

Dataset: Motor Vehicle Collisions – Crashes

- William Andrés Gómez
- Santiago Camilo Rey Benavides
- Daniel Alfredo Vidal de León

City of New York

There is no description for this organization

Topics

Local Government

Publisher

data.cityofnewyork.us

Contact

NYC OpenData

Share on Social Sites

Twitter

Facebook

This is a Non-Federal dataset covered by different Terms of Use than Data.gov. [See Terms](#)

Motor Vehicle Collisions - Crashes

Metadata Updated: July 26, 2024

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least \$1000 worth of damage (https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/ny_overlay_mv-104an_rev05_2004.pdf). It should be noted that the data is preliminary and subject to change when the MV-104AN forms are amended based on revised crash details. For the most accurate, up to date statistics on traffic fatalities, please refer to the NYPD Motor Vehicle Collisions page (updated weekly) or Vision Zero View (updated monthly).

Due to success of the CompStat program, NYPD began to ask how to apply the CompStat principles to other problems. Other than homicides, the fatal incidents with which police have the most contact with the public are fatal traffic collisions. Therefore in April 1998, the Department implemented TrafficStat, which uses the CompStat model to work towards improving traffic safety. Police officers complete form MV-104AN for all vehicle collisions. The MV-104AN is a New York State form that has all of the details of a traffic collision. Before implementing Trafficstat, there was no uniform traffic safety data collection procedure for all of the NYPD precincts. Therefore, the Police Department implemented the Traffic Accident Management System (TAMS) in July 1999 in order to collect traffic data in a uniform method across the City. TAMS required the precincts manually enter a few selected MV-104AN fields to collect very basic intersection traffic crash statistics which included the number of accidents, injuries and fatalities. As the years progressed, there grew a need for additional traffic data so that more detailed analyses could be conducted. The Citywide traffic safety initiative, Vision Zero started in the year 2014. Vision Zero further emphasized the need for the collection of more traffic data in order to work towards the Vision Zero goal, which is to eliminate traffic fatalities. Therefore, the Department in March 2016 replaced the TAMS with the new Finest Online Records Management System (FORMS). FORMS enables the police officers to electronically, using a Department cellphone or computer, enter all of the MV-104AN data fields and stores all of the MV-104AN data fields in the Department's crime data warehouse. Since all of the MV-104AN data fields are now stored for each traffic collision, detailed traffic safety analyses can be conducted as applicable.

Access & Use Information

- Public:** This dataset is intended for public access and use.
- Non-Federal:** This dataset is covered by different Terms of Use than Data.gov. [See Terms](#)
- License:** No license information was provided.

A photograph of a car accident scene. A white car is heavily damaged, with its front end crumpled. Two police officers in white uniforms are visible; one is in the foreground with his back to the camera, holding a clipboard, and another is in the background. The scene is on a wet street with Japanese signs in the background.

Pregunta de investigación

¿Cómo cambia la frecuencia de heridos en el tiempo?

Entendiendo el Dataset

- El dataset se llama 'Motor Vehicle Collisions - Crashes' y contiene 2'106.325 registros de eventos de colisiones informados por la policía de Nueva York. Este dataset se creó debido a que la policía de NY debe completar un informe policial en el caso de colisiones en las que alguien resulte herido o muera, o en las que haya daños por un valor de al menos \$1000.
- Por tanto, en este dataset se encuentran los registros de accidentes que ocurrieron hasta el año 2024 y que tuvieron eventos de muerte, heridos o daños superiores a los \$1000

The image features a magnifying glass held over a bar chart. The chart displays two data series, one in blue and one in green, across four quarters labeled Q1, Q2, Q3, and Q4. The magnifying glass is positioned over the Q2, Q3, and Q4 bars, making them appear larger and more detailed. The text 'Análisis Exploratorio de los Datos (EDA)' is overlaid in the center of the image in a white, sans-serif font. The background is a dark, blurred image of the same chart, emphasizing the data analysis theme.

