UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMÁTICAS

SÍLABO

1. INFORMACIÓN GENERAL

1.1 Curso	:	Análisis exploratorio de datos
1.2 Código	:	CS253
1.3 Programa Académico	:	Computación Científica
1.4 Créditos	:	5
1.5 Prerrequisitos	:	MA351. Estadística y Probabilidades. (6^{to} Sem)
1.6 Tipo de curso	:	Electivo
1.7 Área de Estudios	:	Estudios Generales
1.8 Número de Semanas	:	16
1.9 Horas por semana	:	3 HT; 4 HP;
1.10 Ciclo	:	I
1.11 Modalidad de aprendizaje	:	Presencial
1.12 Periodo Académico	:	2026-I
1.13 Inicio del semestre	:	
1.14 Fin del semestre	:	

2. PROFESORES

Atención previa coordinación con el profesor

3. INTRODUCCIÓN AL CURSO

El Análisis Exploratorio de Datos (EDA) es una etapa crítica en cualquier proyecto de ciencia de datos. Este curso proporciona las herramientas para entender la estructura, detectar anomalías y extraer patrones iniciales de los datos antes del modelado formal. Se enfoca en el uso de estadísticas descriptivas, limpieza de datos y visualización preliminar para validar supuestos y guiar la selección de modelos complejos, asegurando la integridad y calidad de los hallazgos científicos.

4. OBJETIVOS

- Aplicar técnicas de limpieza y preprocesamiento de datos para asegurar su calidad.
- Utilizar métodos estadísticos descriptivos para resumir conjuntos de datos complejos.
- Identificar patrones, valores atípicos y relaciones entre variables mediante análisis visual.
- Formular hipótesis basadas en la exploración inicial de los datos.
- Comunicar hallazgos preliminares de manera efectiva a través de reportes técnicos.

5. RESULTADOS DEL ESTUDIANTE

ABET-3) Habilidad para desarrollar y llevar a cabo experimentos o probar hipótesis, analizar e interpretar datos y utilizar el juicio científico para obtener conclusiones. (Usage)

ABET-2) Habilidad para formular o diseñar un sistema, proceso, procedimiento o programa para el propósito previsto. (Usage)

6. TEMAS

Unidad 1: Fundamentos del Análisis Exploratorio de Datos (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Filosofía y objetivos del análisis exploratorio. ▪ Tipos de datos: estructurados, semi-estructurados y no estructurados. ▪ Escalas de medición: nominal, ordinal, intervalo y razón. ▪ Flujo de trabajo en la exploración de datos.	▪ Identificar la importancia del EDA en el ciclo de vida de los datos [Familiarizarse]. ▪ Clasificar variables según su escala de medición y tipo [Usar].
Lecturas : [**Tukey1977**], [**Peng2016**]	

Unidad 2: Limpieza y Preparación de Datos (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Concepto de "Tidy Dataz manipulación de tablas. ▪ Manejo de valores faltantes: imputación versus eliminación. ▪ Detección y tratamiento de valores atípicos (outliers). ▪ Normalización y estandarización de variables numéricas.	▪ Implementar rutinas de limpieza para conjuntos de datos ruidosos [Usar]. ▪ Evaluar el impacto de diferentes estrategias de imputación [Evaluar].
Lecturas : [**Wickham2014**], [**McKinney2022**]	

Unidad 3: Estadística Descriptiva Univariada (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Medidas de tendencia central, dispersión y forma. ▪ Distribuciones de frecuencia y funciones de densidad. ▪ Análisis de asimetría y curtosis en variables reales. ▪ Visualización univariada: Histogramas, Boxplots y estimación de densidad (KDE).	▪ Resumir numéricamente la distribución de una variable [Usar]. ▪ Interpretar gráficos de densidad para identificar sesgos en los datos [Evaluar].
Lecturas : [**Healy2018**], [**Tukey1977**]	

Unidad 4: Exploración Bivariada y Multivariada (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Correlación de Pearson, Spearman y Kendall. ▪ Tablas de contingencia y medidas de asociación categórica. ▪ Diagramas de dispersión, matrices de correlación y mapas de calor. ▪ Introducción conceptual a la reducción de dimensionalidad (PCA).	▪ Cuantificar la relación entre múltiples variables de un dataset [Usar]. ▪ Utilizar mapas de calor para detectar colinealidad en conjuntos de datos [Evaluar].
Lecturas : [**Healy2018**], [**Wickham2016**]	

Unidad 5: Herramientas Computacionales para Exploración (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Ecosistemas computacionales para análisis (Python/Pandas o R/Tidyverse). ▪ Manipulación eficiente de DataFrames de gran volumen. ▪ Generación automática de reportes de perfilado de datos (Data Profiling).	▪ Implementar procesos de EDA automatizados mediante programación [Usar]. ▪ Optimizar el procesamiento de datos para análisis eficiente [Evaluar].
Lecturas : [**McKinney2022**], [**Wickham2016**]	

Unidad 6: Pruebas de Hipótesis Exploratorias (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Pruebas de normalidad y supuestos de varianza. ▪ Comparación de grupos: T-test y ANOVA en contexto exploratorio. ▪ Métodos no paramétricos y remuestreo (Bootstraping).	▪ Validar supuestos estadísticos necesarios para el modelado posterior [Usar]. ▪ Diseñar experimentos exploratorios para confirmar hallazgos visuales [Evaluar].
Lecturas : [**Peng2016**], [**Healy2018**]	

Unidad 7: Ética de Datos y Responsabilidad (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Identificación de sesgos de selección y confirmación en los datos. ▪ Privacidad de datos: Anonimización y técnicas de enmascaramiento. ▪ Ética en la comunicación de hallazgos preliminares y riesgos asociados.	▪ Identificar posibles sesgos éticos en conjuntos de datos reales [Familiarizarse]. ▪ Aplicar técnicas de anonimización básicas durante el análisis [Usar].
Lecturas : [**Healy2018**]	

Unidad 8: Comunicación de Resultados y Narrativa de Datos (8 horas)	
Resultados esperados: ABET-2,ABET-3	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Estructura de un reporte técnico de calidad de datos. ▪ Narrativa (Storytelling) basada en el descubrimiento exploratorio. ▪ Selección de visualizaciones clave para la toma de decisiones.	▪ Redactar informes técnicos efectivos sobre la calidad de los datos [Evaluar]. ▪ Realizar presentaciones de hallazgos exploratorios ante audiencias técnicas [Evaluar].
Lecturas : [**Wickham2016**], [**Healy2018**]	

7. PLAN DE TRABAJO

7.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

7.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

7.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

8. SISTEMA DE EVALUACIÓN

********* EVALUATION MISSING ********

9. BIBLIOGRAFÍA BÁSICA