

Prueba Técnica HiloTools — Data

Convocatoria interna HiloTools

21 de agosto de 2025

Objetivo general

Construir un pipeline de datos **end-to-end** a partir de los **3 Excel provistos**:

- Ventas
- RRHH
- Inventario

ubicados en: https://drive.google.com/drive/folders/1n2CsVWT1v_zYE8j0dtLJBReMKezGP1i-?usp=sharing

El entregable debe demostrar **capacidad de ingeniería de datos y análisis científico**, integrando los tres dominios en un modelo analítico coherente.

1. Alcance de la prueba

1. **Ingesta**: diseñar un sistema que consuma automáticamente los tres ficheros Excel y los transforme a un formato analítico uniforme.
2. **Modelo estrella**: proponer y materializar un **esquema en estrella** con al menos:
 - Una tabla de hechos de ventas.
 - Dimensiones para clientes, productos, fechas, empleados y almacenes.
3. **Analítica PCA**: generar un conjunto de *features* relevantes y aplicar un análisis de componentes principales para identificar patrones latentes entre ventas, desempeño de empleados y stock de inventarios.

2. Restricciones

- Todo el pipeline debe ser **reproducible** en un entorno local (Docker o Conda).
- Los ficheros originales no pueden modificarse; las transformaciones deben realizarse en la capa de procesamiento.
- Se espera tolerancia a datos incompletos, decimales con coma, duplicados y diferencias de formato.
- El entregable debe estar documentado en un **repositorio GitHub** con estructura clara (`/ingest`, `/model`, `/analytics`, `/report`).

3. Plan de trabajo

Análisis exploratorio y definición de diccionarios de datos.

Diseño del esquema en estrella y normalización de columnas. de datos en el modelo (fact + dimensiones). ón de features integrando ventas, RRHH e inventarios.

Ejecución de PCA y selección de componentes explicativos.

Pruebas de consistencia, documentación final y entrega en GitHub.

4. Criterios de evaluación

1. **Ingeniería de datos (40 %)**: solidez de la ingesta y modelado en estrella.
2. **Analítica (30 %)**: calidad del PCA, justificación de variables y hallazgos.
3. **Documentación (10 %)**: estructura del repositorio, claridad del informe LaTeX y reproducibilidad.

5. Entregables

- Repositorio GitHub con:
 - Código fuente.
 - README.md con instrucciones de instalación/ejecución.
 - Informe técnico en $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ (PDF) o markdown.

Nota final

No se evaluará únicamente la corrección técnica, sino también la **capacidad de estructurar un proyecto de datos de forma profesional**, con buenas prácticas de ingeniería y documentación.