

<!--Simulacro proyecto final-->

## INTEGRANTES:



- Alexandra Andi3n
- Angela Charris
- Aracelly Reguillo
- Bianys Ballesteros
- Cindy Villalobos
- Deisy Su3rez
- Edith Julio
- Grace Lastre
- Jenny Ruiz
- Ketty Ortega
- Nicole Rua
- Paula Florez
- Paula Martinez
- Rosario Peñuela
- Saray Urueta
- Virginia Guerrero



```
<!--Simulacro proyecto final-->
```

Día 1:  
Introducción y  
Preparación de los  
Datos {

<Por=Grupo de trabajo # 3

}



## Objetivo del día:

- 01 Conocer las bases de datos que van a analizar.
- 02 Aprender a cargar, limpiar y explorar los datos con pandas.
- 03 Formulación de preguntas y una hipótesis clara para el análisis.

# Tareas del día {

01

Presentación  
del taller

02

Carga y  
limpieza de  
los datos

03

Exploración  
de los datos

04

Formulación  
de hipótesis

05

Entregable  
del día

}

# Presentación del taller

{

En calidad de equipo consultor en analítica de datos, hemos sido contratados por el CLIENTE para realizar un estudio detallado sobre los accidentes eléctricos ocurridos en Colombia entre los años 2010 y 2020.

Como punto de partida, el CLIENTE nos entregó una base de datos oficial y planteó el siguiente objetivo general para el proyecto:

## Ojetivo general

Diseñar un informe diagnóstico sobre las causas de los accidentes laborales de origen eléctrico en las empresas del sector eléctrico en Colombia, mediante el análisis de datos provenientes de la base Superservicios-Información de Accidentes de Origen Eléctrico, correspondiente al periodo 2010-2021.

## Accidentes eléctricos años 2010 - 2020

Exportar

Minas Y Energía

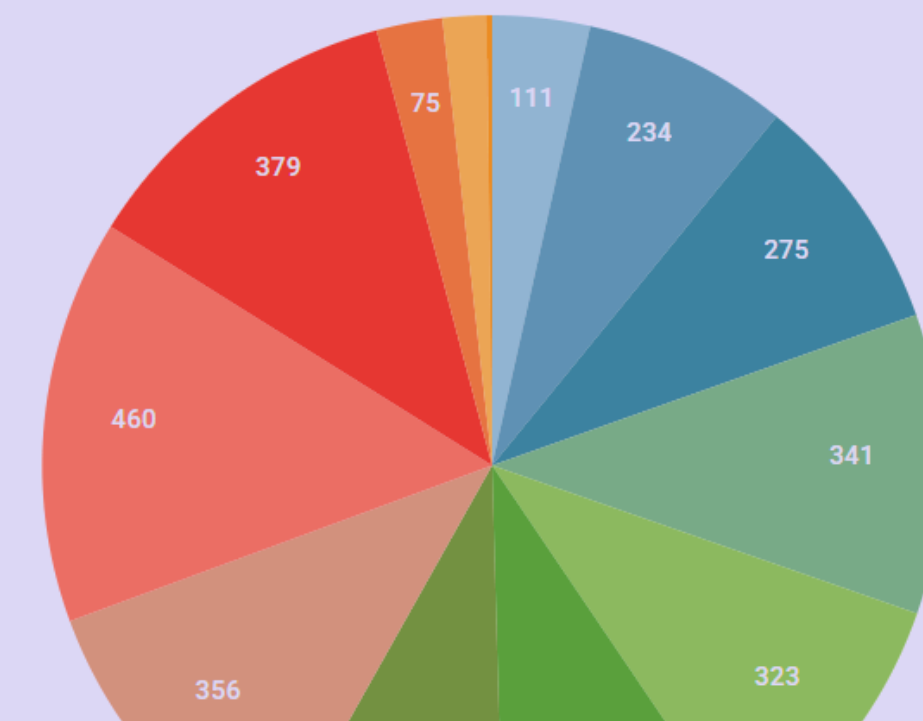
Más información ▾



Un miembro del público creó esta visualización. La plataforma de datos abiertos del gobierno colombiano no ha revisado ni avalado ningún cambio, incluidos los filtros y las actualizaciones del título y la descripción. [Obtener información adicional.](#)

### Accidentes eléctricos 2010-2019

Muertes



}

# Presentación del taller

{

Para guiar este proceso, se establecieron los siguientes objetivos específicos, que orientaron cada etapa del trabajo:

## Ojetivos específicos

- Implementar el uso de herramientas como Pandas para conocer, cargar, limpiar y explorar los datos, y de **Numpy**, **Matplotlib** y **Seaborn** para su visualización y análisis gráfico.
- Visualizar el conjunto de datos con el fin de responder preguntas clave y validar hipótesis formuladas a partir de observaciones preliminares.
- Analizar las principales fuentes que causan los accidentes laborales de origen eléctrico en el sector eléctrico colombiano.
- Formular estrategias orientadas a la prevención de accidentes, útiles para empresas del sector eléctrico que buscan fortalecer sus medidas de seguridad laboral.

## Accidentes eléctricos años 2010 - 2020

Exportar

Minas Y Energía

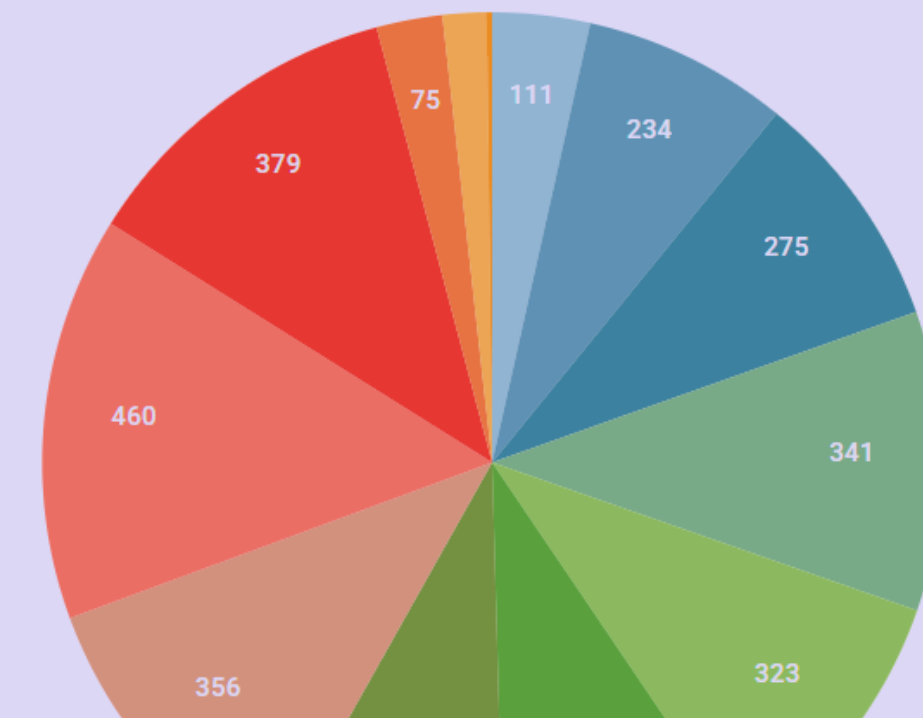
Más información ▾



Un miembro del público creó esta visualización. La plataforma de datos abiertos del gobierno colombiano no ha revisado ni avalado ningún cambio, incluidos los filtros y las actualizaciones del título y la descripción. [Obtener información adicional.](#)

### Accidentes eléctricos 2010-2019

Muertes



}

Revisión inicial del  
contexto {

## # Reglamento Técnico de Instalaciones Eléctricas - RETIE

### Accidente eléctrico.

todo evento no deseado originado por la electricidad que puede generar daño a personas, bienes o animales" (Ministerio de Minas y Energía, 2023).

### Tipos de accidentes.

**contacto directo**, cuando una persona toca partes activas de una instalación; **contacto indirecto**, al entrar en contacto con partes metálicas energizadas por fallas de aislamiento; y **arco eléctrico o cortocircuito**, debido a la ruptura del aislamiento, fallas en el sistema o errores humanos.

Con esta base conceptual, procedimos a realizar el análisis exploratorio de los datos entregados por el CLIENTE, los cuales recopilan información oficial de accidentes eléctricos entre 2010 y 2020, incluyendo variables como: **año, departamento, tipo de red, actividad que realizaba la víctima, y resultado del accidente.**

}



# Carga y limpieza de los datos {

Se ejecutaron comandos desde la terminal integrada de Visual Studio Code para inicializar el repositorio local, enlazarlo con GitHub y subir el proyecto

## # COMANDOS PARA GIT Y GITHUB

### ## Inicializar el repositorio Git local

```
git init
```

### ## Agregar el repositorio remoto de GitHub

```
git remote add origin
```

```
https://github.com/dsuarezm0425/Taller-SQL-Python-Numpy-Matplotlib-y-Seaborn.git
```

### ## Agregar todos los archivos

```
git add .
```

### ## Realizar el primer commit

```
git commit -m "Primer commit del taller - carga inicial"
```

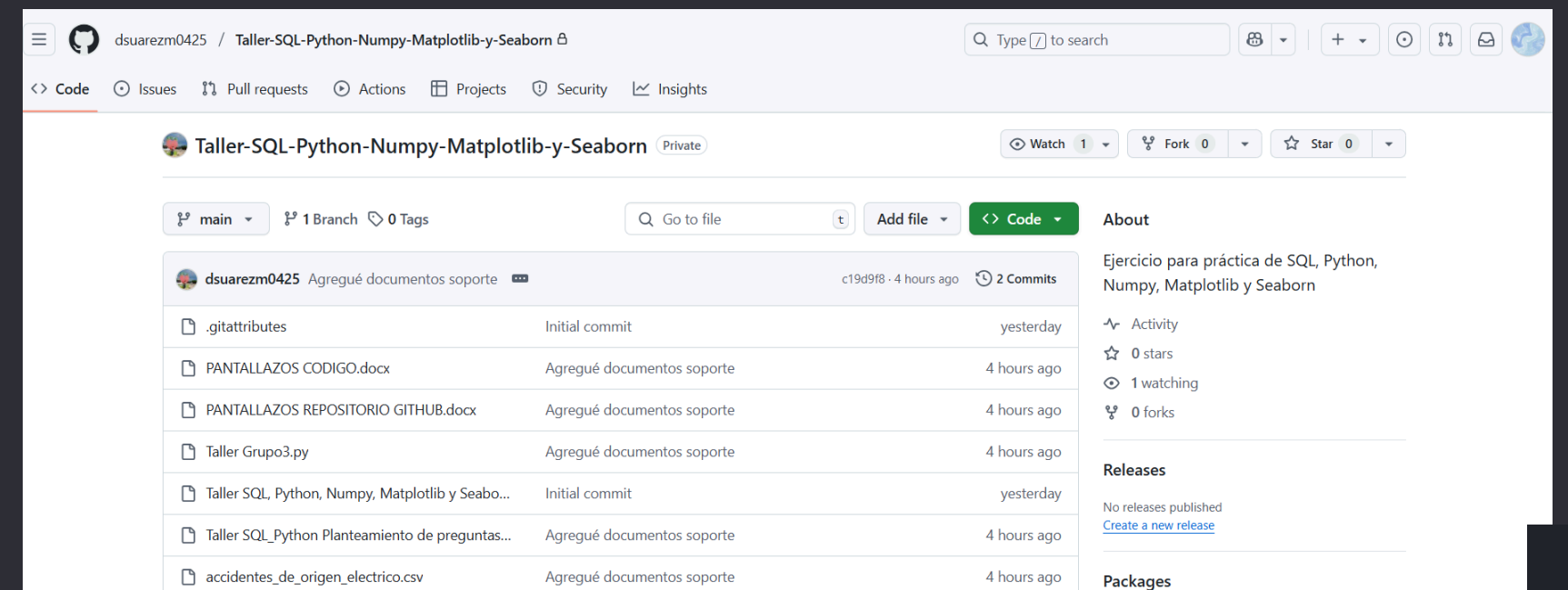
### ## Subir a la rama principal

```
git push -u origin main
```

## # Git y GitHub

Imagen de referencia

```
PS C:\Users\LENOVO\Desktop\DA\Proyecto_Simulacro> git remote add origin
in https://github.com/NicoleARM/Simulacro.git
PS C:\Users\LENOVO\Desktop\DA\Proyecto_Simulacro> git add .
PS C:\Users\LENOVO\Desktop\DA\Proyecto_Simulacro> git commit -m "Primer
commit del taller - carga inicial"
[master (root-commit) 200dbc9] Primer commit del taller - carga inici
al
 1 file changed, 3169 insertions(+)
 create mode 100644 Accidentes_de_Origen_Electrico.csv
PS C:\Users\LENOVO\Desktop\DA\Proyecto_Simulacro> git push origin mai
n
Enumerating objects: 4, done.
Counting objects: 100% (4/4), done.
Delta compression using up to 8 threads
Compressing objects: 100% (2/2), done.
Writing objects: 100% (2/2), 358 bytes | 358.00 KiB/s, done.
Total 2 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
To https://github.com/NicoleARM/Simulacro.git
 3c36a38..32627e8  main -> main
PS C:\Users\LENOVO\Desktop\DA\Proyecto_Simulacro>
```





# Carga y limpieza de los datos {

# Visual Studio Code

Run Python File

Número de filas y columnas: (3168, 23)

```
1 import pandas as pd
# Cargar archivo CSV
df = pd.read_csv ('Accidentes_de_Origen_Electrico.csv', delimiter=',', encoding='latin1')
# Análisis básico
print("Número de filas y columnas:", df.shape)
print("\n Tipos de datos:")
print(df.dtypes)
print("\n Cantidad de valores nulos por columna:")
print(df.isnull().sum())
print("\n Cantidad de valores únicos por columna:")
print(df.nunique())
print("\n Estadísticas descriptivas (solo numéricas):")
print(df.describe())
print("\n Primeras filas de la base:")
print(df.head())
```

Librería PANDAS

Inspección inicial

Tipos de datos:

IDENTIFICADOR_EMPRESA	object
EMPRESA	object
FECHA	object
ANIO	float64
TRIMESTRE	float64
HORA	object
COD_DANE	float64
DEPARTAMENTO	object
MUNICIPIO	object
CENTRO_POBLADO	object
UBICACION	object
SEXO	object
EDAD	float64
TIPO_IDENTIFICACION	object
VINCULADO_EMPRESA	object
TIPO_VINCULACION	object
GRADO_ESCOLARIDAD	object
TIEMPO_VINCULACION	float64
SECCION_EMPRESA	object
TIPO_LESION	object
ORIGEN_ACCIDENTE	object
CAUSA_ACCIDENTE	object
MEDIDAS;	object
dtype:	object

Cantidad de valores nulos por columna:

IDENTIFICADOR_EMPRESA	0
EMPRESA	2643
FECHA	2643
ANIO	2643
TRIMESTRE	2643
HORA	2643
COD_DANE	2643
DEPARTAMENTO	2644
MUNICIPIO	2644
CENTRO_POBLADO	2644
UBICACION	2643
SEXO	2643
EDAD	2643
TIPO_IDENTIFICACION	2643
VINCULADO_EMPRESA	2643
TIPO_VINCULACION	2643
GRADO_ESCOLARIDAD	2643
TIEMPO_VINCULACION	2643
SECCION_EMPRESA	2644
TIPO_LESION	2643
ORIGEN_ACCIDENTE	2643
CAUSA_ACCIDENTE	2643
MEDIDAS;	2643
dtype:	int64

Se realizó la carga del archivo **Accidentes\_de\_Origen\_Electrico.csv** utilizando la librería pandas. Posteriormente, se llevó a cabo una limpieza de los datos que incluyó: conversión de tipos de variables numéricas, eliminación de duplicados y columnas irrelevantes, tratamiento de valores nulos en campos clave como TIPO\_LESION y CAUSA\_ACCIDENTE, así como el ajuste de nombres de columnas para asegurar su correcto uso en análisis posteriores.

