

# ANÁLISIS GEOESPACIAL DE SINIESTROS

---

Gustavo Benitez

# Definición de Objetivos y Requerimientos

- El objetivo de este proyecto es evaluar el riesgo que presenta cada ruta. Para hacerlo, primero hay que definir la palabra **riesgo**.
- Podemos usar la siguiente definición: El **riesgo** se define como la combinación de la probabilidad de que se produzca un evento y sus consecuencias negativas.
- En este caso específico, considero que el riesgo puede ser dividido en 2 subcategorías: el riesgo a los **recursos** de la empresa, como sus vehículos o su dinero, y el riesgo a la **integridad física y emocional** de las personas.
- El riesgo a los recursos de la empresa puede ser calculado utilizando métricas que están presentes en el set de datos, como la cantidad de siniestros en una ruta y el monto perdido por siniestro.
- El riesgo a las personas es más difícil de medir, ya que no existen métricas predefinidas para hacerlo. Sin embargo, podemos asignar valores numéricos que nos ayuden a ponderar los riesgos que presenta cada siniestro. Estos valores pueden ser ajustados según su importancia subjetiva.
- El objetivo de este ejercicio es la definición y el cálculo de un **Índice de Riesgo** para cada ruta, que tome en cuenta tanto métricas cuantificables como variables subjetivas.

# Comprensión de los Datos

Hoja	Descripción	Columna	Explicación columna
Instrucciones	Contiene los requisitos de este ejercicio	NA	NA
Glosario	Contiene el glosario de la información a tratar	Hoja	Nombre de la hoja
		Descripcion	Breve resumen del contenido de la hoja
		Columna	Nombre de una columna dentro de la hoja
		Explicación columna	Breve explicación de la columna dentro de la hoja
PuntoPartida	Contiene la informacion de los puntos de partida de cada ruta	IDRuta	Es un identificador de la ruta a realizar
		NombreRuta	Contiene el tipo de ruta al que se refiere
		TipoUbicacion	Contiene si es una tienda o una bodega
		Empresa	Contiene el nombre de la empresa a la que pertenece la ruta
		HoraProgramadaSalida	Hora que se tiene como estimada para la salida de la ruta
		Longitud	Coordenada geográfica de longitud
		Latitud	Coordenada geográfica de latitud
		ValorMercancia	Monto que tiene la mercancia a mover en la ruta
PuntoLlegada	Contiene la informacion de los puntos de llegada de cada ruta	IDConductor	Identificador del conductor de la ruta
		IDRuta	Es un identificador de la ruta a realizar
		NombreRuta	Contiene el tipo de ruta al que se refiere
		TipoUbicacion	Contiene si es una tienda o una bodega
		Empresa	Contiene el nombre de la empresa a la que pertenece la ruta
		HoraSalida	Hora de salida de la ruta
		Longitud	Coordenada geográfica de longitud
		Latitud	Coordenada geográfica de latitud
Siniestros	Contiene la información de los siniestros ocurridos y registrados	ValorMercancia	Monto que tiene la mercancia a mover en la ruta
		IDConductor	Identificador del conductor de la ruta
		FolioSiniestro	Numero de folio del evento ocurrido
		Evento	Tipo de siniestro que ocurrio
		Subtipo	Puede o no tener valores, se refiere en robos al tipo sin violencia o con violencia
		UnidadOperativa	Se refiere a si fue a una tienda o a un vehiculo
		IDTienda	Identificador de la tienda siniestrada
		IDVehiculo	Identificador del vehiculo siniestrado
		ValorPerdido	Monto de los daños por el siniestro
		Latitud	Coordenada geográfica de latitud
		Longitud	Coordenada geográfica de longitud

La imagen muestra un glosario usado para explicar las distintas tablas del set de datos.

Mi primer instinto fue analizar el glosario para identificar las relaciones entre las tablas. Identifiqué que las tablas PuntoPartida y PuntoLlegada pueden ser usadas para crear una tabla de **Rutas**, que será de utilidad en el análisis.

