Read Source from Excel

```
import pandas as pd
df = pd.read_excel('Salary_DataEmpty.xlsx', sheet_name='RAW')
df.head()
```

```
   Año  Salario Real
0  1.1         39343
1  1.3         46205
2  1.5         37731
3  2.0         43525
4  2.2         39891
```

Linear Regression

Linear Regression for two variables is a simple model or function and can be modeled by the following expressión:

$$Y = \beta_0 + \beta_1 X_i$$

where

$$\beta_1 = r\frac{S_y}{S_x} = \frac{n\sum(x \cdot y) - \sum x \sum y}{n\sum(x^2) - (\sum x)^2}$$

and

$$\beta_0 = \frac{\sum y - \beta_1 \sum x}{n}$$

Calculate varaibles

```
# En nuestro dataset el Año es la variable independiente (x) y el
Salraio Real es la varaible dependeinte (y)

sum_xy = sum(df['Año']*df['Salario Real'])
sum_x = sum(df['Año'])
sum_y = sum(df['Salario Real'])
n = len(df['Año'])
sum_x2 = sum(df['Año']*df['Año'])
sum2_x = sum(df['Año'])**2
```

Calculate the equation of the line

```
def linear_regression(x:float) -> float:
    """Linear regression function, receives an x variable and returns
an estimated variable"""
    beta_1 = (n*sum_xy-sum_x*sum_y)/(n*sum_x2-sum2_x)
    beta_0 = (sum_y - beta_1*sum_x)/n
    return beta_0 + beta_1*x
```

Show the results

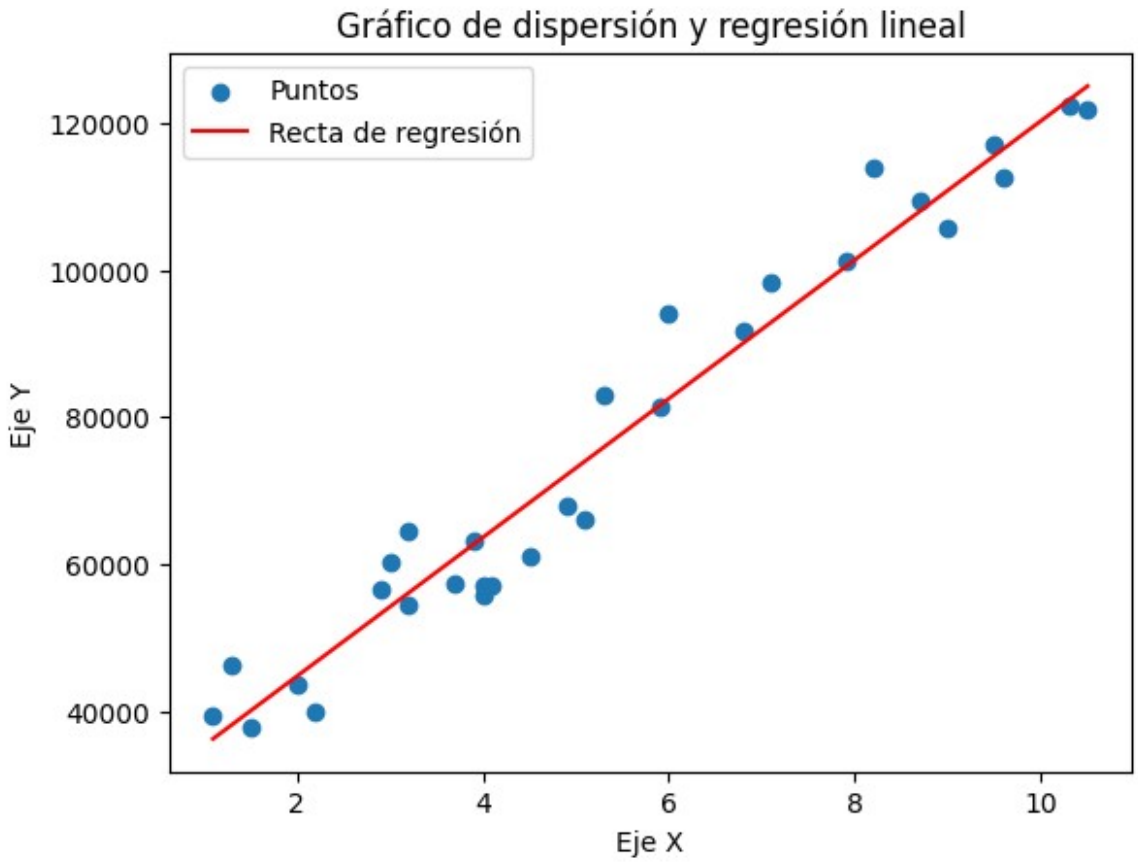
```
import matplotlib.pyplot as plt

# Crear el gráfico de dispersión de puntos
plt.scatter(df['Año'], df['Salario Real'], label='Puntos')

# Trazar la línea de regresión lineal
plt.plot(df['Año'], df['Año'].apply(linear_regression), color='red',
label='Recta de regresión')

# Configurar etiquetas y título
plt.xlabel('Eje X')
plt.ylabel('Eje Y')
plt.title('Gráfico de dispersión y regresión lineal')
plt.legend()

# Mostrar el gráfico
plt.show()
```



Answering questions

What can be the salary of a programmer in 6 years?

```
print(f"The salary would be {linear_regression(6)}")
```

```
The salary would be 82491.97412739915
```

What can be the salary of a programmer in 7 years?

```
print(f"The salaray would be {linear_regression(7)}")
```

```
The salaray would be 91941.93644885422
```

What can be the salary of a programmer in 20 years?

```
print(f"The salary would be {linear_regression(20)}")
```

```
The salary would be 214791.44662777026
```

What can be the salary of a programmer at the beginning of his career?

```
print(f"The salary would be {linear_regression(0)}")
```

```
The salary would be 25792.200198668685
```