CTIC UNI

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

Programa de Iniciación Tecnológica PIT - 2026

CERTIFICADO

Otorgado a

DAVID DANIEL ROMERO SUXE

Por haber aprobado el curso de

CIENCIA DE DATOS 1: EXPLORATORY DATA ANALYSIS

Realizado del 20 de mayo al 01 de junio del presente, con una duración de 18 horas.

Lima, junio de 2026

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

N° Certificado: 018 - 0030597

FIRMA DIGITAL

Firmado digitalmente por: BORJA ROSALES Ruben Arturo FAU 20168004359 hard
Motivo: En señal de conformidad
Fecha: 11/08/2026 00:27:49-0500

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

CTIC UNI

Temario del Curso:

CIENCIA DE DATOS 1: EXPLORATORY DATA ANALYSIS

- 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: 018 - 0030597