NATIONAL UNIVERSITY OF SAN MARCOS

FACULTY OF MATHEMATICAL SCIENCES

SYLLABUS

1. GENERAL INFORMATION

1.1 Course name	:	Exploratory data analysis
1.2 Code	:	CS253
1.3 Academic Program	:	Scientific Computing
1.4 Credits	:	5
1.5 Prerrequisites	:	MA351. Statistics and Probabilities. (6^{th} Sem)
1.6 Type of Course	:	Elective
1.7 Area of Studies	:	Estudios Generales
1.8 Number of Weeks	:	16
1.9 Hours per week	:	3 HT; 4 HP;
1.10 Cycle	:	I
1.11 Learning modality	:	Face to face
1.12 Academic Period	:	2026-I
1.13 Semester Start Date	:	
1.14 Semester End Date	:	

2. PROFESSORS

Meetings after coordination with the professor

3. INTRODUCTION TO THE COURSE

Exploratory Data Analysis (EDA) is a critical stage in any data science project. This course provides the tools to understand the structure, detect anomalies, and extract initial patterns from data before formal modeling. It focuses on the use of descriptive statistics, data cleaning, and preliminary visualization to validate assumptions and guide the selection of complex models, ensuring the integrity and quality of scientific findings.

4. GOALS

- Apply data cleaning and preprocessing techniques to ensure data quality.
- Use descriptive statistical methods to summarize complex datasets.
- Identify patterns, outliers, and relationships between variables through visual analysis.
- Formulate hypotheses based on the initial exploration of the data.
- Communicate preliminary findings effectively through technical reports.

5. COMPETENCES

ABET-3) An ability to develop and conduct experiments or test hypotheses, analyze and interpret data, and use scientific judgment to draw conclusions. (Usage)

ABET-2) An ability to formulate or design a system, process, procedure, or program for the intended purpose. (Usage)

6. TOPICS

Unit 1: Foundations of Exploratory Data Analysis (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Philosophy and objectives of exploratory analysis. • Data types: structured, semi-structured, and unstructured. • Measurement scales: nominal, ordinal, interval, and ratio. • Workflow in data exploration.	• Identify the importance of EDA in the data life cycle [Familiarizarse]. • Classify variables according to their measurement scale and type [Usar].
Readings : [**Tukey1977**], [**Peng2016**]	

Unit 2: Data Cleaning and Preparation (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• "Tidy Data" concept and table manipulation. • Handling missing values: imputation versus removal. • Detection and treatment of outliers. • Normalization and standardization of numerical variables.	• Implement cleaning routines for noisy datasets [Usar]. • Evaluate the impact of different imputation strategies [Evaluar].
Readings : [**Wickham2014**], [**McKinney2022**]	

Unit 3: Univariate Descriptive Statistics (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Measures of central tendency, dispersion, and shape. • Frequency distributions and density functions. • Skewness and kurtosis analysis in real variables. • Univariate visualization: Histograms, Boxplots, and Kernel Density Estimation (KDE).	• Numerically summarize the distribution of a variable [Usar]. • Interpret density plots to identify biases in the data [Evaluar].
Readings : [**Healy2018**], [**Tukey1977**]	

Unit 4: Bivariate and Multivariate Exploration (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Pearson, Spearman, and Kendall correlation. • Contingency tables and measures of categorical association. • Scatter plots, correlation matrices, and heatmaps. • Conceptual introduction to dimensionality reduction (PCA).	• Quantify the relationship between multiple variables in a dataset [Usar]. • Use heatmaps to detect collinearity in datasets [Evaluar].
Readings : [**Healy2018**], [**Wickham2016**]	

Unit 5: Computational Tools for Exploration (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Computational ecosystems for analysis (Python/Pandas or R/Tidyverse). • Efficient manipulation of large-volume DataFrames. • Automatic generation of data profiling reports.	• Implement automated EDA processes through programming [Usar]. • Optimize data processing for efficient analysis [Evaluar].
Readings : [**McKinney2022**], [**Wickham2016**]	

Unit 6: Exploratory Hypothesis Testing (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Normality tests and variance assumptions. • Group comparison: T-test and ANOVA in an exploratory context. • Non-parametric methods and resampling (Bootstrapping).	• Validate statistical assumptions necessary for subsequent modeling [Usar]. • Design exploratory experiments to confirm visual findings [Evaluar].
Readings : [**Peng2016**], [**Healy2018**]	

Unit 7: Data Ethics and Responsibility (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Identification of selection and confirmation biases in data. • Data privacy: Anonymization and masking techniques. • Ethics in communicating preliminary findings and associated risks.	• Identify potential ethical biases in real-world datasets [Familiarizarse]. • Apply basic anonymization techniques during analysis [Usar].
Readings : [**Healy2018**]	

Unit 8: Result Communication and Data Storytelling (8 hours)	
Competences Expected: ABET-2,ABET-3	
Topics	**Learning Outcomes**
• Structure of a technical data quality report. • Storytelling based on exploratory discovery. • Selection of key visualizations for decision-making.	• Write effective technical reports on data quality [Evaluar]. • Perform presentations of exploratory findings to technical audiences [Evaluar].
Readings : [**Wickham2016**], [**Healy2018**]	

7. WORKPLAN

7.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

7.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

7.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

8. EVALUATION SYSTEM

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

9. BASIC BIBLIOGRAPHY