# Carga y limpieza de los datos {

# Visual Studio Code

Proyecto\_Simulacro / Carga\_datos

```
1 import pandas as pd
```

Librería PANDAS

```
# Cargar archivo CSV
df = pd.read_csv ('Accidentes_de_Origen_Electrico.csv', delimiter=',', encoding='latin1')

# Análisis básico
print("Número de filas y columnas:", df.shape)

print("\n Tipos de datos:")
print(df.dtypes)

print("\n Cantidad de valores nulos por columna:")
print(df.isnull().sum())

print("\n Cantidad de valores únicos por columna:")
print(df.nunique())

print("\n Estadísticas descriptivas (solo numéricas):")
print(df.describe())

print("\n Primeras filas de la base:")
print(df.head())
```

Inspección inicial

Run Python File

Cantidad de valores únicos por columna:

IDENTIFICADOR_EMPRESA	2647
EMPRESA	37
FECHA	449
ANIO	12
TRIMESTRE	4
HORA	166
COD_DANE	231
DEPARTAMENTO	27
MUNICIPIO	206
CENTRO_POBLADO	226
UBICACION	3
SEXO	2
EDAD	47
TIPO_IDENTIFICACION	5
VINCULADO_EMPRESA	2
TIPO_VINCULACION	5
GRADO_ESCOLARIDAD	8
TIEMPO_VINCULACION	144
SECCION_EMPRESA	11
TIPO_LESION	9
ORIGEN_ACCIDENTE	24
CAUSA_ACCIDENTE	12
MEDIDAS;	370
dtype: int64	

Estadísticas descriptivas (solo numéricas):

	ANIO	TRIMESTRE	COD_DANE	EDAD	TIEMPO_VINCULACION
count	525.000000	525.000000	5.250000e+02	525.000000	525.000000
mean	2014.171429	2.544762	4.733546e+07	37.468571	57.727619
std	2.685333	1.098401	3.082067e+07	11.834154	97.256017
min	2010.000000	1.000000	5.001000e+06	1.000000	1.000000
25%	2012.000000	2.000000	1.744400e+07	30.000000	6.000000
50%	2014.000000	3.000000	5.200100e+07	36.000000	19.000000
75%	2016.000000	3.000000	7.602000e+07	46.000000	50.000000
max	2021.000000	4.000000	9.520000e+07	99.000000	816.000000

# Carga y limpieza de los datos {

# Visual Studio Code

Proyecto\_Simulacro / Carga\_datos

```
1 import pandas as pd
```

Librería PANDAS

```
# Cargar archivo CSV
df = pd.read_csv ('Accidentes_de_Origen_Electrico.csv', delimiter=',', encoding='latin1')

# Análisis básico
print("Número de filas y columnas:", df.shape)

print("\n Tipos de datos:")
print(df.dtypes)

print("\n Cantidad de valores nulos por columna:")
print(df.isnull().sum())

print("\n Cantidad de valores únicos por columna:")
print(df.nunique())

print("\n Estadísticas descriptivas (solo numéricas):")
print(df.describe())

print("\n Primeras filas de la base:")
print(df.head())
```

Inspección inicial

Run Python File

Primeras filas de la base:

	IDENTIFICADOR_EMPRESA	EMPRESA	...	CAUSA_ACCIDENTE	MEDIDAS;
0	617.0	EMPRESAS MUNICIPALES DE CARTAGO S.A.S. E.S.P.	...	Sobrecarga	Utilizar; debidamente los equipos de proteccii...
1	617.0	EMPRESAS MUNICIPALES DE CARTAGO S.A.S. E.S.P.	...	Equipo defectuoso	SE IMPLEMENTARA UN PROTOCOLO DE TRABAJO SEGURO;
2	3370.0	EMPRESA DE ENERGIA DE CASANARE SA ESP	...	Contacto Directo	Instalacion correcta puesta a tierra Reforzar ...
3	23442.0	COMPAÑIA ENERGETICA DE OCCIDENTE S.A.S. E.S.P.	...	Contacto Indirecto	Realizar un programa de capacitacion de identi...
4	617.0	EMPRESAS MUNICIPALES DE CARTAGO S.A.S. E.S.P.	...	Arcod Eléctricos	CAMBIO DE 2 CORTA CIRCUITOS PODA DE ARBOLES Y ...

[5 rows x 23 columns]

}

# Carga y limpieza de los datos {

# Visual Studio Code

Proyecto\_Simulacro / Carga\_Datos

```
1 import pandas as pd
```

Librería PANDAS

```
## Eliminar Duplicados  
df = df.drop_duplicates()
```

```
## Elimina todas las filas que tengan al menos un Na  
df = df.dropna()
```

```
##Revisar que columnas numericas tenemos  
print(df.dtypes)
```

Limpieza de datos

```
##Convertir tipos de datos
```

```
df['ANIO'] = df['ANIO'].astype('Int64')  
df['TRIMESTRE'] = df['TRIMESTRE'].astype('Int64')  
df['EDAD'] = df['EDAD'].astype('Int64')  
df['TIEMPO_VINCULACION'] = df['TIEMPO_VINCULACION'].astype('Int64')
```

```
##Revisar datos después de limpieza  
print("Número de filas y columnas después de limpieza:", df.shape)
```

Número de filas y columnas después de limpieza: (520, 23)

Run Python File

```
IDENTIFICADOR_EMPRESA    object  
EMPRESA                  object  
FECHA                   object  
ANIO                    float64  
TRIMESTRE               float64  
HORA                    object  
COD_DANE                float64  
DEPARTAMENTO             object  
MUNICIPIO               object  
CENTRO_POBLADO          object  
UBICACION              object  
SEXO                    object  
EDAD                   float64  
TIPO_IDENTIFICACION     object  
VINCULADO_EMPRESA       object  
TIPO_VINCULACION        object  
GRADO_ESCOLARIDAD       object  
TIEMPO_VINCULACION      float64  
SECCION_EMPRESA         object  
TIPO_LESION             object  
ORIGEN_ACCIDENTE        object  
CAUSA_ACCIDENTE         object  
MEDIDAS;                object  
dtype: object
```

}



# Exploración de los datos {

Para dar inicio al análisis del conjunto de datos, se llevó a cabo una fase de **exploración de datos**, cuyo objetivo principal fue comprender la estructura y características generales de las variables relacionadas con los accidentes eléctricos registrados. Para esto, se formularon **preguntas exploratorias** que guiaron la elaboración de visualizaciones básicas mediante herramientas como **gráficos de barras, histogramas y gráficos de torta**, configurados a través de las bibliotecas *matplotlib* y *seaborn* de Python.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Cargar el archivo limpio
df = pd.read_csv("AccidentesElectricos_Limpio.csv")
```

→ Librerías

## # Preguntas exploratorias

- ¿Cuál es la distribución de edad de las personas que han sufrido accidentes eléctricos?
- ¿Cómo varía la edad según el sexo?
- ¿Cómo se distribuyen los accidentes por año?
- ¿Cuál es el trimestre con más accidentes y su porcentaje?
- ¿Cuál es el departamento con más accidentalidad?
- ¿Cuáles son las principales causas, orígenes y tipos de lesión?
- ¿Cuáles son las primeras 5 empresas que reportan más accidentes?
- ¿Cuáles son las principales características de los accidentes en las dos empresas con mayor número de accidentes

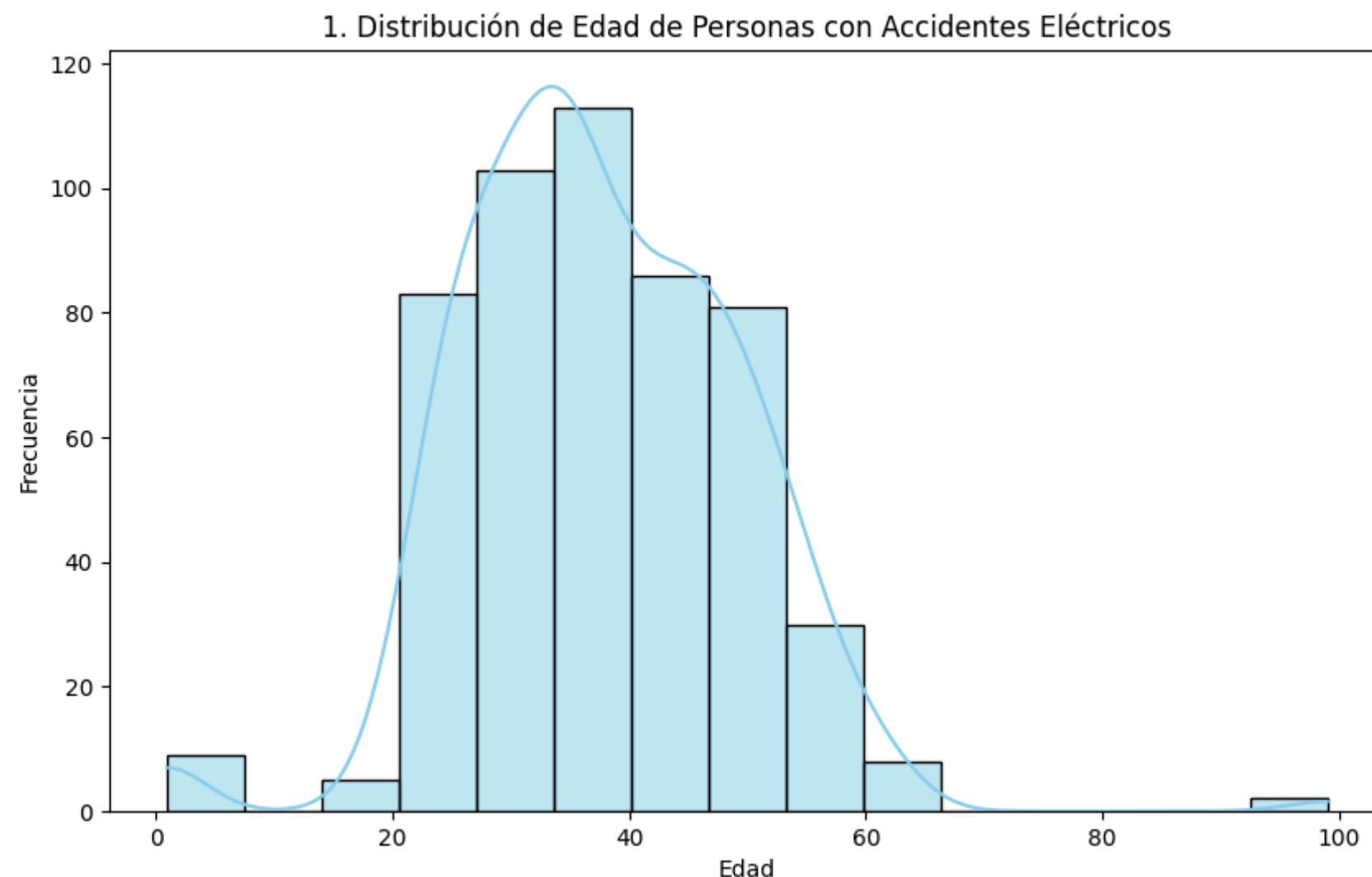
}

# Exploración de los datos {

# Visual Studio Code

```
#1. ¿Cuál es la distribución de edad de las personas que han sufrido accidentes eléctricos? (histograma)
plt.figure(figsize=(10,6))
sns.histplot(df["EDAD"], bins=15, kde=True, color='skyblue')
plt.title("1. Distribución de Edad de Personas con Accidentes Eléctricos")
plt.xlabel("Edad")
plt.ylabel("Frecuencia")
plt.show()
```

¿Cuál es la distribución de edad de las personas que han sufrido accidentes eléctricos?



## Hipótesis

La mayor incidencia de accidentes eléctricos se presentan en la población joven.

## Conclusión

Entre 2010 y 2020, la mayoría de los accidentes eléctricos en Colombia afectaron a personas adultas jóvenes y de mediana edad (25-55 años), con un pico alrededor de los 35-40 años. Estos datos sugieren que la población laboralmente activa fue la más expuesta, lo cual es clave para enfocar estrategias de prevención.

}

# Exploración de los datos {

# Visual Studio Code

```
# 2. ¿Cómo varía la edad según el sexo? (histograma)
plt.figure(figsize=(10,6))
sns.histplot(data=df, x="EDAD", hue="SEXO", bins=15, kde=True, palette="Set2", multiple="stack")
plt.title("2. Distribución de Edad por Sexo en Accidentes Eléctricos")
plt.xlabel("Edad")
plt.ylabel("Frecuencia")
plt.show()
```

¿Cómo varía la edad según el sexo?

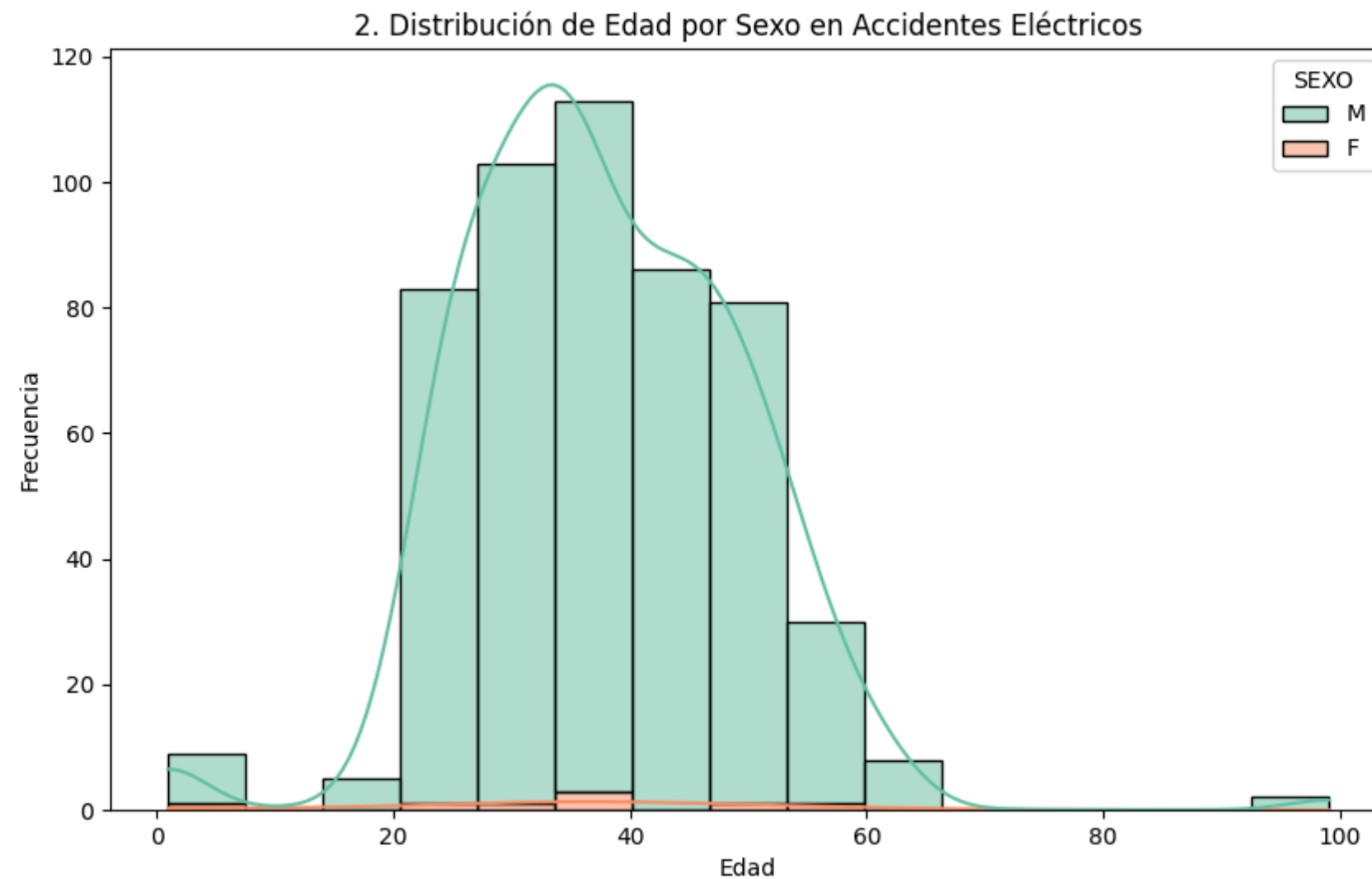
## Hipótesis

Los hombres presentan una mayor tasa de accidentes eléctricos laborales que las mujeres en Colombia entre 2010 y 2020 posiblemente a la relación de que los hombres tiene mayor participación en oficios con riesgos eléctricos que las mujeres.

## Conclusión

La gráfica muestra que, entre 2010 y 2020 en Colombia, los hombres sufrieron muchos más accidentes eléctricos que las mujeres, especialmente entre los 25 y 55 años. Ambos sexos comparten un patrón de mayor incidencia en la edad adulta, pero la frecuencia en hombres es significativamente mayor.

}





# Exploración de los datos {

# Visual Studio Code

