

SEGUNDA ENTREGA DEL PROYECTO

ALEJANDRO ARIAS ORTIZ

LUIS MATEO OCHOA AGUDELO

DUVAN ESNEIDER GALLEGUO JIMENEZ

RAUL RAMOS POLLAN



MODELOS DE SISTEMAS 1 / INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

MEDELLÍN 2023

DESCRIPCION DEL PROGRESO

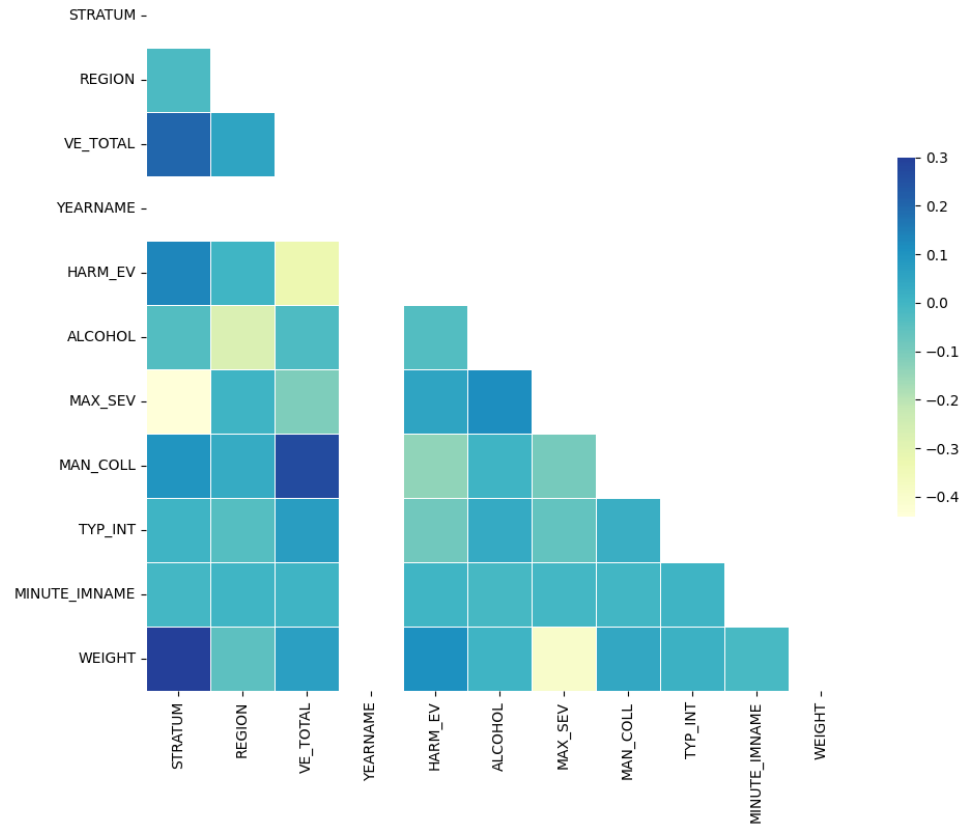
1) PROCESAMIENTO DEL DATASET

Para la primera parte del procesamiento del dataset hicimos la instalación de bibliotecas instalando algunas bibliotecas Python necesarias para el análisis de datos. Se utilizan los comandos: *!pip install* y *!pip install -q* para instalar las bibliotecas *kaggle*, *gdown*, y *seaborn*. Posterior a esto descargamos el conjunto de datos para esto Utiliza la biblioteca *gdown* para descargar un archivo desde Google Drive. La URL del archivo es proporcionada en la variable *url*, y el archivo se guarda en el sistema local con el nombre "*acc_20.csv*". Posteriormente cargamos los datos desde el archivo descargado Se carga el archivo CSV descargado ("*acc_20.csv*") en un DataFrame de pandas utilizando el método *pd.read_csv*. Esto permite la manipulación y exploración de los datos.

Luego de tener los datos disponibles procedemos a la exploración de los mismos

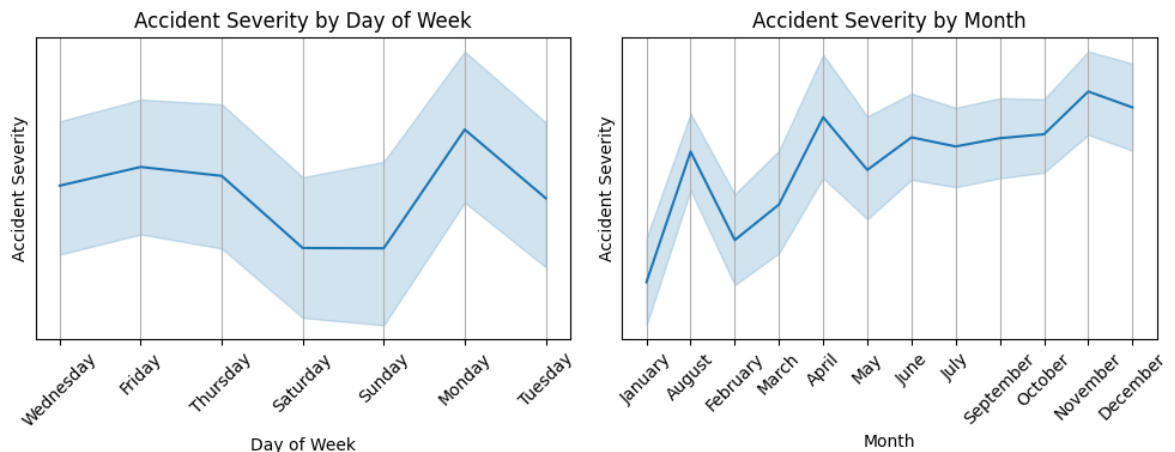
Como primer paso en la exploración de los datos analizamos que no todas las columnas eran necesarias para nuestro análisis y por esto mismo procedemos a eliminarlas, Las columnas que se eliminarán se definen en la lista *columns_to_drop*.

