

## 1. Lenguajes de programación

### Código:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
import numpy as np

# Cargar los datos desde un archivo CSV
archivo_csv = 'Lenguajes_porcentaje.csv'
df = pd.read_csv(archivo_csv)

# Validar columnas
print("Columnas del archivo:", df.columns)

# Renombrar columna si es necesario
if "Porcentaje" in df.columns:
    df.rename(columns={"Porcentaje": "Porcentaje de Uso"}, inplace=True)

# Convertir la columna "Porcentaje de Uso" a flotantes
try:
    df["Porcentaje de Uso"] = df["Porcentaje de Uso"].str.strip().str.rstrip('%')
    df.dropna(subset=["Porcentaje de Uso"], inplace=True) # Elimina filas vacías
    df["Porcentaje de Uso"] = df["Porcentaje de Uso"].astype(float)
except Exception as e:
    print("Error al procesar la columna 'Porcentaje de Uso':", e)
    print(df.head())
    raise

# Crear un índice ficticio como variable independiente
df = df.sort_values(by="Porcentaje de Uso", ascending=False).reset_index(drop=True)
df["Ranking"] = np.arange(1, len(df) + 1)

# Separar las variables independiente (Ranking) y dependiente (Uso)
X = df[["Ranking"]].values # Variable independiente
y = df["Porcentaje de Uso"].values # Variable dependiente

# Crear y ajustar el modelo de regresión lineal
model = LinearRegression()
model.fit(X, y)

# Generar predicciones
y_pred = model.predict(X)

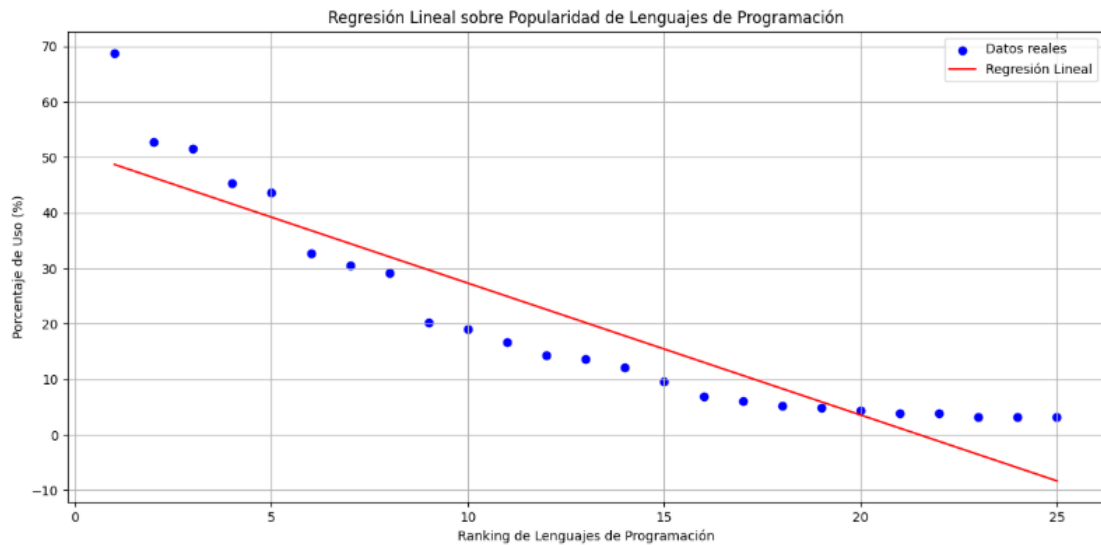
# Mostrar los coeficientes del modelo
print("Pendiente (coeficiente):", model.coef_[0])
print("Intercepto:", model.intercept_)

# Graficar los datos reales y la regresión lineal
plt.figure(figsize=(12, 6))
plt.scatter(df["Ranking"], df["Porcentaje de Uso"], color="blue", label="Datos reales")
plt.plot(df["Ranking"], y_pred, color="red", label="Regresión Lineal")

# Etiquetas y título
plt.xlabel("Ranking de Lenguajes de Programación")
plt.ylabel("Porcentaje de Uso (%)")
plt.title("Regresión Lineal sobre Popularidad de Lenguajes de Programación")
plt.legend()
plt.grid(True)
plt.tight_layout()

# Mostrar el gráfico
plt.show()
```

## Grafica:



## Hipótesis:

El porcentaje de uso de los lenguajes de programación disminuye a medida que aumenta su ranking en popularidad. Los lenguajes con un ranking más bajo (más populares) tienden a concentrar una mayor cuota de uso, mientras que los lenguajes menos populares (ranking más alto) tienen porcentajes de uso considerablemente menores. Esta relación se puede modelar como una tendencia lineal decreciente, aunque puede haber ciertas variaciones debido a factores externos como el auge de lenguajes en sectores específicos o el uso especializado de lenguajes menos conocidos.