```
#3. ¿Cómo se distribuyen los accidentes por año? (countplot)
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x="AÑO", palette="tab10")
plt.title("3. Distribución de accidentes eléctricos por año", fontsize=14)
plt.xlabel("Año", fontsize=12)
plt.ylabel("Número de Accidentes", fontsize=12)
plt.xticks(rotation=45)
plt.show()
```

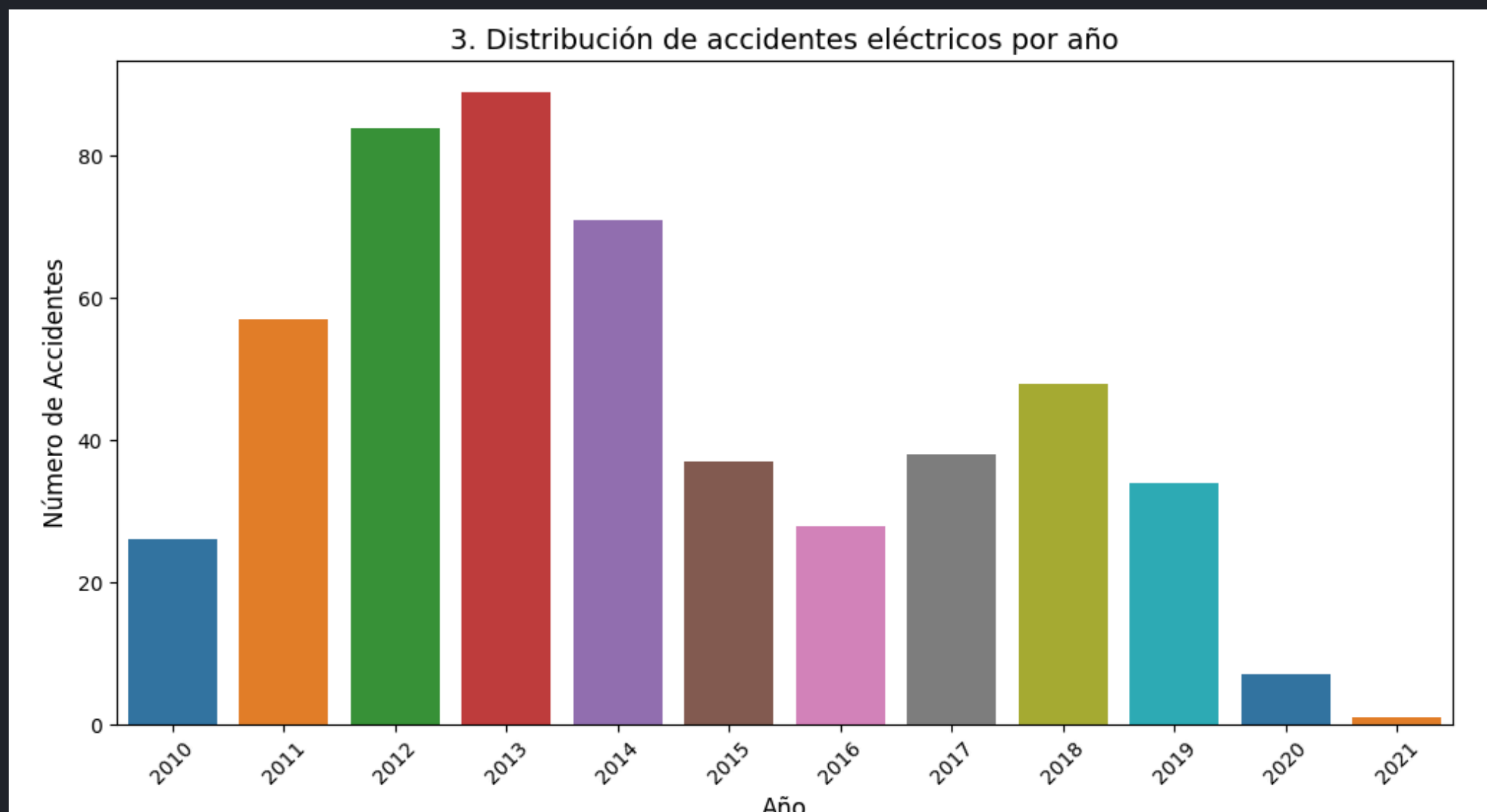
¿Cómo se distribuyen los accidentes por año?

## Hipótesis

Entre 2010 y 2021, las medidas de prevención y control de riesgos laborales en el sector eléctrico contribuyeron a la disminución progresiva de los accidentes eléctricos en Colombia.

## Conclusión

La gráfica muestra una tendencia decreciente en los accidentes eléctricos después de 2013, lo que sugiere que, posiblemente, se han implementado acciones efectivas de prevención y control. Sin embargo, las fluctuaciones en algunos años indican que aún existen factores variables que requieren atención continua.



}

# Exploración de los datos {

# Visual Studio Code

```
#4. ¿Cuál es el trimestre con más accidentes y su porcentaje? (gráfico torta)
trimestre_counts = df['TRIMESTRE'].value_counts().sort_index()      #contar accidentes por trimestre
trimestre_percent = (trimestre_counts / trimestre_counts.sum()) * 100  #para % por trimestre
labels = [f"Trimestre {i}" for i in trimestre_counts.index]         #etiqueta para trimestres
plt.figure(figsize=(8, 8))
plt.pie(trimestre_percent.values,
        labels=labels,
        autopct='%1.1f%%',
        startangle=140,
        colors=plt.cm.Set3.colors)
plt.title("4. Porcentaje de Accidentes Eléctricos por Trimestre")
plt.axis('equal') # Mantener forma circular
plt.tight_layout()
plt.show()
```

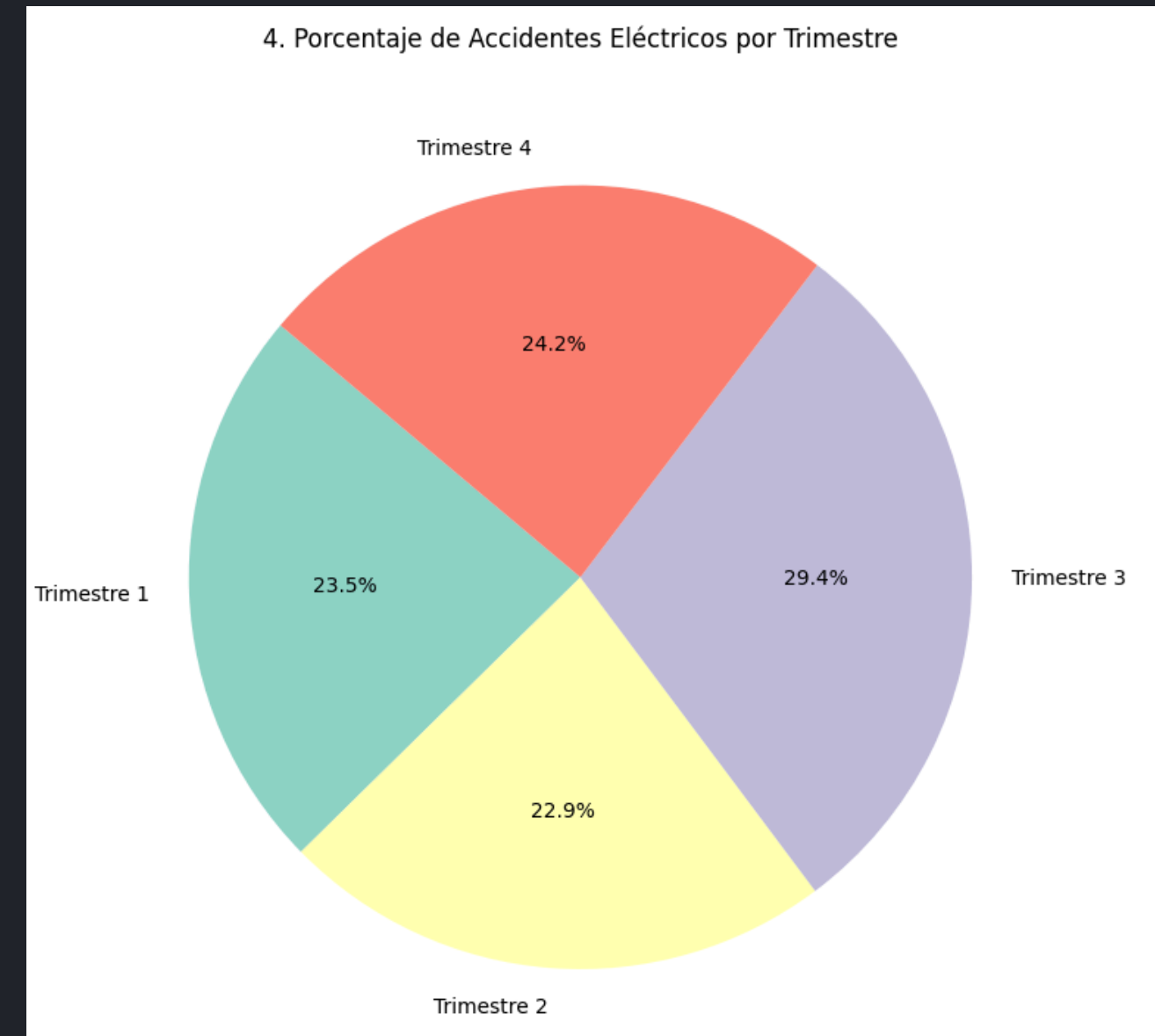
## Hipótesis

Es posible que durante el tercer trimestre del año se presenten más accidentes eléctricos en Colombia por situaciones que suelen repetirse en esa época, como cambios en el clima o mayor uso de energía.

## Conclusiones

El hecho de que el mayor porcentaje de accidentes ocurra en el tercer trimestre indica que algo particular sucede en esos meses que aumenta el riesgo. Por eso, es importante prestar más atención a ese periodo para prevenir incidentes.

¿Cuál es el trimestre con más accidentes y su porcentaje?



}

# Exploración de los datos {

# Visual Studio Code

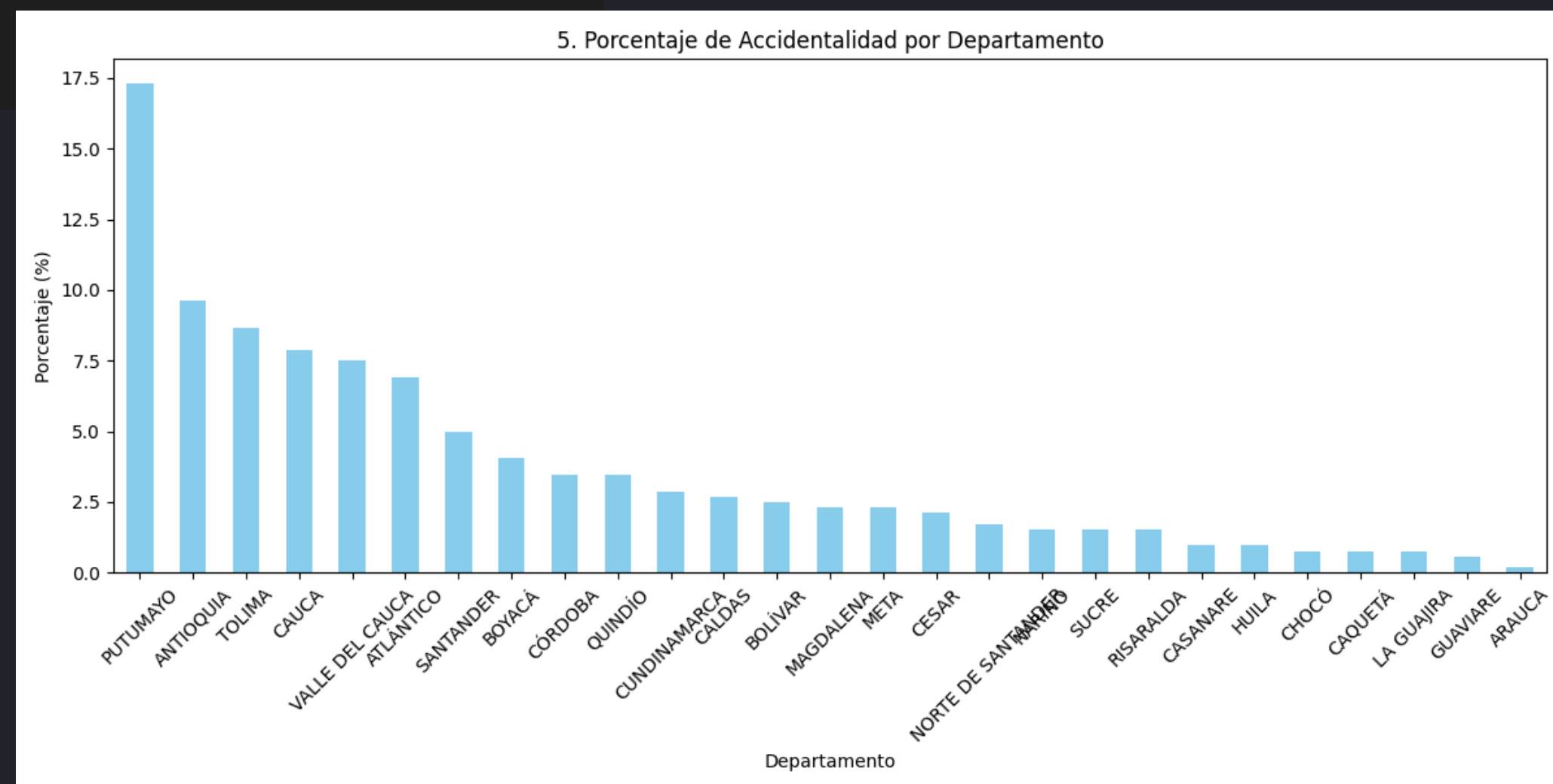
```
#5. ¿Cuál es el departamento con más accidentalidad? (barras de porcentaje)
dept_counts = df['DEPARTAMENTO'].value_counts() #contar departamentos
dept_percent = (dept_counts / len(df)) * 100 # %
anio_counts = df['ANIO'].value_counts().sort_index()
anio_percent = (anio_counts / len(df)) * 100 # lo mismo pero con años
plt.figure(figsize=(12, 6))
dept_percent.sort_values(ascending=False).plot(kind='bar', color='skyblue')
plt.title('5. Porcentaje de Accidentalidad por Departamento')
plt.ylabel('Porcentaje (%)')
plt.xlabel('Departamento')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

¿Cuál es el departamento con más accidentalidad?

Putumayo presenta el mayor porcentaje de accidentes eléctricos en el país durante el periodo analizado. Esto sugiere que en ese departamento hay factores específicos que podrían estar aumentando el riesgo y que deben ser identificados y atendidos para reducir la accidentalidad.

## Hipótesis

Es probable que el alto porcentaje de accidentes eléctricos en Putumayo se relacione con condiciones particulares del departamento, como el tipo de trabajo que se realiza, la infraestructura eléctrica o la formación del personal.



}

# Exploración de los datos {

# Visual Studio Code

```
#6. ¿Cuáles son las principales causas, orígenes y tipos de lesión? (gráfico torta)
# Obtener los valores más frecuentes
causas = df["CAUSA_ACCIDENTE"].value_counts().head(5)
origenes = df["ORIGEN_ACCIDENTE"].value_counts().head(5)
lesiones = df["TIPO_LESION"].value_counts().head(5)
# Crear la figura con 3 gráficos de torta
fig, axes = plt.subplots(1, 3, figsize=(25, 6))
# Torta de causas
axes[0].pie(causas.values, labels=causas.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.tab20.colors)
axes[0].set_title("Principales causas")
# Torta de orígenes
axes[1].pie(origenes.values, labels=origenes.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.Set3.colors)
axes[1].set_title("Principales orígenes")
# Torta de tipos de lesión
axes[2].pie(lesiones.values, labels=lesiones.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.Pastel1.colors)
axes[2].set_title("Tipos de lesión más frecuentes")
plt.tight_layout()
plt.subplots_adjust(wspace=0.5)
plt.show()
```

¿Cuáles son las principales causas, orígenes y tipos de lesión?

