



INSTITUTO
POLITÉCNICO
NACIONAL



ESCUELA SUPERIOR DE CÓMPUTO

Licenciatura en Ciencia de datos

MINERÍA DE DATOS

EJERCICIO DE REGRESIÓN

PRESENTA: Equipo 2

Fernanda Madariaga Villanueva

Alejandro Maldonado López

Ulises Abdiel Cabello Cárdenas

Andrés Zamarripa Altamirano

DOCENTE:

Fabiola Botello Ocampo

GRUPO:

5AM1

Noviembre 2025

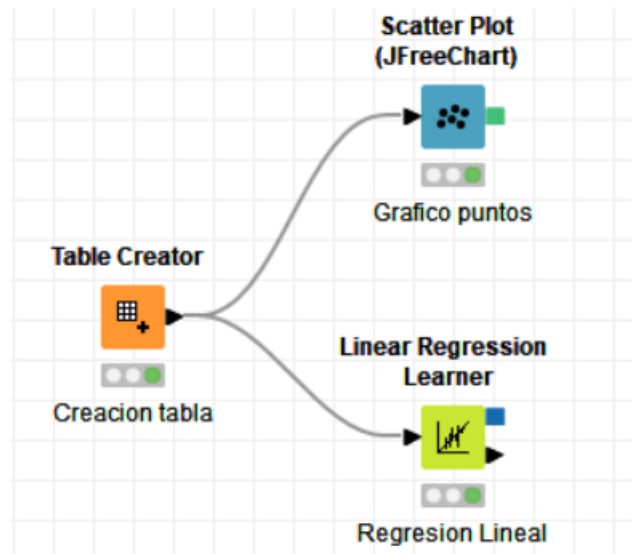


Imagen 1. Flujo en Kettle

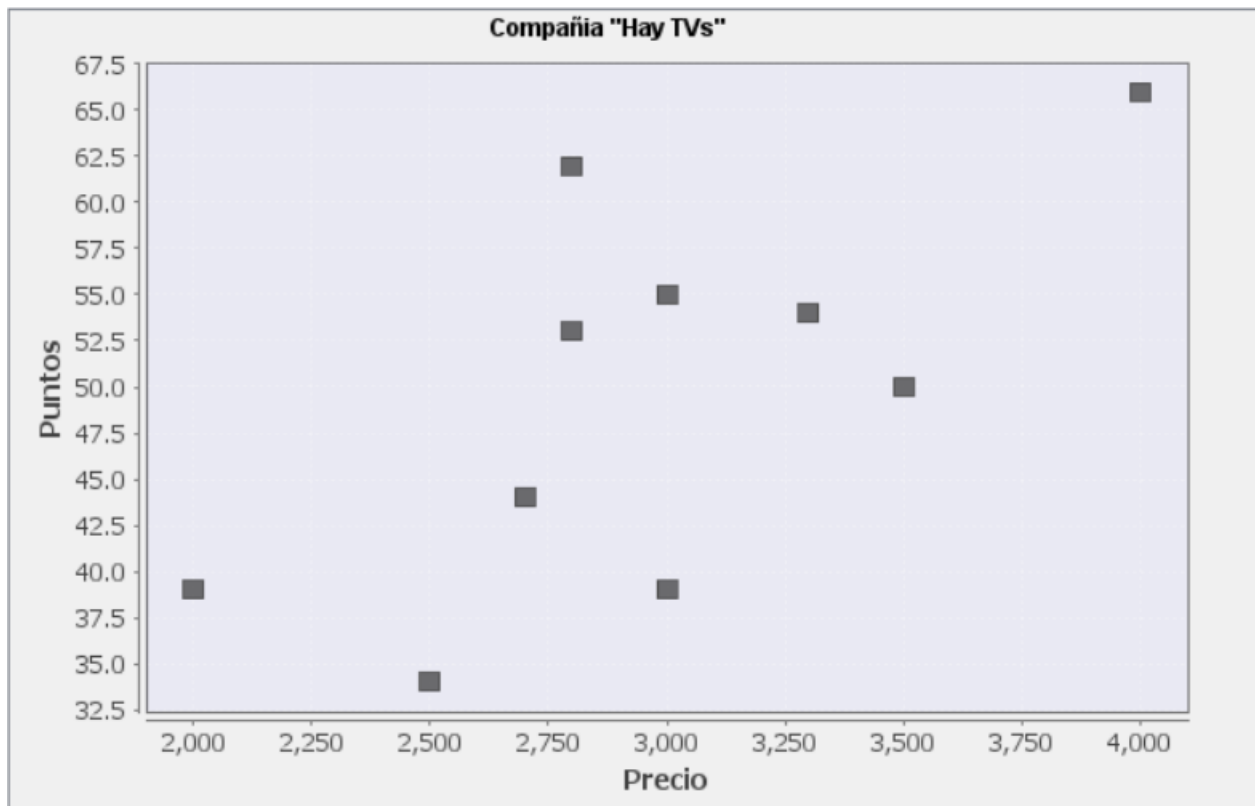


Imagen 2. Precio vs Puntos

Statistics on Linear Regression				
Variable	Coeff.	Std. Err.	t-value	P> t
Precio	0.0127	0.005	2.5599	0.0337
Intercept	12.0175	14.9096	0.806	0.4435
R-Squared: 0.4503				
Adjusted R-Squared: 0.3816				

Imagen 3. Métricas del modelo e información adicional

Interpretación de las métricas:

1. Coeficiente de Determinación (R^2):

$R^2 = 0.4503$. Esto indica que aproximadamente el 45.03% de la variabilidad observada en la variable Puntos puede ser explicada por su relación lineal con la variable Precio. El resto, cerca del 54.97%, se atribuye a otros factores no incluidos en el modelo o a la variabilidad aleatoria inherente a los datos.

