### 

Proyecto Final Análisis de datos 1

**Descripción del problema**

Facilitar a los usuarios directos e indirectos del buró de crédito en Colombia la mejora de su perfil crediticio mediante estrategias efectivas de educación financiera. Como primer paso, se identificará a los usuarios con las calificaciones crediticias más bajas o categorizadas como 'malas', permitiendo así, proponer una intervención personalizada y orientada a corregir comportamientos financieros para potenciar su puntaje y acceso al crédito.

**Justificación**

Se puede usar IA y ciencia de datos analizando los datos crediticios de los históricos de los usuarios del Buro de Credito. Esto permitirá identificar patrones y comportamientos financieros comunes que han llevado a mejoras o deterioros en los puntajes crediticios. Podemos buscar segmentar con IA a los usuarios en diferentes perfiles y personalizar el programa de educación financiera.

Esto puede ayudar a brindar tutorías inteligentes basadas en IA que ofrezcan recomendaciones en tiempo real, a través de simulaciones de decisiones financieras y respuestas a preguntas frecuentes. Con esto podemos ayudar a las personas a entender cómo ciertas acciones pueden impactar su puntaje crediticio.

**Pregunta SMART**

Identificar cuales son las variables que generan mayor impacto en el buro de credito de colombia para ayudar a las personas a mejorar su puntaje crediticio.

**Estructura de los datos**

**Carga de datos inicia:**

Revisamos estadísticas descriptivas y visualizamos relaciones entre las variables numéricas y categóricas usando gráficos de dispersión.

* Edad Género
* Ingreso\_mensual
* Saldo\_total
* Número\_de\_cuentas
* Historial\_crediticio
* Tipo\_de\_empleo
* Porcentaje\_uso\_credito
* Mora\_en\_pagos
* Número\_de\_moras
* Monto\_total\_prestado
* Puntaje\_crediticio
* Score\_Buro

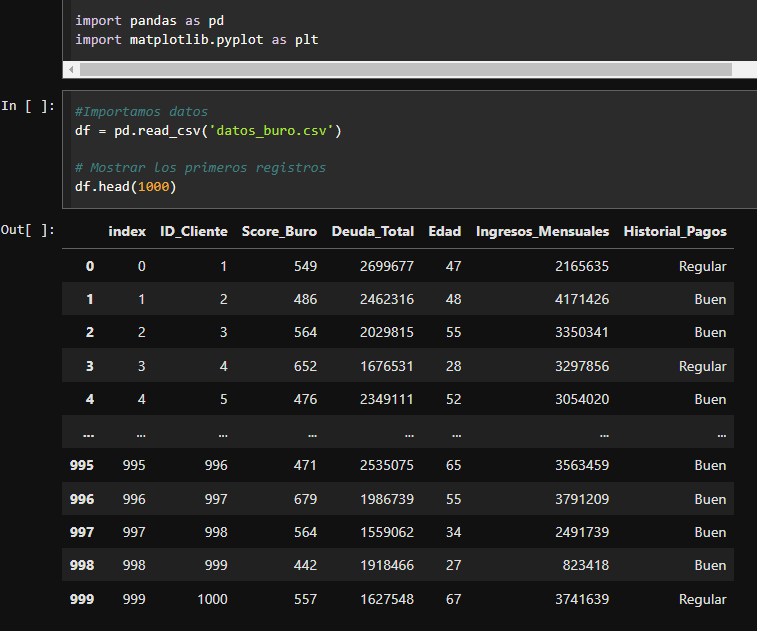
Se observa que las variables numéricas clave son:

* Ingreso\_mensual
* Saldo\_total
* Porcentaje\_uso\_credito
* Historial\_crediticio
* Número\_de\_moras
* Score\_Buro

