

Course Syllabus 2024 Edition
Exploratory Data Analysis
Specialization in Economics, Data Science Option

Instructor(s) and Email:

Guillermo Lezama, guillermo.lezama@cienciassociales.edu.uy

Credits: 4

Course Format: In-person / Hybrid

Estimated Workload and Distribution of Hours:

20 in-person hours and 40 hours of independent work.

Recommended Prerequisites:

Basic Python, Jupyter Notebook, Basic Statistics

Course Content and Structure:

Part 1: Descriptive Statistics and Data Preparation

- Introduction and principles of EDA.
- Descriptive statistics and data visualization.
- Data cleaning and preparation techniques.
- Using SQL for data analysis.

Part 2: Variable Handling and Transformation

- Creation, modification, and deletion of variables.
- Feature engineering.
- Exploratory graphical analysis.
- Group comparison to identify disparities.

Part 3: Advanced EDA Techniques

- Correlation analysis and exploratory statistics.
- Dimensionality reduction for exploration.
- Clustering to describe data.
- EDA applied to time series and text data analysis.
- Hypothesis generation and testing based on EDA.

Teaching Method:

Theoretical classes with practical applications. Homework assignments for each session.

Evaluation System:

- Final Project (60%): Apply EDA techniques to a specified dataset and present the findings.
- Classwork (40%): Practical exercises conducted during the course.

Recommended Bibliography:

No specific textbook will be followed, but these books will be referenced for each unit:

- *"Python for Data Analysis"* by Wes McKinney
- *"Python Data Science Handbook"* by Jake VanderPlas
- *"Learning SQL"* by Alan Beaulieu
- *"Practical Statistics for Data Scientists"* by Peter Bruce and Andrew Bruce
- *"Introduction to Time Series Forecasting with Python"* by Jason Brownlee
- *"The Effect: An Introduction to Research Design and Causality"* by Nick Huntington-Klein, Chapters 3 and 4