

Universidad San Carlos de Guatemala

Facultad de ingeniería.

Ingeniería en ciencias y sistemas

Tarea #1. —

Limpieza y análisis inicial de datos con Python y Pandas



FIUSAC
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



PONDERACIÓN: 1.66 ptos

 **Tiempo estimado: 2 hrs**
1. Marco Formativo
1.1. Valores

Nombre del valor	¿Cómo se aplica en tu laboratorio?
Responsabilidad académica y profesional	Cumplimiento de lineamientos, entrega puntual y documentación clara del análisis realizado.

1.2. Competencia(s)

Tipo de Competencia	General y específicas
Competencia General	Comunica resultados de análisis de datos utilizando diversos formatos y lenguajes dirigidos a diferentes audiencias profesionales y sociales
Competencia Específica	<ul style="list-style-type: none"> Analiza datos estructurados y no estructurados utilizando procesos de preparación, transformación y exploración para identificar patrones, relaciones y oportunidades de mejora Evaluá a calidad y pertinencia de la información empleando criterios de validez, confiabilidad y seguridad para fundamentar la toma de decisiones estratégicas

1.3. Objetivo SMART

SMART	Definición	Objetivo redactado
Específico (¿Qué?)	El objetivo es concreto y tangible.	Importar un dataset asignado desde Kaggle en un entorno Python (Jupyter Notebook/VS Code/Colab) usando Pandas, aplicar un proceso completo de limpieza de datos que incluya eliminación de duplicados, tratamiento de valores faltantes y estandarización de formatos, y generar visualizaciones que permitan analizar relaciones entre las variables.
Medible (¿Cuánto?)	El objetivo tiene una medida objetiva de éxito.	Se considerará cumplido al obtener un notebook (.ipynb) o script .py con las transformaciones aplicadas, el dataset limpio exportado (CSV/Parquet).
Alcanzable (¿Cómo?)	El objetivo debe ser posible con los recursos disponibles.	El objetivo se logrará utilizando Python 3 y la librería Pandas (en Jupyter Notebook/VS Code/Colab) y el dataset proporcionado, siguiendo técnicas básicas de limpieza y exploración de datos vistas en clase.
Realista (¿Para qué?)	El objetivo contribuye a metas más amplias.	El objetivo es contribuir al desarrollo de competencias de análisis, calidad y comunicación de datos, relevantes para procesos de toma de decisiones dentro del ámbito profesional.

A Tiempo (¿Cuándo?)	El objetivo tiene fecha límite o mejor aún un cronograma de hitos de progreso	El objetivo deberá completarse dentro del período establecido para la actividad en el semestre 2026, según el cronograma del curso de Seminario de Sistemas 2.
--------------------------------	---	--

2. Material de Apoyo

- Kaggle. Repositorio de datasets para análisis. <https://www.kaggle.com/datasets>
- Normas Básicas para validación de calidad según ISO 25000.
<https://www.iso25000.com/index.php/normas-iso-25000/iso-25012>

3. Actividad

El estudiante deberá crear un repositorio digital con el nombre **SS21S2026_#carnet**. Dentro de este repositorio se generará una carpeta denominada **Tarea1**, donde se incluirán todos los archivos correspondientes a esta actividad.

En la carpeta **Tarea1** deberán entregarse los siguientes elementos:

1. Notebook de Python (.ipynb) o script (.py) con el dataset asignado desde Kaggle, mostrando las transformaciones realizadas:
 - Eliminación de duplicados
 - Tratamiento de celdas vacías
 - Estandarización de valores y formatos
2. Un archivo **README.md** que documente de forma clara y ordenada:
 - Nombre del dataset utilizado
 - Descripción breve del proceso de limpieza aplicado
 - Capturas de pantalla de las tablas o visualizaciones generadas
 - Interpretación concisa de los resultados obtenidos

El repositorio deberá mantenerse organizado y con los elementos correctamente nombrados para facilitar la revisión y calificación de la actividad.

4. Cronograma

Asignación de tarea	Semana 3
Entrega de Tarea	Semana 4

5. Rúbrica de calificación

CRITERIO	DESCRIPCIÓN	PUNTEO
Preparación y limpieza de datos	El notebook/script en Python demuestra correctamente la eliminación de duplicados, tratamiento de valores faltantes y estandarización adecuada de datos.	35 ptos.
Exploración mediante tablas tipo pivot (Pandas)	Se presenta un dataframe con los datos antes del tratamiento/limpieza (estado original) y un dataframe con los datos después del tratamiento/limpieza (estado depurado), acompañados de interpretaciones coherentes que evidencian los cambios aplicados y su impacto en la calidad de la información.	30 ptos.
Documentación en README.md	El README.md presenta la actividad de forma clara, incluye descripción del proceso, capturas de pantalla e interpretaciones de resultados.	20 ptos.
Organización y entrega	El repositorio sigue el formato solicitado, estructura correcta de carpetas y nomenclatura adecuada. Entrega puntual.	15 ptos.
TOTAL		100 ptos.

■ ■ ■