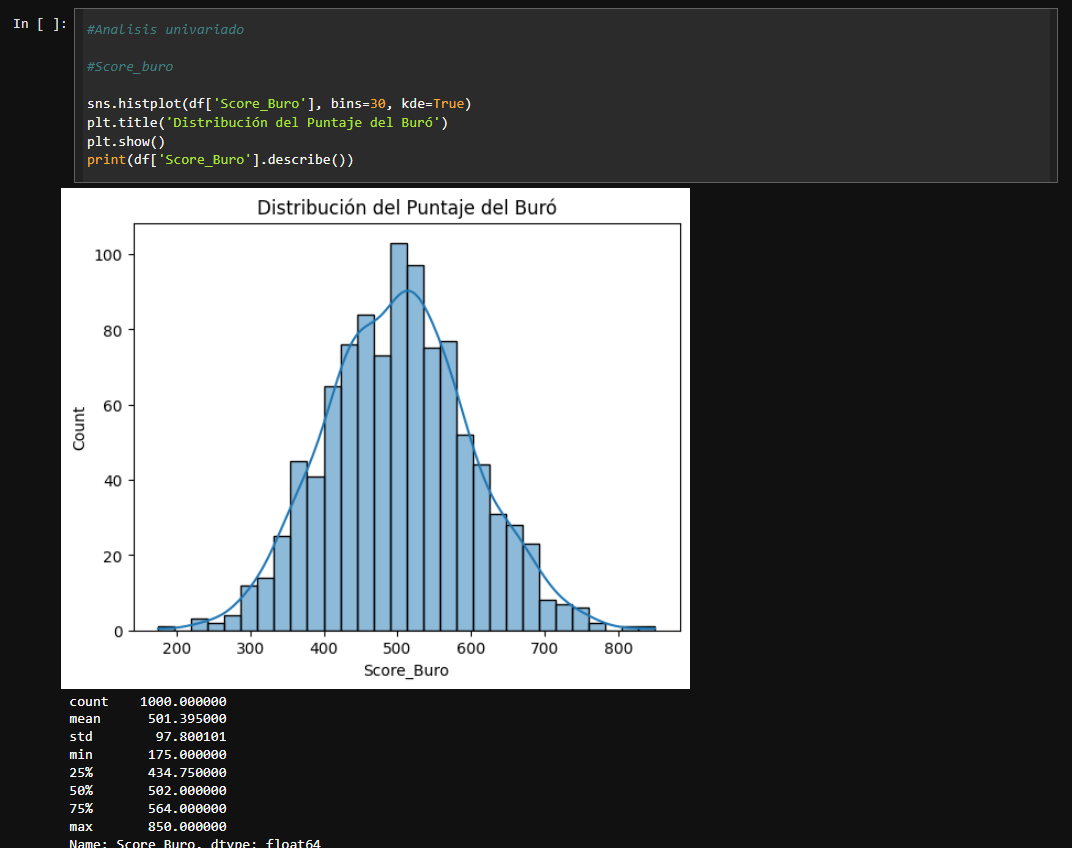
Variables categóricas clave son:

* Género
* Mora\_en\_pagos
* Tipo\_de\_empleo

## **Análisis Descriptivo:**



Se realiza un primer análisis univariado sobre el campo Score\_buro para obtener una visión general de su distribución, el objetivo es observar la distribución del porcentaje crediticio como primera medida para saber el comportamiento de los usuario.



Se logra observar una distribución normal

**Media (501)**: El puntaje promedio es 501, lo que indica que el perfil de crédito de la mayoría de los usuarios es aceptable.

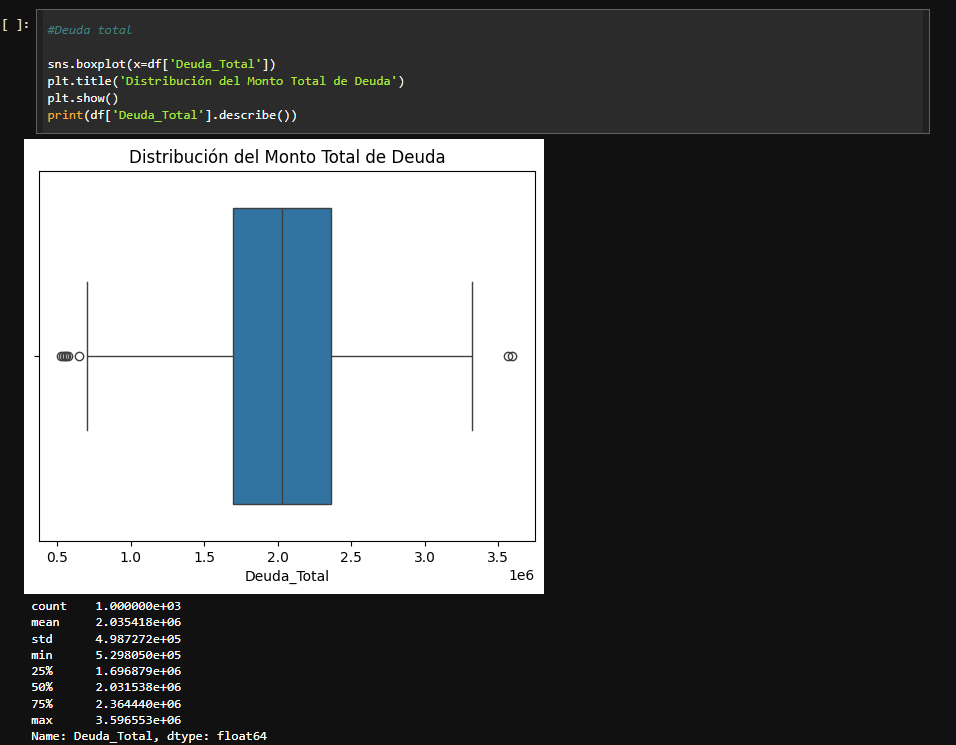
**Desviación estándar (50)**: Los puntajes están bastante concentrados en torno a la media; no hay demasiada dispersión.

