



Accidente de trabajo, Aurelio Arteta Errasti.

Praevidens.

**Predicción con IA
del Riesgo Laboral
para una Gestión
Proactiva**

Anticipando Riesgos para
Proteger Vidas y
Optimizar Recursos.



Andrea Gutiérrez González

Ingeniera Industrial con experiencia en
diversas áreas del sector real.

Praevidens.



Un campo de arroz italiano, Ángelo Morbelli.

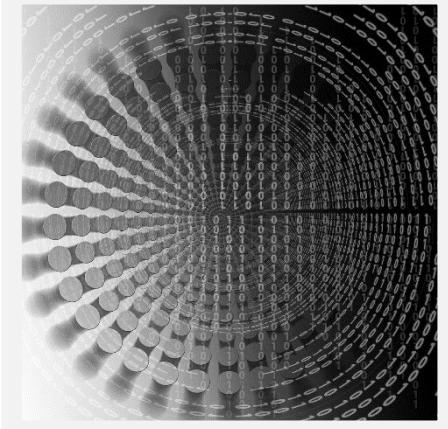
¿Quiénes Somos?

Problema e Importancia

Prevención basada en datos para entornos laborales
más seguros.

El fin de la jornada laboral, Jules Breton.

Antecedentes



Seguridad

La seguridad y salud en el trabajo en USA está regulada principalmente por la **Occupational Safety and Health Administration (OSHA)**.

Desafío

A pesar de décadas de esfuerzos, **Accidentes y enfermedades laborales** siguen representando un **desafío significativo** en diversas industrias.

Incidentes

2023 se registraron **2.6 Millones** de lesiones y enfermedades laborales en el sector privado, un **8,4%** menos que en **2022**. **2,4 casos** por cada **100 trabajadores** a tiempo completo.

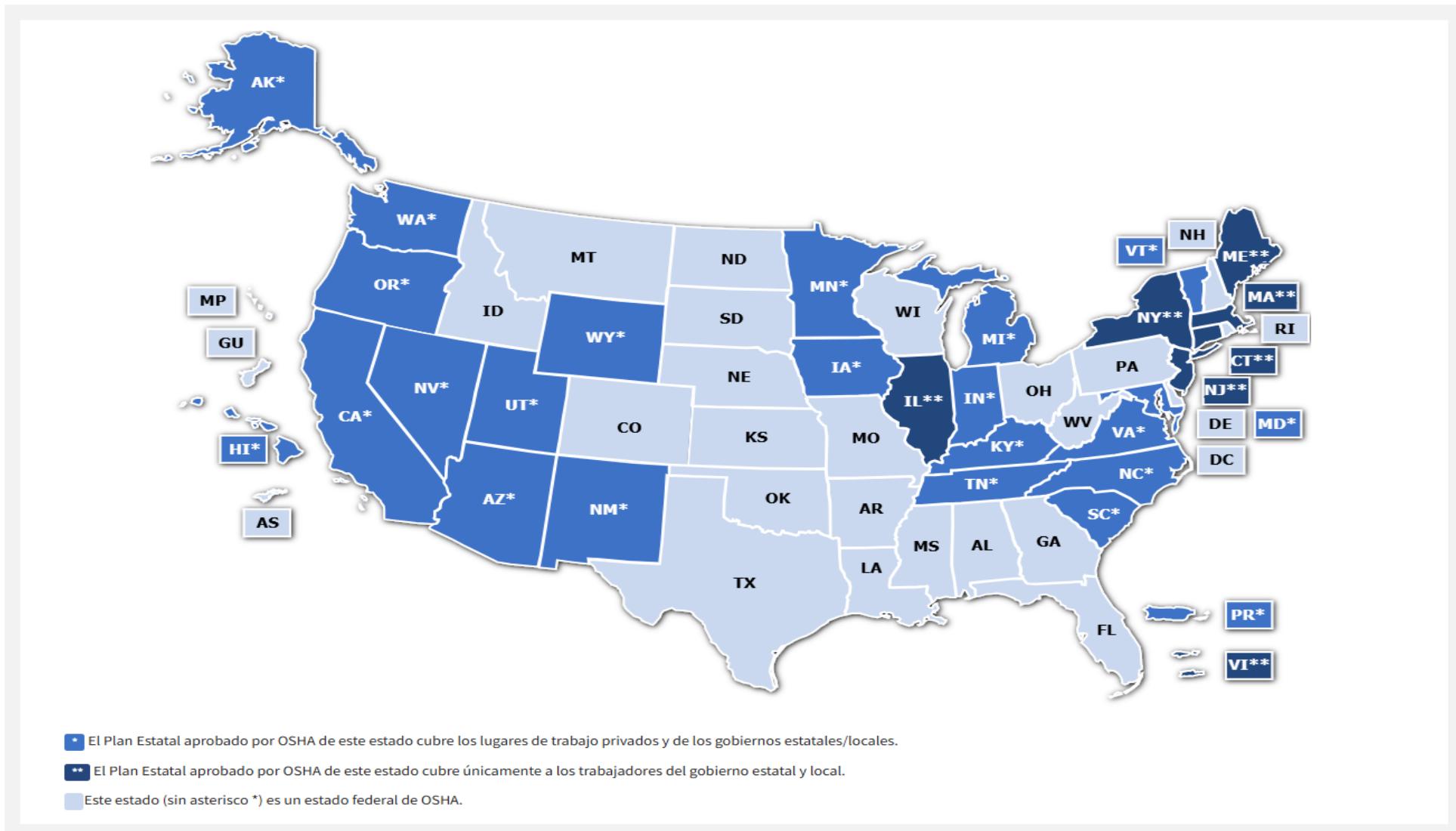
Muertes

5.283 muertes laborales en **2023**, lo que representa una disminución del **3,7%** respecto a las **5.486** de **2022**.

Datos

La existencia de registros brinda una oportunidad para aplicar técnicas de ML que permitan **anticipar lesiones y enfermedades**.

Cobertura OSHA



Aplica al **sector privado** y en algunos estados también al **sector público**.

No cubre autónomos ni sectores regulados por otras agencias como **Minería (MSHA)**, **Transporte Aéreo (FAA)**, **Ferroviario (FRA)**, **Industria Nuclear (NRC)**.

Presente en todos los **estados (50) y territorios** con normativa federal o estatal equivalente.



Problema

Anticipando riesgos para proteger vidas y optimizar recursos.

En el ámbito de la **salud y seguridad ocupacional**, la **capacidad de anticipar** el número de lesiones y enfermedades laborales que pueden ocurrir en una compañía es fundamental para **implementar estrategias** de prevención efectivas y **optimizar** la asignación de recursos.

Se depende principalmente de **datos históricos y reportes reactivos** para gestionar los riesgos asociados al bienestar de los trabajadores.

La ausencia de un mecanismo predictivo limita la proactividad en la identificación de establecimientos **con alto riesgo de incidentes**, lo que puede derivar en consecuencias graves como aumento de accidentes, bajas laborales prolongadas, costos elevados por indemnizaciones y deterioro de la reputación corporativa.

Desarrollar un **modelo ML** capaz de **predecir con precisión el índice y nivel de riesgo** en función de los casos reportados y **características relevantes** como la industria, tamaño, ubicación, etc.

Con ello se busca **habilitar una gestión predictiva del riesgo**, permitiendo anticipar incidentes potenciales y tomar acciones preventivas oportunas.

Su **importancia** radica en el potencial para:

- **Reducir ocurrencia** de accidentes y enfermedades laborales mediante intervenciones focalizadas.
- **Mejorar eficiencia** en asignación de recursos para inspección y capacitación en seguridad.
- Facilitar **diseño de políticas** basadas en evidencia cuantitativa.
- **Optimizar costos** asociados a siniestralidad.



Objetivo

Pronosticando el índice y el nivel de riesgo para tomar decisiones preventivas inteligentes.

Datos: Análisis y Preparación

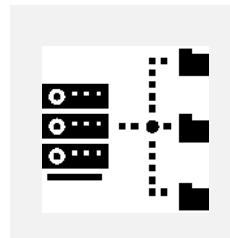
Prevención basada en datos para entornos laborales
más seguros.

El fin de la jornada laboral, Jules Breton.



Datos

Descubriendo patrones que impulsan decisiones.



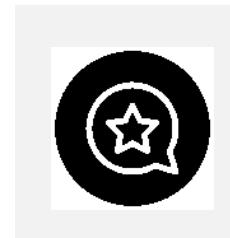
Dataset

Datos Resumidos
de 2024



Instancias

389.874



Features

35

Features

```
RangeIndex: 389874 entries, 0 to 389873
Data columns (total 35 columns):
 #   Column          Non-Null Count  Dtype  
 ---  -- 
 0   id              389874 non-null   object  
 1   establishment_name 389874 non-null   object  
 2   establishment_id   389874 non-null   int64  
 3   ein              347436 non-null   object  
 4   company_name     370747 non-null   object  
 5   street_address   389872 non-null   object  
 6   city             389872 non-null   object  
 7   state            389874 non-null   object  
 8   zip_code         389874 non-null   int64  
 9   naics_code       389874 non-null   int64  
 10  naics_year       389874 non-null   int64  
 11  industry_description 363923 non-null   object  
 12  establishment_type 389324 non-null   float64 
 13  size             389874 non-null   int64  
 14  annual_average_employees 389874 non-null   int64  
 15  total_hours_worked 389874 non-null   int64  
 16  no_injuries_illnesses 389874 non-null   int64  
 17  total_deaths      389874 non-null   int64  
 18  total_dafw_cases  389874 non-null   int64  
 19  total_djtr_cases  389874 non-null   int64  
 20  total_other_cases 389874 non-null   int64  
 21  total_dafw_days   389874 non-null   int64  
 22  total_djtr_days   389874 non-null   int64  
 23  total_injuries    389874 non-null   int64  
 24  total_skin_disorders 389874 non-null   int64  
 25  total_respiratory_conditions 389874 non-null   int64  
 26  total_poisonings  389874 non-null   int64  
 27  total_hearing_loss 389874 non-null   int64  
 28  total_other_illnesses 389874 non-null   int64  
 29  created_timestamp 389874 non-null   object  
 30  change_reason     13687 non-null    object  
 31  year_filing_for  389816 non-null   float64 
 32  sector            389816 non-null   object  
 33  zipcode           389816 non-null   float64 
 34  naics_char        389816 non-null   float64 
dtypes: float64(4), int64(20), object(11)
```

Descripción Features más relevantes:

ein = Identificación Fiscal del empleador.

naics_code = Código del Sistema de Clasificación Industrial de América del Norte (SCIAN) que clasifica el negocio.

naics_year = Año calendario de la versión de los códigos NAICS.

establishment_type = Privado o Gobierno estatal o local.

- 1 No es una entidad gubernamental.
- 2 Entidad del Gobierno Estatal.
- 3 Entidad del Gobierno Local.

no_injuries_illnesses =

- 1 si tuvo lesiones o enfermedades.
- 2 si NO tuvo lesiones o enfermedades.

annual_average_employees = Número promedio anual de empleados en el establecimiento.

size = Tamaño del establecimiento según número máximo de empleados que trabajaron allí durante el año.

- 1 < 20 empleados.
- 2 entre 20 y 249 empleados.
- 21 de 20 a 99.
- 22 de 100 a 249.
- 3 >= 250 o más empleados.

total_hours_worked = Total de horas trabajadas por todos los empleados.

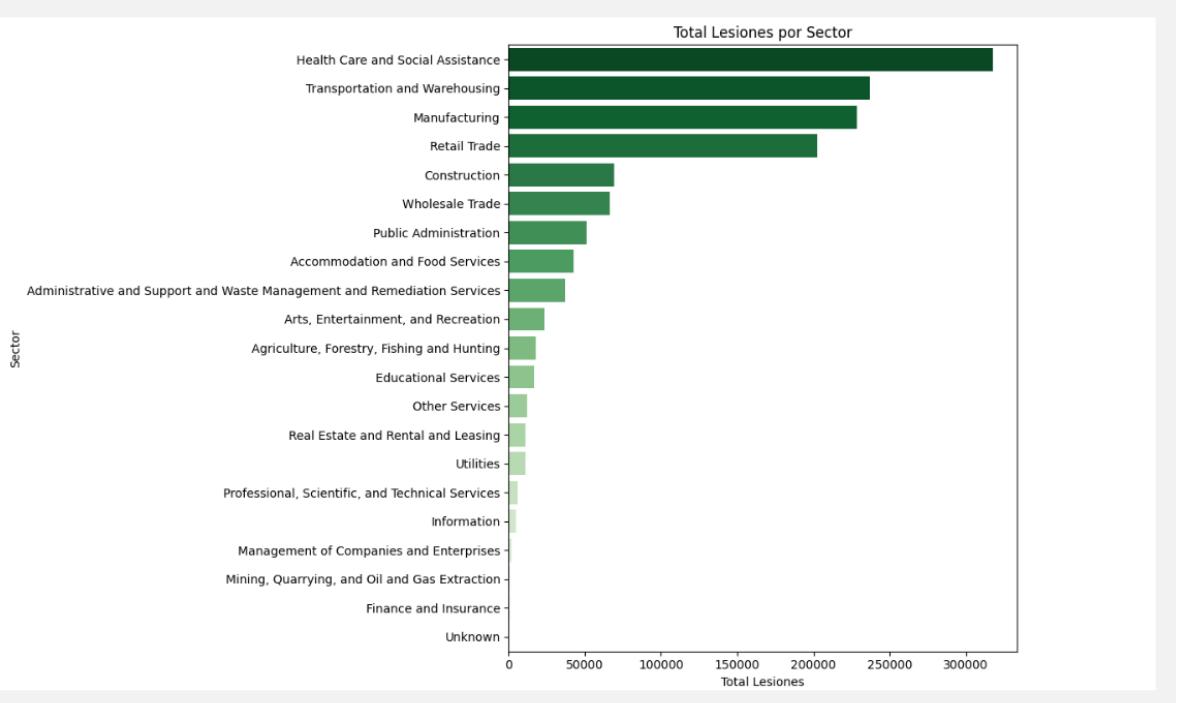
total_deaths = Número total de fallecimientos.

total_dafw_cases = Número total de casos con días de ausencia del trabajo.

total_djtr_cases = Número total de casos con transferencia o restricción de trabajo.

total_injuries = Número total de lesiones.

