

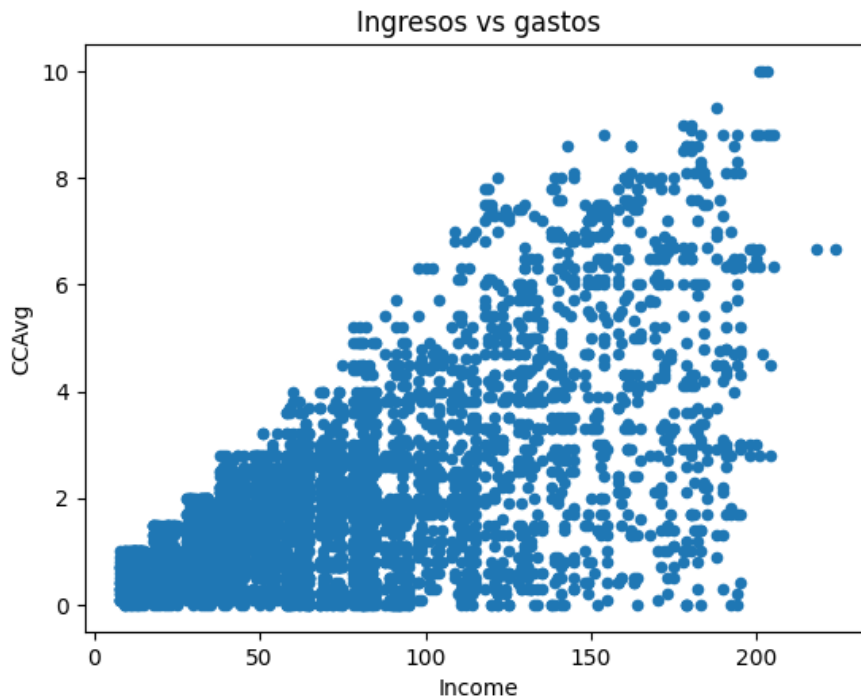
Reporte 2 Regresión Lineal

Para esta práctica, trabajé con un archivo obtenido de Kaggle (<https://www.kaggle.com/datasets/luisenriquesguerrero/creditos-personales-actualizado>), una base de datos donde muestra a clientes de un banco, ordenados por su ID, edad, experiencia laboral, ingreso y entre otros datos, la relación entre sus ingresos y gastos y si tienen algún préstamo, en resumen, un pequeño balance de su situación económica.

Lo interesante para trabajar era la relación de gasto versus ingreso, las cifras están en (miles de dólares estadounidenses anuales)

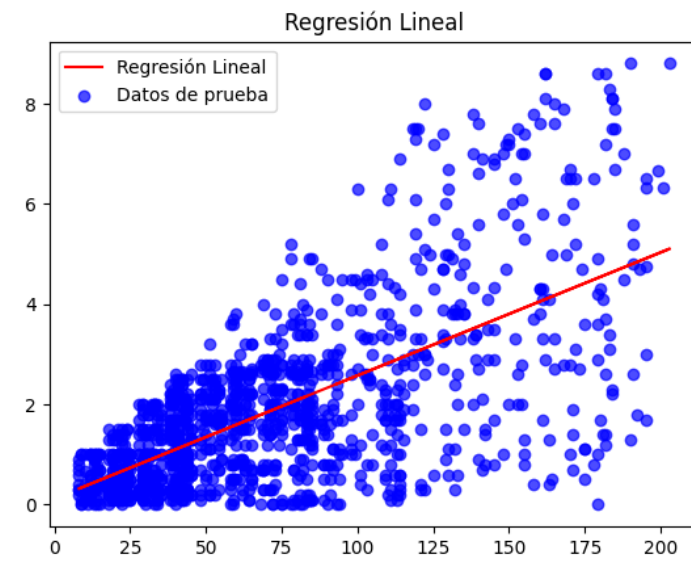
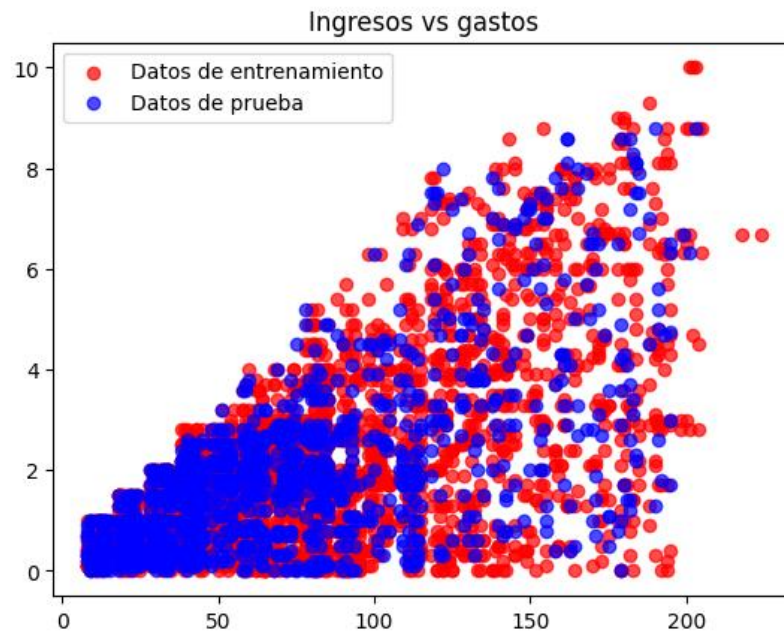
Es intuitivo pensar que los que ganan más tienen más gastos respecto a lo que ganan, sin embargo, también hay valores “atípicos” en este sentido, ya que hay gente que a pesar de tener más ingresos no gasta siempre más que alguien que gane menos, por eso digo atípicos entre comillas, porque por factores externos como el número de dependientes familiares, si tiene o no hipoteca, pueden hacer que una persona gaste más de lo que quisiera y alguien sin estos compromisos puede no gastar, y aunque no viene explícito, se puede dar la oportunidad de ahorrar

Grafica de relación entre el ingreso y el gasto



En el gráfico podemos ver unos saltos algo ordenados en cuanto al crecimiento del gasto según crece el ingreso, y como a partir de 50,000 USD se empieza a distorsionar un poco, agregando variación según se gana más a gastar mucho más o a no gastar a pesar de tener un sueldo mejor a sus predecesores.

Segregación de datos



Hay mucha variabilidad entre los datos, la regresión lineal no es el mejor modelo para estos datos, el puntaje obtenido fue de 0.4252