| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: HUGO ROLANDO | DNI:0503268997 | |  |
| Apellidos: SANCHEZ QUISPE |

| ESTUDIO | ASIGNATURA | CONVOCATORIA |
| --- | --- | --- |
| MÁSTER UNIVERSITARIO EN INGENIERÍA MATEMÁTICA Y COMPUTACIÓN (PLAN 2016) | 4391020006.- TÉCNICAS MULTIVARIANTES | Ordinaria Número periodo 1823 |

| FECHA | MODELO | CIUDAD DEL EXAMEN |
| --- | --- | --- |
| 14-16/01/2022 | Modelo - B |  |

| Etiqueta identificativa |
| --- |
|  |

**INSTRUCCIONES GENERALES**

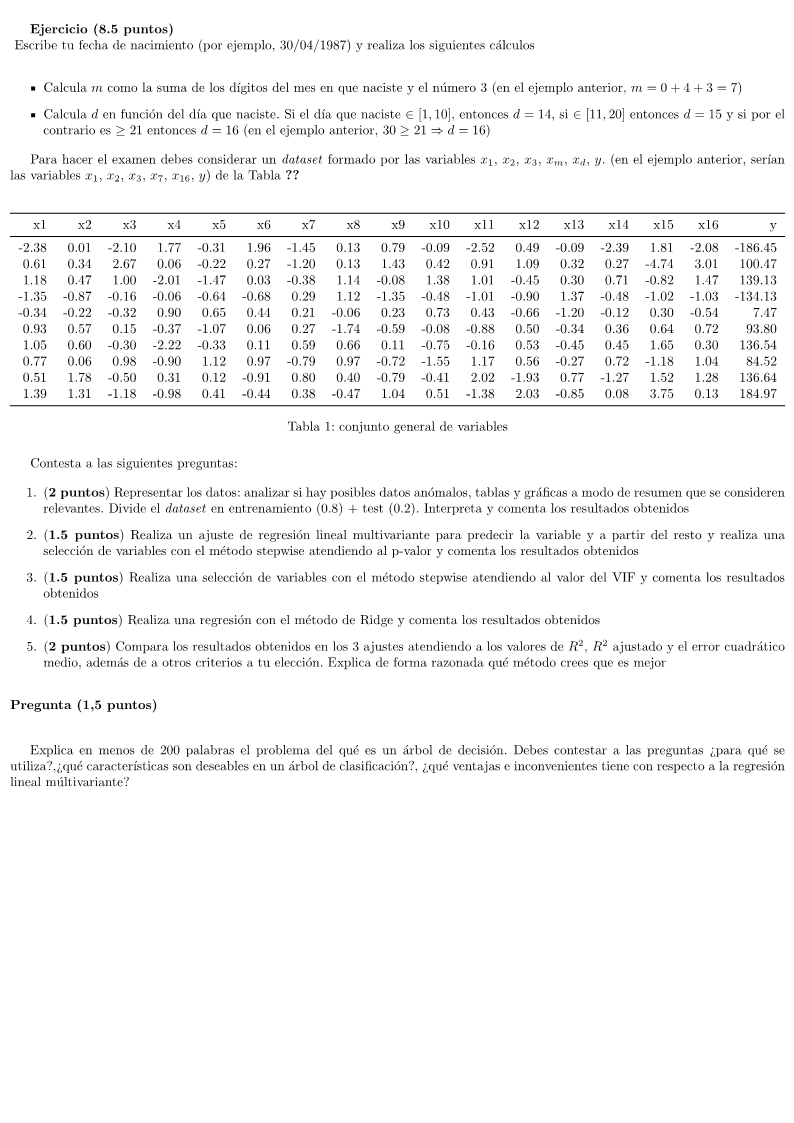
1. Ten disponible tu documentación oficial para identificarte, en el caso de que se te solicite.
2. Si tu examen consta de una parte tipo test, indica las respuestas en la plantilla según las características de este.
3. Debes contestar en el documento adjunto, respetando en todo momento el espaciado indicado para cada pregunta. Si este es en formato digital, los márgenes, el interlineado, fuente y tamaño de letra vienen dados por defecto y no deben modificarse. En cualquier caso, asegúrate de que la presentación es suficientemente clara y legible. Entrega toda la documentación relativa al examen, revisando con detenimiento que los archivos o documentos son los correctos. El envío de archivos erróneos o un envío incompleto supondrá una calificación de “no presentado”.
4. Durante el examen y en la corrección por parte del docente, se aplicará el Reglamento de Evaluación Académica de UNIR que regula las consecuencias derivadas de las posibles irregularidades y prácticas académicas incorrectas con relación al plagio y uso inadecuado de materiales y recursos.
5. No está permitido el uso de Internet ni ningún tipo de comunicación con otra persona.Durante todo el examen tu teléfono móvil debe estar en modo avión.
6. La parte principal de cada pregunta consiste en interpretar y comentar los resultados obtenidos. Si te limitas a hacer los cálculos no vas a poder superar el examen.
7. Es fundamental que las respuestas estén debidamente redactadas, de forma clara y precisa y sin faltas de ortografía.
8. Para hacer el examen puedes utilizar los apuntes del curso y los scripts que hayas preparado y Python para hacer los cálculos.

**Puntuación**

**Preguntas**

El examen constará de un ejercicio práctico (8,5 puntos) y una pregunta teórica (1,5 puntos). Los enunciados están en la página 14 y el espacio para responder el examen está entre las práginas 4 y 13.  
  