**Mínimo (175) y máximo (850)**: El puntaje más bajo es 175 y el más alto es 850, mostrando que la mayoría de los usuarios tienen puntajes en un rango amplio pero controlado.

**Percentil 25 (434), Mediana (502), Percentil 75 (564)**: Lo cual indica que el 50% de los usuarios tiene puntajes entre 434 y 564, lo cual es relativamente estrecho en comparación con el rango total, lo que sugiere que la mayoría de los usuarios tienen un perfil crediticio moderado.

*El análisis del Score\_Buro muestra que la mayoría de las personas con reporte negativo tienen puntajes concentrados en los niveles bajos, lo que refuerza su clasificación como de alto riesgo.*

Se realiza un primer análisis univariado sobre el campo Saldo\_total para obtener una visión general de su distribución de la deuda, el objetivo es observar la distribución del endeudamiento total para saber el comportamiento de los usuario.



El análisis del Score\_Buro muestra que la mayoría de las personas con reporte negativo tienen puntajes concentrados en los niveles bajos, lo que refuerza su clasificación como de alto riesgo.

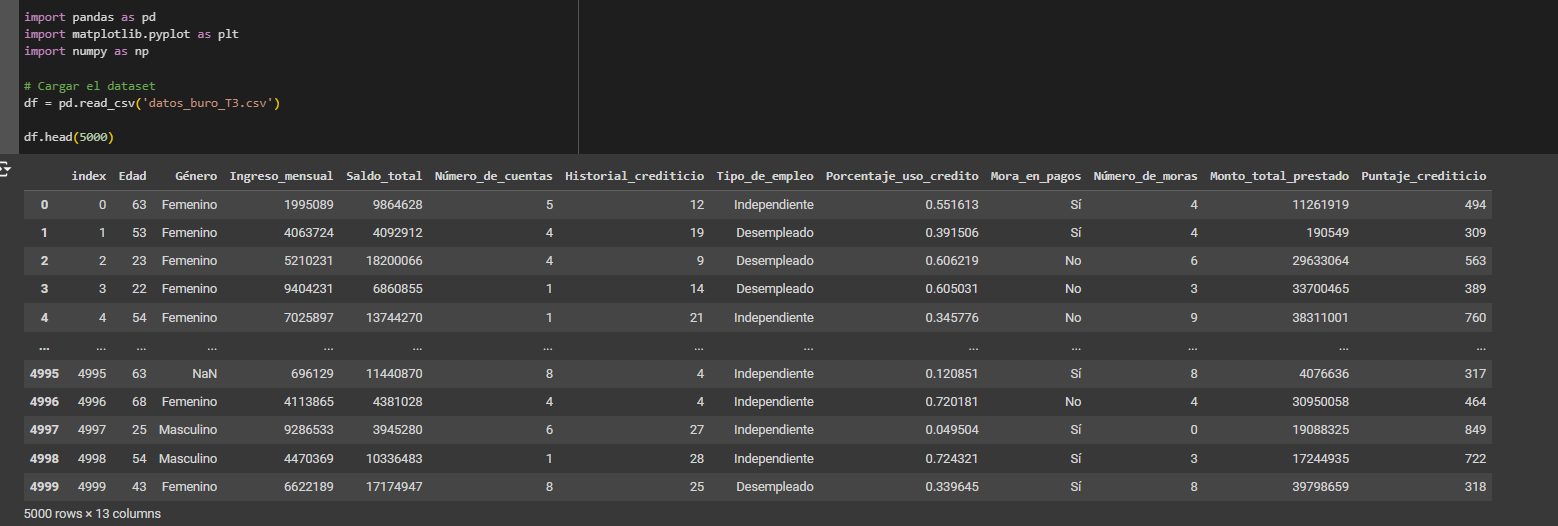
Por otro lado, la Deuda\_Total presenta una distribución con valores extremos que pueden indicar la presencia de deudas significativas en un subconjunto de la población. Esto sugiere que, además del puntaje,

El monto de la deuda podría ser un factor clave en la probabilidad de reporte negativo en el buró de crédito.

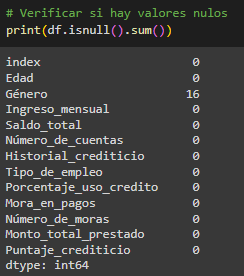
**Detección de nulos o faltantes**

Se determina el porcentaje de valores faltan en cada variable.

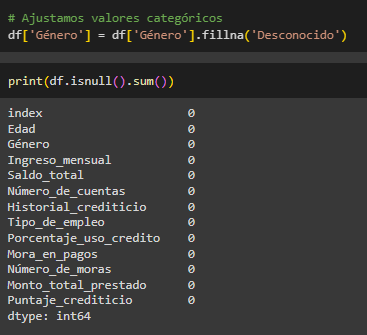
Se toman decisiones: eliminar, ellenar con la media, mediana o crear una categoría específica, mediante Label Encoding y One-Hot Encoding.



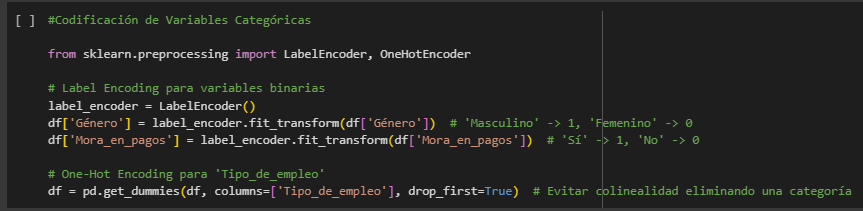
Verificación de valores nulos

