**Avance Proyecto Final**

**Integrantes:**

**Sergio Rojas**

**Cristhian Amaya**

**Andrés Beltrán**

**Gloria Ramos**

**¿Cuál es la probabilidad de que haya un accidente fatal en un accidente automovilístico en México?**

1. **Problema que abordarán y su contexto.**

Contexto:

México es un país latinoamericano con una densa red de carreteras y una alta tasa de movilidad. En 2021, El Sistema Nacional de Seguridad Pública de México registró un total de aproximadamente 296.000 accidentes automovilísticos en el país. De estos accidentes, más de 23.000 resultaron en fatalidades, lo que representó un preocupante aumento en comparación con años anteriores.

Este incremento en las cifras de accidentes fatales ha suscitado una profunda preocupación en las autoridades de tránsito y en la población en general. Se ha identificado que varios factores contribuyen a esta situación, incluyendo el exceso de velocidad, el consumo de alcohol, la falta de cumplimiento de las normas de tránsito y el estado de las carreteras. Además, la falta de implementación efectiva de políticas de seguridad vial ha dejado una huella importante en la seguridad en las carreteras mexicanas.

El impacto social y económico de estos accidentes es significativo. Las fatalidades en accidentes automovilísticos afectan a las familias y comunidades, y también generan costos médicos y rehabilitación considerables, además de pérdidas de productividad en el país.

El presente estudio, que pretende responder a la pregunta de ¿Cuál es la probabilidad de que haya un accidente fatal en un accidente automovilístico en México? Se basa en la necesidad de abordar esta problemática y reducir las consecuencias devastadoras que estos accidentes tienen en México. La información recopilada y analizada proporcionada por El instituto Nacional de Estadística y Geografía (INEGI) de México será fundamental para comprender las tendencias específicas de 2021 y contribuir a la implementación de estrategias efectivas de seguridad vial en el país.

El análisis de datos, respaldado por la información que se capta mediante el SICATUS, permitirá identificar patrones y factores de riesgo asociados a accidentes fatales, lo que, a su vez, proporcionará una base sólida para la toma de decisiones informadas y promoción de prácticas de conducción más seguras en México.

1. **Pregunta de negocio y alcance del proyecto.**

Pregunta de negocio: ¿Cuál es la probabilidad de que algún implicado peatonal, pasajero o conductor resulte en deceso en un accidente automovilístico?

Alcance del proyecto:

El alcance del proyecto se enfocará en los siguientes aspectos:

* Recopilación de datos: Se recopilarán datos de accidentes automovilísticos ocurridos en México durante el año 2021. Estos datos incluirán un total de 46 variables y un total de 3.849 registros.
* Limpieza y preparación de los datos: Preparación de los datos para garantizar la calidad y coherencia de la información recopilada.
* Modelado de los datos: Se aplicarán técnicas estadísticas y modelado de datos para calcular la probabilidad de accidentes fatales en función de las variables recopiladas. Esto incluirá el uso de técnicas de clasificación.
* Identificación de factores de riesgo: Se identificarán los factores de riesgo más influyentes en la probabilidad de accidentes fatales. Esto permitirá comprender qué variables tienen mayor correlación con accidentes mortales.
* Informe final: Construcción de un tablero dinámico que permita evidenciar los resultados.

1. **Conjuntos de datos a emplear.**

Las fuentes de datos son originadas por INEGI por medio de un cuestionario de accidentes de tránsito la cual recopila dentro de las 46 variables cuatro tipos de datos, registro de localización (código de municipio, zona urbana y zona sub urbana), registro de hora y fecha en que sucede el accidente (Año, hora, minutos, día de la semana, día), datos del conductor y datos del vehículo.

El nombre de la fuente de datos original es “Estadística de Accidentes de Tránsito Terrestre en Zonas Urbanas y SubUrbanas”.

