Lecciones Aprendidas

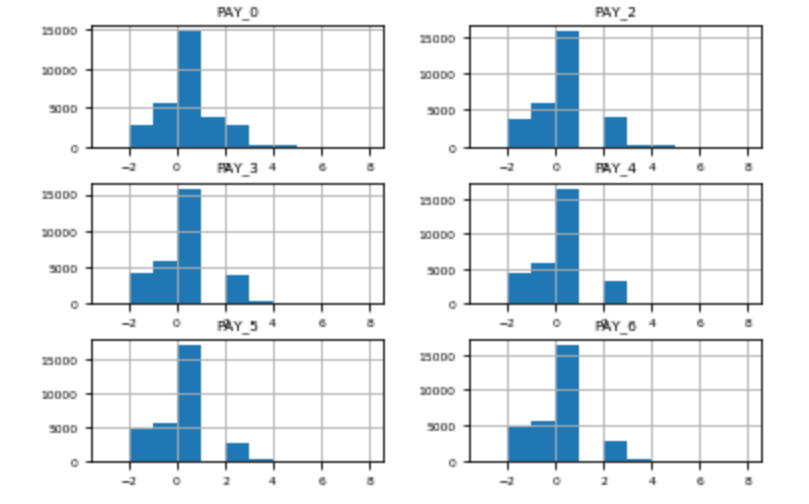
Prepare and Explore the Data

Oscar Quesada Ávalos

Universidad cenfotec

Con el fin de realizar el análisis solicitado es necesario investigar sobre la librería Pandas, matplotlib y seaborn. Generalmente este tipo de investigación es normal cuando hablamos de lenguajes de programación estilo open source, este tipo de lenguaje de programación tiene como común denominador su sintaxis de programación y por otro lado las librerías estas pueden ser migradas de un lenguaje a otro sin con una dificultad técnica elevada aun asi en su trasfondo mantienen mismas funciones a modo general.

Cabe recalcar que la información mas importante es la que vamos a generar de forma visual ya que nos muestra un reflejo del contenido de la data set y podemos realizar análisis para ver nuestra información.   
  
Vamos a mostrar gráficamente los atrasos mensuales de los pagos, en lo cual podemos identificar gráficos que se repiten:



La conclusión va refiriéndose a que los datos provienen de los meses de abril hacia agosto de manera errónea, con lo cual nos surge la duda de consultar al autor de la data. Para verificar si dicha información es correcta o no. Actualmente lo que se hace es ignorar dichos atributos para los pasos siguientes de nuestro análisis.

Es importante mencionar la forma en que vamos a tratar la data o transformar por ejemplo se ignora el id, se transforma atributos a columnas como Sex, Education y Marriage. Para esto se convierte valores numéricos en texto.

Finalizando nuestro proceso, los datos anteriores se eliminan ya que no genera información y tampoco nos afecta. También BILL\_AMT2, BILL\_AMT4, BILL\_AMT5 correlación muy fuerte de las variables por lo cual se eliminan.

Luego de tratar la data y su análisis respectivo, obtenemos un grafico con la matriz de correlación:

