

# A Performance Comparison of Microsoft Excel and Python for Tabular Data Analysis

**Basma Mustafa M. H**

<sup>1</sup>College of science, University of Kerbala, Karabaia, Iraq. Email: [basma.m@uokerbala.edu.iq](mailto:basma.m@uokerbala.edu.iq)

## ARTICLE INFO

### Article history:

Received: 07 /01/2026

Revised form: 11 /02/2026

Accepted : 15 /02/2026

Available online: 30 /03/2026

### Keywords:

Microsoft Excel, Python  
(Pandas),Data Analysis,  
Performance Evaluation, Execution  
Time.

## ABSTRACT

Microsoft Excel is commonly treated as a simple spreadsheet program that is mainly applied to do simple calculations, and Python has become a common solution to complex data analysis based on the programming language. In this work, a technical performance analysis of Microsoft Excel and Python (Pandas) is provided as an attempt to assess their compatibility with basic data analysis functions. The comparison is done which consists of loading datasets, cleaning data, calculating features, aggregating and searching them. The main performance parameter used is execution time and measurements are taken at every processing phase so as to provide a detailed and equitable evaluation. To ensure a robust assessment, the methodology utilized two distinct transactional datasets: a medium-scale over 500000 records and a large-scale dataset over one million records. A key contribution of this work is the implementation of a search loop algorithm as a stress test, where execution time was measured for three specific scenarios: searching for a record at the beginning, at the end, and for a non-existent value. According to the experimental findings, Excel can efficiently handle all of the analyzed tasks and its execution time is as approximate Python as both systems are able to complete the tasks in an approximate time interval. Despite Python having more accuracy in procedures and measurements of time, more automation, and higher reproducibility due to code-based workflows, Excel has competitive analytical capabilities in interactive processing of data. These results underscore the idea that Microsoft Excel can no longer be viewed as an exclusively computational device, and instead of that, it can be viewed as a technically competent data analysis platform capable of performing similarly to Python on large-scaled analytical workloads.

MSC..

<https://doi.org/10.29304/jqcm.2026.18.12585>

## 1. Introduction

Data analysis is a key component of modern decision making. Organizations increasingly rely on data to guide their decisions. Data science focuses on extracting useful knowledge from data (Provost & Fawcett, 2013). Consequently, the selection of proper data analysis platforms has become of growing importance in the academic and industrial worlds. Excel is one of the most widely used tools for data analysis and reporting data since it is easy to use, available and contains strong built-in functions. Excel offers several functionalities, pivot tables are among Excel's most powerful features for summarizing data, built-in formulas enable fast calculations and efficient data manipulation and Excel allows users to analyze data without writing complex programs (Walkenbach, 2015), (Alexander & Kusleika, 2018). Microsoft Excel is a widely used tool for calculations, spreadsheets, and basic statistical analysis in business and educational settings. Excel is often used for data cleaning and descriptive analysis because it is familiar to many users, but it is not suitable for cases where multivariate analysis (Orhani, 2025).

\*Corresponding author: Basma Mustafa M. H

Email addresses: [basma.m@uokerbala.edu.iq](mailto:basma.m@uokerbala.edu.iq)

Communicated by 'sub editor'

Python-based systems are very flexible, highly automated, and can be scaled, which makes them popular in large-scale and research-driven data analysis piping (McKinney, 2012), (VanderPlas, 2016). Pandas provides high-performance, easy-to-use data structures and enables fast data processing and analysis in Python. The library is designed for efficient manipulation of tabular data. (McKinney, 2010). Overall, data science serves as the catalyst for informed decision-making across various industries, providing a solid foundation for predictions and strategies based on empirical evidence rather than intuition (Wang, 2024). The spreadsheet tools are also very effective in interactive and big data analytics have become increasingly important in both the academic and the business communities over the past two decades. exploration data analysis and execution time disparity is insignificant with data sets (Chen, Chiang, & Storey, 2012). Although the use of either Excel or Python is widespread, the view that Excel is only used in simple calculations is still widespread. The body of empirical research that would provide a systematic assessment of whether Excel could accomplish core data analysis functions with the same performance levels as Python given the same experimental conditions is wanting. To fill this gap, this study will compare the performance of Microsoft Excel and Python (Pandas) in controlled performance comparison based on how much they take to perform the same data processing steps in terms of execution time. This is to prove that Excel is not just a tool of calculation, but an effective data analysis platform that can perform to similar performance standards on key analysis processes.

## 2. Related work

There are a number of studies focusing on the productivity of spreadsheet-based tools and data analysis frameworks based on programs to work with and process large data sets. This paper will overview seven pieces of work that are in direct relationship to the area of this work.

In Ref. (Alexander & Kusleika, 2022), Alexander and Kusleika noted in this study that, just as Python is known for its outstanding tool in big data, recent studies have indicated that Excel has become a powerful tool for analyzing large-scale data through some advanced features such as Power Pivot. Excel can process data that exceeds the maximum number of rows, providing a level of performance comparable to that found in programming environments like Python, especially in terms of speed and ease of use.

Python was then proposed by McKinney in Ref. (McKinney, 2011) as a powerful tool of data analysis by creating a library known as Pandas which aims to facilitate working with statistics data sets and to provide a set of fundamental building blocks for implementing statistical models. His work was devoted to the development of high-performance data structures to work with tabular data and with such operations as filtering, aggregation, and statistical analysis.

In the study, it was established that Pandas is designed for working with large datasets efficiently are considered in contrast to the traditional spreadsheet-based methods. Later, McKinney in Ref. (McKinney, 2018) elaborated on the real-life uses of Python Pandas to the data analysis task. The author contrasted Python-based workflows with spreadsheet applications like Microsoft Excel and pointed out that Spreadsheet tools are useful for small-scale analysis and quick experimentation, whereas Pandas excels at data manipulation and analysis of complex datasets.

The comparative study of spreadsheet tools and Python-based data analysis frameworks reported by Chen et al in Ref. (Chen, Li, & Zhang, 2022). Execution time and efficiency are key factors in evaluating data analysis tools. Basic data processing activities were tested in the study and included filtering, aggregation, and statistical calculations. The results indicated that both programs offer strengths in handling data efficiently, providing a solid basis for evaluating the speed of implementation across different data sizes.

Ref. (Knaflic & Few, 2023), Knaflic and Few examined the data analysis processes in the contemporary world through comparing the old-fashioned spreadsheets methods with the programming ones, including Python. Their paper highlighted that Spreadsheets are effective for fast exploratory data analysis, and Programming-based tools excel in automation and reproducibility, and provide better performance for complex analytical tasks. The authors emphasized the growing shift from spreadsheets to programming-based data analysis.

Ref. (Panko, 2015) Panko investigated the shortcomings of spreadsheet program in data analysis especially in cases where large amounts of data are involved. The paper has highlighted that Spreadsheets are particularly vulnerable to human. Error rates increase as spreadsheets grow in size and complexity. It was found that the spreadsheet programs like the Excel are not appropriate when using large data volumes or performing repetitive data processing operations and the author suggested the use of the programming-based solutions.

A comparative experiment of diverse data analysis software, such as Microsoft Excel, SPSS, and Python, was provided by Van der Aalst in Ref. (Van der Aalst, 2016). The research worked on the preprocessing, aggregation and statistical analysis activities. The findings showed that Python-based tools support reproducible and research-oriented data analysis, and reproducible. Automated data analysis workflows improve efficiency and reliability.

Alteryx Research Team in Ref. (Alteryx Research Team, 2020) in business analytics setting, a comparative research study of Excel and Python was undertaken. The analysis has indicated that Excel is preferred by many non-technical users due to its ease of use and offers strong visualization and spreadsheet-based analysis capabilities, whereas Python is faster and more efficient for large-scale and repetitive analysis. The authors have pointed out the benefit of Python enables automation and integration with advanced analytics workflows.

In Ref. (McKinney & Pandas Development Team, 2020), Enterprises Wes McKinney and Pandas contributors tested Python Pandas on retail and e-commerce like transactional data. They showed that group-by and aggregation operations take considerably less time when performed by the Pandas than when performed by spreadsheet-based methods. The research affirmed Pandas provides fast and flexible data structures for working with relational data, designed to handle data-intensive analysis tasks efficiently, widely used for analyzing transactional and time-series data. A summary of related works is provided in Table 1, and it indicates the tools employed, the type of data sets, the area of comparison.