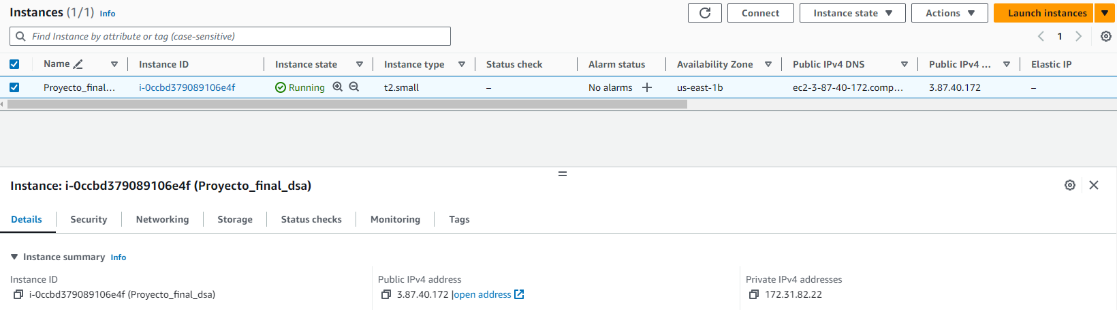
Adicional se cuenta con una tabla de referencia de los Municipios de México, que tiene información del código del municipio y el nombre asociado. Esta base es generada por el Instituto Nacional para el Federalismo y Desarrollo Municipal.

Fuente de datos original y diccionario: <https://www.inegi.org.mx/rnm/index.php/catalog/788/data-dictionary/F1?file_name=ATU>

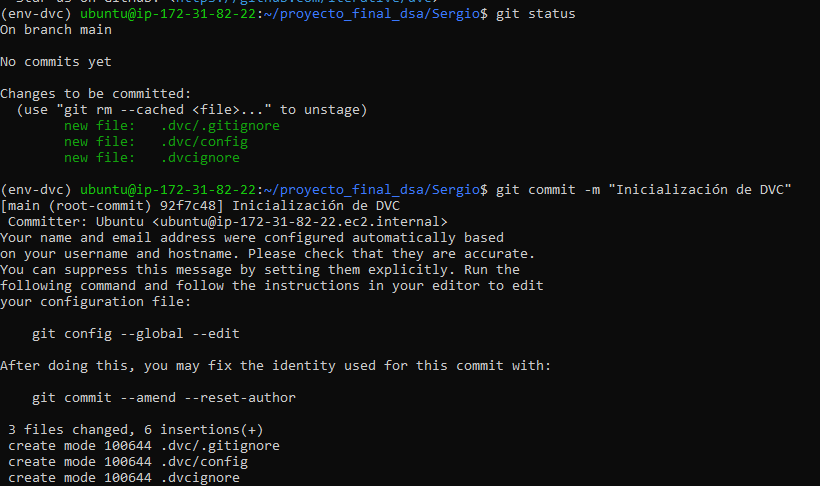
**4. Repositorio Git en uso para el código y Repositorio DVC en uso para los datos.**

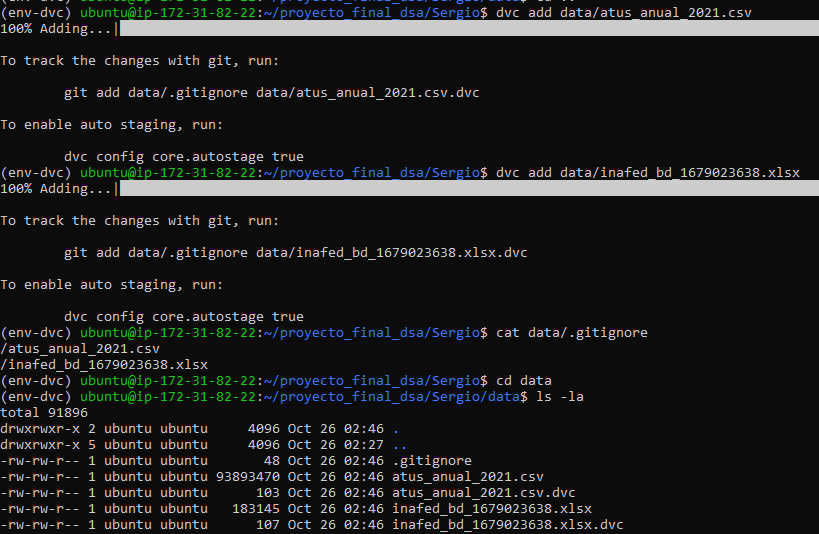
**URL repositorio:** <https://github.com/SergioRojas86/Proyecto_final_Despliegue_de_soluciones_analiticas>

