ITAI 2373 – Natural Language Processing
Professor: Narges DeBary
Student: Natalia Solórzano Perez W207818526
Spring 2025 CN: 16579

## Assignment 03: PANDAS

Pandas is a widely used open-source Python library that provides efficient tools for handling and analyzing structured data. It simplifies tasks related to organizing, cleaning, and transforming data, making it a crucial tool in the data science field.

At its core, Pandas introduces two primary data structures: Series and DataFrame. A Series is a one-dimensional labeled array, while a DataFrame is a two-dimensional table that allows for easy manipulation of data with labeled rows and columns. These structures make it possible to perform operations such as filtering, sorting, and reshaping datasets.

One of Pandas' strengths is its ability to clean data efficiently. It provides functions to handle missing values, remove duplicates, and format data correctly. Additionally, it allows for data transformation through operations like applying custom functions to modify datasets. The library also supports data aggregation using the groupby() function, which helps in analyzing trends by summarizing large amounts of information.

Beyond these capabilities, Pandas seamlessly integrates with other Python libraries such as NumPy for numerical computations and Matplotlib for visualization, making it a key component of the data science workflow. It also supports importing and exporting data from multiple file formats, including CSV, Excel, and SQL databases, which enhances its versatility.

In the field of data science, Pandas plays a fundamental role by simplifying data preparation, an essential step before building predictive models. Its flexible and intuitive functionality allows analysts and scientists to efficiently manage datasets, making data-driven decision-making more effective. With its comprehensive tools for manipulating and analyzing data, Pandas continues to be a valuable asset in modern data science.

## References

-Analytics Vidhya. (2021, March 10). Pandas Functions for Data Analysis and Manipulation. Retrieved from https://www.analyticsvidhya.com/blog/2021/03/pandas-functions-for-data-analysis-and-manipulation/

-McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter (2nd ed.). O'Reilly Media.

-NVIDIA. (2024). What is Pandas in Python? Retrieved from https://www.nvidia.com/en-us/glossary/pandas-python/

-pandas-dev. (2024). pandas documentation. Retrieved from https://pandas.pydata.org/

-VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media. Retrieved from https://jakevdp.github.io/PythonDataScienceHandbook/03.00-introduction-to-pandas.html