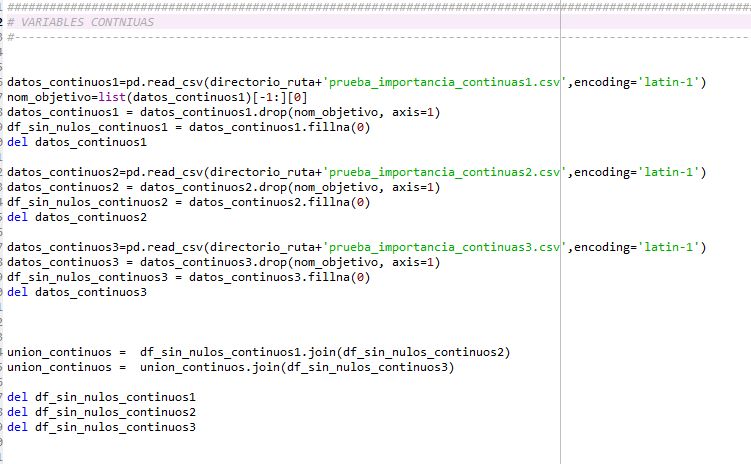
Explicacion del algoritmo

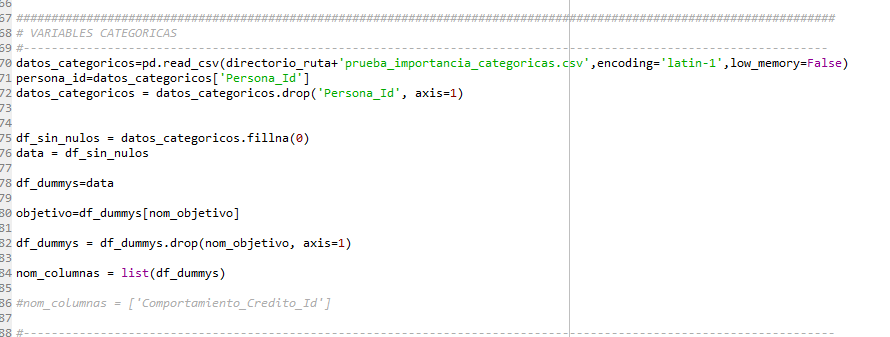
Como origen tomamos los resultados de los csv de la ruta “cvs para algoritmo.str”

Tenemos 4 archivos, 3 para continuas y 1 para las categóricas. Tuvimos que separarlas de este modo ya que rompia por memoria.

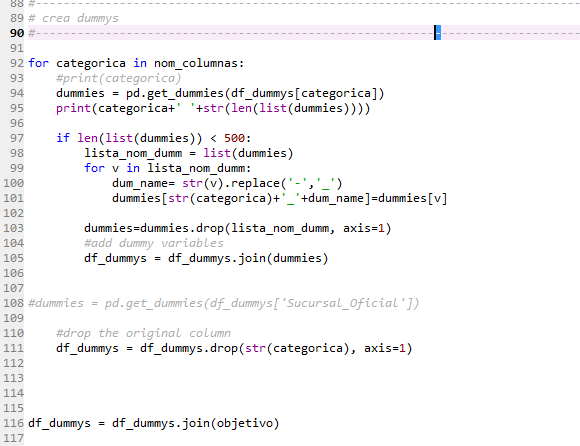
1. Une datasets de variables continuas en uno solo, luego elimna los datasets intermedio para hacer espacio en memoria.



