**LINK PROYECTO:** <https://github.com/Bismarxd/PROYECTO_FINAL>

**NOMBRE: MAYTA TINTAYA BISMARCK**

**CORREO ELECTRÓNICO: bismarckmaytatintaya@gmail.com**

**CELULAR: 69933860**

**DESCRIPCIÓN GENERAL DEL PROBLEMA**

Sabemos que la enfermedad que en estos tiempos el mundo vive en grave problema afectado por una pandemia, y nuestro país no es ajeno a eso conforme pasaba el tiempo las cifras de contagiados aumentaba y en nuestro dataset vemos como aumentaba los casos por día en cada departamento y como esta enfermedad a afectado a nuestro país conforme pasaba el tiempo.

**DESCRIPCIÓN DE LAS COLUMNAS DEL DATA SET**

Nuestro data set cuenta con 11 columnas:

* Fecha: esta columna nos indica las diferentes fechas desde el inicio de la pandemia hasta el 20 de julio.
* Región: es el departamento en los que a ocurrido los casos.
* Casos: es el numero de casos de esa fecha.
* Recuperados: en Nro. de recuperados en ese día.
* Muertes: la cantidad de personas que murieron ese día en esa región.
* Población: es la cantidad de población que habita en esa región.
* Activos: son los pacientes activos que se encuentran hasta esa fecha.
* Activos por 10 k: cuantos activos hay cada 10.000 personas.
* Muertes por 1 millos: cuantas muertes hay cada 1 millo de personas.
* Susceptibles: cuantas personas se encuentran susceptibles a contraer la enfermedad
* Descartados: cuantos casos descartados hay en esa fecha.

**DESCRIPCIÓN BÁSICA DE LAS SOLUCIONES**

1. **EL PROCESO BÁSICO DE ANÁLISIS DE DATOS**

**PREPROCESAMIENTO (TÉCNICAS UTILIZADAS)**

**TRANSFORMACIÓN DE DATOS**

Se hace este pre procesamiento haciendo uso del Scikit-learn que nos incluye varios algoritmos, la cual también esta diseñada para trabajar con las bibliotecas numpy.

Para hacer la transformación de datos estandarizamos los datos eliminando la media y escala los datos para que su varianza sea 1.

**VALORES PERDIDOS Y FILAS DUPLICADAS**

Se hace de la biblioteca panda para trabajar con el dataset y se hace una serie de verificaciones y selección de filas o columnas para poder borrarlas o reemplazarlas.

**PREPROCESAMIENTO WEKA**

Se hace un preprocesamiento que reemplaza los datos faltantes con la media de las columnas.

1. **CONSTRUCCIÓN DE MÉTODOS**

**RED NEURONAL**

Se guarda en dataframe los datos del dataset y se selecciona las variables a predecir.

Haciendo uso del Sklearn para hacer el pre procesamiento.

