

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 27791 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

A magnifying glass is positioned 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 centered over the Q2 and Q3 bars. The text 'Análisis Exploratorio de los Datos (EDA)' is overlaid in white on the chart. A '1,000' label is visible on the y-axis in the upper right corner.

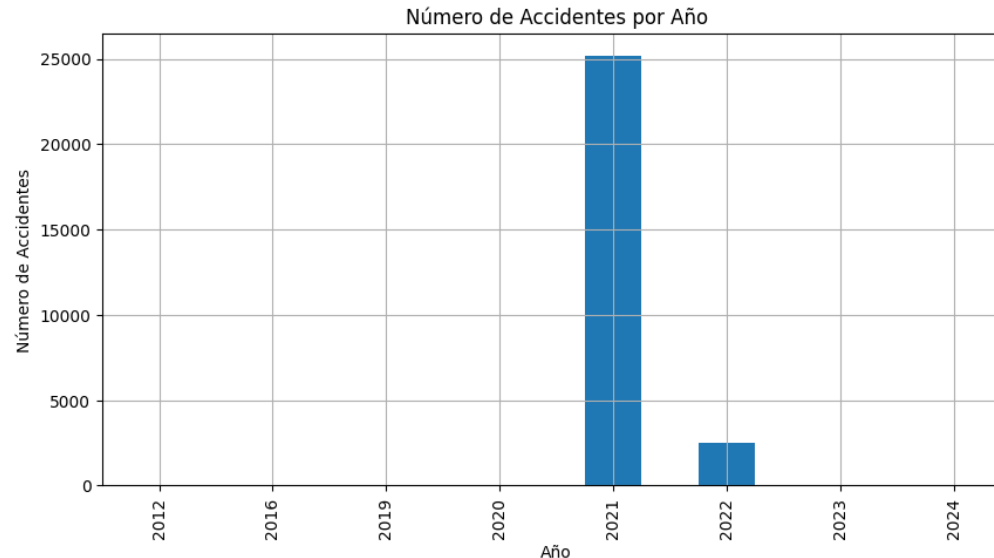
Análisis Exploratorio de los Datos (EDA)

- 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: 27791 entries, 0 to 27790
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   CRASH DATE                            27791 non-null  object
1   CRASH TIME                            27791 non-null  object
2   BOROUGH                               18168 non-null  object
3   ZIP CODE                             18165 non-null  float64
4   LATITUDE                             25518 non-null  float64
5   LONGITUDE                             25518 non-null  float64
6   LOCATION                             25518 non-null  object
7   ON STREET NAME                        20341 non-null  object
8   CROSS STREET NAME                     12918 non-null  object
9   OFF STREET NAME                       7450 non-null   object
10  NUMBER OF PERSONS INJURED              27791 non-null  int64
11  NUMBER OF PERSONS KILLED               27791 non-null  int64
