**ASSIGNMENT 2 FRONT SHEET**

|  |  |  |  |
| --- | --- | --- | --- |
| **Qualification** | **BTEC Level 5 HND Diploma in Computing** | | |
| **Unit number and title** | Unit 14: Business Intelligence | | |
| **Submission date** |  | **Date Received 1st submission** |  |
| **Re-submission Date** |  | **Date Received 2nd submission** |  |
| **Student Name** |  | **Student ID** |  |
| **Class** |  | **Assessor name** |  |
| **Student declaration**  I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice. | | | |
|  |  | **Student’s signature** |  |

**Grading grid**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| P3 | P4 | P5 | P6 | M3 | M4 | D3 | D4 |
|  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **❒ Summative Feedback: ❒ Resubmission Feedback:** | | |
| **Grade:** | **Assessor Signature:** | **Date:** |
| **IV Signature:** | | |

Contents

[I. Intro 5](#_Toc97585300)

[1. What is BI? 5](#_Toc97585301)

[2. Real examples of how to apply BI on business 5](#_Toc97585302)

[3. Techniques Used in BI 6](#_Toc97585303)

[4. Data Visualization 6](#_Toc97585304)

[5. Data Mining 6](#_Toc97585305)

[6. Reporting 6](#_Toc97585306)

[7. Time-series Analysis Including (Predictive Techniques) 7](#_Toc97585307)

[8. On-line Analytical Processing (OLAP) 7](#_Toc97585308)

[9. Statistical Analysis 7](#_Toc97585309)

[10. BI tools (Python) 8](#_Toc97585310)

[11. BI tools(tableau) 9](#_Toc97585311)

[12. Tableau application in business 10](#_Toc97585312)

[II. Related work 10](#_Toc97585313)

[1. Dataset 10](#_Toc97585314)

[2. Filter data 15](#_Toc97585315)

[III. Evaluation 19](#_Toc97585316)

[IV. Conclusion 31](#_Toc97585317)

[Bibliography 32](#_Toc97585318)

[Figure 1. BI (Taylor, 2022) 6](#_Toc97583144)

[Figure 2. BI tools (Nguyễn, 2020) 8](#_Toc97583145)

[Figure 3. tableau 9](#_Toc97583146)

[Figure 4.Gender 12](#_Toc97583147)

[Figure 5.Gender\_bought 13](#_Toc97583148)

[Figure 6.gender\_sold 14](#_Toc97583149)

[Figure 7. load file 15](#_Toc97583150)

[Figure 8. get\_index 15](#_Toc97583151)

[Figure 9.delete column 15](#_Toc97583152)

[Figure 10.get\_col 16](#_Toc97583153)

[Figure 11. Convert the words 16](#_Toc97583154)

[Figure 12.Convert the gender column 17](#_Toc97583155)

[Figure 13. rounded 17](#_Toc97583156)

[Figure 14. Rounded seniorityAsMonths 18](#_Toc97583157)

[Figure 15.Rounded senorityAsYear 18](#_Toc97583158)

[Figure 16.Output file 19](#_Toc97583159)

[Figure 17.Dashboard 1 19](#_Toc97583160)

[Figure 18.gender chart 20](#_Toc97583161)

[Figure 19.Gender\_bought chart 20](#_Toc97583162)

[Figure 20.Gender\_sold chart 21](#_Toc97583163)

[Figure 21.Dashboard 1 21](#_Toc97583164)

[Figure 22.Dashboard 2 22](#_Toc97583165)

[Figure 23.Follower\_sold chart 23](#_Toc97583166)

[Figure 24.Followers\_bought chart 23](#_Toc97583167)

[Figure 25.Gender chart 24](#_Toc97583168)

[Figure 26.Action 24](#_Toc97583169)

[Figure 27.DashBoard 3 25](#_Toc97583170)

[Figure 28. Map chart 26](#_Toc97583171)

[Figure 29.Language chart 26](#_Toc97583172)

[Figure 30.Contry\_sold chart 27](#_Toc97583173)

[Figure 31.Bought\_sold chart 27](#_Toc97583174)

[Figure 32.Action 28](#_Toc97583175)

[Figure 33.action 28](#_Toc97583176)

# I. Intro

The e-commerce platform is growing strongly across the country, the e-commerce market appears everywhere. There are a lot of unknowns when it comes to running an e-Commerce store, even if you have the analytics to make your decision. Users are an important factor in an e-commerce business. This dataset is intended to serve as a benchmark for an e-commerce fashion store. Using this dataset, you may want to try and understand what you can expect from your users and predetermine your level of development.