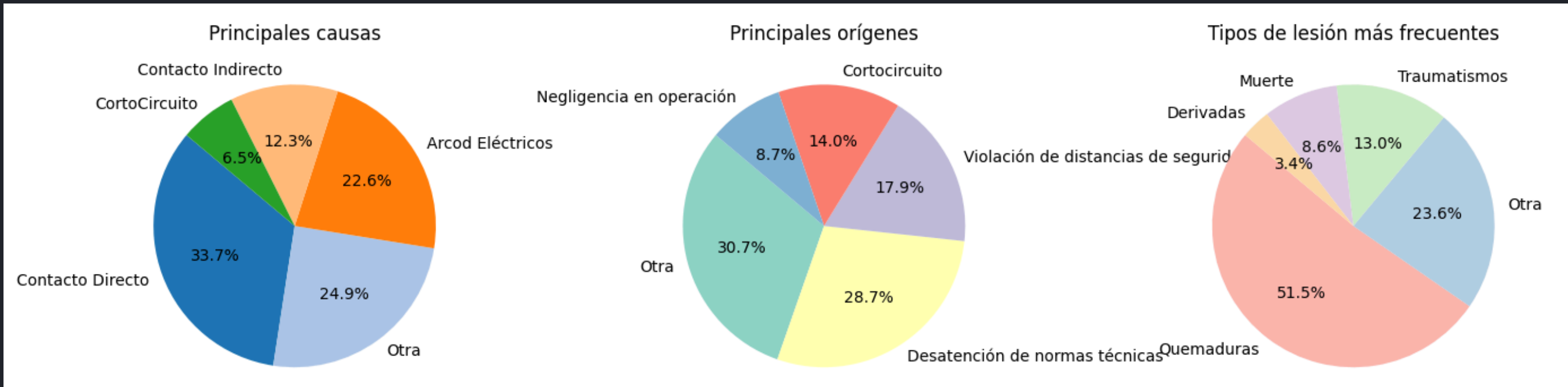
## Hipótesis

- **Causas:** La causa más frecuente de los accidentes eléctricos en Colombia durante el periodo 2010-2021 es el contacto directo con fuentes eléctricas.
- **Orígenes:** La mayoría de los accidentes eléctricos tienen su origen en el

incumplimiento de normas técnicas y de seguridad.

- **Tipos de lesión:** Las quemaduras son el tipo de lesión más común derivado de accidentes eléctricos.

## Conclusión



- **Principales causas:** Muestra que el contacto directo es la causa principal (33.7%), seguido por "otra" (24.9%), y negligencia en operación (22.6%). Una parte importante de los incidentes se debe al contacto físico con fuentes eléctricas activas, ya sea de forma directa o por arco eléctrico, lo que evidencia fallas en los sistemas de protección y aislamiento. }



# Exploración de los datos {

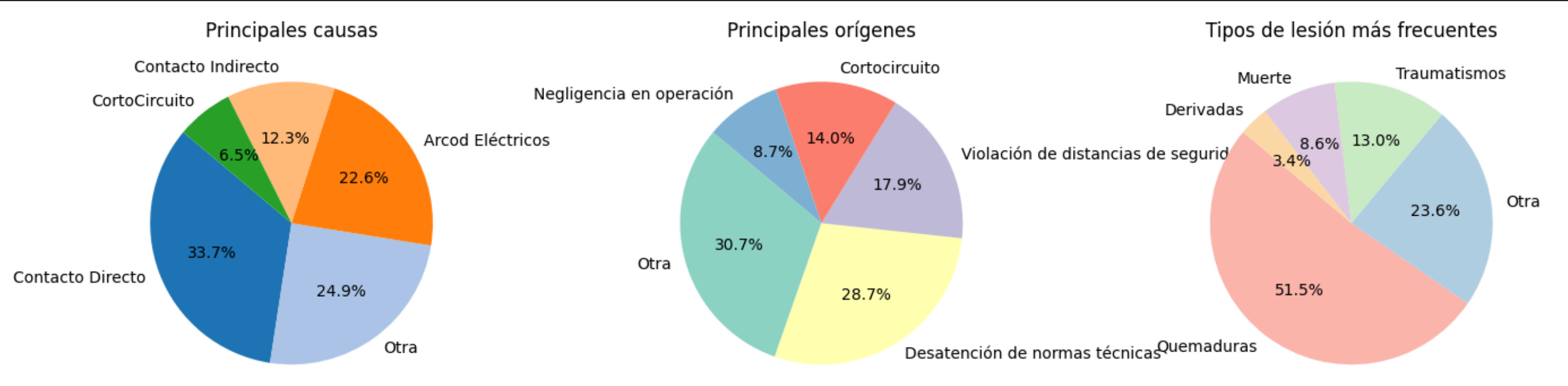
# Visual Studio Code

```
#6. ¿Cuáles son las principales causas, orígenes y tipos de lesión? (gráfico torta)
# Obtener los valores más frecuentes
causas = df["CAUSA_ACCIDENTE"].value_counts().head(5)
origenes = df["ORIGEN_ACCIDENTE"].value_counts().head(5)
lesiones = df["TIPO_LESION"].value_counts().head(5)
# Crear la figura con 3 gráficos de torta
fig, axes = plt.subplots(1, 3, figsize=(25, 6))
# Torta de causas
axes[0].pie(causas.values, labels=causas.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.tab20.colors)
axes[0].set_title("Principales causas")
# Torta de orígenes
axes[1].pie(origenes.values, labels=origenes.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.Set3.colors)
axes[1].set_title("Principales orígenes")
# Torta de tipos de lesión
axes[2].pie(lesiones.values, labels=lesiones.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.Pastel1.colors)
axes[2].set_title("Tipos de lesión más frecuentes")
plt.tight_layout()
plt.subplots_adjust(wspace=0.5)
plt.show()
```

¿Cuáles son las principales causas, orígenes y tipos de lesión?

## Conclusión

- **Principales orígenes:** Indica que la desatención a normas técnicas es la principal causa (28.7%), seguida por violación de distancias de seguridad (17.9%). Más del 45% de los incidentes están relacionados directamente con el incumplimiento de normas de seguridad eléctrica, lo que pone en



evidencia deficiencias en la capacitación, supervisión o aplicación de procedimientos técnicos.

- **Tipos de lesión más frecuentes:** Señala que las quemaduras son la lesión más común (51.5%), seguida por "otra" (23.6%) y traumatismos (13.0%). Las quemaduras representan el principal tipo de lesión, lo cual guarda coherencia con las causas reportadas (como el contacto directo o arcos eléctricos), indicando un alto nivel de riesgo térmico en los entornos analizados. }

# Exploración de los datos {

# Visual Studio Code

```
#7. ¿Cuáles son las primeras 5 empresas que reportan más accidentes? (gráfico torta)
accidentes_por_empresa = df["EMPRESA"].value_counts().head(5) # Las 5 empresas con más accidentes
plt.figure(figsize=(9, 9))
plt.pie(accidentes_por_empresa.values, labels=accidentes_por_empresa.index, autopct="%1.1f%%", colors=plt.cm.Set3.colors)
plt.title("7. Top 5 empresas con más accidentes eléctricos", fontsize=14)
plt.axis("equal") # Para mantener proporciones
plt.show()
```

¿Cuáles son las primeras 5 empresas que reportan más accidentes?

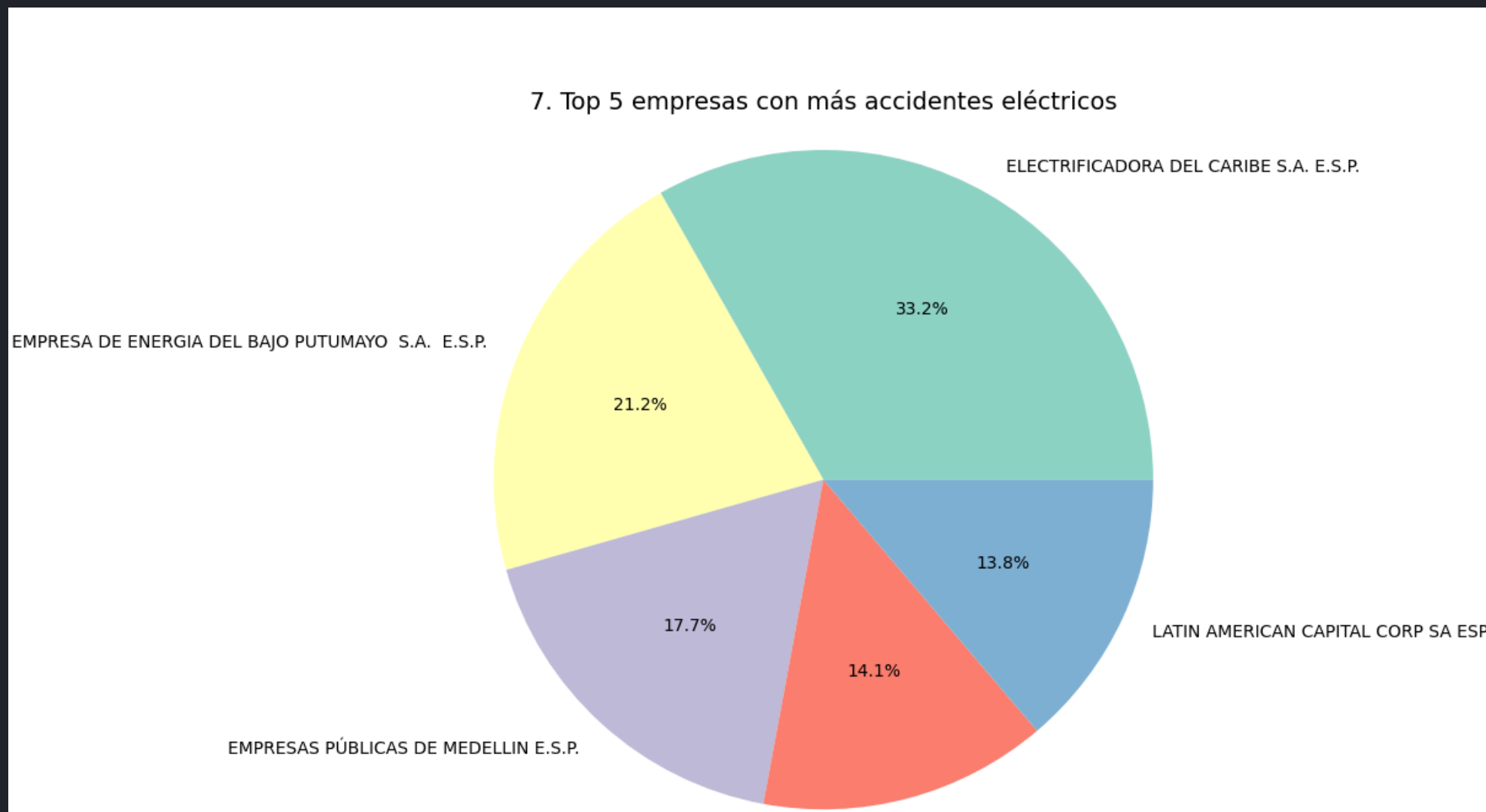
## Hipótesis

Las empresas que reportan más accidentes eléctricos podrían no estar aplicando de forma adecuada las medidas de seguridad necesarias para proteger a sus trabajadores.

## Conclusiones

La Electrificadora del Caribe S.A. E.S.P. es la empresa que más accidentes eléctricos presenta entre las cinco principales, lo que indica que se deben revisar y mejorar sus condiciones de trabajo. Como las tres primeras empresas suman más del 70% de los casos, centrarse en ellas podría ayudar a reducir de manera importante los accidentes en el país.

}



# Exploración de los datos {

# Visual Studio Code

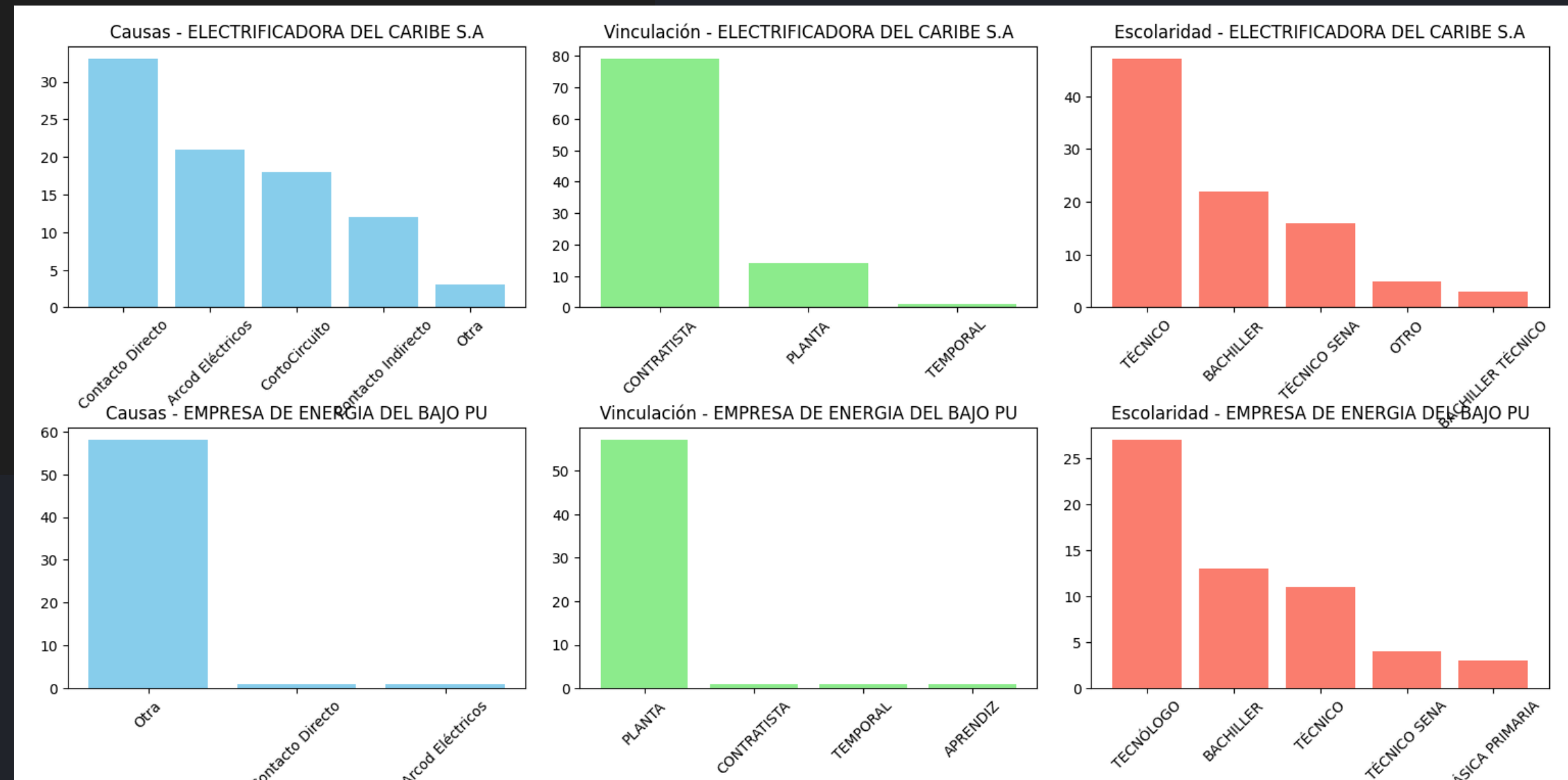
#8. ¿Cuáles son las principales características de los accidentes en las dos empresas con mayor número de accidentes (histograma)

```
# Obtener las 2 empresas con mayor número de accidentes
top_empresas = df["EMPRESA"].value_counts().head(2).index.tolist()
# Crear subplots (3 gráficos por cada empresa)
fig, axes = plt.subplots(2, 3, figsize=(20, 12))
for i, empresa in enumerate(top_empresas):
    empresa_df = df[df["EMPRESA"] == empresa]
    # Gráfico de barras para causas de accidente
    causas = empresa_df["CAUSA_ACCIDENTE"].value_counts().head(5)
    axes[i][0].bar(causas.index, causas.values, color='skyblue')
    axes[i][0].set_title(f"Causas - {empresa[:30]}")
    axes[i][0].tick_params(axis='x', rotation=45)
    # Gráfico de barras para tipo de vinculación
    vinculacion = empresa_df["TIPO_VINCULACION"].value_counts()
    axes[i][1].bar(vinculacion.index, vinculacion.values, color='lightgreen')
    axes[i][1].set_title(f"Vinculación - {empresa[:30]}")
    axes[i][1].tick_params(axis='x', rotation=45)
    # Gráfico de barras para grado de escolaridad
    escolaridad = empresa_df["GRADO_ESCOLARIDAD"].value_counts().head(5)
    axes[i][2].bar(escolaridad.index, escolaridad.values, color='salmon')
    axes[i][2].set_title(f"Escolaridad - {empresa[:30]}")
    axes[i][2].tick_params(axis='x', rotation=45)
plt.tight_layout(pad=3.0)
plt.show()
```

En las dos empresas con más accidentes, se observan diferencias en las causas y en el perfil de los empleados afectados. En la Electrificadora del Caribe, los accidentes parecen estar más relacionados con el contacto directo y los contratistas, mientras que en la Empresa de Energía del Bajo Putumayo predominan otras causas y los trabajadores vinculados a planta. Estos factores sugieren que las condiciones de trabajo y el tipo de vinculación laboral juegan un papel importante en la ocurrencia de accidentes.