Muertes y Lesiones por Sector

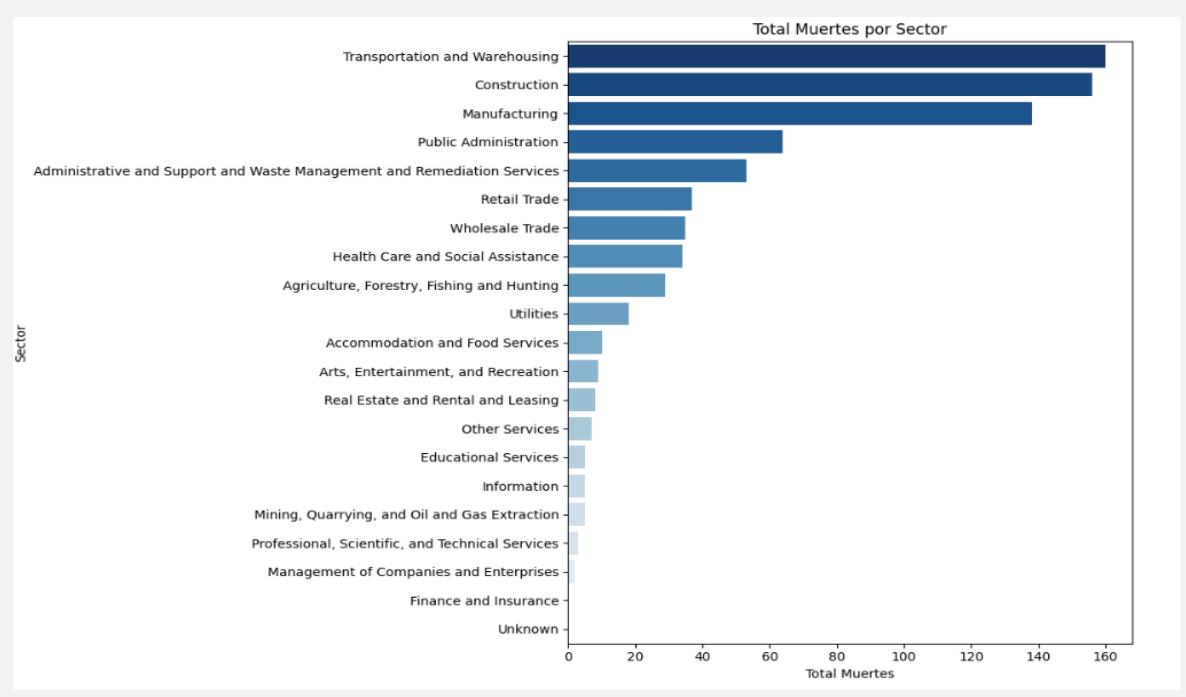


Top 3 Muertes

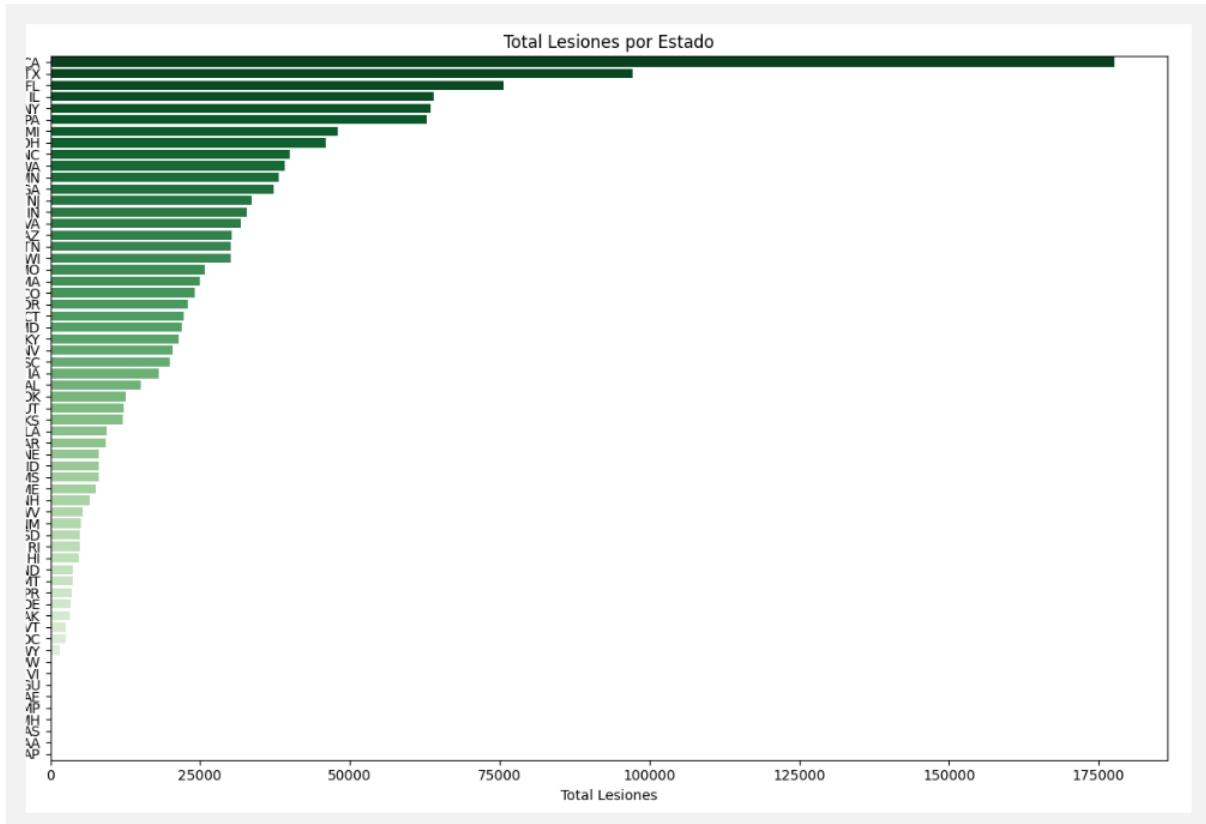
- **Transporte**
- **Construcción**
- **Manufactura**

