**U N I V E R S I D A D D E G U A D A L A J A R A**

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍA

DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

**Seminario de Solución de Problemas de Sistemas Basados en Conocimiento**

Práctica No. 2

Nombre: Hurtado González Edgar Arturo

Código: 212597894

**Introducción**

La regresión lineal simple es un método estadístico utilizado para modelar la relación entre una variable dependiente (la que se quiere predecir) y una variable independiente (la que se utiliza para la predicción) mediante una línea recta.

En su forma más básica, la ecuación de la regresión lineal simple es:

Donde:

* es la variable dependiente que se quiere predecir.
* es la variable independiente.
* es la intersección con el eje Y (el valor de cuando es igual a 0).
* es la pendiente de la línea, que indica cómo cambia cuando aumenta en una unidad.
* es el término de error, que representa la diferencia entre el valor real de y el valor predicho por el modelo.

El objetivo de la regresión lineal es encontrar los valores de y que minimizan la suma de los cuadrados de los errores (mínimos cuadrados), es decir, encontrar la línea que mejor se ajuste a los datos observados.

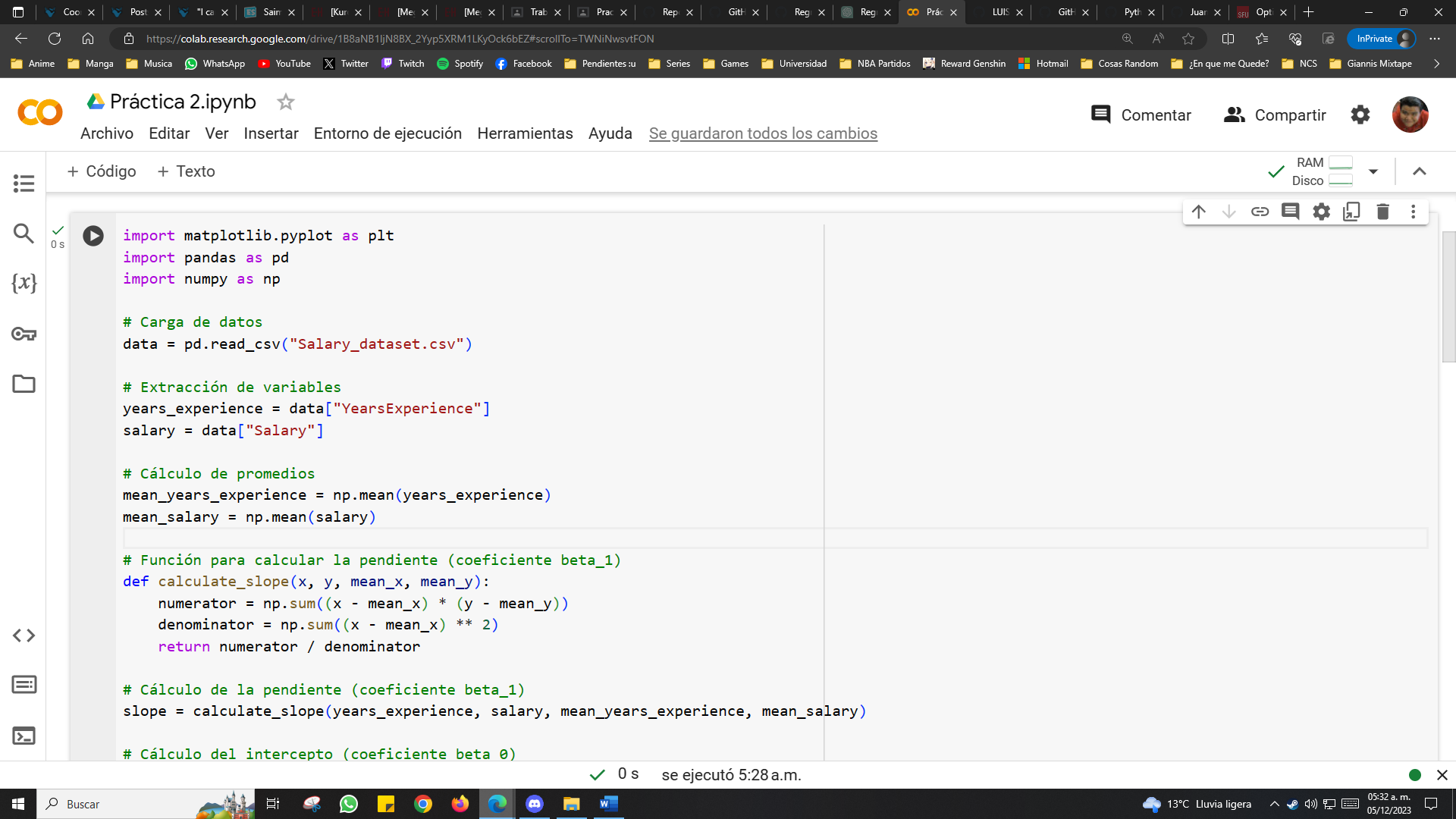
Para hacer esto, se utilizan técnicas estadísticas como el método de mínimos cuadrados o algoritmos de optimización para calcular los coeficientes de y . Una vez que se ha ajustado el modelo, se puede utilizar para predecir los valores de la variable dependiente para nuevos valores de la variable independiente .