12  NUMBER OF PEDESTRIANS INJURED          27791 non-null  int64
13  NUMBER OF PEDESTRIANS KILLED           27791 non-null  int64
14  NUMBER OF CYCLIST INJURED              27791 non-null  int64
15  NUMBER OF CYCLIST KILLED               27791 non-null  int64
16  NUMBER OF MOTORIST INJURED              27791 non-null  int64
17  NUMBER OF MOTORIST KILLED              27791 non-null  int64
18  CONTRIBUTING FACTOR VEHICLE 1           27651 non-null  object
19  CONTRIBUTING FACTOR VEHICLE 2           21623 non-null  object
20  CONTRIBUTING FACTOR VEHICLE 3           2887 non-null   object
21  CONTRIBUTING FACTOR VEHICLE 4            773 non-null   object
22  CONTRIBUTING FACTOR VEHICLE 5            213 non-null   object
23  COLLISION_ID                           27790 non-null  float64
24  VEHICLE TYPE CODE 1                     27458 non-null  object
25  VEHICLE TYPE CODE 2                     18701 non-null  object
26  VEHICLE TYPE CODE 3                     2694 non-null   object
27  VEHICLE TYPE CODE 4                      728 non-null   object
28  VEHICLE TYPE CODE 5                      203 non-null   object
```


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

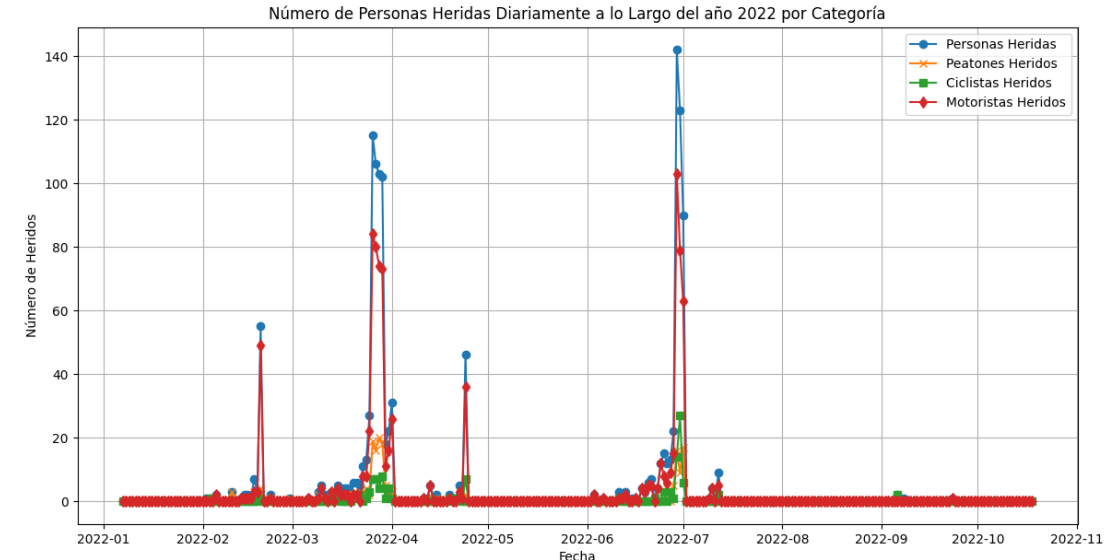
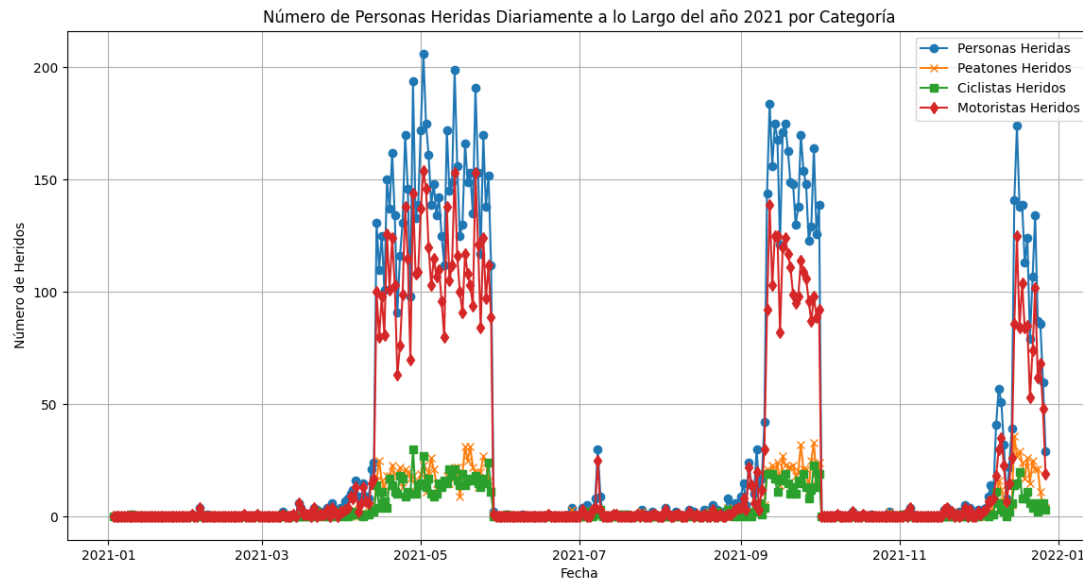
- La característica '**CRASH DATE**' tiene la fecha de ocurrido del accidente.
