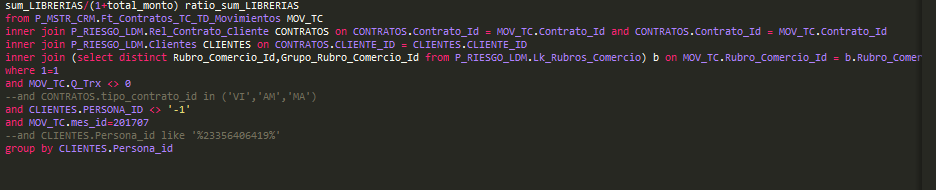
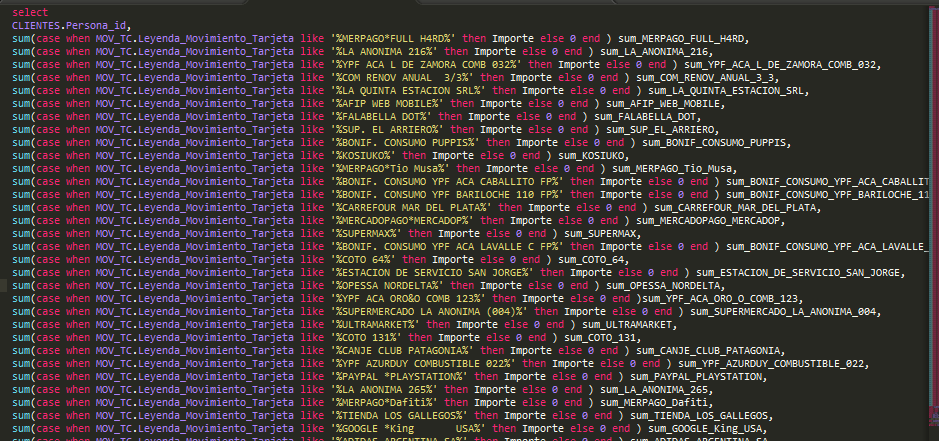
Se analiza la variable de tipo de consumo por tarjeta de debito y crédito para ver como se correlaciona con el sueldo

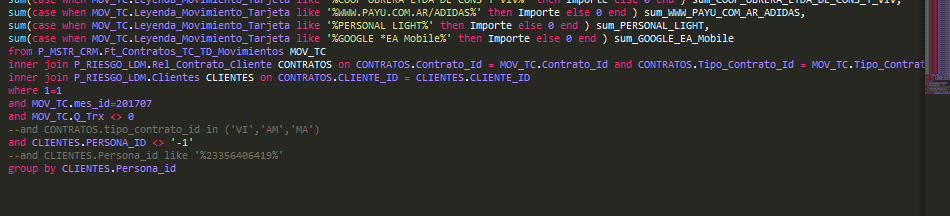
En

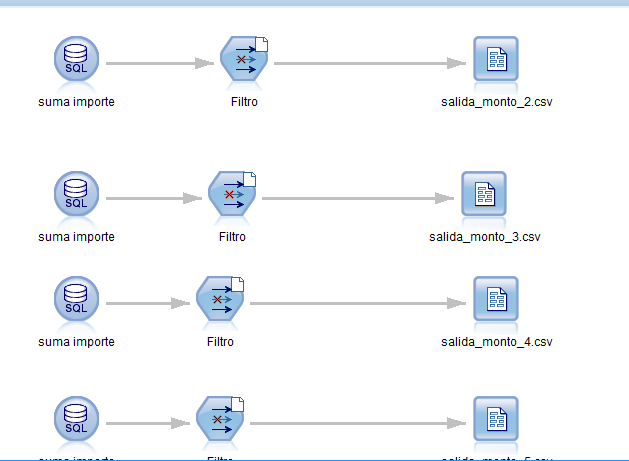
[\\srappw001v011.terceros.banco.com.ar\Disco D\COBRANZAS\prueba\_claiun\NSE\Desagregado\_tipo\_consumo\Ruta2.str](file:///\\srappw001v011.terceros.banco.com.ar\Disco%20D\COBRANZAS\prueba_claiun\NSE\Desagregado_tipo_consumo\Ruta2.str)

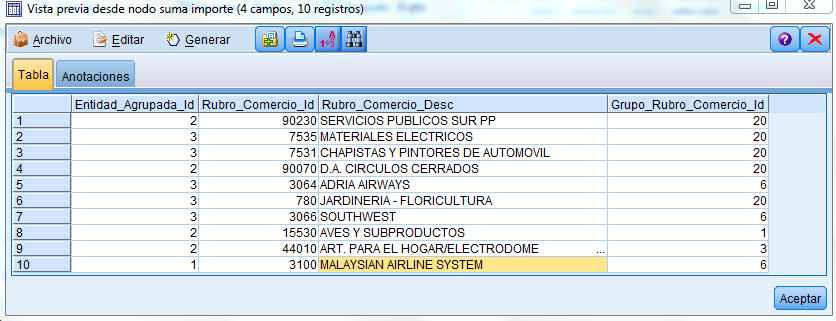
Se corren las queries para generar flags de tipo de consumo y también los agrupados







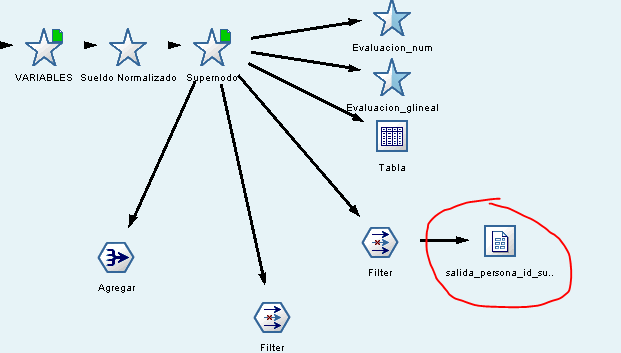




En

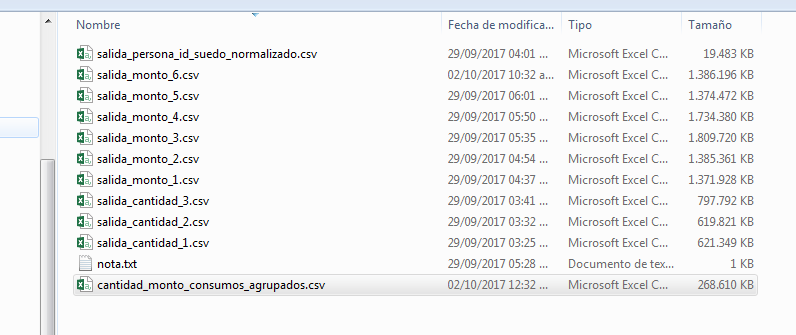
[\\srappw001v011.terceros.banco.com.ar\Disco D\COBRANZAS\prueba\_claiun\NSE\NSE 2017\PREP\_NSE-E\_analisis\_variables\_v13.str](file:///\\srappw001v011.terceros.banco.com.ar\Disco%20D\COBRANZAS\prueba_claiun\NSE\NSE%202017\PREP_NSE-E_analisis_variables_v13.str)

