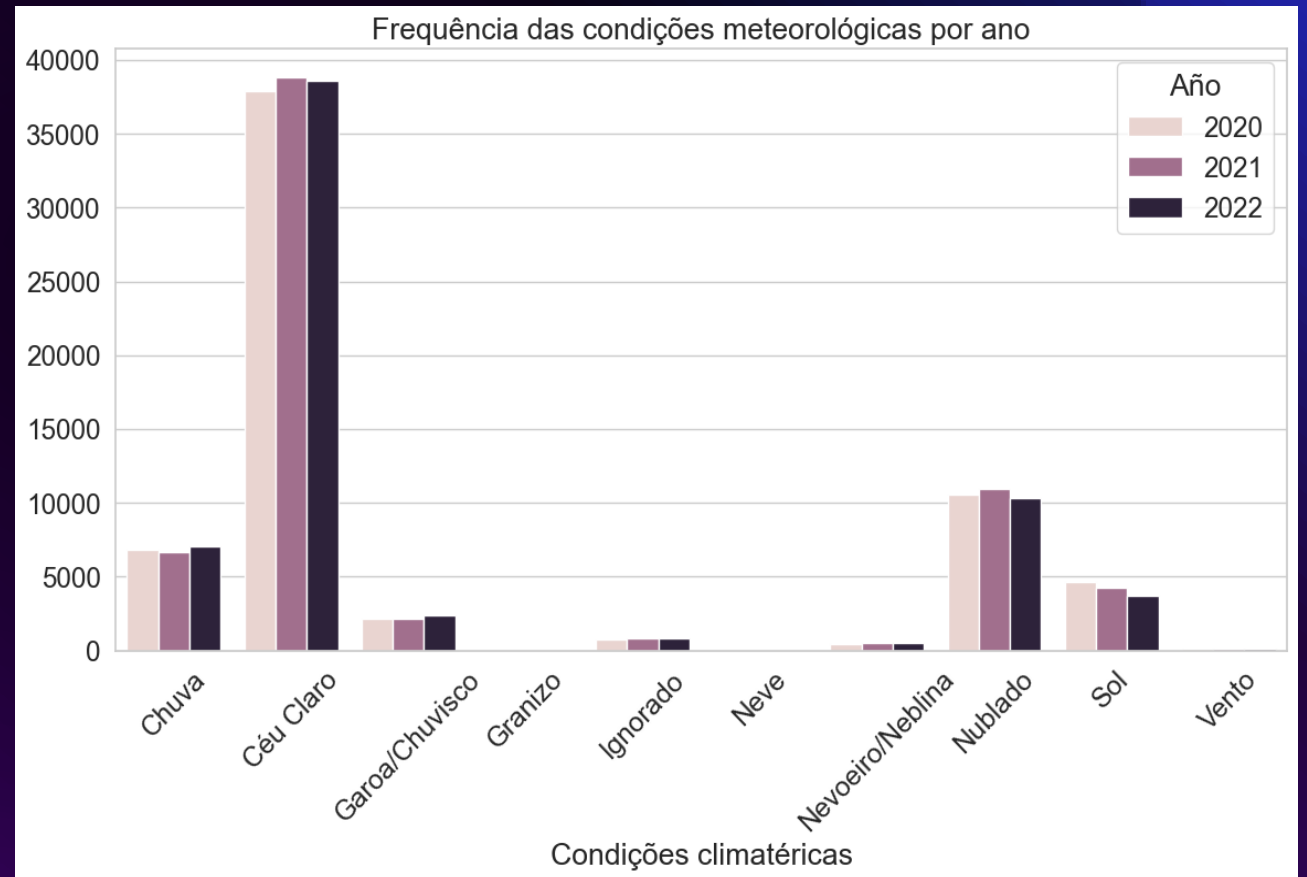
Primavera: Alrededor de 40,000 accidentes