Hipótesis:

¿Cuáles son las principales características de los accidentes en las dos empresas con mayor número de accidentes



Es probable que las características de los accidentes en las dos empresas más afectadas, como las causas y el tipo de vinculación de los trabajadores, estén relacionadas con los perfiles laborales y el tipo de tareas que realizan, lo que podría influir en la frecuencia y naturaleza de los accidentes. }



```
<!--Simulacro proyecto final-->
```

# Día 2: Análisis y Visualización de los Datos {

<Por=Grupo de trabajo # 3

}



Tareas del día {

01

Análisis de  
los datos

02

Visualización  
de los datos

03

Hypothesis  
testing

04

SQL y NoSQL

05

Documentación  
y Reporte

06

Entregable  
del día

}

1. ¿Cuál es la distribución anual de los accidentes eléctricos?

2. ¿Qué departamentos y municipios registran más accidentes?

3. ¿Cuál es la causa más frecuente de los accidentes eléctricos?

4. ¿Qué tipo de lesión ocurre con mayor frecuencia?

5. ¿Existe relación entre la vinculación laboral y el tipo de lesión?

6. ¿Las fallas de mantenimiento o de operación son la causa principal de muertes?

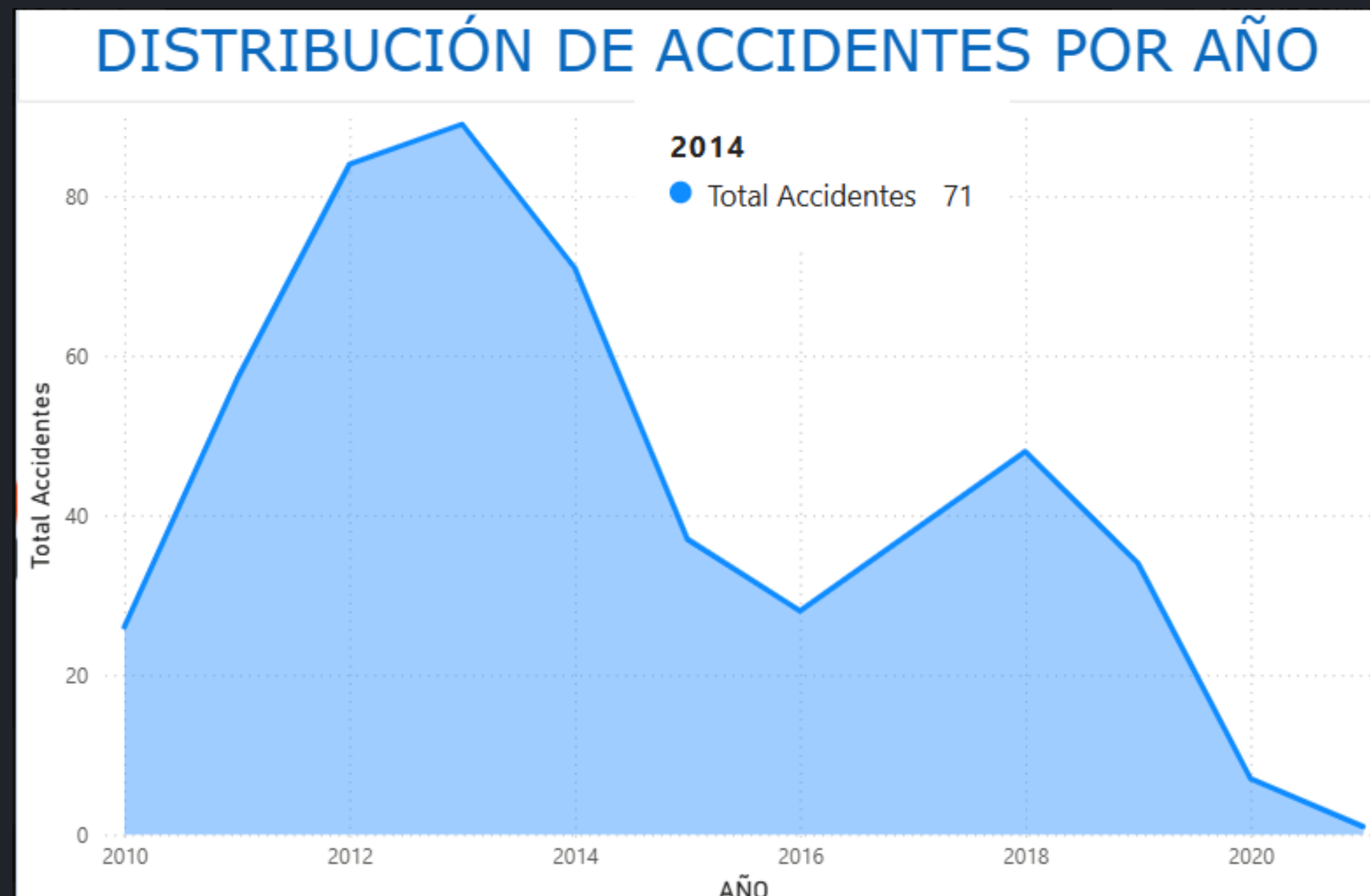
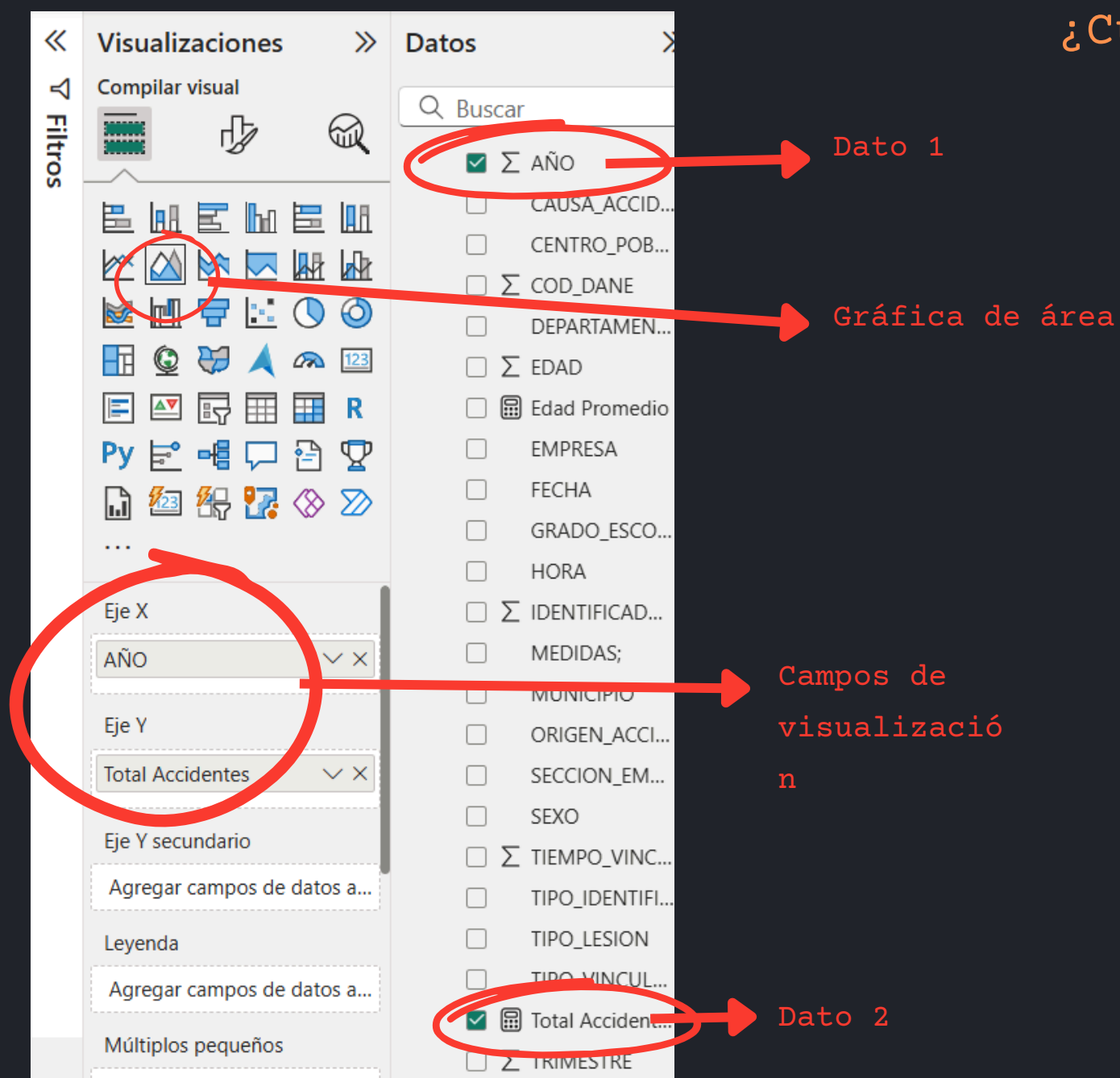
7. ¿Qué empresas presentan más accidentes?

}

# Visualización de datos {

# Power BI

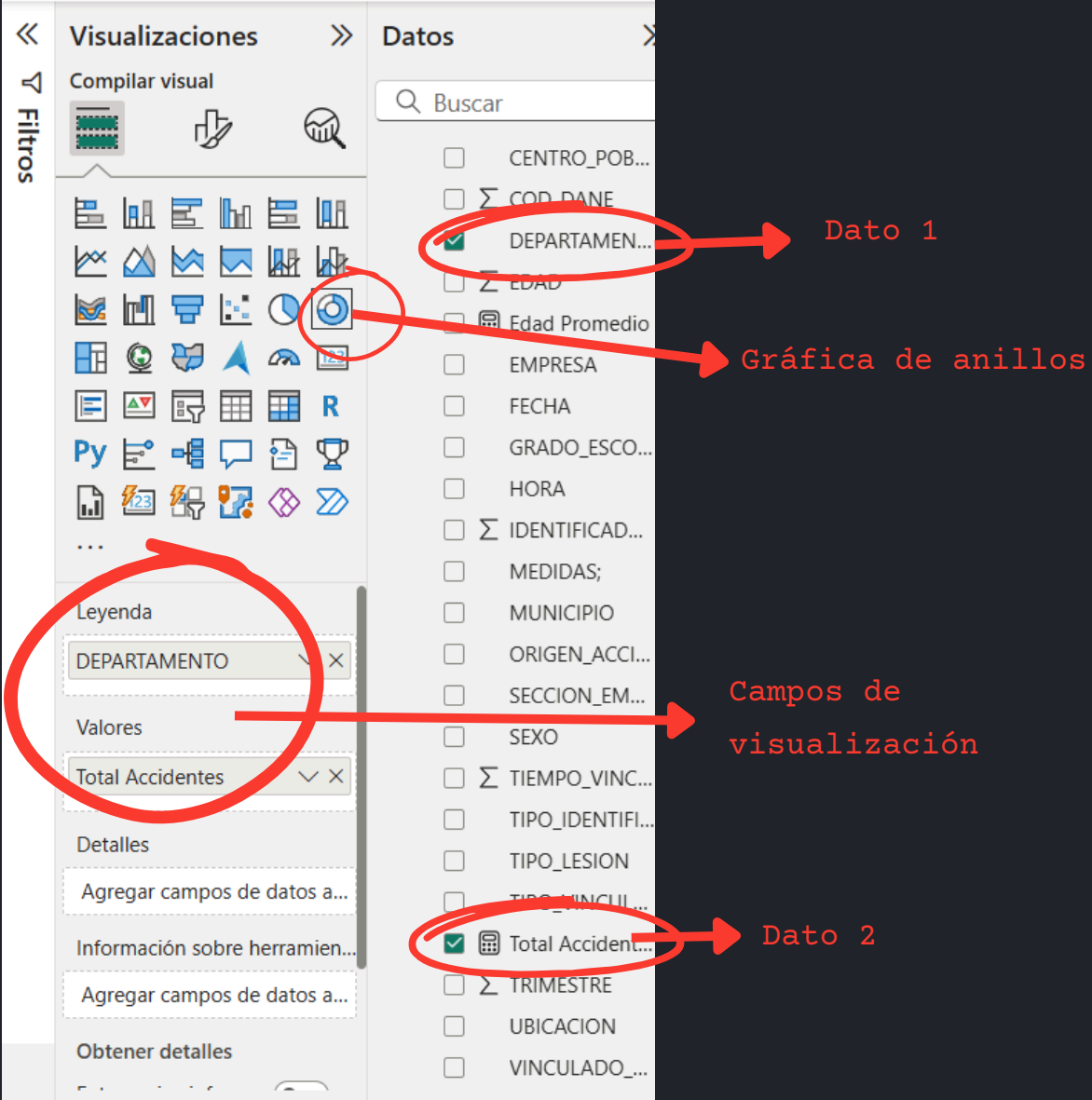
¿Cuál es la distribución anual de los accidentes eléctricos?



La gráfica presenta cómo se distribuyen los accidentes eléctricos por año entre 2010 y 2021. Se observa que el mayor número de casos se registró en 2013, superando los 90 incidentes. A partir de 2015, la tendencia muestra una disminución más notable, llegando a cifras considerablemente bajas en los años 2020 y 2021. }

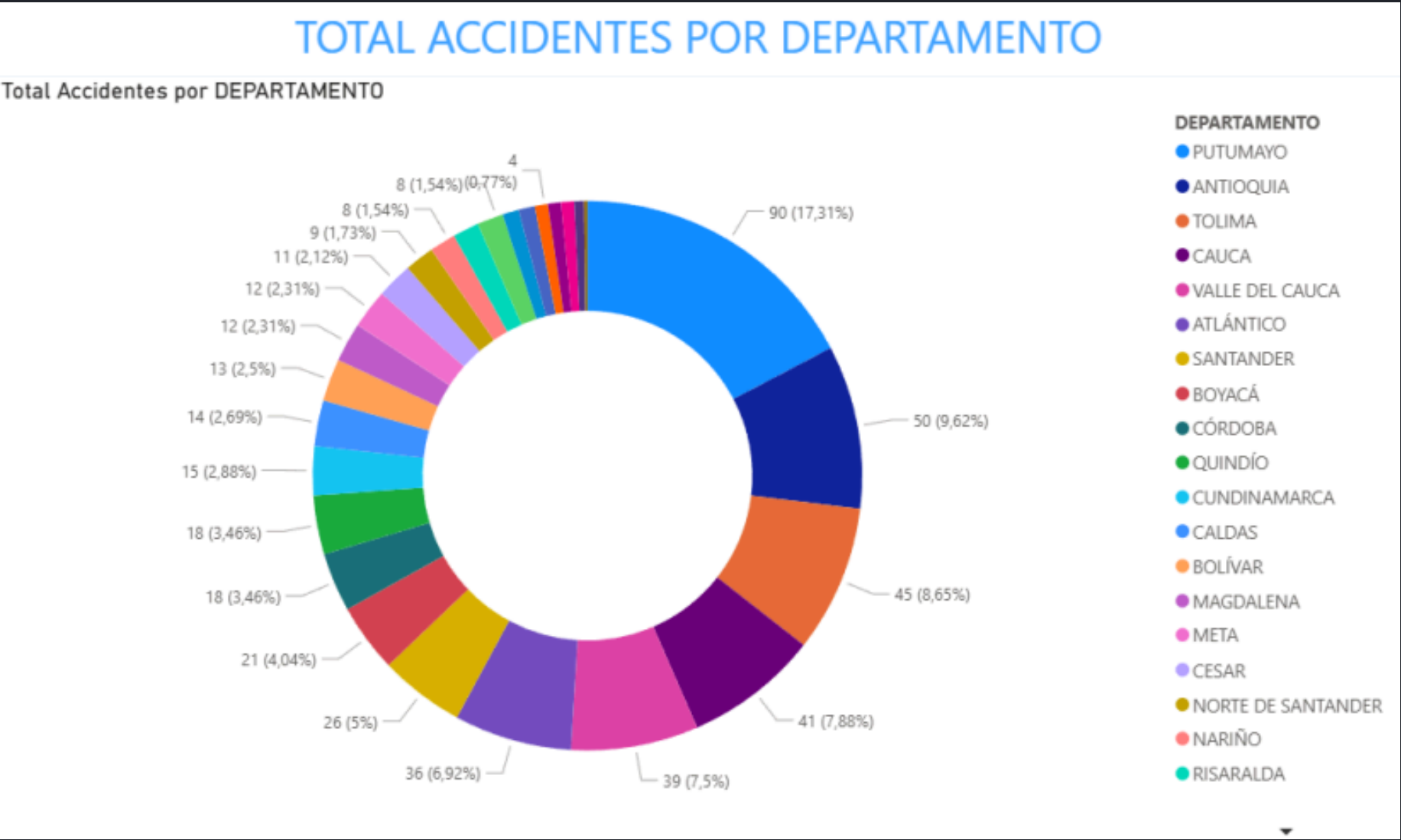
# Visualización de datos {

# Power BI



¿Qué departamentos y municipios registran más accidentes?