**Table 1: shows that previous studies relied on different types of datasets.**

Related work	Tools Used	Dataset Type	Comparison Focus
[10]	Excel, Python (Pandas)	Benchmark and real-world datasets	Tool performance
[11]	Python (Pandas)	Synthetic and statistical datasets	Traditional spreadsheets
[12]	Excel, Python (Pandas)	Retail and e-commerce datasets	Workflow efficiency
[13]	Excel, Python	Transactional datasets	Execution time
[14]	Spreadsheets, Python	Retail and e-commerce datasets	Analysis workflows
[15]	Excel	Retail and e-commerce datasets	Programming tools
[16]	Retail and e-commerce datasets	Retail and e-commerce datasets	Analysis tools
[17]	Excel, Python	Business analytics datasets	Business analytics
[18]	Python (Pandas)	Retail and e-commerce datasets	Spreadsheet methods

### 3. Methodology

This part explains the approach to be used to assess and compare the two programs in terms of processing and analyzing tabular data Microsoft Excel and Python (Pandas). The performance measure used to compare the two entities is the execution time; in addition, efficiency is also taken into consideration given the same conditions of processing. In order to have a controlled experimental design, use two data sets medium scale dataset and large scale dataset, and the specified data analysis procedures, such as data cleaning, calculating features, aggregating them, and searching, were used in both tools. A search loop was applied to both types of data to evaluate processing efficiency. The search task was performed in three cases: searching for a record at the beginning, another at the end, and a record that did not exist, in order to demonstrate the rate of time calculation as the data size increased. Figure 1 illustrates the steps of the experimental work process followed in performance comparison Excel and Python.

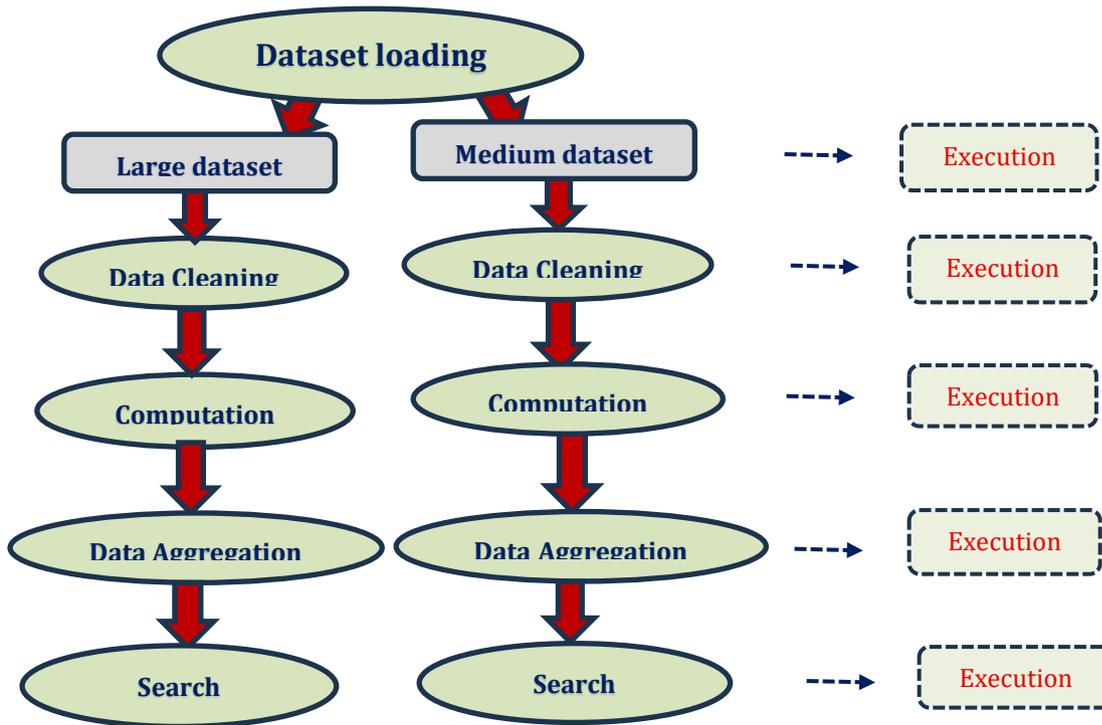


Fig. 1 Phases of the suggested data analysis process.

### 3.1 Dataset

The experiments were conducted using two different data sets to evaluate the scalability and efficiency of the selected platforms.

**1. medium scale dataset:** the Online Retail Dataset which is a feasibility of actual transactional records of a web-based retail outlet. The data came out of UCI Machine Learning Repository and contained details of invoice numbers, product identifiers, quantity, unit price, date of transaction, identification of customer, and country. The Online Retail dataset consists of more than 500,000 records, which depict the real-world retail data set apt in performance analysis (UCI Machine Learning Repository, n.d.).

**2. large scale dataset:** A large database was obtained from Kaggle, a leading data platform, containing over one million records and including banking information suitable for conducting tests.

In this study, two data analysis tools were utilized:

Microsoft Excel is a popular data analysis and visualization spreadsheet. Python (Pandas), a programming-based data analysis framework that is created to handle data processing at scale in an efficient manner. The dataset was transformed into CSV format in order to be compatible with both tools. All experiments in the Microsoft Excel and Python (Pandas) used the same data to ensure a level playing field.

### 3.2 Steps of Data Analysis in Microsoft Excel

#### 3.2.1 Loading the Datasets in Microsoft Excel

The Online Retail data was downloaded in the UCI Machine Learning Repository and the second large dataset downloaded in the Kaggle; both datasets were opened first in Microsoft Excel. These data were then stored as CSV to make it compatible and uniform to all steps and other tools of analysis applied in this research. As illustrated in Figure 2, a and b the online retail data and large dataset were loaded into Microsoft Excel where it can be seen as a preview.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/2010 08:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	01/12/2010 08:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/2010 08:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/2010 08:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/2010 08:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	01/12/2010 08:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	01/12/2010 08:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	01/12/2010 08:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01/12/2010 08:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01/12/2010 08:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	01/12/2010 08:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	01/12/2010 08:34	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01/12/2010 08:34	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	01/12/2010 08:34	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BLOCK WORD	3	01/12/2010 08:34	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BLOCK WORD	3	01/12/2010 08:34	5.95	13047	United Kingdom
536367	21777	RECIPE BOX WITH METAL HEART	4	01/12/2010 08:34	7.95	13047	United Kingdom
536367	48187	DOORMAT NEW ENGLAND	4	01/12/2010 08:34	7.95	13047	United Kingdom

(a) Medium dataset

Date	Domain	Location	Value	Transaction_count
01/01/2022	RESTRAUNT	Bhuj	365554	1932
01/01/2022	INVESTMENTS	Ludhiana	847444	1721
01/01/2022	RETAIL	Goa	786941	1573
01/01/2022	INTERNATIONAL	Mathura	368810	2049
01/01/2022	RESTRAUNT	Madurai	615681	1519
01/01/2022	INTERNATIONAL	Daman	1191092	1813
01/01/2022	INTERNATIONAL	Buxar	968883	2098
01/01/2022	PUBLIC	Trichy	1030297	806
01/01/2022	RESTRAUNT	Kullu	688655	1463
01/01/2022	MEDICAL	Hyderabad	1174302	1463
01/01/2022	PUBLIC	Lucknow	912902	1035
01/01/2022	INVESTMENTS	Bikaner	436534	1093
01/01/2022	PUBLIC	Amritsar	849803	2013
01/01/2022	INVESTMENTS	Mathura	1180043	2068
01/01/2022	RETAIL	Doda	1003497	1654
01/01/2022	MEDICAL	Jaipur	616724	1686
01/01/2022	PUBLIC	Kannur	933938	1362
01/01/2022	RETAIL	Lunglei	521003	2304
01/01/2022	INVESTMENTS	Bombay	849779	1202
01/01/2022	INVESTMENTS	Kannur	770080	2431
01/01/2022	MEDICAL	Pune	762330	1492

(b) Large dataset

Fig. 2 Overview of loaded datasets in Microsoft Excel.

