

1) Zonas de riesgo / puntos negros (Objetivo 1)

El proyecto busca identificar zonas o tramos de calle donde la probabilidad de ocurrencia de accidentes graves sea alta, dadas ciertas condiciones (hora, día, contexto vial, etc.), y usar esa información como base para el diseño de rutas más seguras.

Esta noción de “punto negro” se entiende como:

Zona o tramo de calle donde la probabilidad de tener un accidente grave es elevada, considerando frecuencia de accidentes, severidad (muertos/heridos) y la dimensión temporal (horas del día, días de semana, años).

2. Datos de red vial

El proyecto utiliza **OpenStreetMap (OSM)** como fuente de la red vial, accediendo a través de **OSMnx** en Python.

- **Obtención de la red vial**
 - OSMnx realiza internamente una consulta a **Overpass** para descargar la red vial de la zona de estudio (por ejemplo, CDMX o un estado completo).
 - El resultado es directamente un **grafo vial**:
 - Nodos: intersecciones.
 - Aristas: tramos de calle, con atributos como longitud, tipo de vía, límite de velocidad (cuando está disponible), entre otros.
- **Almacenamiento y reutilización**
 - La red descargada se almacena para su reutilización durante todo el proyecto, por ejemplo:
 - En **MongoDB**.
 - En archivos como **GraphML** o **pickle**.
 - De esta manera, se evita descargar repetidamente la red y se garantiza consistencia a través de todos los análisis.

3. Datos de accidentes

El proyecto se apoya en los datos de accidentes de tránsito provenientes de **INEGI (ATUS)**, para el periodo **2019–2023**.

- **Unificación de años**

- Todos los años (2019, 2020, 2021, 2022, 2023) se integran en un único conjunto de datos, manteniendo una columna de **año** ([año/año](#)) consistente.
- Esta integración permite:
 - Aumentar el número de observaciones para análisis y modelos.
 - Conservar la posibilidad de filtrar o ponderar por año en etapas posteriores.

- **Variables básicas a construir**

- **datetime**: combinación de año, mes, día, hora y minutos en un solo campo de fecha-hora (la llamé fechahora/fechadora).
- severidad: métrica compuesta de cuán grave fue el accidente. Por ejemplo:
 - Versión simple:
 - severidad = $10 * \text{totmuertos} + 3 * \text{totheridos}$
 - = $10 * (\text{condmuerto} + \text{pasamuerto} + \text{peatmuerto} + \text{ciclmuerto} + \text{otromuerto}) + 3 * (\text{condherido} + \text{pasaherido} + \text{peatherido} + \text{ciclherido} + \text{otroherido})$
- **hay_muertos**: indicador booleano que vale 1 si en el accidente hubo al menos una persona fallecida ($\text{totmuertos} > 0$) y 0 en caso contrario.
- **hay_heridos**: indicador booleano que vale 1 si en el accidente hubo al menos una persona herida ($\text{totheridos} > 0$) y 0 en caso contrario.
- **solo_daños_materiales**: indicador booleano que vale 1 cuando no hay muertos ni heridos ($\text{totmuertos} = 0$ y $\text{totheridos} = 0$), es decir, solo se reportan daños materiales.
- **severidad_cat**: versión categórica de la gravedad del accidente, construida a partir de totmuertos y totheridos . Por ejemplo:
 - muy grave: si $\text{totmuertos} \geq 1$
 - grave: si $\text{totmuertos} = 0$ y $\text{totheridos} \geq 3$
 - moderada: si $\text{totmuertos} = 0$ y $1 \leq \text{totheridos} \leq 2$
 - leve: si $\text{totmuertos} = 0$ y $\text{totheridos} = 0$
- **franja_horaria**: categoría que agrupa la hora del accidente (hora) en bloques del día:
 - Madrugada: 0–5 h
 - Mañana: 6–11 h
 - Tarde: 12–17 h
 - Noche: 18–23 h