**1.** Pregunta

 (Responder en 10 caras)



RESOLUCION DEL EXAMEN

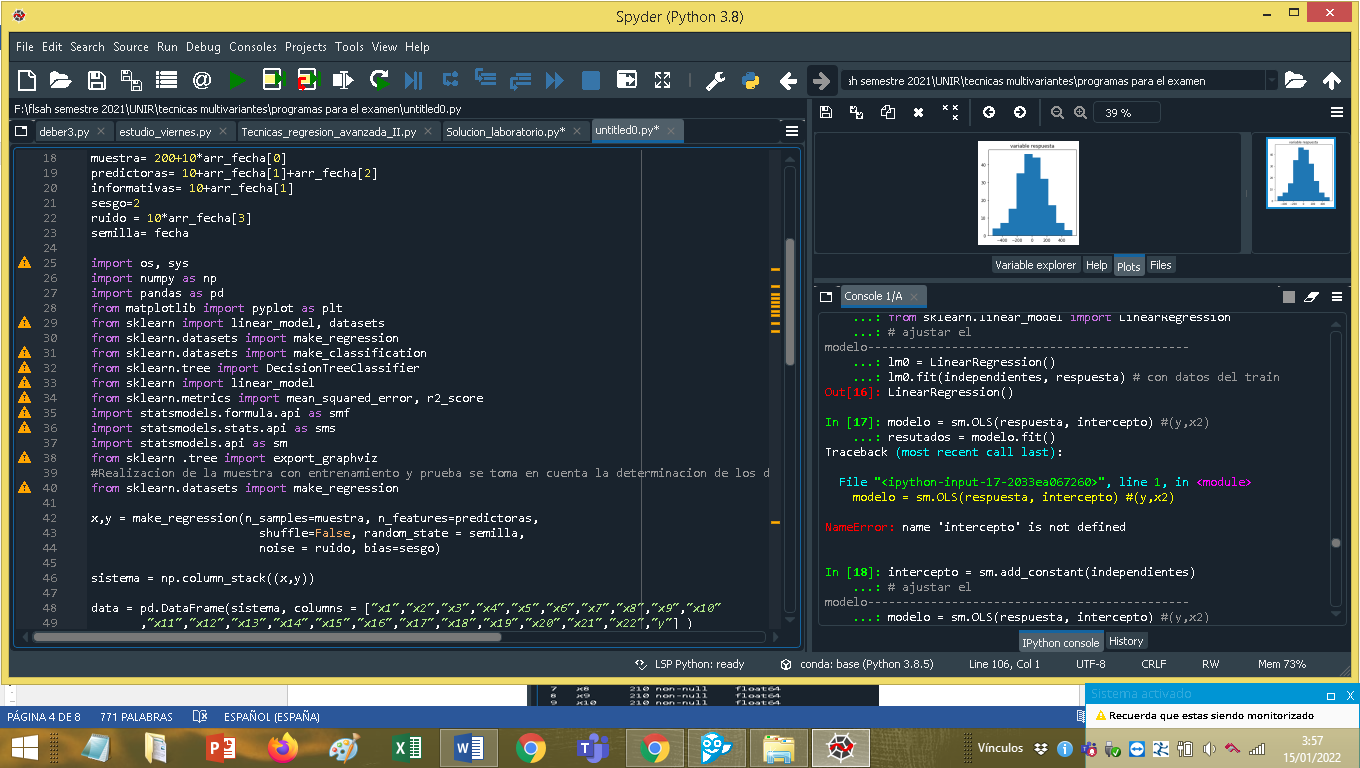
1. Representacion de datos en base a las indicaciones dadas

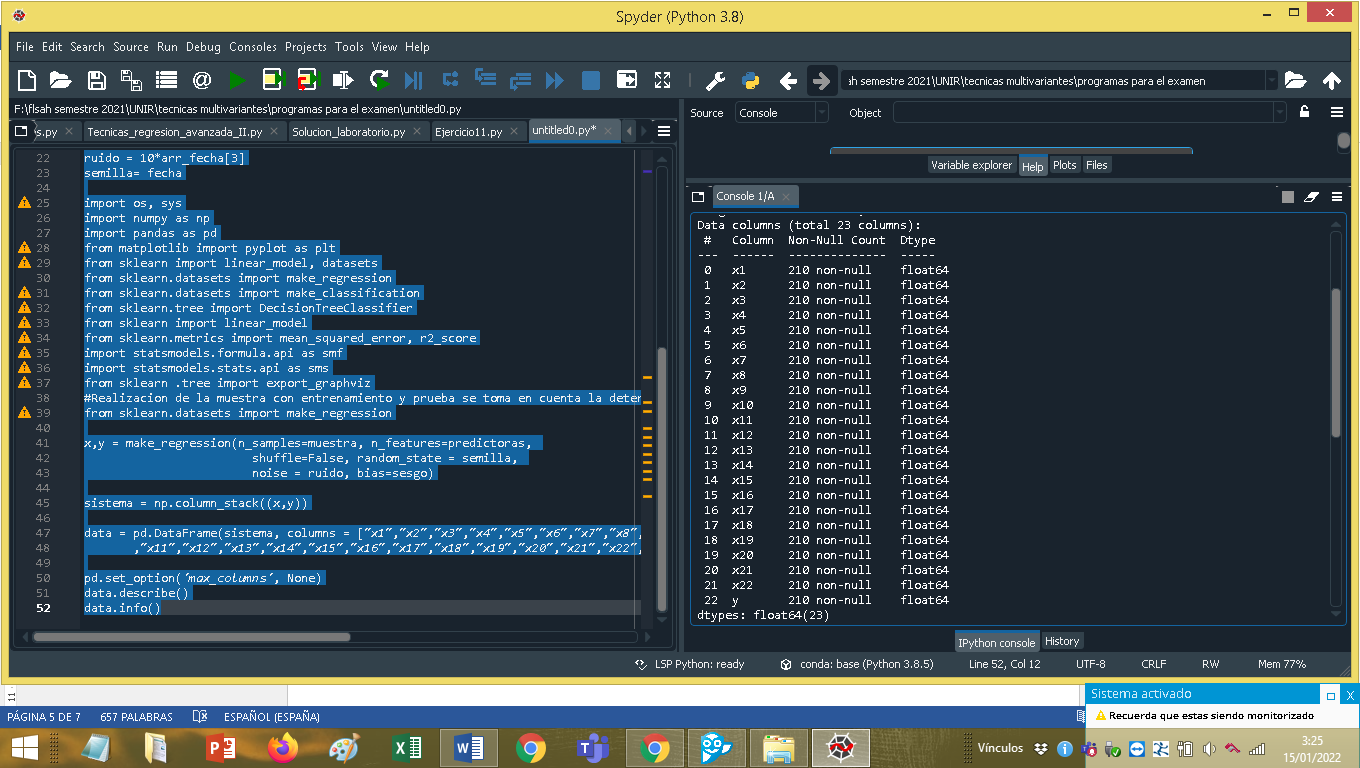
Fecha de nacimiento 07/05/2020

m=0+5+8

d=14 ya que el dia de mi nacimiento es el 07 y esta en el intervalo entre [1,10]