La regresión lineal simple es una herramienta poderosa y ampliamente utilizada en análisis de datos, predicción y modelado estadístico debido a su simplicidad y facilidad de interpretación. Sin embargo, es importante tener en cuenta sus suposiciones, como la linealidad y la independencia de los errores, para obtener resultados precisos y significativos.

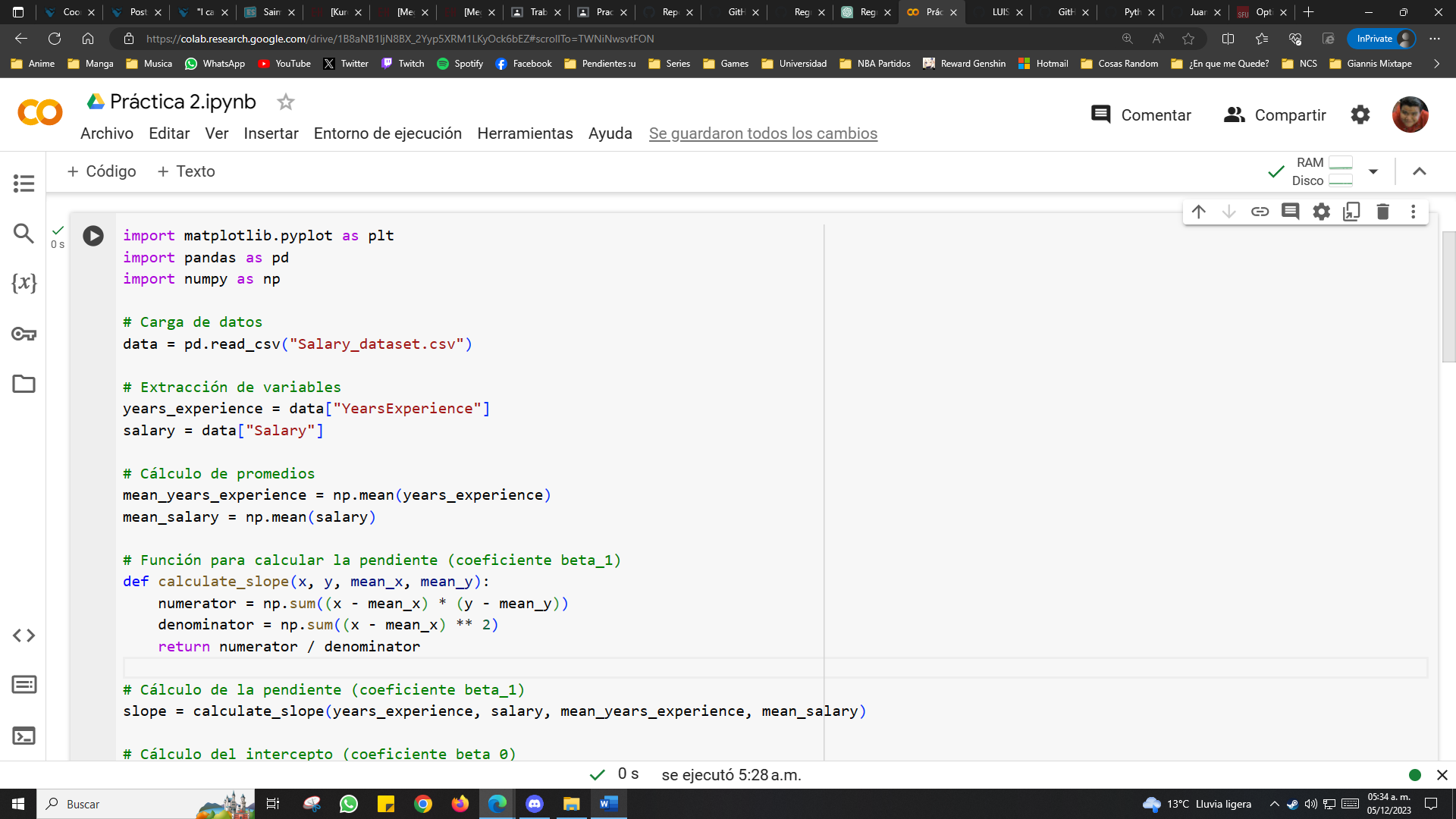
**Desarrollo**

En la Práctica se realizó lo siguiente:

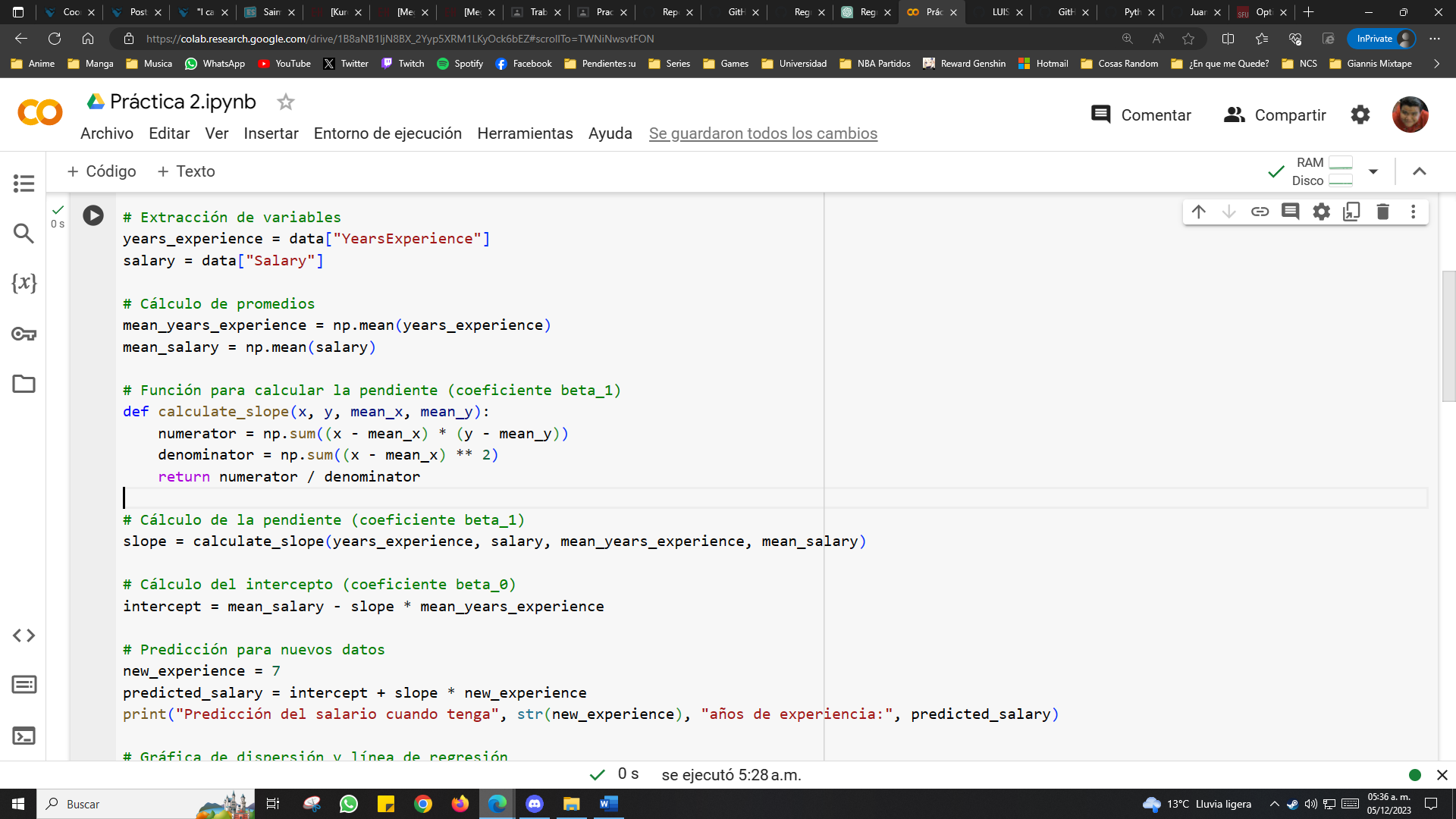
1. **Carga de datos:** Se utilizó la librería **“pandas”** para cargar un conjunto de datos desde un archivo CSV, que contiene información sobre años de experiencia y salarios.