For example, if you find that most of your users are inactive, you can look at this dataset to compare store performance.

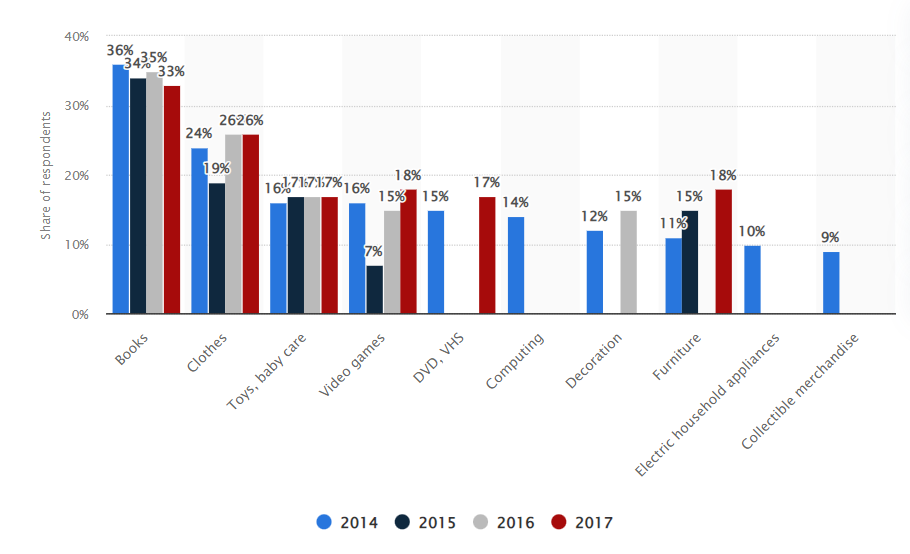


Figure .E-commerce (Dux, 2021)

This statistic displays results of a survey about C2C e-commerce which took place in France from the 28th of August to the 11th of September 2017. It reveals that books where the most common type of product to be purchased from private individuals. Indeed, over one third of respondents declared that they had already purchased a book from a private online seller in 2017. Clothes are the second sales category that had remained stable over the years, with a fall to 19 percent in 2015 but overall staying at around 25 percent. (Dux, 2021)

## 1. What is BI?

Business intelligence (BI) is a collection of procedures, systems, and technology that transform raw data into relevant information that drives lucrative business operations. It is a collection of tools and services designed to turn data into actionable intelligence and knowledge. BI has a direct impact on a company's strategic, tactical, and operational business decisions. BI promotes fact-based decision-making based on historical data rather than guesswork and intuition (Taylor, 2022)

## 2. Real examples of how to apply BI on business

A hotel owner uses BI analytical applications to gather statistical information regarding average occupancy and room rate. It helps to find aggregate revenue generated per room. It also collects statistics on market share and data from customer surveys from each hotel to decide its competitive position in various markets. Analyzing these trends year by year, month by month and day by day helps management to offer discounts on room rentals (Taylor, 2022)

## 3. Techniques Used in BI



Figure 2. BI (Taylor, 2022)

## 4. Data Visualization

Data is precise but difficult to grasp when represented as a set or matrix of integers. For example, are sales increasing, decreasing, or remaining stable? This becomes significantly more difficult when looking at more than one dimension of the data. As a result, data visualization in charts is a useful way to quickly grasp how to interpret the data (dataintegration, 2022)

## 5. Data Mining

Data mining is a computer-assisted process for discovering previously undiscovered or undetected relationships between data elements. Shopping basket analysis, measurement of products consumers buy together in order to promote other products; in the banking sector, client risk assessment is used to evaluate whether the client is likely to repay the loan based on historical data; in the insurance sector, fraud detection based on behavioral and historical data; and in medicine and health, analysis of complications and/or common diseases may help to reduce mortality. (dataintegration, 2022)

## 6. Reporting

Business users can benefit from BI technologies in the design, scheduling, and generation of performance, sales, reconciliation, and savings reports. BI tool reports efficiently gather and show information to support management, planning, and decision making. Once prepared, the report can be distributed automatically to a predetermined distribution list in the desired format, providing daily/weekly/monthly statistics (dataintegration, 2022).