- Los datos están desbalanceados por año.



| YEAR | |
|------|-------|
| 2012 | 1 |
| 2016 | 1 |
| 2019 | 2 |
| 2020 | 22 |
| 2021 | 25191 |
| 2022 | 2523 |
| 2023 | 21 |
| 2024 | 30 |

El 90.6% de los datos corresponden al año 2021 y el 9.08% al año 2022.

¿Cuándo ocurrieron los accidentes y cuántos heridos dejaron?

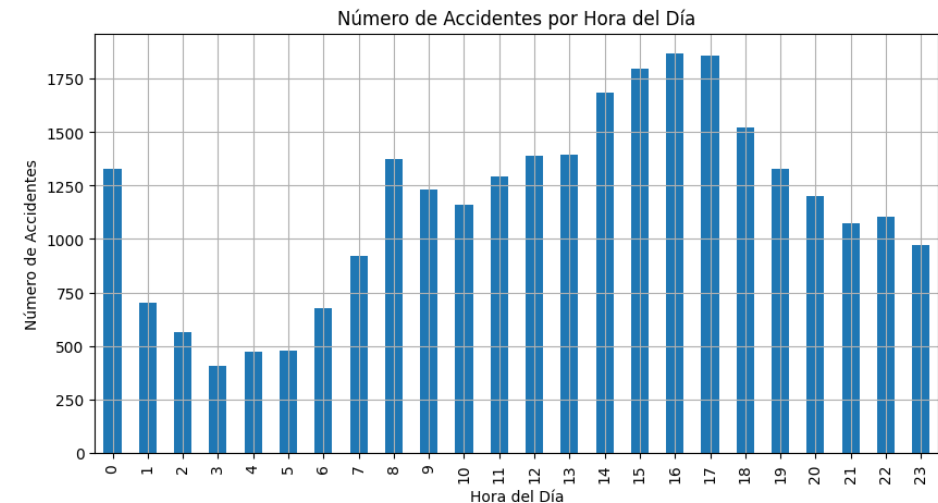
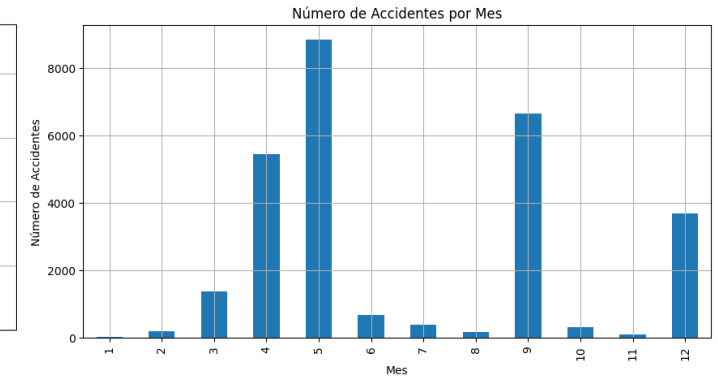
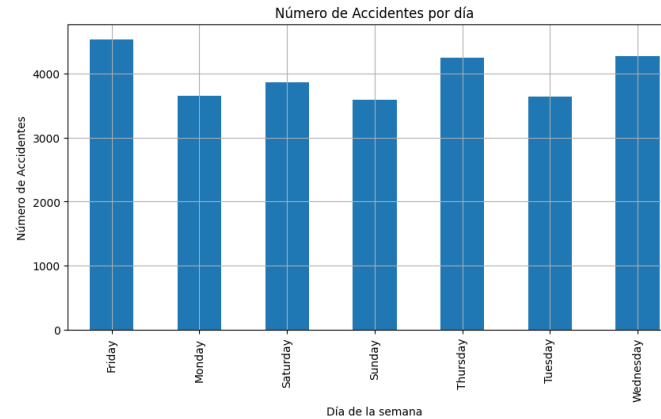


- Los datos muestran el número de heridos por día para 4 categorías de individuos involucrados en los años 2021 y 2022.
