



INFORME FINAL: ANÁLISIS EXPLORATORIO DE DATOS

Estudiante

KEVIN SANTIAGO CUESTA HERNANDEZ

PROGRAMA INGENIERÍA DE SISTEMAS

MINERÍA DE DATOS

Profesor

JULIAN ANDRES QUIMBAYO CASTRO

12 Noviembre 2025



Introducción

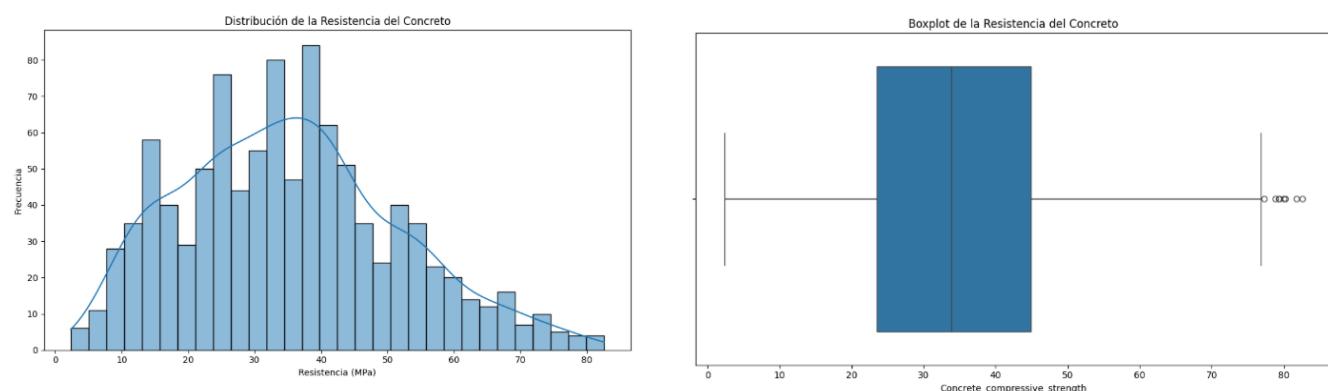
El presente documento presenta los resultados del análisis exploratorio de datos (EDA) realizado sobre el dataset "Concrete Compressive Strength" del repositorio UCI Machine Learning. El objetivo principal fue investigar la relación entre los componentes del concreto y su resistencia compresiva, específicamente enfocándose en responder la pregunta de investigación: "¿Qué relación existe entre la cantidad de cemento y la resistencia del concreto, y es posible predecir la resistencia compresiva usando únicamente esta variable?"

El dataset contiene 1,030 muestras de mezclas de concreto con 8 variables predictoras (cemento, escoria de alto horno, ceniza volante, agua, superplastificante, agregado grueso, agregado fino y edad) y una variable objetivo (resistencia compresiva medida en MPa).

Hallazgos Principales

Calidad de Datos: Dataset de alta calidad: cero valores faltantes, duplicados mínimos eliminados, 1,030 muestras válidas para análisis robusto.

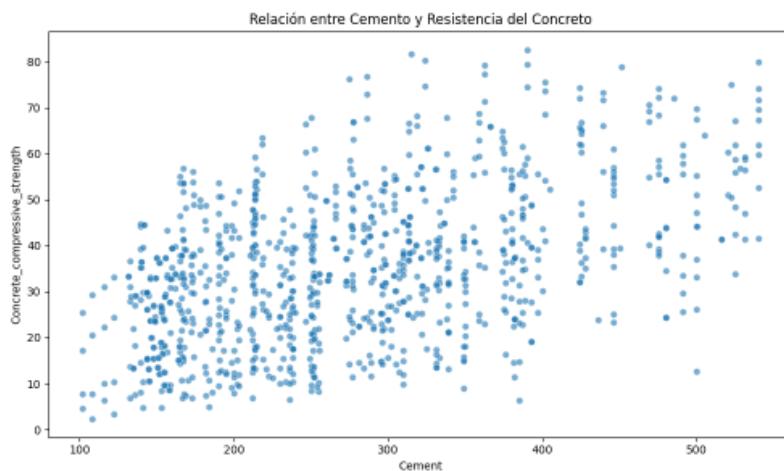
Distribución de Resistencia



Distribución aproximadamente normal con ligero sesgo derecho, concentrada entre 20-50 MPa (pico en 35-40 MPa). Mediana ~35 MPa. Outliers identificados: bajos (0-10 MPa, posibles fallos) y altos (>70 MPa, mezclas optimizadas).



Relaciones Clave



Cemento vs Resistencia: Relación positiva moderada ($r=0.46$). Mezclas con $>400 \text{ kg/m}^3$ alcanzan hasta 80 MPa; $<200 \text{ kg/m}^3$ raramente superan 40 MPa. Dispersión considerable indica que otros factores influyen.

Insights principales:

- El concreto es un sistema multifactorial: la resistencia depende de interacciones entre componentes, no de variables individuales
- Aditivos modernos son cruciales: permiten reducir agua sin perder trabajabilidad
- Relación agua/cemento más determinante que cantidades absolutas

Implicaciones prácticas:

- No confiar en fórmulas univariadas para diseño de mezclas
- Optimizar relación agua/cemento es prioritario
- Control de calidad debe ser multivariable
- Validación experimental necesaria antes de aplicación en obra

Recomendaciones

Mejora del análisis:

- Desarrollar modelo de regresión múltiple (R^2 estimado >0.70)
- Crear variable relación agua/cemento como predictor directo
- Explorar modelos no lineales
- Analizar términos de interacción entre variables



Conclusión

El cemento es un predictor importante pero insuficiente por sí solo para determinar la resistencia del concreto. Existe relación positiva moderada ($r=0.46$), pero un modelo basado únicamente en cemento explica solo 21% de la variabilidad.

Los hallazgos confirman que el concreto es un sistema complejo donde las interacciones entre componentes son más determinantes que cantidades absolutas. Para aplicaciones prácticas en ingeniería civil, se recomienda enfáticamente el desarrollo de modelos de regresión múltiple que consideren todas las variables disponibles, lo cual podría mejorar sustancialmente la capacidad predictiva y proporcionar herramientas confiables para diseño y control de calidad.