Análisis Exploratorio de los Datos (EDA)

- Tenemos 27791 filas y 29 columnas.
- Existen características de tiempo (0 y 1), de ubicación (2-9), del número de heridos o Muertos (10-17), de la contribucion en el accidente de hasta 5 vehiculos (18-22) y codigos que identifican la collision y el tipo de vehiculos (23-28).
- Para responder la pregunta de investigación nos son de gran utilidad las columnas de la fecha y la hora en que ocurrio el accidente (0 y 1) y las columnas con información del número de heridos (10, 12, 14 y 16).

```

RangeIndex: 2106325 entries, 0 to 2106324
Data columns (total 29 columns):
 #   Column                                Dtype
---  -
 0   CRASH DATE                           object
 1   CRASH TIME                           object
 2   BOROUGH                              object
 3   ZIP CODE                             object
 4   LATITUDE                             float64
 5   LONGITUDE                            float64
 6   LOCATION                             object
 7   ON STREET NAME                       object
 8   CROSS STREET NAME                   object
 9   OFF STREET NAME                     object
10   NUMBER OF PERSONS INJURED            float64
11   NUMBER OF PERSONS KILLED             float64
12   NUMBER OF PEDESTRIANS INJURED        int64
13   NUMBER OF PEDESTRIANS KILLED         int64
14   NUMBER OF CYCLIST INJURED            int64
15   NUMBER OF CYCLIST KILLED             int64
16   NUMBER OF MOTORIST INJURED           int64
17   NUMBER OF MOTORIST KILLED            int64
18   CONTRIBUTING FACTOR VEHICLE 1        object
19   CONTRIBUTING FACTOR VEHICLE 2        object
20   CONTRIBUTING FACTOR VEHICLE 3        object
21   CONTRIBUTING FACTOR VEHICLE 4        object
22   CONTRIBUTING FACTOR VEHICLE 5        object
23   COLLISION_ID                         int64
24   VEHICLE TYPE CODE 1                  object
25   VEHICLE TYPE CODE 2                  object
26   VEHICLE TYPE CODE 3                  object
27   VEHICLE TYPE CODE 4                  object
28   VEHICLE TYPE CODE 5                  object

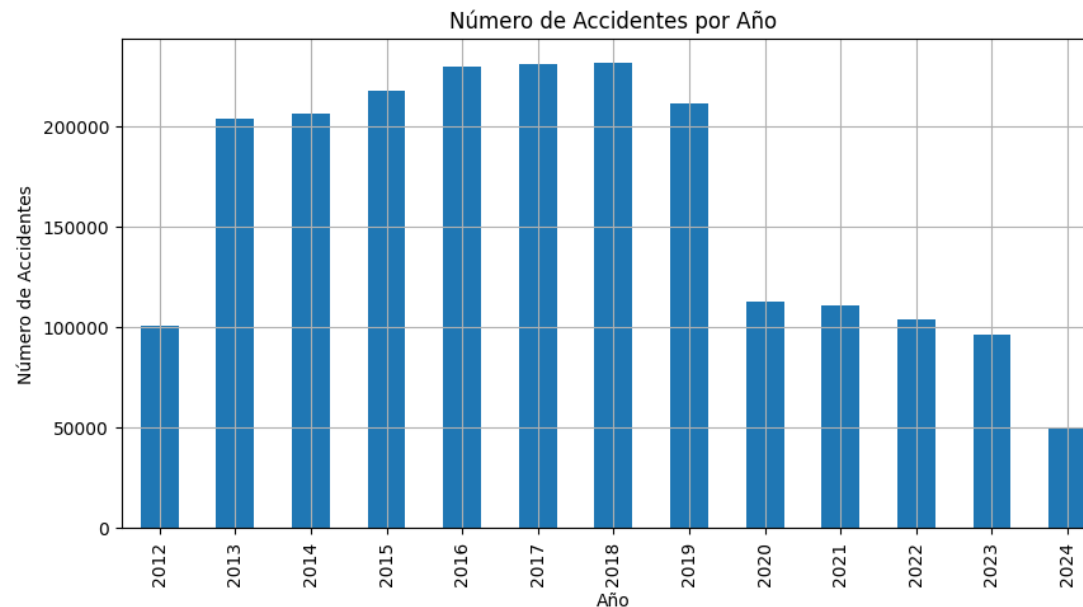
```