- Pensamos que los picos de los años 2021 y 2022 deberían reflejarse periódicamente a lo largo de todos los años, causados por ejemplo por eventos sociales como las temporadas de vacaciones cuando más vehículos se movilizan. Efectos político-económicos pueden tener un factor diferencial en ciertos años, pero la tendencia general debería mantenerse. Existe una fuerte relación entre el número de heridos, el número de accidentes y el número de vehículos que circulan las calles, ya que hay más probabilidad de que exista un accidente.
- Por lo tanto, pensamos que hay gran riesgo de accidentalidad en los meses de **Abril a Mayo**, el mes de **Septiembre** y el mes de **Diciembre**. También, con una menor cantidad de heridos por día, pero igualmente de gran importancia está el mes de **Julio**, por lo evidenciado en el año 2022.

Los heridos ocurrieron especialmente en los meses de Abril, Mayo, Julio, Septiembre y Diciembre y dejaron cientos de heridos.

¿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 en **Mayo** y **Septiembre**, además **Marzo**, **Abril** y **Diciembre** presentan un nivel importante de heridos.



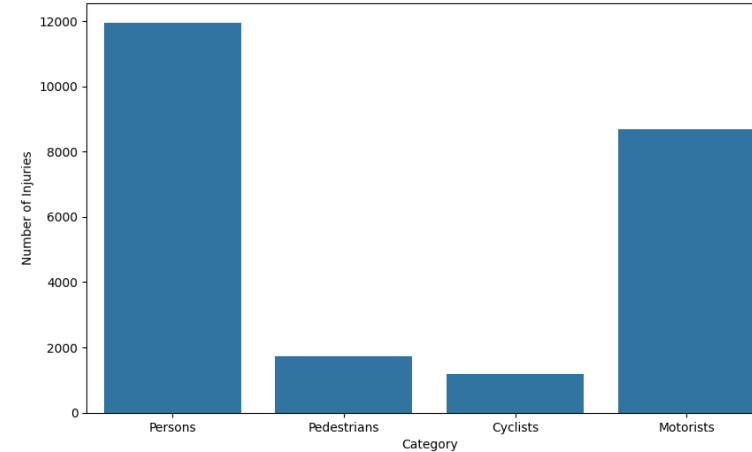
Los días Viernes tienen el mayor número de heridos, así como los meses de Mayo y Septiembre. Igualmente, entre las 3 de la tarde y las 5 de la tarde se presenta el mayor número de heridos.

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

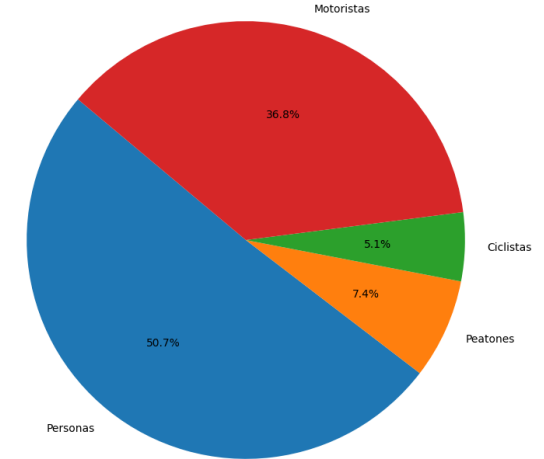
- Finalmente, se observó el número de heridos por categoría. Las **personas** son las más afectadas en un **50.7%** de los casos, y los menos afectados fueron los ciclistas, con el 5.1% de los casos.
- También se registraron en este año un total de **25,191 accidentes**, en los cuales quedaron **23,576 heridos**. Esto sugiere que, por cada accidente, hay **0.9 heridos en promedio**.