VISUALIZACION DE GRÁFICOS

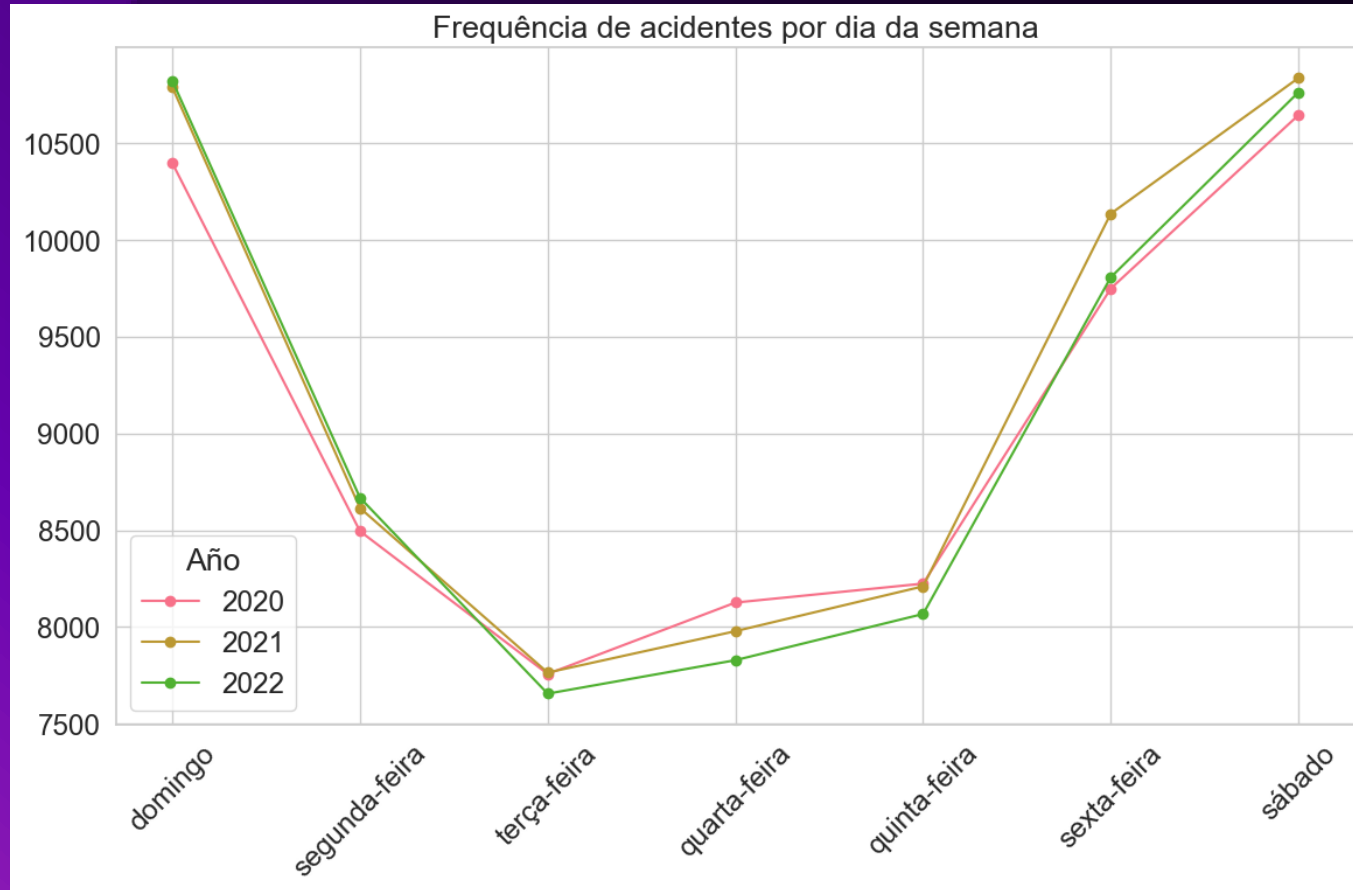
Condiciones meteorológicas más comunes en los accidentes

Este gráfico de barras permite comparar la distribución de accidentes por tipo de pista en diferentes años, lo que puede ayudar a identificar patrones o cambios en la preferencia de tipo de pista.