- Hemos reducido el número de características a las mencionadas en el paso anterior. Además, hemos utilizado la fecha y hora del accidente para crear un nuevo conjunto de características que indican el día del accidente, la hora, el mes y el año, para su posterior análisis.

```
RangeIndex: 2106325 entries, 0 to 2106324
Data columns (total 11 columns):
#   Column                                Dtype
---  -
0   CRASH DATE                            object
1   CRASH TIME                            object
2   NUMBER OF PERSONS INJURED             float64
3   NUMBER OF PEDESTRIANS INJURED         int64
4   NUMBER OF CYCLIST INJURED             int64
5   NUMBER OF MOTORIST INJURED            int64
6   CRASH DATETIME                        datetime64[ns]
7   YEAR                                  int32
8   MONTH                                 int32
9   HOUR                                  int32
10  DAY_OF_WEEK                           object
```

¿Cómo se distribuyen las observaciones a lo largo del tiempo?

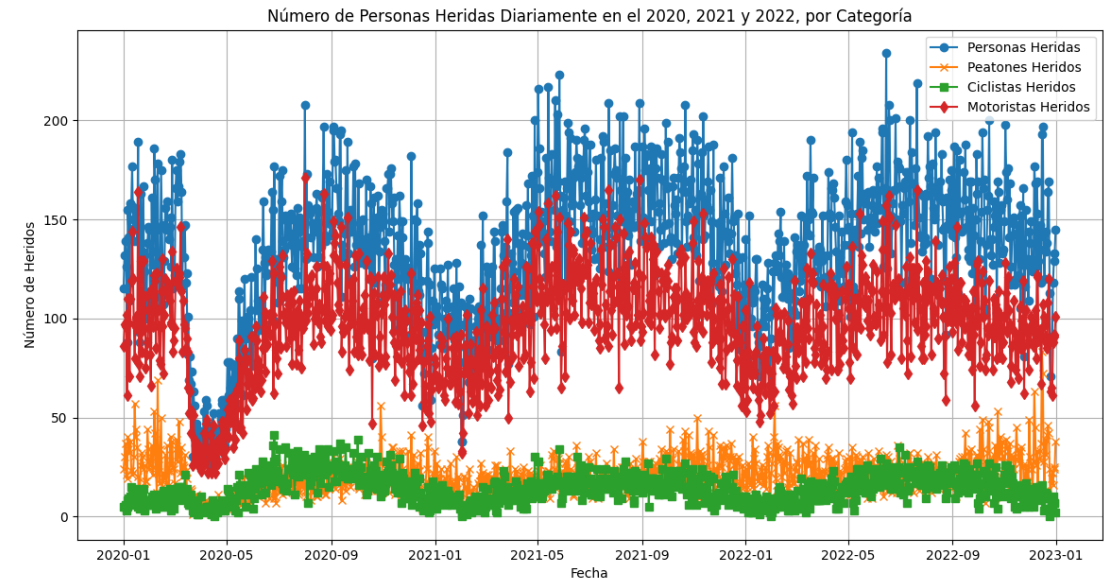
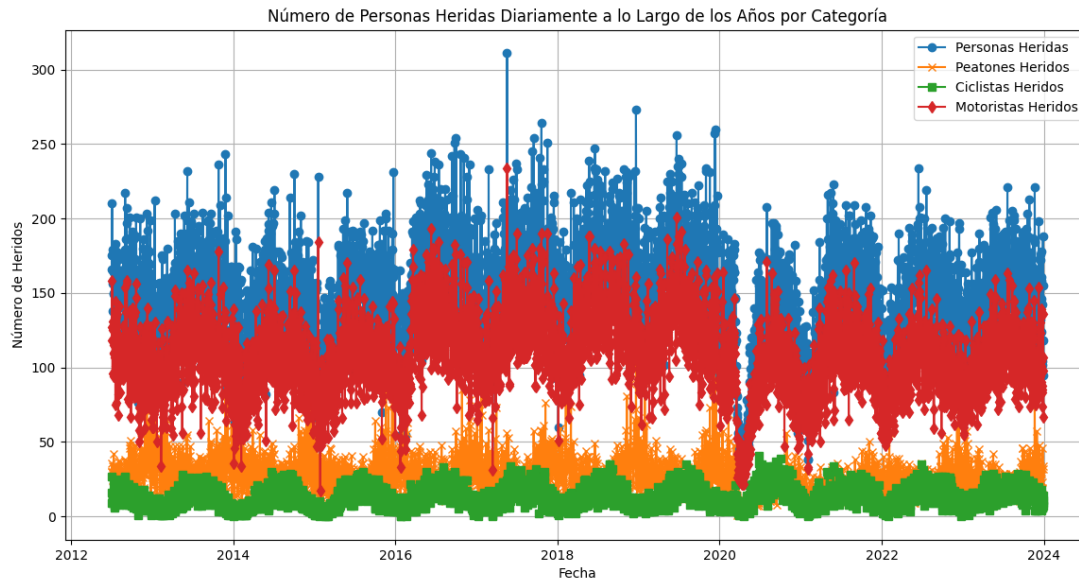
- La característica '**CRASH DATE**' tiene la fecha de ocurrido del accidente.
- Los datos representan los accidents ocurridos desde el año 2012 hasta el 2024.
- Podemos decir que a partir del 2020, la cantida de heridos en accidentes han reducido considerablemente con lo observada en la epoca antes de la pandemia.



