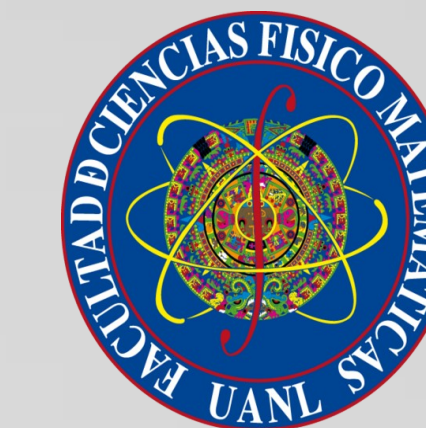


# Variables que influyen en la severidad de un accidente automovilístico

## Analizando los accidentes automovilísticos en Estados Unidos



### Introducción

Como futuros profesionistas de la carrera de Actuaría, la severidad de los accidentes en cualquier ámbito suele despertar un gran interés en nosotros, es por esto que decidimos realizar nuestro análisis basándonos en esta variable.

Día a día es mas común que la gente cuente con un auto propio, lo cual incrementa de forma directa la frecuencia de accidentes automovilísticos por todo el mundo. Para este análisis nos enfocaremos en los choques que tuvieron lugar en Estados Unidos entre los años 2016 y 2019 para ver si esta frecuencia que vemos se encuentra en aumento, afecta a la severidad de los accidentes automovilísticos y si algunas hay variables a las que todos tengamos acceso que nos ayuden a predecir que tan severo será un accidente.



### Objetivo

Encontrar en que estado y en que condiciones climáticas la severidad tiende a ser mayor y revisar si existe correlación entre la cantidad de accidentes y su severidad.

Analizar si la humedad, la temperatura y el tiempo de duración del choque tienen un impacto significativo en la severidad.

### Recursos



Base de datos: US Accidents (3.0 million records)

### Metodología

Pasos a seguir en nuestro análisis:

- 1) Se limpiaron los datos para quedarnos solamente con aquellos meramente necesarios.
- 2) Se realizó un análisis estadístico, obteniendo indicadores como la media, la varianza, cuartiles, etc.
- 3) Con ayuda del análisis, se generaron subgrupos basados en la severidad media por estado y por tipo de clima.
- 4) Se realizan técnicas de predicción con las variables de interés, para ver que tan significativas son a la hora de predecir la severidad.