2. R^2 Ajustado (Adjusted R-Squared):

R^2 Ajustado = 0.3816. Esta métrica ajusta el R^2 teniendo en cuenta el número de predictores en el modelo y el tamaño de la muestra. Su valor de 38.16% confirma que, tras este ajuste, una parte significativa de la variación en las pulgadas es explicada por el precio, aunque refina la estimación a la baja, lo cual es común en modelos con pocos datos.

3. Intercepto (Constant):

Coeficiente = 12.0175**. Este valor representa el número de Puntos (pulgadas) previsto por el modelo cuando el Precio es cero. Si bien esta interpretación carece de sentido práctico en este contexto (un televisor no puede tener un precio de cero), es un parámetro necesario para definir la línea de regresión.

Valor-p = 0.4435. Dado que este valor-p es mayor que el nivel de significancia común de 0.05, no contamos con evidencia suficiente para afirmar que el intercepto es estadísticamente diferente de cero.

4. Variable Precio:

Coeficiente = 0.0127. Este es el parámetro central de nuestro modelo. Indica que por cada unidad monetaria que aumenta el Precio, la variable Puntos (pulgadas) se incrementa, en promedio, en 0.0127

unidades. En términos prácticos, existe una relación positiva entre ambas variables: televisores más caros tienden a tener un mayor tamaño de pantalla.

Valor-p = 0.0337. Este valor-p es menor que 0.05, lo que nos permite concluir que la relación observada entre el Precio y los Puntos es estadísticamente significativa. Es decir, es muy improbable que esta relación se deba únicamente al azar.

En conclusión, el modelo identifica una relación positiva y estadísticamente significativa entre el precio y el tamaño en pulgadas de los televisores, explicando una parte considerable de su variabilidad.

Ecuación de la Recta de Regresión

Puntos = 12.0175 + 0.0127 * Precio

$$y = 12.0175 + 0.0127 * x$$

¿Por qué esta ecuación?

Esta es la forma estándar de una línea recta ($Y = a + bX$). En nuestro contexto, el Intercepto (12.0175) es el valor base de "Puntos" cuando el "Precio" es cero, y el Coeficiente de Precio (0.0127) es la pendiente, que cuánto esperamos que cambien los "Puntos" por cada unidad que aumente el "Precio". El modelo calculó estos valores para encontrar la línea recta que "mejor se ajusta" a nuestros datos, minimizando las diferencias entre los puntos reales y los predichos por la línea.

Evaluación del Ajuste del Modelo (R^2)

Con un $R^2 = 0.4503$, podemos afirmar que la ecuación de regresión estimada proporciona un ajuste moderado, pero no fuerte.

Interpretación: El modelo logra explicar el 45.03% de la variación en las pulgadas de los televisores. Si bien esto indica una relación discernible, significa que más de la mitad (54.97%) de la variabilidad en el tamaño de las pantallas se debe a otros factores que no están incluidos en este modelo, como la marca, la resolución, las características "smart", etc. Es un punto de partida útil, pero sugiere que el modelo podría mejorarse significativamente incorporando más variables predictoras.

Error Estándar de la Estimación (SEE)

El Error Estándar de la Estimación (también llamado Error Estándar del Residuo) mide la desviación promedio entre los valores reales de "Puntos" y los valores predichos por nuestra línea de regresión. Se

puede calcular a partir de la Desviación Estándar de los residuos, que a su vez se deriva del Error Estándar de los coeficientes y los grados de libertad.

Un cálculo aproximado para nuestro modelo arroja un Error Estándar de la Estimación de aproximadamente 14.91.

¿Qué significa?

En promedio, las predicciones de "Puntos" realizadas por nuestro modelo se desvían aproximadamente ± 14.91 pulgadas del valor real. Este valor nos da una idea de la precisión de nuestras predicciones. Un SEE más bajo indicaría un modelo más preciso. Dado que el rango de "Puntos" en nuestro dataset probablemente es de varias decenas de pulgadas, un error de casi 15 pulgadas es considerable, lo que refuerza la conclusión de que el ajuste del modelo es solo moderado.

Predicción de un Televisor de Precio 3200

Utilizando nuestra ecuación de regresión, estimamos la puntuación (pulgadas) para un televisor con un precio de 3200:

$$\text{Puntos} = 12.0175 + 0.0127 * 3200$$

$$\text{Puntos} = 12.0175 + 40.64$$

$$\text{Puntos} \approx 52.66$$

Estimación: Para un televisor con un precio de 3200, el modelo predice un tamaño de pantalla de aproximadamente 52.7 pulgadas.

Es crucial recordar que esta es una estimación sujeta al error estándar de la estimación calculado anteriormente (± 14.91 pulgadas), por lo que el valor real podría variar significativamente.