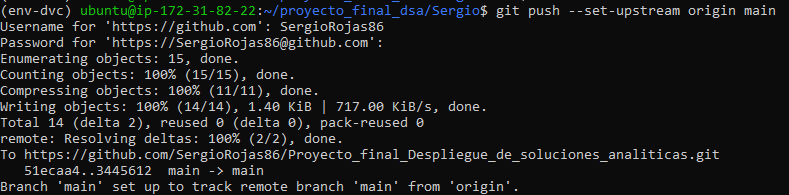
Creación Instancia



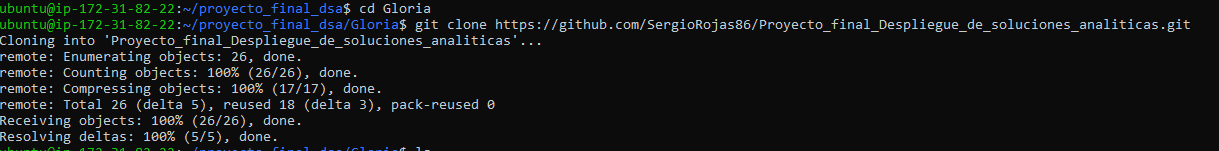
Inicialización git y dvc







Clonación repositorio Gloria



Clonación repositorio Cristhian

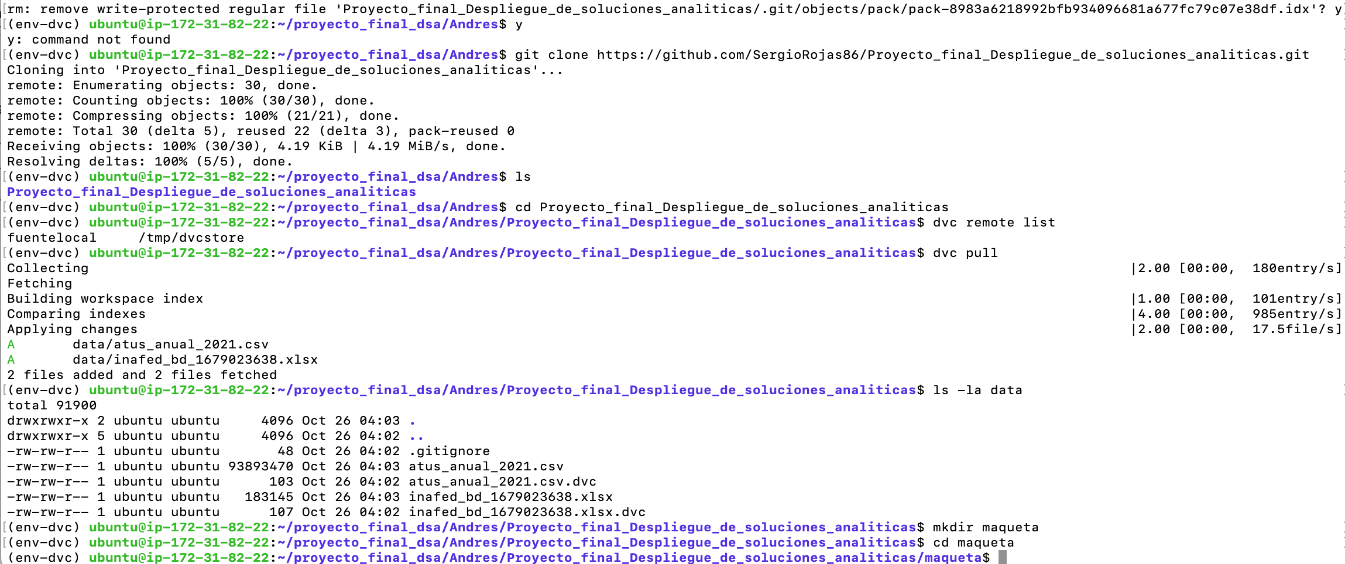
**Texto

Descripción generada automáticamente**

Texto

Descripción generada automáticamente

Clonación repositorio Andrés



1. **Exploración de los datos.**

Realizamos un análisis de distribución de la variable de respuesta.