Posterior a esto hacemos una visualización de las columnas que nos quedan después de la eliminación y con estas calculamos una matriz de correlación.



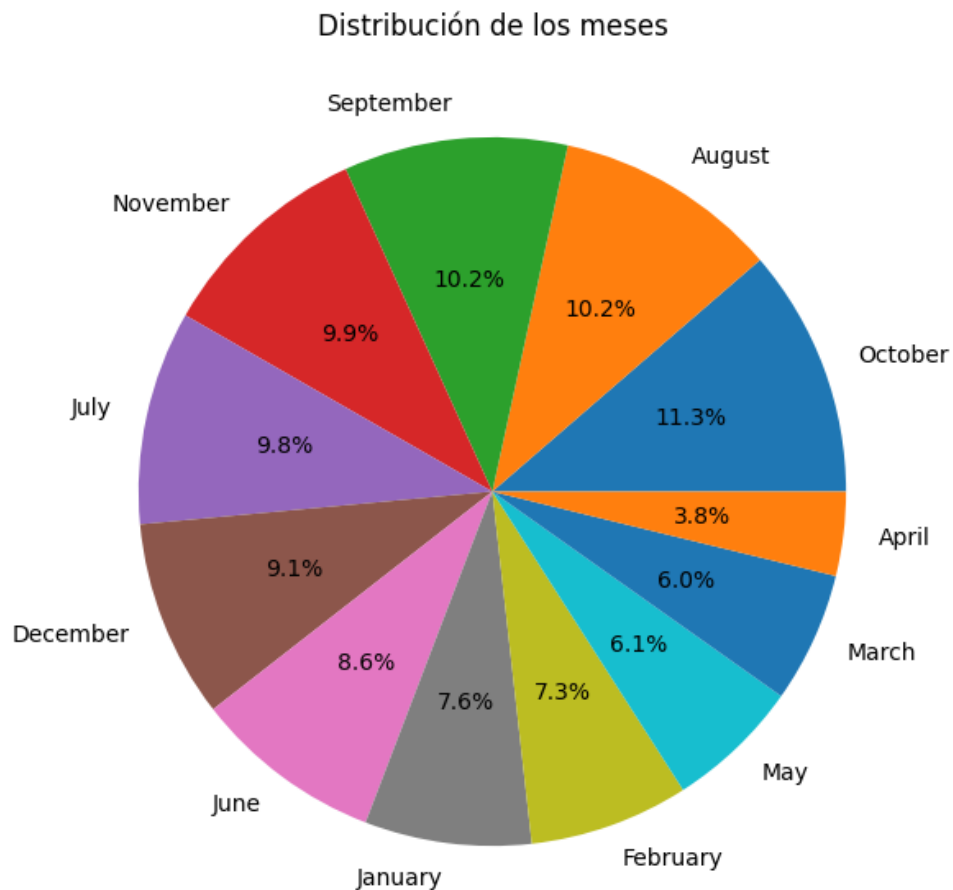
Esta matriz de correlación entre las variables numéricas y crea un mapa de calor de esta matriz de correlación utilizando *seaborn* y *matplotlib*. El mapa de calor es una herramienta visual que muestra la relación entre las variables.

La matriz de correlación muestra cómo las diferentes variables del conjunto de datos se relacionan entre sí. Esto es esencial para identificar las relaciones fuertes o débiles entre las variables



también creamos unos gráficos adicionales lineales que muestran la relación entre la gravedad de los accidentes y el día de la semana y el mes.

los gráficos de líneas que muestran la relación entre la gravedad de los accidentes y variables como el día de la semana y el mes pueden proporcionar información sobre patrones temporales.



también creamos un gráfico circular para mostrar la distribución de los meses en el conjunto de datos. Esto se hace contando la frecuencia de cada mes y representándola en el gráfico.

La distribución de los meses en un gráfico circular puede ayudar a identificar si existen patrones estacionales en la tasa de accidentes

Estas graficas son útiles para entender la relevancia de los datos y pueden ser pasos importantes en la creación de un modelo predictivo para la tasa de accidentalidad en el futuro.