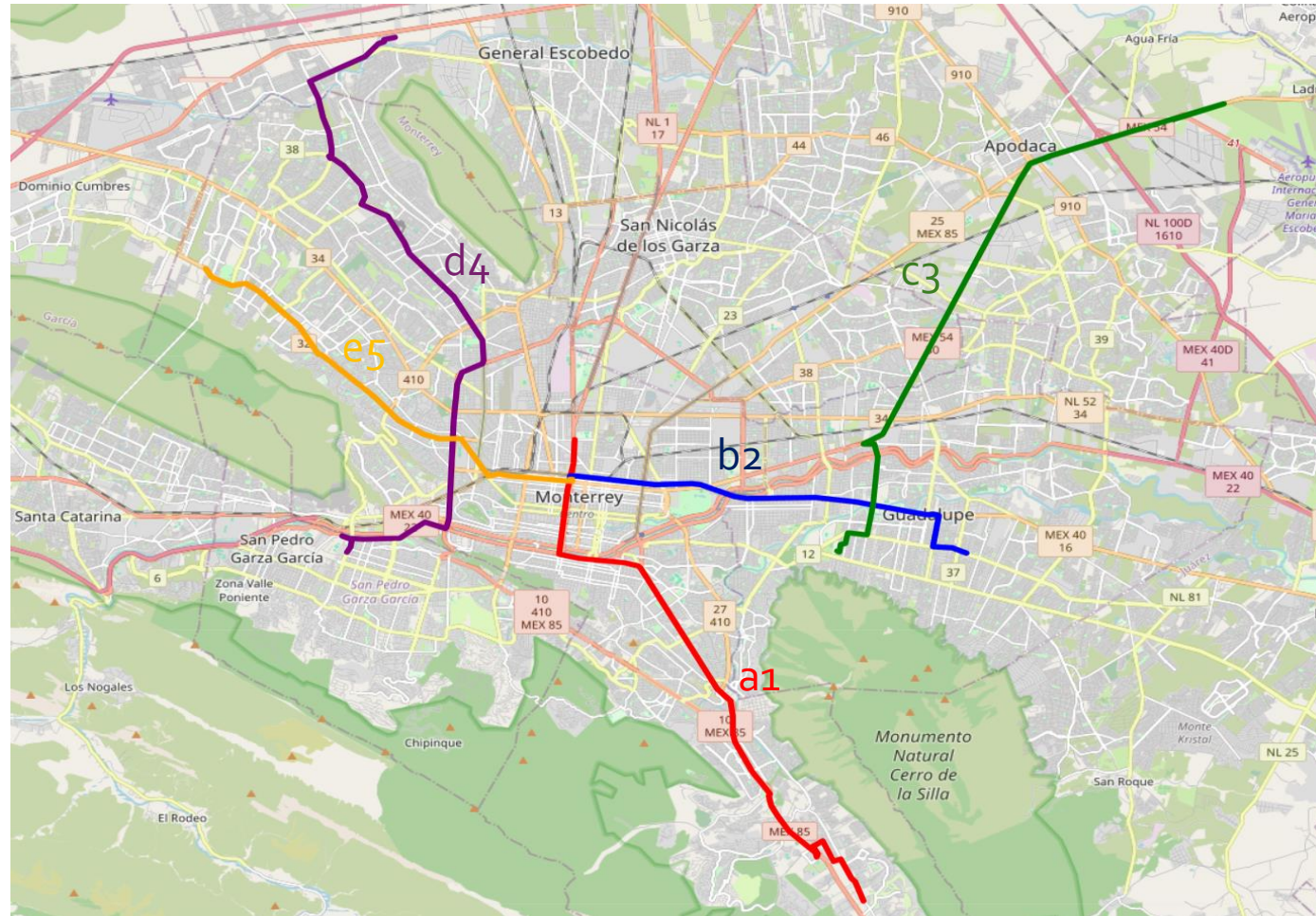
# Limpieza y Preparación de los datos

- Los datos estaban relativamente limpios. Identifiqué un valor que no estaba en el formato correcto, y lo corregí.
- Existen datos nulos en la tabla de siniestros, pero decidí que estos nulos no eran impedimentos para mi análisis, por lo que no fueron modificados.
- Utilicé la librería OSMnx para extraer datos de OpenStreetMap, una plataforma de código abierto que cuenta con una base de datos geográfica.
- Estos datos me permitieron hacer el cálculo de las rutas óptimas, junto con sus respectivas distancias y sus tiempos.
- La tabla resultante de Rutas se ve de esta manera:

IDRuta	NombreRuta	TipoUbicacionPartida	EmpresaPartida	HoraProgramadaSalida	LongitudPartida	LatitudPartida	ValorMercancia	IDConductor	TipoUbicacionLlegada	EmpresaLlegada	HoraSalida	LongitudLlegada	LatitudLlegada	DistanciaRutaKilometros	TiempoRutaMinutos	HoraEstimadaLlegada
a1	Relleno de Inventario	Bodega	Empresa 1	10:00:00	-100.316540	25.696300	150000	1	Tienda	Empresa 1	10:03:00	-100.228542	25.550476	20.517	19.483333	10:22:28
b2	Traslado de efectivo	Tienda	Empresa 1	9:00:00	-100.203823	25.663144	300000	1	Banco	Banco 1	08:58:00	-100.318156	25.685163	12.909	13.750000	09:11:45
c3	Relleno de Inventario	Bodega	Empresa 3	14:00:00	-100.145293	25.788839	1500000	1	Tienda	Empresa 3	14:10:00	-100.242414	25.664271	22.584	21.516667	14:31:31
d4	Relleno de Inventario	Bodega	Empresa 2	11:00:00	-100.367606	25.814914	200000	1	Tienda	Empresa 2	11:00:19	-100.381548	25.664200	23.733	25.216667	11:25:32
e5	Traslado de efectivo	Tienda	Empresa 3	15:00:00	-100.421277	25.744508	123000	1	Banco	Banco 1	16:00:00	-100.318156	25.685163	14.047	14.000000	16:14:00

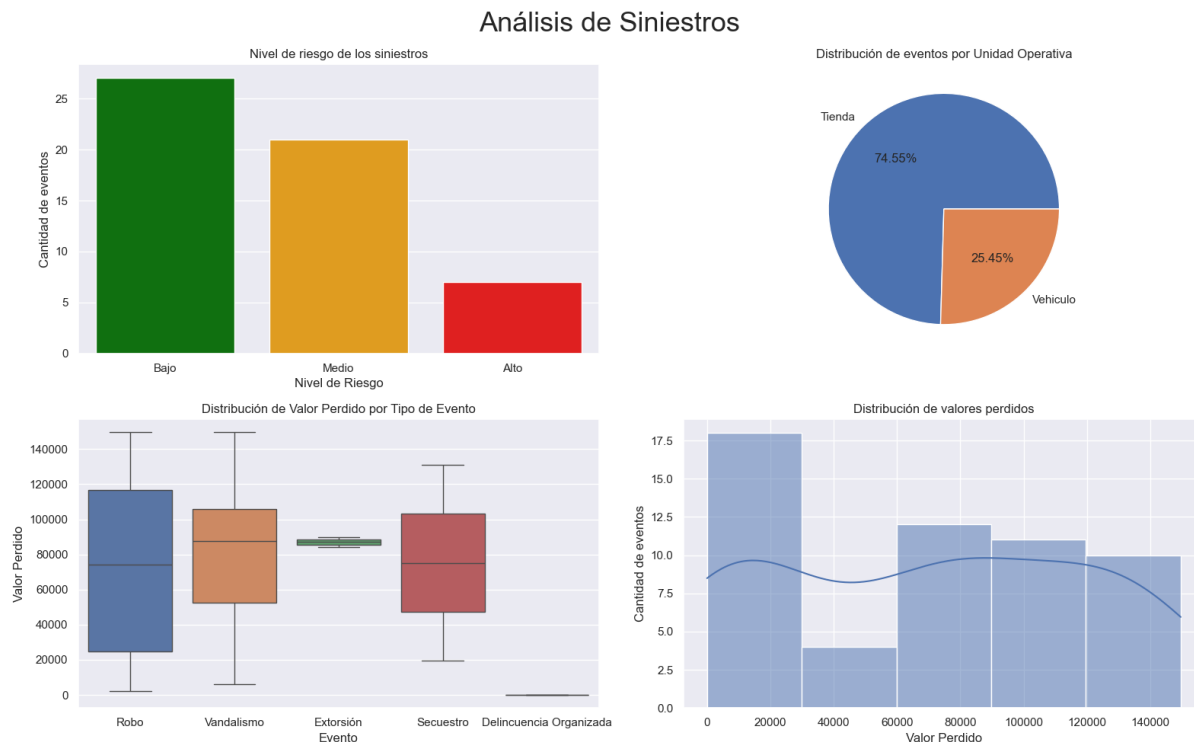
# Análisis Exploratorio de Datos

Con el uso de la librería Folium, logré graficar un mapa con las rutas óptimas extraídas de OSM.