VISUALIZACION DE GRÁFICOS

- Días de la semana tienen la mayor cantidad de accidentes

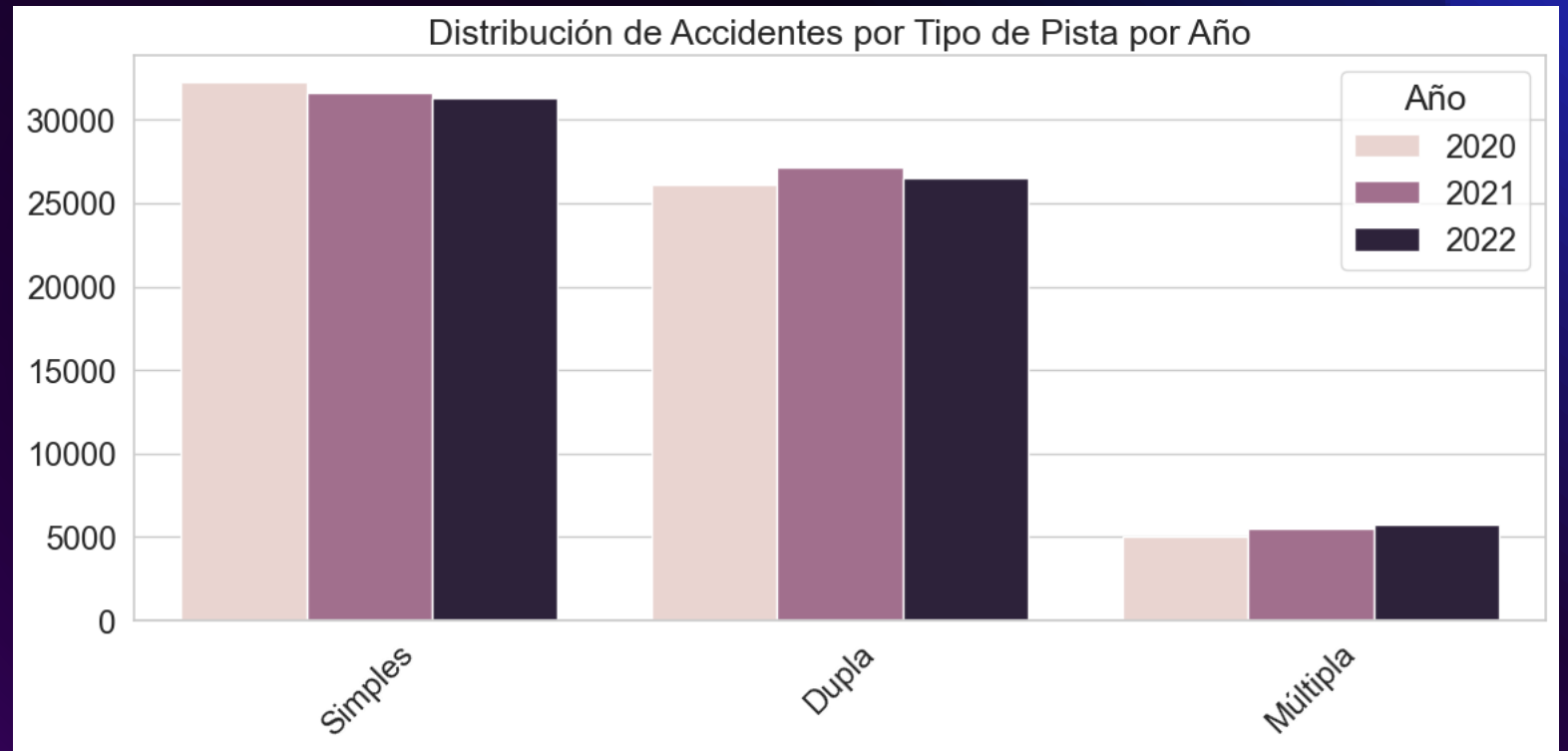


Permite comparar la frecuencia de accidentes en diferentes días de la semana a lo largo de los años entre 2020 a 2022.

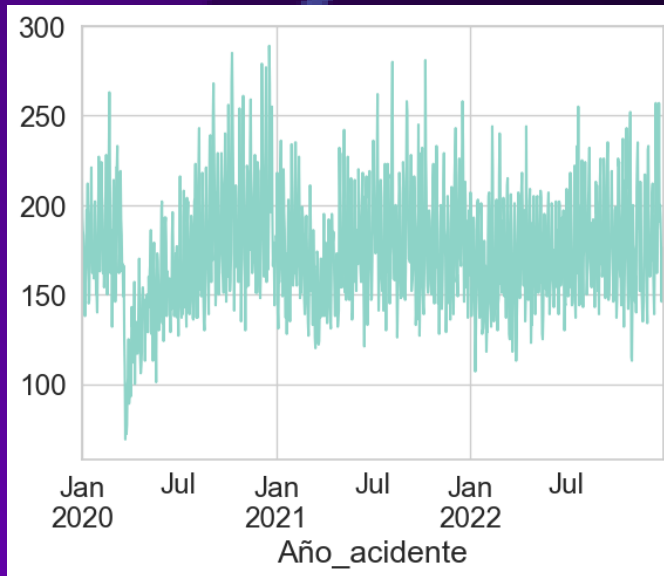
VISUALIZACION DE GRÁFICOS

○ Tipos de carreteras en relación con los accidentes

Este gráfico de barras permite comparar la distribución de accidentes por tipo de pista en diferentes años, lo que puede ayudar a identificar patrones o cambios en la preferencia de tipo de pista.



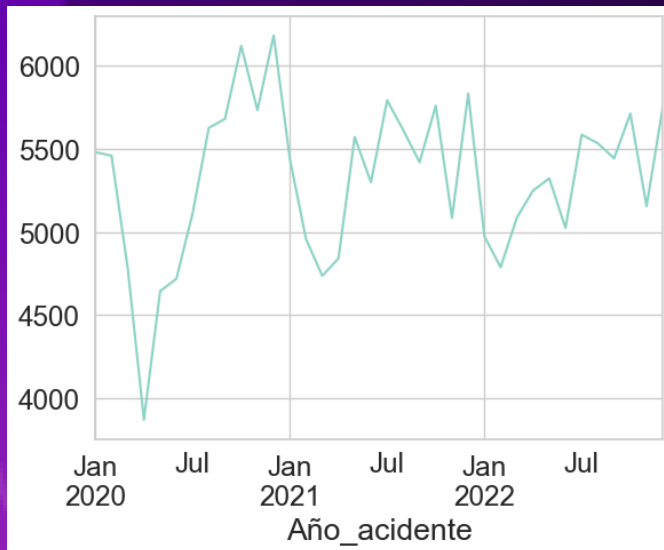
VISUALIZACION DE GRÁFICOS



Análisis mensuales de los accidentes

La gráfica muestra la tendencia de accidentes mes a mes durante aproximadamente dos años y medio, desde enero de 2020 hasta julio de 2022. Aquí está mi análisis como estudiante de ciencia de datos