### 3.2.2: Data Cleaning in Microsoft Excel

The Duplicate in Microsoft excel was used to delete duplicate records from both datasets medium and large. Data cleaning was done in Microsoft Excel to ensure quality and consistency of data before analysis. The conditional filtering was used to eliminate the transaction of non-positive quantities, which are returns or invalid sales records. Further, the duplicate entries were removed with the help of the Remove Duplicates tool to prevent biased aggregation outputs. Consequently, a clean and consistent dataset was obtained the execution time for this step was calculated for both datasets. Figure 3 below demonstrates both data sets following the data cleaning process in a and b.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	MissingFlag	PositiveQtyFlag	Total
536365	85123A	WHITE HANGING HI	6	01/12/2010 08:26	2.55	17850	United Kingdom	OK	valid	6 2.55
536365	71053	WHITE METAL LAN	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	6 3.39
536365	84406B	CREAM CUPID HEA	8	01/12/2010 08:26	2.75	17850	United Kingdom	OK	valid	8 2.75
536365	84029G	KNITTED UNION FL	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	6 3.39
536365	84029E	RED WOOLLY HOT	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	6 3.39
536365	22752	SET 7 BABUSHKA I	2	01/12/2010 08:26	7.65	17850	United Kingdom	OK	valid	2 7.65
536365	21730	GLASS STAR FRO	6	01/12/2010 08:26	4.25	17850	United Kingdom	OK	valid	6 4.25
536366	22633	HAND WARMER UN	6	01/12/2010 08:28	1.85	17850	United Kingdom	OK	valid	6 1.85
536366	22632	HAND WARMER RE	6	01/12/2010 08:28	1.85	17850	United Kingdom	OK	valid	6 1.85
536367	84879	ASSORTED COLOI	32	01/12/2010 08:34	1.69	13047	United Kingdom	OK	valid	32 1.69
536367	22745	POPPYS PLAYHOL	6	01/12/2010 08:34	2.1	13047	United Kingdom	OK	valid	6 2.1
536367	22748	POPPYS PLAYHOL	6	01/12/2010 08:34	2.1	13047	United Kingdom	OK	valid	6 2.1
536367	22749	FELTCTRAF PRINC	8	01/12/2010 08:34	3.75	13047	United Kingdom	OK	valid	8 3.75
536367	22310	IVORY KNITTED ML	6	01/12/2010 08:34	1.65	13047	United Kingdom	OK	valid	6 1.65
536367	84969	BOX OF 6 ASSORT	6	01/12/2010 08:34	4.25	13047	United Kingdom	OK	valid	6 4.25
536367	22623	BOX OF VINTAGE J	3	01/12/2010 08:34	4.95	13047	United Kingdom	OK	valid	3 4.95
536367	22622	BOX OF VINTAGE A	2	01/12/2010 08:34	9.95	13047	United Kingdom	OK	valid	2 9.95
536367	21754	HOME BUILDING BL	3	01/12/2010 08:34	5.95	13047	United Kingdom	OK	valid	3 5.95
536367	21755	LOVE BUILDING BL	3	01/12/2010 08:34	5.95	13047	United Kingdom	OK	valid	3 5.95
536367	21777	RECIP BOX WITH	4	01/12/2010 08:34	7.95	13047	United Kingdom	OK	valid	4 7.95
536367	48187	DOORMAT NEW ET	4	01/12/2010 08:34	7.95	13047	United Kingdom	OK	valid	4 7.95

(a) Medium dataset

Date	Domain	Location	Value	Transaction_count	Trim
01/01/2022	RESTRAUNT	Bhuj	365554	1932	365554
01/01/2022	INVESTMENTS	Ludhiana	847444	1721	847444
01/01/2022	RETAIL	Goa	786941	1573	786941
01/01/2022	INTERNATIONAL	Mathura	368610	2049	368610
01/01/2022	RESTRAUNT	Madurai	815681	1519	815681
01/01/2022	INTERNATIONAL	Daman	1191092	1813	1191092
01/01/2022	INTERNATIONAL	Buxar	968883	2098	968883
01/01/2022	PUBLIC	Trichy	1030297	806	1030297
01/01/2022	RESTRAUNT	Kullu	688655	1463	688655
01/01/2022	MEDICAL	Hyderabad	1174302	1463	1174302
01/01/2022	PUBLIC	Lucknow	912902	1035	912902
01/01/2022	INVESTMENTS	Bikaner	436534	1093	436534
01/01/2022	PUBLIC	Amritsar	849803	2013	849803
01/01/2022	INVESTMENTS	Mathura	1180043	2068	1180043
01/01/2022	RETAIL	Doda	1003497	1654	1003497
01/01/2022	MEDICAL	Jaipur	616724	1686	616724
01/01/2022	PUBLIC	Kannur	933938	1362	933938
01/01/2022	RETAIL	Lunglei	521003	2304	521003
01/01/2022	INVESTMENTS	Bombay	849779	1202	849779
01/01/2022	INVESTMENTS	Kannur	770080	2431	770080
01/01/2022	MEDICAL	Pune	762330	1492	762330

(b) Large dataset

Fig.3 Data set once data cleaning procedure is completed in Microsoft Excel.

### 3.2.3: Computation

The data in medium dataset were cleaned and then a calculated column called TotalAmount was developed in Microsoft Excel by multiplying Quantity with Unitprice. This operation was employed to measure the arithmetic operation performance on the cleansed data set.

$$=quantity * unitPrice$$

In large dataset after cleaning data a new column has been added to calculate the tax 5%.

Figure 4 (a) presents the calculated with the creation of the TotalAmount feature in medium dataset, (b) calculation the tax of large dataset.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	MissingFlag	Posit	TotalAmount
536365	85123A	WHITE HANGING HI	6	01/12/2010 08:26	2.55	17850	United Kingdom	OK	valid	15.3
536365	71053	WHITE METAL LAN	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	20.34
536365	84406B	CREAM CUPID HEA	8	01/12/2010 08:26	2.75	17850	United Kingdom	OK	valid	22
536365	84029G	KNITTED UNION FL	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	20.34
536365	84029E	RED WOOLLY HOT	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	20.34
536365	22752	SET 7 BABUSHKA F	2	01/12/2010 08:26	7.65	17850	United Kingdom	OK	valid	15.3
536366	21730	GLASS STAR FROX	6	01/12/2010 08:26	4.25	17850	United Kingdom	OK	valid	25.5
536366	22633	HAND WARMER UN	6	01/12/2010 08:28	1.85	17850	United Kingdom	OK	valid	11.1
536366	22632	HAND WARMER RE	6	01/12/2010 08:28	1.85	17850	United Kingdom	OK	valid	11.1
536367	84879	ASSORTED COLOI	32	01/12/2010 08:34	1.69	13047	United Kingdom	OK	valid	54.08
536367	22745	POPPY'S PLAYHOL	6	01/12/2010 08:34	2.1	13047	United Kingdom	OK	valid	12.6
536367	22748	POPPY'S PLAYHOL	6	01/12/2010 08:34	2.1	13047	United Kingdom	OK	valid	12.6
536367	22749	FELTCRAFT PRINC	8	01/12/2010 08:34	3.75	13047	United Kingdom	OK	valid	30
536367	22310	IVORY KNITTED ML	6	01/12/2010 08:34	1.65	13047	United Kingdom	OK	valid	9.9
536367	84869	BOX OF 6 ASSORT	6	01/12/2010 08:34	4.25	13047	United Kingdom	OK	valid	25.5
536367	22623	BOX OF VINTAGE J	3	01/12/2010 08:34	4.95	13047	United Kingdom	OK	valid	14.85
536367	22622	BOX OF VINTAGE A	2	01/12/2010 08:34	9.95	13047	United Kingdom	OK	valid	19.9
536367	21754	HOME BUILDING BL	3	01/12/2010 08:34	5.95	13047	United Kingdom	OK	valid	17.85
536367	21755	LOVE BUILDING BL	3	01/12/2010 08:34	5.95	13047	United Kingdom	OK	valid	17.85
536367	21777	RECIPE BOX WITH	4	01/12/2010 08:34	7.95	13047	United Kingdom	OK	valid	31.8
536367	48187	DOORMAT NEW EI	4	01/12/2010 08:34	7.95	13047	United Kingdom	OK	valid	31.8