# Análisis Exploratorio de Datos

Adicionalmente, generé este pequeño dashboard para tener un mejor entendimiento de las variables de la tabla Siniestros



## Análisis

**Nivel de Riesgo de Siniestros:** La gran mayoría de los eventos se clasifican como de riesgo bajo o medio.

**Distribución de Eventos por Unidad Operativa:** El 74.55% de los eventos ocurren en tiendas, lo que sugiere que las tiendas son más vulnerables o que hay más recursos en ellas.

**Distribución del Valor Perdido por Tipo de Evento:** Los robos, vandalismos y secuestros tienen una gran variabilidad en el valor perdido. Las consecuencias de la delincuencia organizada no se reflejan en el valor perdido.

**Distribución de Valores Perdidos:** Existe una gran cantidad de eventos en los que el valor perdido fue menor a 20,000 pesos. Hay una disminución en la cantidad de eventos a medida que aumenta el valor perdido, con una fluctuación en el medio.



# Cálculo del Índice de Riesgo

- Esta es una breve descripción de los pasos que tomé para calcular el índice de riesgo para cada ruta.
1. Asigné una ponderación numérica para clasificar el riesgo de cada siniestro. Decidí usar un **2** a los siniestros de riesgo bajo, un **5** para los de riesgo medio, y un **10** para los de riesgo alto. Categoricé a los robos sin violencia y a los vandalismos como siniestros de riesgo bajo, al robo con violencia y a la extorsión como siniestros de riesgo medio, y al secuestro y delincuencia organizada de riesgo alto.
  2. Hice una relación de los siniestros con las rutas. para esto, decidí que una ruta se puede ver afectada por un siniestro que sucede a una distancia aproximada de **650 metros**. Este es un número arbitrario, y lo escogí porque considero que así también se puede tomar en cuenta las variaciones a la ruta que puedan tomar los conductores. Al hacer la relación de esta manera, **un sólo siniestro puede estar relacionado a varias rutas**, si sucede dentro de la intersección de estas. Considero que esto refleja algo que puede suceder en el mundo real.
  3. Calculé la **cantidad de siniestros** asociados a cada ruta, junto con las **sumas de sus ponderaciones y de los valores perdidos**.

# Cálculo del Índice de Riesgo

4. Decidí que Índice de Riesgo se calcularía de la siguiente manera: La suma ponderada del total de siniestros, más la suma del valor perdido en cientos de miles de pesos.

$$\text{IndiceRiesgo} = \text{SumaSiniestrosPonderada} + (\text{SumaValorPerdido} / 100,000)$$

5. Adicionalmente, clasifiqué los índices de riesgo resultantes en 3 categorías:

- Riesgo Bajo – Riesgo en el percentil 20 o inferior
- Riesgo Medio – Riesgo en el percentil 60 o inferior, pero superior a 20
- Riesgo Alto – Riesgo arriba del percentil 60

