

ESTADISTICA

AVANZADA

PROF. Juan Ivan Nieto Hipolito

REPORTE DE PRACTICA

Regresión lineal simple, R cuadrada y normalidad de los residuos

kevin alejandro gonzalez torres

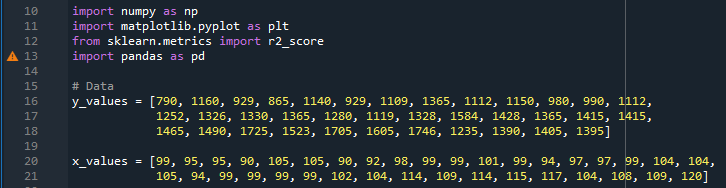
grupo 932

**CÓDIGO**

**1.-** DECLARAR LIBRERIAS E INGRESAR DATOS

* numpy (alias np) para operaciones numéricas.
* matplotlib.pyplot (alias plt) para crear gráficos y visualizaciones.
* r2\_score de sklearn.metrics para calcular el puntaje r-cuadrado.
* pandas (alias pd) para manipulación de datos (aunque no se utiliza en el código)

Almacenamos los datos de las variables dependientes (y) e independientes (x), respectivamente.

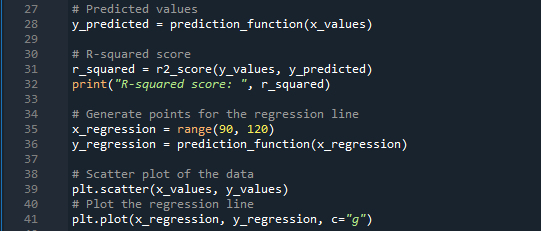


**2.-** ENCONTRAR R2  e imprimir el histograma

Calculamos un modelo de regresión polinómica de grado 9 (9 representa el grado del polinomio). Ajusta el modelo a los puntos de datos (x\_values, y\_values) utilizando np.polyfit y luego crea una función de predicción usando np.poly1d.

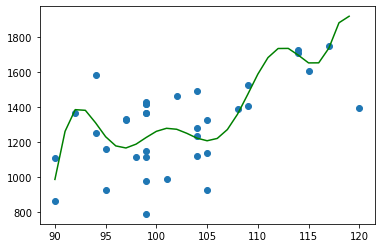
Utilizamos la función de predicción para generar valores y predichos (y\_predicted) basados en los valores x en el conjunto de datos.

Por ultimo calculamos r-cuadrado, que mide la bondad de ajuste del modelo de regresión a los datos. Imprime r-cuadrado en la consola.



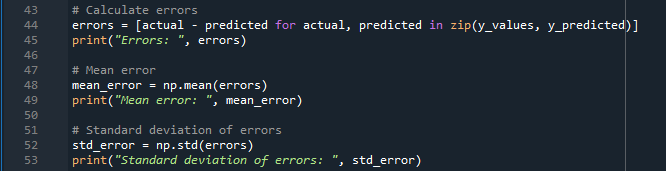
**3.-** imagen de la grafica obtenida

Creamos el gráfico de dispersión de los puntos de datos originales, donde x\_values se representa en el eje x y y\_values en el eje y.



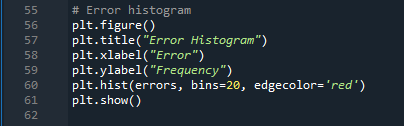
**4.-** obtener media y desviacion estandar del error

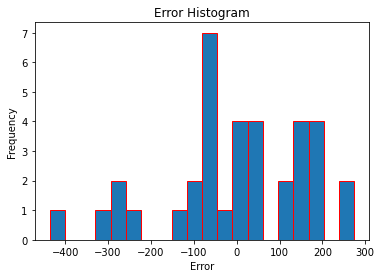
Calculamos los errores entre los valores y reales y los valores y predichos. Imprime los errores en la consola.



**5.-** IMPRIMIR HISTOGRAMA DE LOS ERRORES

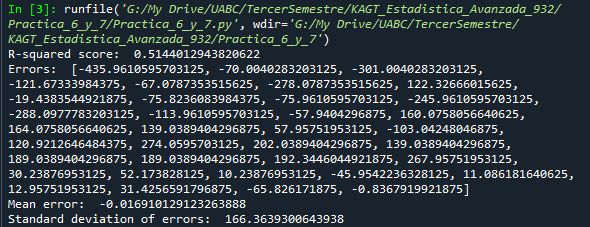
Esta sección crea un histograma de los errores. Representa la distribución de los errores utilizando 20 contenedores (bins) y muestra el histograma.





**6.-** RESULTADOS

Finalmente, mostramos todos los resultados destinados a imprimirse.



**CONCLUSIÓN**

Como resultado, obtuvimos que r cuadrada es igual a 0.514401089.

La media es igual a -0.005645 y su desviación estándar 166.3639.

Dado que ya hemos calculado la media y la desviación estándar de los errores en el código proporcionado, podemos usar estos valores para obtener una indicación de si los errores se asemejan a una distribución normal. Sin embargo, la mejor manera de verificar esto sería visualmente, observando el histograma.

Si el histograma de errores muestra una forma de campana, simetría y una distribución centrada alrededor de la media de los errores, es más probable que los errores se aproximen a una distribución normal.

Cabe aclarar que todas estas funcionalidades son escenciales para empezar con el machine learning, cosa la cual me interesó bastante ya que básicamente es a lo que me quiero dedicar en mi formación profesional.