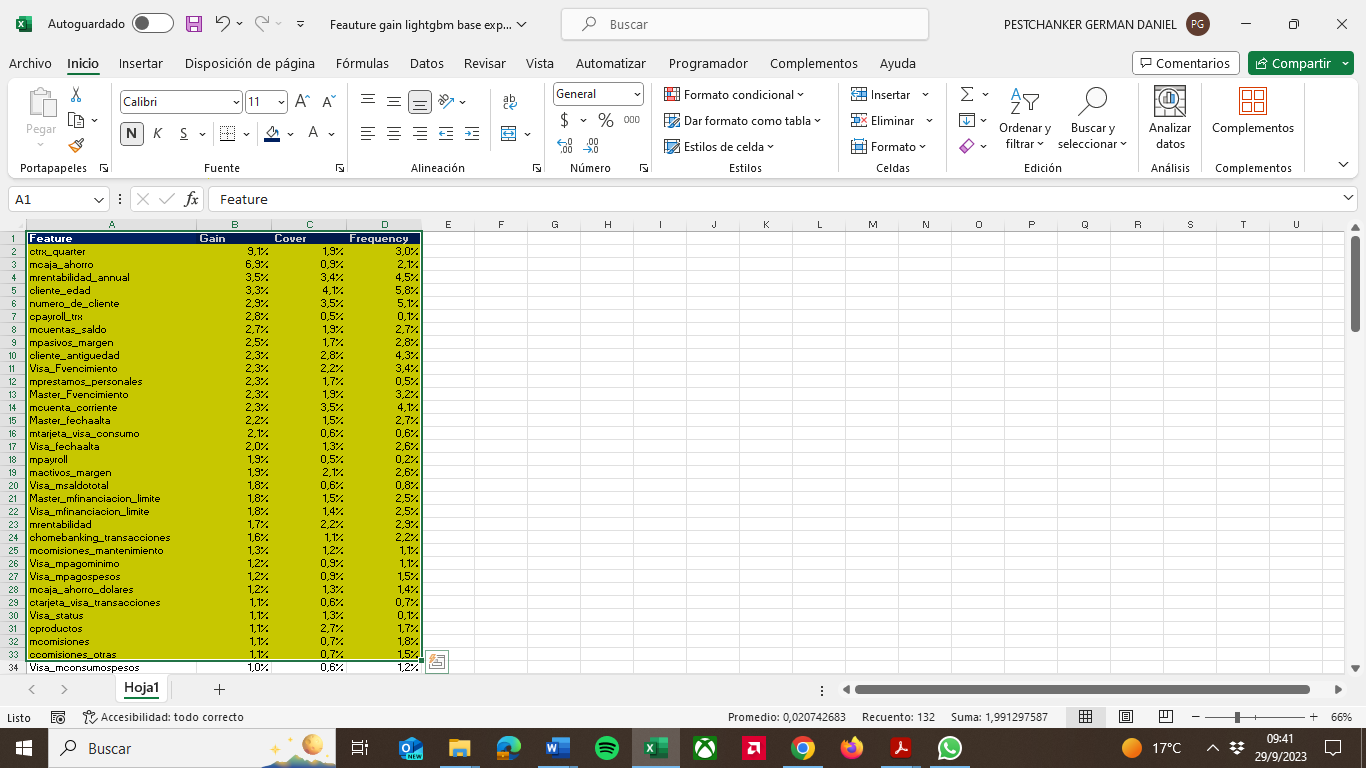
* En esta primera etapa, para la “Transformación del Dataset” me voy a concentrar en las variables que suman el 70% de la importancia en el lightgbm base, y cada atributo individual tiene más de 1.1% de importancia aprox (si hay alguna en el límite pero que puede tener una transformación con otra variable la agrego).



* Data drifting Se puede normalizar (por media y desvío). En este ejercicio, voy a hacer ajuste por inflación (inflacto al 202107= en totas las variables monetarias (las que dice “pesos” en el diccionario de datos), y luego elimino la variable de inflación acumulada.
* Transformación de variables: voy a hacer una transformación manual

(Ej de la presentación: Visa\_consumo / Visa\_limite\_compra;

mcaja\_ahorro + mcuenta\_corriente + mplazo\_fijo + minversiones1 + minversiones2

mpayroll / cliente\_edad )

* ctrx\_quarter -- Cantidad de movimientos voluntarios en las cuentas bancarias ( no tarjeta de credito ) que el cliente realizó en los ultimos 90 dias.

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

Acá se puede hacer media móvil de 6 períodos par ver la tendencia

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente

* mcaja\_ahorro -- Monto total de la caja de ahorro del Paquete Premium

Acá se puede hacer: mcaja\_ahorro + mcuenta\_corriente + mplazo\_fijo + minversiones1 + minversiones2 + mcaja\_ahorro\_adicional + mcaja\_ahorro\_dolares + mcuenta\_corriente\_adicional + mcuentas\_saldo + mcuenta\_debitos\_automaticos

Captura de pantalla de computadora

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente (hay que agregar meses pero sacar la época de la pandemia)

* mrentabilidad -- Ganancia total que ha obtenido el banco de ese cliente, en ese mes.; y mrentabilidad\_annual -- Ganancia total que el banco ha obtenido de ese cliente en el último año de relación cliente-banco, o meses desde que ingresó si es cliente reciente.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Gráfico, Histograma

Descripción generada automáticamente Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

Se podría juntar la media móvil de 12 períodos (1 año) y mrentabildiad\_annual, con una suma, y solucionar el drifting con la inflación

* cliente\_edad: está ok está variable, no hay que tocar nada
* cpayroll\_trx -- Cantidad de Acreditaciones de Haberes en relación de depencia que le hicieron al cliente en ese mes. Un cliente puede estar en relacion de dependencia con mas de una empresa. Una empresa puede hacerle VARIOS depósitos al mismo empleado durante el mes. Soalmente se consideran las acreditaciones de empresas que tienen un contrato con el banco.

mpayroll pesos Monto total que le acreditaron los empleadores “acreditados” al cliente durante el mes.

mpayroll2 pesos Monto total que le acreditaron fuera de archivo de los empleadores “acreditados” al cliente durante el mes.

cpayroll2\_trx -- Cantidad de transacciones de acreditacion de haberes en el mes

transformación:

cpayroll\_trx + cpayroll2\_trx

Drifting en mpayroll

mpayroll2: fijo, no cambia nada hacer una transformación

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Zero cpayroll y mpayroll: fue bajando, pero parece una tendencia por cambio de sistema: se puede dejar estable en 0.5

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

* mpasivos\_margen pesos Monto total de la ganancia que el banco ha obtenido por el dinero/inversiones que el cliente tiene en el banco.:

Hay drifting. Arreglar los ceros.

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

* Cliente\_antiguedad: todo ok.
* -Master\_Fvencimiento dias Dias para el vencimiento del plastico de la tarjeta de crédito, contados a la fecha de la foto.

-Master\_fechaalta dias Dias desde de alta de la cuenta de la tarjeta de crédito, contados a la fecha de la foto.

-Visa\_Fvencimiento dias Dias para el vencimiento del plastico de la tarjeta de crédito, contados a la fecha de la foto

- Visa\_fechaalta dias Dias desde el alta de la cuenta de la tarjeta de crédito, contados a la fecha de la foto.