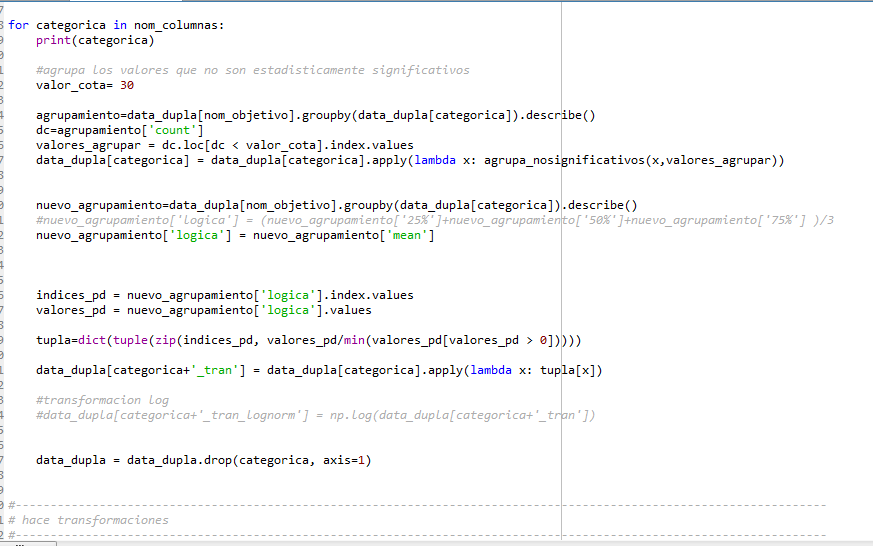
1. Levanta archivo de categóricas



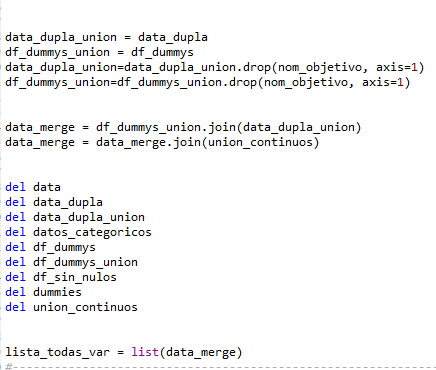
1. Crea variables dummies para cada variable categorica



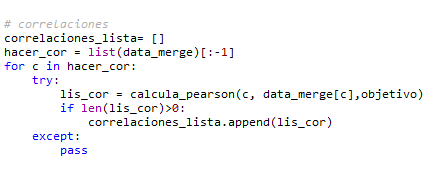
1. Por cada variable cateogica se le aplica una transformación de medias. La idea es que por cada valor categorica se le asigne la media respecto la variable objetivo. Se agrupan las observaciones menores a 30 en una bolsa comunitaria. Una vez con este agrupamiento por cada categorica estadísticamente significativa calculamos la media, luego buscamos de todas estas medias mayores a 0, la menor. Paso posterior, dividimos cada media por este valor minimo para que cada categoría ahora represente un incremental de la media respecto a la minima encontrada. De este modo la nueva variable puede ajustarse mejor a la regresión.



1. Unimos todas las variables ahora quedando un dataset de pura variables continuas

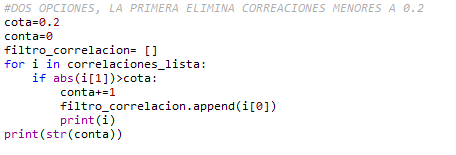


1. Se hace la correlacion de cada vasriable con la variable objetvo



1. Ahora tenemos que filtrar las variables con mayor relevancia. Tenemos dos opciones, filtrar aquellas que cumplan alguna cota.

En la imagen siguiente se puso una cota de 0.2 para eliminar aquellas variables que queden por debajo de esta correlacion (valor absoluto)



La segunda opción es analizar en un Excel todas las variables y asignar un filtro manual luego de analizarlo. Este es el mas adecuado si se tiene conocimiento del negocio



Hay que elegir uno de estos dos métodos, una vez asignado elfiltro, filtramos.

1. Nos quedamos con un conjunto de variables reducidos



1. Con la lista de variables candidatas lo que hacemos ahora es buscar la correlacion de cada variable respecto de todas las otras en el prefiltro, de esta manera buscamos aquellas variables que estén correlacionadas entre si, si vemos que alguna variable esta correlacionada por mas de 0.45, eliminamos alguna de las dos variables, en este caso, eliminas la variable menos correlacionada con el variable objetivo.
2. Guardamos dataset con las variables candidatas

