

TRABAJO PRÁCTICO

Visualización de datos

Nombre: Carlos Alberto Mejía Rodríguez

Propuesta: Análisis de accidentes de tránsito en la ciudad de Bucaramanga, Colombia.

Enlaces:

[Proyecto en StackBlitz](#)
[GitHub](#)
[Google Colab](#)

Introducción

ACCIDENTES DE TRÁNSITO DESDE ENERO 2012 A FEBRERO DE 2023

En este informe se realiza un análisis gráfico y estadístico utilizando el conjunto de datos obtenido durante el proceso KDD en la asignatura de "Minería de Datos". Puedes acceder al cuaderno de Google Colab con el código del preprocesamiento de los datos a través de este [enlace](#).

La base de datos seleccionada, titulada originalmente "03. ACCIDENTES DE TRÁNSITO DESDE ENERO 2012 A FEBRERO DE 2023", ha sido descargada de DATOS ABIERTOS, una plataforma del gobierno colombiano que proporciona acceso a información pública de Colombia. Esta base de datos contiene información georreferenciada sobre los accidentes de tránsito ocurridos en el municipio de Bucaramanga, incluyendo el tipo de accidente, año, mes, día, tipo de vehículo involucrado y otras variables. Puedes acceder a la base de datos original y ampliar información de la misma a través de este [enlace](#).

Además, se ha generado una base de datos resultante del preprocesado realizado en la asignatura de minería de datos. Esta última es con la que se trabajará en la presente práctica, puedes acceder a ella a través de este [enlace](#).

Es importante destacar que este conjunto de datos también será utilizado en el trabajo final del Máster en Big Data y Data Science y que la visualización de datos desempeñará un papel fundamental en este proyecto, ya que permitirá representar de manera clara y efectiva la información contenida en el dataset.

Resumen del dataset

Nombre de la Variable	Tipo de Variable	Nombre de la Variable	Tipo de Variable	Nombre de la Variable	Tipo de Variable
ORDEN	int64	AUTOMOVIL	int64	VOLQUETA	int64
FECHA	string	CAMPAERO	int64	MOTO	int64
AÑO	int64	CAMIONETA	int64	BICICLETA	int64
MES	string	MICRO	int64	OTRO	int64
DÍA	string	BUSETA	int64	VIA 1	string
GRAVEDAD	string	BUS	int64	BARRIO	string
PEATON	int64	CAMION	int64	HORA	string

Contexto y motivación

En Colombia, el índice de accidentes vehiculares y de tránsito es muy alto en comparación con otros países. Por lo tanto, el objetivo es disminuir estas cifras identificando dónde, cuándo y cómo se producen los accidentes de tránsito. Para lograrlo, se utilizará una base de datos que contiene información sobre los accidentes registrados en la ciudad de Bucaramanga, Colombia, desde enero de 2012 hasta febrero de 2023.

Objetivo

El objetivo de esta investigación es utilizar los datos adecuados y proporcionar una visualización de datos que permita observar y comprender la problemática de los accidentes de tránsito en Bucaramanga, Colombia.

Metodología

En este trabajo de visualización de datos, se han seleccionado dos herramientas clave para abordar los desafíos de presentar de manera efectiva y comprensible la información extraída. **Python, con su librería Seaborn**, brinda la flexibilidad y potencia necesarias para manipular y analizar los datos, mientras que **Power BI** permite crear visualizaciones interactivas y dinámicas para transmitir de manera impactante los hallazgos.

Según Brown (2020), Python ha emergido como un lenguaje de programación ampliamente utilizado en la comunidad de ciencia de datos debido a su capacidad para manipular datos, realizar análisis estadísticos y generar visualizaciones de alta calidad. La librería Seaborn, en particular, ofrece una amplia gama de opciones de gráficos estadísticos y personalización, lo que permite presentar de manera efectiva los patrones y tendencias en los datos.

Por otro lado, Power BI ha demostrado ser una herramienta poderosa en el campo de la visualización de datos. Según Smith (2019), Power BI permite conectar diferentes fuentes de datos, realizar transformaciones y crear visualizaciones interactivas en un entorno intuitivo. Su capacidad para crear paneles dinámicos, informes interactivos y compartirlos con facilidad, ayuda a comunicar de manera efectiva los resultados del análisis de datos a diferentes audiencias.

