Fundamentos de Programación Grafica

Universidad de la salle bajio | itsn 201

Reporte del Análisis

Fernando Krauss fukunaga luis javier origel gonzalez

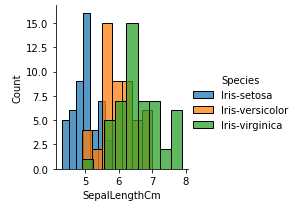
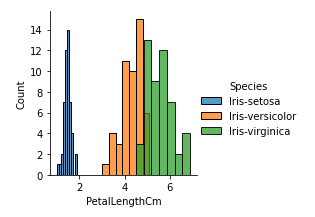
Introducción

Las gráficas son formas representativas de datos, no de información, la información se da al leer estas graficas e interpretarlas, por lo tanto, la gráfica es la intermediaria entre los datos y la información, por lo tanto ocupamos que las gráficas tengan los datos correctos para evitar malinterpretaciones, ya que, en caso de no hacer bien una gráfica, podemos caer en ambigüedades y cambiando completamente la idea, por eso es que el buen manejo de los recursos visuales es vital para que las personas no profesionales puedan entender los datos.

Así que nosotros analizamos diferentes archivos csv, los cuales interpretaremos con gráficas para que así determinamos que datos son claves para entender y convertir los datos en información

Análisis de la Base de Datos de las Flores Iris

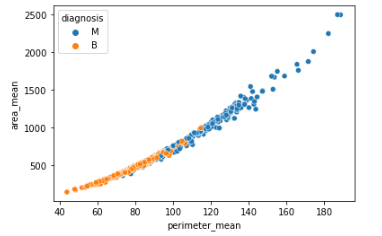
En este archivo csv se encontraban 3 tipos de iris, estas tienen datos bastante similares, esto debido a que son flores de la misma especie, al ver los datos nos dimos cuenta que la iris-setosa es muy diferente a las otras dos, la iris-versicolor y la iris-virginica, ya que al estudiar los 4 parámetros que tenemos, estos son el largo y ancho del sépalo y del petalo, en el sépalo, la iris versicolor y la iris virginica son muy similares en cambio la setosa tiene medidas completamente distintas, en el caso del petalo si están separas la versicolor y virginica, y la setosa tiene mediciones completamente distintas.



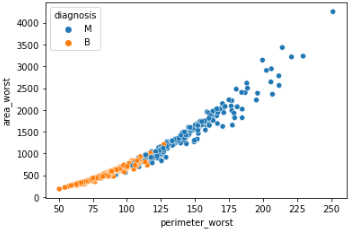
En este caso tenemos la altura del petalo y el sépalo, en el caso del sépalo este es algo parecido entre cada una de las flores, pero en el petalo las cosas cambian, su altura es muy diferente entre cada una, la versicolor y virginica tienen muy poco parecido entre ellas.

Análisis de la Base de Datos de Cáncer de Mama

Este archivo tiene 31 valores distintos, esto hace que sea una base muy completa, un archivo el cual nos puede tomar meses y meses de estudio, pero al verlo, nos podemos dar cuenta que hay bastantes valores los cuales tienden a ser similares, tenemos el caso del promedio de área y la peor área, estos son valores los cuales al verlos en la gráfica tienen valores muy similares, pero aun así podemos decir que son necesarios estos datos, ya que al ver estas graficas lo que uno tiene que preguntarse es, ¿Para qué sirve esta grafica?, en este caso el propósito es determinar cuándo un cáncer de mama es maligno o benigno, y al ver el peor área, notamos que es más fácil ver cuando es bueno o malo, pero por lo general, buscar los datos en un histograma es complicado, pero al ver la gráfica de dispersión del área y perímetro promedio, uno se puede dar cuenta de que es una muy buena forma de ver los datos.



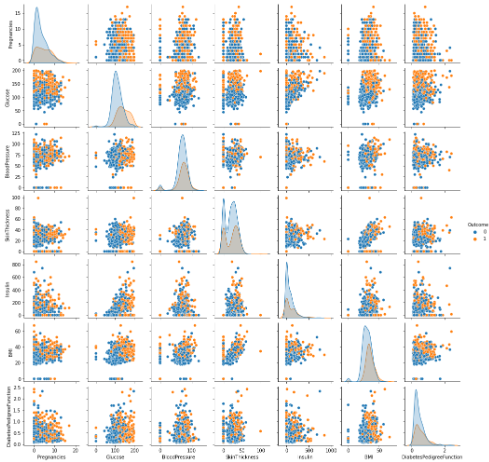
Ahora viendo esto no con el promedio si no con el peor de área y perímetro se ve muy parecido



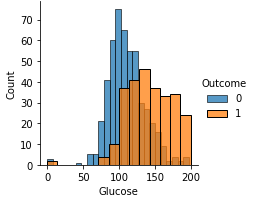
Al ser las peores áreas y perímetros, las distancias son mucho mayores que del promedio, y a parte si hay una diferencia, la cual es que los datos no se mezclan tanto entre ellos

Análisis de la Base de Datos de Diabetes

La diabetes es una enfermedad la cual ya está muy bien estudiada, ya hay muchas formas de determinar si una persona tiene diabetes o no, pero al analizar estos datos uno puede darse cuenta cuales son los datos más importantes para determinar la diabetes, uno de estos y muy obviamente, es el caso de la glucosa

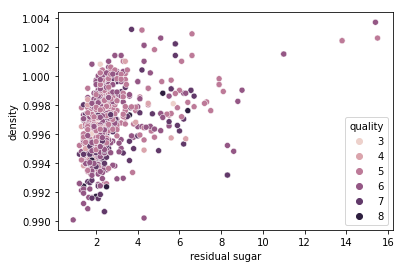


En todas estas graficas de dispersión los casos de la glucosa son los que más separados están uno de los otros y en base a estos datos junto con todos los demás podemos determinar si es o no diabetes, esta me parece que es la base de datos más complicada dado que los datos tienden a ser parecidos por separado, pero en conjunto ya se empiezan a ver cambios.

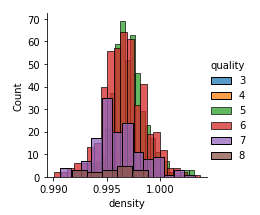


Análisis de la Base de Datos de Vino

Como ultima base tenemos la calidad del vino, esta se mide del 3 siendo la peor que se tiene, al 8, esto se define en base a 13 columnas, se toman valores como densidad, pH y alcohol, pero hay valores que no pensé que fueran relevantes, como por ejemplo el ácido cítrico o los sulfatos, y veamos el caso del residuo de azúcar y la densidad



Aquí los datos parecen que son muy parecidos pero el estudio de todas las demás variables nos puede dar resultados que determinaran la calidad del vino



En el caso de la densidad los datos están muy empalmados entre ellos, por lo cual esto termina haciendo una tarea complicada analizar un buen vino con gráficas, un buen vino es más fácil distinguirlo con el sabor u olor, no con valores visuales.

Conclusión

Al crear estas graficas llegue a la conclusión que hay cosas las cuales no pueden ser medidas con gráficas, o más bien hay métodos más fáciles que graficar para entender la información que se nos presenta, dependiendo del tema, tengamos por ejemplo el cáncer, este es más fácil de determinar con los valores que se nos dieron, que con el vino, que es algo que se entiende mejor con el olor y sabor, también mientras más datos tengamos, es más probable que se tengan diferentes opciones para entender esta información.

Referencias