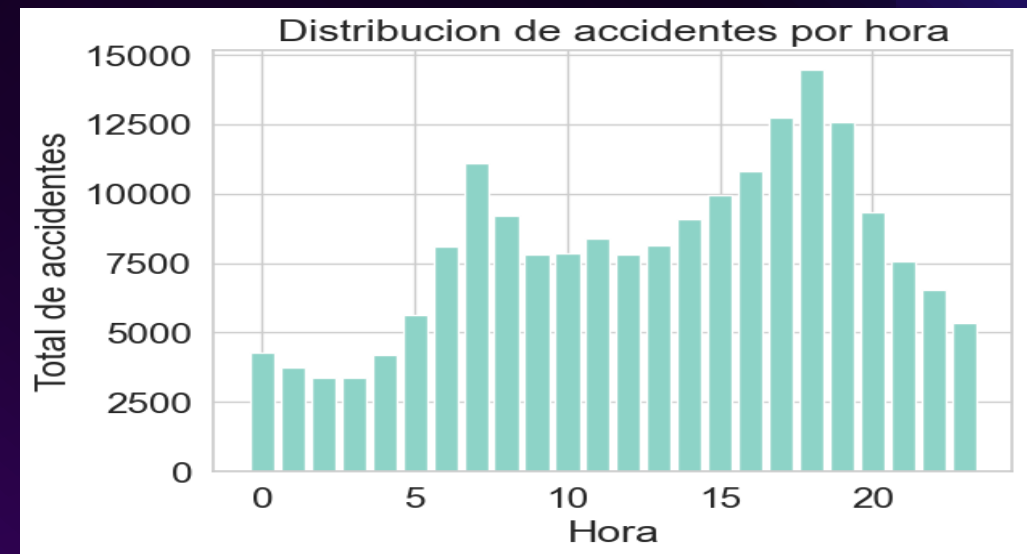
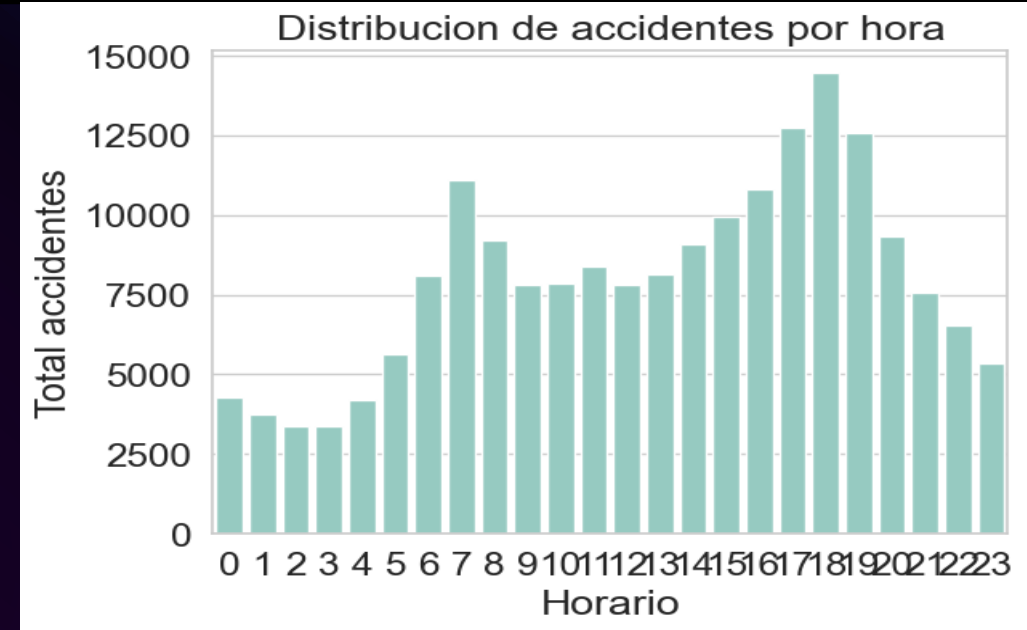
Existe una marcada disminución inicial en el número de accidentes a principios de 2020, lo que podría corresponder a las restricciones de viaje y confinamientos debidos a la pandemia de COVID-19.



VISUALIZACION DE GRÁFICOS

○
.
Análisis mensuales de los accidentes por distribución de horarios.

En estos gráficos se visualizan las variaciones de los números de accidentes por distribución de horarios.