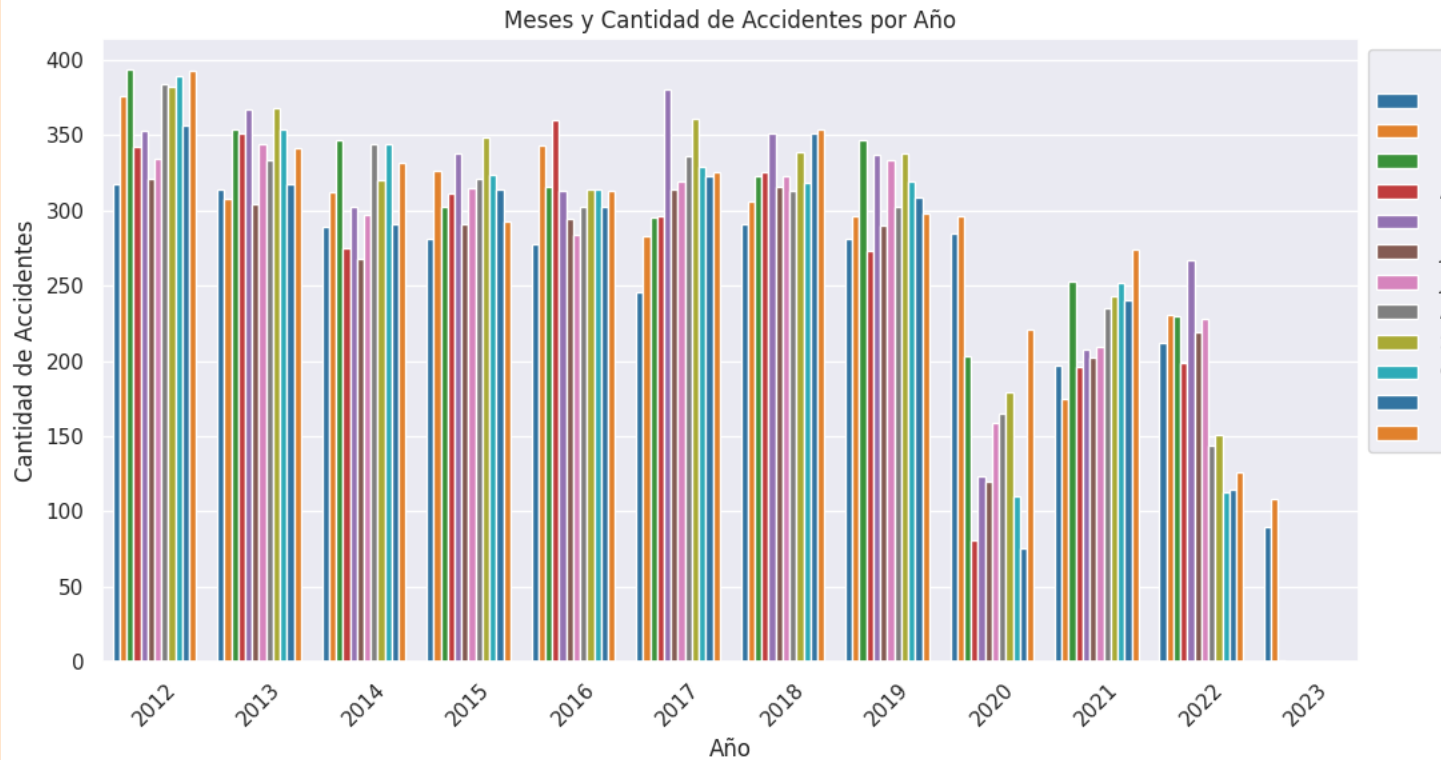
La combinación de Python con Seaborn y Power BI proporciona una solución integral para el análisis y la visualización de datos. Según lo planteado por Johnson (2018), el uso de herramientas especializadas en visualización, como Seaborn, junto con la potencia y la facilidad de

Resultados

A continuación se presentan los principales resultados y se discuten los resultados obtenidos en el trabajo.