Esta la corrida del se calcula el sueldo normalizados para estas personas en julio y se guarda por persona id el sueldo en un csv



Aca están todas las salidas

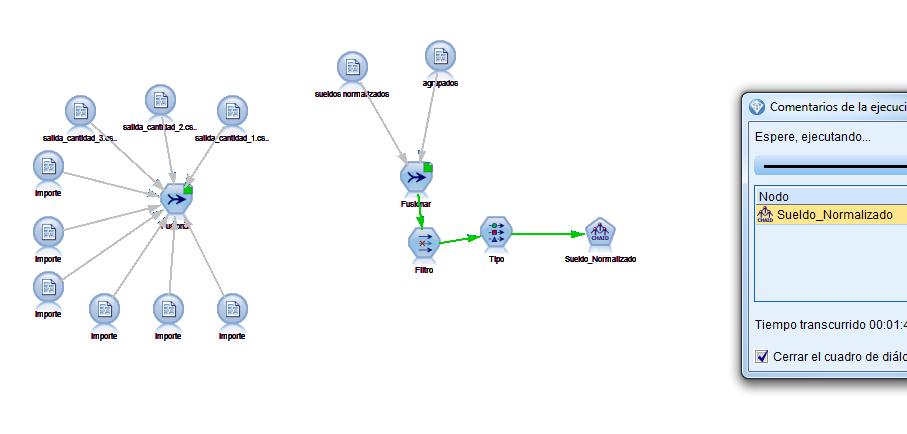
\\srappw001v011.terceros.banco.com.ar\Disco D\COBRANZAS\prueba\_claiun\NSE\Desagregado\_tipo\_consumo



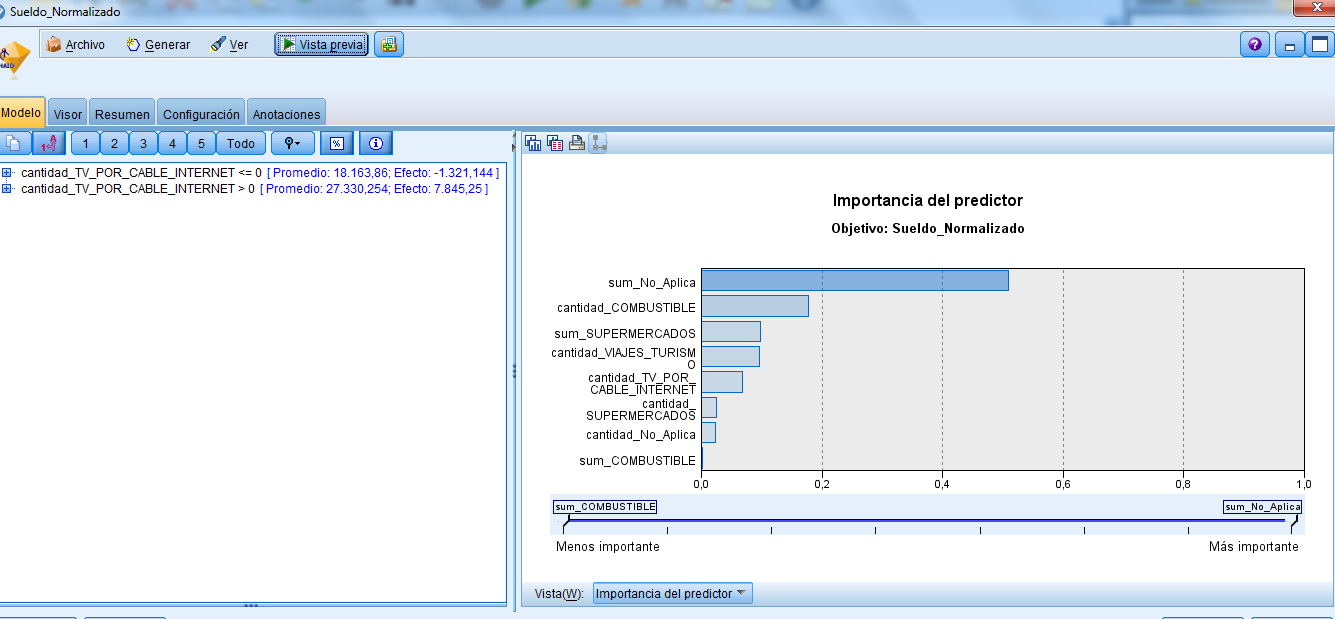
en

C:\Users\LAC40641\Documents\0-Proyectos\Avances\NSE\Desagregado\_tipo\_consumo\join\_csv\_evaluacion.str

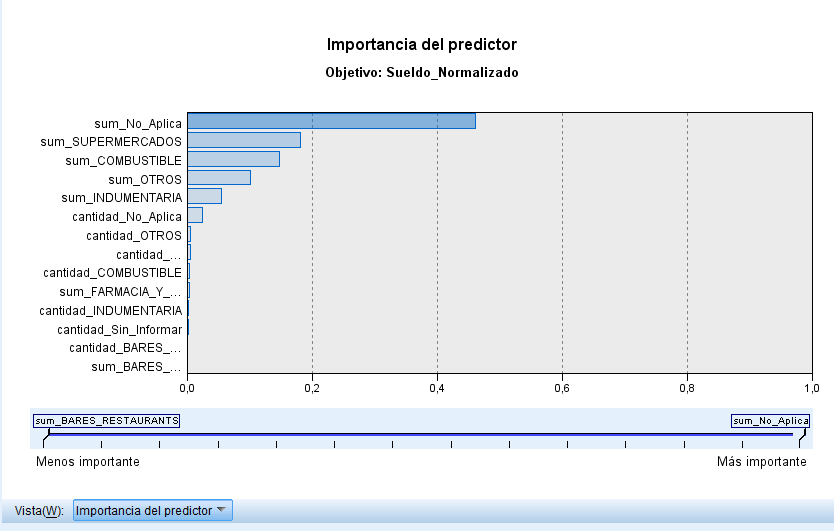
Están los joins para analizar las variables y su correlacion