- Putumayo concentra el mayor número de accidentes, con 90 casos (17,31%), lo que lo posiciona como el departamento con mayor incidencia.
- Le siguen Antioquia (50 accidentes, 9,62%), Tolima (45 accidentes, 8,65%) y Cauca (41 accidentes, 7,88%).



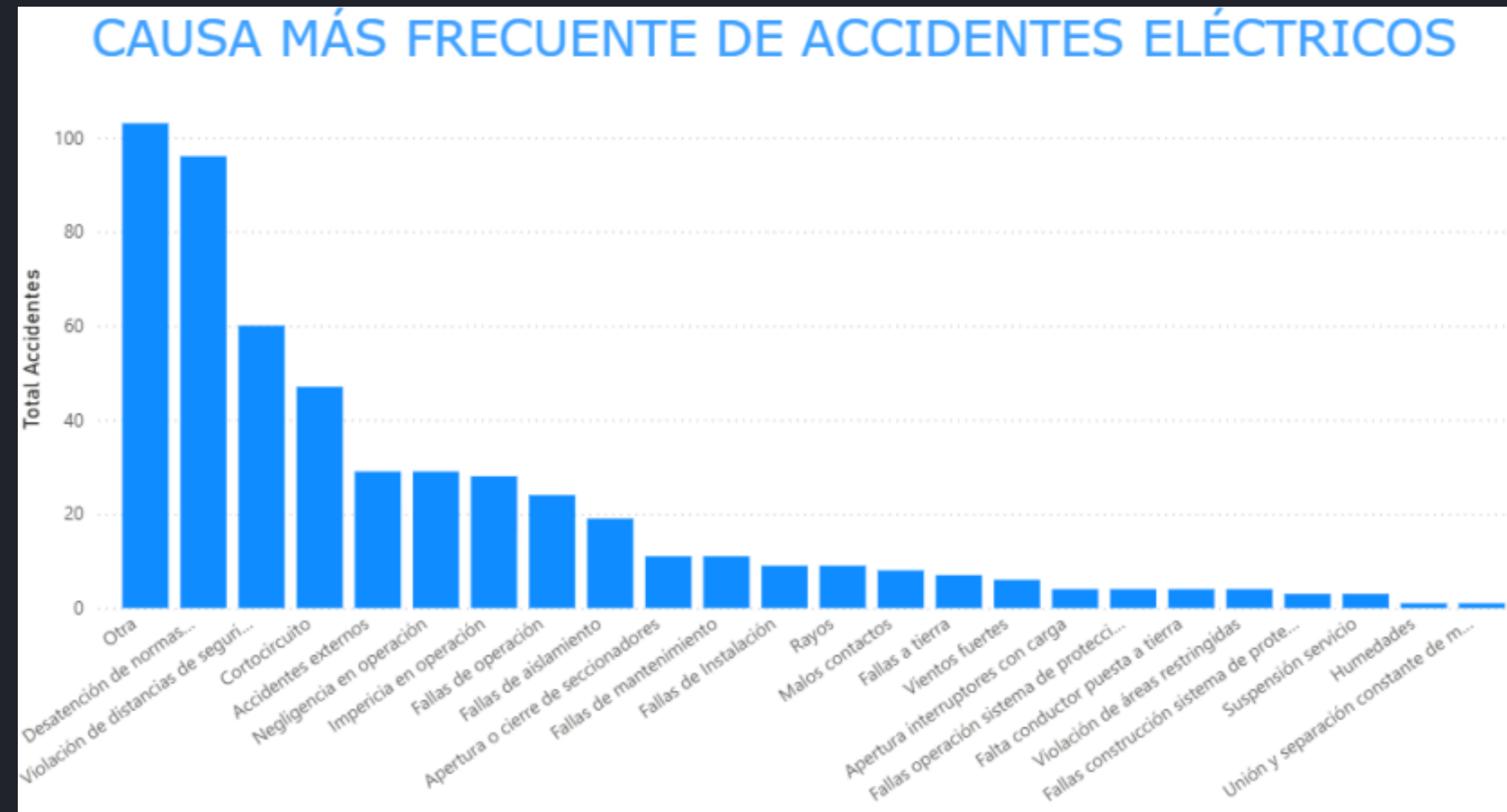
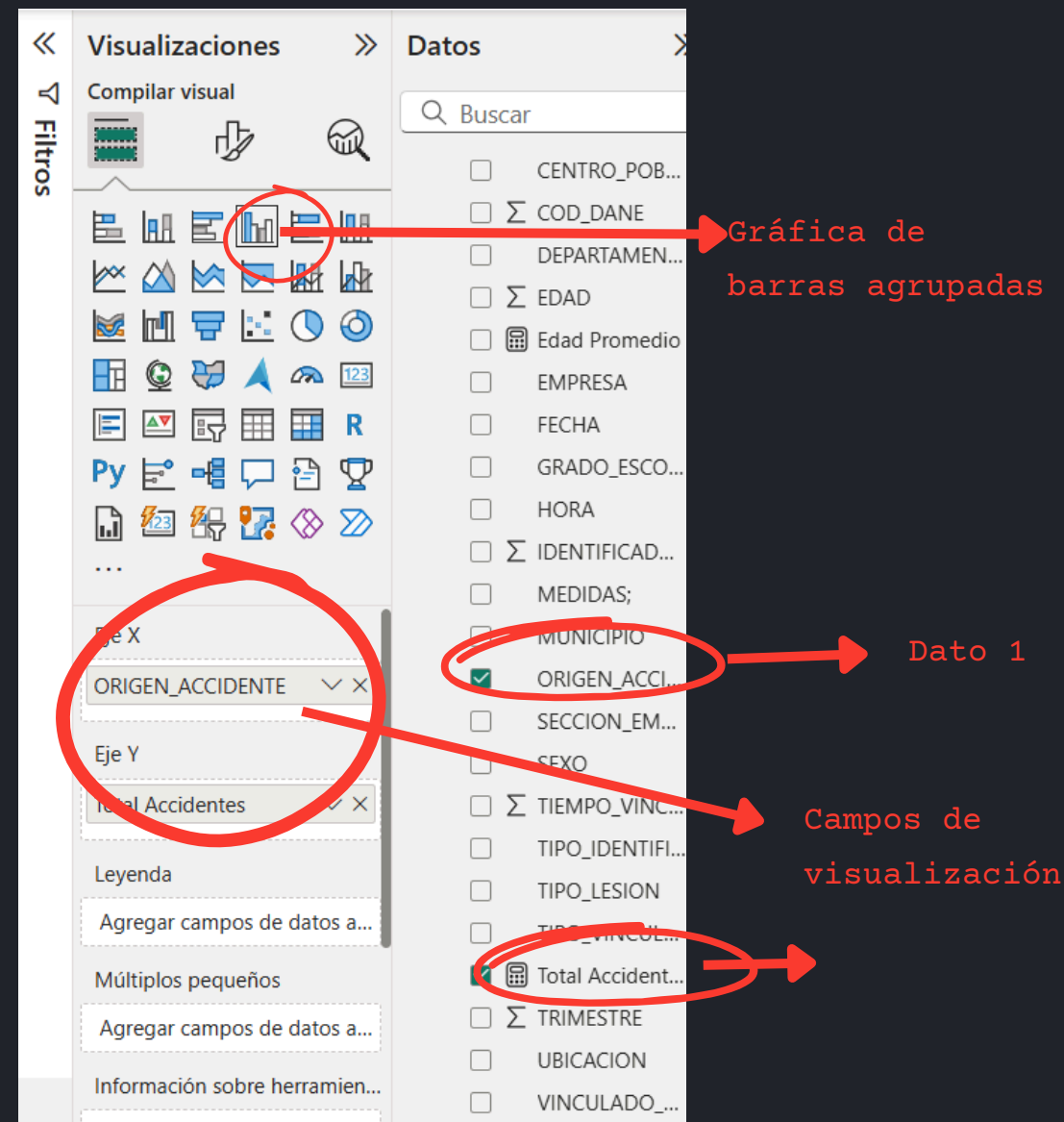
- La distribución no es uniforme. Algunos departamentos tienen una carga desproporcionadamente alta de accidentes en comparación con otros, lo que podría reflejar diferencias en factores
- Nivel de tránsito o actividades industriales

}

# Visualización de datos {

# Power BI

¿Cuál es la causa más frecuente de los accidentes eléctricos?



- Por esta grafica podemos observar una desatención de normas y violación de distancias de seguridad son las siguientes causas más frecuentes, con cifras cercanas a los 100 accidentes, lo que indica problemas de falta de cumplimiento de procedimientos y normas.
- También resaltan el contacto eléctrico, accidentado externo, y negligencia en operación, lo que señala fallas humanas como una causa predominante.

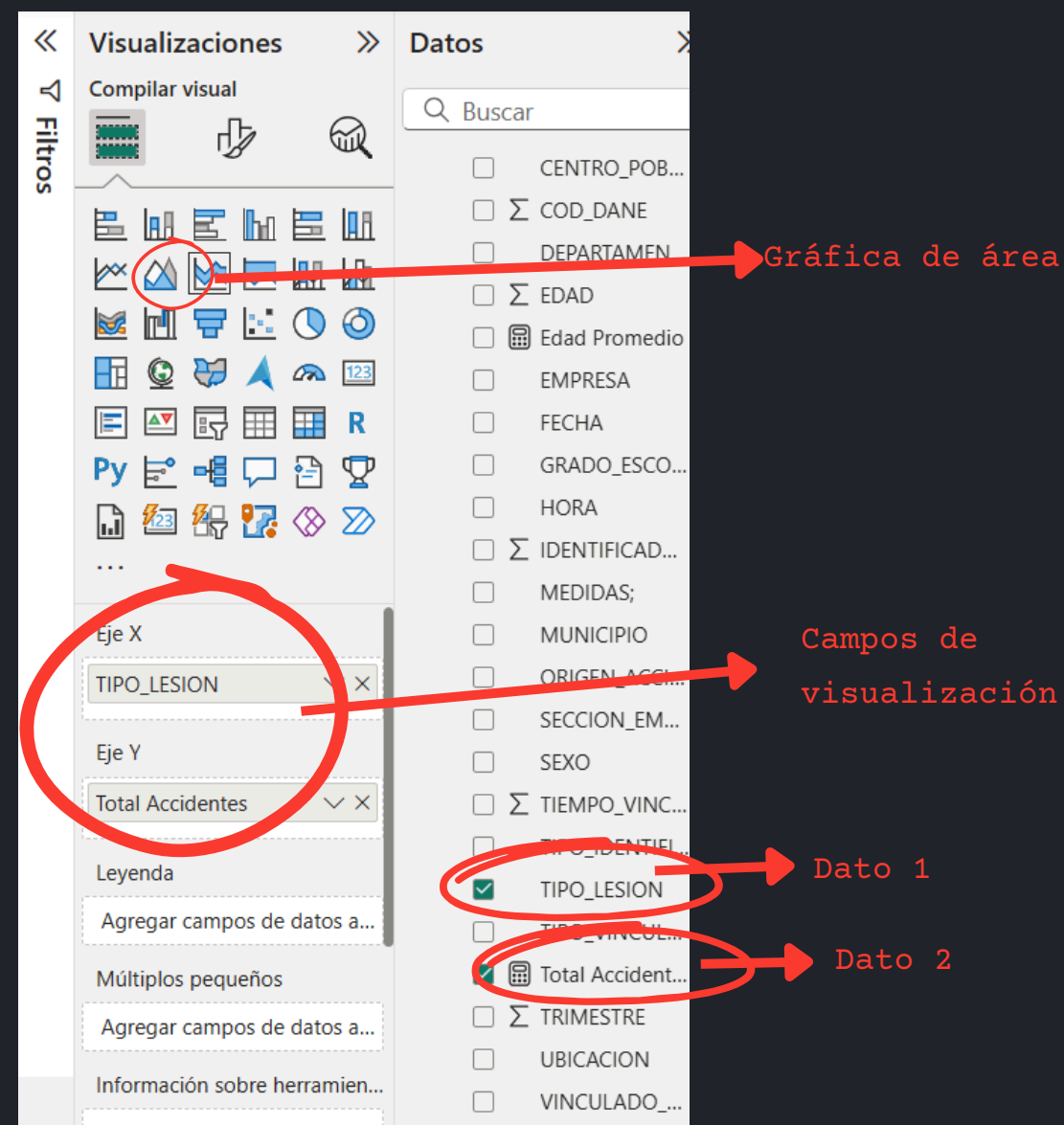
}



# Visualización de datos {

# Power BI

¿Qué tipo de lesión ocurre con mayor frecuencia?



Las quemaduras son la lesión más frecuente en accidentes eléctricos, lo que indica que la exposición a altas temperaturas o arcos eléctricos es el principal riesgo físico. Además, la existencia de un número importante de muertes y traumatismos evidencia que estos accidentes pueden ser altamente graves o incluso fatales. }

