

1. Estadísticas de la Externalización de la Ingeniería de Software

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

# Convertir la columna "Frecuencia_Outourcing(%)" a valores numéricos
df["Frecuencia_Outourcing(%)"] = df["Frecuencia_Outourcing(%)"].astype(float)

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

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

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

# Generar predicciones
y_pred = model.predict(X)

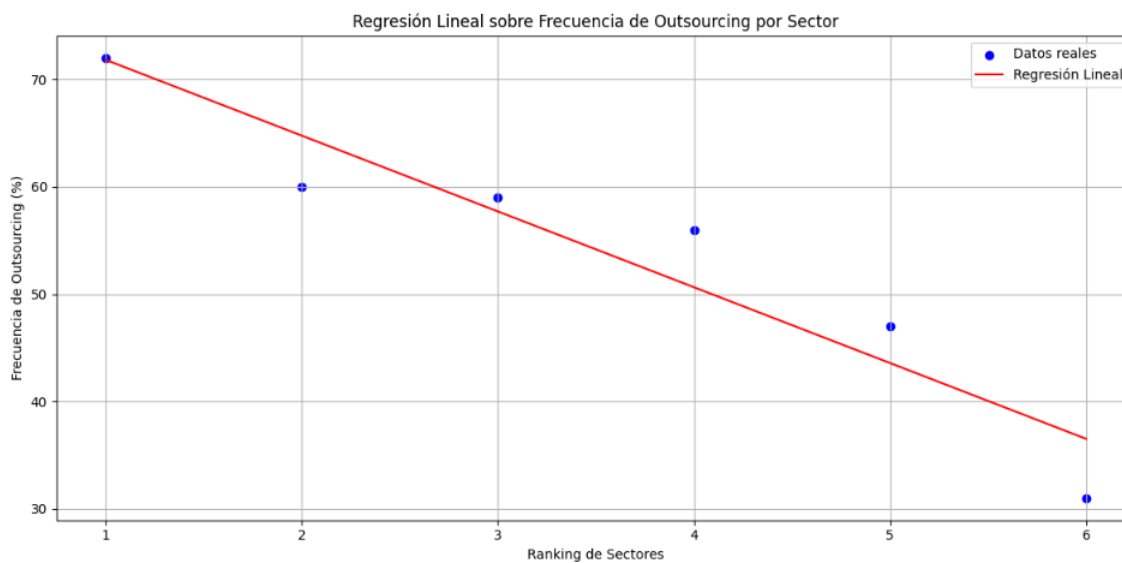
# Mostrar los coeficientes del modelo
print("Pendiente (coeficiente):", model.coef_[0])
print("Intercepto:", model.intercept_)

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

# Etiquetas y título
plt.xlabel("Ranking de Sectores")
plt.ylabel("Frecuencia de Outsourcing (%)")
plt.title("Regresión Lineal sobre Frecuencia de Outsourcing por Sector")
plt.legend()
plt.grid(True)
plt.tight_layout()

# Mostrar el gráfico
plt.show()
```

Grafica:



Hipótesis:

La frecuencia de outsourcing en los sectores varía inversamente con el ranking, lo que indica que los sectores con mayor porcentaje de outsourcing tienden a ser los más dependientes de servicios externos debido a la naturaleza de sus operaciones o la necesidad de optimización de recursos.