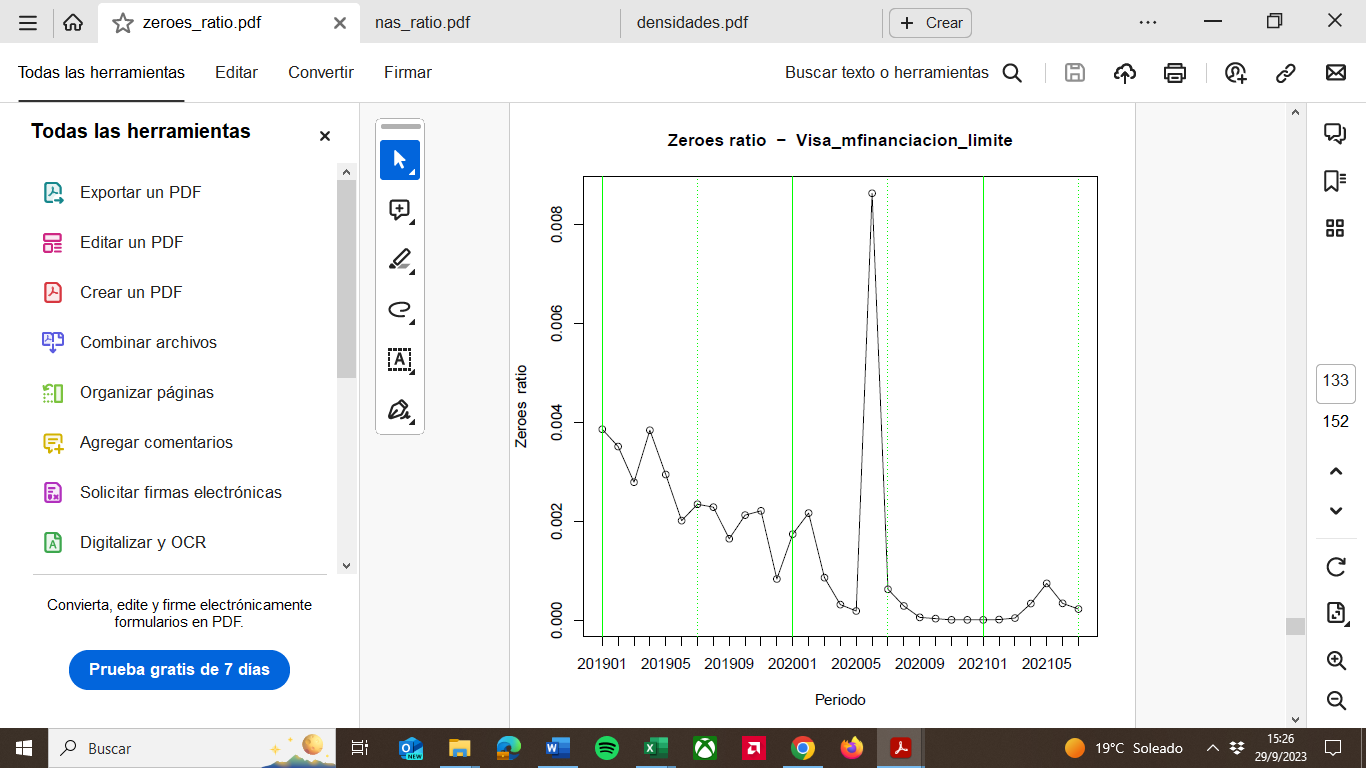
Sumar la diferencia entre vencimiento y alta entre ambos

Están ok las variables.

Por ahora, probar hasta acá-

* Master\_mfinanciacion\_limite; Visa\_mfinanciacion\_limite

Sumar las dos (en visa parece raro , en jun-20 error)



Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

* “Catastrophe Analysis” (Revisión de Calidad). Para este caso, chequeo de que no haya nulos, y que los ceros tengan un comportamiento adecuado
* Variables históricas: (Si entreno en un dataset union de varios períodos, el algoritmo no se da cuenta que un numero\_de\_cliente es la misma persona a lo largo de los meses). En variables monetarias se puede hacer una media móvil de 2 o 3 períodos, que sintetiza la tendencia que buscan otras medidas (delta, aceleración, regresión).