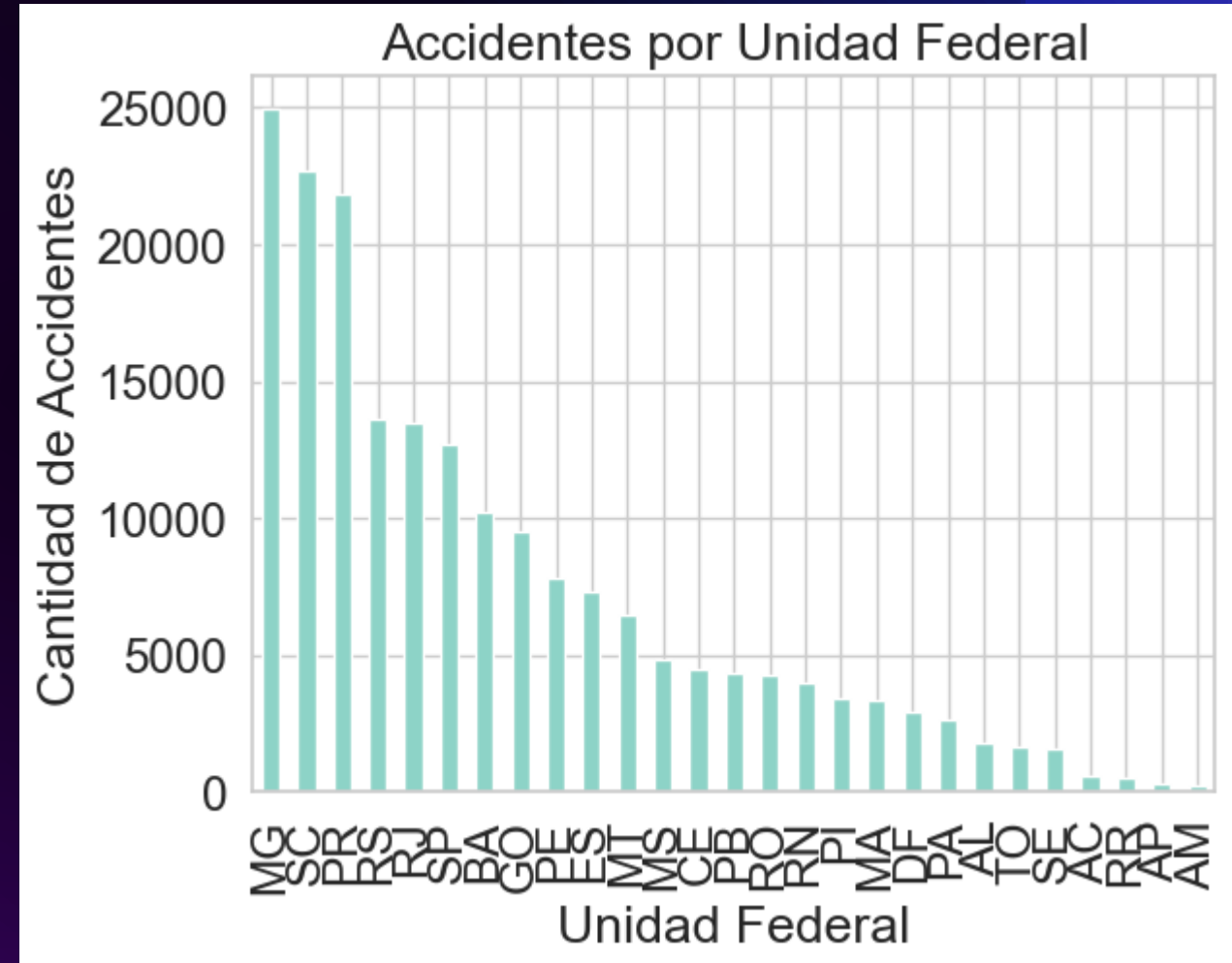


VISUALIZACION DE GRÁFICOS

Unidades Federales con mayores índices de accidentes.

Este informe refleja la distribución de los accidentes por Unidad Federal en Brasil, donde se puede observar que Minas Gerais (MG) tiene el mayor número de accidentes reportados con 24,951, seguido de cerca por Santa Catarina (SC) con 22,669 accidentes y Paraná (PR) con 21,818 accidentes.

En contraste, los estados con menos accidentes reportados son Roraima (RR) y Amazonas (AM), con 514 y 234 accidentes respectivamente.





ARBOL DE DECISIÓN

Arbol de decision

Vizualización de duplicados

```
unique_weapons = df['Classificacao_acidente'].unique()
print(unique_weapons)
```

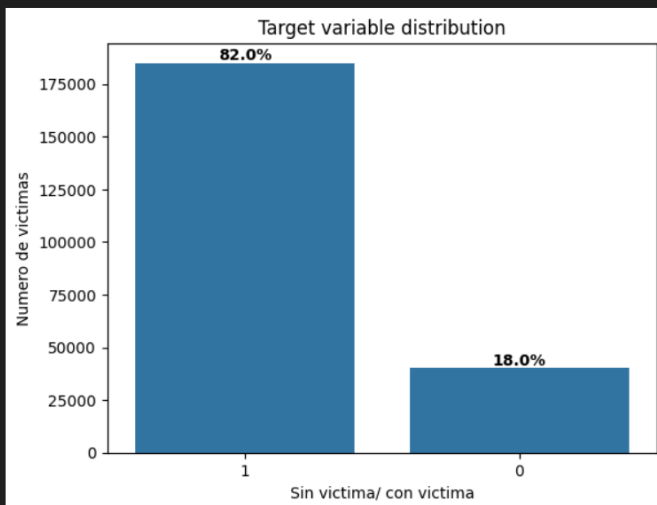
```
[54]
... ['Com Vítimas Feridas' 'Com Vítimas Fatais' 'Sem Vítimas']
```

```
# Voy a realizar una agrupación para que mis gráficos no sean tan grandes

vitima_mapping = {
    'Com Vítimas Feridas': '1',
    'Com Vítimas Fatais': '1',
    'Sem Vítimas': '0'
}

df['C_vitimas'] = df['Classificacao_acidente'].map(vitima_mapping) # Creo una nueva columna con áreas agrupadas
```

```
[55]
```

