Tarea #2  
Proceso ETL: Extracción, Transformación y Carga básica

línea corta

**PONDERACIÓN: 2 puntos**

**Horas Aproximadas: 4 horas**

Universidad San Carlos de Guatemala

Facultad de ingeniería.

Ingeniería en ciencias y sistemas

# Introducción

El proceso denominado como ETL (extracción, transformación y carga) es clave para el correcto desarrollo de soluciones de inteligencia de negocios, esto debido a que los cubos OLAP los data warehouse y otras partes de este proceso requieren datos estructurados y normalizados, los cuales el estudiante utilizando las bases del proceso ETL y sus fases podrá controlar y lograr una correcta manipulación de los datos.

# Competencia(s)

# Implementa procesos ETL eficientes empleando herramientas como SSIS, Pentaho o Talend asegurando la transformación y carga efectiva de datos para su posterior análisis

# Contenido

# El estudiante realizará un pequeño proceso ETL usando datos públicos (CSV) de su elección: realizará la extracción, aplicará una limpieza o transformación simple (normalización de campos, eliminación de datos nulos) y cargará el resultado en un entorno como Excel o Power BI. Entregará evidencia (capturas y breve informe).

# Recursos Adicionales

Fuentes de datos para la tarea: https://www.kaggle.com/

# Forma de entrega

Enlace del repositorio del curso nombrado “SS2\_2S2025\_#carné” en la carpeta Tarea\_2 con un archivo README con toda la documentación requerida y los archivos necesarios (código, imágenes, JSON, XML) dependiendo del criterio del estudiante.

# Fecha de entrega

Fecha límite para realizar la entrega de la tarea

# Rúbrica de Calificación

|  |  |  |
| --- | --- | --- |
| **Criterio** | **Descripción** | **Puntos Máximos** |
| **Extracción de datos** | Selección de una fuente de datos válida y carga correcta del archivo CSV. Evidencia de extracción clara (código, screenshot, fuente). | 15 pts |
| **Transformación aplicada** | Aplicación de al menos una transformación significativa: normalización, limpieza de nulos, tipificación de datos, etc. | 25 pts |
| **Carga de datos** | Exportación o carga del resultado en una herramienta (Excel, Power BI, base de datos, etc.). Evidencia clara de la carga. | 15 pts |
| **Uso de herramienta ETL o código** | Uso de un entorno como SSIS, Talend, Pentaho o Python/pandas. Se valora ejecución y uso correcto, aunque sea simple. | 15 pts |
| **Informe/documentación técnica (README)** | Breve explicación del proceso (qué hizo en cada fase), con capturas y estructura clara. | 20 pts |
| **Formato de entrega** | Repositorio bien nombrado, carpeta correcta, archivos legibles y bien organizados. | 5 pts |
| **Claridad y presentación** | Ortografía, presentación, formato legible y comprensible del informe. | 5 pts |

# Valores

|  |  |  |
| --- | --- | --- |
| **Valor** | **Descripción esperada en la tarea** | **Consecuencia si no se cumple** |
| **Organización** | Evidencia clara y ordenada del proceso ETL (fase por fase, en carpetas o imágenes bien etiquetadas). | ❌ Hasta **-10 puntos** por entregas confusas, desordenadas o poco legibles. |
| **Responsabilidad** | Entrega completa, dentro del plazo, y con cumplimiento de los pasos requeridos (extracción, transformación, carga). | ❌ **0 puntos** si el proceso ETL no está desarrollado o está vacío. |
| **Disciplina** | Muestra de que el proceso fue realizado paso a paso, con constancia, sin saltarse fases ni soluciones improvisadas. | ❌ Hasta **-15 puntos** si no hay orden lógico o hay ausencia de evidencia. |
| **Pensamiento lógico** | Las transformaciones aplicadas deben tener sentido (por qué normalizar, qué campos eliminar, etc.). | ❌ Hasta **-10 puntos** por procesos incoherentes o sin justificación. |