Severidad Media vs Frecuencia por estado

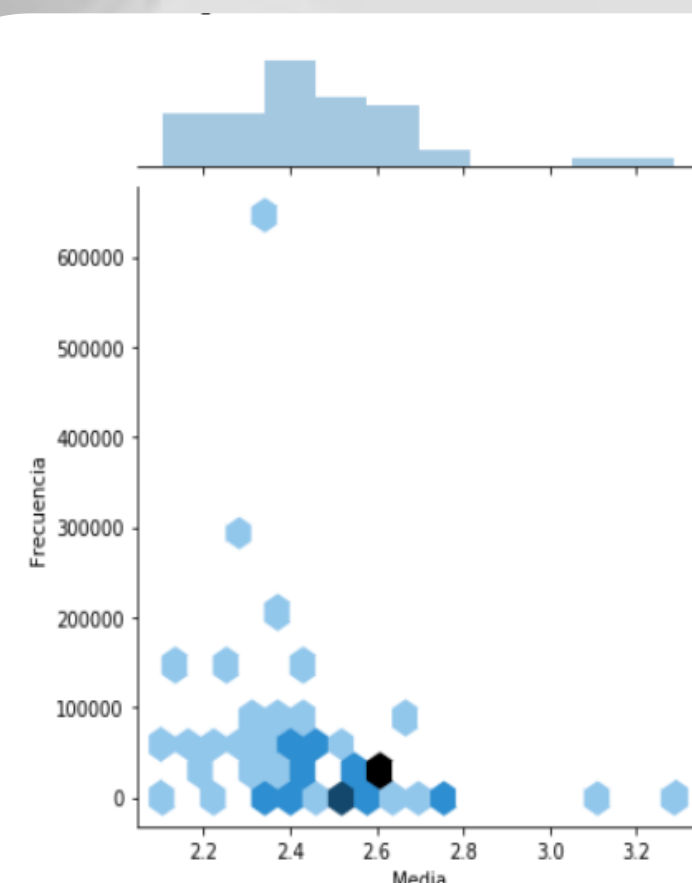


Figura 1.1

Severidad Media vs Frecuencia por clima

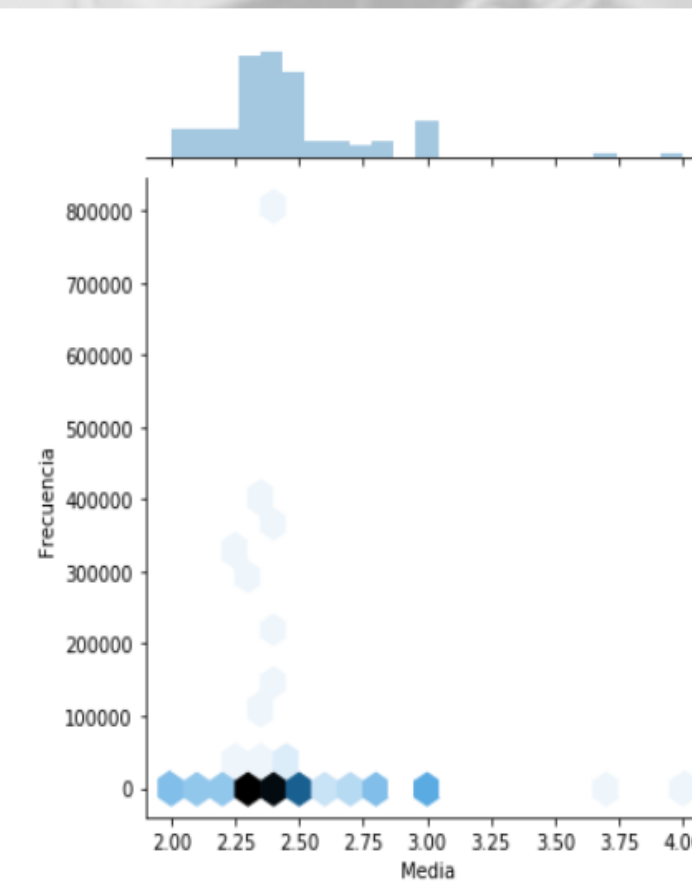


Figura 1.2

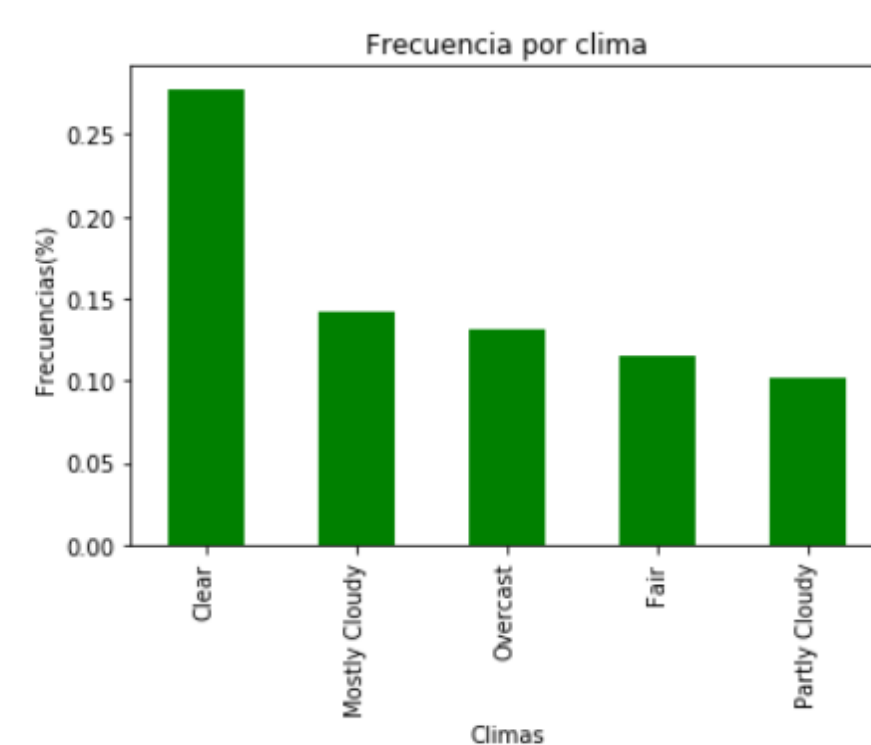


Figura 2.1

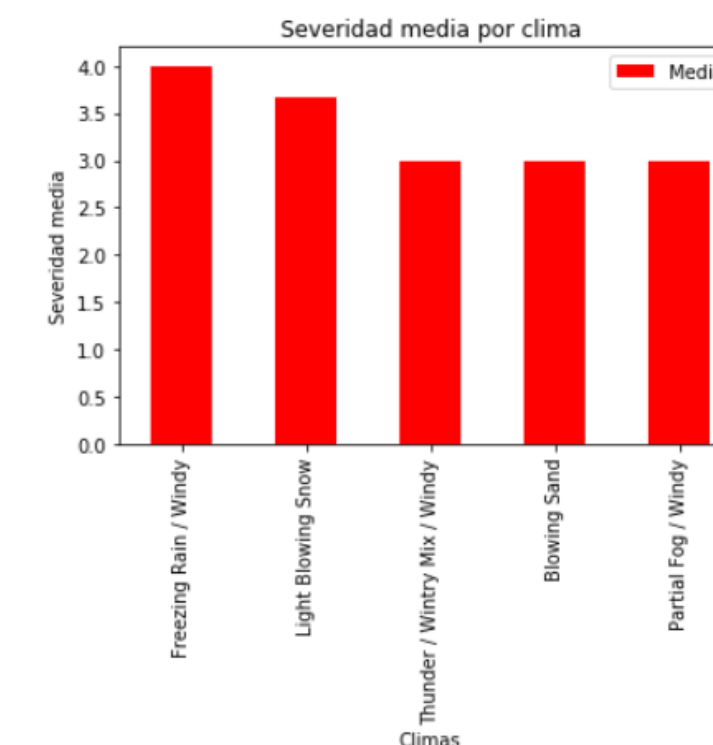


Figura 2.2

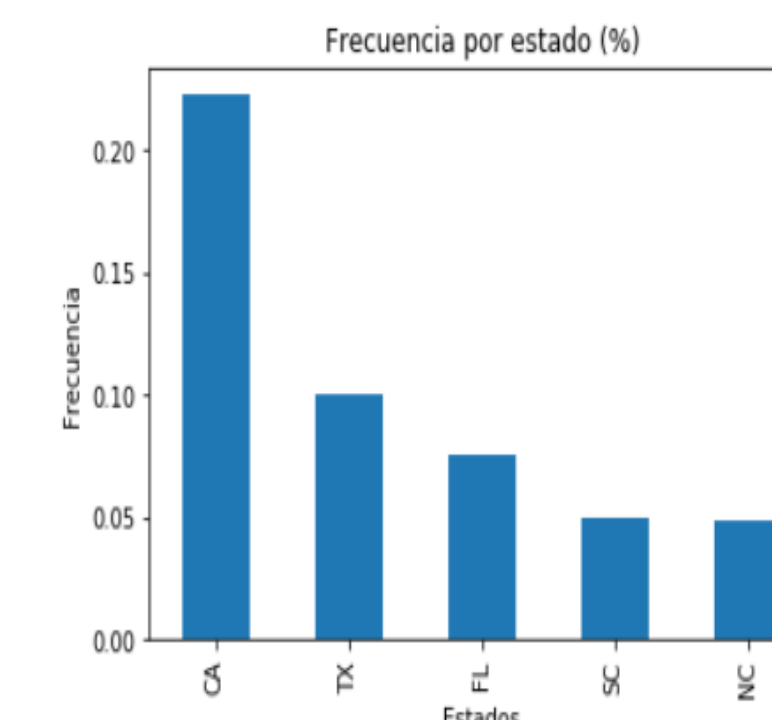


Figura 3.1

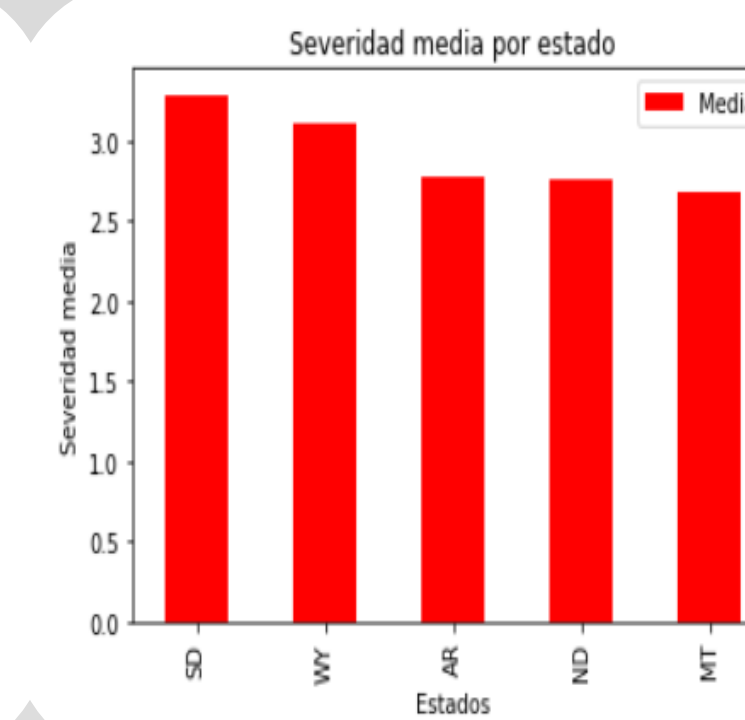


Figura 3.2

### Resultados

Decidimos probar dos técnicas, la de K-vecinos y la regresión logística, en ambas utilizamos la librería scikit-learn que ya tiene integrados los algoritmos. Con estas técnicas intentamos predecir la severidad a partir de diferentes variables predictoras.

