



UANL

LA SEVERIDAD EN LOS ACCIDENTES AUTOMOVILÍSTICOS

Plantamiento de Hipótesis

Contreras Garcia G.A., Sifuentes Soto A., Solis Agundis V.



FCFM

INTRODUCCIÓN

En Estados Unidos se presentan más de 6 millones de accidentes automovilísticos al año y además se estima que, después de presentarse un choque, la vialidad se ve directamente afectada, provocando un paro en el tráfico de 90 minutos en promedio. Es importante poder localizar las zonas que presentan accidentes con mayor severidad, ya que esto ayudaría al cuerpo policiaco a definir en qué áreas hay que tener una mayor vigilancia, y así, a la hora de presentarse un accidente, podrían atender el llamado más rápido.

Se tiene como hipótesis que existen ciertas variables que influyen directamente en el valor de la severidad de los choques.

OBJETIVOS

Localizar los estados donde se presenta un mayor número de accidentes así como los lugares donde los accidentes tienen una severidad mayor.

Realizar un modelo predictivo que dé como respuesta la severidad que presentará un accidente, además de reconocer cuales son las variables que más afectan a ese valor.



METODOLOGÍA



Primero se realizó la limpieza de nuestra base de datos. Eliminamos columnas que no brindaban información relevante y generamos otras (año, mes, día de la semana y hora).

Como segundo paso se hizo un análisis de los valores faltantes. Se realizó un segundo filtro donde eliminamos columnas que no mostraban suficiente información, y eliminamos filas con datos nulos, además de cambiar los valores de las variables binarias por 0's y 1's con la finalidad de poder usarlas como variables predictoras.

Se generaron pruebas estadísticas para medir el tiempo promedio que se mantenía parado el tráfico, su varianza y los cuartiles.

Finalmente se elaboraron diferentes gráficas para analizar la severidad a lo largo del tiempo, donde de todas ellas podemos destacar la **Figura 1** referente al conteo de los accidentes producidos por año y separados por su severidad, la cual nos da una visión preliminar de cual severidad es la que se produce con más frecuencia en cada choque.

Conteo de accidentes por Año

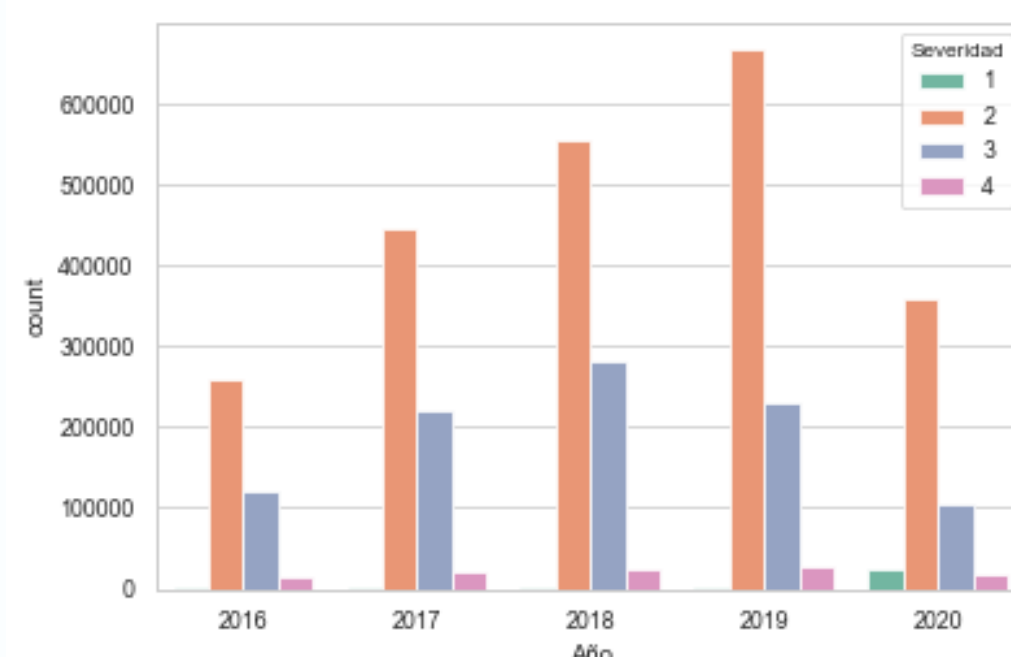


FIGURA 1

RESULTADOS

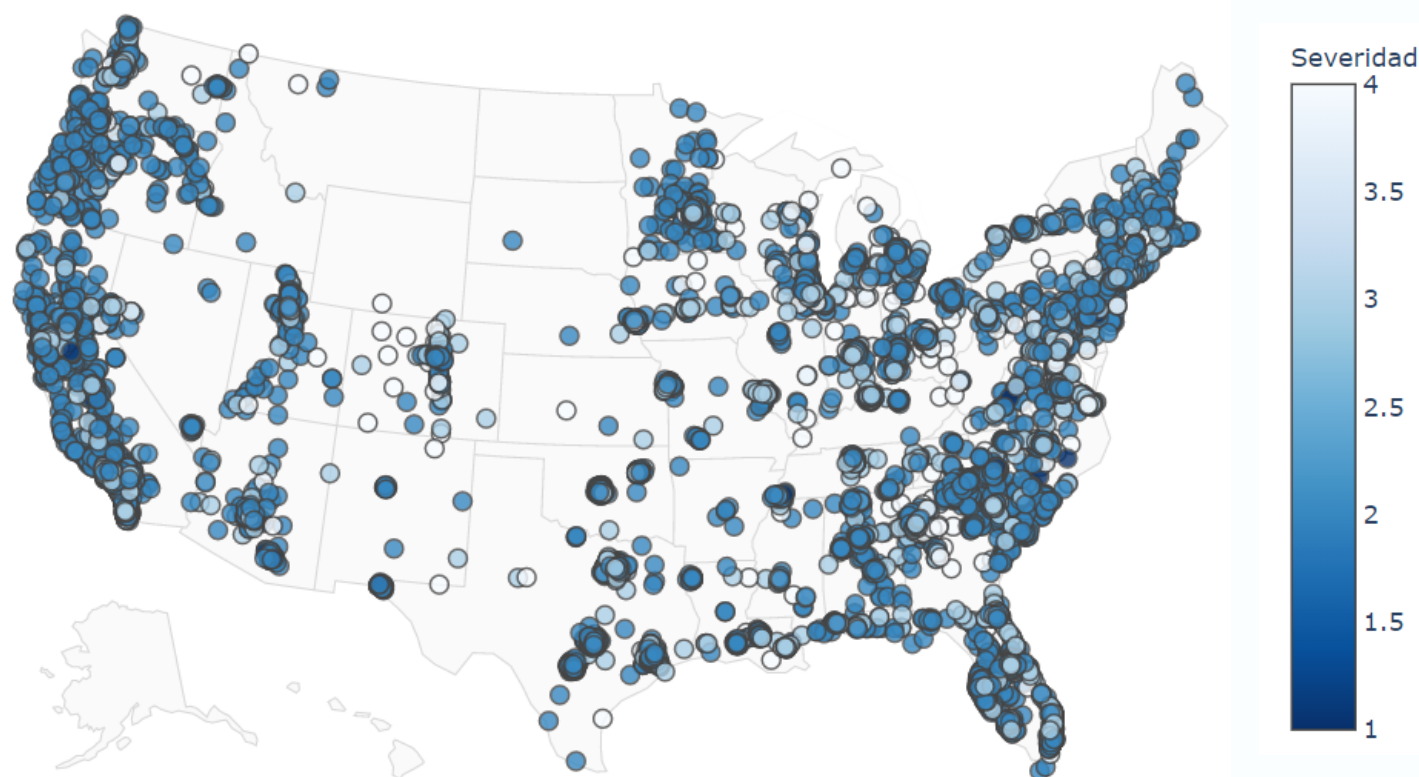


FIGURA 2

Utilizando las variables de longitud, latitud y severidad logramos generar un mapa de calor (**Figura 2**), que nos permitió visualizar los estados en donde más se producen accidentes, así como la severidad de cada uno de ellos.

Concluimos de esta manera que los estados con mayor frecuencia y severidad son **California, Texas y Florida**.

Estas 3 zonas son las que presentan un mayor riesgo.



Generamos un bosque aleatorio con la finalidad de producir como respuesta la severidad de un accidente tomando como variables predictoras sus características.

Se obtuvo con este modelo una exactitud del **93.5%** en las predicciones que genera.

Además de ello, nos arrojó una lista con la puntuación de la importancia de cada una de estas características, teniendo así en la **Figura 3** enlistadas las 10 características que mayor influencia, siendo la de mayor puntaje la variable de **distancia**, esta variable muestra la longitud de la extensión de la carretera afectada por el accidente.

FIGURA 3

Visualización de características importantes



RECURSOS



ANACONDA



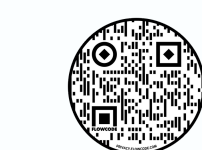
Base de datos:
US Accidents
(3.5 million records)

Seaborn

NumPy

matplotlib

Pandas



Base de datos
manipulada.

Gracias a este modelo, se puede realizar una predicción de la severidad del choque luego de que se reciba el informe del mismo. Esto se puede utilizar para llamar a un mayor número de patrullas al lugar del accidente si la severidad es muy grande, también gracias a este proyecto se tiene una visión más clara de las variables que determinan dicha severidad, además de lograr distinguir las zonas con mayor peligro, con ello se espera que se realice una mayor vigilancia en estos estados, con el fin de estar más preparados a la hora de atender un llamado de emergencia.

Como objetivo se tenía el predecir donde se producirá un choque, no logramos llegar a ello debido a la dificultad que implicaba el predecir una localización específica, pero sería una aplicación importante que podría cubrirse en algún trabajo futuro, además de ello, se podría crear una aplicación que en tiempo real registre donde pasó un accidente, y uno como usuario pueda ver en que calle se produjo el choque y la severidad de este.

CONCLUSIÓN