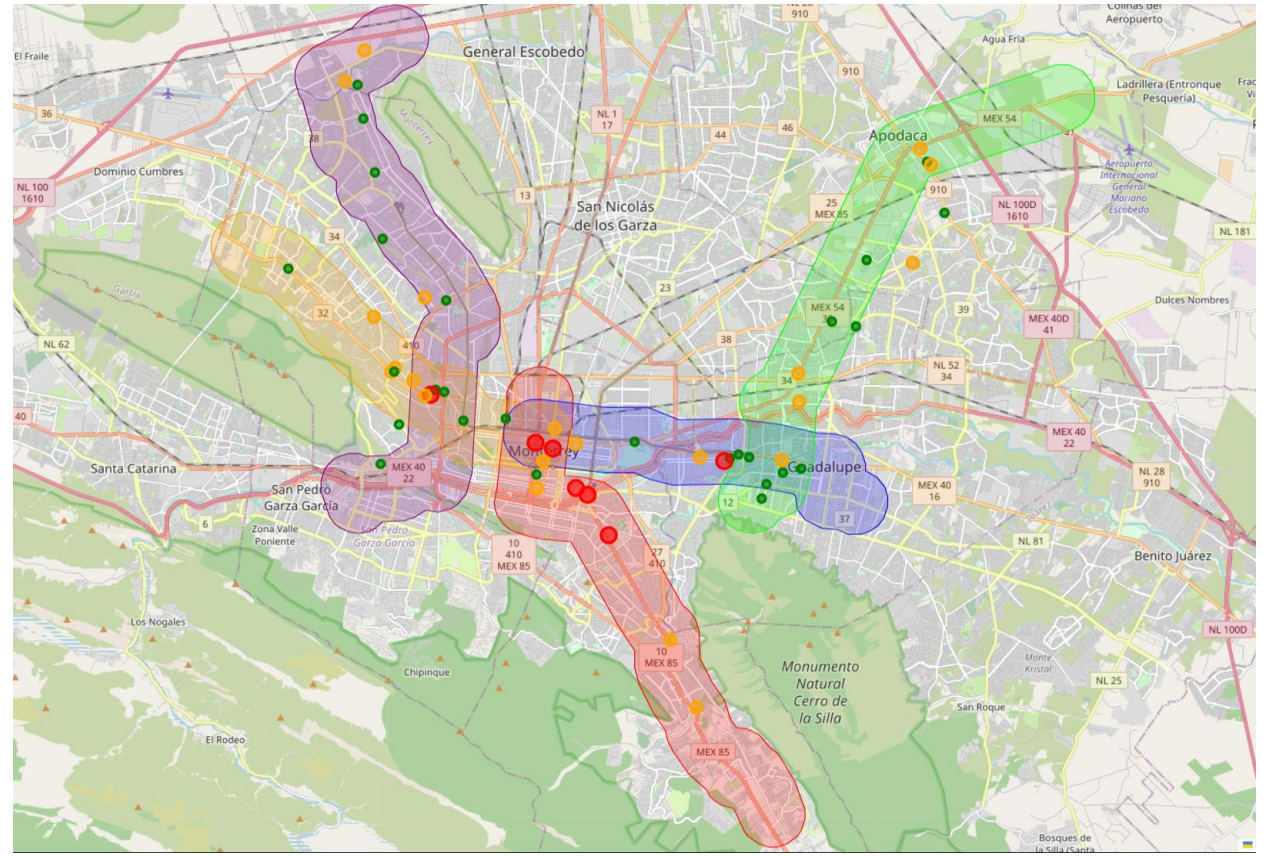


# Mapa con Análisis de Riesgo

En este mapa podemos observar cada una de las rutas. Su versión interactiva está disponible en el archivo **"mapa\_rutas\_riesgo.html"**.

Cada ruta es representada por un color distinto, e incluye el radio de 650m que mencionamos anteriormente.

Cada punto representa un siniestro, y se usa el tamaño y color para identificar el nivel de riesgo de cada uno.



# Preguntas y Respuestas

- a) Calcula la driving distance (distancia en automóvil) y un estimado del tiempo de llegada de cada ruta
- e) Con la información de los resultados del inciso a) de la Parte 1, genera una tabla que muestre toda la información generada de las rutas (Driving Distance y Tiempo Estimado)

IDRuta	DistanciaRutaKilometros	TiempoRutaMinutos	HoraSalida	HoraEstimadaLlegada
a1	20.517	19.466667	10:03:00	10:22:28
b2	12.909	13.750000	08:58:00	09:11:45
c3	22.584	21.516667	14:10:00	14:31:31
d4	23.733	25.200000	11:00:19	11:25:31
e5	14.047	14.000000	16:00:00	16:14:00

# Preguntas y Respuestas

d) En el inciso a) de la Parte 1, ¿es posible realizar un mapa de estas rutas generadas dentro del software estadístico? ¿Qué librería o paquetería utilizarías para realizar este mapa? (Es importante mencionar el nombre de esta librería)

Para graficar un mapa de las rutas generadas, utilicé las siguientes librerías:

- OXMMnx: Para obtener el grafo de rutas y calcular las distancias y tiempos de viaje
- GeoPandas: Para manejar datos geoespaciales y realizar análisis espacial
- Shapely: Para crear geometrías de puntos y líneas
- Folium: Para visualizar el mapa con las rutas y los siniestros