- **Uso de la columna `oid`**
 - La columna `oid` se interpreta como un identificador interno geográfico del shapefile o geobase, no como característica del accidente.
 - Dado que:
 - Solo aparece en algunos años (p. ej. 2022–2023).
 - Ya existe una columna `id` como identificador de accidente.
 - En el análisis y en los modelos, `oid` se considera prescindible y puede eliminarse (droppearla) para evitar ruido, especialmente antes de concatenar los dataframes de distintos años.
-

4. Elección de la unidad espacial

El proyecto contempla dos enfoques complementarios para definir la unidad espacial de análisis:

4.1. Opción A – Cuadrícula / celdas

- La zona de estudio se divide en celdas regulares, por ejemplo:
 - **100 m x 100 m o 200 m x 200 m.**
- Cada accidente se asigna a una celda según su latitud y longitud.
- Para cada celda se calcula:
 - `accidentes_totales`
 - `accidentes_graves` (por ejemplo, donde `totmuertos > 0` o `severidad` por encima de un umbral)
 - `severidad_total` (suma de la severidad de todos los accidentes en la celda)

Ventajas

- Resultados fáciles de visualizar en **mapas de calor (heatmaps)**.
- Enfoque sencillo y útil para análisis exploratorio y detección preliminar de puntos negros.

Desventajas

- Las celdas no siempre se alinean exactamente con los tramos de calle, aunque siguen siendo muy informativas.

4.2. Opción B – Tramos de calle (segmentos de la red vial)

Este enfoque se alinea directamente con el objetivo final de generar **rutas seguras**.

- Cada arista (segmento) del grafo vial de OSMnx se considera una unidad espacial.
- Cada accidente se “pega” al tramo más cercano, realizando un **snap al edge** más próximo del grafo.
- Para cada tramo se calculan:
 - `accidentes_totales`
 - `severidad_total`
 - `accidentes_por_año`
 - Otras métricas derivadas, según sea necesario.

4.3. Estrategia combinada

- Para la **presentación inicial** y visualización global de los puntos negros, se utilizará el enfoque de **celdas** (cuadrícula).
- Para conectar con el módulo de **ruteo seguro**, el proyecto migrará progresivamente a la unidad de análisis basada en **segmentos de calle**, reutilizando la misma lógica de riesgo.

5. Definición y cálculo del “Índice de Riesgo”

El proyecto define un **Índice de Riesgo** por zona (celda o tramo) que combine frecuencia, gravedad y, cuando corresponda, fatalidad.

5.1. Métricas base por zona

Para cada zona z y un periodo T (por ejemplo, todo 2019–2023):

- N_z : número total de accidentes.
- S_z : severidad total (suma de `severidad` de todos los accidentes en la zona).
- F_z : número de accidentes fatales (por ejemplo, donde `totmuertos > 0`).

5.2. Normalización

Cada métrica se normaliza dentro de la ciudad, por ejemplo mediante **min–max**:

$$\begin{aligned} \bullet \quad N'_z &= \frac{N_z - \min(N)}{\max(N) - \min(N)} \\ \bullet \quad S'_z &= \frac{S_z - \min(S)}{\max(S) - \min(S)} \\ \bullet \quad F'_z &= \frac{F_z - \min(F)}{\max(F) - \min(F)} \end{aligned}$$

5.3. Índice de riesgo compuesto

Se define un índice de riesgo para cada zona:

$$\begin{aligned} \text{riesgo}_z &= 0.4 \cdot N'_z + 0.4 \cdot S'_z + 0.2 \cdot F'_z \\ \text{riesgo_z} &= 0.4 \cdot N'_z + 0.4 \cdot S'_z + 0.2 \cdot F'_z \end{aligned}$$

- Los pesos son **ajustables**, pero el ejemplo prioriza:
 - Zonas con muchos choques (N).
 - Zonas con accidentes de alta severidad o con muertos (S, F).

En los entregables, se documenta la elección de estos pesos y, si se realizan pruebas alternativas, se explican las variantes.

6. Dimensión temporal del riesgo

El carácter dinámico del riesgo se incorpora en varias escalas temporales.

