CTIC UNI

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

Programa de Iniciación Tecnológica PIT - 2025

CERTIFICADO

Otorgado a

JEAN CARLOS ZUBIATE HUAMANI

Por haber aprobado el curso de

CIENCIA DE DATOS 1: EXPLORATORY DATA ANALYSIS + IA

Realizado del 25 de octubre al 22 de noviembre del presente, con una duración de 18 horas.

Lima, noviembre de 2025

Mag. Ing. Rubén Arturo Borja Rosales
JEFE OTI - UNI

N° Certificado: 017 - 0132315

FIRMA DIGITAL

Firmado digitalmente por: BORJA ROSALES Ruben Arturo FAU 20169004359 hard
Motivo: En señal de conformidad
Fecha: 23/12/2025 23:27:54-0500

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

CTIC UNI

Temario del Curso:

CIENCIA DE DATOS 1: EXPLORATORY DATA ANALYSIS + IA

- Fundamentos del EDA.
- Rol del EDA en el ciclo de datos y objetivos.
- Tipos de variables y problemas comunes.
- Herramientas base (Python, pandas, matplotlib).
- Tipos y conversiones “trampa” (Fechas, strings numéricos, nulos encubiertos).
- Buenas prácticas de reproducibilidad (Semillas, notebooks limpios, control de rutas).
- Descriptivas, visualización y distribución de clases.
- Medidas de tendencia central y dispersión.
- Histogramas, boxplots, KDE y visualización responsable.
- Distribución de clases y diagnóstico de desbalance.
- Segmentación exploratoria (Cohortes, región, canal).
- Partición estratificada preliminar.
- Relaciones y EDA por tipo de datos.
- Correlaciones (Pearson, Spearman, Kendall).
- Mapas de calor y relaciones numéricas ↔ categóricas.
- Series temporales: Estacionalidad, re-sampling, ventanas rodantes.
- Mini-EDA de texto y geodatos.
- Calidad y preparación de datos.
- Valores faltantes: MCAR, MAR, MNAR e implicancias.
- Outliers robustos: IQR, MAD, winsorización.
- Duplicados y unicidad de llaves.
- Normalización, estandarización y transformaciones (Log, Box-Cox).
- Consolidación de tipos y unidades.
- Feature Engineering y selección.
- Creación de variables: Ratios, binning, agregaciones.
- Selección por varianza, correlación y redundancia.
- Manejo de categóricas de alta cardinalidad.
- PCA introductorio como puente a modelado.
- Riesgos de fuga de información (Data leakage).
- Integración y cierre.
- Validación de calidad y linaje de datos.
- Evaluación de desbalance y partición estratificada final.
- Drift preliminar: Comparación train vs. muestra reciente.
- Storytelling de hallazgos y handoff a modelado.
- Reproducibilidad: Notebooks → scripts, control de versiones.

Nota final: **20 (Veinte)**

N° Certificado: 017 - 0132315