YEAR	
2012	100545
2013	203742
2014	206038
2015	217694
2016	229831
2017	231007
2018	231564
2019	211486
2020	112916
2021	110553
2022	103886
2023	96591
2024	50472

A partir del año 2020 los heridos han disminuido considerablemente

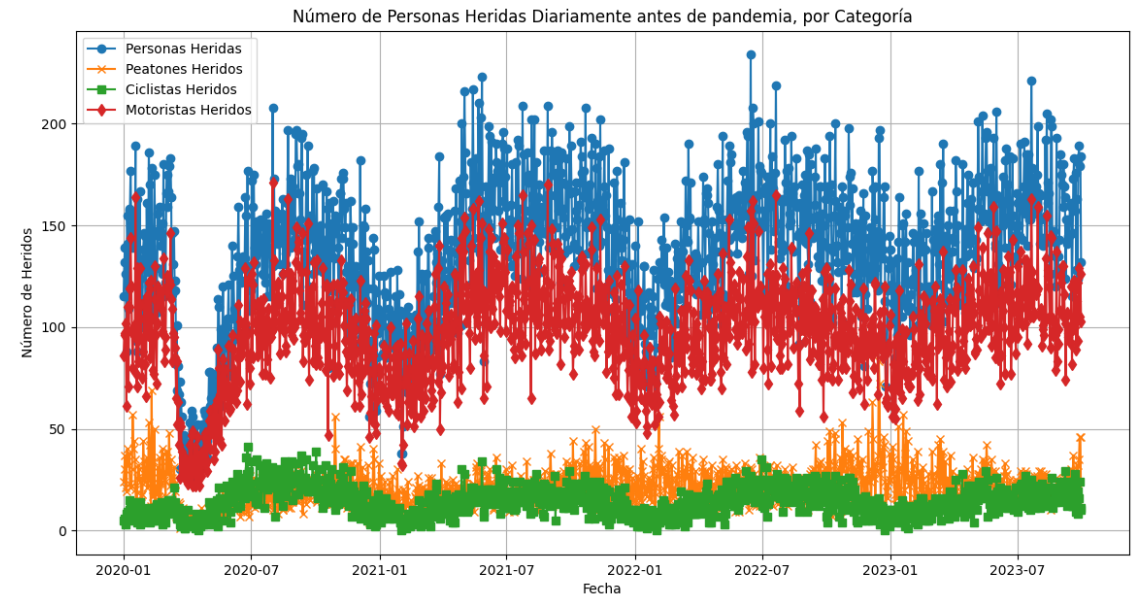
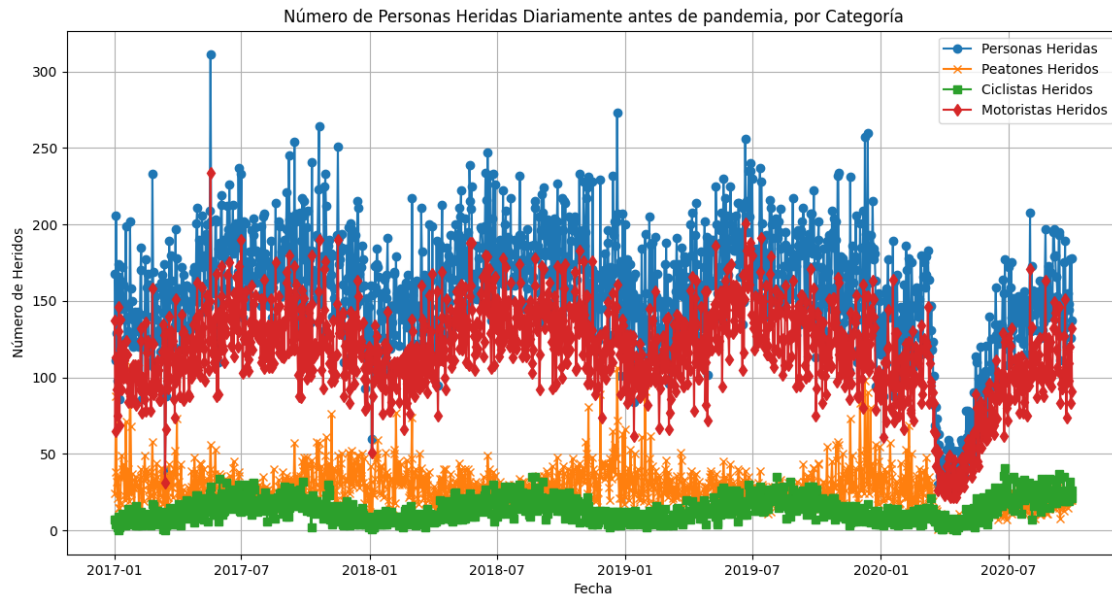
¿Cuándo ocurrieron los accidentes y cuántos heridos dejaron?



- Los datos muestran el número de heridos por día para 4 categorías de individuos involucrados a lo largo de los años. Se puede observar un comportamiento de periodicidad en todas las categorías. Hipotetizamos que las 4 categorías presentan los mismos picos de subida y bajada al mismo tiempo, teniendo la misma frecuencia y que se diferencien es en su amplitud y desplazamiento vertical.
- En la imagen de la izquierda vemos que la cantidad de heridos se comporta como una función sinusoidal, la cuál se ve interrumpida en el 2020 debido al COVID-19, y recupera su forma original post-pandemia. En la imagen de la derecha, podemos ver el efecto del confinamiento durante la pandemia del COVID-19, que provocó una fuerte disminución en el número de heridos.

¿Ocurren los picos de heridos en los mismos meses antes y después de la pandemia?

¿Cuándo ocurrieron los accidentes y cuántos heridos dejaron?



