**Tema 5**

**Task 2: Prepare and explore the data**

**Francisco Paniagua Barrantes**

**Lessons Learned**

1. Se recomienda utilizar abreviaciones para nombrar a los paquetes instalados, esto con el fin de facilitar su llamado o utilización con los comandos.

2. Utilizo credit.head para visualizar los datos y se carga el data set pero se visualiza de una forma poco amigable, lo anterior se debe a que no coloque los paréntesis (), volví a correr credit.head() y se visualizó muy bien.

3. Utilice credit para visualizar el set de datos y se carga de forma similar al head() y también como un tail().

5. Si requiero realizar un pronóstico de clasificación requiero variables de tipo nominal, por lo que se va a necesitar el tipo de las variables dependientes e independiente.

6.Para agregar el título y los nombres de los ejes intente varias formas, pero al final me resulto lo siguiente:

plt.title() plt.xlabel() y plt.ylabel()

7. Para el boxplot nótese que A es para la variable a graficar, 0 es para la mediana y 'gD' para el color, en este caso verde

8. Al generar la correlación, donde se observa la mayor correlación es con los pagos atrasados.

9. Se recomienda cambiar el nombre de la variable dependiente. En el futuro el nombre de las variables no debe estar compuesto de varias palabras y se debe verificar que las palabras no sean comandos reservados de Python.

Se utilizo: df.rename(columns = {'y':'year'})

Me dio un error porque no estaba guardando el cambio en la variable credit, lo corregí de la siguiente forma:

credit = credit.rename(columns = {'default payment next month' : 'DEFAULT'})

10. Parece ser que al aumentar el monto del pago disminuye el Default. Se observa que los créditos entre 0 y 200000 dólares son los que presentan mayor tendencia al default, por lo tanto, se recomienda revisar los filtros para los créditos de montos de entre 0 y 200000 dólares.