Tirando árbol chaid solo con variables agrupadas, profunidad 5

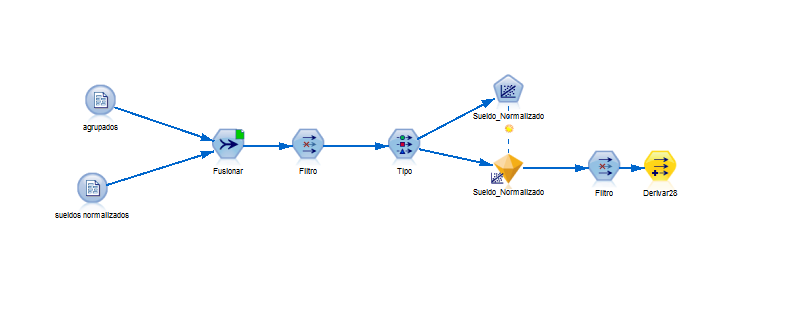


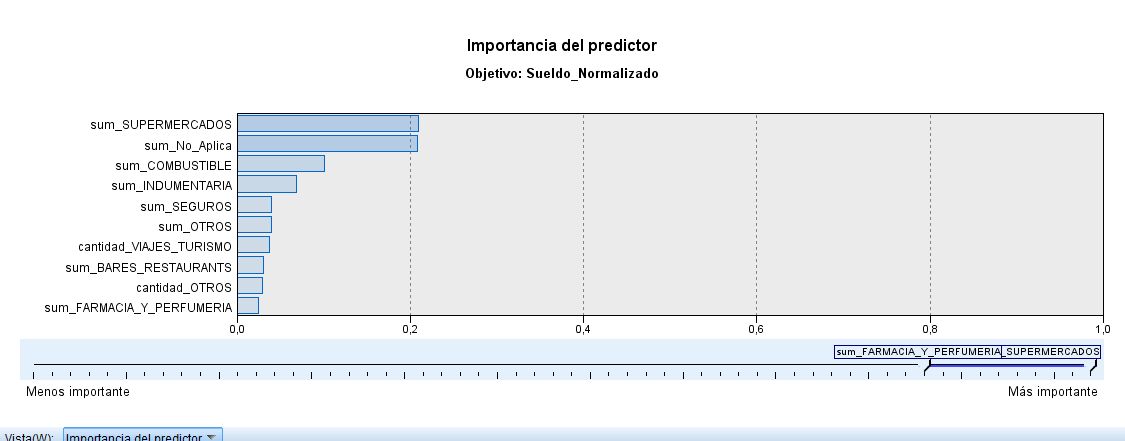
Profundidad 10

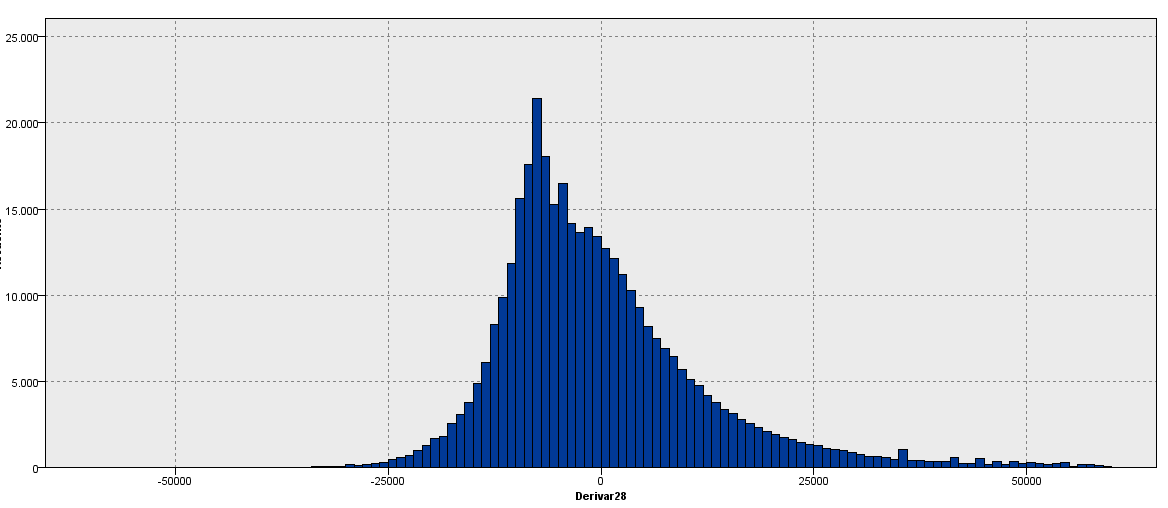


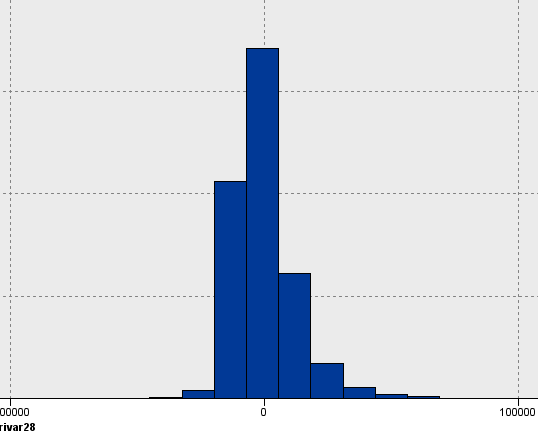
Probamos con una regresión en

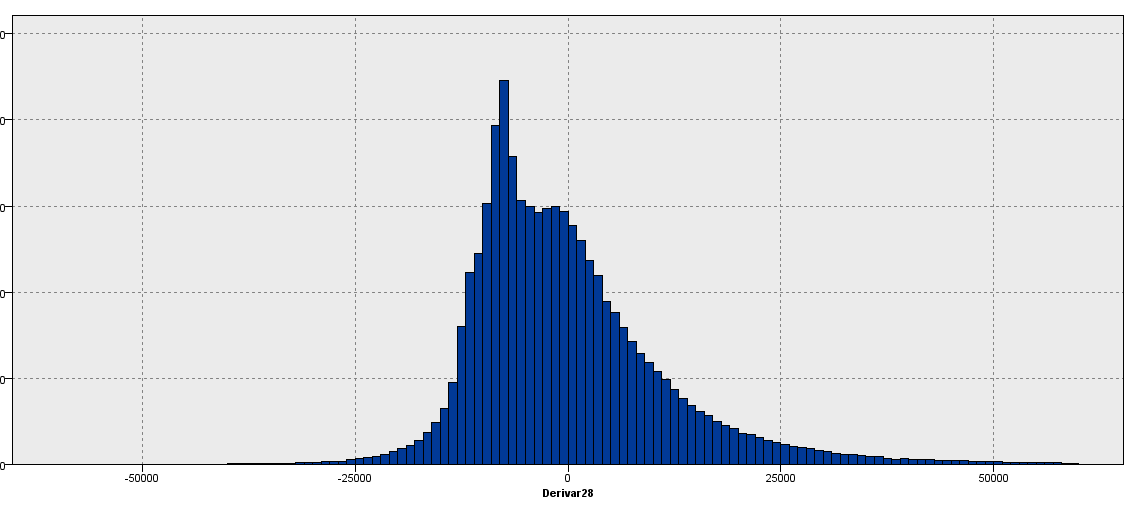
C:\Users\LAC40641\Documents\0-Proyectos\Avances\NSE\Desagregado\_tipo\_consumo\prueba\_reg\_lineal\_agrupados.str