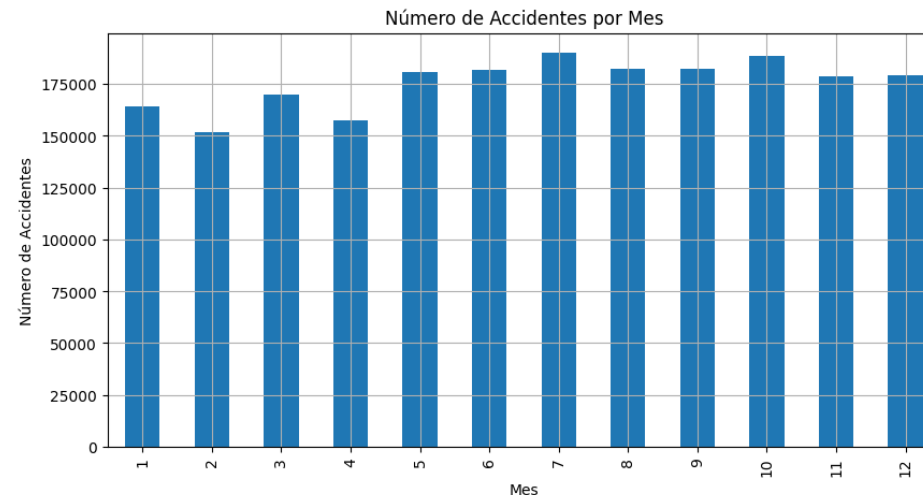
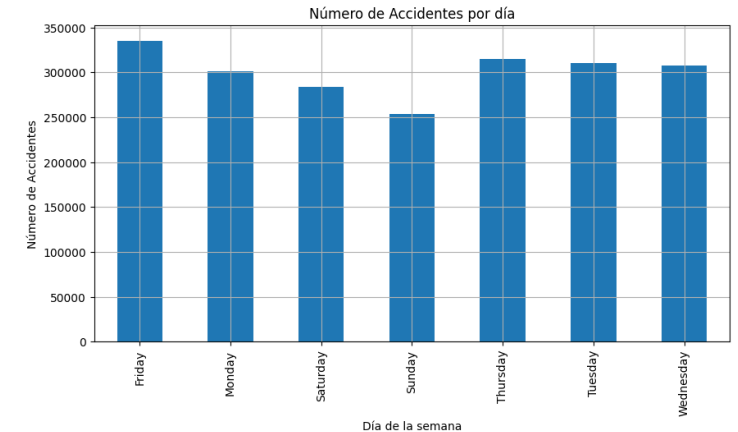
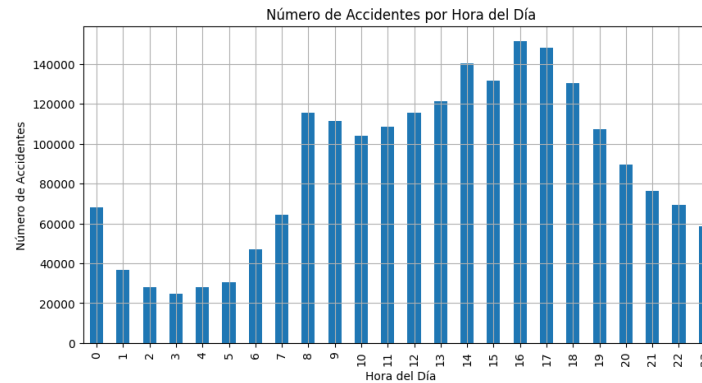
- A la izquierda tenemos el número de heridos antes de la pandemia, Podemos ver que los picos se presentaron aproximadamente en el mes 7 de los años 2017, 2018 y 2019.
- A la derecha tenemos lo ocurrido después de la pandemia del COVID-19, Podemos ver que los picos vuelven a ocurrir en el mes 7 del año 2021, 2022 y 2023. Además vemos que el pico de Julio del 2020 se vio afectado por la cuarentena y ocurrió aproximadamente en el mes 8, pero después la tendencia que se observa desde el año 2012, se estabilizó.

Los picos ocurren especialmente en el mes 7 del año (Julio), y esto ha sido así antes y después de la pandemia, el único efecto que dejó la cuarentena del año 2020 fue hacer que el pico de ese año se desplazara a el mes 8, pero todo se normalizó después de esto.