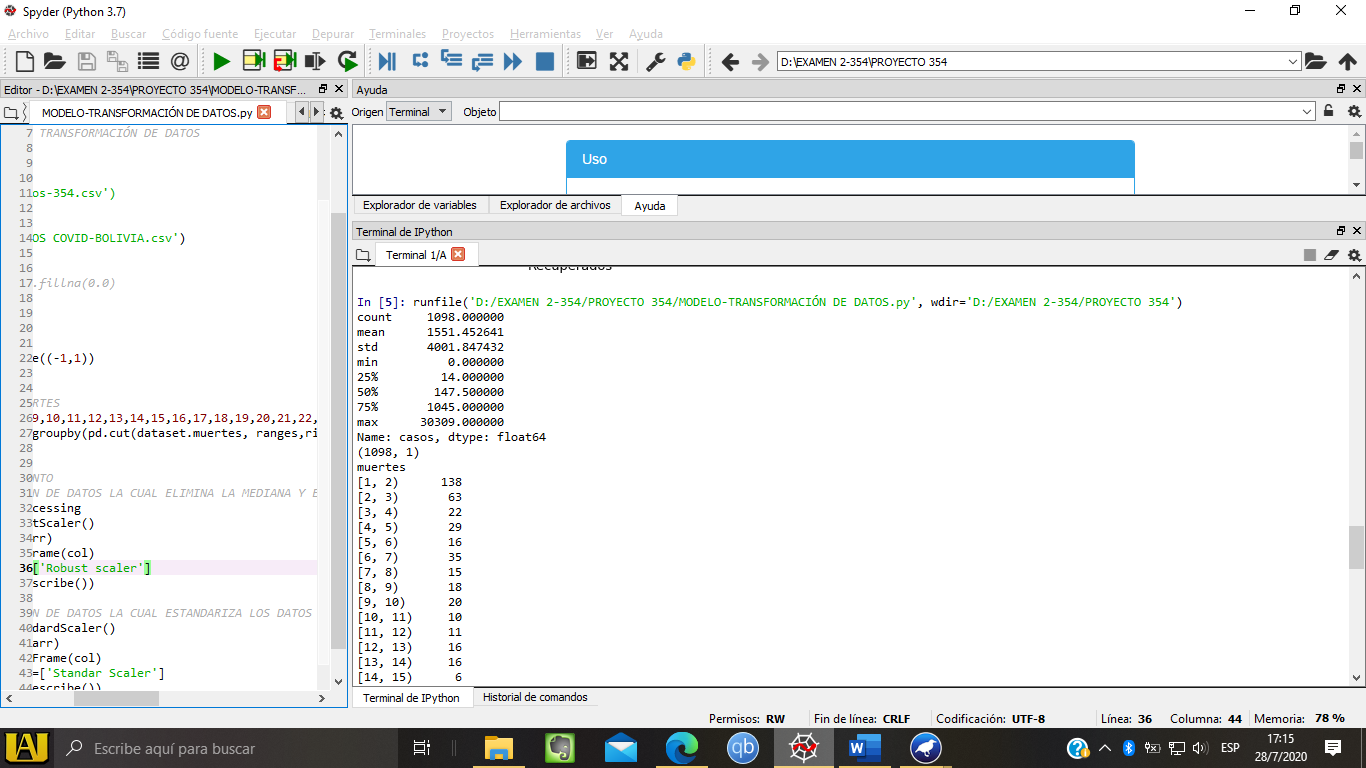
Se hace uso de la biblioteca keras la cual es una biblioteca de redes neuronales, por ultimo se optimiza y evalúa el modelo.

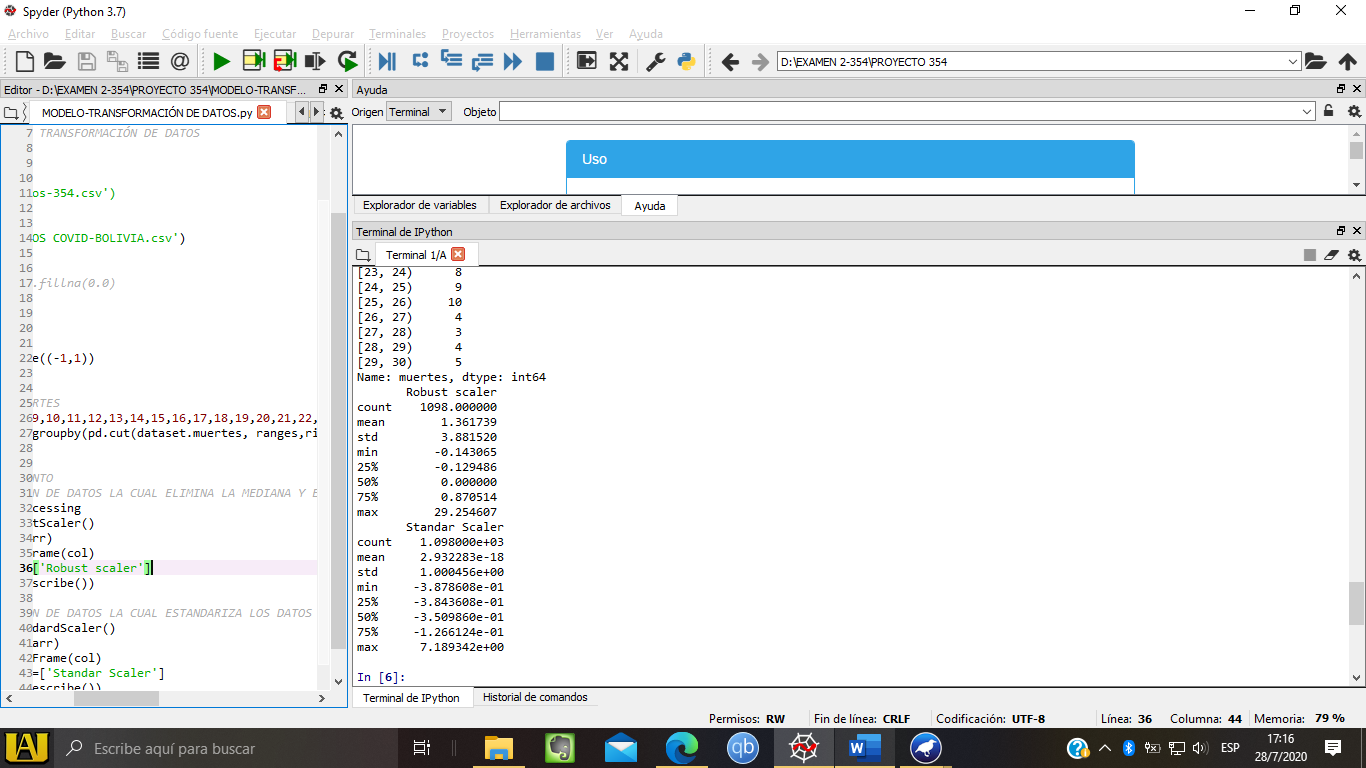
**REGRESIÓN QUE VARIABLES SE TOMAN EN CUENTA**