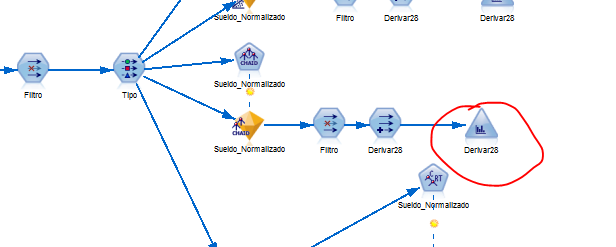








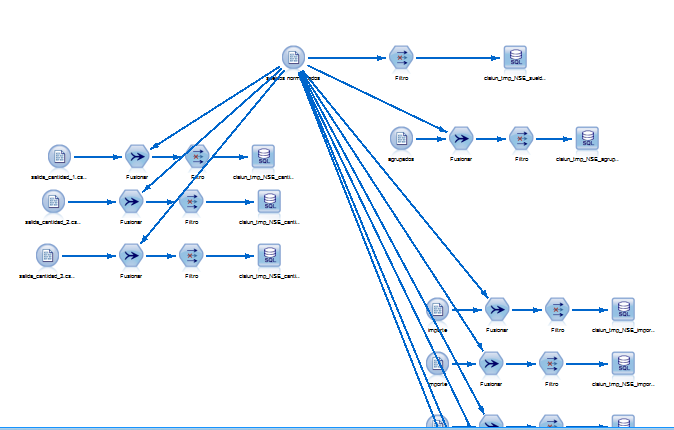
Árbol



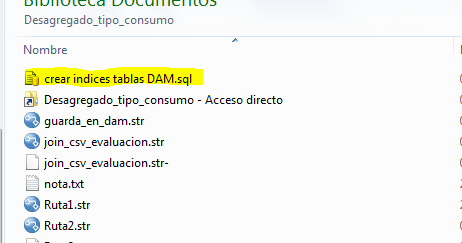
Guada en DAM

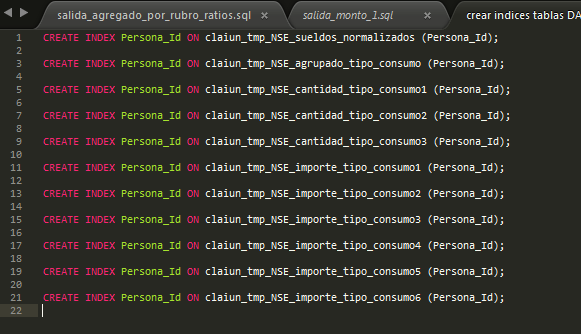
En

C:\Users\LAC40641\Documents\0-Proyectos\Avances\NSE\Desagregado\_tipo\_consumo\guarda\_en\_dam.str



Guardo en sql las tablas para luego joinear mas rápidamente ahí, ya que se crean índices por tabla



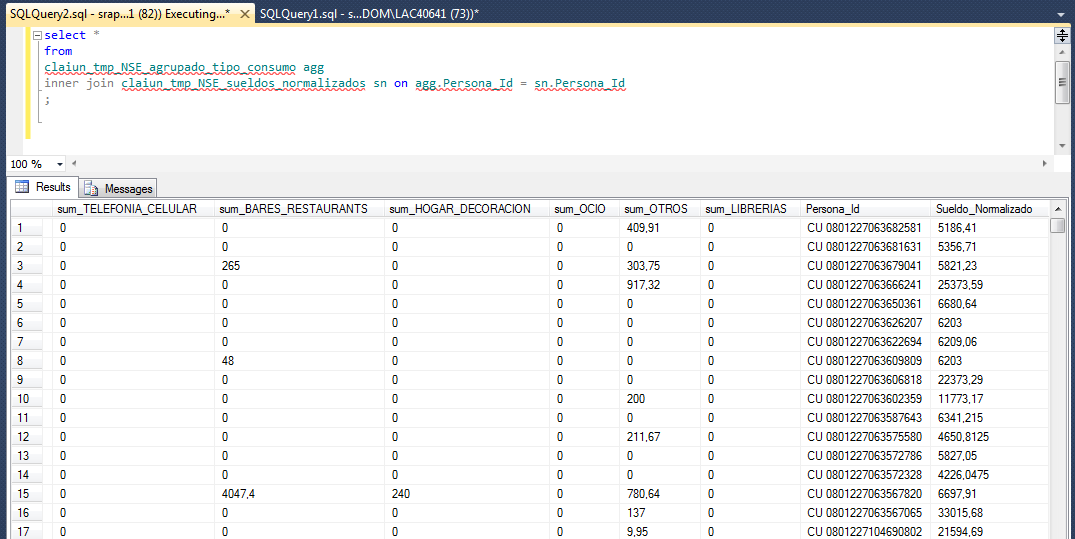


Exploro correlacion de algunas variables con el sueldo normalizado para ver la importancia y que enfecto tienen las medias por ejemplo.