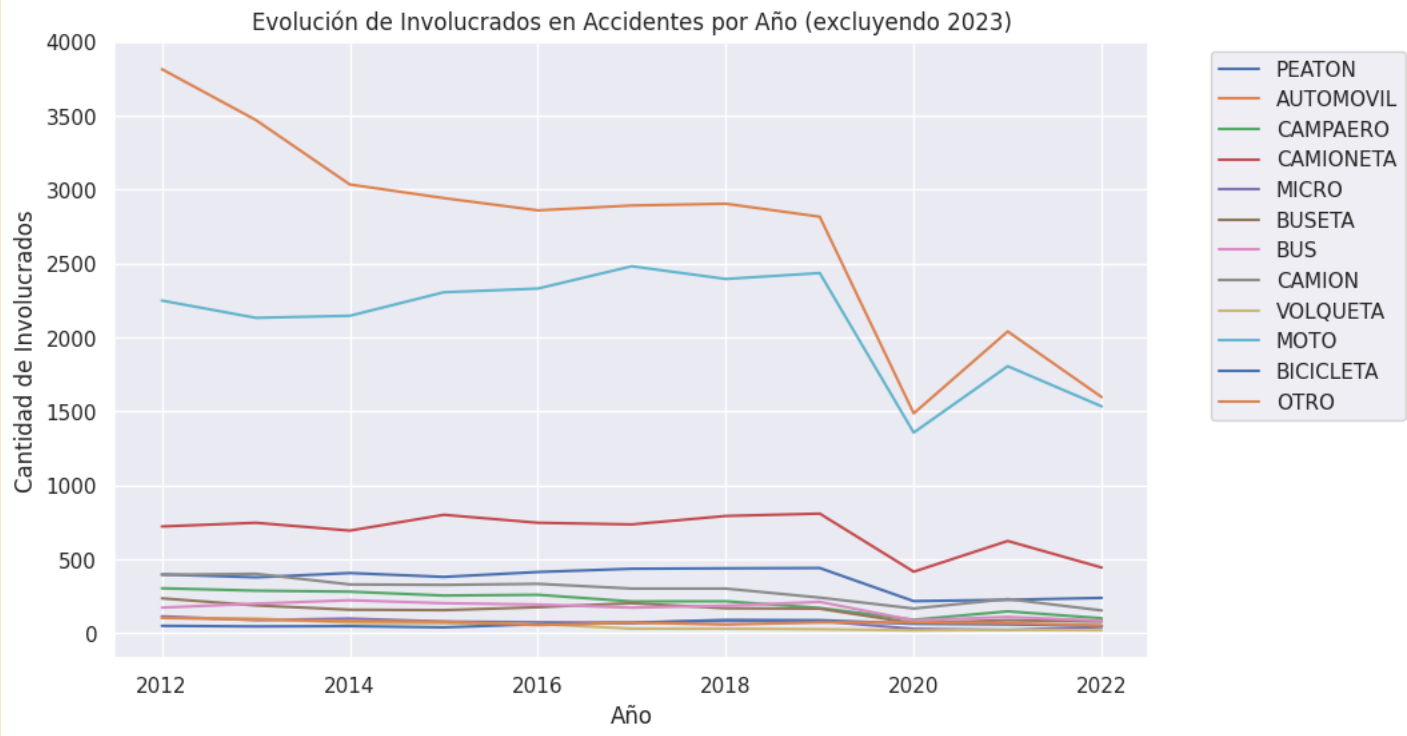
Gráficos utilizando Python (Seaborn)

Meses y cantidad de accidentes por Año



La gráfica titulada **Meses y cantidad de accidentes por Año** muestra un panorama general de la cantidad de accidentes segmentados por meses y años. Proporciona información sobre la evolución anual de la cantidad de accidentes, así como a nivel mensual. Por ejemplo, se puede observar que la cantidad de accidentes ha disminuido desde 2012 hasta febrero de 2023. El año 2012 tuvo la mayor cantidad de accidentes, principalmente en los meses de marzo, octubre y diciembre.