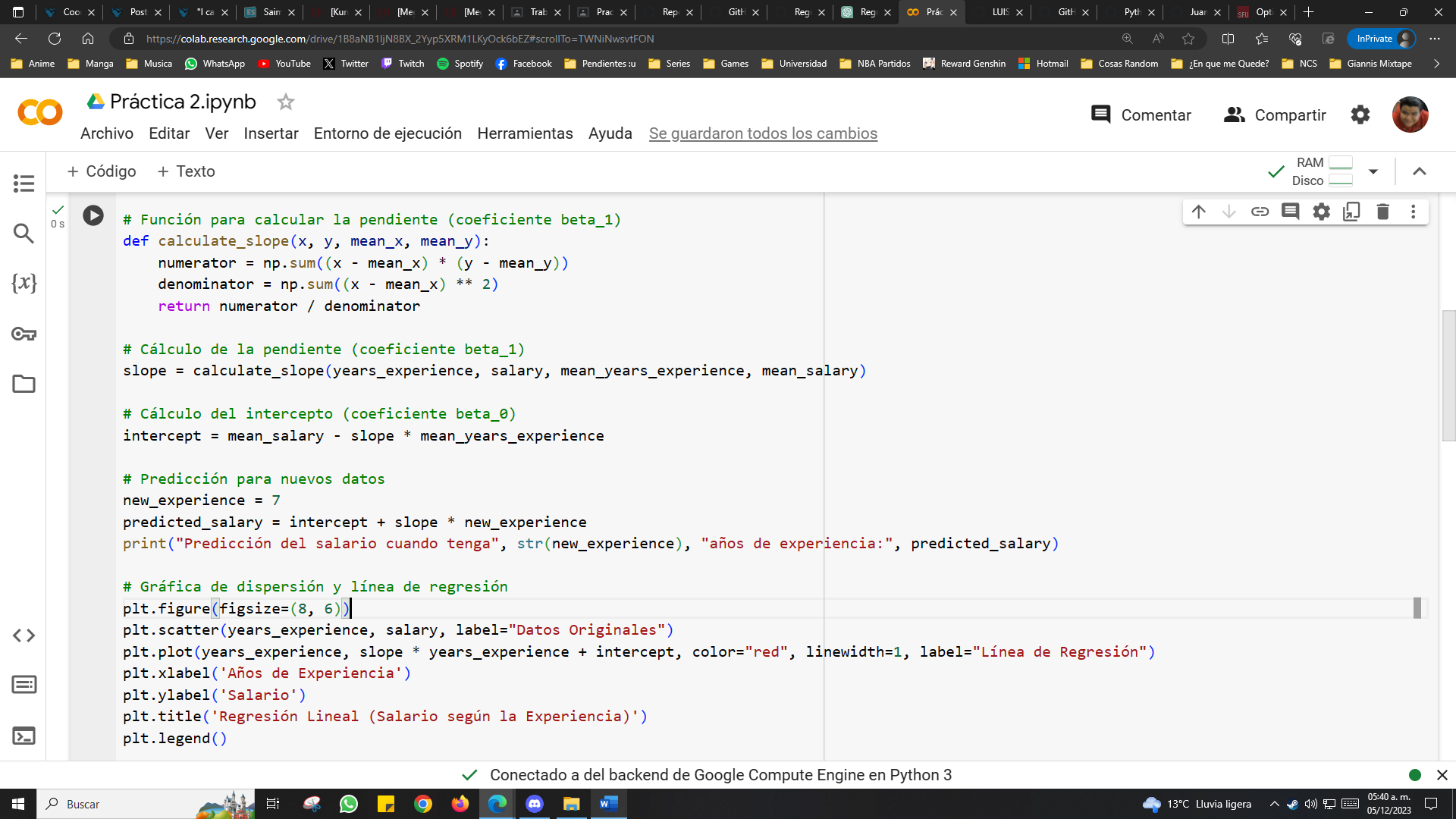
1. **Extracción de variables:** Se tomaron las columnas relevantes del conjunto de datos para las variables **“years\_experience”** y **“salary”**.



