**Logotipo, nombre de la empresa

Descripción generada automáticamente**

**Actividad Evaluable:**

**Mapas de calor y boxplots**

**Herramientas computacionales: el arte de la analítica**

Javier Piña Camacho

A01701478

12 de enero de 2022

**Una introducción al conjunto de datos ¿Qué es? ¿De dónde se obtuvo? ¿Qué representa?**

Los datos que se utilizaron en esta actividad fueron obtenidos de choques de automóviles en Estados Unidos. Cuentan con una descripción del evento y datos relevantes sobre el ambiente, duración, fecha y lugar del suceso. Cada registro representa un accidente y recopila las variables que lo acompañaron.

Base de datos de: <https://www.kaggle.com/sobhanmoosavi/us-accidents>

**Cantidad de datos, las variables que contiene cada vector de datos y el tipo de variables.**

La base de datos cuenta con 499 registros, los vectores contienen 47 variables de tipo entero, flotante, objeto y booleano.

1. ID object
2. Severity int64
3. Start\_Time object
4. End\_Time object
5. Start\_Lat float64
6. Start\_Lng float64
7. End\_Lat float64
8. End\_Lng float64
9. Distance(mi) float64
10. Description object
11. Number float64
12. Street object
13. Side object
14. City object
15. County object
16. State object
17. Zipcode object
18. Country object
19. Timezone object
20. Airport\_Code object
21. Weather\_Timestamp object
22. Temperature(F) float64
23. Wind\_Chill(F) float64
24. Humidity(%) float64
25. Pressure(in) float64
26. Visibility(mi) float64
27. Wind\_Direction object
28. Wind\_Speed(mph) float64
29. Precipitation(in) float64
30. Weather\_Condition object
31. Amenity bool
32. Bump bool
33. Crossing bool
34. Give\_Way bool
35. Junction bool
36. No\_Exit bool
37. Railway bool
38. Roundabout bool
39. Station bool
40. Stop bool
41. Traffic\_Calming bool
42. Traffic\_Signal bool
43. Turning\_Loop bool
44. Sunrise\_Sunset object
45. Civil\_Twilight object
46. Nautical\_Twilight object
47. Astronomical\_Twilight object

**Los rangos de las variables escogidas para el análisis**

Humidity(%), con un rango de valores de 26 a 100

Visibility(mi), con un rango de valores de 0.5 a 20.0

**Basado en la media, mediana y desviación estándar de cada variable:**

Los valores de la humedad varían mucho de acuerdo a su desviación estándar de 17.42. Se puede notar un porcentaje de humedad alto en los accidentes de autos si vemos la media de 75.65%. La mediana de 80% apoya lo anterior, pues al menos la mitad de los accidentes contó con un alto porcentaje de humedad.

En cuanto a la visibilidad, la dispersión es menor con un valor de 3.597. Es por ello que se repiten más valores que en la columna de humedad. Tomando en cuenta que la media es de 7.029 y la mediana de 10, es decir se encuentran cercanos a la zona media media del rango (0.5 a 20), los accidentes no parecen ser muy influenciados por la visibilidad pero tienden ligeramente a suceder cuando la visibilidad disminuye.

Existe una pequeña relación inversa entre estos dos valores. Puesto que la humedad tiende a ser alta y la visibilidad se inclina ligeramente a ser menor en los accidentes de coche.

**Representación gráfica:**

Gráfico, Histograma

Descripción generada automáticamente

Figura 1. Histograma de la columna de visibilidad.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Figura 2. Diagrama de cajas y bigotes de la columna de visibilidad.

Gráfico, Histograma

Descripción generada automáticamente

Figura 3. Histograma de la columna de humedad.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Figura 4. Diagrama de cajas y bigotes de la columna de humedad.

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

Figura 5. Mapa de calor de las correlaciones entre todas la variables númericas de la base de datos.

**En base a las gráficas:**

Se pueden hacer mejores observaciones de ambas variables con ayuda de las gráficas presentadas en cada figura. El mapa de calor de la figura 5 resulta ser lo más significativo, dandonos confirmación de una leve correlación inversa entre ambas variables. No obstante, podemos comprobar con los histogramas de las figuras 1 y 3, que aunque la visibilidad vaya disminuyendo gradualmente mientras que la humedad aumenta, la visibilidad tiene un incremento que representa la mayor parte de los datos cuando se llega a la visibilidad media. De acuerdo a los diagramas de cajas y bigotes (figura 2 y 4), podemos ver que los datos son significativos puesto que no se encuentran outliers. Además, muestran visualmente la correlación inversa al tener la humedad su rango intercuartil en la parte baja y el rango de la visibilidad estar en la parte alta.