## 7. Time-series Analysis Including (Predictive Techniques)

A temporal dimension is present in nearly all data warehouses and enterprise data. For instance, product sales, phone calls, patient hospitalizations, and so forth. It is critical to highlight changes in user behavior over time, product relationships, or changes in sales contracts based on marketing promotion. We may also attempt to forecast future patterns or outcomes based on historical data (dataintegration, 2022)

## 8. On-line Analytical Processing (OLAP)

OLAP is best known for its OLAP-cubes, which are used to visualize multidimensional data. Dimensions are displayed on the cube edges in OLAP cubes (e.g. time, product, customer type, customer age etc.). The cube's values represent the measured facts (e.g. value of contracts, number of sold products etc.). The drill-up, drill-down, and drill-across functions allow the user to navigate across OLAP cubes. The drill-up feature allows the user to effortlessly zoom out to additional fine-grained details. Drill-down, on the other hand, displays the information in greater detail. Finally, drilling-across implies that the user can move to another OLAP cube to view the relationships on a different dimension (s). All functionality is available in real-time (dataintegration, 2022)

## 9. Statistical Analysis

The theoretical foundations of statistical analysis are used to qualify the significance and reliability of observed relationships. Distribution analysis and confidence intervals are the most intriguing characteristics (for example for changes in user behaviors, etc). Statistical analysis is used to design and analyze data mining findings.

## 10. BI tools (Python)

****

Figure 3. BI tools (Nguyễn, 2020)

Python, a complex, high-level, object-oriented programming language, was created by Guido Van Rossum. It's simple to learn and is quickly becoming one of the best starting programming languages. Python makes use of automatic memory allocation and is completely dynamically typed. Python has complex high-level data structures as well as a simple and effective object-oriented programming paradigm. Python's command syntax is a big advantage since its clarity, ease of comprehension, and flexible typing make it ideal for scripting and application development across a wide range of domains and platforms.

There are many libraries in Python for numerical science and computing, such as SciPy and NumPy, which are used for general purposes in computing. And, there are specific libraries like: EarthPy for earth science, AstroPy for Astronomy,... In addition, Python is also widely used in machine learning, data mining and deep learning (Yang, 2020).

## 11. BI tools(tableau)

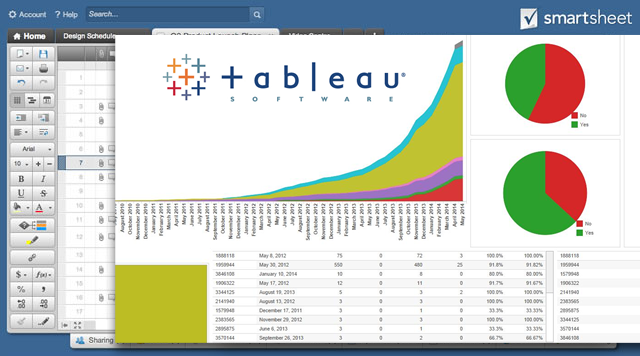
****

Figure 4. tableau

Tableau is a software program developed by the same-named firm in Washington, D.C. This application is widely used in the field of Business Intelligence. This software is concerned with data visualization, and it allows developers to visually depict the data and information they collect. In this way, it facilitates and smooths the process of analyzing numerical data and drawing conclusions. They are presented in the form of dashboards and sheets (Mai, 2021).

Compare Tableau to other business intelligence (BI) solutions like Power BI and Qlik Sense and the difference is clear. The three most prominent features of Tableau are:

* Real-time data analysis
* Combine data
* Data Collaboration

• Tableau facilitates objective data presentation. This tool also supports users in producing the most accurate trend forecasts. This is accomplished through the use of algorithm-based analysis. Tableau then lays the groundwork for users to make the best decisions possible. The application facilitates data retrieval in both raw and graphical versions (Mai, 2021).