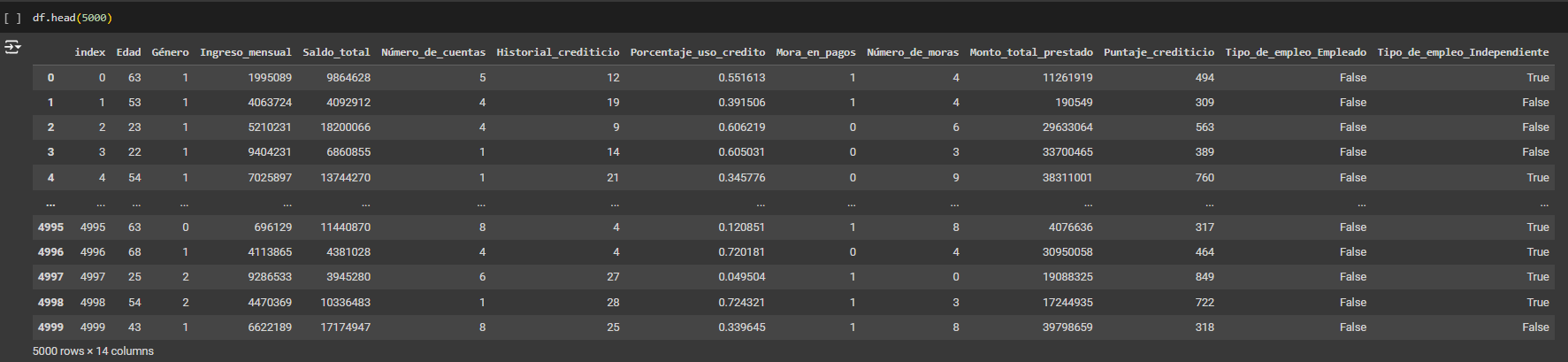


Se Ajustan valores categóricos



***Label encoding y one-hot Encoding***

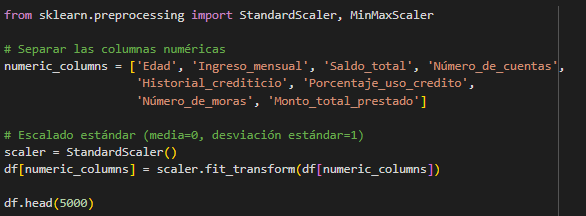


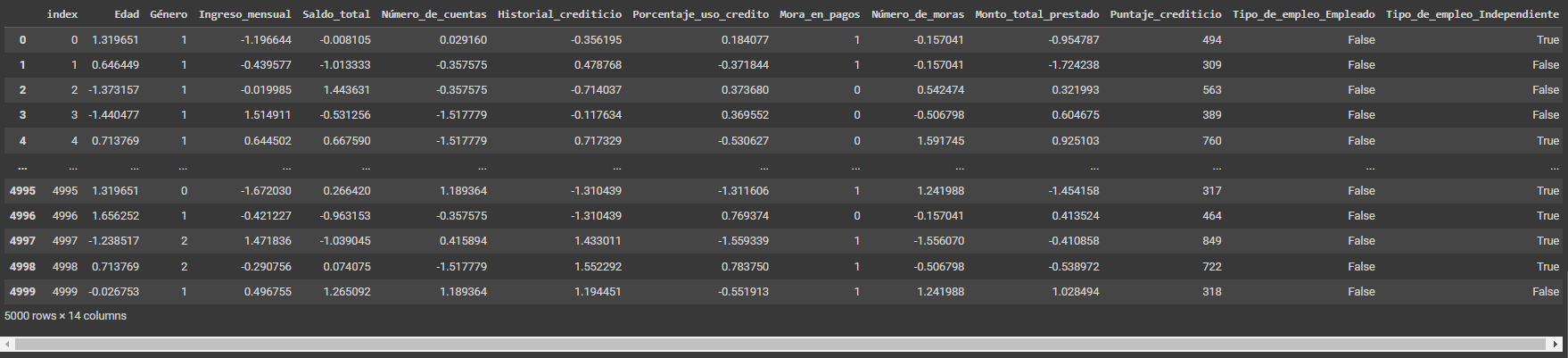


**Análisis y distribución de variables.**

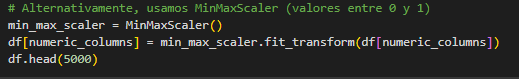
A continuación ajustamos las variables numéricas escaladas utilizando StandardScaler.

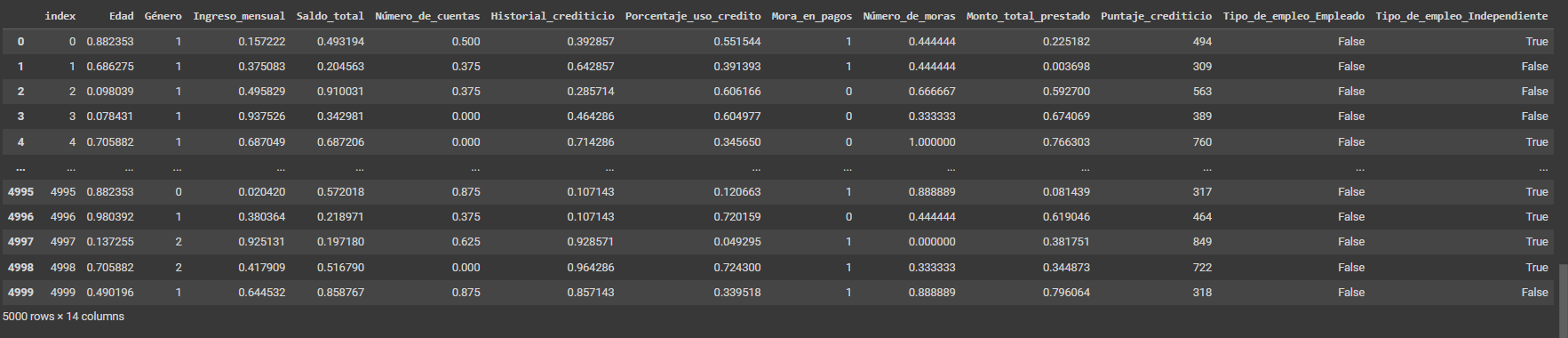
Todas las columnas innecesarias (como identificadores únicos o categorías redundantes) se eliminan o transforman.