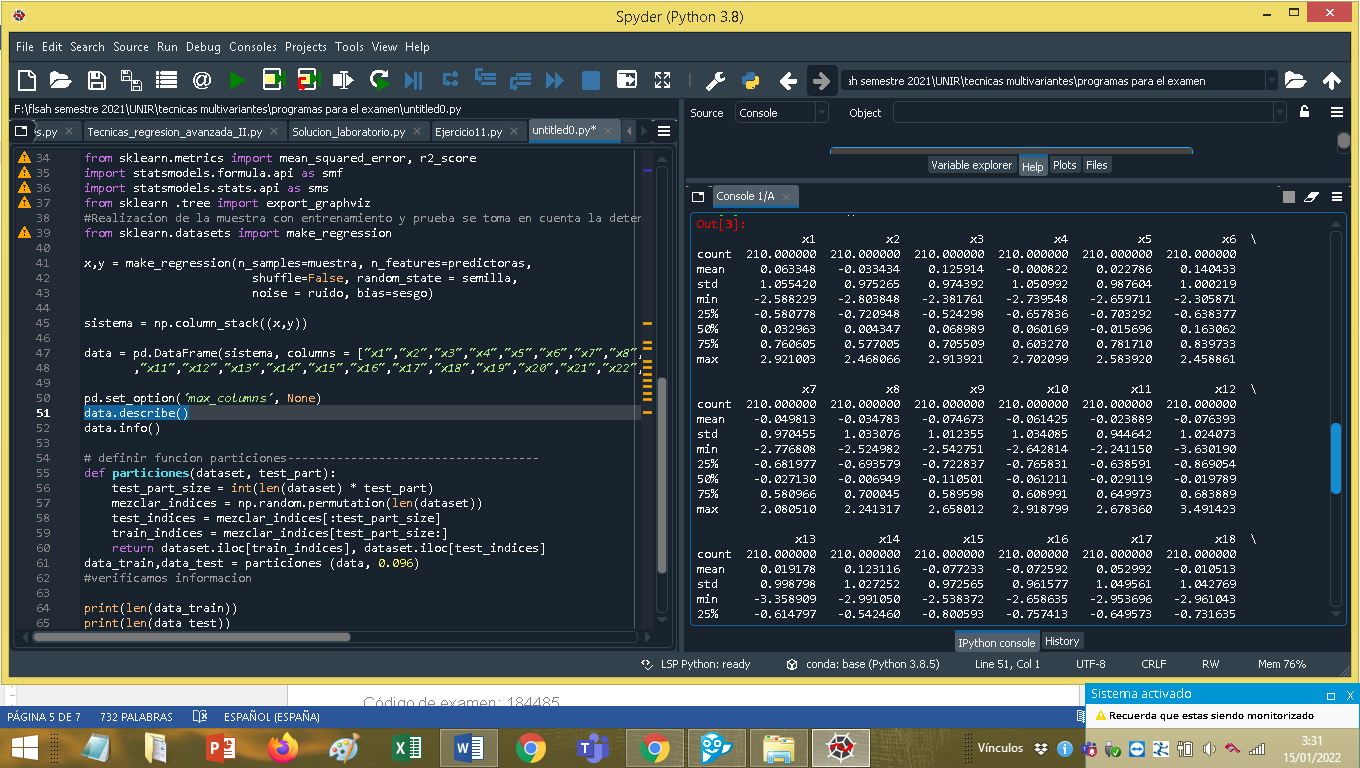
definimos as variables encontradas para posteriormente crear e data set y su respectiva particion de 0,8 y 0,2





