

Informe Taller 2 Análisis Univariado

Elaborado por:

Holman Sanchez

Esteban Oliveros

Jesús Córdoba

1. Introducción y selección de la variable

Para este análisis se trabajó con la columna **porcentaje**, creada a partir del cruce entre la base de datos de **contratos** y la de **facturas**. Su cálculo refleja el nivel de ejecución financiera de cada contrato, es decir, el porcentaje del valor efectivamente pagado frente al valor inicialmente estimado.

Esta variable es altamente relevante porque permite identificar:

- **Sobre ejecuciones** (cuando el pago supera lo estimado).
- **Subejecuciones** (cuando se paga menos de lo proyectado).
- **Errores de registro** o inconsistencias en la información.

Además, se realizaron pasos previos de **limpieza de datos**, como el tratamiento de **fechas** (para estandarizar formatos y asegurar la correcta comparación temporal de contratos y facturas) y la **eliminación de valores nulos (NA)** en columnas clave, lo que permitió un análisis más confiable.

2. Análisis univariado de porcentaje

2.1 Estadísticas descriptivas

Después de limpiar el dataset y eliminar un valor atípico extremo (96.000%), se obtuvieron los siguientes resultados:

Medida	Valor
Recuento (n)	581
Media	85,12
Mediana	100,0
Desviación estándar	43,95
Mínimo	0,0
25% (Q1)	65,14
50% (Q2)	100,0
75% (Q3)	100,0
Máximo	273,84

2.2 Visualización

El **boxplot** (Figura 1) muestra que la mayoría de los valores de porcentaje se concentran alrededor del 100%, con algunos casos de **sobreejecuciones** (por encima de 150%) y **subejecuciones** (cercanas a 0%).

Figura 1. Boxplot de la variable porcentaje después de depuración de outliers.

2.3 Hallazgos clave

- La **mediana en 100%** indica que la mitad de los contratos fueron ejecutados exactamente en lo estimado.
- La **media en 85%** revela que, en promedio, los contratos se ejecutan por debajo de lo proyectado.
- Existen **outliers** algunos contratos sobrepasan el 200%, lo que sugiere pagos superiores al valor inicial.
- También hay registros con ejecución cercana a 0%, lo cual puede deberse a contratos sin avance o cancelados.

3. Interpretación de los resultados

El análisis muestra que la variable porcentaje es crítica para comprender la eficiencia y transparencia en la ejecución contractual:

- Los valores cercanos al 100% reflejan cumplimiento esperado.
- Las desviaciones significativas (muy por debajo o por encima) señalan **anomalías operativas o administrativas** que requieren revisión.
- La presencia de outliers, incluso tras depuración, sugiere la necesidad de controles adicionales en la validación de facturación y en la gestión financiera de los contratos.

4. Conclusiones

1. La creación de la variable porcentaje permitió vincular pagos reales con estimaciones contractuales, aportando una **visión clara de la ejecución financiera**.
2. Se identificó que la mayoría de contratos alcanzan el 100% de ejecución, aunque en promedio tienden a subejecutarse.
3. Los outliers detectados son indicadores de posibles **errores de registro, sobre ejecuciones o contratos ineficientes**, y su corrección mejora la calidad analítica.
4. Como próximos pasos, se recomienda:
 - Realizar un análisis bivariado, relacionando porcentaje con variables como **sector o modalidad de contratación**.
 - Implementar reglas de negocio para la detección automática de anomalías.
 - Explorar la distribución temporal del porcentaje (por año o por trimestre) para analizar la evolución de la ejecución de contratos.