

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_programacion.csv' # Asegúrate de tener el archivo 'datos.csv' en el mismo directorio
df = pd.read_csv(archivo_csv)

# Convertir la columna "Porcentaje de Uso" a flotantes (removiendo el símbolo '%')
df["Porcentaje de Uso"] = df["Porcentaje de Uso"].str.rstrip('%').astype(float)

# 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)

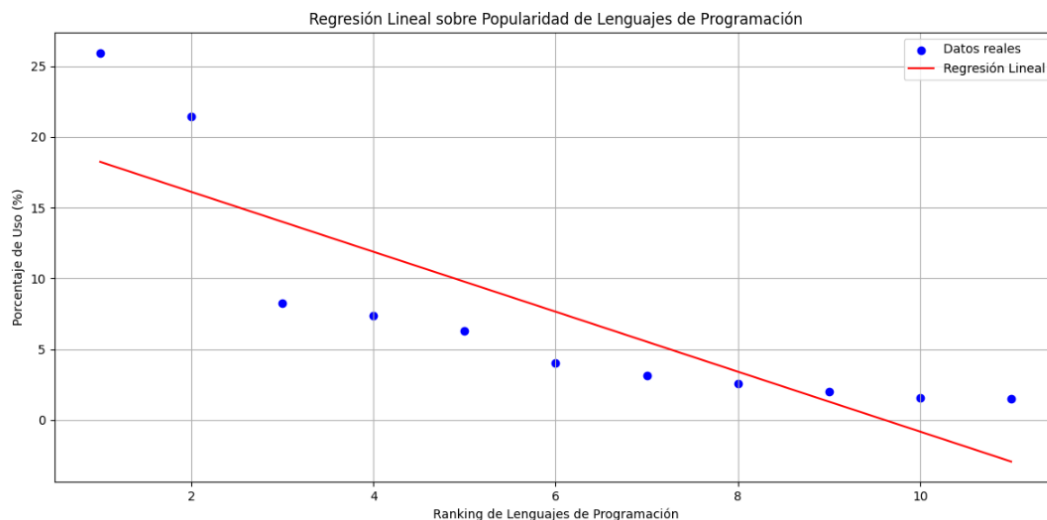
# Mostrar los coeficientes del modelo
print("Pendiente (coeficiente):", model.coef_[0])
print("Intercepto:", model.intercept_)

# 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:

Existe una relación inversa entre el ranking de los lenguajes de programación y su porcentaje de uso, lo que significa que los lenguajes más populares (ranking más bajo) tienen un porcentaje de uso más alto, mientras que los lenguajes menos populares (ranking más alto) tienden a tener un porcentaje de uso menor. La tendencia general puede representarse mediante una línea de regresión lineal, aunque es probable que existan desviaciones debido a factores externos como el dominio de aplicaciones específicas o la creciente popularidad de ciertos lenguajes en nichos particulares.