(a) Medium dataset

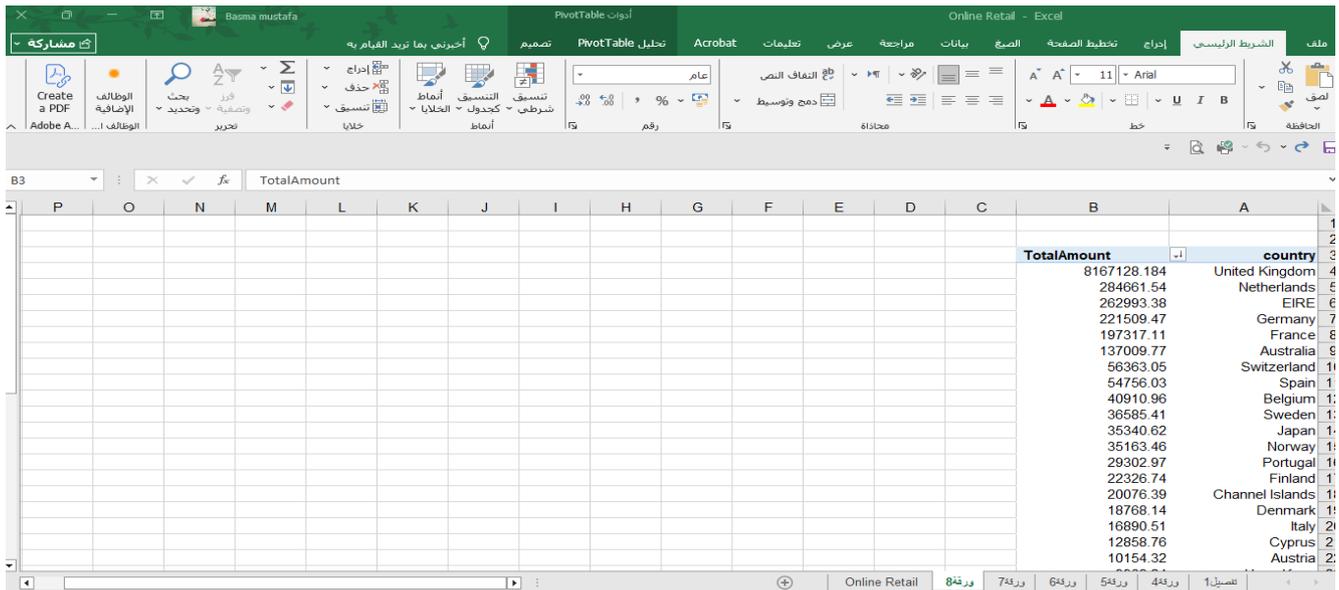
Date	Domain	Location	Value	Transaction_count	Trim	Text result
01/01/2022	RESTRAUNT	Bhuj	365554	1932	365554	18277.7
01/01/2022	INVESTMENTS	Ludhiana	847444	1721	847444	42372.2
01/01/2022	RETAIL	Goa	786841	1573	786841	39347.05
01/01/2022	INTERNATIONAL	Mathura	368610	2049	368610	18430.5
01/01/2022	RESTRAUNT	Madurai	615681	1519	615681	30784.05
01/01/2022	INTERNATIONAL	Daman	1191092	1813	1191092	59554.6
01/01/2022	INTERNATIONAL	Buxar	968883	2098	968883	48444.15
01/01/2022	PUBLIC	Trichy	1030297	606	1030297	51514.85
01/01/2022	RESTRAUNT	Kullu	688655	1463	688655	34432.75
01/01/2022	MEDICAL	Hyderabad	1174302	1463	1174302	58715.1
01/01/2022	PUBLIC	Lucknow	912902	1035	912902	45645.1
01/01/2022	INVESTMENTS	Bikaner	436534	1093	436534	21826.7
01/01/2022	PUBLIC	Amritsar	849803	2013	849803	42490.15
01/01/2022	INVESTMENTS	Mathura	1180043	2068	1180043	58002.15
01/01/2022	RETAIL	Doda	1003497	1654	1003497	50174.85
01/01/2022	MEDICAL	Jaipur	616724	1686	616724	30836.2
01/01/2022	PUBLIC	Kannur	933938	1362	933938	46696.9
01/01/2022	RETAIL	Lunglei	521003	2304	521003	26050.15
01/01/2022	INVESTMENTS	Bombay	849779	1202	849779	42488.95
01/01/2022	INVESTMENTS	Kannur	770080	2431	770080	38504
01/01/2022	MEDICAL	Pune	762330	1492	762330	38116.5

(b) Large dataset

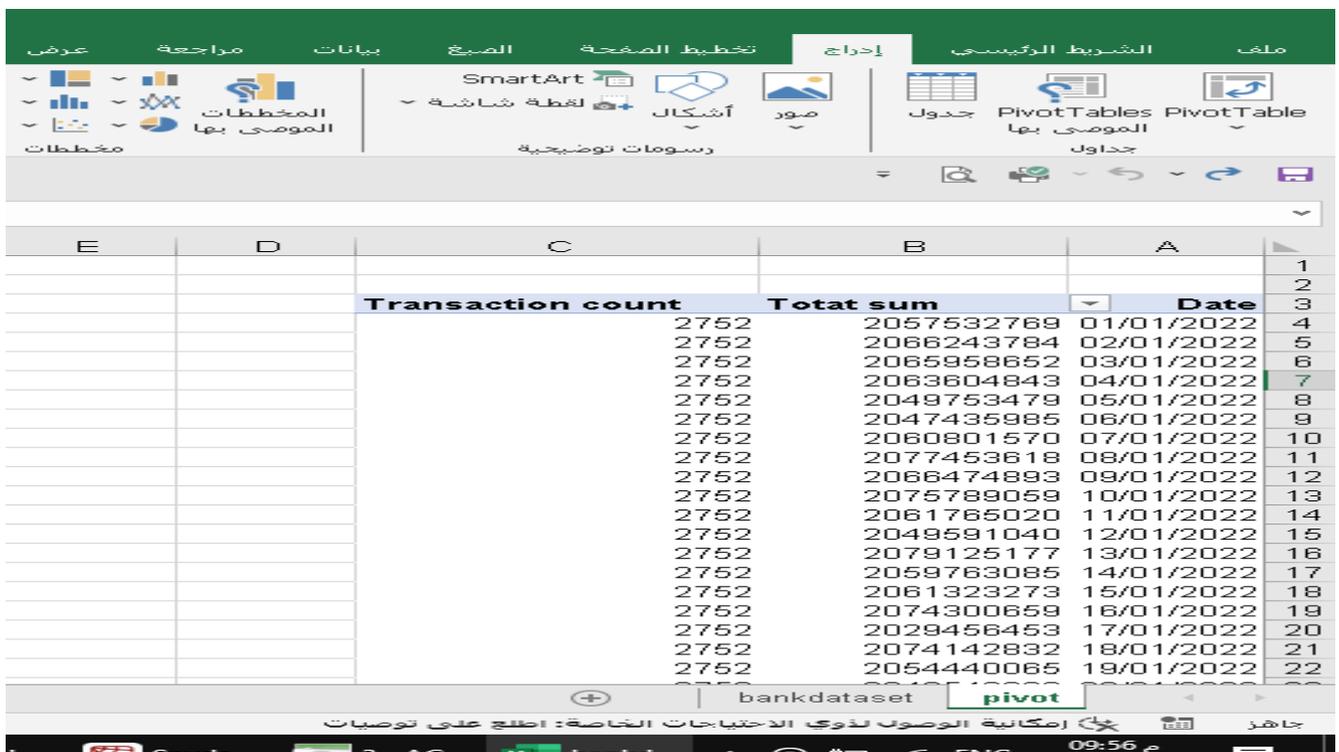
Fig. 4: Database following calculation step in datasets.

### 3.2.4: Data Aggregation Pivot Table

In Microsoft excel, a Pivot table was prepared to summarize the cleansed data. All the range of data was chosen, and the Pivot Table tool was employed to group sales data. The obtained table is a brief summary of the total sales by countries and was taken to compare the efficiency of data aggregation operations. The Pivot table results were then arranged in descending order in terms of the amount of sum Totals in order to determine the most performing countries. At large dataset the same aggregation was applied, pivot table summarized large dataset, and in this step execution time was recorded for each datasets Figure 5 (a)and (b)demonstrates how the steps of aggregation are used for both datasets.



(a) Medium dataset

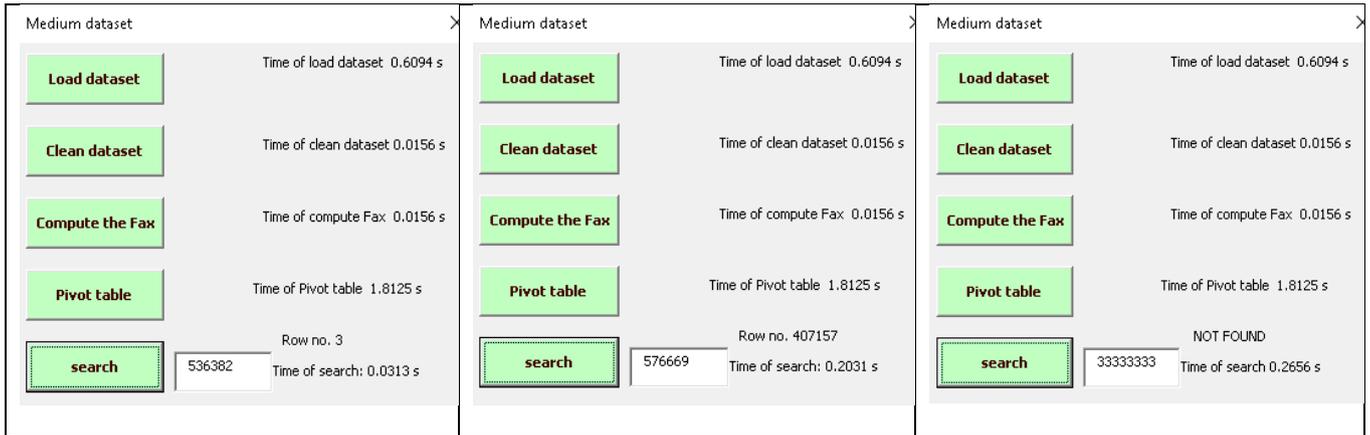


(b) Large dataset

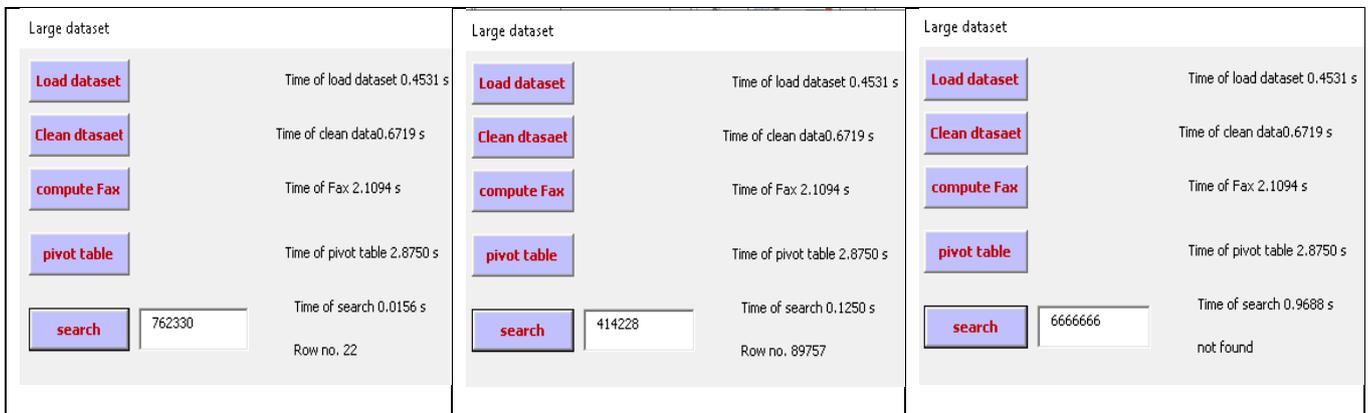
Fig. 5 Aggregated data obtained in Pivot Tables in Microsoft Excel for datasets.

### 3.2.5 Search for number of row

At this stage, an algorithm was developed to search for the row number using the Visual Basic Application program, and the execution time was measured using a timer that gives accurate results in order to ensure comparison of the results with Python. The search operation was performed three times in order to ensure the accuracy of the results. The results of applying the search process three times in Excel to the two datasets are shown in Figure 6.



(a) Medium dataset



(b) Large dataset

**Fig. 6** searching for no. of row in both datasets.

## 3.3 Steps of Data Analysis in Python

### 3.3.1 Loading the Dataset in Python

The dataset Online Retail and large dataset were imported to Python via the Pandas library. The datasets were imported with the read csv function and the initial records were presented in order to ensure that the data was imported and is now ready to be analyzed. The choice of panda was based on its effective data structures and optimized functions to process and handle tabular large-scale data. The time used in executing each stage of the process was kept. Figure 7 shows the two datasets after being loaded in Python using the Pandas library.

df - DataFrame

Index	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/2010 08:26	2.55	17850	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	01/12/2010 08:26	3.39	17850	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/2010 08:26	2.75	17850	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/2010 08:26	3.39	17850	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/2010 08:26	3.39	17850	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08:26	7.65	17850	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	01/12/2010 08:26	4.25	17850	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	01/12/2010 08:28	1.85	17850	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	01/12/2010 08:28	1.85	17850	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01/12/2010 08:34	1.69	13047	United Kingdom
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08:34	2.1	13047	United Kingdom
11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08:34	2.1	13047	United Kingdom
12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01/12/2010 08:34	3.75	13047	United Kingdom
13	536367	22310	IVORY KNITTED MUG COSY	6	01/12/2010 08:34	1.65	13047	United Kingdom
14	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	01/12/2010 08:34	4.25	13047	United Kingdom
15	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01/12/2010 08:34	4.95	13047	United Kingdom
16	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	01/12/2010 08:34	9.95	13047	United Kingdom
17	536367	21754	HOME BUILDING BLOCK WORD	3	01/12/2010 08:34	5.95	13047	United Kingdom
18	536367	21755	LOVE BUILDING BLOCK WORD	3	01/12/2010 08:34	5.95	13047	United Kingdom

Format    Resize     Background color     Column min/max

(a) Medium dataset

df - DataFrame

Index	Date	Domain	Location	Value	Transaction_count
0	01/01/2022	RESTAUNT	Bhuj	365554	1932
1	01/01/2022	INVESTMENTS	Ludhiana	847444	1721
2	01/01/2022	RETAIL	Goa	786941	1573
3	01/01/2022	INTERNATIONAL	Nathura	368610	2049
4	01/01/2022	RESTAUNT	Madurai	615681	1519
5	01/01/2022	INTERNATIONAL	Daman	1191092	1813
6	01/01/2022	INTERNATIONAL	Buzar	968883	2098
7	01/01/2022	PUBLIC	Trichy	1030297	606
8	01/01/2022	RESTAUNT	Kullu	688655	1463
9	01/01/2022	MEDICAL	Hyderabad	1174302	1463
10	01/01/2022	PUBLIC	Lucknow	912902	1035
11	01/01/2022	INVESTMENTS	Bikaner	436534	1093
12	01/01/2022	PUBLIC	Amritsar	849803	2013
13	01/01/2022	INVESTMENTS	Nathura	1180043	2068
14	01/01/2022	RETAIL	Doda	1003497	1654
15	01/01/2022	MEDICAL	Jaipur	616724	1686
16	01/01/2022	PUBLIC	Kannur	933938	1362
17	01/01/2022	RETAIL	Lunglei	521003	2304
18	01/01/2022	INVESTMENTS	Bombay	849779	1202

Format    Resize     Background color     Column min/max    Save and Close    Close

(b) Large dataset

Fig. 7 Two dataset loaded in Python using the Pandas library.

### 3.3.2 Data Cleaning in Python

In this step, the following data cleaning operations were performed using Python (Pandas):

- Removal of missing values
- Elimination of duplicate records to avoid redundant transactions.

The cleaned dataset after applying the data cleaning operations in Python is illustrated in Figure 8 for both datasets.

Index	InvoiceNC	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	MissingFlag	PositiveQtyFlag	عمود1	عمود2
297646	A563186	B	Adjust bad debt	1	12/08/2011 14:51	-11062...	nan	United Kingdom	OK	valid	1	-11062.1
297647	A563187	B	Adjust bad debt	1	12/08/2011 14:52	-11062...	nan	United Kingdom	OK	valid	1	-11062.1
605	536414	22139	nan	56	01/12/2010 11:52	0	nan	United Kingdom	OK	valid	56	0
1934	536545	21134	nan	1	01/12/2010 14:32	0	nan	United Kingdom	OK	valid	1	0
1935	536546	22145	nan	1	01/12/2010 14:33	0	nan	United Kingdom	OK	valid	1	0
1936	536547	37509	nan	1	01/12/2010 14:33	0	nan	United Kingdom	OK	valid	1	0
1951	536549	85226A	nan	1	01/12/2010 14:34	0	nan	United Kingdom	OK	valid	1	0
1952	536550	85044	nan	1	01/12/2010 14:34	0	nan	United Kingdom	OK	valid	1	0
1984	536552	20950	nan	1	01/12/2010 14:34	0	nan	United Kingdom	OK	valid	1	0
1985	536553	37461	nan	3	01/12/2010 14:35	0	nan	United Kingdom	OK	valid	3	0
1986	536554	84670	nan	23	01/12/2010 14:35	0	nan	United Kingdom	OK	valid	23	0
2362	536589	21777	nan	-10	01/12/2010 16:50	0	nan	United Kingdom	OK	invalid	-10	0
4286	536764	84952C	nan	-38	02/12/2010 14:42	0	nan	United Kingdom	OK	invalid	-38	0
4287	536765	84952C	nan	19	02/12/2010 14:43	0	nan	United Kingdom	OK	valid	19	0
6294	536941	22734	amazon	20	03/12/2010 12:08	0	nan	United Kingdom	OK	valid	20	0
6295	536942	22139	amazon	15	03/12/2010 12:08	0	nan	United Kingdom	OK	valid	15	0
7085	536995	35951	nan	57	03/12/2010 15:29	0	nan	United Kingdom	OK	valid	57	0
7086	536996	22712	nan	-20	03/12/2010 15:30	0	nan	United Kingdom	OK	invalid	-20	0
7087	536997	22028	nan	-20	03/12/2010 15:30	0	nan	United Kingdom	OK	invalid	-20	0