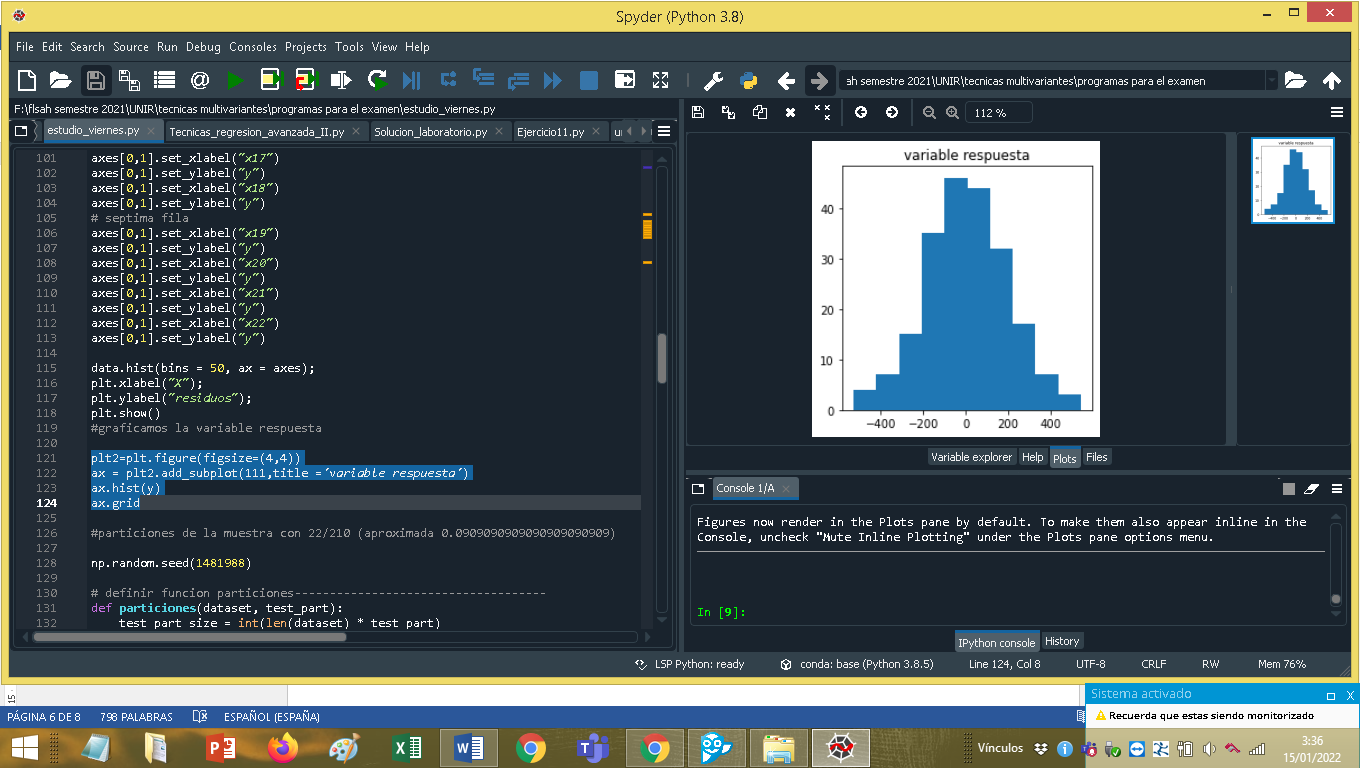
En la figura se puede apreciar las 210 observaciones que tiene cada una de las variables X1……..X22, en e mismo se puede evidenciar que no existen datos faltantes por lo que no havria la necesidad de completarlos mediante la imputación.

Con describe(), podemos identificar las caracteriticas de cada dato como se indica en la siguiente figura.

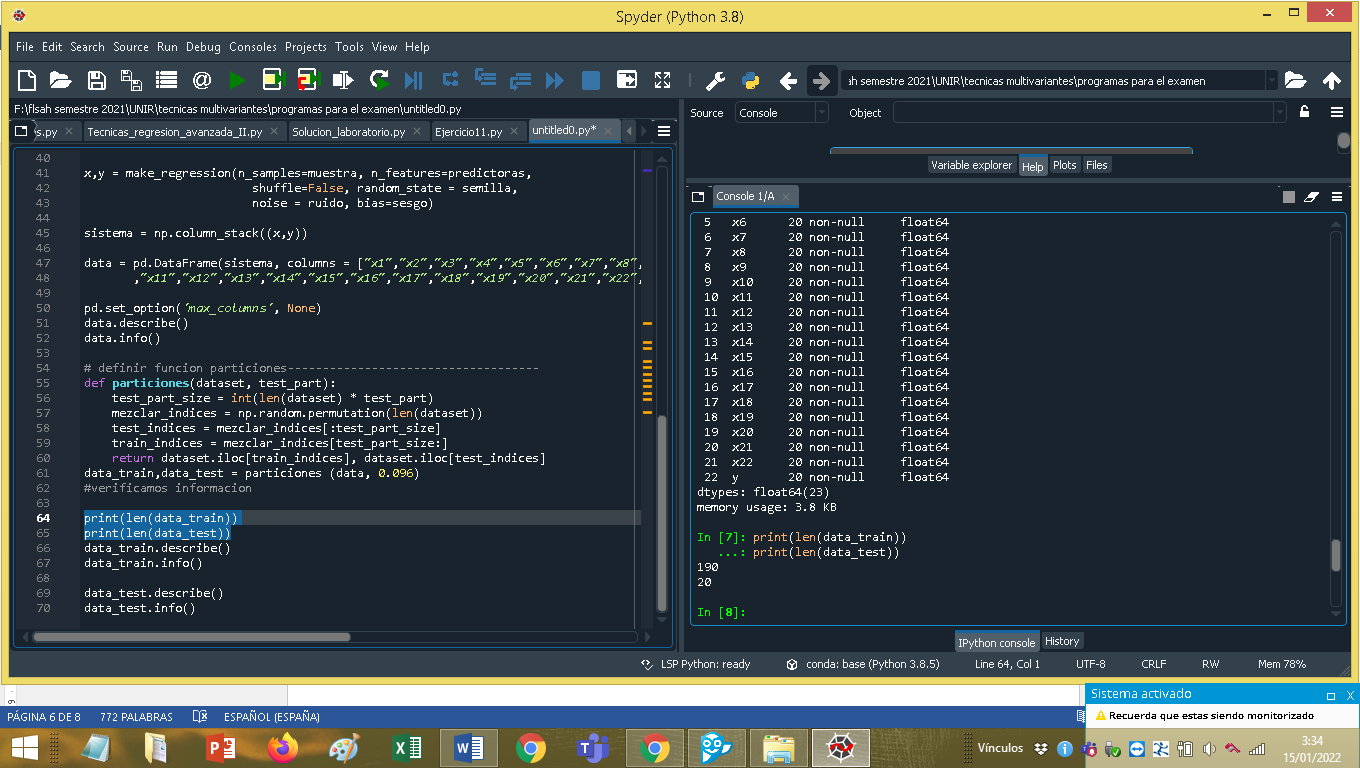


En el grafico anterior se puede observaro los valores caracteriticos de cada una de las variables como la desvicion estandar, la media entre otras

Adicional podemos ver el comportamiento de los datos a traves de la variable respuesta identificando un comprtamiento aceptable.