En

[\\srappw001v011.terceros.banco.com.ar\Disco D\COBRANZAS\prueba\_claiun\NSE\Desagregado\_tipo\_consumo\analisis\_python.py](file:///\\srappw001v011.terceros.banco.com.ar\Disco%20D\COBRANZAS\prueba_claiun\NSE\Desagregado_tipo_consumo\analisis_python.py)

Analizamos por correlaciones esta consulta en Python

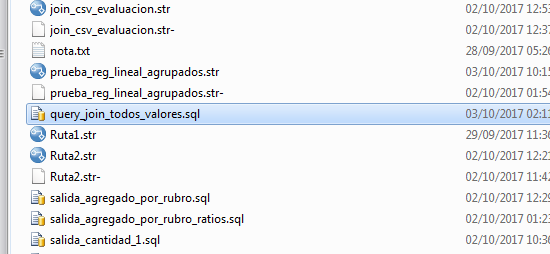


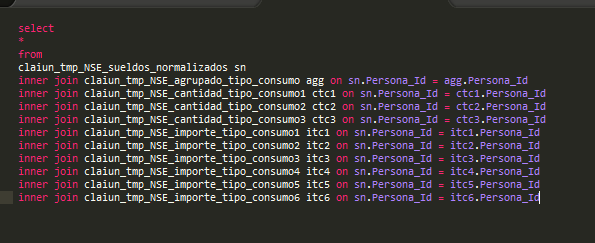
Correlaciones y boxplots

En

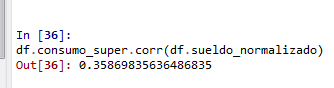
C:\Users\LAC40641\Documents\0-Proyectos\Avances\NSE\Desagregado\_tipo\_consumo\query\_join\_todos\_valores.sql

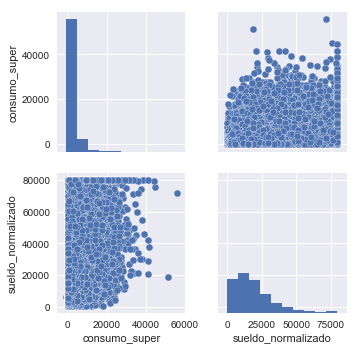
Esta el join





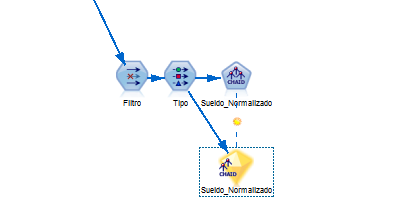


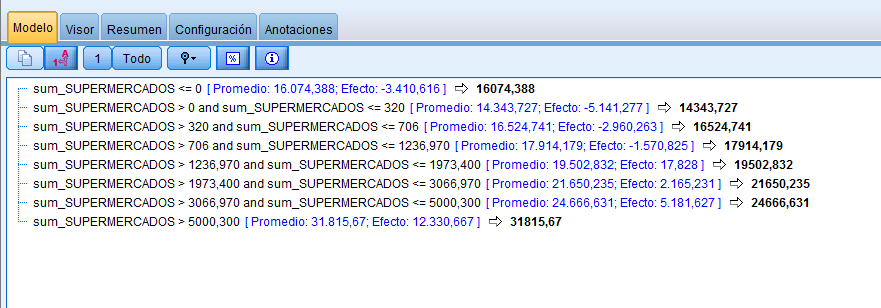




Esta poco correlacionado

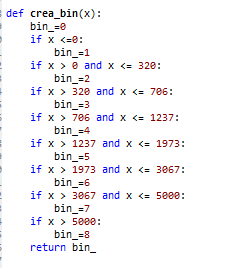
Aplico árbol para que haga bines y asi ver en un box plot separado por bines como varia la media de cada variable, si hay diferencias de medias o algún estadístico aplicarlo para ver como esa variable tiene o no peso

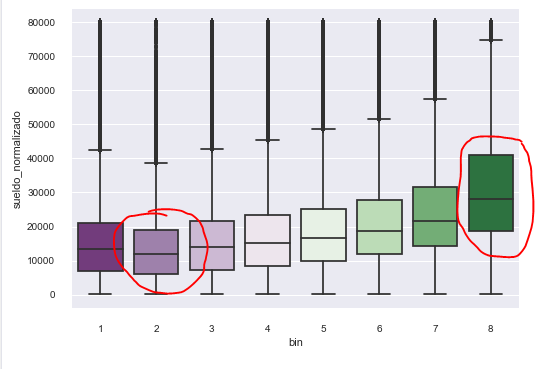




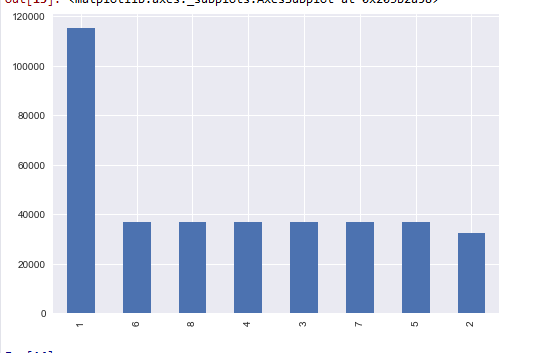
Para esta variable no hay correlacion significativa pero si cortamos en distintos intervaloes hay diferencias de medias entre los bines encontrados.

Probando con intervalos



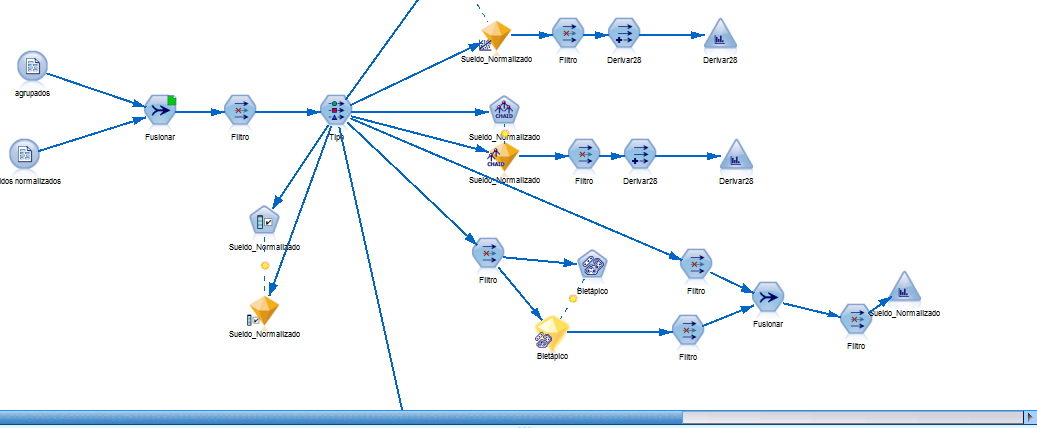


Hay diferencias de medias por lo tanto esta variable por más que no tenga correlación puede separar poblaciones y de hecho lo hace ya que es una variable importante tanto en el árbol como en la regresión

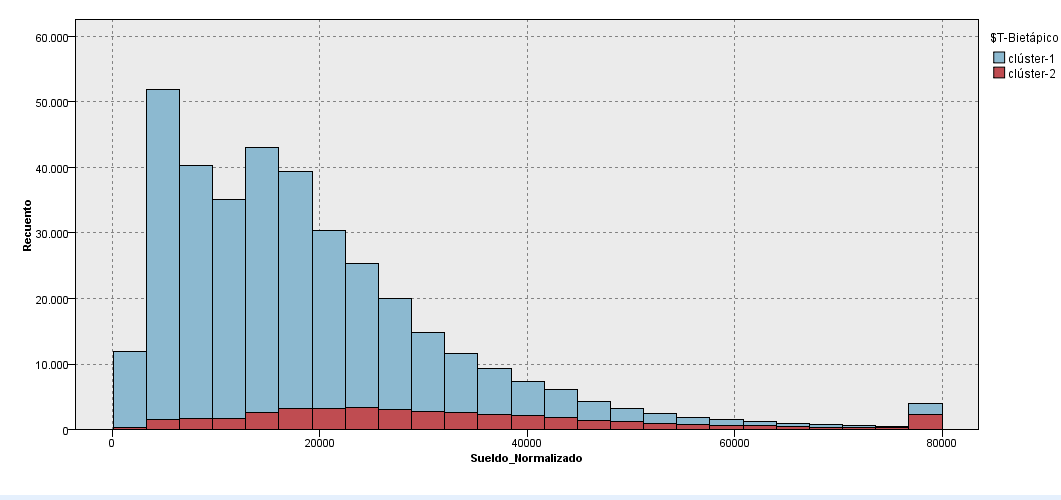


Tiro modelo con todas las variables de los consumos para ver como afectan

Investigamos efecto de algún cluster con estas variables agrupadas

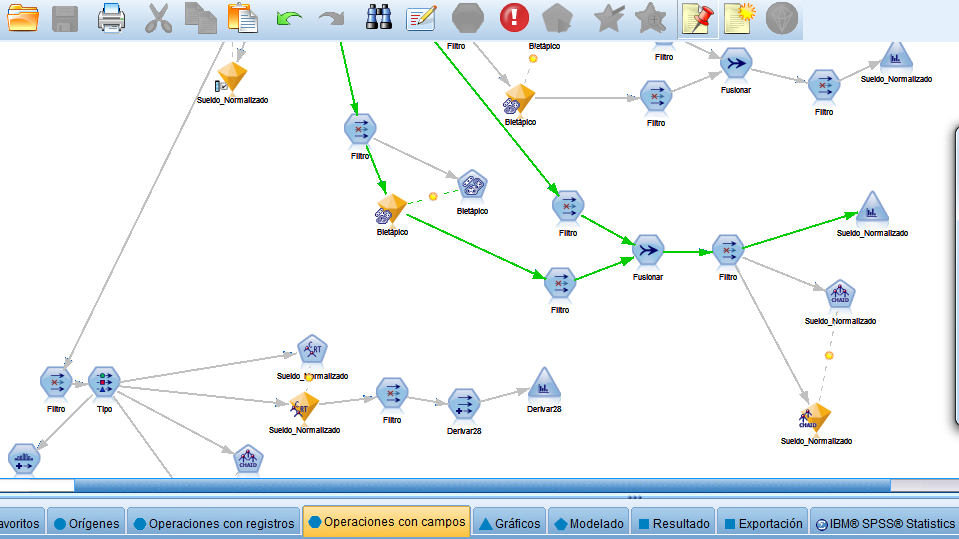


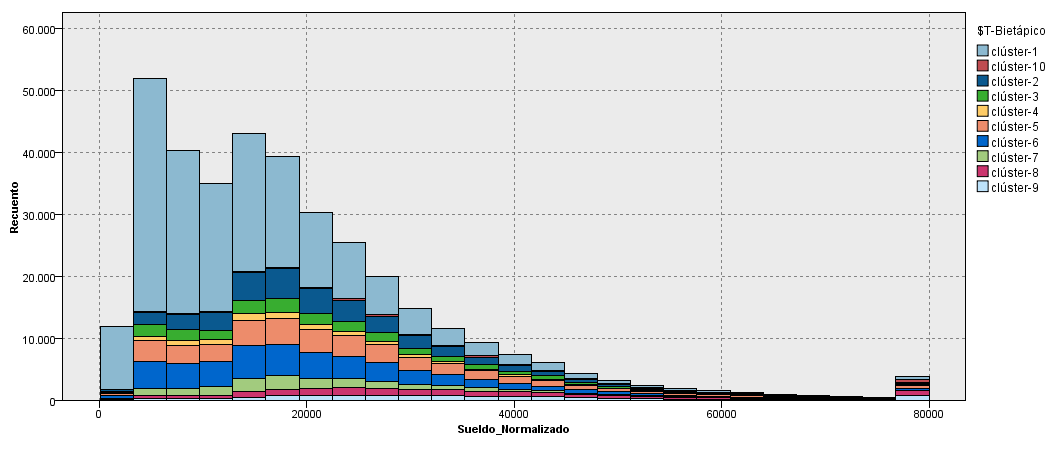
El algoritmo nos da dos clusters



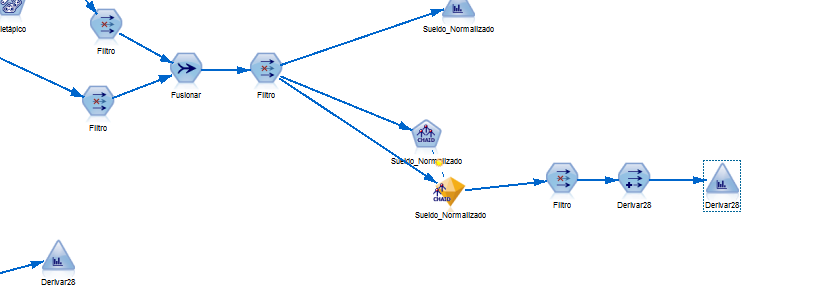
Los rojos parecieran tener un mejor nivel de vida observando los consumos, vemos los sueldos según el cluster

