

Reto Final - "*De los datos a la decisión: Proyecto Analítico Integrador*"

Descripción general

Este reto integrador articula los aprendizajes de los cursos **Fundamentals Data**, **Booster Python** y **Procesamiento de datos**, simulando un flujo completo de trabajo en ciencia de datos: desde la definición de una problemática hasta la visualización de insights para la toma de decisiones en Power BI.

Objetivo

Que los estudiantes:

- Identifiquen una problemática relevante.
- Generen y administren **dos fuentes de datos** (mínimo: 1 archivo Excel + 1 base MySQL).
- Extraigan, limpien y transformen datos con **Python** y **SQL**.
- Integren y analicen información aplicando criterios estadísticos básicos.
- Construyan un **dashboard en Power BI** con visualizaciones y KPIs útiles para la toma de decisiones.
- Presenten hallazgos y propuestas de acción en un **informe ejecutivo** breve.

Entregables

1. **Descripción de la problemática** (1–2 páginas).
2. **Fuentes de datos:**
 - a. Archivo **Excel** (.xlsx) con datos simulados y diccionario de datos.
 - b. Esquema y **dump** de **MySQL** (.sql) con tablas y registros.
 - c. API Pública. Eje: <https://api-colombia.com/>
3. **Código en Python** (Notebook .ipynb o script .py) que:
 - a. Conecte a MySQL y lea Excel.
 - b. Limpie y transforme datos.
 - c. Integre datasets y genere tablas finales para Power BI (CSV o tabla en MySQL).
4. **Dashboard Power BI** (.pbix) con:
 - a. 3–5 visualizaciones clave.
 - b. 2–3 KPIs.
 - c. Segmentadores y filtros útiles.

5. **Informe ejecutivo** (2–3 páginas): contexto, metodología, hallazgos, decisiones sugeridas, limitaciones.
6. **Presentación oral** (8–10 minutos).

Alcance mínimo

- **Volumen de datos:** ≥ 500 registros totales y ≥ 8 columnas relevantes.
- **SQL:** ≥ 3 consultas (JOINS, agregaciones, filtros) + creación de tablas.
- **Python:** manejo de nulos, tipos, duplicados; creación de variables derivadas; al menos 2 agregaciones y 1 visualización exploratoria (opcional).
- **Power BI:** al menos 1 gráfico de tendencia/serie temporal, 1 comparativo, 1 mapa o matriz (según caso), y 2 KPIs.

Criterios de evaluación

1. **Problema y contexto (15 pts)**
Claridad, relevancia, objetivos y métricas definidas.
2. **Calidad de datos (20 pts)**
Diseño coherente de fuentes, diccionario de datos, integridad y consistencia.
3. **Procesamiento (Python + SQL) (25 pts)**
Limpieza, transformaciones, buenas prácticas, reproducibilidad.
4. **Análisis y métricas (15 pts)**
Indicadores útiles, estadística básica, trazabilidad de resultados.
5. **Dashboard Power BI (20 pts)**
Diseño, claridad, interactividad, narrativa visual, utilidad para decisiones.
6. **Informe y presentación (5 pts)**
Síntesis ejecutiva, comunicación, conclusiones y acciones.

Hitos y fechas

- **Sesión 1:** elección de problemática y diseño de fuentes.
- **Sesión 2:** creación de Excel y schema MySQL; primer ETL en Python.
- **Sesión 3:** integración y dataset final; prototipo de dashboard.
- **Sesión 4:** refinamiento, informe y presentación.

Temas sugeridos

- Retail: ventas, inventario, devoluciones y satisfacción de clientes.
- Educación: rendimiento por curso, asistencia y recursos.

- Salud: citas, tiempos de espera, satisfacción y costos.
- Logística: tiempos de entrega, rutas, costos y quejas.
- Marketing: campañas, leads, conversiones y ROI.