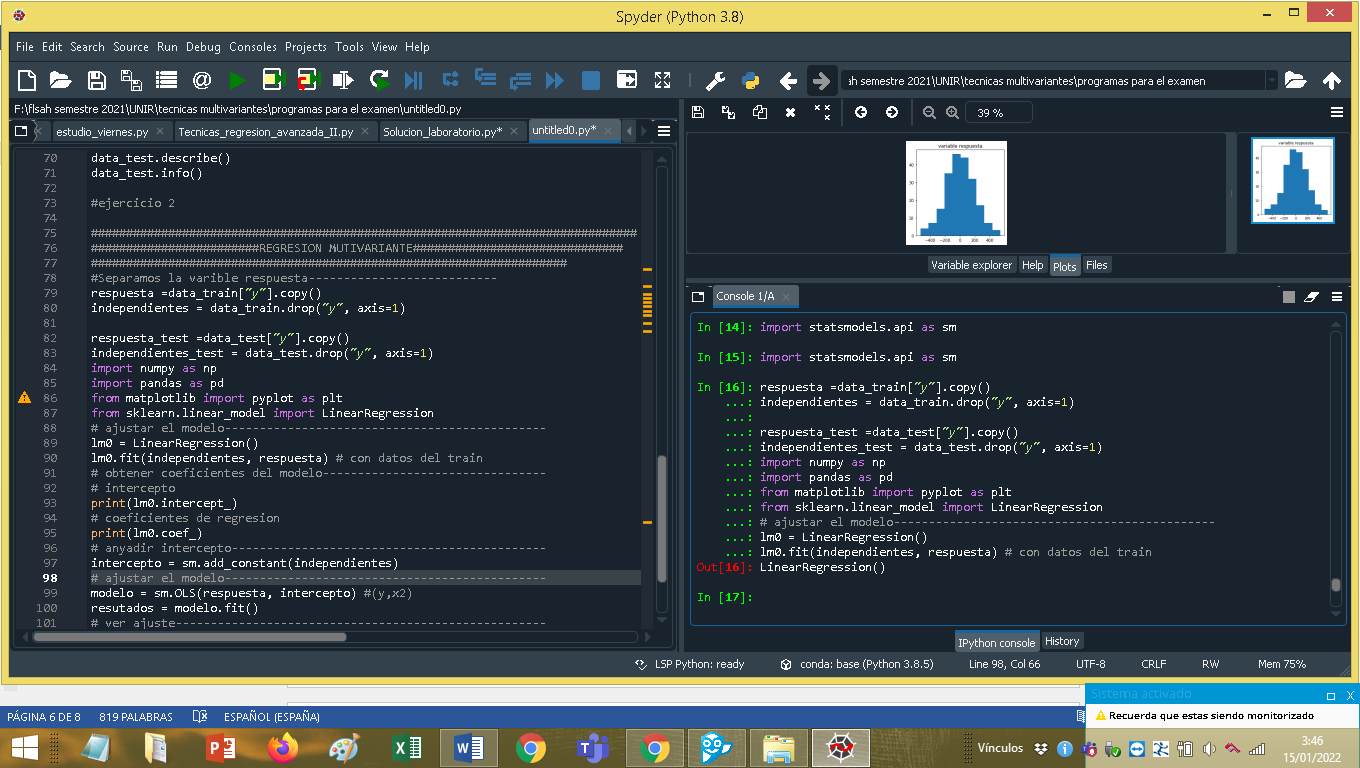


Procedemos a la particion del data set según lo que nos solicita la pregunta1 0,8 para datos de entrenamiento y 0,2 para datos de validacion.



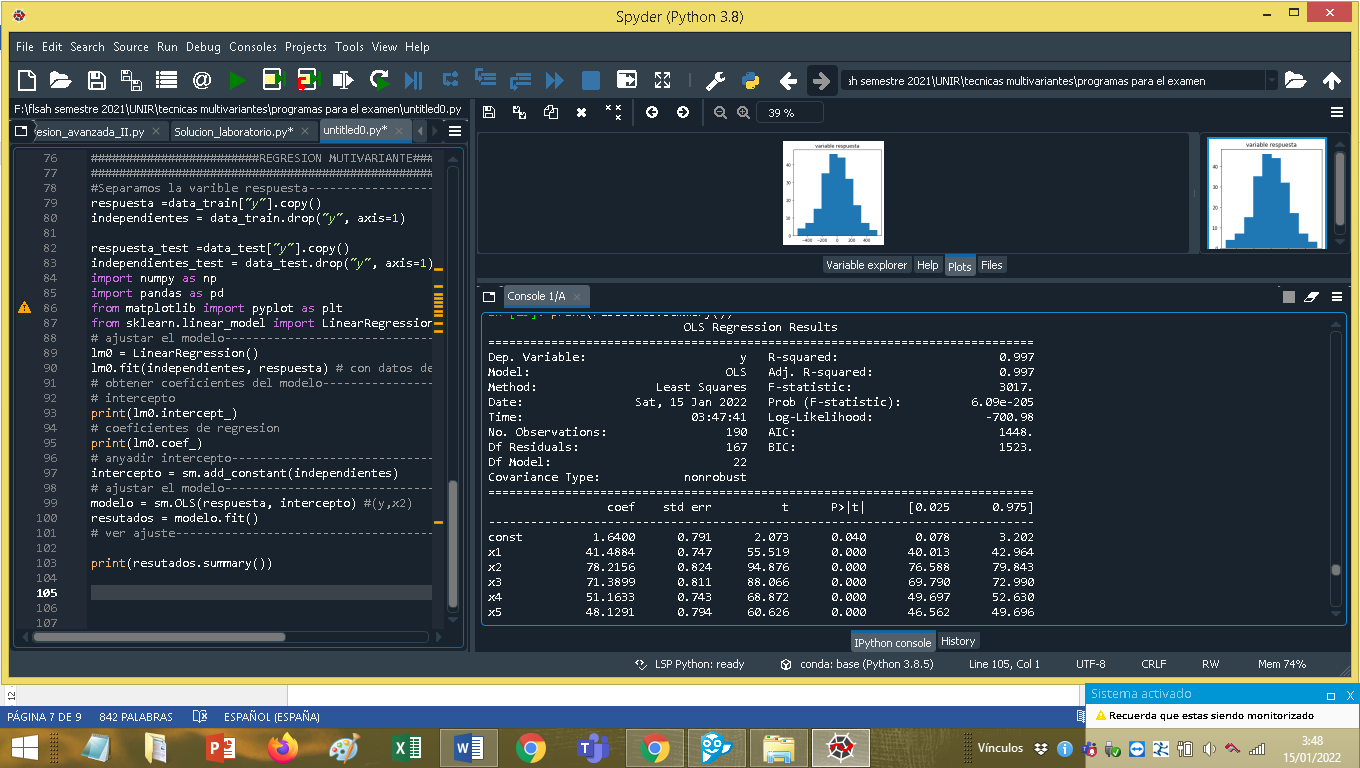
En la anterior imagen se puede obsrvar la particion solicitada por parte del ejercicio en donde se tiene 190 datos para entrenamiento y 20 para test.