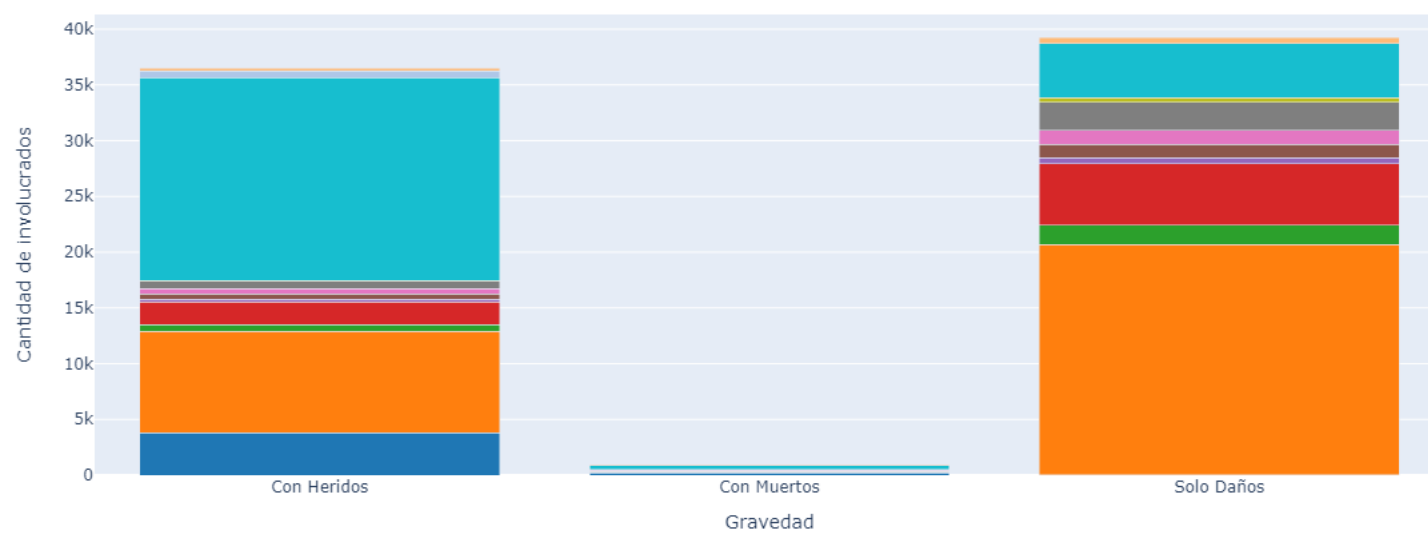
Evolución de Involucrados en Accidentes por Año (excluyendo 2023)



En el gráfico **Evolución de Involucrados en Accidentes por Año (excluyendo 2023)** se puede apreciar la evolución de los involucrados en accidentes a lo largo de los años 2012 al 2022. Se observa que los tipos de involucrados más frecuentes son "Automóvil" y "Moto" (motocicleta), aunque su cantidad ha disminuido a medida que también lo han hecho los accidentes en general. Estos dos tipos de involucrados ocupan consistentemente los primeros lugares a lo largo de la serie temporal, lo que indica que los automóviles y las motocicletas son los principales protagonistas de los accidentes en ese periodo.