• Tableau provides a lot of assistance to users when it comes to displaying data on a map. This program saves a wide range of information, such as postal codes, location names, contact information, and so on. In addition, visual representations of the maps, such as a flow map, a heat map, a secret map score, and so on, will be offered (Mai, 2021).

• Tableau software can handle a variety of data sources, including cloud data, SQL and NoSQL data, files, and so on (Mai, 2021)

## 12. Tableau application in business

Coca-Cola Bottling Company maximized operational efficiency

Company: Coca-Cola Bottling Company (CCBC), Coca Cola’s largest independent bottling partner Problem: Manual reporting processes restricted access to real-time sales and operations data.

Solution: Coca-Cola's business intelligence team handles reporting for all sales and delivery operations at the company. With their BI platform, the team automated manual reporting processes, saving over 260 hours a year—more than six 40-hour work weeks. Report automation and other enterprise system integrations put customer relationship management (CRM) data back into the hands of sales teams in the field through mobile dashboards that provide timely, actionable information and a distinct competitive advantage. A self-service BI implementation fosters more effective collaborations between IT and business users that maximize the expertise of participants. Analysts and IT can focus on big-picture strategy and long-term innovations such as enterprise data governance rather than manual research and reporting tasks (tableau, 2022). (tableau, 2022)

# II. Related work

## 1. Dataset

The data in our dataset is about a successful online fashion store with more than 9 million registered users. This dataset includes identifierHash, type, country, language, socialNbFollowers, socialNbFollows, socialProductsLiked, productsListed, productsSold, productsPassRate, productsWished, productsBought, gender, civilityGenderId, civilityTitle, hasAnyApp, hasAndroidApp,hasIosApp, seniorLastLogin, daysSince seniorityAsYears, countryCode. However, this dataset still needs editing. First we will remove 3 columns: type, identifierHash,civilityGenderId. Second, Modify the data in the column gender, language to the correct type, avoiding abbreviations. And finally, we'll round to the decimals of the senorityAsYears , senorityAsMonths columns.

|  |  |
| --- | --- |
| Columns | Meaning |
| identifierHash | Hash of the user's id |
| Type | The type of entity |
| Country | User's country (written in french) |
| Language | The user's preferred language |
| socialNbFollowers | Number of users who subscribed to this user's activity. New accounts are automatically followed by the store's official |
| socialNbFollows | Number of user account this user follows. New accounts are automatically assigned to follow the official partners |
| socialProductsLiked | Number of products this user liked |
| productsListed | Number of currently unsold products that this user has uploaded. |
| productsSold | Number of products this user has sold |
| productsPassRate | % of products meeting the product description. (Sold products are reviewed by the store's team before being shipped to the buyer.) |
| productsWished | Number of products this user added to his/her wishlist. |
| productsBought | Number of products this user bought |
| gender | user's gender |
| civilityGenderId | civility as integer |
| civilityTitle | Civility title |
| hasAnyApp | user has ever used any of the store's official app |
| hasAndroidApp | user has ever used the official Android app |
| hasIosApp | user has ever used the official iOS app |
| hasProfilePicture | user has a custom profile picture |
| daysSinceLastLogin | Number of days since the last login |
| seniority | Number of days since the user registered |
| seniorityAsYears | See seniority in years |
| seniorityAsMonths | see seniority in months |
| countryCode | user's country (ISO-3166-1) |

What is inside is more than just rows and columns. Make it easy for others to get started by describing how you acquired the data and what time period it represents, too. The data was scraped from a successful online C2C fashion store with over 9M registered users. The store was first launched in Europe around 2009 then expanded worldwide. I decided to choose this dataset because the value and information of the data is huge and useful for market analysis. And this dataset was collected from a large and successful e-commerce store many years ago. This is an opportunity for me to learn from successful people.

**List of chart**

**Gender chart**



Figure 5.Gender

We drag and drop gender into rows and drop gender into text to display Female and male from there we get the table gender.

