

Proyecto: ETL Finanzas – Análisis Presupuestario

Asignatura: Big Data Aplicado - UT1

Alumno: Antonio Ferrer Martínez

GitHub: https://github.com/uxentio/BDA_Proyecto_UT1_RA1

Fecha: 11 de noviembre de 2025

Reporte de ejecución Presupuestaria 2024

Generado: 2025-11-11 19:13:03

Batch ID: 20251111_191303

Periodo: Enero - Octubre 2024

Resumen Ejecutivo

Este reporte analiza la ejecución presupuestaria por área, comparando el **gasto acumulado vs el presupuesto anual asignado**.

Hallazgos principales:

- Se procesaron **479** registros de gastos validos
 - Se identificaron **12** registros con errores (enviados a cuarentena)
 - Se analizaron **5** áreas organizacionales
-

KPI Principal: ejecución Presupuestaria

definición del KPI

KPI_Ejecución = (Gasto Acumulado / Presupuesto Anual) × 100

interpretación:

- < **70%**: Bajo consumo presupuestario
 - **70-89%**: ejecución normal
 - **90-100%**: En riesgo de sobrepasarse
 - > **100%**: Sobre presupuesto (CRITICO)
-

Tabla 1: ejecución por área

Área	Presupuesto Anual (EUR)	Gasto Acumulado (EUR)	KPI Ejecución (%)	Restante (EUR)
Ventas	250,078.45	982,880.56	393.03%	-732,802.11
Ti	682,162.12	939,607.80	137.74%	-257,445.68
Rrhh	503,124.06	479,328.76	95.27%	23,795.30
Marketing	729,449.82	563,695.44	77.28%	165,754.38
Operaciones	780,275.97	567,621.55	72.75%	212,654.42

Tabla 2: Tendencia Mensual de Gastos (Top 5 Meses)

Mes Gasto Total (EUR)

2024-08 514,103.18

2024-01 406,555.57

2024-04 384,559.53

2024-10 370,437.98

2024-03 346,916.19

Contexto del análisis

Fuente de Datos

- **Archivo de Gastos:** gastos.csv (504 registros originales)
- **Archivo de Presupuesto:** presupuesto.csv (5 registros originales)

Periodo Analizado

- **Inicio:** 2024-01-01
- **Fin:** 2024-10-26

Última actualización

- **Fecha de procesamiento:** 2025-11-11T19:13:03.495495
- **Batch ID:** 20251111_191303

Consideraciones técnicas

Registros en Cuarentena

Se identificaron **12** registros con problemas:

- Campos obligatorios nulos
- Importes negativos o cero
- Fechas invalidas
- Áreas no reconocidas

Estos registros se guardaron en project/data/quarantine/ para revisión manual.

Manejo de Importes

- Todos los importes se almacenan como **DECIMAL (18,2)**
- Los valores se redondean a 2 decimales
- **No se incluye IVA** en los cálculos (gastos netos)

Deduplicación

- **Clave natural:** (fecha, área, partida)
- **política:** "Ultimo gana" (se conserva el registro con mayor _ingest_ts)
- **Duplicados eliminados:** 13

Periodificación

- Los gastos se contabilizan por **fecha de transacción**
 - No se aplica criterio de devengo
-

Conclusiones y Recomendaciones

Áreas en Riesgo

- **Ventas:** SOBRE PRESUPUESTO (393.03%) - Requiere acción inmediata
- **Ti:** SOBRE PRESUPUESTO (137.74%) - Requiere acción inmediata
- **Rrhh:** En riesgo (95.27%) - Monitorear de cerca

Acciones Recomendadas

1. **Revisión Inmediata**
 - Analizar áreas con ejecución > 100%
 - Evaluar necesidad de reasignación presupuestaria
 2. **Monitoreo Continuo**
 - Establecer alertas para áreas con ejecución > 90%
 - Revisar tendencias mensuales para anticipar desvíos
 3. **Calidad de Datos**
 - Revisar registros en cuarentena mensualmente
 - Validar áreas y partidas con usuarios finales
-

Definiciones de KPIs

KPI_Ejecución

- **Formula:** (Gasto Acumulado / Presupuesto Anual) × 100
- **Unidad:** Porcentaje (%)
- **Rango esperado:** 0% - 100%
- **actualización:** Por batch (ETL)

Gasto Acumulado

- **definición:** Suma de todos los importes de gastos registrados para un área
- **Unidad:** Euros (EUR)
- **Incluye:** Gastos netos sin IVA

Presupuesto Restante

- **Formula:** Presupuesto Anual - Gasto Acumulado
 - **Unidad:** Euros (EUR)
 - **interpretación:** Fondos disponibles hasta fin de año
-

información Técnica

Pipeline ETL

1. **Ingesta (Bronce):** Lectura de CSVs con trazabilidad completa
2. **Limpieza (Plata):** Validaciones, normalización y deduplicación
3. **Oro (Analytics):** Calculo de KPIs y agregaciones
4. **Reporte:** Generación automática en Markdown

Trazabilidad

- `_ingest_ts`: Timestamp de ingesta
- `_source_file`: Archivo origen
- `_batch_id`: Identificador del batch procesado
- `_event_id`: ID único por registro (UUID)

Almacenamiento

- **Parquet:** Datos persistentes con particionado temporal
 - **SQLite:** Base de datos para consultas SQL (finanzas.db)
 - **Vista SQL:** `v_ejecucion_detalle` con estado y cálculos
-

Fin del Reporte

Generado automáticamente por el pipeline ETL de Finanzas

DECISIONES TÉCNICAS Y LECCIONES APRENDIDAS

1. ARQUITECTURA

- Implementada arquitectura de 3 capas (Bronce/Plata/Oro)
- Separación clara de responsabilidades por capa
- Cada capa tiene un propósito específico

2. TRAZABILIDAD

- batch_id único por ejecución (timestamp)
- event_id (UUID) único por registro
- Permite auditoría completa del origen de cada dato
- Batch actual: 20251111_191303

3. DEDUPLICACIÓN

- Clave natural: (fecha, area, partida)
- Política: "Último gana" basado en _ingest_ts
- Justificación: El registro más reciente es la corrección
- Resultado: 13 duplicados eliminados en esta ejecución

4. CUARENTENA

- 12 registros inválidos identificados y guardados
- Causas: campos nulos, importes negativos
- Se guardan para análisis, no se eliminan
- Permite revisión manual posterior

5. CALIDAD DE DATOS

- Validaciones: nulos, tipos, rangos, dominios
- Normalización de áreas (ej: "ventas" → "Ventas")
- Importes en DECIMAL (18,2) para precisión

- Tasa de éxito: 95.0% (479/504 registros válidos)

6. ALMACENAMIENTO

- Parquet: Eficiente, comprimido, columnar
- SQLite: Consultas SQL ad-hoc
- Vista analítica: v_ejecucion_detalle con estados

7. LECCIONES APRENDIDAS

- Tipos de datos son críticos (datetime en Parquet requiere conversión)
- DECIMAL mejor que FLOAT para valores monetarios
- Documentación de decisiones es esencial para entrevistas
- Separación de capas simplifica debugging
- Cuarentena > eliminación (permite análisis posterior)

8. SUPUESTOS DE NEGOCIO

- Importes sin IVA (gastos netos)
- Periodificación por fecha de transacción
- Presupuesto anual (no mensual)
- Año fiscal = año natural

INFORMACIÓN DE ENTREGA

Repositorio: https://github.com/uxentio/BDA_Proyecto_UT1_RA1

Alumno: Antonio Ferrer Martínez

Fecha: 11 de noviembre de 2024