

Seminario de Proyectos

4A. Resumen. Generalidades de una Base de Datos de Investigación

Nombre: David Aaron Ramirez Olmeda

Programa: Maestría en Ciencia de Datos e Información



Reporte de Recopilación de Datos para el Proyecto de Predicción de Tasa de Accidentes Viales en Sinaloa, México

Fuentes de Información:

- Los datos se han obtenido de dos fuentes principales:
 - La Estadística de Accidentes de Tránsito Terrestre en Zonas Urbanas y Suburbanas del INEGI. [Link \(https://www.inegi.org.mx/programas/accidentes/#documentacion\)](https://www.inegi.org.mx/programas/accidentes/#documentacion)
 - Información acerca del número de eventos individuales de victimización delictiva reportados. [Link \(https://www.inegi.org.mx/app/descarga/?t=155\)](https://www.inegi.org.mx/app/descarga/?t=155)

Manera o algoritmo para obtener los datos:

- Los datos se han descargado directamente del portal del INEGI. No se ha aplicado ningún algoritmo específico para la extracción de datos, ya que están disponibles para descarga pública.

Sistema de Gestión de Base de Datos a Utilizar:

- Por el momento, hemos optado por almacenar los datos en archivos CSV. Esto nos proporciona la flexibilidad necesaria para manipular los datos en Python. No consideramos necesario procesar los datos en una base de datos SQL u otro sistema similar en esta etapa del proyecto.

Variables a Incluir en la Base de Datos:

- Para predecir la tasa de accidentes viales, y con el potencial de adaptarlas durante el análisis exploratorio de datos, utilizaremos varias variables. Ejemplificamos algunas de las variables más significativas:

Variables de Accidentes de Tránsito:

- **COBERTURA:** Área geográfica a la que están referidos los indicadores estadísticos, como "Zona Urbana" o "Zona Suburbana".
- **ID_ENTIDAD:** Clave de la entidad federativa (ejemplo: "Sinaloa").
- **ANIO:** Año en que ocurrió el accidente (ejemplo: "2022").
- **MES:** Mes de referencia en que ocurrió el accidente (ejemplo: "03" para marzo).
- **ID_HORA:** Hora en que ocurrió el accidente (ejemplo: "15" para las 3 PM).
- **ID_DIA:** Día del mes en que ocurrió el accidente (ejemplo: "07" para el séptimo día).

- **TIPACCID:** Tipo de accidente de tránsito, como "Colisión con vehículo automotor" o "Volcadura".

Variables de Incidencia Delictiva:

- **TIPO_DELITO:** Tipo de delito reportado, como "Robo" o "Homicidio".
- **ANIO:** Año en que se registró el delito (ejemplo: "2022").
- **MES:** Mes en que se cometió el delito (ejemplo: "03" para marzo).
- **HORA:** Hora en que se cometió el delito (ejemplo: "21:30" para las 9:30 PM).

Estas variables son ejemplos iniciales que podrían ser utilizados en nuestro modelo de predicción de la tasa de accidentes viales y la incidencia delictiva en Sinaloa. Durante la etapa de análisis exploratorio de datos, evaluaremos la relevancia de estas variables y podremos adaptarlas, agregar nuevas o eliminar algunas según sea necesario.

-