

Hipótesis

La popularidad de una base de datos está positivamente correlacionada con su uso: las bases de datos más populares tienen un mayor porcentaje de uso, mientras que las menos populares tienen un porcentaje de uso más bajo.

Justificación

El análisis de regresión lineal muestra que existe una relación negativa entre el ranking de las bases de datos (con un ranking mayor representando menos popularidad) y su porcentaje de uso. Esto implica que las bases de datos más populares, como PostgreSQL, MySQL y SQLite, ocupan las primeras posiciones en el ranking y tienen un uso significativamente mayor. Por otro lado, a medida que las bases de datos caen en el ranking (es decir, son menos populares), su uso tiende a disminuir. Esto sugiere que la popularidad de una base de datos está fuertemente correlacionada con su adopción en proyectos reales y en el mercado, reflejando una tendencia común en el ecosistema de bases de datos.

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

# Leer el archivo CSV
archivo_csv = "BD.csv" # Cambia esto a la ruta de tu archivo real
df = pd.read_csv(archivo_csv)

# Limpiar los datos de la columna "Usage" (remover '%')
df['Usage'] = df['Usage'].str.rstrip('%').astype(float)

# Ordenar por uso y resetear el índice
df = df.sort_values(by="Usage", ascending=False).reset_index(drop=True)
df["Ranking"] = np.arange(1, len(df) + 1)

# Crear variables independientes y dependientes para el modelo
X = df[["Ranking"]].values # Variable independiente (Ranking)
y = df[["Usage"]].values # Variable dependiente (Uso)

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

# Generar predicciones
y_pred = model.predict(X)

# Analizar el modelo y generar un análisis textual
pendiente = model.coef_[0]
intercepto = model.intercept_

# Imprimir análisis del modelo
print("\nAnálisis del Modelo de Regresión Lineal:")
print(f"- La pendiente del modelo (coeficiente) es: {pendiente:.2f}")
print(f"- El intercepto del modelo es: {intercepto:.2f}")
print("\nInterpretación del comportamiento:")
print("- Las bases de datos más populares, como PostgreSQL, MySQL y SQLite, tienen un mayor porcentaje de uso.")
print("- A medida que descendemos en el ranking de popularidad, el uso de las bases de datos disminuye.")
print("- El modelo de regresión muestra una relación negativa, lo que indica que a medida que el ranking de popularidad disminuye, el uso también disminuye.")
print("- Esto sugiere que la popularidad de una base de datos está correlacionada con su uso en el mercado.")
```

```
# Configurar el gráfico con los nombres de las bases de datos en el eje y
plt.figure(figsize=(12, 8))
plt.scatter(df["Usage"], df["Database"], color="blue", label="Datos reales")
plt.plot(y_pred, df["Database"], color="red", label="Regresión Lineal")

# Etiquetas y título
plt.xlabel("Uso (%)")
plt.ylabel("Base de Datos")
plt.title("Relación entre Bases de Datos y Uso")
plt.legend()
plt.grid(True)
plt.tight_layout()

# Mostrar el gráfico
plt.show()
```

Análisis del Modelo de Regresión Lineal:

- La pendiente del modelo (coeficiente) es: -0.98
- El intercepto del modelo es: 26.25

Interpretación del comportamiento:

- Las bases de datos más populares, como PostgreSQL, MySQL y SQLite, tienen un mayor porcentaje de uso.
- A medida que descendemos en el ranking de popularidad, el uso de las bases de datos disminuye.
- El modelo de regresión muestra una relación negativa, lo que indica que a medida que el ranking de popularidad disminuye, el uso también disminuye.
- Esto sugiere que la popularidad de una base de datos está correlacionada con su uso en el mercado.