(a) Medium dataset

Index	Date	Domain	Location	Value	Transaction_count
0	01/01/2022	RESTRAUNT	Bhuj	365554	1932
1	01/01/2022	INVESTMENTS	Ludhiana	847444	1721
2	01/01/2022	RETAIL	Goa	786941	1573
3	01/01/2022	INTERNATIONAL	Mathura	368610	2049
4	01/01/2022	RESTRAUNT	Madurai	615681	1519
5	01/01/2022	INTERNATIONAL	Daman	1191092	1813
6	01/01/2022	INTERNATIONAL	Buxar	968883	2098
7	01/01/2022	PUBLIC	Trichy	1030297	606
8	01/01/2022	RESTRAUNT	Kullu	688655	1463
9	01/01/2022	MEDICAL	Hyderabad	1174302	1463
10	01/01/2022	PUBLIC	Lucknow	912902	1035
11	01/01/2022	INVESTMENTS	Bikaner	436534	1093
12	01/01/2022	PUBLIC	Amritsar	849803	2013
13	01/01/2022	INVESTMENTS	Mathura	1180043	2068
14	01/01/2022	RETAIL	Doda	1003497	1654
15	01/01/2022	MEDICAL	Jaipur	616724	1686
16	01/01/2022	PUBLIC	Kannur	933938	1362
17	01/01/2022	RETAIL	Lunglei	521003	2304
18	01/01/2022	INVESTMENTS	Bombay	849779	1202

(b) Large dataset

Fig. 8. Dataset after applying the data cleaning process in Python.

### 3.3.3: Computation in Python

In this step, a new feature called TotalAmount was computed using a vectorized operation for medium dataset and compute the tax for large dataset. The two datasets after computing step is shown in Figure 9.

Index	invoiceNo	stockCode	Description	Quantity	InvoiceDate	UnitPrice	customer	Country	MissingFlag	InvoiceQtyF	عمود1	عمود2	TotalAmount
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/2010 08:26	2.55	17850	United Kingdom	OK	valid	6	2.55	15.3
1	536365	71053	WHITE METAL LANTERN	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	6	3.39	20.34
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/2010 08:26	2.75	17850	United Kingdom	OK	valid	8	2.75	22
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	6	3.39	20.34
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/2010 08:26	3.39	17850	United Kingdom	OK	valid	6	3.39	20.34
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08:26	7.65	17850	United Kingdom	OK	valid	2	7.65	15.3
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	01/12/2010 08:26	4.25	17850	United Kingdom	OK	valid	6	4.25	25.5
7	536366	22633	HAND WARMER UNION JACK	6	01/12/2010 08:28	1.85	17850	United Kingdom	OK	valid	6	1.85	11.1
8	536366	22632	HAND WARMER RED POLKA DOT	6	01/12/2010 08:28	1.85	17850	United Kingdom	OK	valid	6	1.85	11.1
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01/12/2010 08:34	1.69	13047	United Kingdom	OK	valid	32	1.69	54.08
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08:34	2.1	13047	United Kingdom	OK	valid	6	2.1	12.6
11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08:34	2.1	13047	United Kingdom	OK	valid	6	2.1	12.6
12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01/12/2010 08:34	3.75	13047	United Kingdom	OK	valid	8	3.75	30
13	536367	22310	IVORY KNITTED MUG COSY	6	01/12/2010 08:34	1.65	13047	United Kingdom	OK	valid	6	1.65	9.9
14	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	01/12/2010 08:34	4.25	13047	United Kingdom	OK	valid	6	4.25	25.5
15	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01/12/2010 08:34	4.95	13047	United Kingdom	OK	valid	3	4.95	14.85
16	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	01/12/2010 08:34	9.95	13047	United Kingdom	OK	valid	2	9.95	19.9
17	536367	21754	HOME BUILDING BLOCK WORD	3	01/12/2010 08:34	5.95	13047	United Kingdom	OK	valid	3	5.95	17.85
18	536367	21755	LOVE BUILDING BLOCK WORD	3	01/12/2010 08:34	5.95	13047	United Kingdom	OK	valid	3	5.95	17.85

(a) Medium dataset

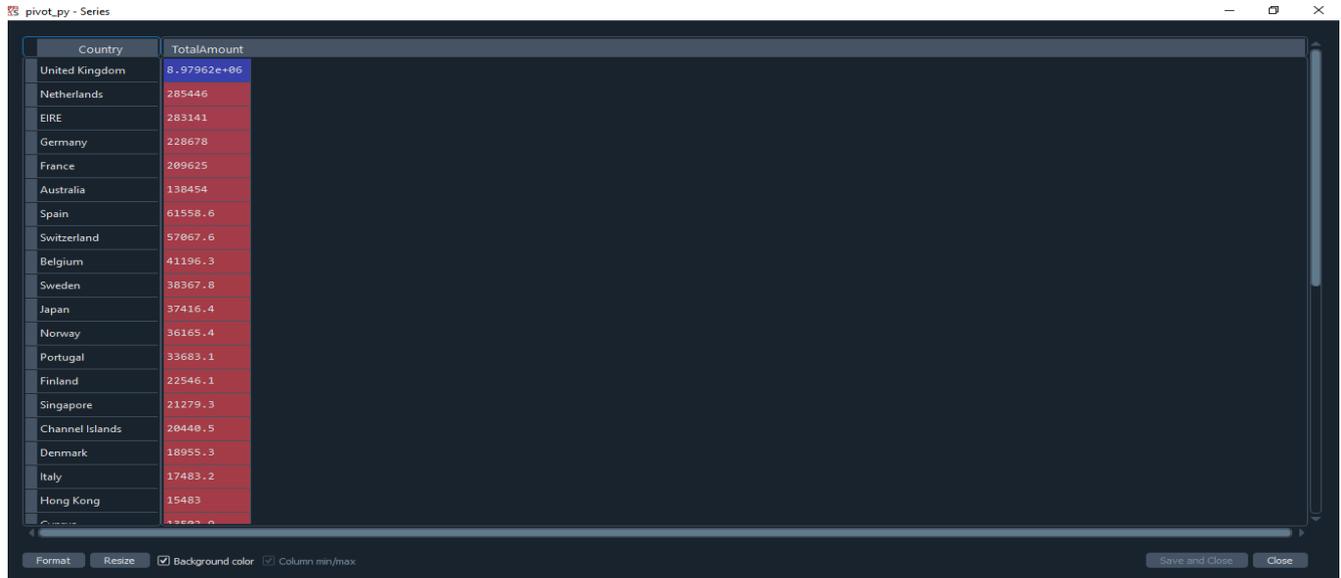
Index	date	domain	location	value	transaction_count	tax_result
410547	30/05/2022	EDUCATION	Ahmedabad	298423	1148	14921.2
451736	14/06/2022	PUBLIC	Buxar	298425	2376	14921.2
855257	07/11/2022	EDUCATION	Pune	298425	2108	14921.2
765583	06/10/2022	RETAIL	Akola	298428	973	14921.4
851771	06/11/2022	PUBLIC	Jaipur	298428	1156	14921.4
202264	15/03/2022	EDUCATION	Kolkata	298429	2046	14921.5
70833	26/01/2022	PUBLIC	Varanasi	298430	2200	14921.5
107508	09/02/2022	INTERNATIONAL	Kolkata	298432	1547	14921.6
438038	09/06/2022	PUBLIC	Bhuj	298432	1596	14921.6
227072	24/03/2022	RETAIL	Kolkata	298433	2250	14921.7
999898	30/12/2022	MEDICAL	Doda	298434	797	14921.7
605616	09/08/2022	RETAIL	Konark	298435	603	14921.8
686027	07/09/2022	RESTRAUNT	Ara	298435	1449	14921.8
411855	30/05/2022	RESTRAUNT	Ludhiana	298437	1954	14921.9
119781	13/02/2022	INTERNATIONAL	Konark	298438	813	14921.9
27867	11/01/2022	MEDICAL	Mon	298440	1792	14922
298253	19/04/2022	RESTRAUNT	Hyderabad	298440	1635	14922
942993	09/12/2022	PUBLIC	Konark	298441	1031	14922.1
151798	25/02/2022	PUBLIC	Tirumala	298442	1157	14922.1