Top 3 Lesiones

- **Salud**
- **Transporte**
- **Manufactura**

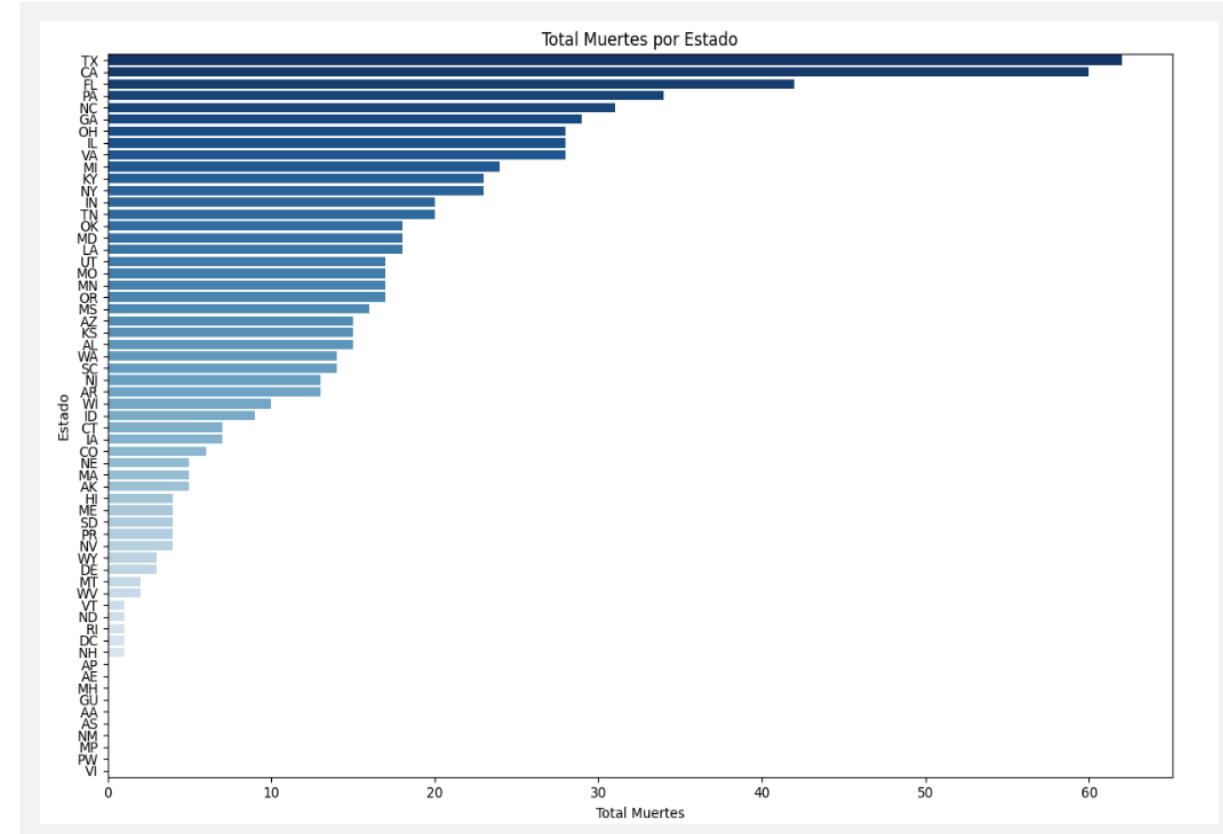


Muertes y Lesiones por Estado

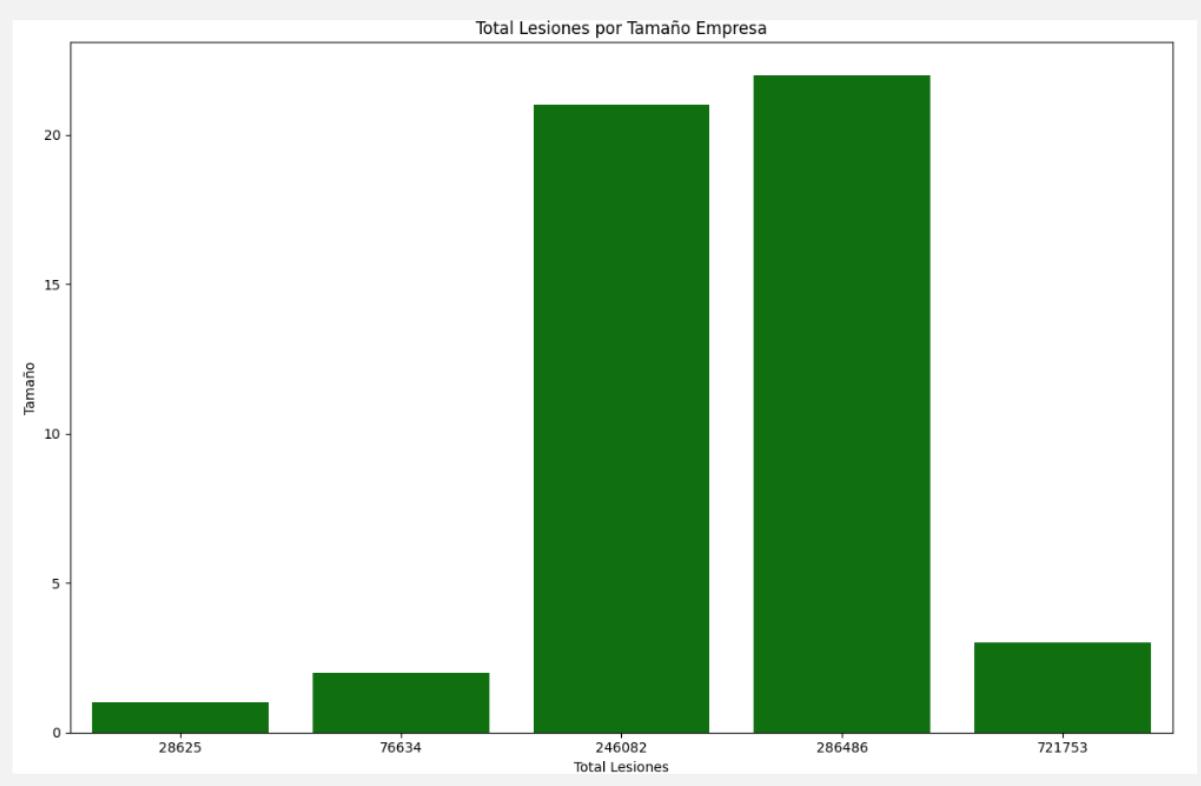


Top 3 Muertes

- Texas
- California
- Florida



Muertes y Lesiones por Tamaño

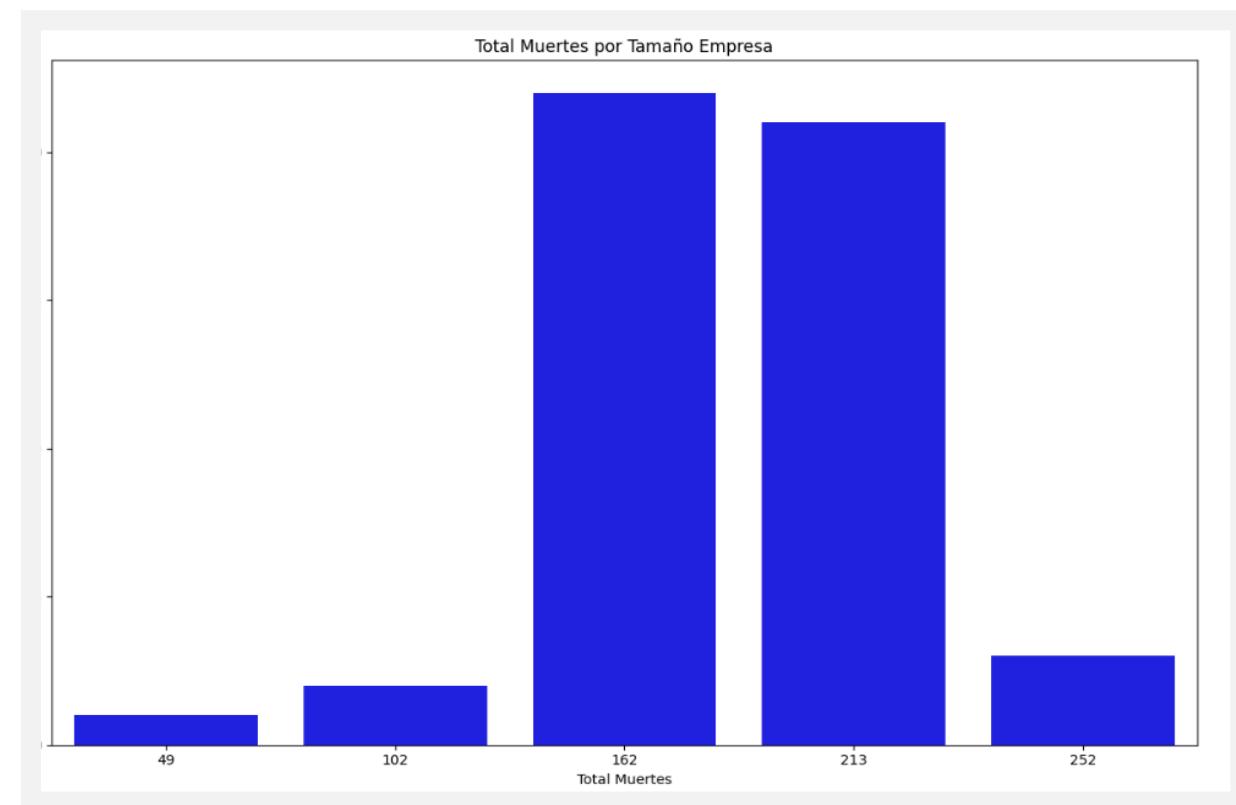


Top 3 Muertes

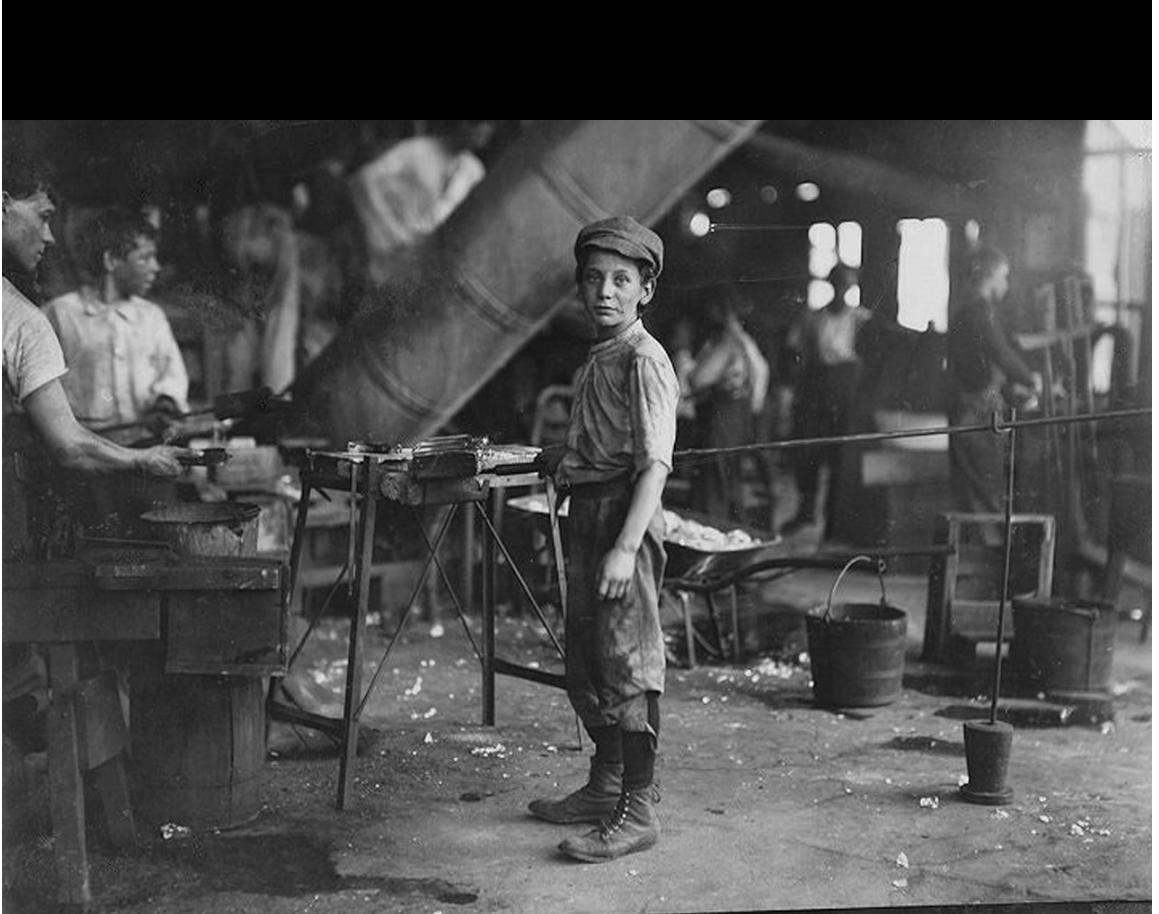
- **20 a 99**
- **100 a 249**
- **250 o más**

Top 3 Lesiones

- **100 a 249**
- **20 a 99**
- **250 o más**



- No hay **instancias duplicadas**.
- No hay **valores sucios**.
- **Eliminación** de features **poco relevantes** o redundantes.
- Conversión de **ceros** a valores nulos (**NaN**), en features requeridos para cálculos**.
- Transformación de **variables numéricas** a categóricas (size)*.
- 6 Features con **valores nulos**:
 - city = **2**
 - industry_description = **25951**
 - establishment_type = **550***
 - annual_average_employees = **1635****
 - total_hours_worked = **1571****
 - sector = **58**



Preparación

Convertiendo datos crudos en información confiable.