This chart is used to choose to compare female and male in the dashboards.

**Gender\_ Bought Chart**

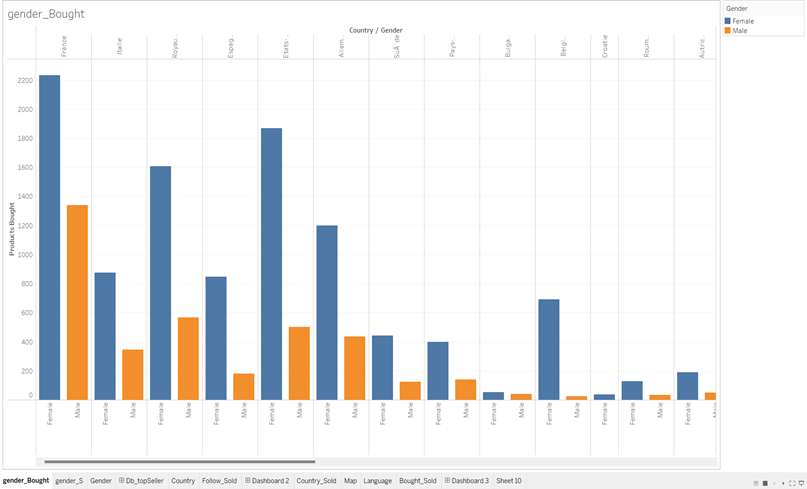


Figure 6.Gender\_bought

We drag country and gender into the columns. Next drag Product Bought into rows and filter for Product Bought descending. Finally, put gender in the color to distinguish Female and male.

This chart helps to see the difference between products bought by each female or male in different countries.

**Gender\_Sold Chart**

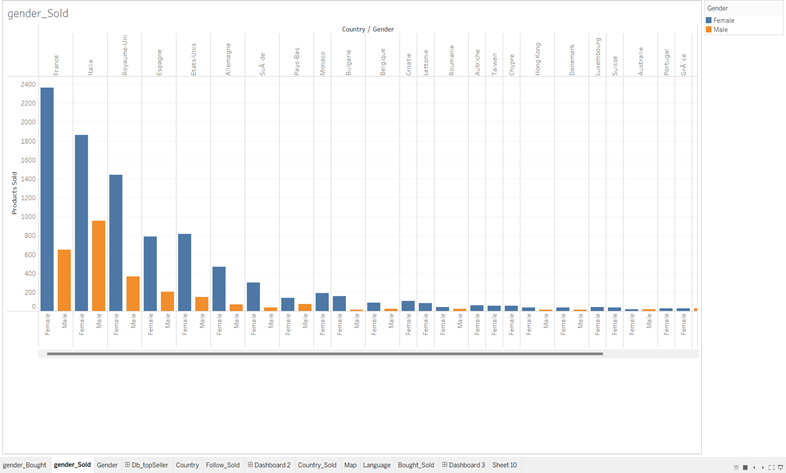


Figure 7.gender\_sold

We drag country and gender into the columns. Next drag Product Sold into rows and filter for Product Bought descending. Finally, put gender in the color to distinguish Female and male.

This chart helps to see the difference between products sold by each female or male in different countries.

## 2. Filter data

We need to read the csv file, I will import csv module, then use open() function to read the file, and then use the for loop to add all the data into a new array.

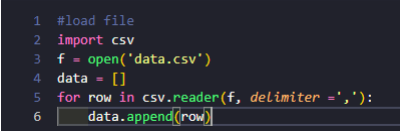


Figure 8. load file

get\_index:

Parameter a pass this function, save the index to variable current, and then runs a for loop in the data. It will compare a with each element in the data row header and return current.

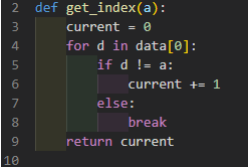


Figure 9. get\_index

Delete column:

Import the numpy library. To remove the column using delete() function , the delete function will receive an array, the position to be deleted (Get it by the get\_index function).



Figure 10.delete column

get\_column function:

The default parameter with the data of the file is the third parameter. Run a for loop in this function will in the data array and then find out which column has the same position as the second argument passed. Append that column to the given empty array.