(b) Large dataset

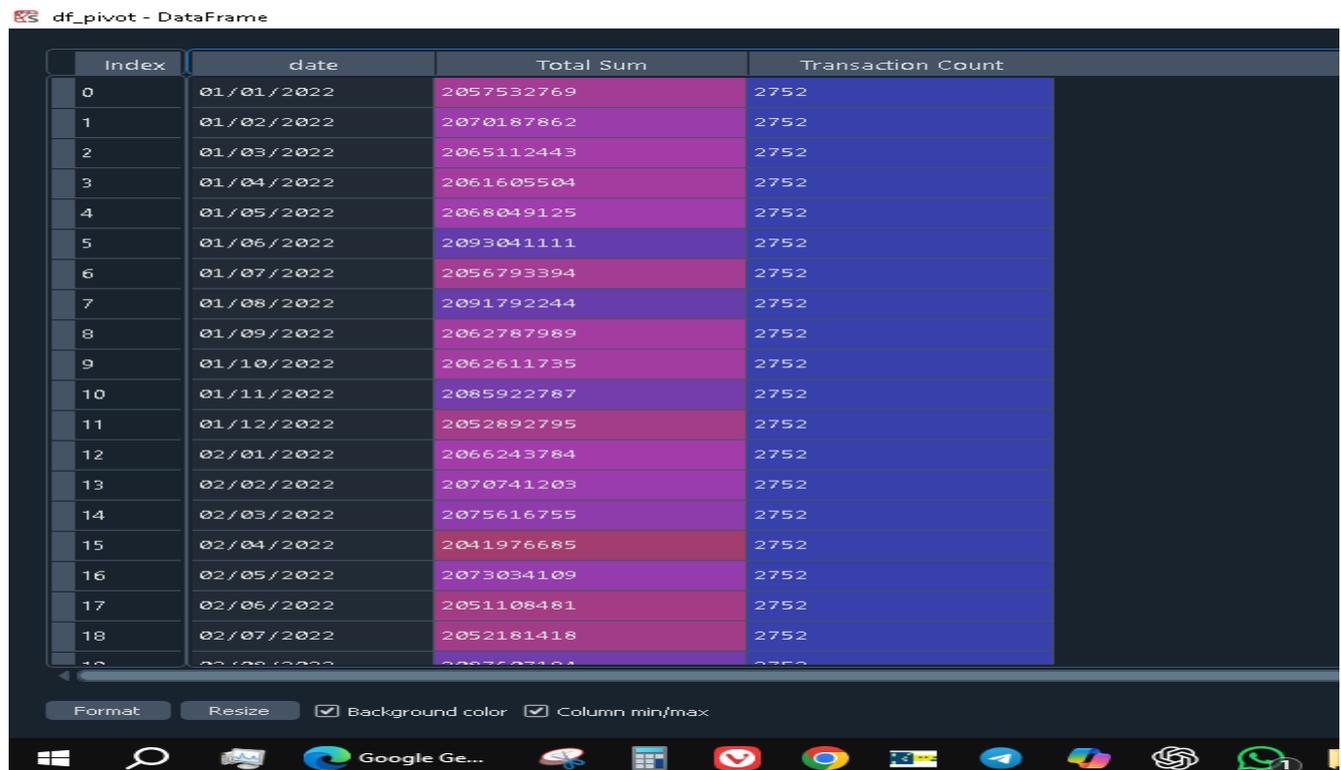
Fig 9 Dataset after computation step in Python.

### 3.3.4: Aggregation Pivot Equivalent in Python

The aggregation of data was performed using the (group by) function from the Pandas library, which is used to perform operations equivalent to a pivot table. This ensures high efficiency when processing large datasets. The execution time for both datasets was calculated using python time library. The aggregated and sorted results are illustrated in Figure 10 for both datasets.



(a) Medium dataset

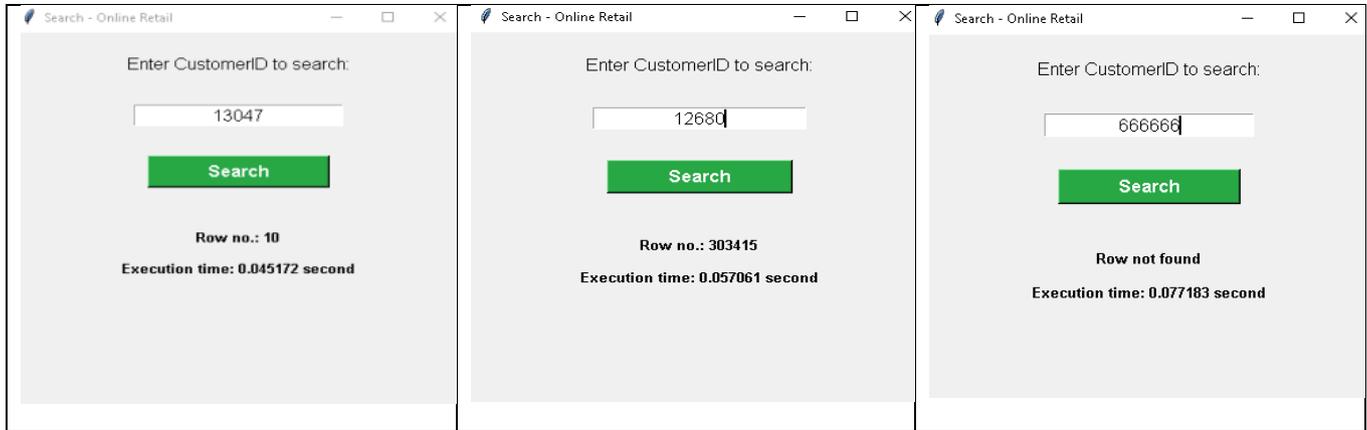


(b) Large dataset

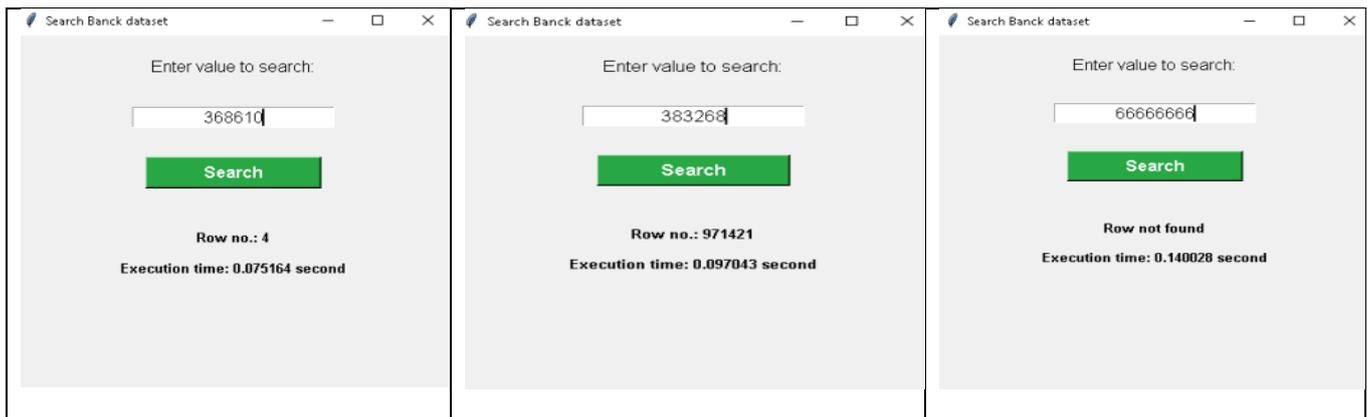
**Fig. 10** Aggregated results obtained using Python for datasets.

### 3.3.5 Search for number of row

In this step, a search loop designed to search for no. of row to evaluate the efficiency and execution time in python for both datasets. The search was performed three times according to the classroom location in order to measure response time. Searching for the row number is illustrated in Figure 11 for two data sets in Python Panda.



