

1. Porcentaje de empleos por rol

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

# 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)

# Verificar si hay valores NaN y eliminarlos si es necesario
if df.isnull().values.any():
    print("Advertencia: Hay valores NaN en los datos. Se eliminarán las filas con valores NaN.")
    df = df.dropna()

# 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

# Verificar el shape de X y y
print(f"Shape de X: {X.shape}, Shape de y: {y.shape}")

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

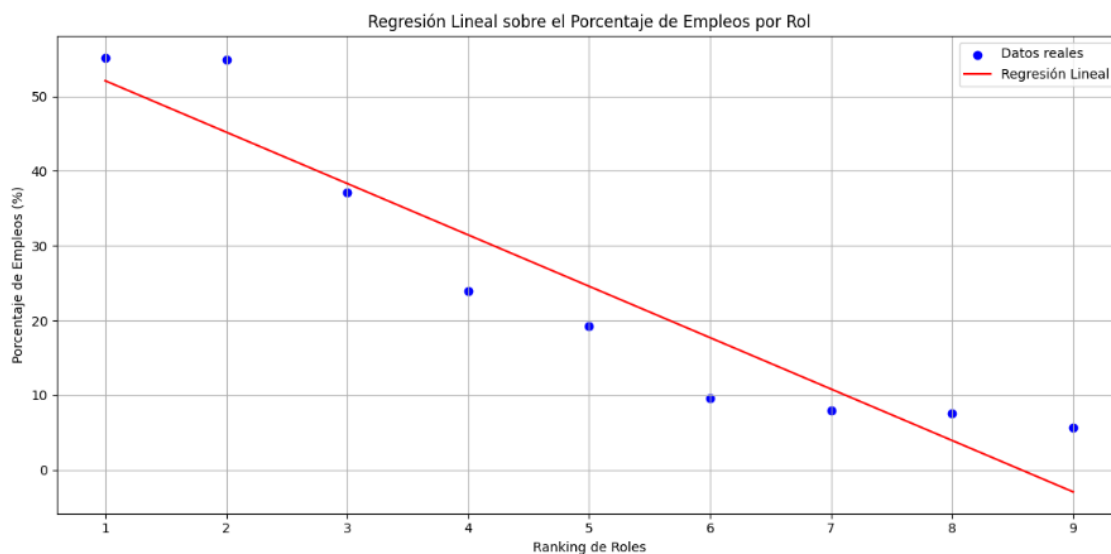
# Generar predicciones
y_pred = model.predict(X)

# 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 Roles")
plt.ylabel("Porcentaje de Empleos (%)")
plt.title("Regresión Lineal sobre el Porcentaje de Empleos por Rol")
plt.legend()
plt.grid(True)
plt.tight_layout()
```

Grafica:



Hipótesis:

El porcentaje de empleos disponibles para diferentes roles tiende a disminuir a medida que aumenta su ranking numérico (es decir, a medida que son menos demandados). Los roles más populares y de mayor demanda en el mercado concentran un porcentaje significativo de las oportunidades de empleo, mientras que los roles menos comunes presentan una menor participación. La relación entre el ranking de roles y el porcentaje de empleos muestra una tendencia lineal decreciente, aunque puede haber variaciones debido a la especialización o demandas específicas en ciertos sectores.