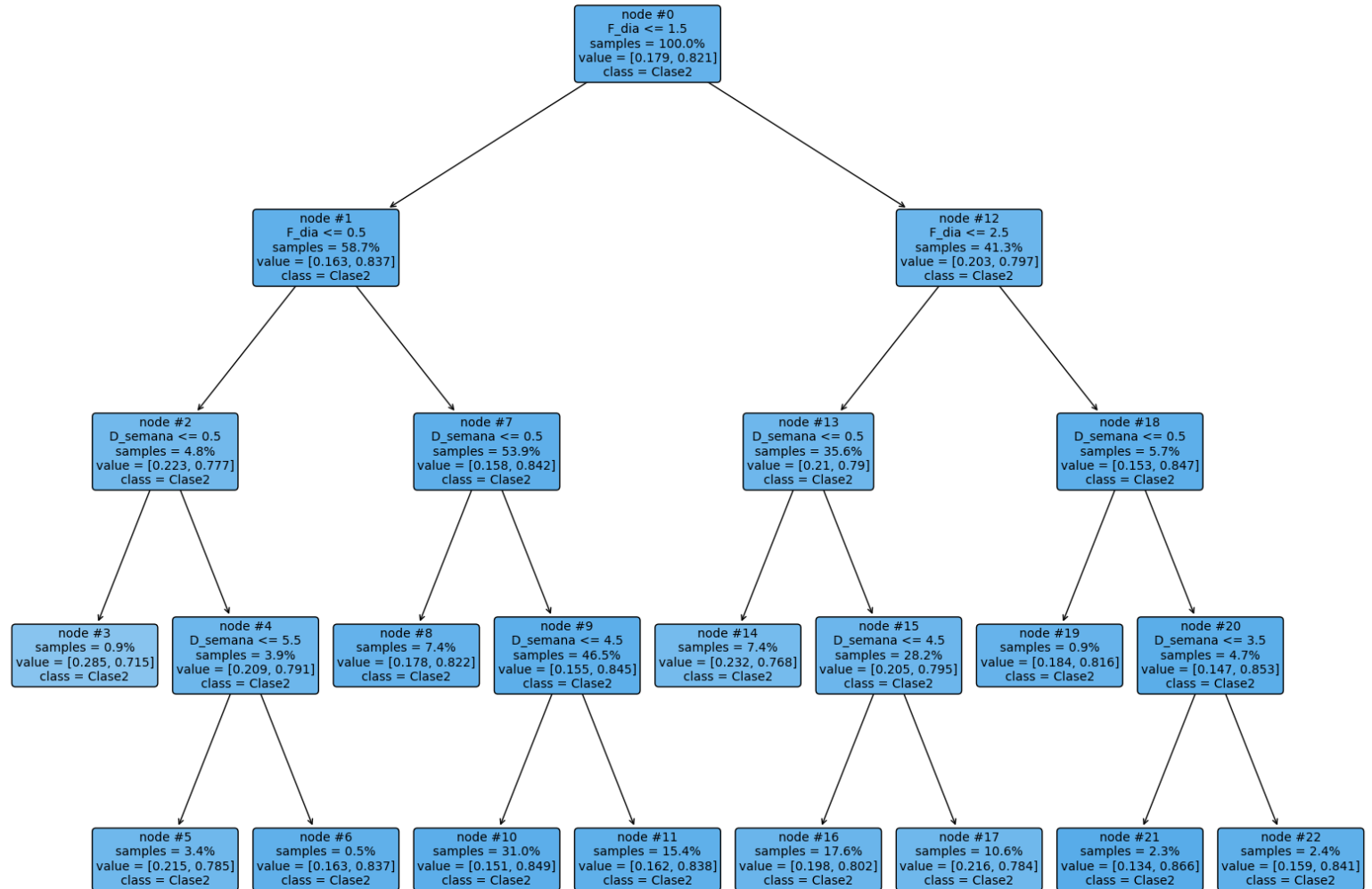


Se seleccionó como variable target
“Classificacao de acidente”

Se utilizará esta variable como variable target ya que en ella se observa que en cada accidente se producen el 82% de víctimas y sólo el 18% de los casos no poseen víctimas.

Arbol de decision

Convertí las variables categóricas 'Fase_dia' y 'Sentido_via' en numéricas mediante codificación, para que pudieran ser utilizadas por mi modelo. Además, creé la matriz de predictores X y el vector objetivo y para los siguientes pasos.



Arbol de decision

Generación de Predicciones

Usé el modelo entrenado `clf` para predecir dos cosas: la clase de accidente (con víctimas o sin víctimas) y el nodo del árbol al que pertenecen las observaciones basadas en las variables predictoras `F_dia` y `S_via`.

Resultados del Paso 1:

La predicción de la clase (`Predict_Arbol_Clase`) sugiere que la gran mayoría de las observaciones se clasifican en la clase 1. Esto puede indicar que según el modelo, la mayoría de los accidentes en el conjunto de datos tienen víctimas.

La predicción del nodo (`Predict_Arbol_Nodo`) muestra una distribución desigual de las observaciones a través de los nodos, con algunos nodos conteniendo mucha

```
[83] df['Predict_Arbol_Clase'].value_counts()

... Predict_Arbol_Clase
1      224500
0         602
Name: count, dtype: int64
```

```
Analisis de Distribución de la Clase Predicha

Conté la cantidad de observaciones en cada clase predicha para obtener una idea de cómo el modelo está clasificando los accidentes.

Resultados del Paso 2:

La mayoría de las observaciones se clasificaron en la clase 1 (224,500 observaciones), mientras que solo una pequeña porción en la clase 0 (602 observaciones).
+ Code + Markdown

[84] df['Predict_Arbol_Nodo'].value_counts()

... Predict_Arbol_Nodo
13      64564
12      56826
17      42706
16      36985
20       6770
10       5777
19       5770
9        5102
2         372
4         200
5          30
Name: count, dtype: int64
```

Arbol de decision

Análisi de las Predicciones

Análisis de Distribución por Nodo del Árbol

Conté la cantidad de observaciones que cada nodo del árbol decidió.

Resultados del Paso 3:

Algunos nodos, como el 13 y el 12, contienen un gran número de observaciones, lo que podría indicar caminos comunes en la toma de decisiones del árbol. Esto podría reflejar patrones comunes o prominentes en los datos.

▶

```
# Groupby por prediccion de clase:
resultados = df.groupby('Predict_Arbol_Clase').agg(
    Cant = ('Predict_Arbol_Clase', 'count'),
    Cant_Victimas = ('C_vitimas', 'sum'),
    Tasa_Victimas = ('C_vitimas', 'mean')
).reset_index()
resultados
```

[85]

...

	Predict_Arbol_Clase	Cant	Cant_Victimas	Tasa_Victimas
0	0	602	251	0.416944
1	1	224500	184761	0.822989

Agregación por Clase Predicha

Realicé una agregación por clase predicha para entender mejor el número total de accidentes y víctimas junto con la tasa promedio de víctimas por clase.