¿En el año 2021, cuándo ocurrieron los accidentes y cuántos heridos dejaron?

- Se contó el número de heridos por día de la semana, también se contaron los heridos por mes y finalmente se observe la cantidad de heridos para cada hora del día.
- De los accidentes que ocurrieron, el día con mayor número de heridos fue el **viernes** con una tendencia que incrementa desde los lunes y que tiene un pico el **miércoles**. Además, los días **domingo** son el día con menor número de heridos.
- Los accidentes ocurren con mayor frecuencia entre las **3 pm** y las **5 pm**, con una tendencia que incrementa de forma casi constante desde las **3 am** pero que presenta un fuerte pico de heridos a las **8 am**. Igualmente las **0 am**, horas representan un momento de altos heridos.
- El mayor número de heridos ocurren efectivamente en el mes 7, aunque vemos otro pico fuerte en el mes 10

Los días Viernes tienen el mayor número de heridos, así como los meses de Julio y Octubre. Igualmente, entre las 3 de la tarde y las 5 de la tarde se presenta el mayor número de heridos.

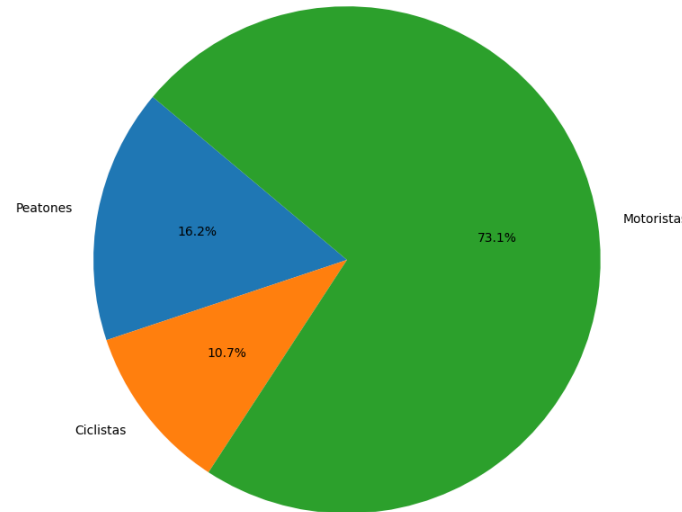


¿En el año 2021, cuándo ocurrieron los accidentes y cuántos heridos dejaron?

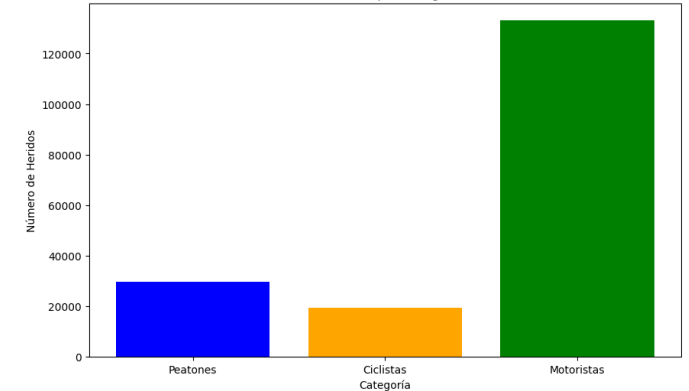
- Finalmente, se observó el número de heridos por categoría. Los **Motociclistas** son los más afectados en un **73.1%** de los casos, y los menos afectados fueron los ciclistas, con el 10.7% de los casos.
- También se registraron en este año un total de **399.381 accidentes**, en los cuales quedaron **182.104 heridos**. Esto sugiere que, por cada accidente, hay **0.45 heridos**.

En el 2021 hubo 399.381 accidentes reportados, dejando a 182.104 heridos con una media de 0.45 heridos por accidente. De los accidentados, el 73.1% son motociclistas y el restante son ciclistas y peatones. En el 2021 hubo más de 180 mil personas heridas por accidentes y más de 120 mil motociclistas igualmente heridos.