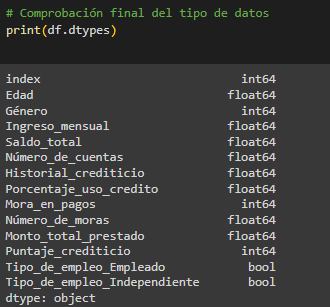


Usamos minMaxScaler en las columnas numéricas





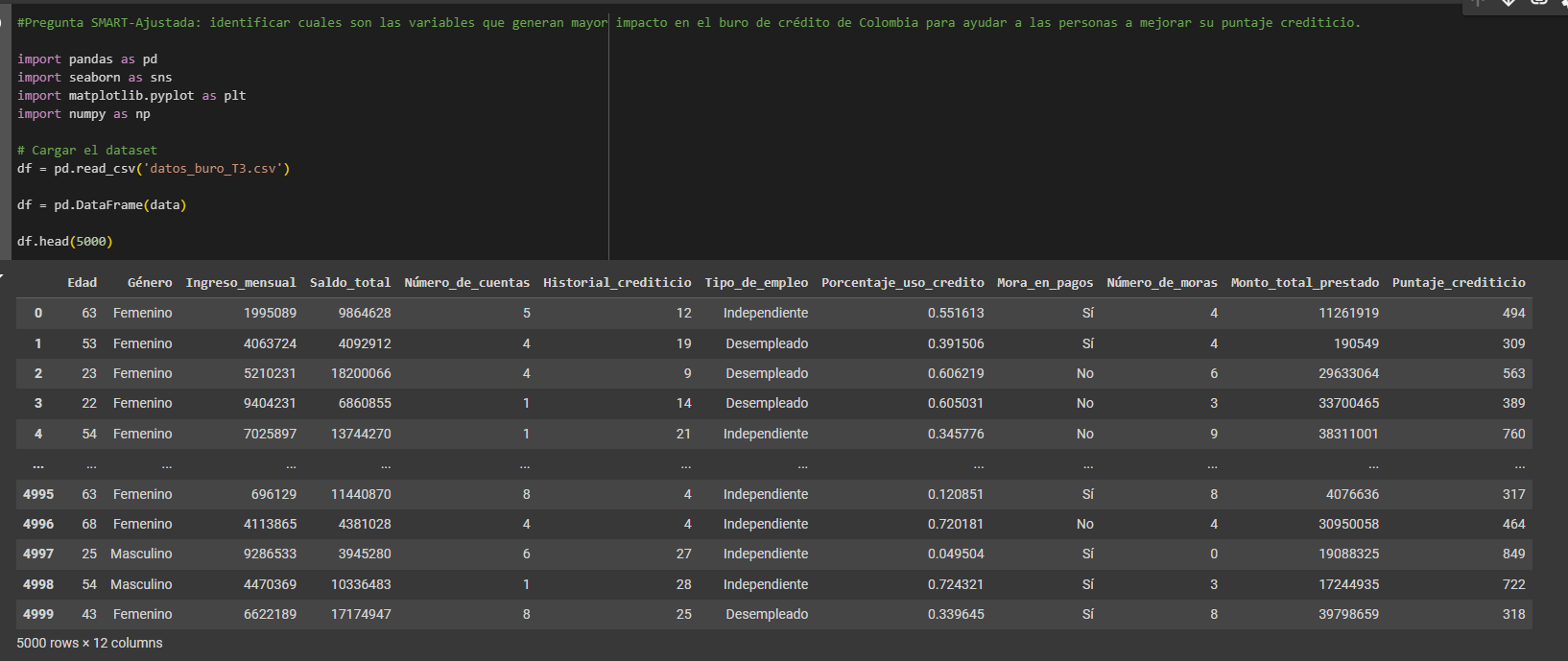
Comprobamos los datos finales

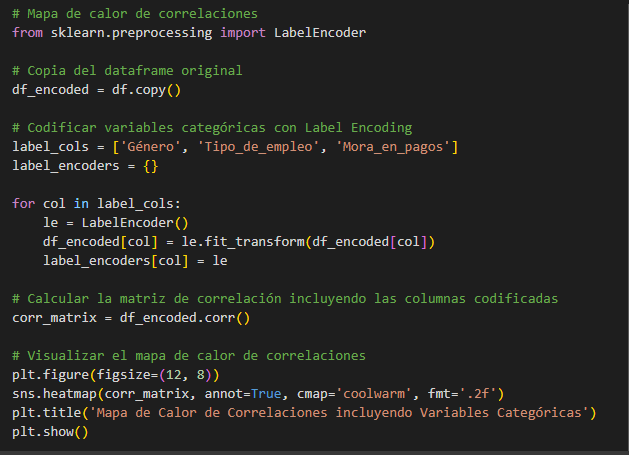


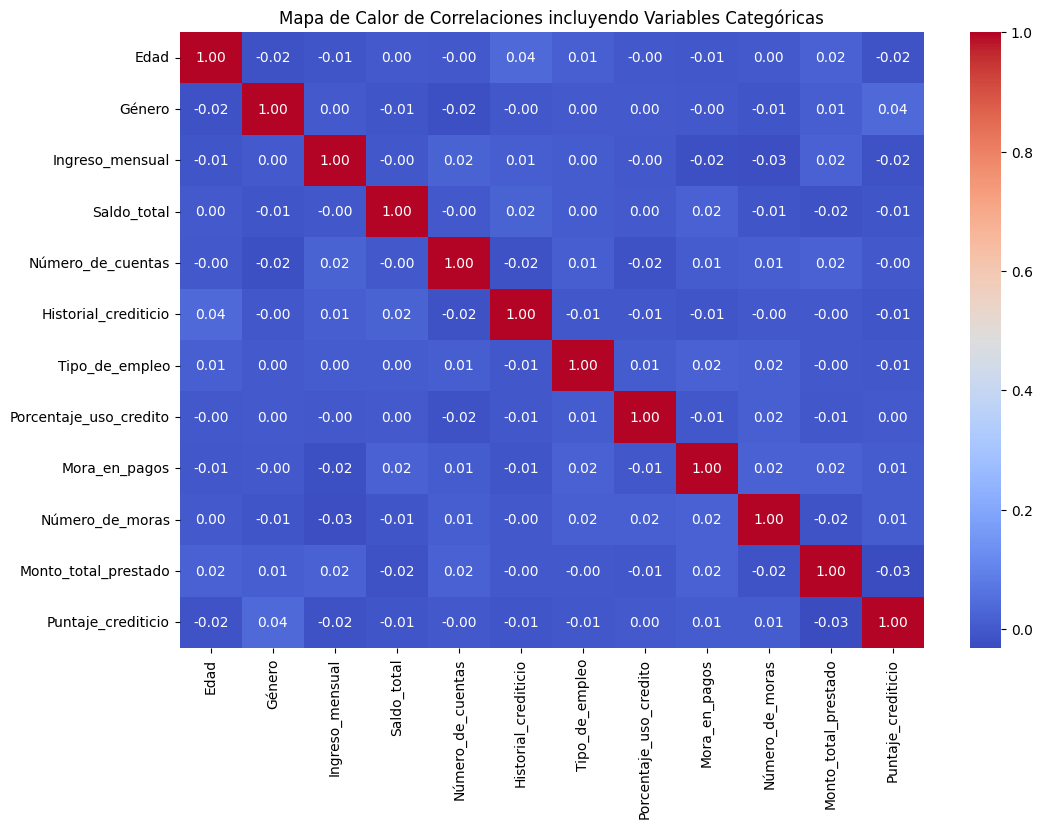
**Análisis de correlaciones**

En el análisis de correlaciones buscamos identificar relaciones lineales entre variables numéricas y categóricas.

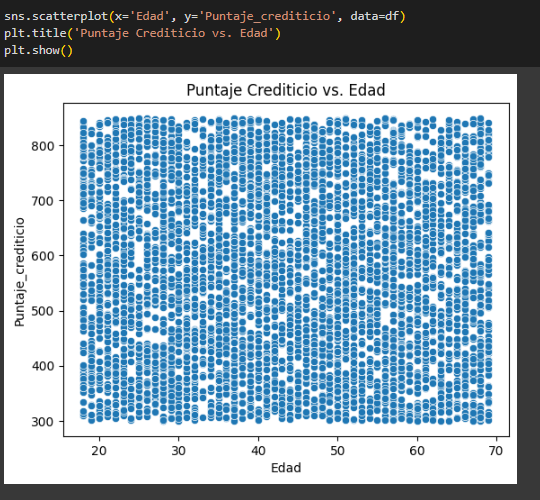
En este caso se carga un dataset con más columnas y datos para poder mejorar el análisis.

