

Herramientas computacionales: el arte de la analítica (Gpo 570)



**Tecnológico
de Monterrey**

Actividad Evaluable: Mapas de calor y boxplots

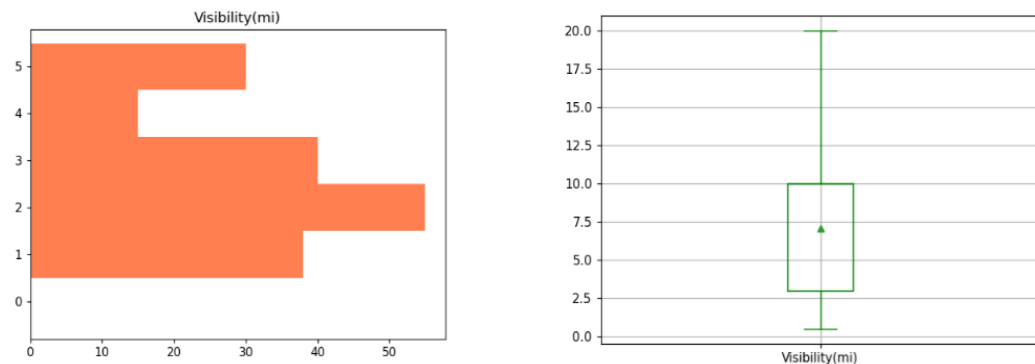
Diego Andrés Moreno Molina

A01283790

Introducción

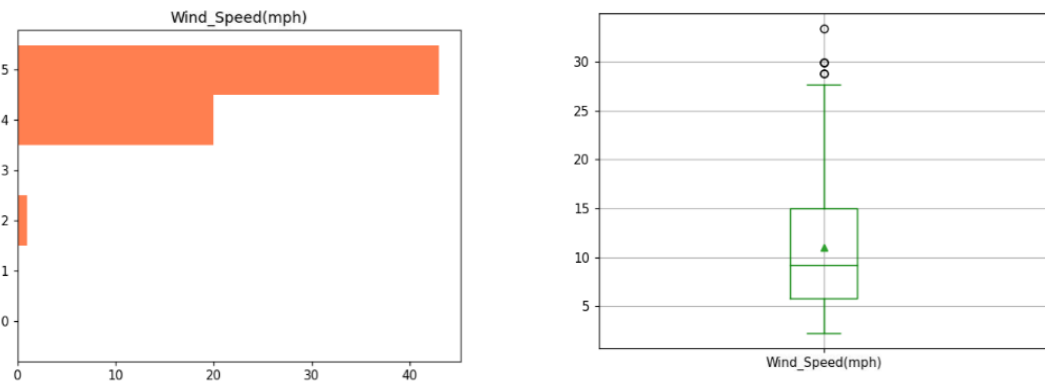
Para esta etapa se eligieron 2 variables que potencialmente tienen relación directa con la posibilidad de choques, por lo que nuestra hipótesis indica que estos valores deberían ser en su mayoría altos y con una fuerte correlación a la gravedad y cantidad de registros. Se analizarán individualmente las siguientes dos variables; Visibilidad y velocidad del viento.

- Visibilidad

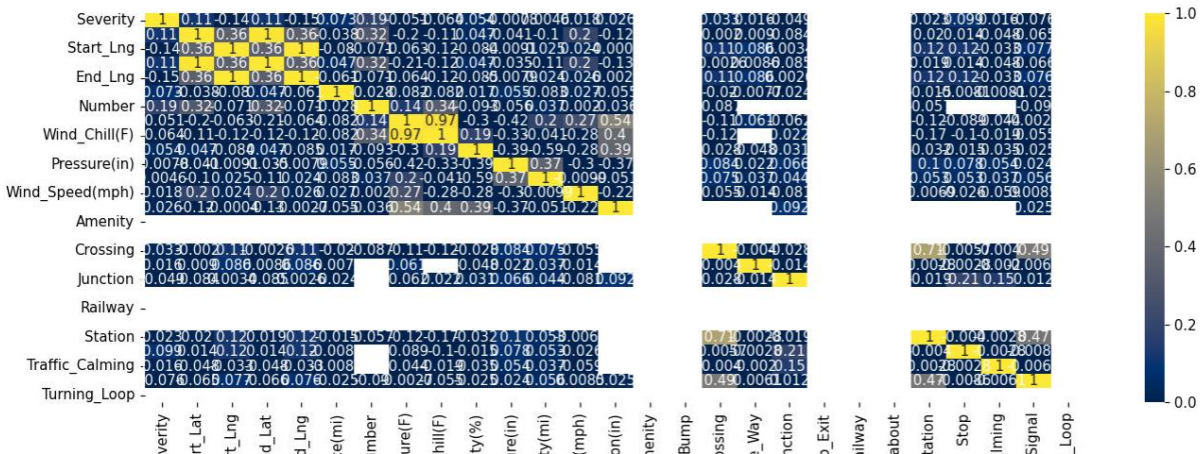


Los resultados de estos diagramas en efecto confirman nuestra hipótesis. Podemos observar que entre menos sea la distancia de visibilidad, mayor la cantidad de registros de choques que se encuentran. De igual manera esto se observa en el diagrama de bigotes, donde la mayoría de los registros están en los primeros cuartiles donde la visibilidad es considerablemente menor. Las dos columnas con mayor visibilidad son los que menor cantidad de registros tienen, por lo que se confirma que a menor visibilidad mayor probabilidad de choque.

- Velocidad del viento



Ahora bien, los resultados de la velocidad del viento también van acorde a nuestra hipótesis, donde podemos ver que entre mayor la velocidad del viento, mayor la cantidad de choques. Sin embargo, se debe a tomar a consideración que esta métrica es un poco engañosa, ya que en el mundo real es casi imposible que exista el viento a velocidades menores a 2 mph. Aun así encontramos un salto considerable de registros cuando la velocidad del viento incrementa de 4mph a 5 mph, por lo que se confirma que en efecto hay una relación directa entre la velocidad del viento y probabilidad de choque.



Por último, analizando el mapa de calor para encontrar correlaciones entre variables podemos darnos cuenta de que realmente no hay muchas variables con una correlación fuerte. Aunque esto si es sorprendente al mismo tiempo tiene sentido, ya que este conjunto de datos se conforma de variables completamente diferentes que modelan un choque. No nos interesa que haya correlación entre estas variables. Los diagramas anteriormente vistos nos son de mayor utilidad, ya que para nuestro análisis el verdadero valor se encuentra en la cantidad de registros que contiene una cierta categoría de datos.