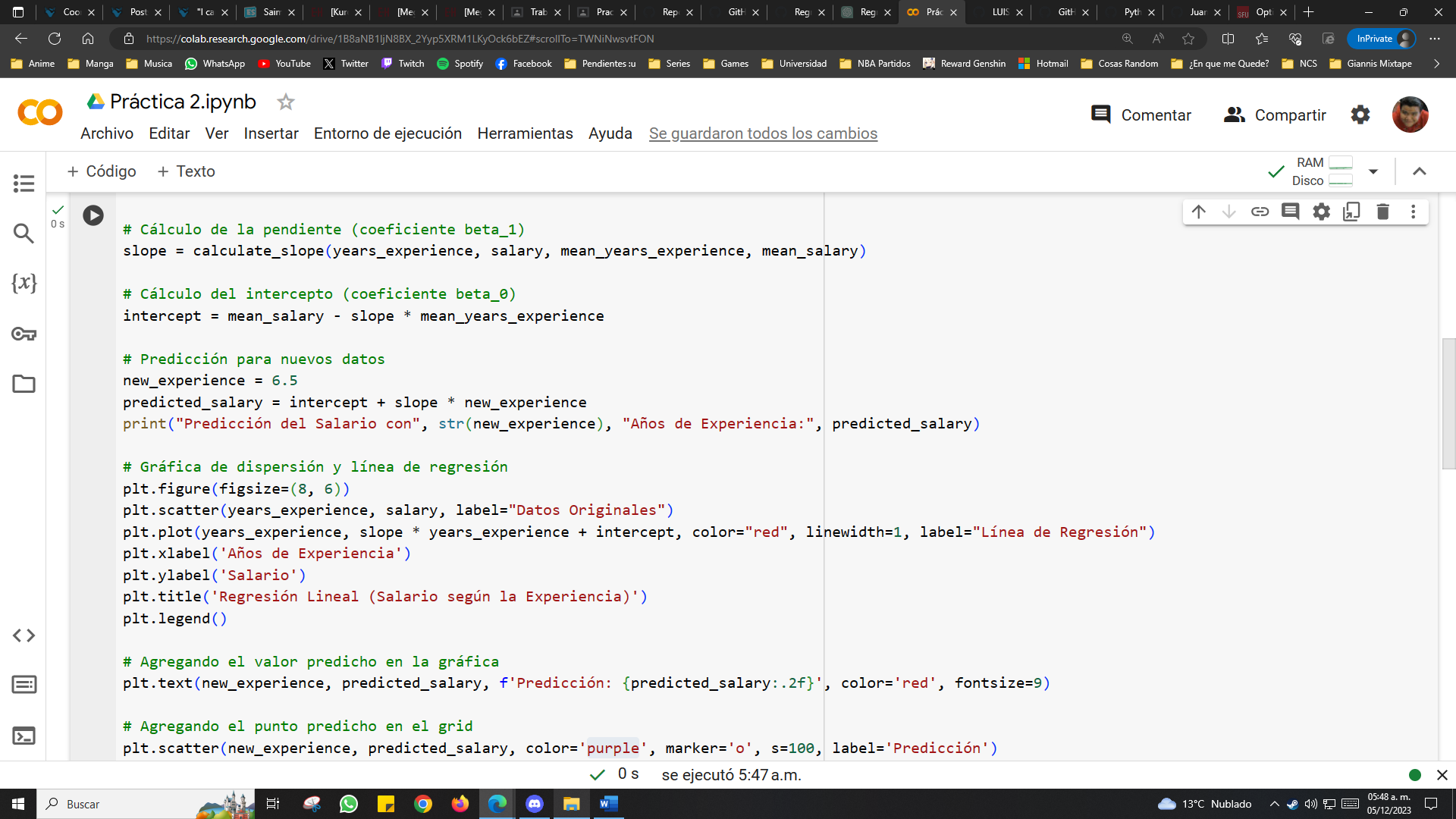
1. **Cálculo de promedios:** Se calcularon los promedios de años de experiencia y salarios utilizando la función **“np.mean()”** de la librería **“numpy”**.