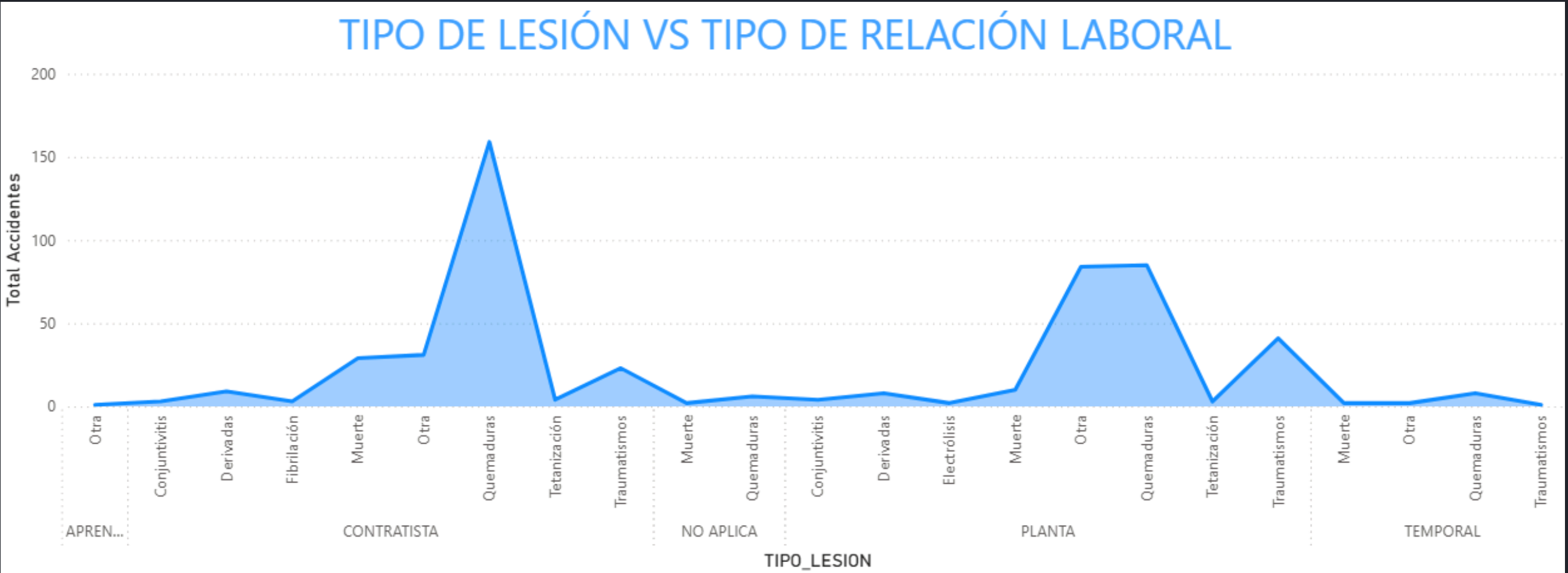
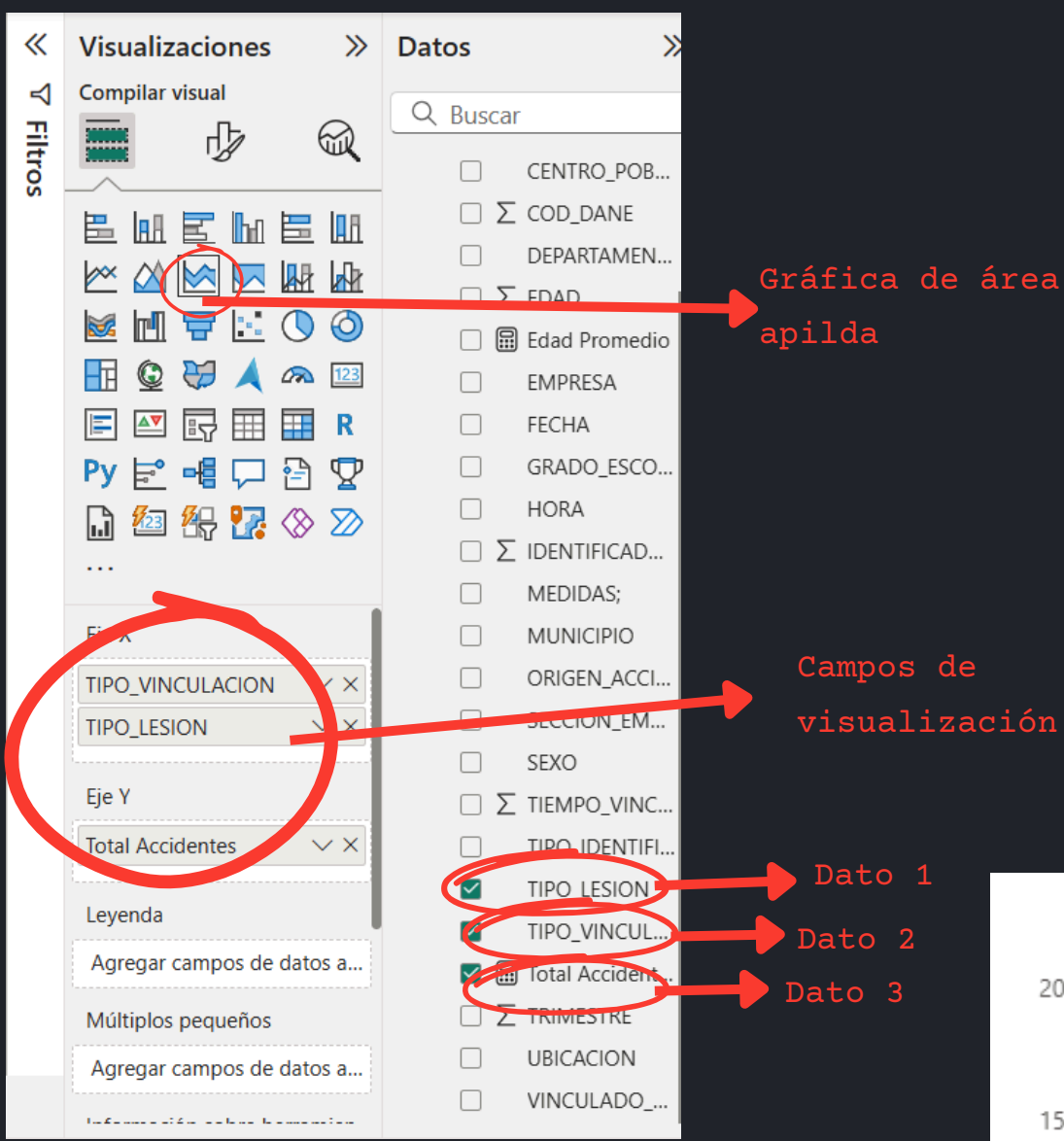


# Visualización de datos {

# Power BI

¿Existe relación entre la vinculación laboral y el tipo de lesión?

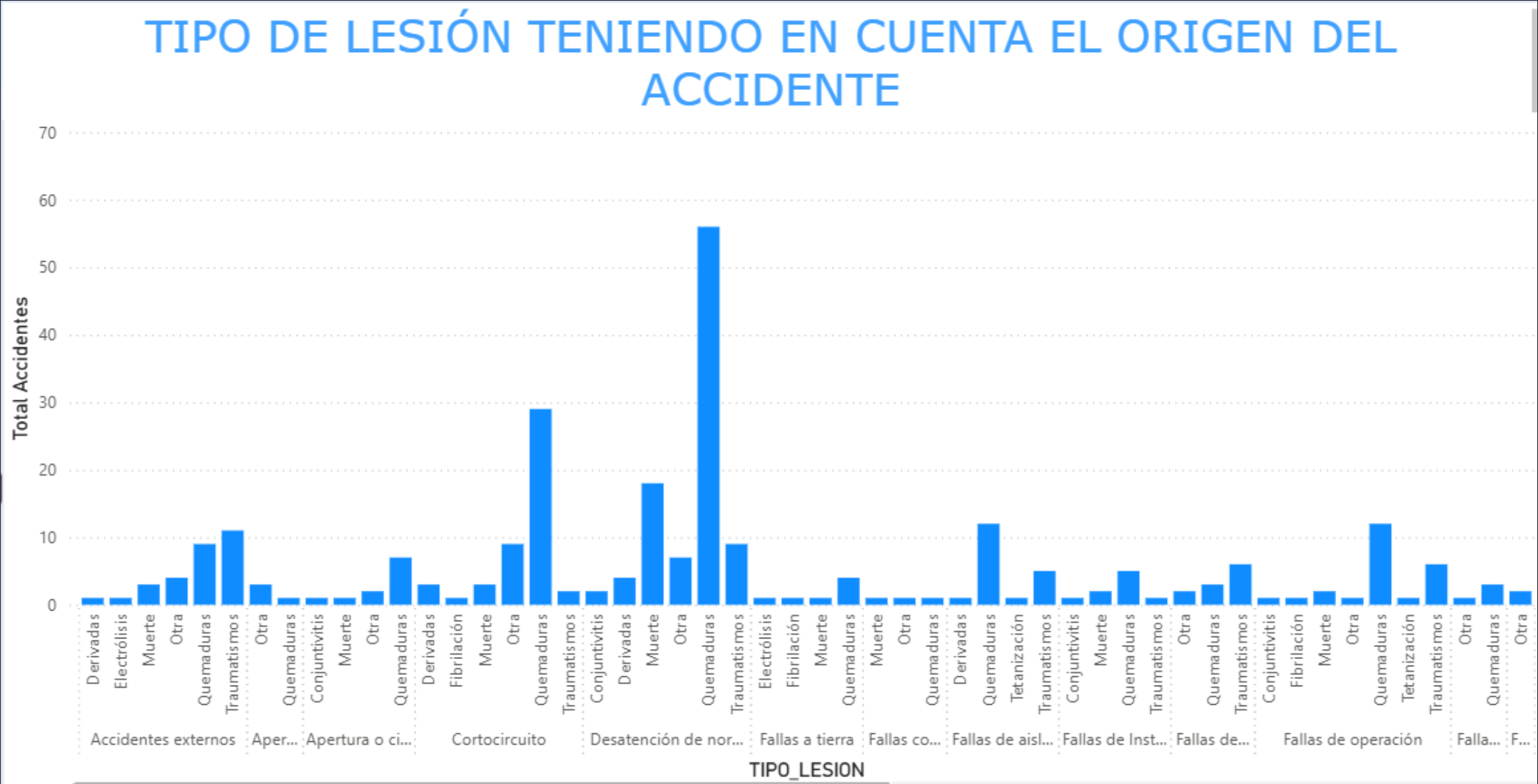
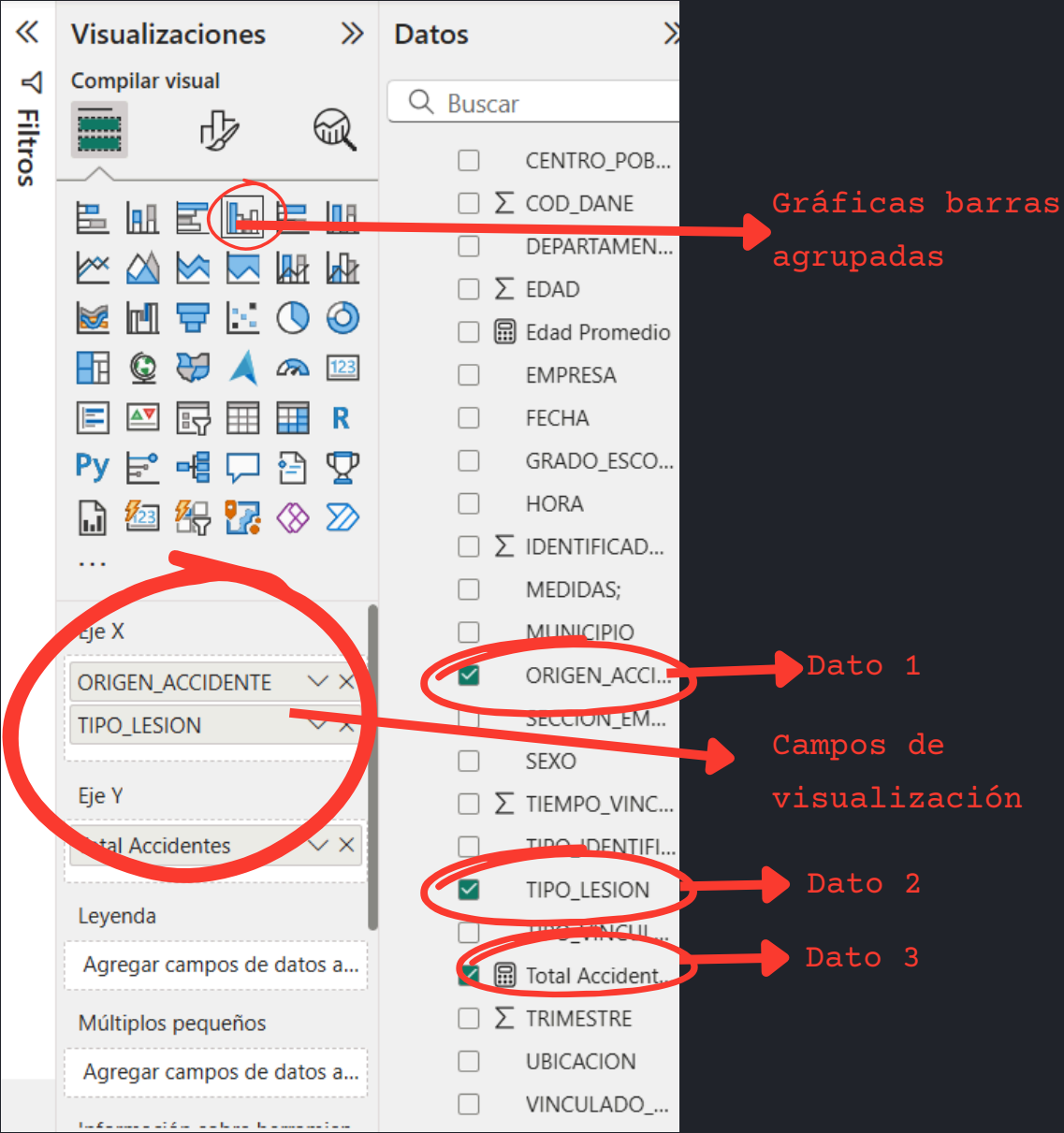
Los trabajadores contratistas son el grupo con mayor cantidad de lesiones, especialmente quemaduras, seguidos por los trabajadores de planta. Esto indica que los riesgos eléctricos afectan más intensamente a quienes tienen vínculos laborales externos o están expuestos a trabajos operativos.



# Visualización de datos {

# Power BI

¿Las fallas de mantenimiento o de operación son la causa principal de muertes?

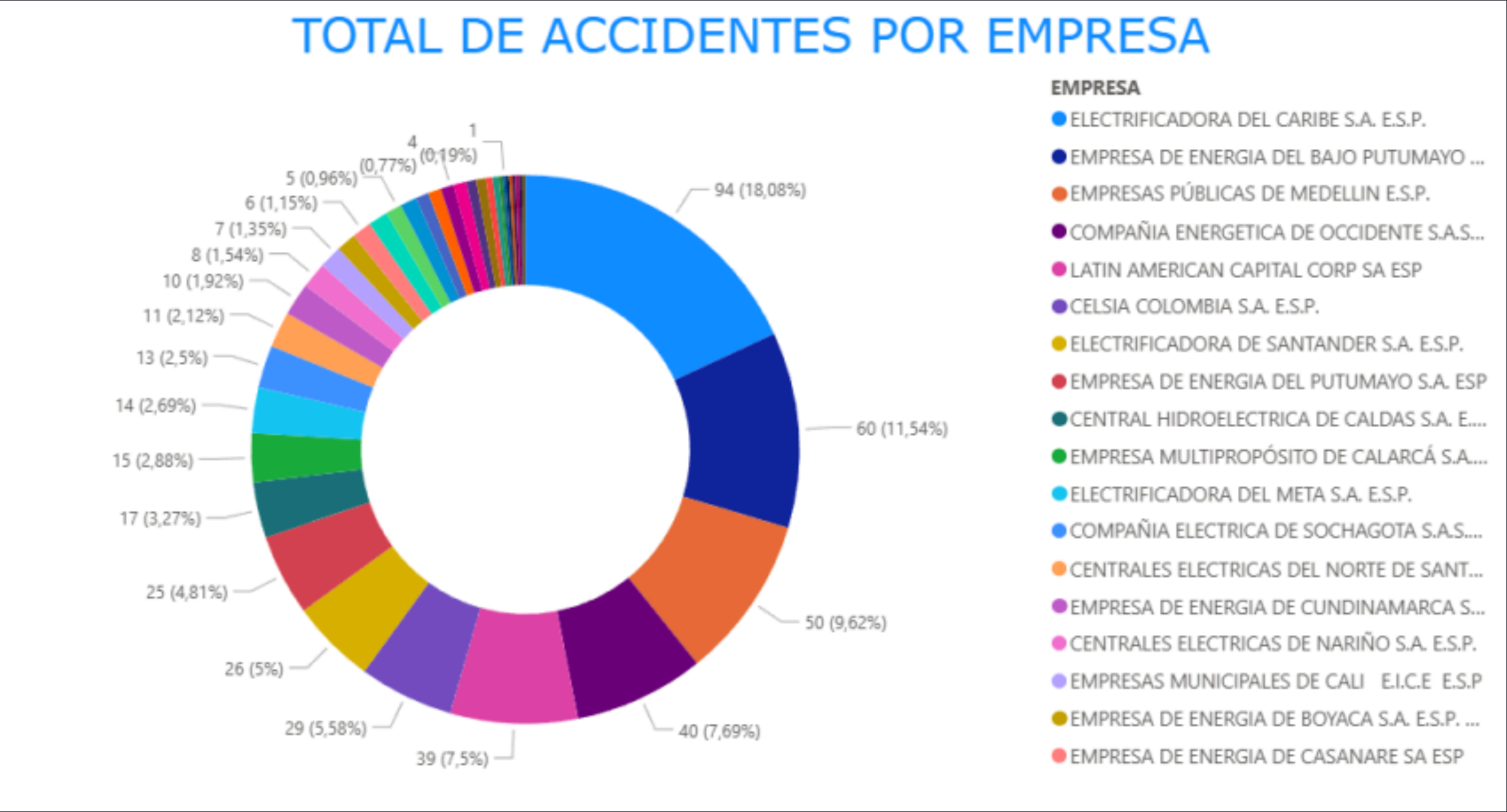
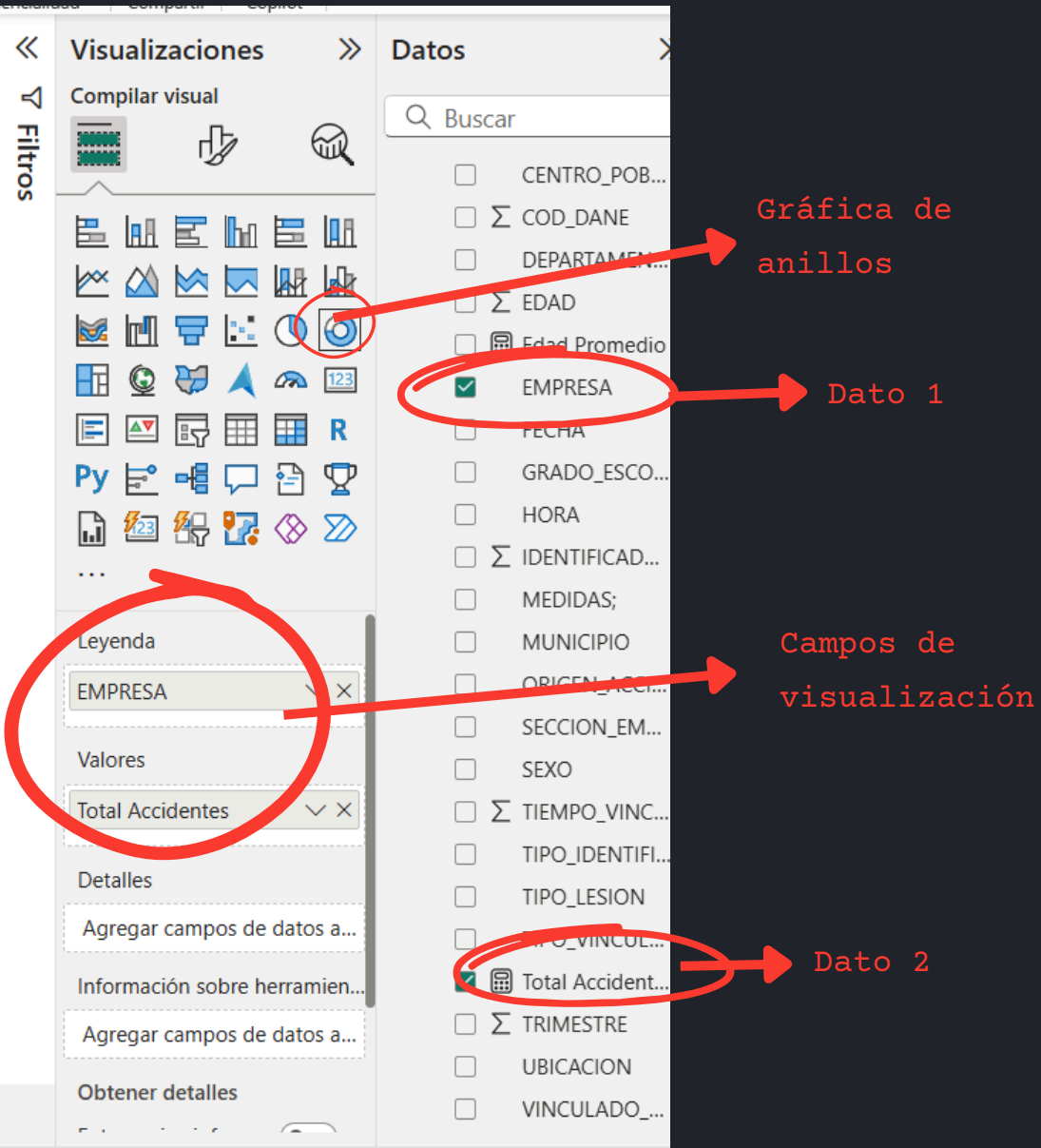


