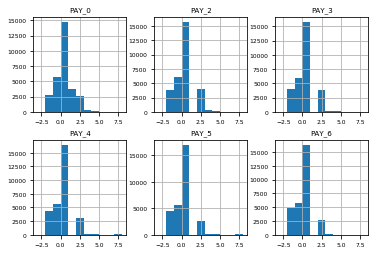
# Lecciones Aprendidas Credit One

Para realizar el presente análisis fue necesario dedicarle bastante tiempo a la parte técnica, y aprender aún más del lenguaje de programación Python y sus diversas librerías, entre ellas Pandas, matplotlib y seaborn.

Es muy importante trabajar con gráficos, ya que esto nos da una perspectiva más detallada de las variables que tenemos en el dataset, y podemos realizar comparaciones para tener un mejor análisis.

Por ejemplo, al graficar los atrasos mensuales en los pagos, pude observar que habían gráficos que se repetían, como se puede apreciar en los gráficos siguientes:



Mi hipótesis al revisar esa información es que, muy probablemente los datos que corresponden a los meses de abril a agosto se fueron obtenidos de manera errónea, por lo que en este caso sería bueno consultar con el cliente o el dueño del dataset, si la información suministrada es correcta o no. De momento, lo que hice fue omitir dichos atributos para los siguientes pasos de mi análisis.

Otras consideraciones que tomé en cuenta en mi análisis, fue la de eliminar la columna de identificador, ya que como en casos anteriores, no es un atributo que pueda generar valor a mi análisis.

También procedí a transformar los valores de ciertos atributos a columnas, como fue el caso de los valores de las variables SEX, EDUCATION y MARRIAGE.

Para hacer esta transformación procedí a convertir los valores numéricos a textos a través de un diccionario, y posteriormente utilizar la función para asignar variables dummy en nuevas columnas.

Al final del proceso, procedí a eliminar los atributos SEX, EDUCATION y MARRIAGE, ya que no me aportaban valor en el análisis de correlación.

Otros atributos eliminados fueron los BILL\_AMT2, BILL\_AMT4 y BILL\_AMT5, esto debido a que la correlación entre estas variables era muy fuerte.

Posterior al análisis de los datos, generé un gráfico con la matriz de correlación se muestra de la siguiente forma:

