

# ANÁLISIS DE DATOS CON STREAMLIT

Esta aplicación es una herramienta interactiva para la exploración y análisis de datos.

## Análisis de Datos con Streamlit

### Lectura de Datos

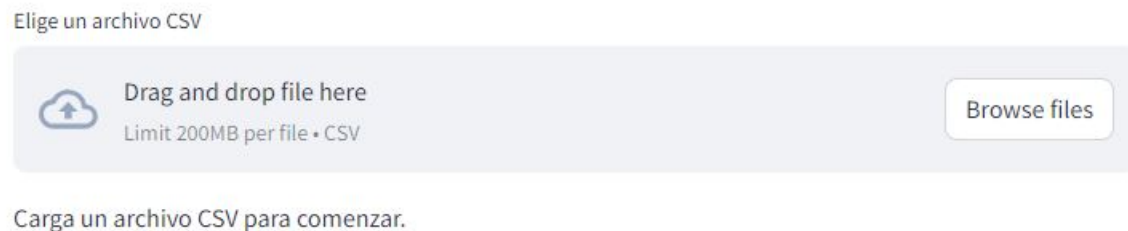


Figure 1:

Sus funcionalidades:

## 1. Lectura de Datos

Esta funcionalidad permite al usuario cargar un archivo CSV en la aplicación. La aplicación admite diferentes codificaciones de archivos para asegurar que los datos se carguen correctamente.

Propósito:

Permitir a los usuarios cargar sus propios conjuntos de datos. Manejar diferentes codificaciones y delimitadores para asegurar la compatibilidad con distintos formatos de archivos CSV.

## 2. Resumen de Datos

Una vez cargados los datos, la aplicación genera un resumen estadístico descriptivo de los datos, mostrando medidas como la media, la mediana, el mínimo, el máximo, los cuartiles, etc.

Propósito:

Proporcionar una visión general rápida de la distribución y las características principales del conjunto de datos. Ayudar a los usuarios a identificar posibles valores atípicos, tendencias generales y la variabilidad de los datos.

### 3. Visualizacion de Datos

La aplicación permite a los usuarios visualizar la distribución de los datos a través de gráficos. En este caso, se utiliza un histograma con una estimación de densidad de núcleo (KDE) superpuesta.

Propósito:

Facilitar la comprensión de la distribución de una variable específica. Ayudar a los usuarios a identificar patrones, sesgos, y distribuciones multimodales. Proveer una representación visual que puede ser más intuitiva que los resúmenes numéricos.

### 4. Interactividad

La aplicación es interactiva, lo que significa que los usuarios pueden seleccionar la columna de datos que desean visualizar y ajustar parámetros como el número de bins para el histograma.

Propósito:

Ofrecer flexibilidad al usuario para explorar diferentes aspectos del conjunto de datos. Permitir un análisis más detallado y ajustado a las necesidades específicas del usuario.

## Objetivo de la aplicacion

El objetivo principal de esta aplicación es proporcionar una herramienta simple y eficaz para la exploración y el análisis preliminar de datos. Esta herramienta es útil para:

Análisis Exploratorio de Datos (EDA): Antes de realizar análisis más complejos, es crucial entender las características básicas de los datos. La aplicación ayuda a identificar patrones, tendencias y anomalías.

Educación: Es una excelente herramienta educativa para enseñar conceptos básicos de análisis de datos y visualización.

Toma de Decisiones: Ayuda a los analistas y científicos de datos a tomar decisiones informadas sobre el tratamiento y análisis de los datos basándose en las distribuciones y resúmenes estadísticos.

## Ejemplo de uso practico

Supongamos que eres un profesor que desea analizar los datos de los estudiantes, incluyendo horas de estudio y calificaciones. Puedes usar esta aplicación para:

Cargar los datos de los estudiantes desde un archivo CSV.

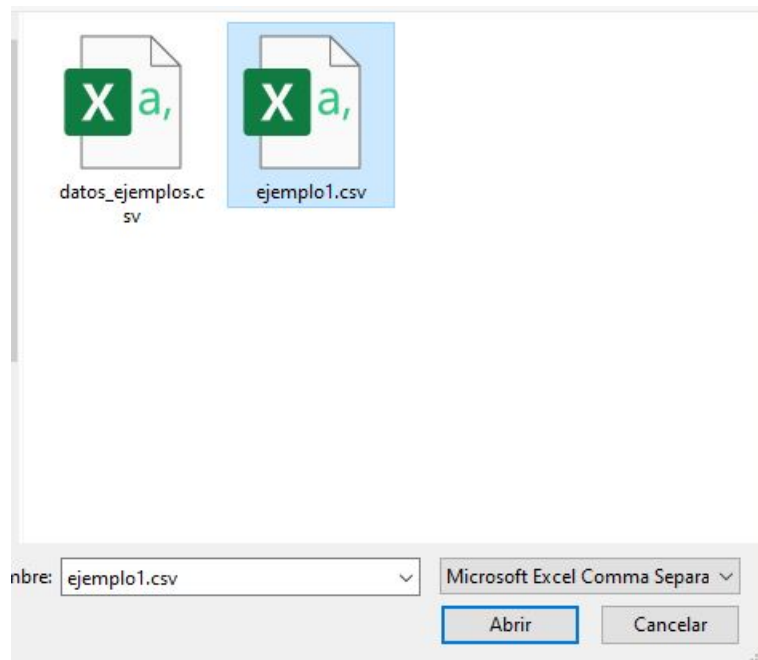


Figure 2:

Revisar un resumen estadístico para entender las características generales, como la media de las calificaciones y las horas de estudio. Visualizar la distribución de las calificaciones para ver si siguen una distribución normal, están sesgadas o tienen múltiples picos. Ajustar y explorar diferentes aspectos de los datos de manera interactiva.

 ejemplo1.csv 121.1KB ✕

Ingresa el delimitador del archivo CSV

;

Datos cargados con éxito

	Horas_Estudio	Calificaciones
0	4.4159	4.0858
1	1.5997	4.1755
2	5.7595	6.2559
3	3.1882	4.6952
4	3.1354	3.7308

Figure 3:

Esta aplicación no solo facilita el análisis inicial de los datos, sino que también ofrece una base sólida para realizar análisis más avanzados o preparar datos para modelos predictivos y otras técnicas de análisis más complejas.

## Resumen de Datos

	Horas_Estudio	Calificaciones
count	4,999	4,999
mean	4.9621	5.9904
std	2.8974	2.5105
min	0.001	-0.3302
25%	2.4184	4.0353
50%	4.9673	5.9485
75%	7.4968	7.9748
max	9.9994	13.5879

Figure 4:

## Aqui el codigo en Python:

```
import streamlit as st
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

st.title("Análisis de Datos con Streamlit")

st.header("Lectura de Datos")
uploaded_file = st.file_uploader("Elige un archivo CSV", type="csv")
if uploaded_file is not None:
    delimiter = st.text_input("Ingresa el delimitador del archivo CSV", value=";")
    try:
        data = pd.read_csv(uploaded_file, delimiter=delimiter, encoding='utf-8')
    except UnicodeDecodeError:
        try:
            data = pd.read_csv(uploaded_file, delimiter=delimiter, encoding='latin1')
        except UnicodeDecodeError:
            try:
                data = pd.read_csv(uploaded_file, delimiter=delimiter,
                                   encoding='iso-8859-1')
            except UnicodeDecodeError:
                st.error("No se pudo leer el archivo CSV. Por favor,
                        verifica la codificación del archivo.")
                st.stop()
    st.write("Datos cargados con éxito")
    st.write(data.head())

# Resumen de Datos
st.header("Resumen de Datos")
st.write(data.describe())

# Visualización de Datos
st.header("Visualización de Datos")

# Grafico de distribucion
st.subheader("Grafico de Distribucion")
col = st.selectbox("Selecciona la variable para
el grafico de distribucion", options=data.columns)
num_bins = st.slider("Numero de bins", min_value=10, max_value=100, value=30)
if st.button("Generar Grafico de Distribucion"):
    fig, ax = plt.subplots()
    sns.histplot(data[col], bins=num_bins, kde=True, ax=ax)
    st.pyplot(fig)
else:
    st.write("Carga un archivo CSV para comenzar.")
```