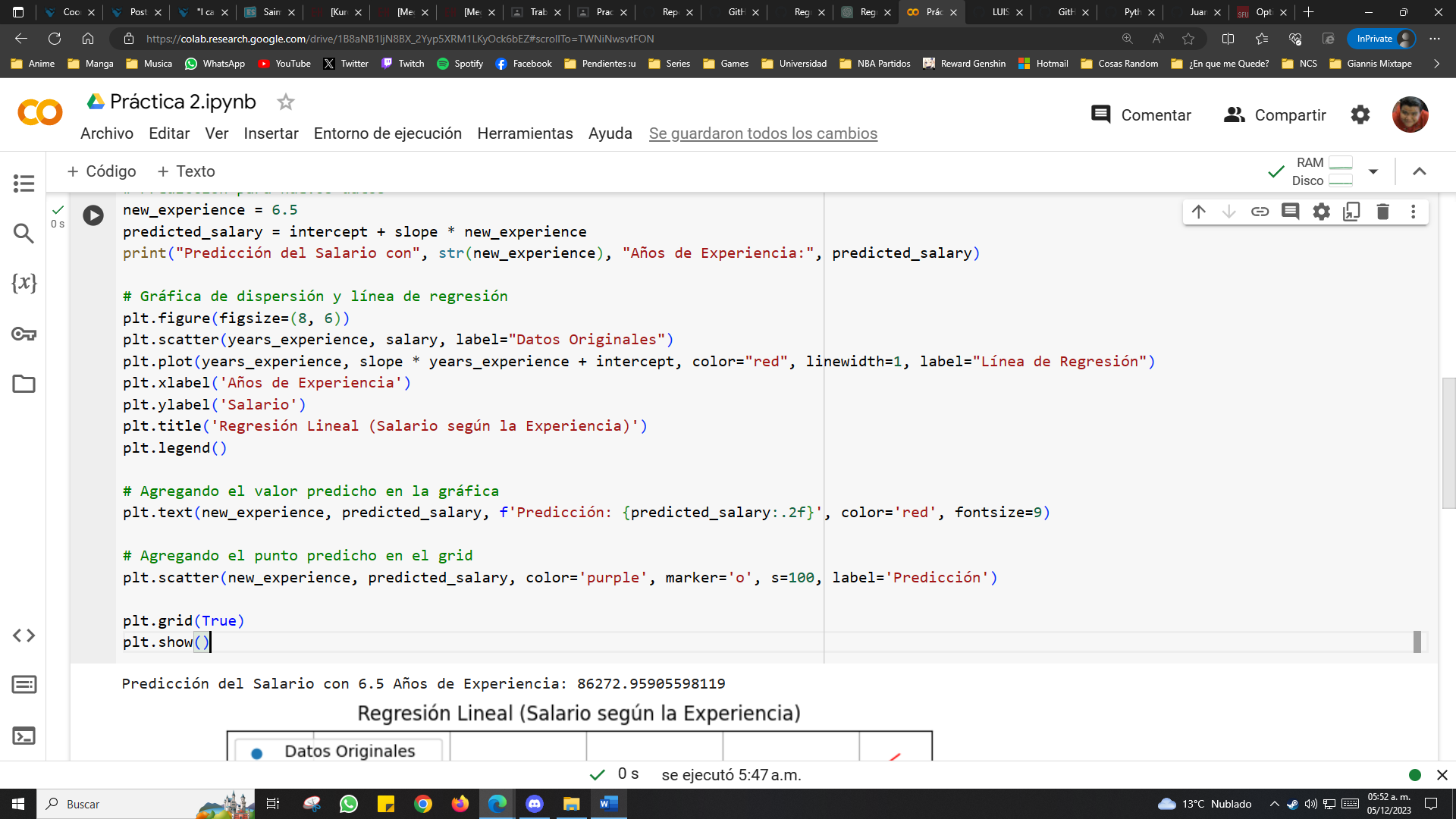
1. **Cálculo de la pendiente y el intercepto:** Se implementó una función **“calculate\_slope()”** para calcular la pendiente (coeficiente ) utilizando la fórmula de regresión lineal. También se calculó el intercepto (coeficiente).



1. **Predicción para nuevos datos:** Se definió un nuevo valor de años de experiencia **“new\_experience”** y se utilizó la ecuación de la línea de regresión (basada en la pendiente y el intercepto calculados) para predecir el salario correspondiente a ese nuevo valor.



1. **Gráfica de dispersión y línea de regresión:** Se utilizó **“matplotlib”** para visualizar los datos originales como una gráfica de dispersión y trazar la línea de regresión lineal obtenida sobre estos datos. Además, se añadió el punto predicho y su valor en la gráfica para ilustrar la predicción realizada.



**Conclusión/Resultados**

El ejercicio realizado muestra un proceso básico de regresión lineal simple para predecir salarios basados en años de experiencia. A través del uso de datos reales, se ilustró cómo se pueden calcular la pendiente y el intercepto de una línea de regresión para modelar la relación entre dos variables. La regresión lineal simple es una herramienta poderosa para analizar y predecir relaciones entre variables cuando se asume una relación lineal.