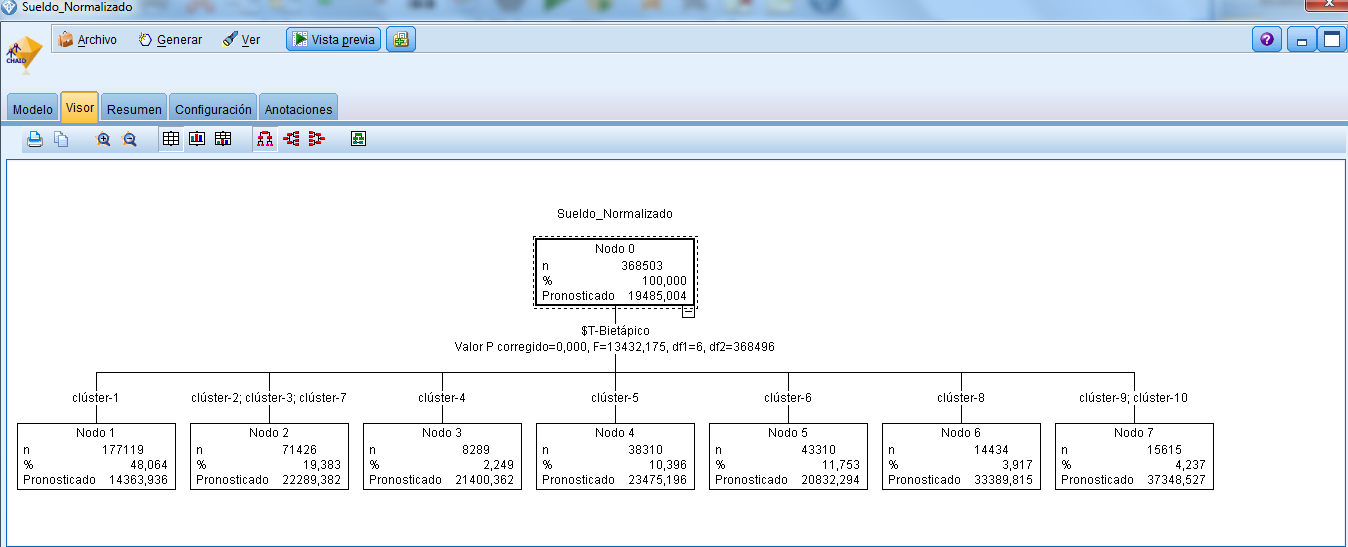
Ahora aplico 10 clusters y tiro un chaid a ver que dice como separa