**2-3 REGRESION INEAL MULTIVARIANTE YSELECCION DE VARIABLES**



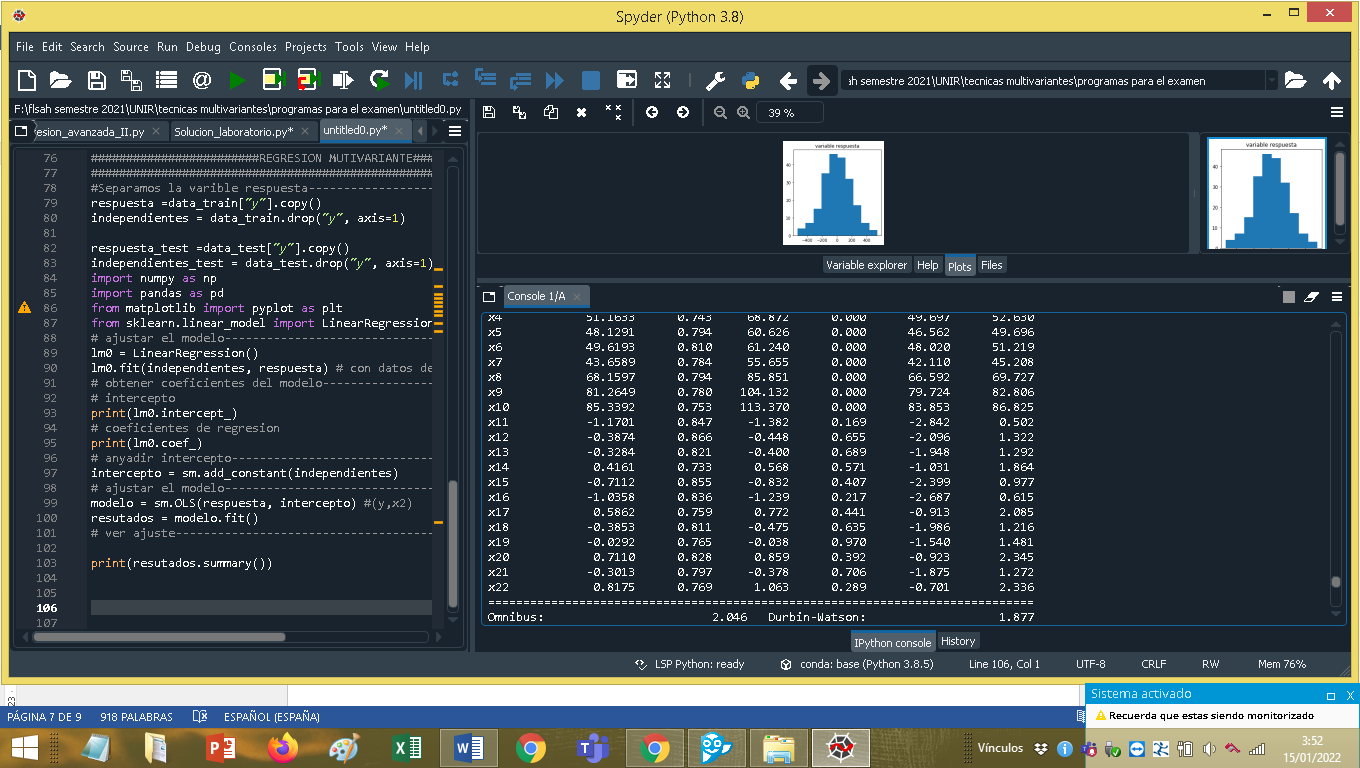
Utilizamos las librerias y programs realizados en clase separando variables, encontranto el intercepto, añadiendo el mismo e imprimos los resultados para su analisis.