(a) Medium dataset



(b) Large dataset

**Fig. 11** Searching for row number with time taken for two datasets.

### Tools and Techniques Used in This Study

Approximate Microsoft Excel and Python (Pandas) were used as the main instruments of data analysis in this study. Excel was chosen because it is widely used and has inbuilt analysis capabilities including a filtering tool inbuilt analytical capabilities including formula-driven computational and a PivotTable is a powerful tool to calculate, summarize, and analyze data that lets you see comparisons, patterns, and trends in your data. (Microsoft, 2023), (Few, 2022). Visual Basic Applications (VBA) was used; it is a programming language that can interact with Excel, employing a set of algorithms such as execution time measurement and providing highly accurate results (Alexander & Kusleika, 2019). Microsoft Excel and Python (Pandas) were used as the main instruments of data analysis in this study. Microsoft Excel was chosen because it is very common, and has built-in functionality to analyze information, including filtering, formulas-based computation, and Pivot Tables, which can be used to summarize and analyses tabular data efficiently and exploratorily (McKinney, 2023), (VanderPlas, 2023). Microsoft Excel is a widely used tool for calculations and spreadsheet-based analysis and has been proven to be a viable and efficient interactive data analysis tool (Orhani, 2025).

## 4. Results and Evaluation

This paper considers execution time as the main performance measure to be used to compare and contrast the performance of Microsoft Excel and Python (Pandas). Execution time is the amount of time required by each data processing operation, such as data cleaning, computing features, aggregation and searching. All operations were performed on a Windows 10,64-bit computer with 8 GB of RAM and an Intel Core i5 processor. The two tools were tested on the two datasets and with same steps of processing to achieve a fair and consistent comparison. To ensure measurement accuracy compared to Python, the execution time of each operation in Excel was calculated using VBA Timer and the time was recorded in seconds for each stage, in Python use Python time library for computer the execution time .The metric gives a clear and practical idea of how much is computed efficient as large scale datasets are dealt with.

### 4.1: Execution Time Recording for Excel

The performance evaluation was done through the measurement of the time taken through by each step in the processing by using VBA timer for both datasets and the performance results are given in Table 2.

**Table 2. Microsoft Excel gives results on the execution time per processing step for both datasets.**

Task	Medium dataset	Large dataset
Data Loading	0.6094	0.4531
Data Cleaning	0.0156	0.6719
Calculation	0.0156	2.1094
Pivot and Sorting	1.8125	2.8750
searching	0.0313	0.0156

### 4.2: Execution Time Recording for Python

Processing activities in the Python were carried out through the Pandas library, and the programmatic timing was used to establish the execution time through Python time library that gives high-resolution timing as a rose to measure performance. Table 3 summarizes the results of the recorded execution time for both datasets.

**Table 3. Python gets the execution time results per step for both datasets.**

Task	Medium dataset	Large dataset
Data Loading	0.3716	0.5397
Data Cleaning	0.0942	0.0333
Calculation	0.0155	0.0259
Pivot and Sorting	0.0482	0.2909
Searching	0.0451	0.0751

The search process was repeated three times to evaluate the performance of both tools on the two dataset types. In the first case, the search was for a record at the beginning of the dataset; in the second case, it was for a record at the end of the database; and in the last case, it was for a non-existent record. In this last case, a full scan of the dataset was performed, and the result was "No record found."

The execution time for each case was recorded in Microsoft Excel and Python Panda. Table 4 shows the execution time for the three cases on medium dataset and table 5 for large dataset.

**Table 4. Execution time and performance of Excel for search step for three times.**

<b>Medium dataset</b>		
<b>Stats</b>	<b>Microsoft Excel</b>	<b>Python</b>
At the beginning of dataset	0.0313	0.0451
At the end of dataset	0.2031	0.0570
Value not found in dataset	0.2656	0.0771

**Table 5. Execution time and performance of Python for search step for three times.**

<b>Large dataset</b>		
<b>Stats</b>	<b>Microsoft Excel</b>	<b>Python</b>
At the beginning of dataset	0.0156	0.0751
At the end of dataset	0.1250	0.0970
Value not found in dataset	0.9688	0.1400

All the processing action took Microsoft Excel approximate execution time compared to Python that is known for its high ability to handle various types of data and this was because the operation in spreadsheet was interactive. The experimental findings indicate that Microsoft Excel can perform the same evaluated processing operations but with a performance that undergoes comparison to Python (Pandas). In the tested of two types of datasets, both tools took data cleaning, computing features, aggregating and searching within an equal range of execution time. Specifically, the execution times of Excel were approximated these with the accuracy of times in Python. These results are to suggest that Excel could be used to give efficient performance on medium and large data analysis tasks that same as to Python, particularly in the case of interactive analysis and exploratory analysis. The primary aim of the study is to prove that Microsoft Excel is not the simple computational tool or spreadsheet, but the powerful platform of data analysis that can execute the essential data processing operations just like Python (Pandas). These results emphasize that Excel can be used as a viable and effective substitute of programming-based products to help in analyzing data.

## 5. Conclusion and Future Work

### 5.1 Conclusion

This paper provided a comparative analysis between Microsoft Excel and Python (Pandas) to perform some of the core data analysis tasks working on a real-life transactional dataset. The experimental outcomes based on Microsoft Excel proved that it can conduct data cleaning, feature computations, aggregations, sorting and searching operations in approximate execution time as Python when working with medium and large datasets. The findings address the general assumption that Excel is a simple computational tool and emphasize its usefulness as an interactive and exploratory platform of data analysis.

### 5.2 Future Work

As future research, this comparison can be applied to huge sets contains 5 million row and more and more in other application areas to further determine scalability and generalizability. Further research can also include the measures of memory use and parallel processing when the data load 5 million and more data to ensure a more detailed evaluation of the performance.

## References

- [1] Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media.
- [2] Walkenbach, J. (2015). *Excel 2016 Bible*. Wiley.
- [3] Alexander, M., & Kusleika, D. (2018). *Excel 2019 Power Programming with VBA*. Wiley.
- [4] Orhani, S. (2025). Comparative analysis of statistical results generated by Python, R, SPSS, and Excel. *International Journal of Progressive Research in Engineering Management and Science*.
- [5] McKinney, W. (2012). *Python for data analysis*. O'Reilly Media.

- 
- [6] VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media.
- [7] McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.
- [8] Wang, S.-F. (2024). The secret behind Big Data – Data Science. *CASE, National Taiwan University*. <https://case.ntu.edu.tw/blog/?p=43964>.
- [9] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics. *MIS Quarterly*, 36(4), 1165–1188.
- [10] Alexander, M., & Kusleika, D. (2022). *Excel Data Analysis: Your Visual Blueprint for Analyzing Data, Charts, and PivotTables*. Wiley.
- [11] McKinney, W. (2011). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.
- [12] McKinney, W. (2018). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.
- [13] Chen, J., Li, Y., & Zhang, H. (2022). A comparative study of spreadsheet tools and Python-based frameworks for data analysis efficiency. *Journal of Big Data*, 9(1), 1–18.
- [14] Knafflic, S. N., & Few, S. (2023). Modern data analysis workflows: From spreadsheets to programming-based tools. *IEEE Access*, 11, 45678–45690.
- [15] Panko, R. R. (2015). What we know about spreadsheet errors. *Journal of End User Computing*, 10(2), 15–21.
- [16] Van der Aalst, W. (2016). *Data Science in Action*. Springer International Publishing.
- [17] Alteryx Research Team. (2020). *Excel vs Python for Data Analytics*. Alteryx Inc.
- [18] McKinney, W., & Pandas Development Team. (2020). *Pandas Documentation*. Available online.
- [19] UCI Machine Learning Repository. (n.d.). *Online Retail dataset*.
- [20] Microsoft. (2023). *Analyze data with Excel PivotTables*. Microsoft Documentation.
- [21] Few, S. (2022). *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring*. Analytics Press.
- [22] Alexander, M., & Kusleika, R. (2019). *Excel 2019 Power Programming with VBA*. John Wiley & Sons.
- [23] McKinney, W. (2023). *pandas: powerful Python data analysis toolkit*. Pandas Documentation.
- [24] VanderPlas, J. (2023). *Python Data Science Handbook* (2nd ed.). O'Reilly Media.
- [25] Orhani, S. (2025). Comparative analysis of statistical results generated by Python, R, SPSS, and Excel. *International Journal of Progressive Research in Engineering Management and Science*.