Involucrados en accidentes por gravedad

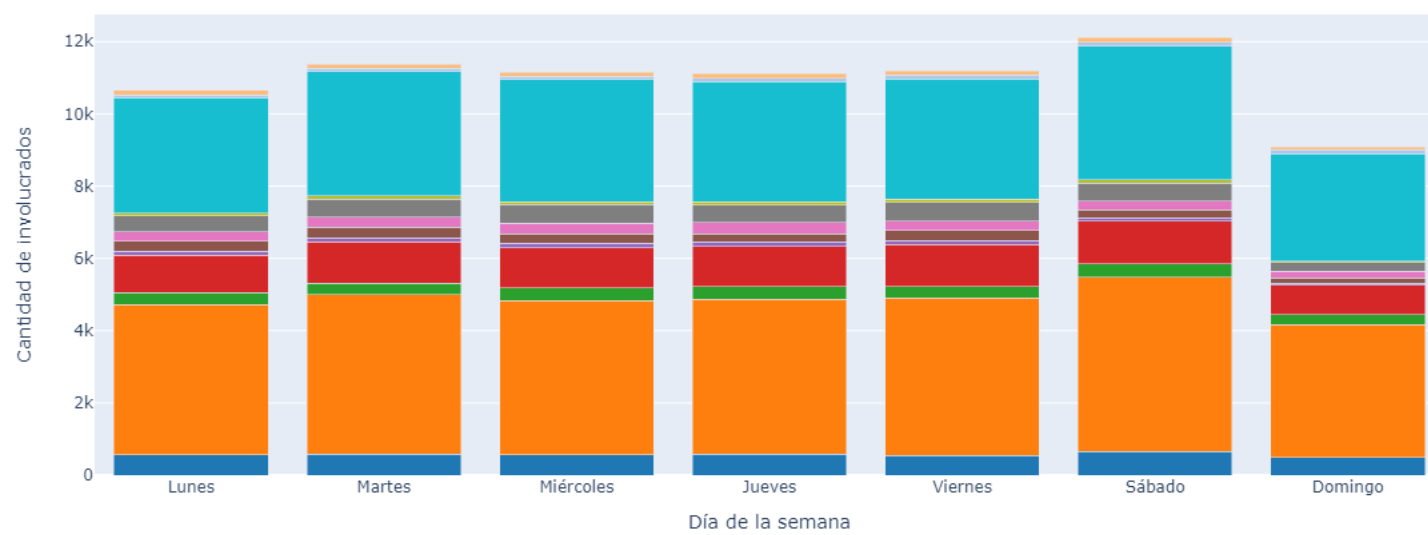
Involucrados en accidentes por gravedad



En el gráfico **Involucrados en accidentes por gravedad** se puede identificar claramente que los accidentes en los que hay heridos están ampliamente involucradas las motocicletas, mientras que en los accidentes en los que solo hay daños materiales están principalmente involucrados los automóviles. Esto es lógico, ya que los motociclistas están más expuestos a sufrir lesiones en un accidente que los pasajeros de un automóvil.

Cantidad de accidentes e involucrados por día de la semana

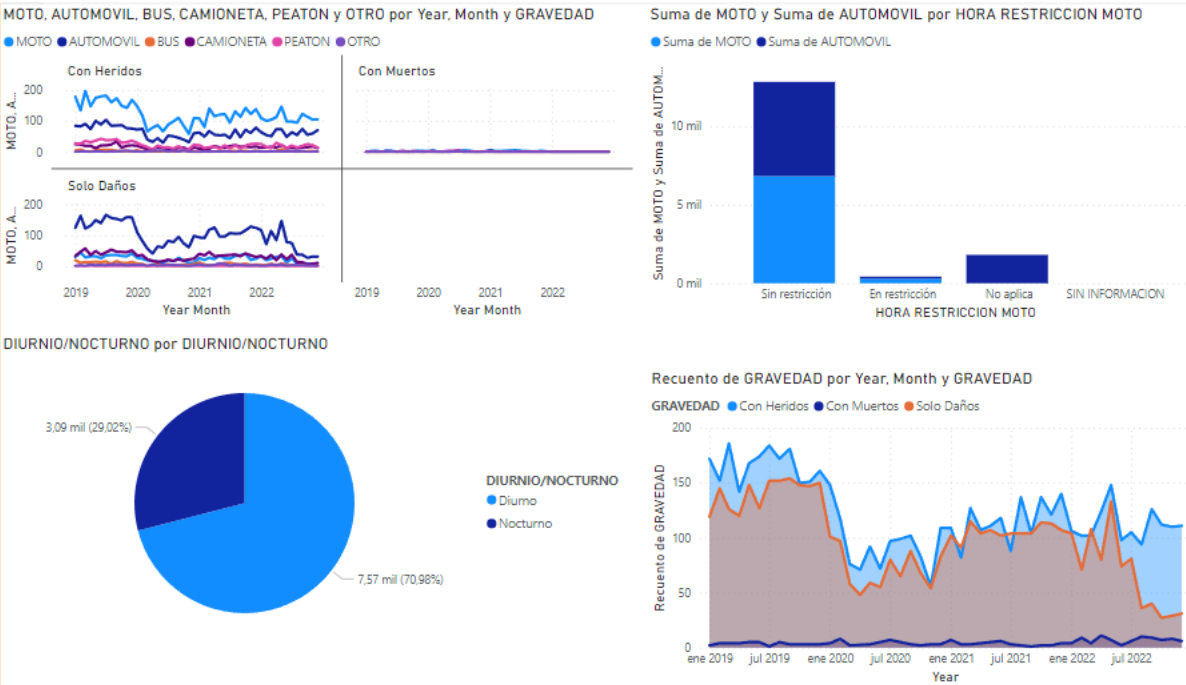
Cantidad de accidentes e involucrados por día de la semana



