Task “Data Analytics with Python and Data Storage Reflection” Unit 5

Introduction

Organizations today rely on insight from data analysis to guide strategic decision-making. Assets, and even business profitability are among the key factors at stake, therefore, data plays a crucial role in business today. With data, one of the key challenges in working is finding an approach that ensures accurate results, leading to correct and effective conclusions. A robust methodology and strategy are required to deliver quality, integration, and scalability; for this reason, selecting the optimal architectural roadmap is a fundamental requirement.

UCI Online Retail Dataset Analysis “UCI online retail dataset analysis.py”

(Chen, 2015; Chambers, and Zaharia, 2018; Databricks.com, 2018; McKinney, 2010)

Result

PS C:\Users\AmnonMalka> &

C:/Users/AmnonMalka/AppData/Local/Programs/Python/Python313/python.exe

"c:/Users/AmnonMalka/Documents/Code/Unit 5/UCI online retail dataset analysis.py"

Revenue: $9,747,747.93

Avg Transaction: $17.99

Items Sold: 5,176,450

Unique Customers: 4372

Unique Products: 4070

Date Range: 2010-12-01 to 2011-12-09

Largest Transaction: $168469.60

Returns/Refunds: 9290 (1.7%)

Best Sales Day: Thu ($20.34/txn)

Quantity-UnitPrice Correlation: -0.00

Storage Comparison Cloud-Based vs Local Storage Based

Choosing a storage model has a considerable impact on accessibility, flexibility, and performance, all of which are essential considerations. In terms of scalability and collaboration, cloud-based storage offers many advantages. Cloud providers, such as AWS, Google Cloud Storage, and Azure, enable large teams to collaborate, even those spread across multiple geographies. These can also be integrated with cloud-native analytics tools, for instance, Google BigQuery and Databricks. Cloud storage simplifies backup, versioning, and security management, which is vital for data. On the other hand, local storage is better for smaller analyses or prototyping. Using tools like Pandas in Python, the dataset can be loaded locally. This is faster for moderate file sizes and gives complete control over data privacy. However, local storage is limited in size, and as demand grows, it becomes slower to share, increasing the risk of data loss. For the moderately sized but detailed UCI Online Retail Dataset, cloud computing is well-suited for exploration; however, when demand rises, local storage may be more suitable.

References

Chen, D., (2015). Online Retail Dataset Online ‘Retail Dataset’. UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/datasets/Online+Retail [Accessed 24 August. 2025]

Chambers, B. and Zaharia, M., (2018). Spark: *The Definitive Guide* – ‘Online Retail Dataset’ chapter 3. Available at: https://essex.primo.exlibrisgroup.com/permalink/44UOES_INST/o3t9un/cdi_safari_books_v2_9781491912201 [Accessed 24 August. 2025]

Databricks.com hosted at GitHub (Spark: *The Definitive Guide* (2018). Dataset ‘retail-data’. Available at: https://github.com/databricks/Spark-The-Definitive-Guide/tree/master/data/retail-data [Accessed 24 August. 2025]

McKinney, W., 2010. Data structures for statistical computing in Python. In: van der Walt, S. and Millman, J., eds. Proceedings of the 9th Python in Science Conference, pp. 51–56. Available at: https://pandas.pydata.org/ [Accessed 24 August. 2025]

The pandas development team, (2023). *pandas-dev/pandas*: Pandas Zenodo (version 2.3.2) Available at: https://pandas.pydata.org/ [Accessed 24 August. 2025]

Matplotlib Development Team, 2024. Matplotlib (Version 3.8.4) [software]. Available at: https://matplotlib.org/ [Accessed 24 August. 2025]

This document has been written solely for educational purposes. All references, names, and trademarks mentioned here remain the property of their respective owners and are used here strictly for the educational context. Grammarly was used exclusively for proofreading and enhancing the clarity and language of the text. All academic writing, analysis, argumentation, and conclusions are entirely the original work of the author.