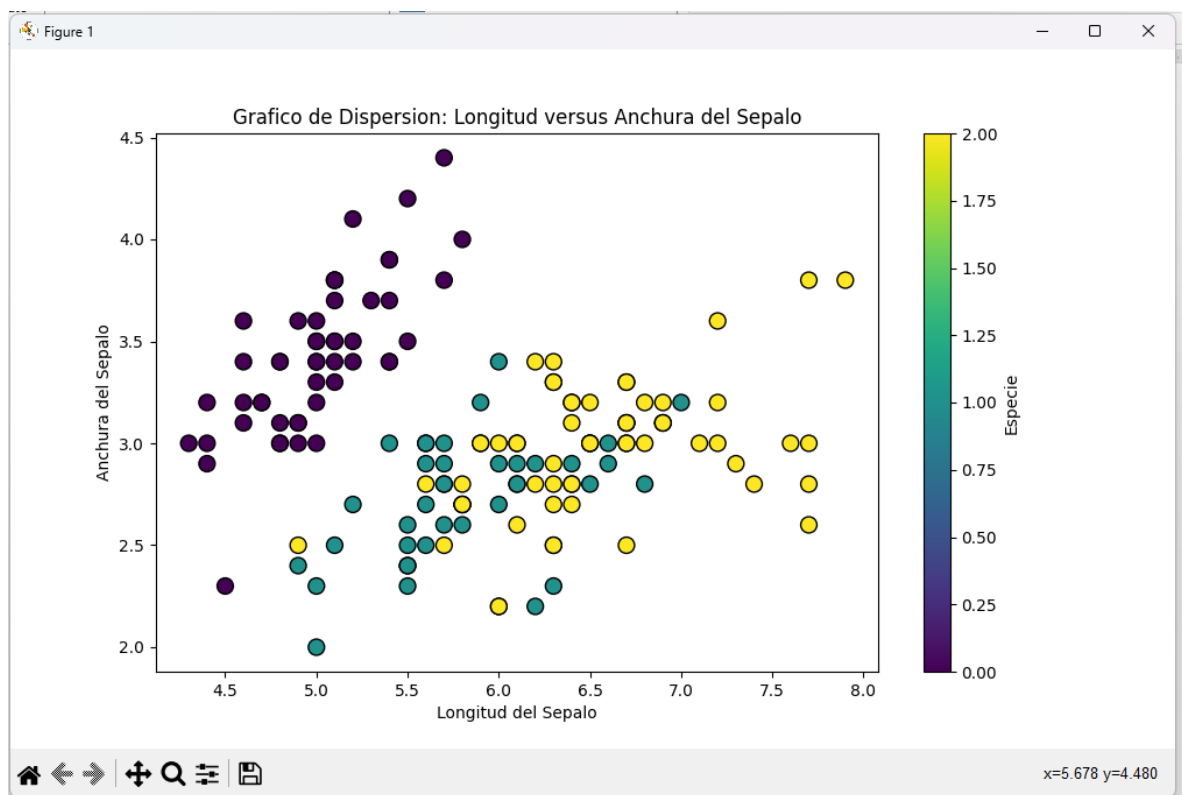


## Informe sobre la Visualización y Análisis del Dataset Iris según los graficos.

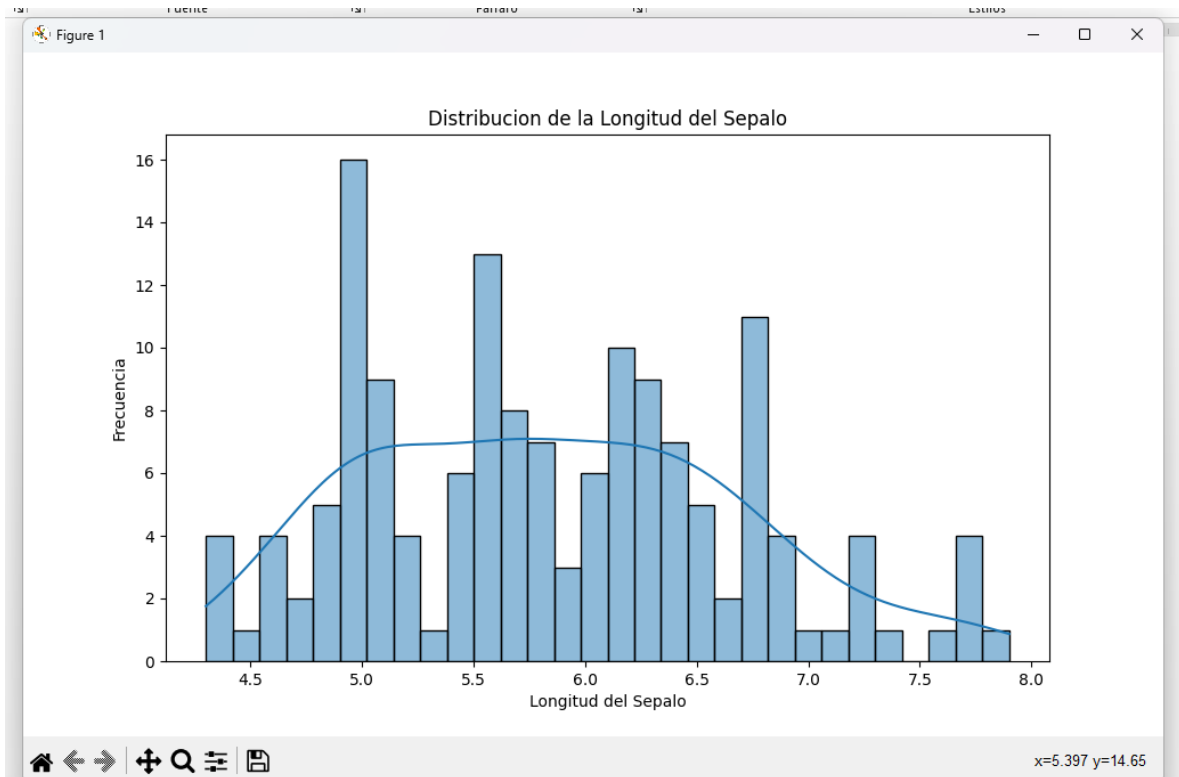
El dataset Iris es uno de los más populares en el ámbito de la ciencia de datos y el aprendizaje automático. Fue introducido por el estadístico y biólogo Ronald A. Fisher en 1936 y contiene mediciones de cuatro características de flores de tres especies de Iris: Iris setosa, Iris versicolor, y Iris virginica. Las características registradas incluyen la longitud y anchura del sépalo y la longitud y anchura del pétalo. Este dataset es frecuentemente usado para ilustrar técnicas de clasificación y visualización debido a su estructura y simplicidad.