Para la regresión lineal se toman en cuenta las variables de “muertes” y “recuperados”

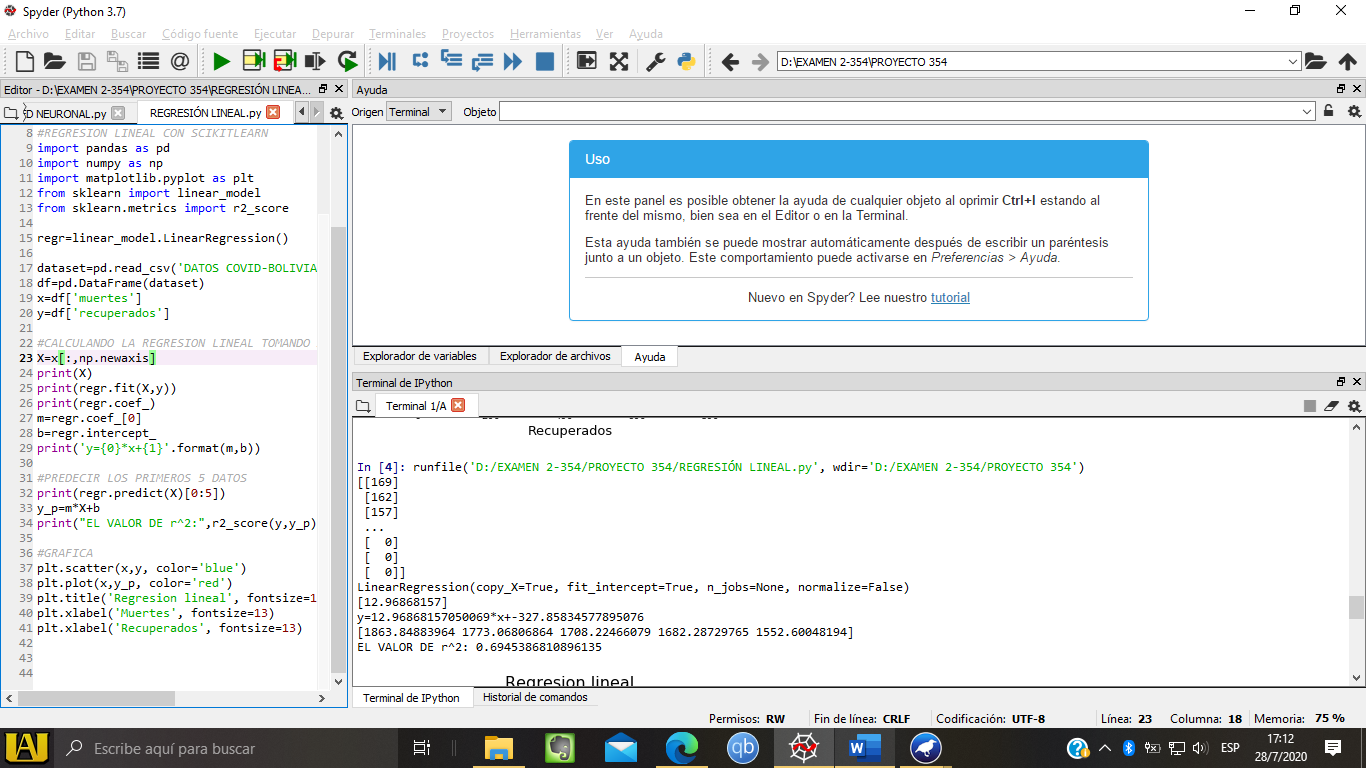
**PRUEBAS**

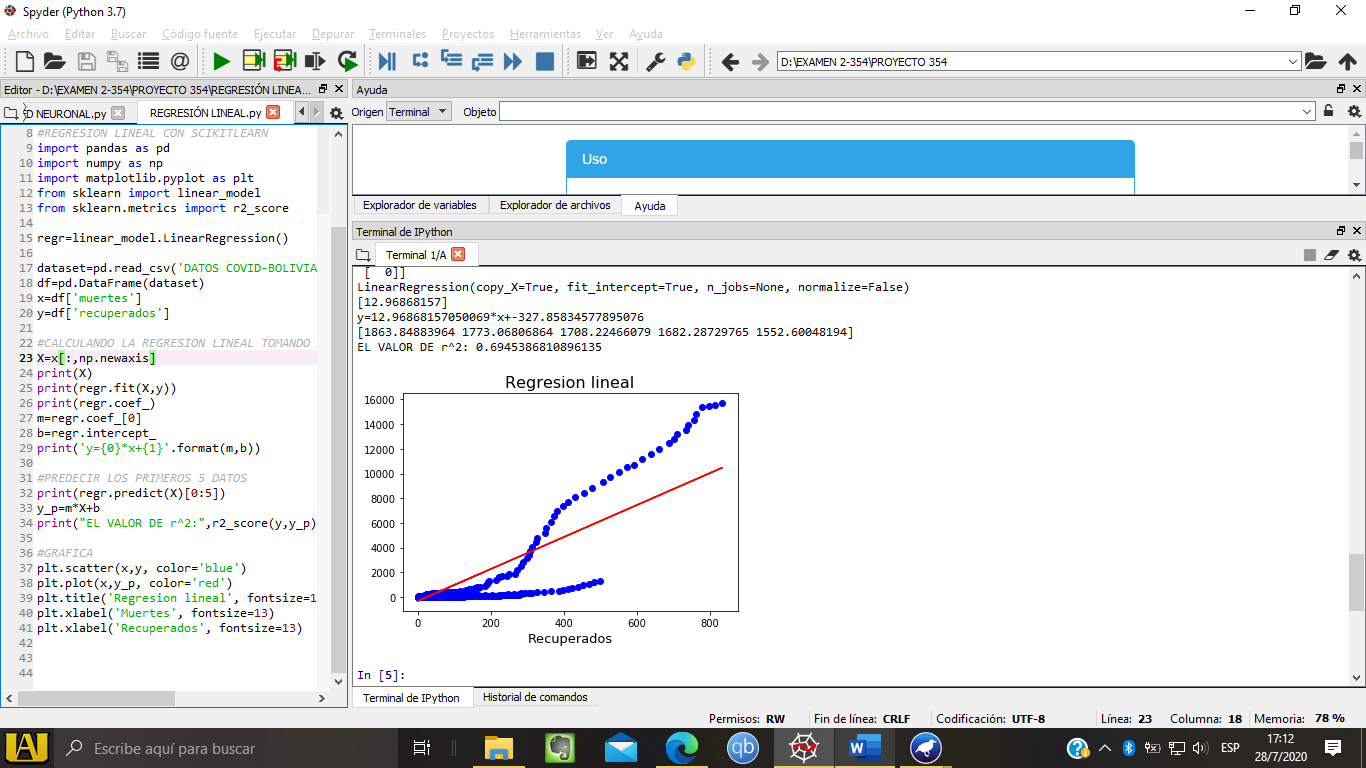
**PREPROCESAMIENTO-TRANSFORMACIÓN DE DATOS**





**REGRESIÓN LINEAL**





**APRENDIZAJE SE HA TENIDO EN EL MODELO PLANTEADO**

en el modelo planteado se aprendí a usar la biblioteca keras la cual nos ayuda en las redes neuronales, la cual tiene el código alojado en git hub y a calcular la regresión lineal haciendo uso de 2 atributos y graficando con la biblioteca matplotlib.