Dataset Limpio

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 389874 entries, 0 to 389873
Data columns (total 23 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   city             389872 non-null   object  
 1   state            389874 non-null   object  
 2   zip_code          389874 non-null   int64  
 3   naics_code        389874 non-null   int64  
 4   industry_description 363923 non-null   object  
 5   establishment_type 389324 non-null   object  
 6   size              389874 non-null   object  
 7   annual_average_employees 389874 non-null   int64  
 8   total_hours_worked 389874 non-null   int64  
 9   no_injuries_illnesses 389874 non-null   int64  
 10  total_deaths       389874 non-null   int64  
 11  total_dafw_cases   389874 non-null   int64  
 12  total_djtr_cases   389874 non-null   int64  
 13  total_other_cases  389874 non-null   int64  
 14  total_dafw_days    389874 non-null   int64  
 15  total_djtr_days    389874 non-null   int64  
 16  total_injuries      389874 non-null   int64  
 17  total_skin_disorders 389874 non-null   int64  
 18  total_respiratory_conditions 389874 non-null   int64  
 19  total_poisonings     389874 non-null   int64  
 20  total_hearing_loss    389874 non-null   int64  
 21  total_other_illnesses 389874 non-null   int64  
 22  sector             389816 non-null   object  
dtypes: int64(17), object(6)
memory usage: 68.4+ MB
```



Instancias

389.874



Features

23

▶ Dataset_Limpio.shape
→ (389874, 23)

Técnicas y Modelos Usados

Prevención basada en datos para entornos laborales más seguros.

El fin de la jornada laboral, Jules Breton.

Imputación

	0
city	2
state	0
zip_code	0
naics_code	0
industry_description	25951
establishment_type	550
size	0
annual_average_employees	1635
total_hours_worked	1571
no_injuries_illnesses	0
total_deaths	0
total_dafw_cases	0
total_djtr_cases	0
total_other_cases	0
total_dafw_days	0
total_djtr_days	0
total_injuries	0
total_skin_disorders	0
total_respiratory_conditions	0
total_poisonings	0
total_hearing_loss	0
total_other_illnesses	0
sector	58

dtype: int64



establishment_type, city, tipo objeto
imputación con la **Moda**.



sector imputación con la **Moda basada**
en naics_code.



annual_average_employees,
total_hours_worked imputación con
Moda basada en 'naics_code', 'size',
'sector'.



industry_description imputación con
KNN.

	0
city	0
state	0
zip_code	0
naics_code	0
industry_description	0
establishment_type	0
size	0
annual_average_employees	0
total_hours_worked	0
no_injuries_illnesses	0
total_deaths	0
total_dafw_cases	0
total_djtr_cases	0
total_other_cases	0
total_dafw_days	0
total_djtr_days	0
total_injuries	0
total_skin_disorders	0
total_respiratory_conditions	0
total_poisonings	0
total_hearing_loss	0
total_other_illnesses	0
sector	0

dtype: int64

Risk_index = (total_injuries + total_deaths * Peso) * 200000) /
total_hours_worked

Risk_level = Percentiles q25, q50, q75, q95

**total_deaths + total_dafw_cases + total_djtr_cases +
total_other_cases = total_illnesses + total_injuries**

```
Percentiles y Valores risk_index
0.00    0.000000
0.25    0.000000
0.50    1.459082
0.75    5.355039
0.95    16.549250
Name: risk_index, dtype: float64

Categorías e Instancias risk_level
risk_level
Safe      169599
High      97470
Critical   77973
Moderate   25338
Severe     19494
Name: count, dtype: int64
```



Cálculos Riesgo

Cuantificando el riesgo antes de que se convierta en daño.

Mapa de Correlación

total_illnesses tiene correlación con:

- total_other_illnesses (0.94)
- total_other_cases (0.59)

total_injuries tiene correlación con:

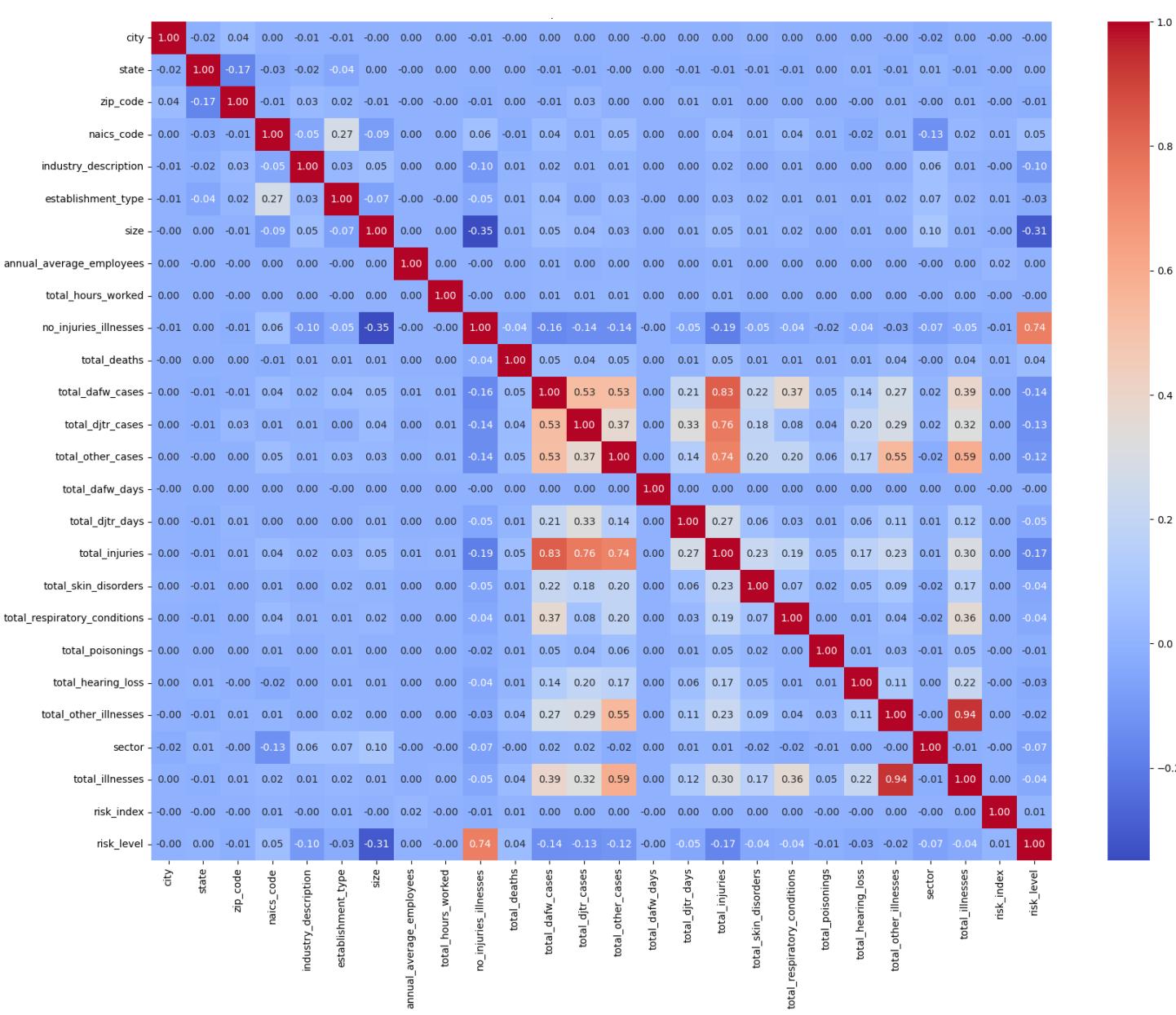
- total_dafw_cases (0.83)
- total_djtr_cases (0.76)
- total_other_cases (0.74)

risk_level tiene correlación con:

- No_injuries_illnesses (0.74)

size tiene correlación negativa con:

- no_injuries_ilnesses (-0.35)
- Risk_level (-0.31)





Datos Predicción

Riesgo bajo control con datos y predicción.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 389874 entries, 0 to 389873
Data columns (total 26 columns):
 #   Column           Non-Null Count   Dtype  
--- 
 0   city             389874 non-null    object  
 1   state            389874 non-null    object  
 2   zip_code          389874 non-null    int64   
 3   naics_code        389874 non-null    int64   
 4   industry_description 389874 non-null    object  
 5   establishment_type 389874 non-null    float64 
 6   size              389874 non-null    object  
 7   annual_average_employees 389874 non-null    float64 
 8   total_hours_worked 389874 non-null    float64 
 9   no_injuries_illnesses 389874 non-null    int64   
 10  total_deaths      389874 non-null    int64   
 11  total_dafw_cases  389874 non-null    int64   
 12  total_djtr_cases  389874 non-null    int64   
 13  total_other_cases 389874 non-null    int64   
 14  total_dafw_days   389874 non-null    int64   
 15  total_djtr_days   389874 non-null    int64   
 16  total_injuries    389874 non-null    int64   
 17  total_skin_disorders 389874 non-null    int64   
 18  total_respiratory_conditions 389874 non-null    int64   
 19  total_poisonings   389874 non-null    int64   
 20  total_hearing_loss  389874 non-null    int64   
 21  total_other_illnesses 389874 non-null    int64   
 22  sector            389874 non-null    object  
 23  total_illnesses    389874 non-null    int64   
 24  risk_index         389874 non-null    float64 
 25  risk_level         389874 non-null    object  
dtypes: float64(4), int64(16), object(6)
memory usage: 77.3+ MB
```