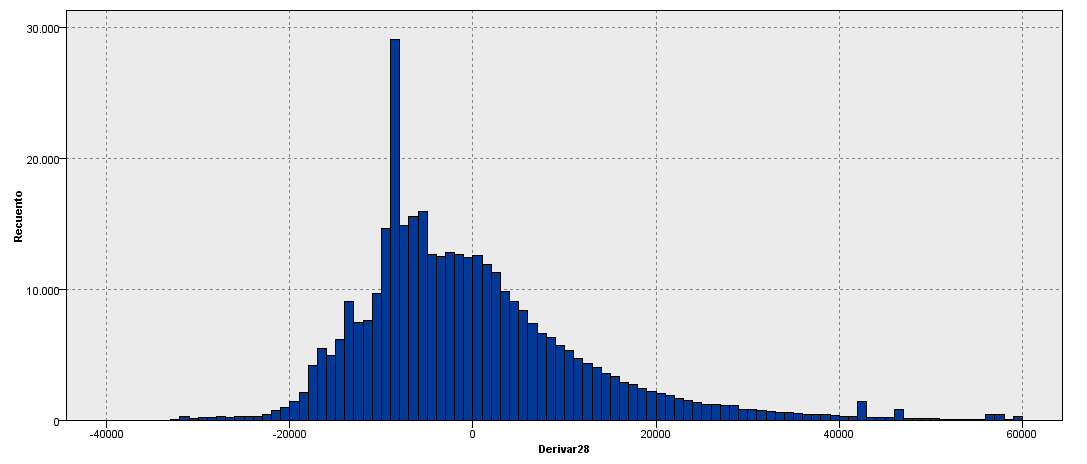




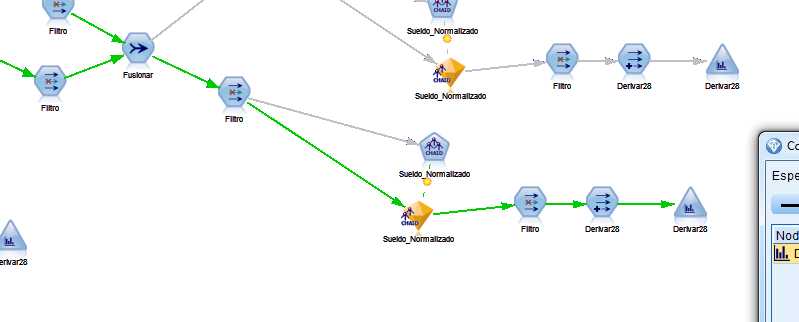
Performance solo con clusters

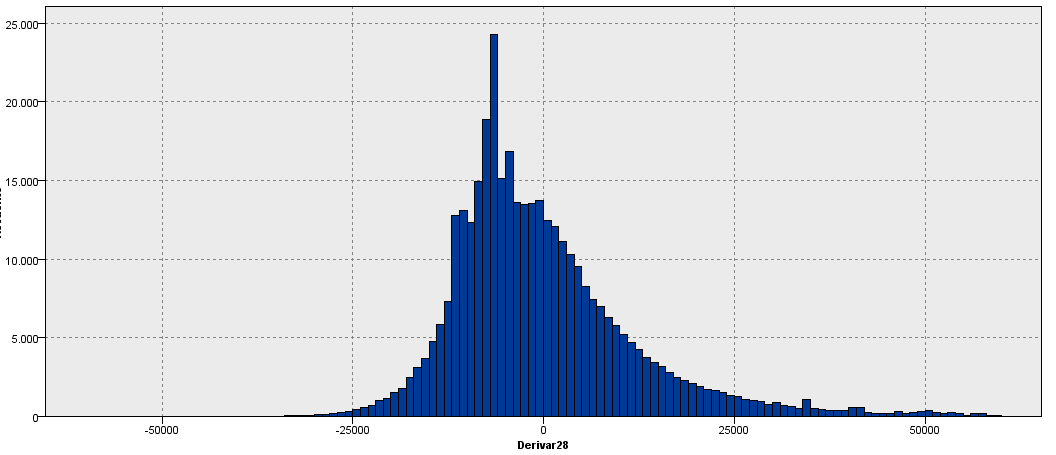




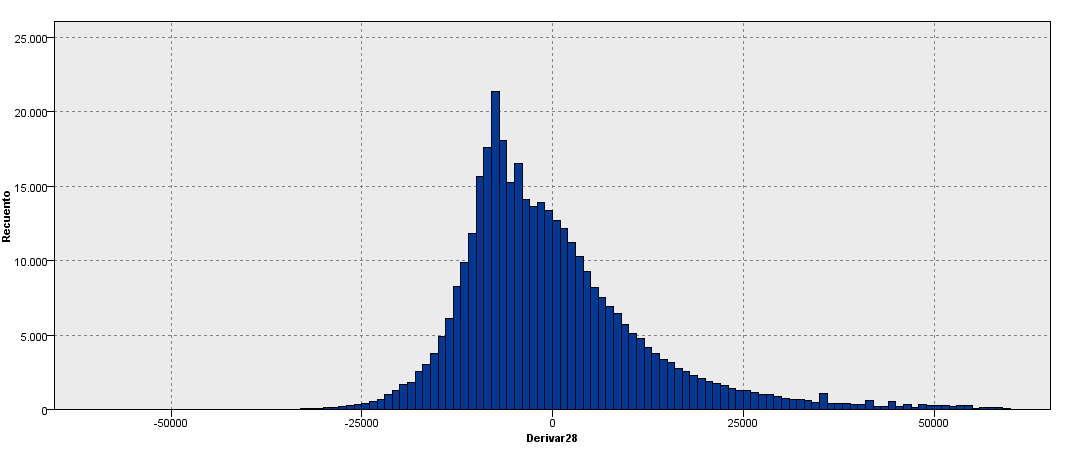


Pruebo con todas las variables mas el clusters

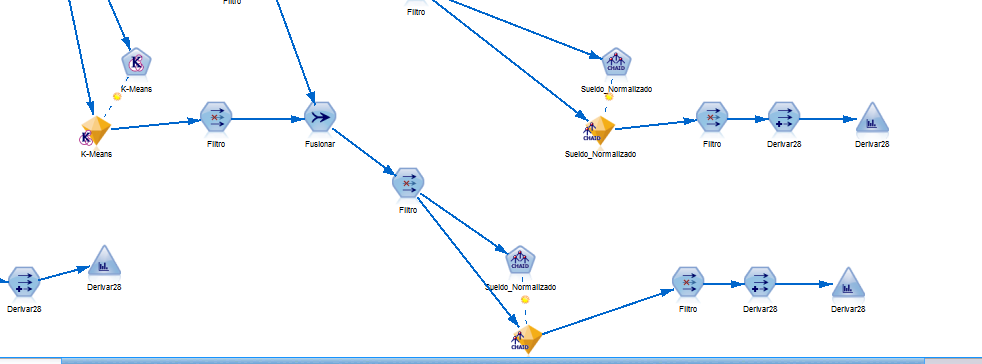


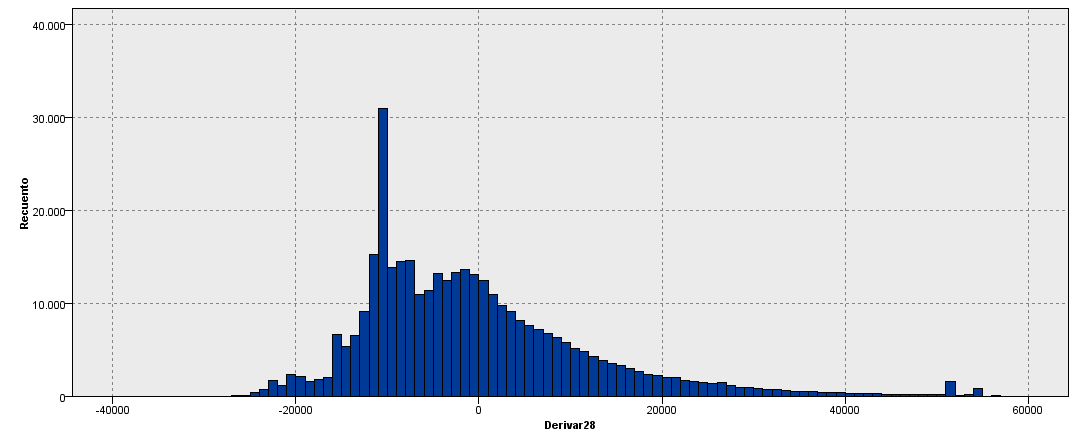


Sin cluster



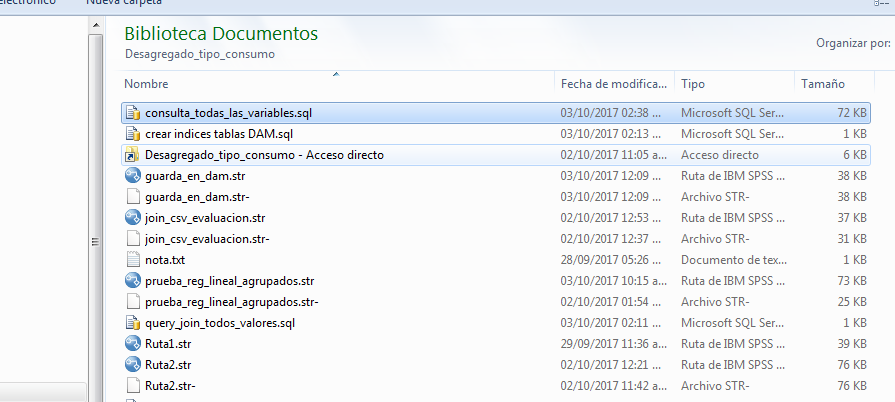
Ahora con kmeadias solamente

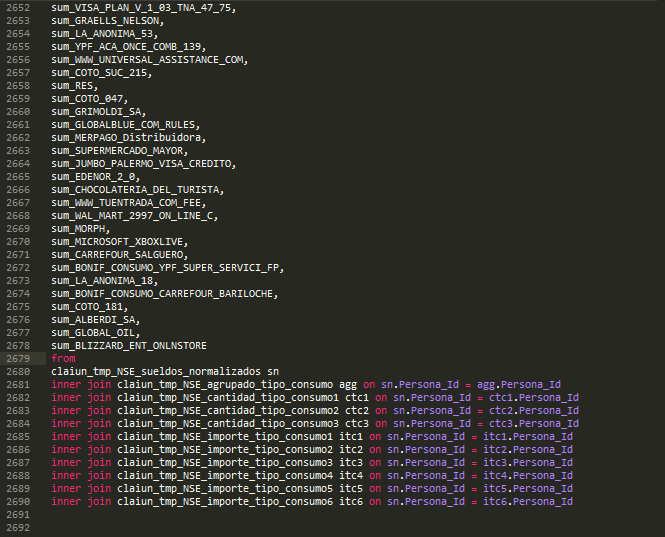




Ahora se prueba corriendo un modelo con todas las variables tirando un árbol y quedando con las variables mas impotantes

C:\Users\LAC40641\Documents\0-Proyectos\Avances\NSE\Desagregado\_tipo\_consumo\consulta\_todas\_las\_variables.sql





Trabajo en

C:\Users\LAC40641\Documents\0-Proyectos\Avances\NSE\Desagregado\_tipo\_consumo\prueba\_modelo\_todas\_variables.str

Prueba con todas las variables y ratios