### 1. Gráfico de Dispersión (Longitud vs. Anchura del Sépalo).



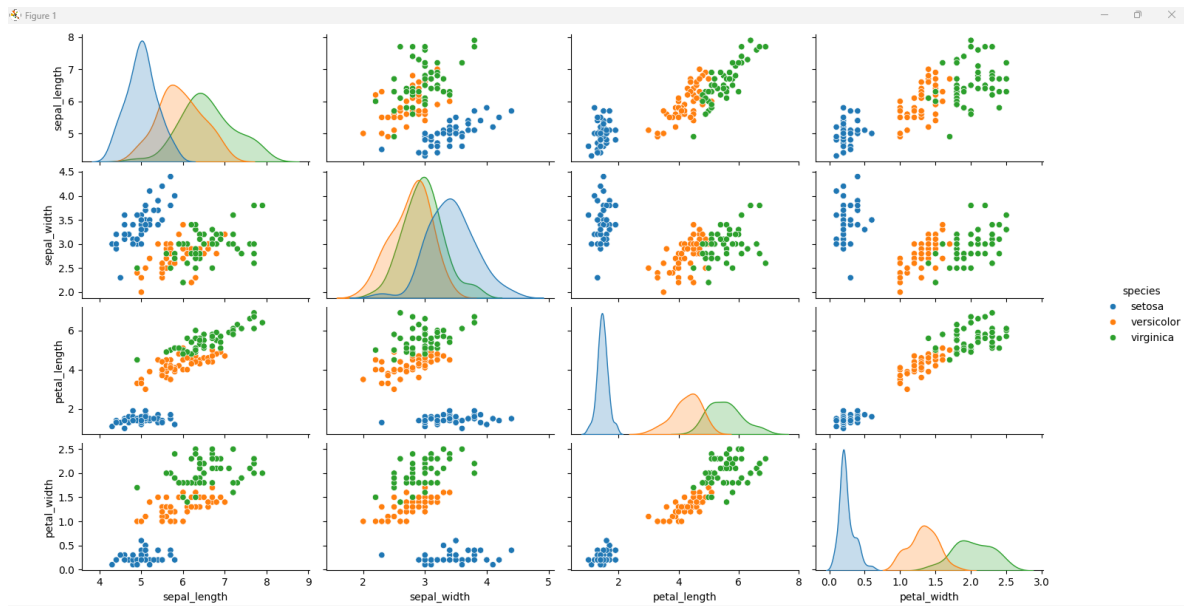
La primera imagen muestra un gráfico de dispersión que compara la longitud del sépalo con su anchura, donde los puntos están coloreados según la especie. Este tipo de gráfico es útil para identificar patrones de agrupación entre las especies, así como para observar relaciones entre las variables. En este caso, el gráfico revela que las flores de la especie Iris setosa tienden a tener sépalos más anchos y relativamente cortos en comparación con las otras dos especies, lo que sugiere que esta característica podría ayudar a diferenciar entre Iris setosa y las otras especies. Sin embargo, las otras dos especies, Iris versicolor y Iris virginica, presentan una superposición considerable en las dimensiones del sépalo, lo que indica que solo estas dos características no son suficientes para distinguirlas claramente.

## 2. Histograma de la Longitud del Sépalo.



La segunda imagen es un histograma de la longitud del sépalo, que proporciona una visión sobre la distribución de esta variable en el dataset. Los histogramas son herramientas fundamentales para entender la frecuencia de valores específicos dentro de una variable y permiten identificar posibles sesgos o tendencias. En este caso, observamos una distribución multimodal, donde se pueden identificar picos en ciertas longitudes. Este tipo de análisis permite inferir que cada especie podría tener un rango característico para la longitud del sépalo, lo que facilita la identificación preliminar de las especies en función de este atributo. Además, el histograma muestra una curva de densidad superpuesta, que ayuda a entender la distribución de la longitud del sépalo de manera continua, resaltando las áreas de mayor concentración de datos.

### 3. Matriz de Gráficos de Dispersión.



La tercera imagen es una matriz de gráficos de dispersión para todas las combinaciones de las cuatro características. En cada gráfico de dispersión de esta matriz, los datos están coloreados según la especie, lo que permite identificar relaciones entre cada par de características y observar cómo se agrupan los datos. Esta matriz es útil porque revela que algunas combinaciones de características, como la longitud y anchura del pétalo, son efectivas para separar las tres especies. Específicamente, Iris setosa se diferencia claramente de las otras dos especies en casi todas las combinaciones de características, mientras que Iris versicolor y Iris virginica muestran cierta superposición, especialmente en las características de sépalos. La inclusión de las curvas de densidad en la diagonal de la matriz proporciona información adicional sobre la distribución de cada variable.

### Conclusión.

Las visualizaciones generadas a partir del dataset Iris permiten obtener insights clave sobre la relación entre las características y la posibilidad de clasificar las especies de manera precisa. Observando los gráficos, es evidente que las características de los pétalos son más efectivas para diferenciar entre las especies en comparación con las características de los sépalos. Estas visualizaciones no solo facilitan la comprensión de las propiedades de cada especie, sino que también sirven como punto de partida para construir modelos de clasificación más precisos. En aplicaciones prácticas, este análisis visual es crucial, ya que permite identificar patrones y posibles variables relevantes antes de aplicar algoritmos de aprendizaje automático.