En el 2021 hubo 25,191 accidentes reportados, dejando a 23,576 heridos con una media de 0.93 heridos por accidente. De los accidentados, el 50% son personas, el 37% motociclistas y el restante son ciclistas y peatones. En el 2021 hubo cerca de 12 mil personas heridas por accidentes y casi 8 mil motociclistas igualmente heridos.

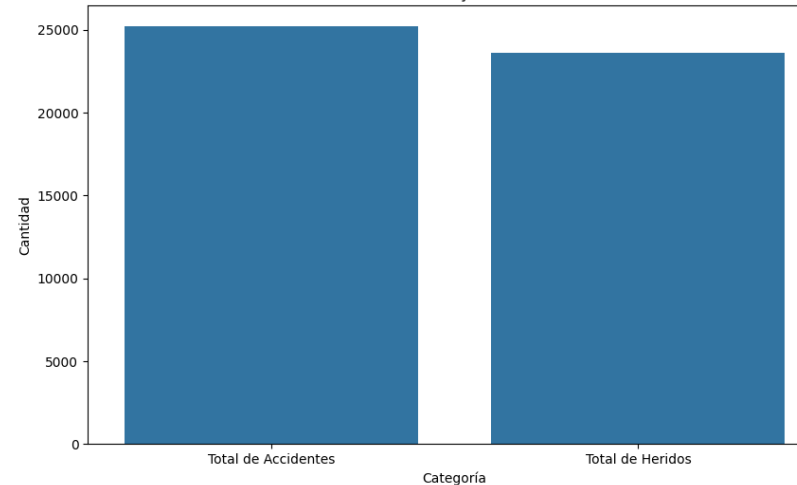
Número de Heridos por categoría en el 2021



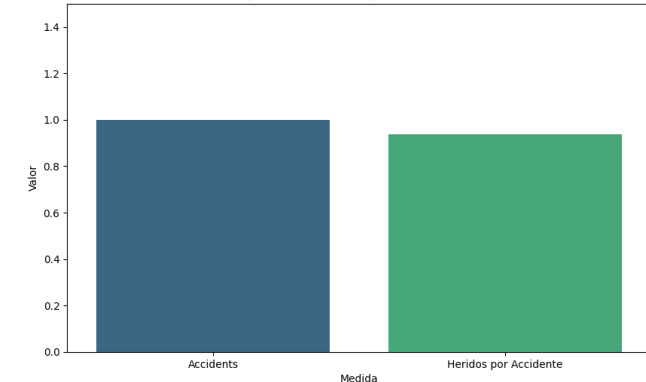
Número de Heridos por Categoría en 2021



Total de Accidentes y Heridos en 2021



Proporción de Heridos por Accidente 2021



CONCLUSIONES

- El 90.6% de los datos corresponden al año 2021 y el 9.08% al año 2022.
- Los heridos ocurrieron especialmente en los meses de Abril, Mayo, Julio, Septiembre y Diciembre y dejaron cientos de heridos.
- Los días Viernes tienen el mayor número de heridos, así como los meses de Mayo y Septiembre. Igualmente, entre las 3 de la tarde y las 5 de la tarde se presenta el mayor número de heridos.
- En el 2021 hubo 25,191 accidentes reportados, dejando a 23,576 heridos con una media de 0.93 heridos por accidente. De los accidentados, el 50% son personas, el 37% motociclistas y el restante son ciclistas y peatones. En el 2021 hubo cerca de 12 mil personas heridas por accidentes y casi 8 mil motociclistas igualmente heridos.

¿Cómo cambia la frecuencia de heridos en el tiempo?

Los datos representan a los heridos en accidentes para el año 2021 en un 90% y para el 2022 en un 9%. No todos los meses del año presentan un número grande de heridos, pero existen algunos meses como abril, mayo, junio, septiembre o diciembre donde el número de heridos es de varios miles. 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.