6.1. Por franja horaria

Se definen franjas horarias, por ejemplo:

- Madrugada: 0–5 h
- Mañana: 6–11 h
- Tarde: 12–17 h
- Noche: 18–23 h

Para cada franja y zona se recalculan Nz,Sz,FzN_z, S_z, F_zNz,Sz,Fz y el **índice de riesgo** correspondiente.

Esto permite generar **mapas de puntos negros por hora del día**.

6.2. Por día de la semana

Se utilizan variables como `diasemana` para:

- Comparar **lunes–viernes** vs **sábado/domingo**, o
- Generar siete mapas distintos (uno por cada día de la semana).

6.3. Por mes o temporada

Se evalúan posibles patrones estacionales:

- Temporada de lluvias.
- Vacaciones, feriados u otros periodos con patrones de movilidad distintos.

En general, se puede considerar una función del tipo:

```
riesgoz(hora,dia_semana)riesgo_z(hora,  
dia\semana)riesgoz(hora,dia_semana)
```

Esta función se conecta posteriormente con el módulo de ruteo, de modo que una petición de ruta (por ejemplo, lunes 8:00 am) utilice el **riesgo correspondiente a esa franja y día**.

7. Tratamiento del periodo 2019–2023 y ponderación por año

La unión de los años 2019–2023 se considera adecuada, pero se diferencia su uso según el objetivo:

7.1. Para modelos y patrones generales

- Se utilizan **todos los años unidos** para entrenar modelos y entender patrones globales.
- La mayor cantidad de datos aporta estabilidad a las estimaciones.
- Se conserva la columna de **año** para posibles análisis por periodo.

7.2. Para el mapa de puntos negros “actuales”

- Se reconoce que la ciudad cambia (obras, cambios de sentido, nueva señalización, etc.) y que 2020–2021 fueron años atípicos por pandemia.
- Por ello, se definen dos capas o enfoques:
 - **Riesgo histórico 5 años (2019–2023)**
 - Utiliza todos los años de forma combinada, con pesos iguales o casi iguales.
 - **Riesgo reciente (por ejemplo, últimos 2–3 años)**
 - Da más peso a los años más cercanos al presente.
- También se contempla un esquema de **ponderación anual**, por ejemplo:
 - 2019 → $w=0.10w = 0.10w=0.10$
 - 2020 → $w=0.15w = 0.15w=0.15$
 - 2021 → $w=0.20w = 0.20w=0.20$
 - 2022 → $w=0.25w = 0.25w=0.25$
 - 2023 → $w=0.30w = 0.30w=0.30$
- (Los pesos suman 1.0.)

Para cada zona zzz, puede definirse un riesgo ponderado como:

riesgoz= $\sum_{an} \text{owanz} \cdot \text{severidadz}_{an}$ riesgo_z = $\sum_{año} w_{año} \cdot \text{severidad}_{z,año}$
severidad_z,año riesgoz=an~o $\sum_{wan} \text{owanz} \cdot \text{severidadz}_{an}$
donde severidadz,an~oseveridad_z,año severidadz,an~o es la suma de severidad en esa zona durante ese año.

En el reporte se explicita que:

“Se usaron datos 2019–2023, con mayor ponderación a los años más recientes para reflejar mejor el estado actual de la red vial”.

8. Herramientas y operaciones espaciales clave

El proyecto se apoya en el ecosistema Python para geodatos:

- **OSMnx**: descarga y manejo de la red vial desde OSM/Overpass, construcción del grafo.
 - **GeoPandas**: manejo de datos espaciales de accidentes y operaciones de georreferenciación.
 - **Operaciones con OSMnx + GeoPandas**:
 - Georreferenciación de accidentes en la red vial.
 - Asignación de cada punto de accidente al tramo de calle más cercano (snap al edge).
 - Cálculo de densidades (por ejemplo, kernel density estimation) y conteos por tramo o por celda.
-

9. Capas temáticas y productos de visualización

A partir de las columnas disponibles en los datos de INEGI, se construyen varias capas analíticas y de visualización:

- **Gravedad**
 - Indicadores: `totmuertos`, `totheridos`, `severidad`.
 - Mapas específicos de:

- Accidentes graves.
 - Accidentes fatales.
- **Tipo de accidente y causa**
 - Variables como `tipaccid` y `causaaccci` permiten:
 - Mapas de puntos negros de atropellos.
 - Mapas de puntos negros de choques por exceso de velocidad, etc.
 - **Tipo de vehículo y usuarios vulnerables**
 - A partir de variables como motocicleta, bicicleta, peatones muertos/heridos (`peatmuerto`, `peatherido`), se elaboran:
 - Mapas de riesgo específicos para **motos, bicis y peatones**.
 - **Contexto vial**
 - A partir de variables como urbana, suburbana, carretera, clase de vía:
 - Comparación de patrones entre zonas urbanas y carreteras.
 - Identificación de tramos de carretera particularmente riesgosos.

9.1. Productos esperados

Entre los productos visuales y analíticos se consideran:

1. **Mapa de calor global 2019–2023**
 - De puntos negros por celdas o segmentos, utilizando el índice de riesgo.
2. **Mapas interactivos con filtros**
 - Filtros por:
 - Hora del día.
 - Día de la semana.
 - Tipo de accidente.
 - Tipo de usuario (moto, peatón, bicicleta, etc.).

3. Tabla de “Top 20” zonas más peligrosas

- Campos sugeridos:
 - Coordenadas aproximadas o referencia de calle1–calle2.
 - Accidentes totales.
 - Número de muertos.
 - Número de heridos.
 - Índice de riesgo.

4. Gráficas de soporte

- Accidentes por hora del día.
- Accidentes por día de la semana.
- Comparación antes y después de la pandemia (2019 vs 2020–2021).
- Distribuciones de severidad, tipos de accidente, etc.

10. Flujo de trabajo resumido

De forma integrada, el plan de trabajo para el equipo (equipo 3–4) se puede sintetizar así:

1. Descarga y preparación de la red vial

- Obtención del grafo vial de la zona de estudio mediante OSMnx.
- Almacenamiento del grafo (MongoDB, GraphML, pickle) para reutilización.

2. Integración y limpieza de datos de accidentes (INEGI ATUS 2019–2023)

- Unificación de años en un solo dataset, armonizando nombres de columnas.
- Eliminación de columnas puramente internas como `oid`.
- Construcción de variables clave: `datetime`, `severidad`, `tipo_via/contexto`, `diasemana`, etc.

3. Definición de unidad espacial

- Construcción de una **cuadrícula** (100–200 m) para análisis exploratorio y mapas de calor.
- Preparación paralela para trabajar con **tramos de calle** (edges del grafo) como unidad para el ruteo seguro.

4. Asignación espacial de accidentes

- Asignación de cada accidente a una celda de la cuadrícula.
- Asignación de cada accidente al tramo de calle más cercano (snap al edge), cuando se utilice la unidad de tramos.

5.

5. Cálculo de métricas e índices

- Cálculo de N_z , S_z , F_z por zona y periodo (global, por año, por franja horaria, por día de semana).
- Normalización de métricas y construcción del **Índice de Riesgo** compuesto.
- Definición de capas:
 - Riesgo histórico 2019–2023.
 - Riesgo reciente ponderado con mayor peso a años actuales.

6. Análisis temporal y temático

- Construcción de mapas de riesgo por franja horaria, día de semana y temporada.
- Generación de mapas específicos por tipo de accidente, causa y tipo de usuario.

7. Visualización y reporte

- Desarrollo de mapas (estáticos e interactivos) de puntos negros.
- Elaboración de tablas de zonas más peligrosas.
- Redacción de la explicación metodológica, incluyendo:
 - Definición formal de “punto negro”.
 - Descripción del índice de riesgo y sus pesos.

- Justificación del uso de datos 2019–2023 con ponderación temporal.
- Preparación de resultados para su futura integración con el módulo de **rutas seguras**, utilizando el riesgo dinámico por hora y día.

Con este plan, todas las piezas (OSM/OSMnx, INEGI ATUS, cuadrícula, segmentos de calle, índices de riesgo, análisis temporal y temático) se articulan en un flujo de trabajo coherente y directamente implementable en Python.