Dataset_Index.shape

(389874, 26)

```
RandomForestRegressor
RandomForestRegressor(random_state=27)
```

Random Forrest Regressor: risk_index

- **Target** = risk_index
- **Variables** = annual_average_employees, total_hours_worked, total_deaths, total_dafw_cases, total_djtr_cases, total_other_cases, total_illnesses
- **Variables de Contexto** = size, state, sector

Random Forrest Classifier: risk_level

- **Target** = risk_level
- **Variables** = annual_average_employees, total_hours_worked, total_deaths, total_dafw_cases, total_djtr_cases, total_other_cases, total_illnesses
- **Variables de Contexto** = size, state, sector

```
RandomForestClassifier
RandomForestClassifier(random_state=27)
```



Modelos

Random Forest combina múltiples árboles para brindar una visión robusta del riesgo.

Predicciones y Rendimiento

Prevención basada en datos para entornos laborales
más seguros.

El fin de la jornada laboral, Jules Breton.

Random Forrest Regressor

Target: risk_index

	RandomForest_Etrenamiento	RandomForest_Prueta
R2	0.93	0.91
mae	1.99	1.78
mse	88617.59	22346.96
rmse	297.69	149.49

Importancia de Features:	
total_dafw_cases	0.493706
total_other_cases	0.241145
total_hours_worked	0.060807
total_djtr_cases	0.054901
annual_average_employees	0.048648
sector	0.034359
state	0.027495
size	0.025367
total_illnesses	0.013489
total_deaths	0.000084
dtype:	float64

Conclusión

Excelente rendimiento (R2). Bajo sobreajuste (pequeña diferencia R2 prueba vs. entrenamiento) El **risk_index** con **total_hours_worked** es una **métrica sólida** y el modelo aprende a predecirlo con mucha **precisión**.

R2 Coeficiente de Determinación

El modelo explica el **91%** de la variabilidad de los datos nuevos, esto indica que el modelo logra predecir muy bien. Explicando el 91% de la varianza del **risk_index**. Esto indica que tiene muy **buenas capacidades de generalización** y un muy **buen rendimiento**.

En los datos de entrenamiento explica el **93%** lo cual es muy bueno para el conjunto de entrenamiento.

MAE Error Absoluto Medio

1.78 indica que las predicciones están a 1.78 unidades del valor real. Dado que el MAE es bajo tanto en entrenamiento como en prueba se infiere que el modelo tiene un **buen ajuste**.

MSE Error Cuadrático

22,346 en prueba y **88,617**, el MSE en prueba es mas bajo lo que indica **mejor desempeño** y que el modelo **no** esta **sobre-ajustado**.

RMSE Raíz Cuadrada de MSE

149 en prueba significa que en promedio el modelo se equivoca en ± 149 unidades. En entrenamiento es de **297**. En el contexto de un R2 alto, estos valores indican que los **errores de predicción** son relativamente **pequeños** y consistentes.

Random Forrest Classifier

Target: risk_level

	RandomForest_Etrenamiento	RandomForest_Prueba
Accuracy	1.0	0.98
Precision	1.0	0.98
Recall	1.0	0.97
F1-score	1.0	0.97

Importancia de Features:	
total_dafw_cases	0.247574
total_other_cases	0.226717
total_djtr_cases	0.196991
total_hours_worked	0.183439
annual_average_employees	0.068920
total_illnesses	0.024765
size	0.021505
state	0.013990
sector	0.012865
total_deaths	0.003233
dtype: float64	

Conclusión

El modelo **aprendió muy bien** en los datos de entrenamiento.

Muy **buen desempeño** en datos nuevos, la ligera caída es normal y mínima, muestra **buena generalización**.

La diferencia entre **100% vs 97% o 98%** es pequeña pero podría indicar que el modelo estaría **levemente sobreajustado** en entrenamiento; si es así el sobre ajuste es muy leve y no está afectando significativamente la generalización del modelo. Las **métricas** de prueba son **excepcionalmente buenas**.

Accuracy (Exactitud)

El modelo en prueba tiene un **98%** lo que significa que el modelo **es muy bueno para clasificar** los nuevos datos para **risk_level** y en entrenamiento un **100%** lo que indica que no hay error.

Precision (Precisión)

98% en prueba lo que indica que hay **pocos falsos positivos** y un buen manejo del error. en Entrenamiento el **100%** sin falsos positivos.

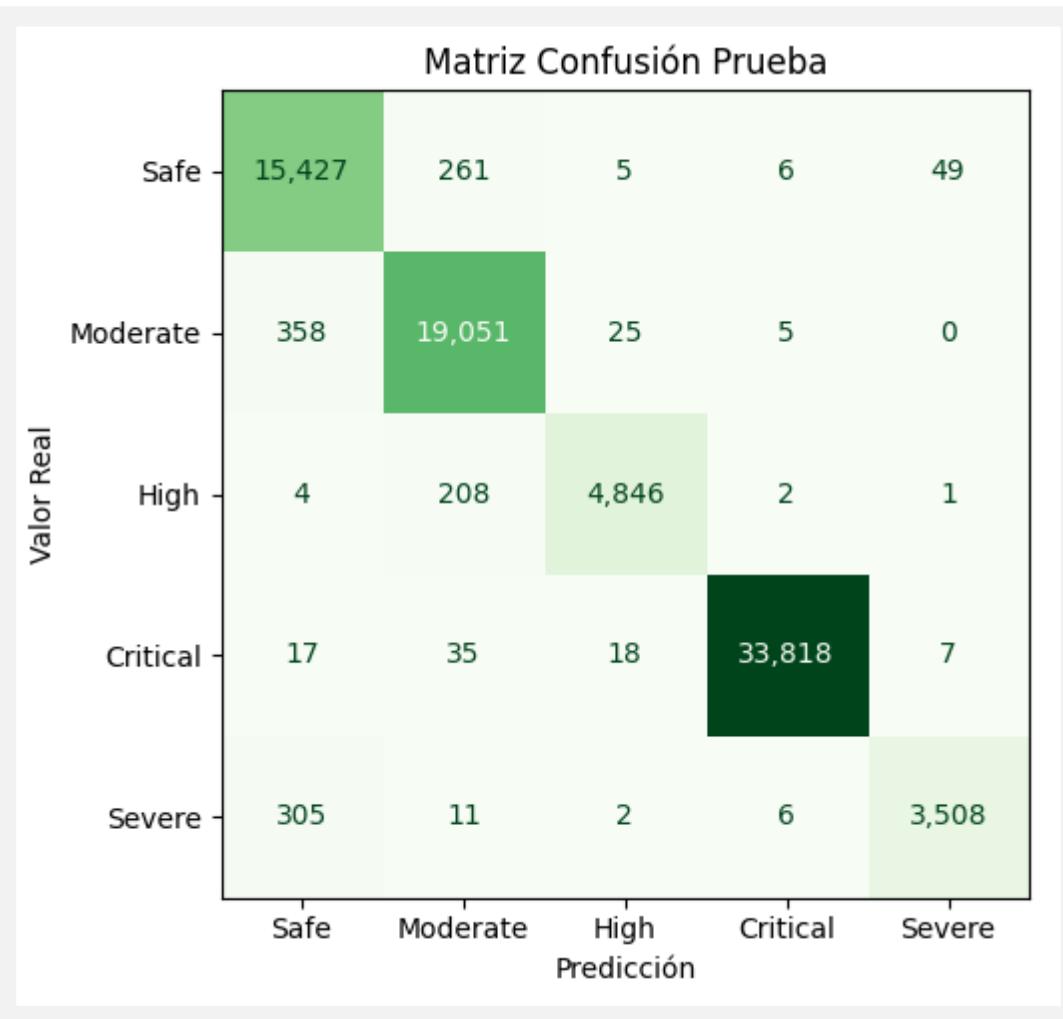
Recall (Sensibilidad)