RESUTADOS

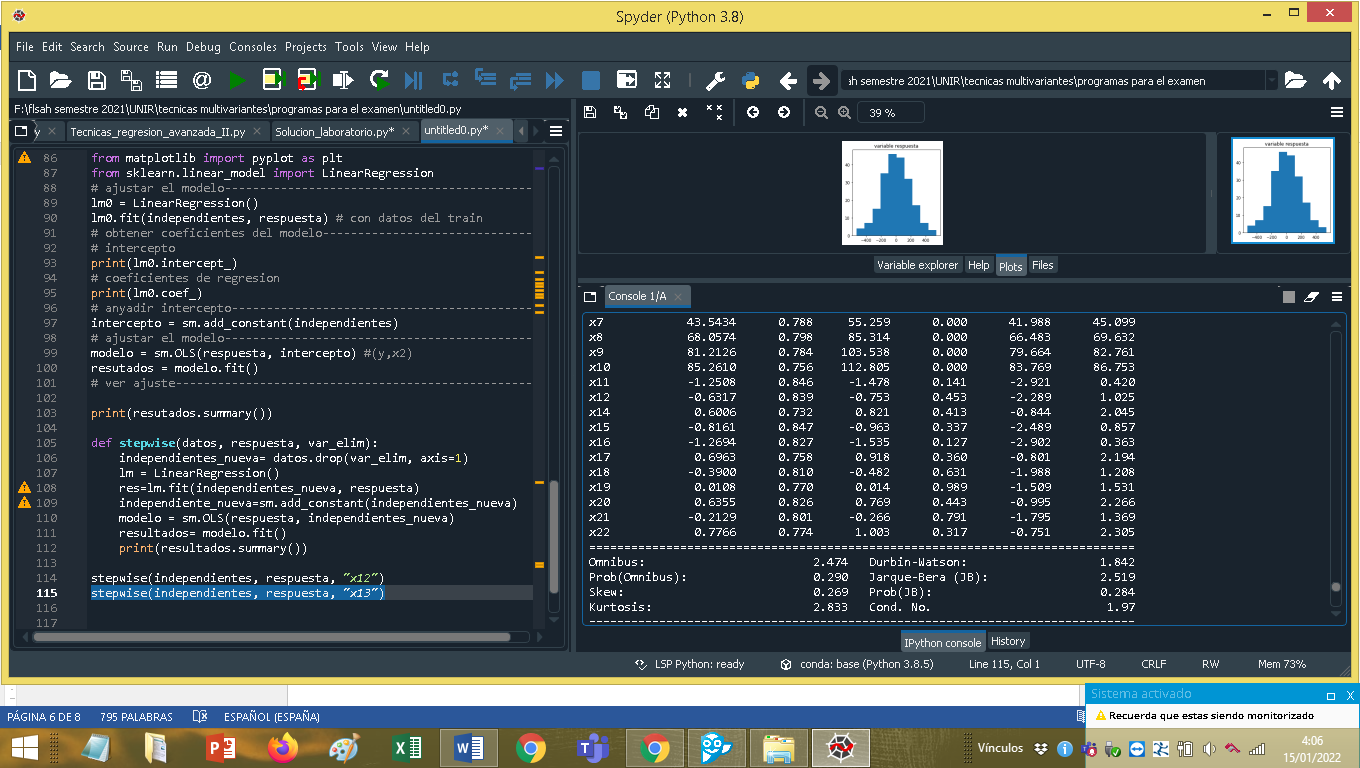


Con este ajuste se `puede observar valores del r^2 y R^2 ajustado bastante considerables e iguales equivalente al 0,997 considerando que estos valores deben estar entre tendiendo a 0 para un ajuste malo y 1 con ajuste perfecto el valor estaria cercano, lo perjudicial es que al tener un exactitud casi perfecta tiene una tendencia a un sobreaste.

Los p valores son muy elvados en una gran cantidad conmo se ve en la siguinte figura.



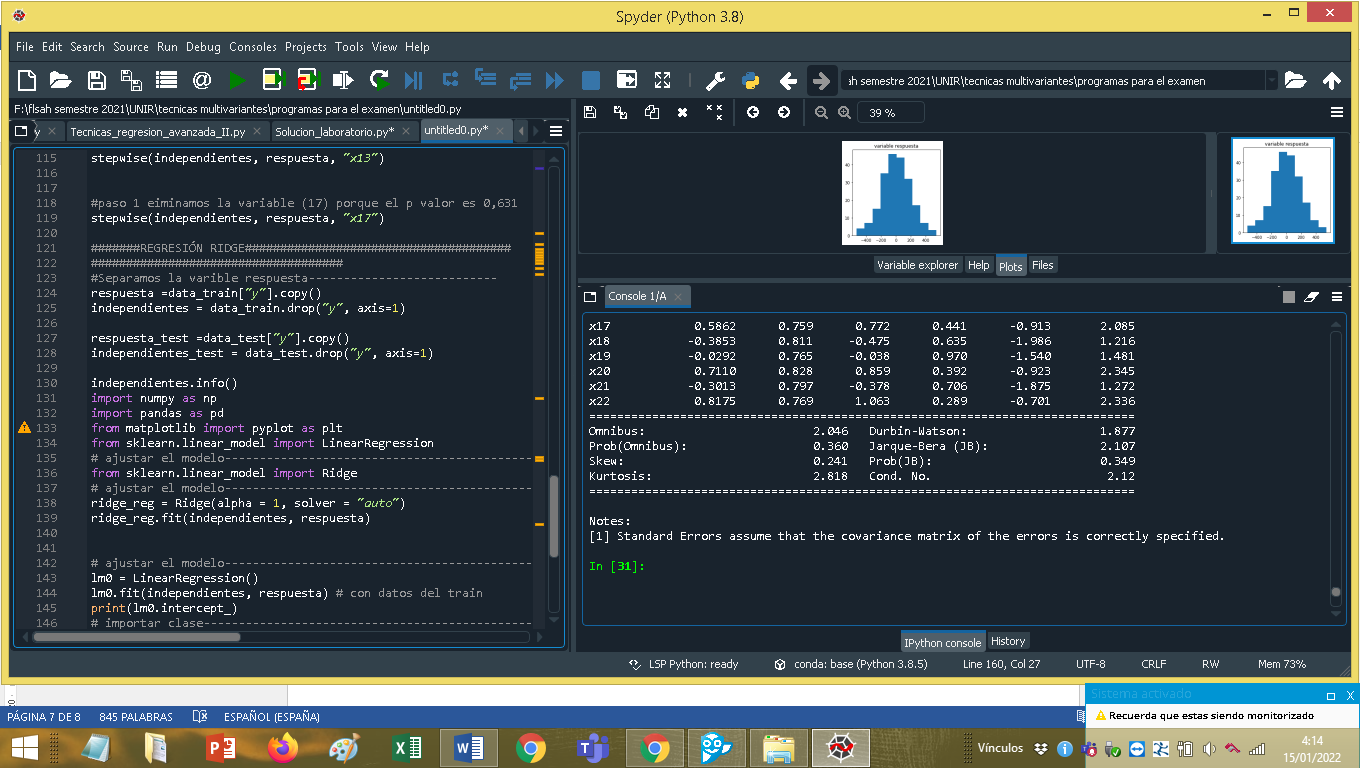
Eliminamos algunos valores con spetweise y contratar los reultados para ejemplo eliminamos la varible con un p valor muy elevado la x13 con 0,6



Los r2 siguen siendo iguales.

