

1. Porcentajes de soluciones

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 = 'solucion_porcentaje.csv' # Asegúrate de tener el archivo con este nombre
df = pd.read_csv(archivo_csv, encoding='latin1') # Usamos 'latin1' para manejar los caracteres incorrectos

# Mostrar los primeros registros para asegurarnos de que se cargó correctamente
print(df.head())

# Limpiar los nombres de las columnas (eliminar espacios extras)
df.columns = df.columns.str.strip()

# Convertir el porcentaje a formato numérico, quitando el signo de '%' y convirtiendo a float
df['Porcentaje'] = df['Porcentaje'].replace('%', '', regex=True).astype(float)

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

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

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

# Generar predicciones
y_pred = model.predict(X)

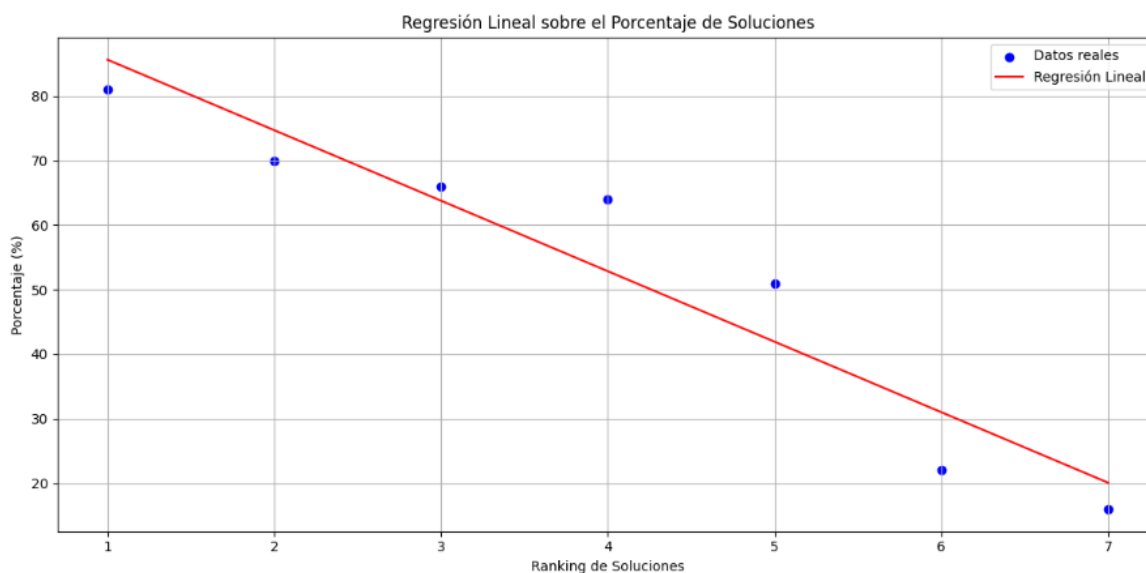
# Mostrar los coeficientes del modelo
print("Pendiente (coeficiente):", model.coef_[0])
print("Intercepto:", model.intercept_)

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

# Etiquetas y título
plt.xlabel("Ranking de Soluciones")
plt.ylabel("Porcentaje (%)")
plt.title("Regresión Lineal sobre el Porcentaje de Soluciones")
plt.legend()
plt.grid(True)
plt.tight_layout()

# Mostrar el gráfico
plt.show()
```

Grafica:



Hipótesis:

A medida que el ranking de las soluciones desciende (es decir, las soluciones son menos populares o utilizadas), el porcentaje de preferencia o adopción también disminuye. Esto sugiere que las soluciones más utilizadas probablemente destacan por su eficacia, accesibilidad, o soporte en la industria, mientras que las menos adoptadas podrían estar limitadas a casos de uso específicos, carecer de soporte o no cumplir completamente con las necesidades de los usuarios.