97% en prueba y **100%** en entrenamiento lo cual indica que hay una muy **buena captura de casos**. El modelo es capaz de identificar el 97% de los casos reales de cada nivel de riesgo.

F1 Score

97% en prueba lo que indica que hay un muy **buen equilibrio entre precisión y recuperación**. Con el **100%** en entrenamiento el equilibrio es perfecto.

¿Como Mejorar el Modelo?



Target: risk_level

Próximos Pasos

Reducir Falsos positivos:

- **Balancear clases** aplicando métodos smote + tomek links para equilibrar el risk_level, mejorar la separación entre clases y eliminar muestras ruidosas.
- Aplicar **Análisis de Componentes Principales** (PCA) para reducir la dimensionalidad y transformar el conjunto de features.
- **Optimizar hiper-parametros** (número de árboles, profundidad del random Forrest) para mejorar el desempeño de los datos de prueba.
- Realizar **nuevos experimentos** con otros modelos.
- Realizar **análisis de las clases** mal predichas para entender que podría fallar en el modelo y así ajustar los features o crear nuevos features derivados.

Resultados Obtenidos

Prevención basada en datos para entornos laborales más seguros.

El fin de la jornada laboral, Jules Breton.



Conclusiones

El potencial se multiplica cuando el modelo se integra con datos mejorados y actualizados.

Enriquecimiento del Dataset:

- **Integración con fuentes externas:** Combinar con inspecciones, condiciones ambientales, horarios de trabajo o rotación de personal permitiría entender **causas estructurales** del riesgo.
- **Historial de inversión:** Cruzar con montos invertidos en seguridad, permitiría **medir el impacto real** de las acciones y estrategias.
- **Contexto Normativo:** Ajustar el análisis según el **estado, sector o marco regulatorio** mejoraría la toma de decisiones específicas.

Modelos por Industria:

- Sectores como educación, salud, construcción, transporte, manufactura tienen **dinámicas y patrones de riesgo distinto**.

- Entrenar modelos ajustados a cada sector permitiría generar **recomendaciones precisas** y con gran impacto.
- Los **planes de prevención segmentados** serían más fáciles de ejecutar, mas eficientes y con mayor ROI en el corto plazo.

Actualización Continua:

- La **automatización** permitiría actualizar modelos e indicadores de riesgo cada vez que se incorporen datos nuevos.
- Las políticas podrían **actualizarse dinámicamente** alineándose mejor con la realidad empresarial y operativa.



Conclusiones

Mayor foco mejor ROI en prevención.



Beneficios

Transformar datos en acción para prevenir,
optimizar y proteger.

Prevención Ocurrencia del Riesgo:

- La predicción permite anticipar áreas, puestos, turnos, regiones donde exista mayor incide o nivel de riesgo y tomar **acciones para prevenir**.
- A través de los datos históricos los modelos pueden detectar **combinaciones de variables que aumentan el riesgo** no evidentes a simple vista.

Asignación Eficiente de Recursos:

- **Focalizar capacitaciones e inspecciones** donde el modelo indique mayor riesgo.
- **Medir si el índice disminuye** luego de implementar acciones para realizar una gestión basada en resultados.

Diseño de Políticas Inteligentes:

- Las áreas de salud ocupacional y seguridad en el trabajo pueden **sustentar sus acciones con datos**.
- Políticas que **evolucionan acorde a las nuevas características involucradas** y los nuevos valores generados por el modelo.

Reducción de Costos Siniestralidad:

- Al reducir **los accidentes laborales** se reducen los gastos por seguros, sanciones, atención médica, paros, etc.
- **Saber donde invertir** para lograr un mayor impacto reduce el gasto y maximiza el beneficio (ROI).



Beneficios

Diseñar políticas inteligentes y reducir costos con decisiones basadas en evidencia.

Resumen Ejecutivo

Prevención basada en datos para entornos laborales
más seguros.

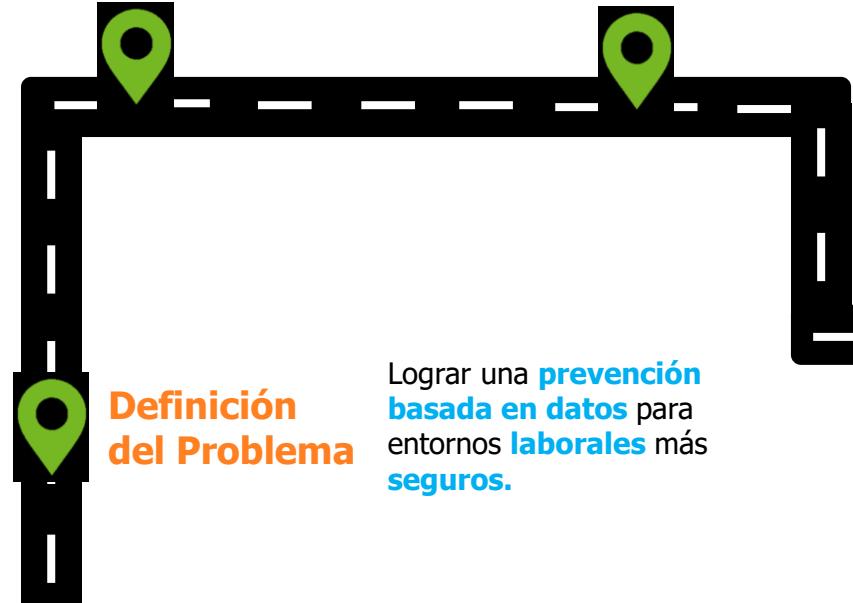
El fin de la jornada laboral, Jules Breton.

Journey del Proyecto

Un modelo valioso no es el que predice mejor, sino el que impulsa mejores decisiones.

- Valores Inconsistentes
- Valores Nulos
- Correlaciones
- Variables de Contexto
- Limpieza de Datos
- Imputaciones

EDA



Primeros Experimentos

- 6 Modelos de Regresión Lineal.
- Variables Correlacionadas
- Variable Objetivo total_injuries
- No captura el riesgo de manera contextual y accionable.

Cambio Objetivo

- Definir features que permitan capturar la magnitud del riesgo y clasificarlo.
- Facilitar la comparación (tamaño de empresas, sectores, estados)

Índice y Nivel Riesgo

- Modelo Delaware
- total_injuries, total_deaths, total_hours_worked
- Percentiles

- Random Forrest
- Regressor
 - Classifier

Modelos Finales

- App en streamlit para ingresar variables y obtener risk_index y risk_level

App Streamlit

- Balanceo Clases
- PCA
- Hiper-parametros
- Nuevos Experimentos
- Análisis de Clases.

Próximos Pasos



Lecciones Aprendidas

- Importancia:
- Variable Objetivo
 - Variables de Contexto
- Un buen Modelo no siempre significa una mejor acción desde el negocio.

Demo

Prevención basada en datos para entornos laborales
más seguros.