Gráfico, Gráfico de barras

Descripción generada automáticamente

La variable respuesta, representa el 1% del total de los datos. Ahora, miramos la distribución de las variables predictoras.

Gráfico, Gráfico de barras

Descripción generada automáticamente

El gráfico de barras permite identificar que los lunes y los sábados son los días con mayor número de accidentes.

Gráfico, Gráfico de barras

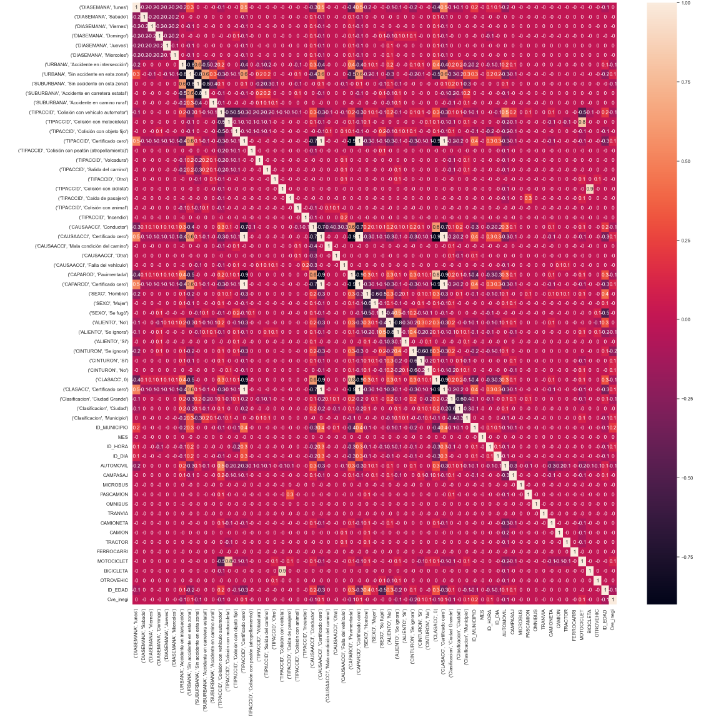
Descripción generada automáticamente

Asimismo, que el automóvil es el medio con mayores accidentes terrestres registrados.

Un dibujo de una persona con un paraguas

Descripción generada automáticamente con confianza media

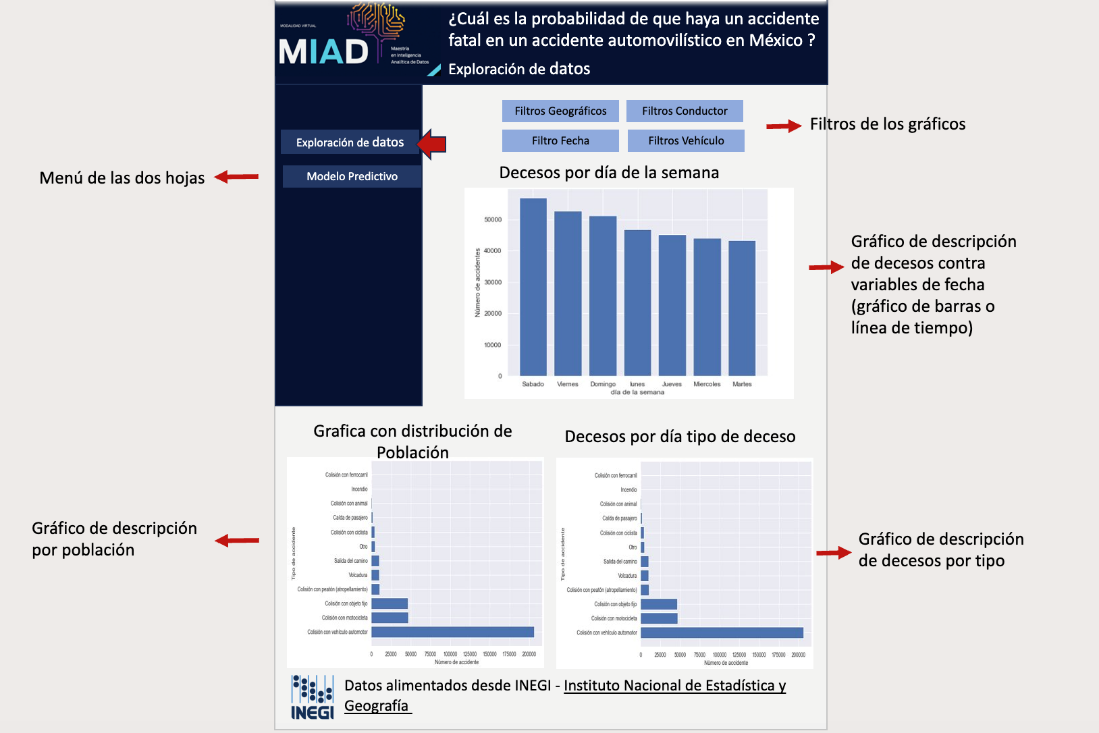
La mediana tanto para hombres como para mujeres se encuentra en los 35 años de edad, donde se observa una alta distribución de la edad de los conductores entre los 23 y 30 años.