4. REGRESION CON LE METODO DE RIDGE

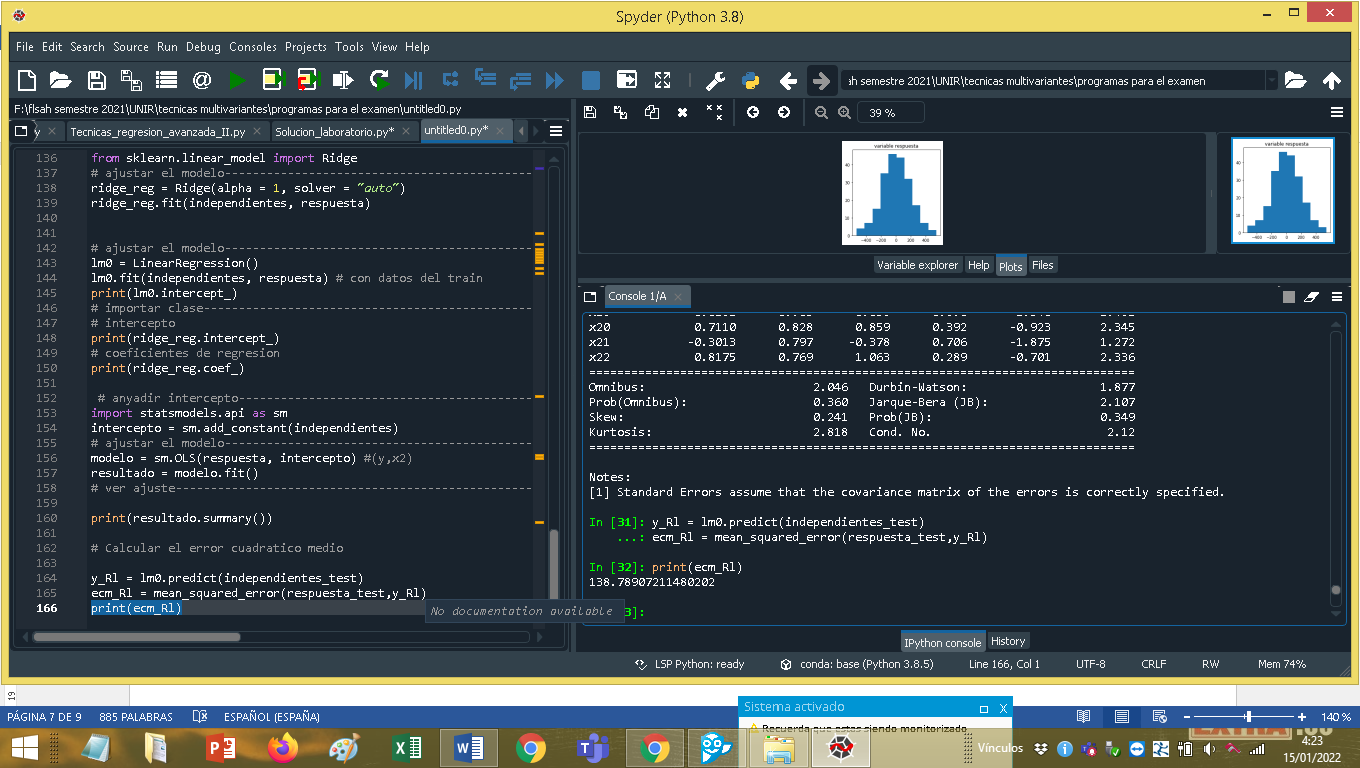
Con el metodo de ridge los valores del r2 y r2 ajustado se mantiene ademas los p valor del conjunto de datos sigue siendo muy a partir del x11 es asi que en el x19 tiende a 1.



5.ANALISIS DEL MEJOR METODO EN BASE A LOS UTILIZADOS.

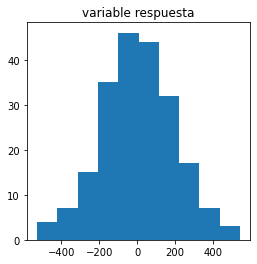
En comparacion con el valor del r^2 los metodos tienen un valor muy sililar de 0,997 por lo que realiza un analisis en funcion del error cuadratico medio.

Para el metodo RIDGE



Podemos abservar valor del eror cuadratico medio de 138.78907211480202

Se procede de la misma maner a con el resto de metodos obteniedo valores muy equivalentes según e critrio que puedo dar es que la tendencia de los datos tiende a tener una curva simetrica ya que la variable respuesta se tiene



El comportamiento de los p2 valores es muy similar exixte valores que se diparan desde el x=11 por lo elegir un metodo adecuador seria complicado, personalmente en relacion de esta anomalias probablemente elijiera el lineal multivariante por su coste.

PREGUNTA DE TEORIA

ARBOL DE DECISION

Un arbol de decision es aquel que me permite clasificar un conjunto de datos en relacion a tamaño, color o amplitud del petalo o tallo, generalmente un arbol de decisión es aquel que esta formado por niveles

En donde hay la pregunta de clasificacion, la medida de impurezas , numero de observaciones en cada no numero de cada tipo y su clsificaion, lo principal caracteristica deseable para un clasificador mediante el arbol de decisión son las Gn (medidas de impuezas ) que lo idoneo tengan un valor de gini=0 indicando una pureza en la clasificacion, una de las desventajas es que sus capacidades de ajuste y prediccion son peores que otros metodos.