El fin de la jornada laboral, Jules Breton.

Praevidens App:



Praevidens: Predicción de Riesgo Laboral OSHA



Promedio Anual de Empleados 174	Total de Horas Trabajadas 368960	Casos DJTR (con Restricción) 1
Total de Muertes 6	Casos DAFW (con Ausencias) 1	Selección Estado Maryland
Otros Casos 7	Total de Enfermedades 0	Selección Sector Construction
Tamaño Empresa 20 a 249 Empleados		

Predecir

Praevidens App: Predicción



Praevidens: Predicción de Riesgo Laboral OSHA



Praevidens.
**Predicción con IA
del Riesgo Laboral
para una Gestión
Proactiva**

Promedio Anual de Empleados	Total de Horas Trabajadas	Casos DJTR (con Restricción)
174	368960	1
Total de Muertes	Casos DAFW (con Ausencias)	Selecciona Estado
6	1	Maryland
Otros Casos	Total de Enfermedades	Selecciona Sector
7	0	Construction
Tamaño Empresa		
20 a 249 Empleados		

Predecir

Índice de Riesgo: 127.77

Nivel de Riesgo: Severe

Praevidens App: Predicción



Praevidens: Predicción de Riesgo Laboral OSHA



Praevidens.

**Predicción con IA
del Riesgo Laboral
para una Gestión
Proactiva**

Promedio Anual de Empleados 3340 - +	Total de Horas Trabajadas 6132584 - +	Casos DJTR (con Restricción) 113 - +
Total de Muertes 1 - +	Casos DAFW (con Ausencias) 64 - +	Selecciona Estado Kansas ▼
Otros Casos 62 - +	Total de Enfermedades 120 - +	Selecciona Sector Manufacturing ▼
Tamaño Empresa Más 250 Empleados ▼		

Predecir

Índice de Riesgo: 6.43

Nivel de Riesgo: Critical

Praevidens App: Predicción



Praevidens: Predicción de Riesgo Laboral OSHA



Praevidens.

**Predicción con IA
del Riesgo Laboral
para una Gestión
Proactiva**

👤 Promedio Anual de Empleados

35

⌚ Total de Horas Trabajadas

68943

👤 🚗 Casos DJTR (con Restricción)

0

💀 Total de Muertes

0

⚠️ Casos DAFW (con Ausencias)

0

🌐 Selecciona Estado

Nevada

📋 Otros Casos

0

😢 Total de Enfermedades

0

🏭 Selecciona Sector

Wholesale Trade

🏢 Tamaño Empresa

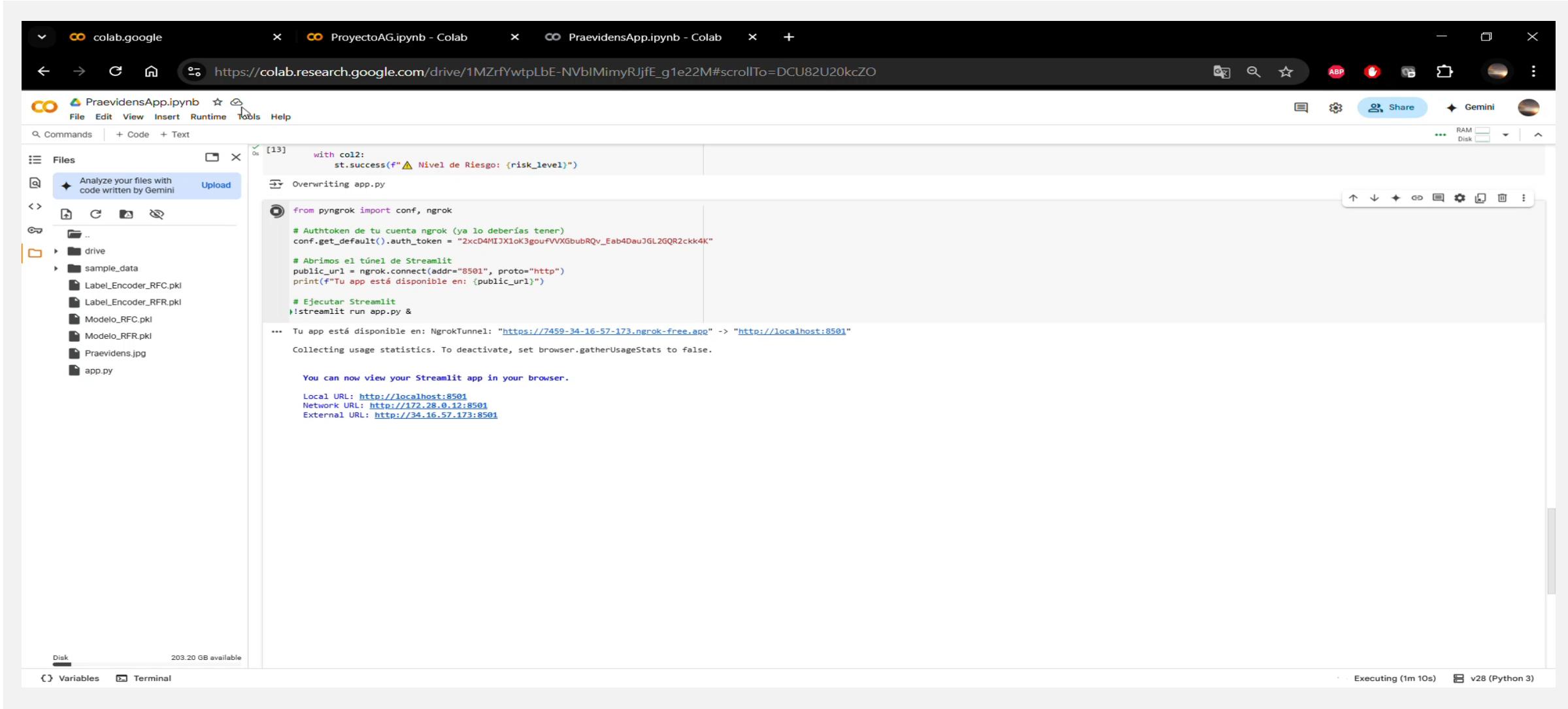
20 a 99 Empleados

👤 Predecir

📈 Índice de Riesgo: 0.00

⚠️ Nivel de Riesgo: Safe

Praevidens App: Demo



The screenshot shows a Google Colab interface with three tabs open: 'colab.google', 'ProyectoAG.ipynb - Colab', and 'PraevidensApp.ipynb - Colab'. The current tab is 'PraevidensApp.ipynb - Colab'.

The notebook code in cell [13] is as follows:

```
[13]: with col2:
    st.success(f"⚠️ Nivel de Riesgo: {risk_level}")

Overwriting app.py

from pyngrok import conf, ngrok

# Authtoken de tu cuenta ngrok (ya lo deberías tener)
conf.get_default().auth_token = "2xD4MIJX1ok3goufVvXGhubRQv_Eab4DauJGL2GQR2ckk4K"

# Abrimos el túnel de Streamlit
public_url = ngrok.connect(addr="8501", proto="http")
print(f"Tu app está disponible en: {public_url}")

# Ejecutar Streamlit
!streamlit run app.py &

... Tu app está disponible en: NgrokTunnel: "https://7459-34-16-57-173.ngrok-free.app" -> "http://localhost:8501"
Collecting usage statistics. To deactivate, set browser.gatherUsageStats to false.

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://172.28.0.12:8501
External URL: http://34.16.57.173:8501
```

The sidebar shows a file tree with 'Files' containing 'app.py', 'Label_Encoder_RFC.pkl', 'Label_Encoder_RFR.pkl', 'Modelo_RFC.pkl', 'Modelo_RFR.pkl', 'Praevidens.jpg', and 'sample_data'. The 'Variables' and 'Terminal' sections are at the bottom.



Urls y Colab

Gestión preventiva del riesgo con evidencia y eficiencia.

Dataset:

- <https://www.osha.gov/Establishment-Specific-Injury-and-Illness-Data>
- [2024 Summary Data](#)

Bibliografía:

- <https://www.osha.gov/>

Github:

- <https://github.com/Andregg05/Praevidens>

Colab:

- ProyectoAG.ipynb
- PraevidensApp.ipynb



Accidente de trabajo, Aurelio Arteta Errasti.

Praevidens.

Gracias

👤 Andrea Gutiérrez González

✉ andregg05@gmail.com