

## 1. Base de datos

### 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 = 'Base_de_Datos_Porcentajes.csv' # Nombre del archivo
df = pd.read_csv(archivo_csv)

# Convertir la columna "Porcentaje_Todos" a valores numéricos
df["Porcentaje_Todos"] = df["Porcentaje_Todos"].str.rstrip('%').astype(float)

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

# Separar las variables independiente (Ranking) y dependiente (Porcentaje_Todos)
x = df[["Ranking"]].values # Variable independiente
y = df["Porcentaje_Todos"].values # Variable dependiente

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

# Generar predicciones
y_pred = model.predict(x)

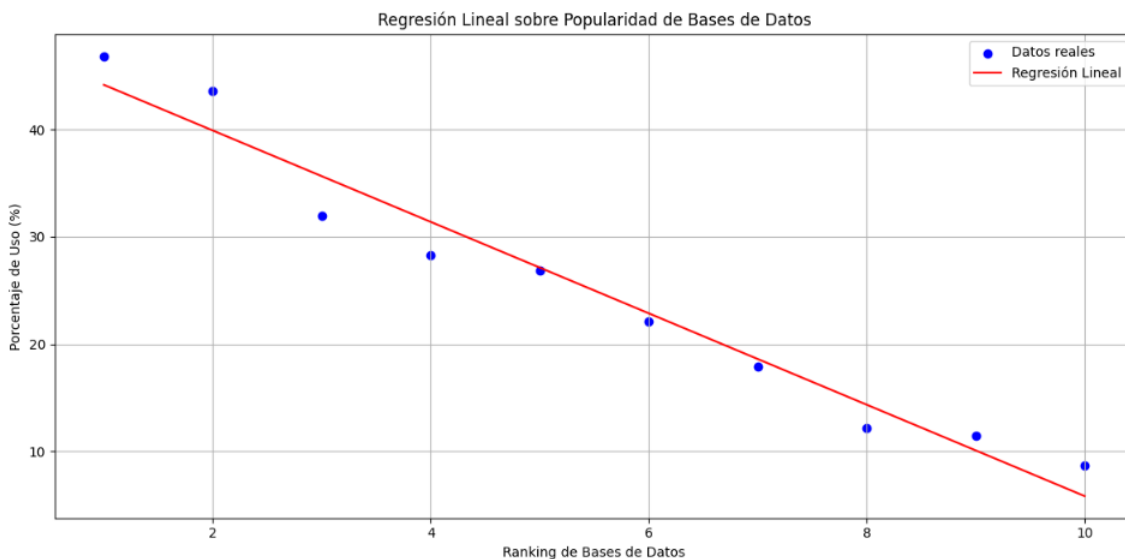
# Mostrar los coeficientes del modelo
print("Pendiente (coeficiente):", model.coef_[0])
print("Intercepto:", model.intercept_)

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

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

# Mostrar el gráfico
plt.show()
```

## Grafica:



## Hipótesis:

A medida que aumenta el ranking numérico de una base de datos (indicando menor popularidad), su porcentaje de uso tiende a disminuir. Las bases de datos más populares concentran un porcentaje significativo del mercado, mientras que las menos populares tienen una menor cuota de uso. Esta relación sigue una tendencia lineal decreciente, aunque pueden existir excepciones debido a factores como la especialización, las necesidades específicas de la industria o la adopción de tecnologías emergentes.