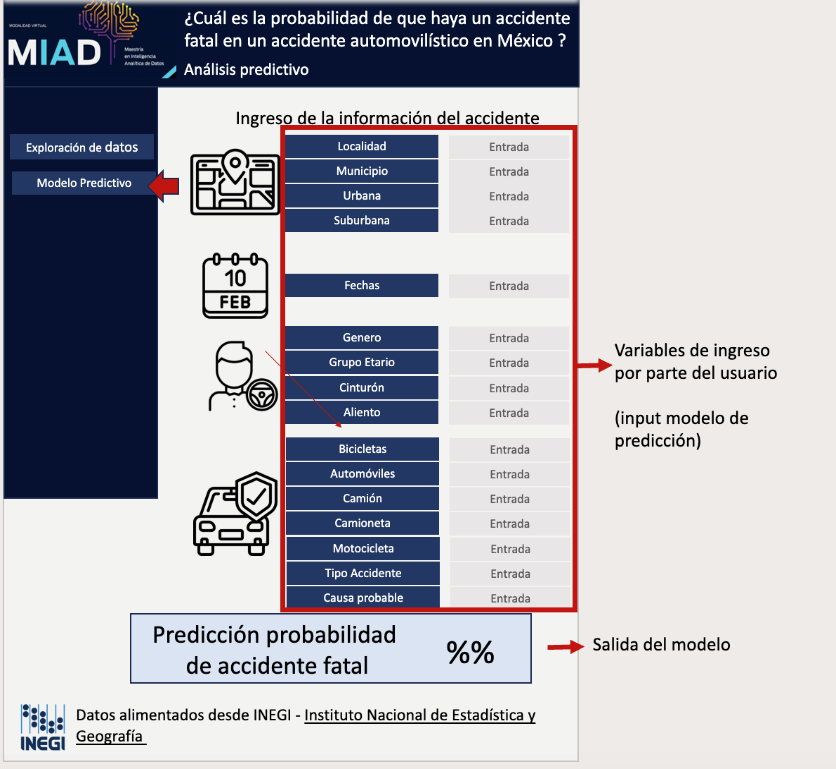
Por ahora, no se evidencian correlaciones significativas con la variable de respuesta.

1. **Maqueta del prototipo.**

La herramienta desplegada comprende dos secciones, desarrolladas por medio de Dash, la primera comprende un análisis de la exploración de datos, donde las gráficas representan el comportamiento de los datos según las tres clasificaciones de las variables.



La segunda hoja comprende el modelo predictivo, donde el usuario ingresa las variables de entrada del modelo y como resultado el modelo entregará la probabilidad de que el accidente sea fatal.



1. **Reporte de trabajo en equipo**

El entregable se dividió en partes iguales, donde cada miembro del equipo tomó dos puntos para agregar. Gloria trabajó la redacción del problema a abordar y del problema de negocio. Sergio, trabajó la parte que lanzar la instancia del grupo y realizar los repositorios de DVC. Cristian y Andrés, trabajaron la parte de exploración de los datos y de construir la maqueta. Al final, cada uno se unió a la instancia creada y clonó el repositorio subiendo cada uno su propia asignación.

Creando un commit cada uno de los integrantes dentro de la misma rama y actualizando la información en el repositorio de GitHub del proyecto.