Cantidad de accidentes e involucrados por día de la semana. En esta gráfica se muestran los días de la semana con mayor y menor índice de accidentes, así como el tipo de involucrados. Según los datos del estudio, el sábado es el día con más accidentes, seguido del martes, mientras que el domingo es el día con menos accidentes. Esta última observación es lógica, ya que el tráfico disminuye considerablemente los domingos. Sin embargo, el alto índice de accidentes los martes podría generar diversas hipótesis, como el estrés, por ejemplo. En cuanto al tipo de involucrado, los automóviles y las motocicletas se mantienen como los principales protagonistas de los accidentes.

Gráficos utilizando Power BI

Cuadro de mando o Dashboard Accidentes de tránsito en Bucaramanga del 2012 al 2023



Dashboard hecho con PowerBI permite analizar varios aspectos de la problemática de los accidentes enfocando las cistas de manera interactiva sobre si los accidentes fueron “Diurno” o “nocturno”, se puede apreciar también la relevancia que tiene la variable “Restricción de motos” ya que cuando hay restricción de motos se reduce notablemente la cantidad de accidentes.

Conclusiones

En conclusión, en este trabajo de la asignatura de Visualización de Datos, se utilizó Python con Seaborn y Power BI para analizar una base de datos relacionada con los registros de accidentes de tránsito en la ciudad de Bucaramanga, Colombia. A través de la generación de diversos gráficos, se obtuvieron resultados relevantes. Uno de los hallazgos principales es que la cantidad de accidentes ha disminuido a partir del año 2019. Esto puede ser indicativo de mejoras en las medidas de seguridad vial o cambios en los patrones de comportamiento de los conductores en la ciudad. Además, se identificó que los automóviles y las motocicletas son los tipos de vehículos más frecuentemente involucrados en los accidentes. Estos dos tipos de vehículos ocupan consistentemente los primeros lugares a lo largo del periodo estudiado, lo que destaca su relevancia en los accidentes de tránsito en Bucaramanga. En cuanto a la gravedad de los accidentes, se observó que la mayoría de los accidentes registrados presentan únicamente heridos, seguidos por la categoría de solo daños materiales. Es interesante destacar que en los accidentes con heridos, las motocicletas están ampliamente involucradas. Esto puede atribuirse a la mayor exposición de los motociclistas a sufrir lesiones en comparación con los ocupantes de automóviles. En relación con los días de la semana, se determinó que el sábado es el día con mayor número de accidentes de tránsito en general, seguido por el martes. Esta información puede ser útil para dirigir esfuerzos de prevención y seguridad vial en los momentos más críticos y en los días de mayor incidencia de accidentes. En resumen, el análisis de la base de datos de accidentes de tránsito en Bucaramanga utilizando Python con Seaborn y Power BI ha permitido obtener información valiosa sobre la evolución de los accidentes, los tipos de vehículos involucrados, la gravedad de los accidentes y los patrones diarios de ocurrencia. Estos resultados proporcionan una base sólida para la toma de decisiones y la implementación de medidas de seguridad vial más efectivas en la ciudad.