Las quemaduras son el tipo de lesión más común y están fuertemente asociadas a accidentes causados por desatención de normas y contacto directo. Esto evidencia que las fallas humanas y la falta de cumplimiento de protocolos de seguridad son factores determinantes en la ocurrencia de accidentes eléctricos graves, por lo que es urgente fortalecer la capacitación, supervisión y cumplimiento normativo en ambientes laborales eléctricos. }

# Visualización de datos {

# Power BI

¿Qué empresas presentan más accidentes?



La gráfica evidencia que una minoría de empresas concentra la mayoría de los accidentes eléctricos, siendo ELECTRO COSTA S.A. E.S.P. la que presenta el mayor número, con un 17,31% del total, seguida por ELECTRIFICADORA DEL CARIBE S.A. E.S.P. y EMPRESAS PÚBLICAS DE MEDELLÍN E.S.P. Esto muestra posibles deficiencias en la gestión de riesgos eléctricos en estas organizaciones o una mayor exposición debido a su tamaño o cobertura geográfica. Por otro lado, muchas empresas reportan pocos accidentes, lo que podría indicar mejores prácticas de seguridad o menor nivel de exposición.

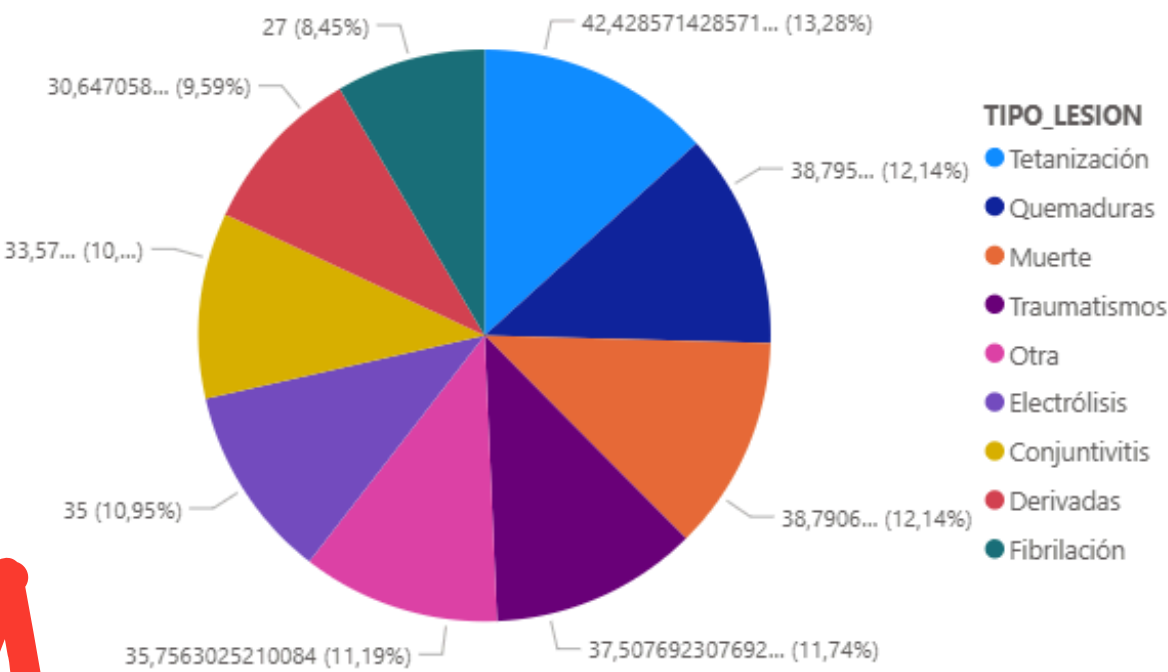
}



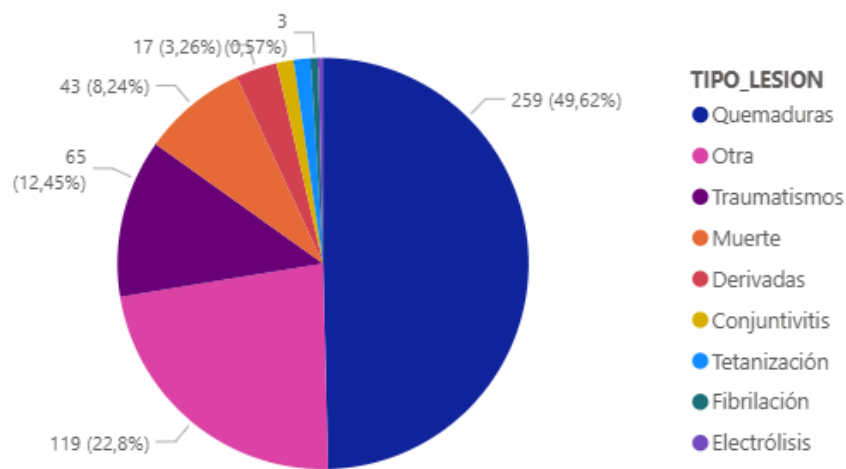
Las edades promedio para los tipos de lesiones son similares, por lo que la edad no parece ser un factor decisivo en el tipo de lesión en los accidentes analizados. Las medidas de seguridad deberían enfocarse en toda la población trabajadora, sin distinciones por grupos de edad. Es importante considerar la frecuencia de cada tipo de lesión. Aunque la tetanización tiene la edad promedio más alta, su baja incidencia sugiere que la prevención debe centrarse en lesiones más comunes, como las quemaduras, donde la edad promedio también es alta.



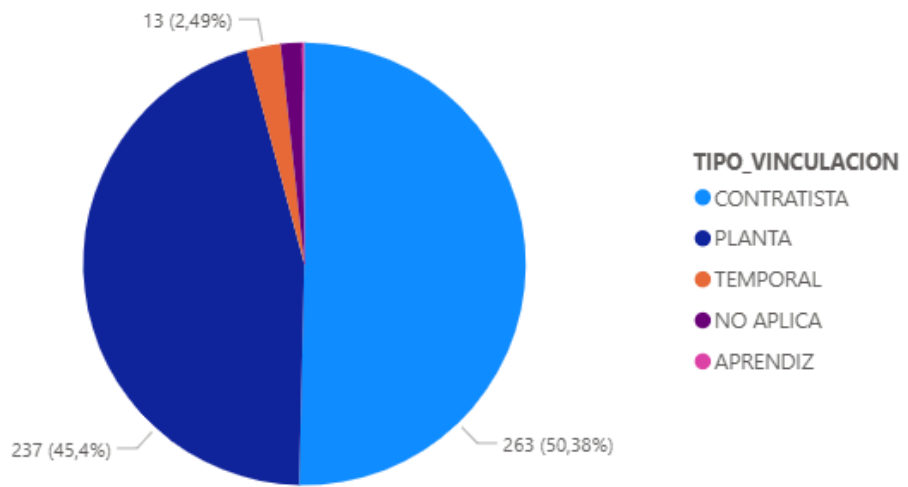
Las quemaduras son el tipo de lesión más común en los accidentes registrados, seguidas por otras lesiones y traumatismos. Es importante notar la presencia de casos fatales, aunque en menor proporción.



Total Accidentes por TIPO\_LESION



Total Accidentes por TIPO\_VINCULACION



```
<!--Simulacro proyecto final-->
```

# Día 3: Conclusiones

<Por=Grupo de trabajo # 3

}



Aplicación de metodologías de análisis de datos sobre la base de datos Superservicios-Información de Accidentes de Origen Eléctrico (2010-2021) como diagnóstico de las causas de los accidentes laborales en el sector eléctrico colombiano. {

## # Conclusiones

- El **Riesgo eléctrico** es una de las principales causas de accidentalidad en el sector eléctrico colombiano, afectando tanto a trabajadores como a personas ajenas a las empresas prestadoras del servicio
- Los reportes incluyen incidentes con consecuencias directas o indirectas por contacto con cargas eléctricas, evidenciando un alto nivel de exposición al peligro.
- Según la normativa actual en **Seguridad y Salud en el Trabajo (SST)**, es fundamental estudiar a fondo cada caso reportado para entender su causa y gravedad
- Muchos accidentes ocurren cuando el personal capacitado **no sigue adecuadamente los protocolos** o **subestima el nivel de riesgo** de sus funciones
- Este análisis de datos permite identificar los factores de riesgo y sus causas, comprender las consecuencias más comunes de los accidentes eléctricos y proponer acciones correctivas y preventivas que ayuden a reducir su ocurrencia en el sector.



}

## # Conclusiones

Las estadísticas del sector eléctrico en Colombia muestran que el riesgo eléctrico tiene altos índices de accidentalidad. Los reportes incluyen tanto a empleados de empresas energéticas como a personas afectadas por cargas eléctricas.

Según la legislación de Seguridad y Salud en el Trabajo (SST), es esencial examinar estos casos para proteger a quienes trabajan con electricidad. Los accidentes pueden ser graves debido a la falta de precaución o incumplimiento de normas. A menudo, los trabajadores no son conscientes del riesgo.

El análisis de estos datos identificará factores de riesgo y acciones para prevenir o reducir accidentes eléctricos, ayudando a implementar mejoras.





## # Conclusiones

### Aspectos de mejora {

Durante el análisis de la base de datos Superservicios-Información\_de\_Accidentes\_de\_Origen\_Eléctrico-Formato19\_20250507, se observa que el tipo de vinculación no es relevante para las entidades proveedoras del servicio eléctrico desde una perspectiva legal, siempre que el personal cumpla con las normativas de seguridad social y formación necesarias. Las causas comunes de accidentes incluyen la violación de distancias de seguridad respecto a las redes eléctricas y la falta de autocuidado, manifestada como exceso de confianza. Estas causas son importantes en los informes, ya que influyen en la descripción del origen de cada accidente. Con base en estas observaciones, se identifican fallos operativos y de control, por lo que se proporcionarán sugerencias para mitigar los riesgos asociados a los accidentes eléctricos.

}

# Oportunidad de mejora.

## # Conclusiones



- **Capacitar periódicamente al personal operativo**, reforzando y actualizando sus conocimientos con un enfoque en la prevención para minimizar los accidentes laborales.
- **\*\*Divulgación de riesgos\*\***: Las empresas deben informar sobre los riesgos a los que están expuestos los trabajadores y proporcionar un historial de incidentes para fomentar una cultura de prevención entre el personal.
- **\*\*Revisión y mantenimiento\*\***: Es fundamental revisar y mantener los equipos de protección personal (EPP), así como las herramientas y maquinaria pesada. Detectar y corregir fallas puede prevenir accidentes y salvar vidas.
- **\*\*Seguimiento de actividades\*\***: Llevar a cabo auditorías, controlar las funciones del personal técnico y asegurar condiciones de trabajo óptimas.
- **\*\*Ajuste de turnos\*\***: Modificar los turnos de trabajo para garantizar un descanso adecuado, evitando que se vean afectadas la capacidad física e intelectual de los empleados.



## # Conclusiones

### # Concienciación sobre la Seguridad en Actividades de Alto Riesgo

Es vital que el personal operativo reconozca los riesgos de sus actividades. Ignorar los protocolos de seguridad puede causar accidentes.

**\*\*Protección Eléctrica\*\***: Proteger al trabajador, al sistema eléctrico y a los usuarios es esencial.

**\*\*Normativas de Seguridad\*\***: Implementar normativas adecuadas es fundamental para la seguridad de las instalaciones eléctricas.

**\*\*Evaluación del Riesgo\*\***: Realizar una evaluación precisa del riesgo y señalar sus fuentes es indispensable para prevenir accidentes.



## # Conclusiones

- Proponer mejoras adecuadas al sistema de seguridad que se ajusten a las necesidades existentes.
- Implementar el uso obligatorio de los EPP (equipos de protección personal) como palitos calientes, guantes, zapatos, trajes y cascos, sin excepciones y con tolerancia cero.
- Crear equipos de inspección dedicados a supervisar las condiciones laborales de los equipos técnicos y verificar el cumplimiento de las secuencias de seguridad.
- Ante una implementación incorrecta o prácticas inadecuadas, se deben imponer sanciones y requerir la asistencia a seminarios educativos obligatorios.



# Conclusiones {

## # Identificación de Errores Operativos

Durante la ejecución de actividades, se han detectado errores operativos, como el uso inadecuado de los Equipos de Protección Personal (EPP) y el incumplimiento de los protocolos de seguridad. La planificación es crucial para una ejecución correcta. Cumplir con las normas es esencial, pero aún más importante es sensibilizar a los trabajadores y al público sobre la prevención de riesgos.

## ## Puntos Clave

**\*\*Información a los Trabajadores\*\*:** Es vital informar a todos los empleados sobre los peligros de trabajar en condiciones de alto voltaje.

**\*\*Atención a Trabajadores con Experiencia\*\*:** Se debe prestar especial atención a los empleados más veteranos, ya que podrían subestimar los riesgos y volverse insensibles al peligro. }



## Conclusiones {

La experiencia, en ocasiones, puede llevar a descuidar un procedimiento operativo adecuado, lo que disminuye la atención y el cuidado necesarios. Esto incrementa la confianza en la capacidad para enfrentar cualquier situación, sin considerar las posibles repercusiones, lo cual es un grave error.

- \*Consecuencias de un accidente eléctrico:\*\*
- No se pueden evaluar de inmediato.
- Pueden resultar en un incidente leve o incluso en una fatalidad.
- Las circunstancias siempre son diferentes.

Por lo tanto, es esencial adherirse a las normas establecidas sin omitir ningún procedimiento.

}

Gracias {



}