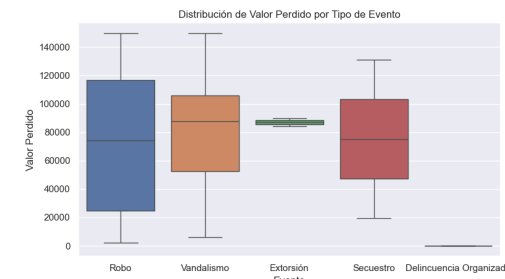
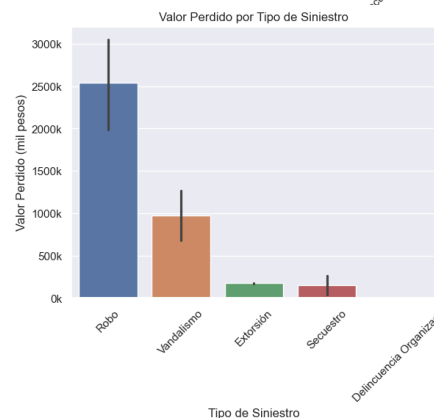
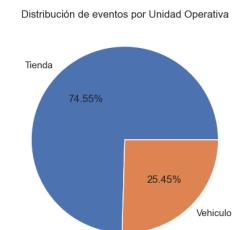
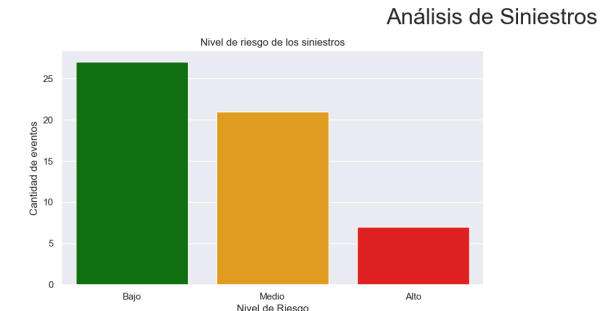
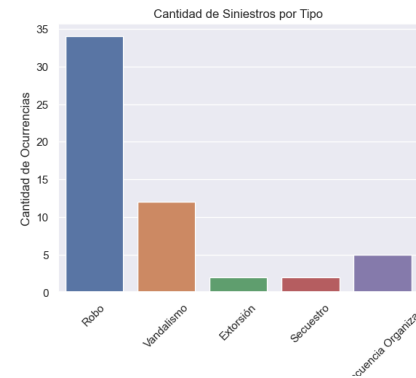
Folium es la librería que me permitió dibujar las líneas en el mapa a partir de un set de coordenadas. Algunas alternativas son Plotly y Cartopy. Me parece que GeoPandas también cuenta con funciones de visualización, pero no estoy tan familiarizado con la herramienta.

# Preguntas y Respuestas

b) Realiza un gráfico que muestre una descripción de la tabla encontrada en la hoja de Siniestros

f) Con la información de los resultados del inciso b) de la Parte 1, ¿Cuál es el siniestro más común? ¿Cuál crees que sea el siniestro con el mayor impacto? ¿Por qué?

- El siniestro con más ocurrencias es el robo, con 34.
- El de mayor impacto en términos de dinero perdido también es el robo, con un valor perdido superior a los 250 mil pesos.
- Sin embargo, otros crímenes de mayor severidad como la extorsión y el crimen organizado puedan ser considerados como los de mayor impacto, a pesar de que este impacto no necesariamente se vea reflejado en términos de dinero.



# Preguntas y Respuestas

c) Con la información de Siniestros y las Rutas calculadas, evalúa el nivel de riesgo de cada ruta (también pueden ser variables cualitativas si se argumenta en la presentación)

g) Con la información de los resultados del inciso c) de la Parte 1, genera una tabla ordenada por la ruta de mayor riesgo a menor riesgo

h) Con la información de los resultados del inciso c) de la Parte 1, describe brevemente cómo y por qué llegaste a la conclusión del nivel de riesgo en las rutas

- Esta tabla muestra cada ruta, junto a sus índices de riesgo. Podemos observar que la ruta **e5** puede ser considerada como la más riesgosa, con un índice de riesgo de **93.38**.
- Como expliqué anteriormente, este índice se compone de métricas objetivas y ponderaciones subjetivas. Cambiar alguna de estas variables arbitrarias podría alterar el resultado significativamente.
- Pienso que el índice de riesgo de la ruta **a1** es un buen ejemplo de cómo las ponderaciones subjetivas terminan afectando el resultado final. Dicha ruta tiene la menor cantidad de siniestros y de valor perdido, pero podemos observar en el mapa que es la ruta donde hay una mayor concentración de siniestros graves, la cual impulsa su índice hasta **80.56**, posicionándola en el segundo lugar en el ranking.

IDRuta	Ranking	IndiceRiesgo	NivelRiesgo	NumSiniestros	TotalValorPerdido
e5	1.0	93.38424	Alto	16	1638424
a1	2.0	80.55986	Alto	12	155986
b2	3.0	72.44580	Medio	15	344580
d4	4.0	70.83572	Medio	15	1783572
c3	5.0	51.54135	Bajo	14	854135