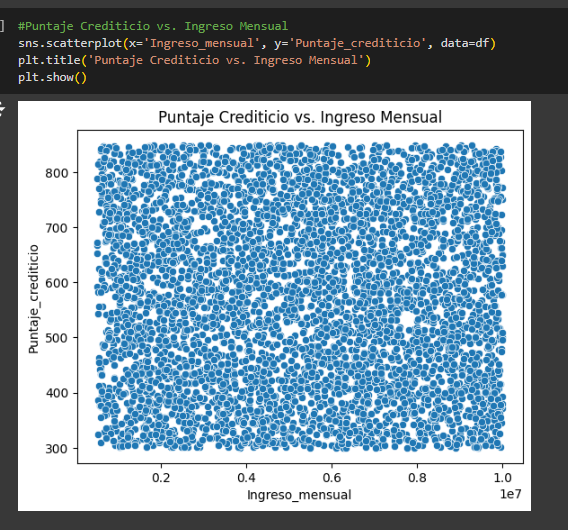






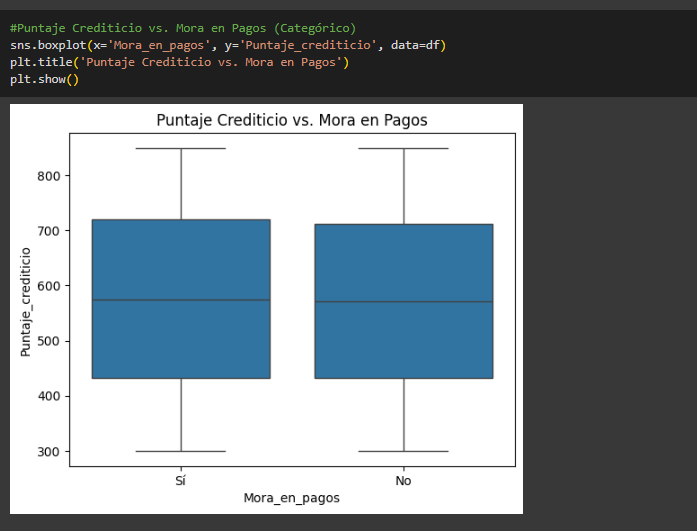
Del análisis de correlación, podríamos concluir que variables como el saldo total, el ingreso mensual, y el porcentaje de uso de crédito tienen una correlación significativa con el puntaje crediticio. Esto sugiere que las personas con mayores ingresos y menores saldos tienden a tener un mejor puntaje.





Edad y Historial Crediticio: Aunque la edad puede tener alguna relación con el puntaje, es probable que el historial crediticio sea un predictor más fuerte,

dado que un historial más largo puede indicar mayor experiencia en el manejo de crédito.



Variables Categóricas: En el análisis bivariado, se podría observar que aquellos que han estado en mora en pagos tienen un puntaje significativamente más bajo,

lo que indica que la mora en pagos es una variable crítica para el puntaje crediticio.

Conclusión

En el análisis exploratorio de la pregunta SMART, "Identificar cuáles son las variables que generan mayor impacto en el buró de crédito de Colombia para ayudar a las personas a mejorar su puntaje crediticio", se observan las siguientes conclusiones clave:

Distribución del Puntaje Crediticio: La mayoría de los usuarios presentan puntajes entre 434 y 564, lo que indica una concentración de personas con puntajes moderados o bajos. Existen valores extremos que representan usuarios con puntajes muy altos o muy bajos.

Variables con Mayor Impacto:

Porcentaje de Uso del Crédito: Una alta correlación negativa con el puntaje crediticio sugiere que cuanto mayor es el uso del crédito, peor es el puntaje.

Historial Crediticio: Los usuarios con un historial más largo y sin moras suelen tener puntajes más altos.

Mora en Pagos: La presencia de moras impactó negativamente y de manera significativa en el puntaje.

Ingresos Mensuales: Usuarios con mayores ingresos tienden a tener puntajes más altos, pero esta relación es moderada.

Variables Menos Relevantes:

Género y Edad: Estas variables muestran una baja correlación con el puntaje crediticio, lo que indica que no son factores determinantes.

Este análisis permite identificar áreas clave donde los usuarios podrían mejorar su comportamiento financiero, como reducir el uso del crédito y evitar las moras en pagos.