Resultados del Paso 4:

En la clase 0, hay 602 accidentes con 251 víctimas, lo que resulta en una tasa media de víctimas de 0.416944. En la clase 1, hay 224,500 accidentes con 184,761 víctimas, lo que resulta en una tasa media de víctimas de 0.822989.

```
# Groupby por prediccion de nodo:
resultados_nodo = df.groupby('Predict_Arbol_Nodo').agg(
    Cant = ('Predict_Arbol_Nodo', 'count'),
    Cant_Victimas = ('C_vitimas', 'sum'),
    Tasa_Victimas = ('C_vitimas', 'mean')
).reset_index()
resultados_nodo.rename(columns={'Tasa_victimas': 'Predict_Prob_Arbol'}, inplace=True)
resultados_nodo
```

[86]

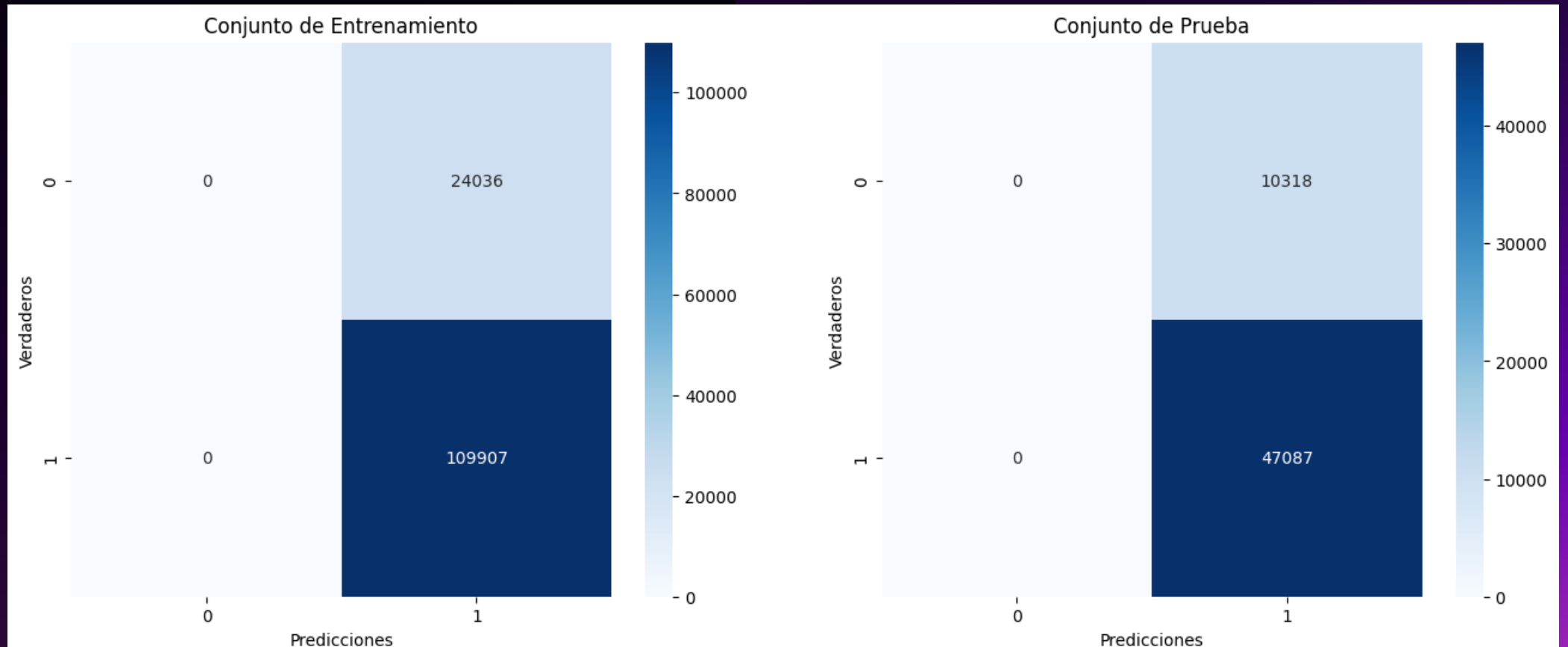
Python

...

	Predict_Arbol_Nodo	Cant	Cant_Victimas	Tasa_Victimas
0	2	372	149	0.400538
1	4	200	90	0.450000
2	5	30	12	0.400000
3	9	5102	3996	0.783222

Matriz de confusión para el conjunto de entrenamiento

En el entrenamiento, el modelo tuvo una precisión del 82% y en la prueba, anduvo cerca del 82%. Aunque parece que está acertando bastante, está claro que tiende a pasar por alto algunos casos positivos, razón por el cuál el modelo debe ser mejorado.

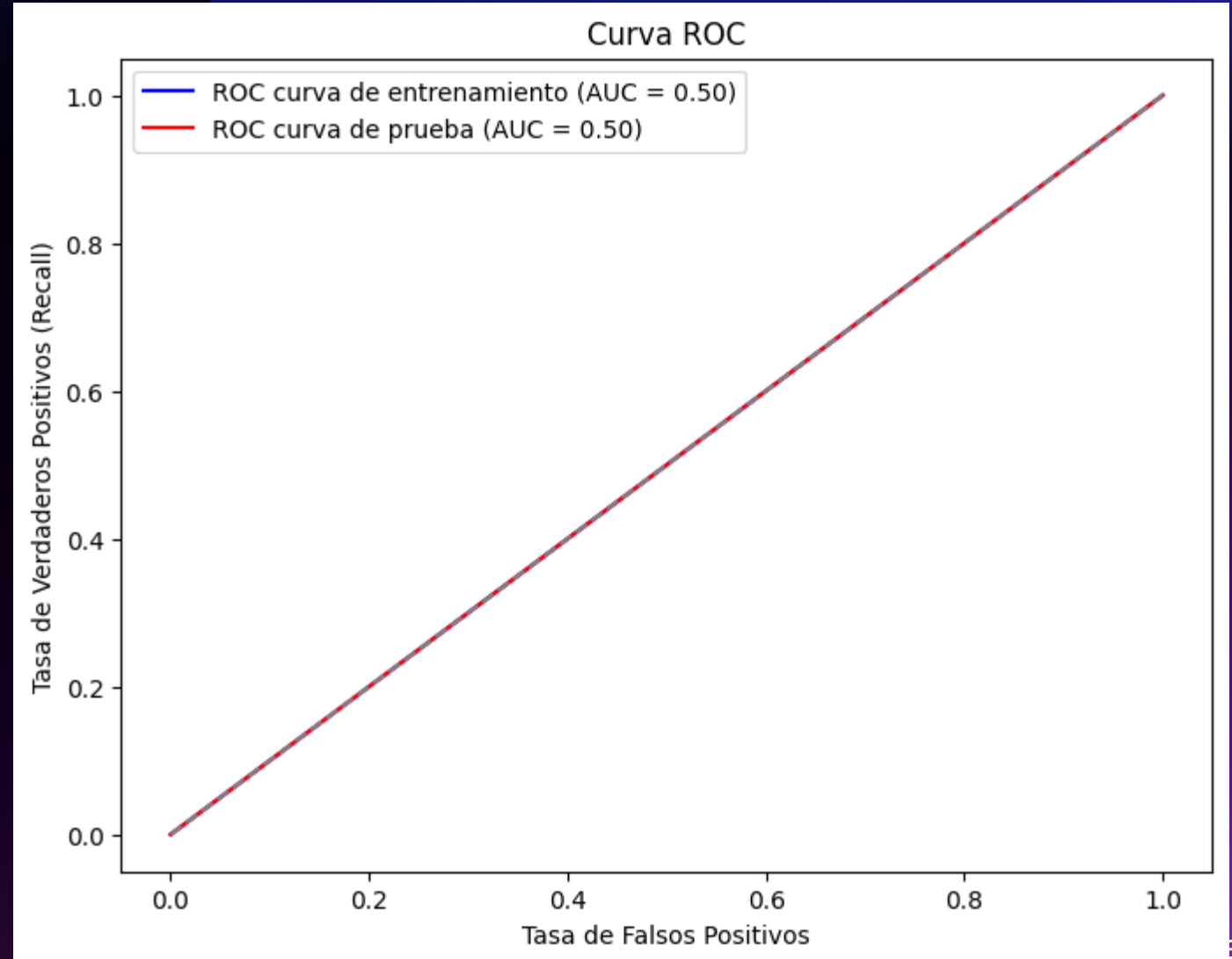




MODELO 2: RANDOM FOREST

Curva ROC

El modelo de Random Forest muestra una alta precisión del 82%, pero falla en detectar la clase negativa, indicando un sesgo hacia la clase positiva. La curva ROC, con un AUC de 0.56, sugiere que el modelo apenas supera el rendimiento aleatorio. Para mejorar, es necesario abordar el desequilibrio de las clases.





OPTIMIZACIÓN

Matriz de confusión para el conjunto de entrenamiento

Luego de utilizar el modelo con
Hiperparámetros se puede observar que el
modelo con mayor accuracy fue el 2 - Random
Forest con Accuracy: 0.82

```
• # Entreno el modelo utilizando la búsqueda en cuadrícula  
grid_search.fit(X_train, y_train)
```

```
▸ GridSearchCV  
▸ estimator: RandomForestClassifier  
  ▸ RandomForestClassifier
```

CONCLUSION

A través del análisis detallado, he llegado a varias conclusiones clave que nos ayudan a comprender los accidentes desde una perspectiva más amplia:

Patrones de Accidentes: Los nodos identificados por el modelo de árbol de decisiones destacan ciertas características que podrían ser útiles para prevenir futuros accidentes.

Constancia en la Ocurrencia: A diferencia de lo que se podría esperar, los accidentes son bastante uniformes a lo largo del año, independientemente de las estaciones o los meses. Esta constancia nos desafía a mantener medidas de seguridad vial consistentes y proactivas todo el año.

Condiciones Climáticas: A pesar de que la lluvia y el clima despejado predominan durante los accidentes, la mayoría ocurre en condiciones normales, lo que puede indicar otros factores en juego como el volumen de tráfico o el comportamiento humano al volante.

Variabilidad Semanal: Los fines de semana muestran un aumento en los accidentes, lo que sugiere que actividades recreativas y viajes no laborales influyen significativamente en los riesgos de tráfico.

CONCLUSION

Tipos de Carretera: Las carreteras simples y dobles son más propensas a los accidentes, lo que indica una posible falta de infraestructura de seguridad o de medidas preventivas efectivas.

Fluctuaciones Temporales: Los picos durante ciertos periodos pueden estar ligados a vacaciones y fechas específicas, lo que apunta a la importancia de la gestión y planificación del tráfico en estos tiempos.

Distribución Horaria: El incremento de accidentes en horas de la tarde y noche nos recuerda que factores como la visibilidad y el tráfico juegan un rol crítico en la seguridad vial.

Diferencias Regionales: La prevalencia de accidentes en ciertas Unidades Federales subraya la necesidad de políticas adaptadas a cada región, considerando las condiciones y necesidades locales.

En conclusión, este análisis me ha provisto de valiosas lecciones sobre la seguridad vial y la importancia de una política pública bien informada. La combinación de técnicas estadísticas y modelos predictivos con una interpretación cuidadosa de los resultados puede iluminar el camino hacia carreteras más seguras en Brasil.

GRACIAS