## Resultado Grafico del Código

### Gráfico de Distribución

Selecciona la variable para el gráfico de distribución

Calificaciones



Número de bins



Generar Gráfico de Distribución

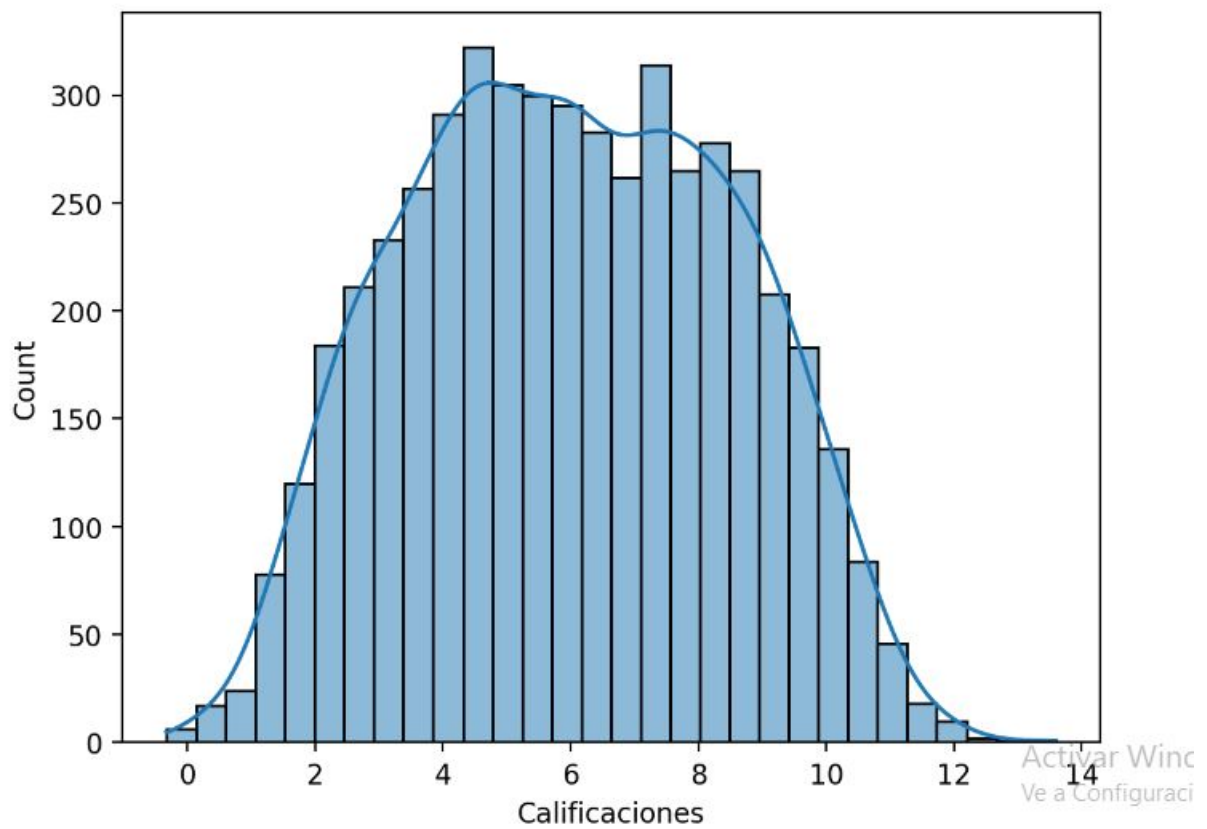


Figure 5:

## Interpretacion del Grafico

El gráfico generado es un histograma con una curva de densidad (KDE) superpuesta, y representa la distribución de las calificaciones de los estudiantes.

### Componentes del Grafico

Eje X (Calificaciones):

Representa las calificaciones de los estudiantes, que varían de 0 a 14.

Eje Y (Count):

Muestra el conteo (frecuencia) de estudiantes que obtuvieron una determinada calificación. Barras del Histograma:

Cada barra representa el número de estudiantes cuyas calificaciones caen dentro de un rango específico. El rango está determinado por el número de bins (en este caso, 30).

### Curva de KDE (Kernel Density Estimation)

La curva azul suave superpuesta al histograma proporciona una estimación de la densidad de probabilidad de las calificaciones. Esto ayuda a visualizar la distribución subyacente de los datos de una manera más continua.

### Interpretación:

Forma de la Distribución:

La distribución parece ser unimodal, con un único pico principal alrededor de las calificaciones de 6 a 8. La distribución muestra cierta simetría, aunque hay una ligera cola a la derecha (sesgo positivo), indicando que hay algunos estudiantes con calificaciones excepcionalmente altas. Moda:

La moda, o el valor más frecuente de las calificaciones, se encuentra entre 6 y 8, ya que las barras más altas están en este rango.

### Dispersion:

Las calificaciones están bastante dispersas, cubriendo un rango de 0 a aproximadamente 12. La mayoría de las calificaciones están concentradas entre 4 y 10. Frecuencia:

La altura de las barras indica que hay muchos estudiantes con calificaciones alrededor de 6 a 8. Las barras son más bajas en los extremos, lo que sugiere que hay menos estudiantes con calificaciones muy bajas (0-2) o muy altas (12-14).

### Análisis Adicional:

Rendimiento Académico:

La mayoría de los estudiantes tienen calificaciones medias, lo que podría indicar un rendimiento académico promedio en general. Pocos estudiantes tienen calificaciones muy bajas o muy altas, lo que puede ser interesante para un análisis más profundo sobre las causas de estos extremos.

Intervención Educativa:

Si el objetivo es mejorar el rendimiento académico, se podría enfocar en los estudiantes en los extremos de la distribución (calificaciones muy bajas) para proporcionarles apoyo adicional.

**Conclusión:**

El gráfico de distribución de calificaciones te proporciona una visión clara de cómo están distribuidas las calificaciones entre los estudiantes. Esta información es útil para identificar patrones generales en el rendimiento académico y puede ayudar a tomar decisiones informadas sobre intervenciones educativas y estrategias de enseñanza.