La de k vecinos nos sirvió más al utilizar todas las variables de la base de datos, utilizamos un vecino, lo que nos dio un 75% de acierto, por otro lado, la regresión logística nos sirvió más al momento de predecir severidad a partir de humedad, tiempo y temperatura, obtuvimos un 67% de acierto

### Conclusiones e investigación a futuro

En conclusión, la frecuencia no tiene relación a la severidad de un accidente (figuras 1.1 y 1.2), además concluimos que el estado de Dakota del Sur (figura 3.2) y el clima de lluvia congelada (figura 2.2) son los que presentan mayor severidad, además llegamos a la conclusión de que la humedad, temperatura y duración del accidente si afectan a la severidad. Como investigación a futuro nos quedaría revisar otras variables y revisar cuáles presentan mayor correlación con la severidad. Ya teniendo esas variables, podríamos volver a utilizar el algoritmo de predicción y con base en el conocimiento de esas variables en la vida real, poder predecir qué tan severo será un accidente. Adjuntamos código QR con la libreta de jupyter que muestra todo nuestro trabajo.



\*\*Las figuras 1.1 y 1.2 nos muestran que los datos con mayor severidad promedio se encuentran dentro de una frecuencia baja de accidentes, y viceversa, aquellos que cuentan con una frecuencia alta de accidentes se encuentran con una severidad media baja.