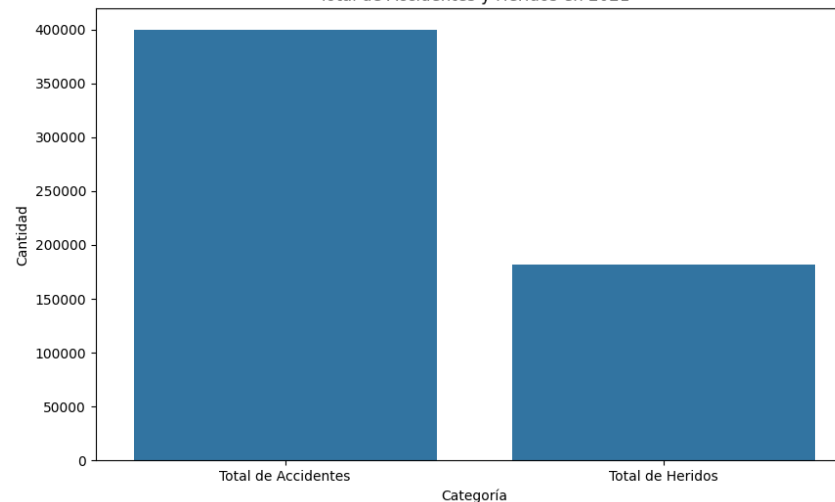
Número de Heridos por Categoría en 2021



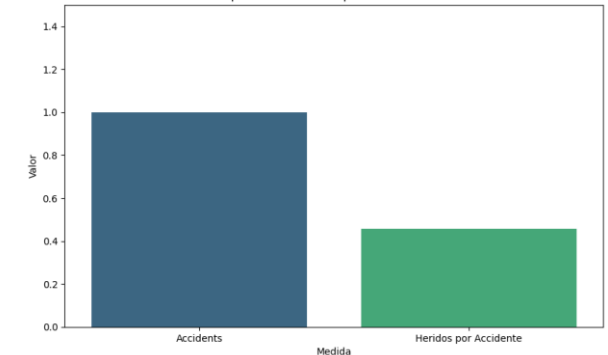
Número de Heridos por Categoría en 2021



Total de Accidentes y Heridos en 2021



Proporción de Heridos por Accidente 2021



CONCLUSIONES

- A partir del año 2020 los heridos han disminuido considerablemente
- Los picos ocurren especialmente en el mes 7 del año (Julio), y esto ha sido así antes y después de la pandemia, el único efecto que dejó la cuarentena del año 2020 fue hacer que el pico de ese año se desplazara a el mes 8, pero todo se normalizó después de esto.
- Los días Viernes tienen el mayor número de heridos, así como los meses de Julio y Octubre. Igualmente, entre las 3 de la tarde y las 5 de la tarde se presenta el mayor número de heridos.
- En el 2021 hubo 399.381 accidentes reportados, dejando a 182.104 heridos con una media de 0.45 heridos por accidente. De los accidentados, el 73.1% son motociclistas y el restante son ciclistas y peatones. En el 2021 hubo más de 180 mil personas heridas por accidentes y más de 120 mil motociclistas igualmente heridos.

¿Cómo cambia la frecuencia de heridos en el tiempo?

Los datos representan a los heridos en accidentes desde el año 2012 hasta la fecha actual. No todos los meses del año presentan el mismo número de heridos, sino que existen unos meses del año con picos de gran cantidad de heridos y otros meses donde este número disminuye en cierta medida. Observando los días de la semana en los que hubo accidentes, los días viernes presentan la mayor cantidad de heridos, mientras que el día domingo presenta la menor cantidad. Finalmente, a medida que transcurre la jornada del día, hay un mayor número de heridos reportados, alcanzando su máximo entre las 3 p.m. y las 5 p.m., con un elevado número de heridos a las 8 a.m.