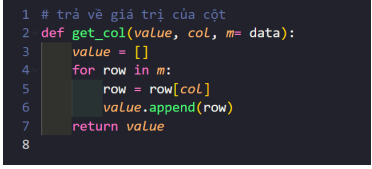


Figure 11.get\_col

Convert the words in the language column from abbreviations to full writing:



Figure 12. Convert the words

Create two arrays, new\_language and set\_language . Language column in data will be equal Set\_language. Create a dictionary containing the key as the abbreviation and the value as the full name of the language. Run the for loop in the set language from position 1 (ignoring the header). Run the for loop in the dictionary dict1. Compare the key word in the dictionary with the word element of the language column. If they are equal, append the value of that key to the new array. After finishing we get the new\_language array with the non-abbreviated letters. Replace the new\_languagearray in the language column of data.

The purpose of this function is to convert the abbreviated gender into full text:



Figure 13.Convert the gender column

Create gender array and assign it equal to the gender column in the data using the get\_col function. Create a array to assign the new value you want to change into. Let the for loop run through the gender array, for any element with a value equal to 'M', append the value 'Male' to new and otherwise append 'Female'. Once done, we get a new array with gender as the full text. Then replace the new array with the gender column in data.

Funtion rounded

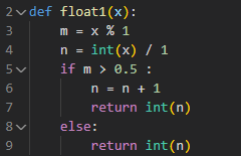


Figure 14. rounded

This function takes para x so compare the odd part after comma of odd if greater than 0.5 add 1 if less than 0.5 keep the same

Rounded senorityAsMonth

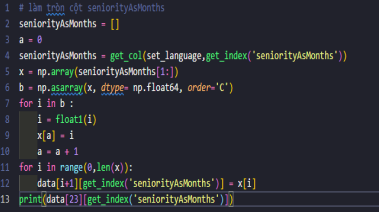


Figure 15. Rounded seniorityAsMonths

Assign senorityAsMonths to the senorityAsMonths column in data. Since I want to convert the data to float, I will turn senorityAsMonths into an array of numpyand convert to float64 using the asarrayfunction. Then run a for loop in the newly obtained array after converting to convert all to integers by function float1() . Then replace the new array in the senorityAsMonthscolumn to get the rounded numbers.

Rounded senorityAsYear

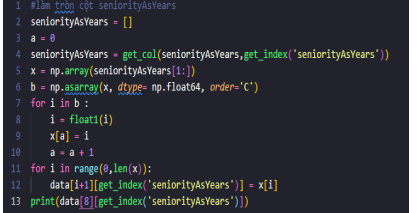


Figure 16.Rounded senorityAsYear

Assign senorityAsYears to the senorityAsYears column in data. Since I want to convert the data to float, I will turn senorityAsMonths into an array of numpy and convert to float64 using the as array() function. Then run a for loop in the newly obtained array after converting to convert all to integers by function float1(). Then replace the new array in the senorityAsYears column to get the rounded numbers.

Last step,

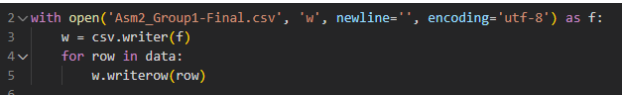


Figure 17.Output file

# III. Evaluation

I applied the tableau knowledge I learned to filter the dataset for e-commerce. Remove unnecessary (type, identityHash, civilityGenderId) columns, round to odd numbers (seniorityAsMonths, senorityAsYear). In addition, I changed some of the numeric abbreviated columns to write the full name for that type (gender and language). My short point is that I have not used libraries to optimize the code as much as possible, still writing pure code to filter data. In the near future, I will learn more knowledge about python and code support libraries, because optimizing code needs 3 criteria including: the code I make takes the least amount of space, the fastest and the short code most compact.

**Definition of dashboard:**

The main data visualization and analysis tool is the dashboard. Integral component of most BI software platforms and are widely used to deliver analytics information to business executives and workers are dashboards.



Figure .defination dashboard

Dashboard 1:

This dashboard shows purchased and sold products by gender and country.

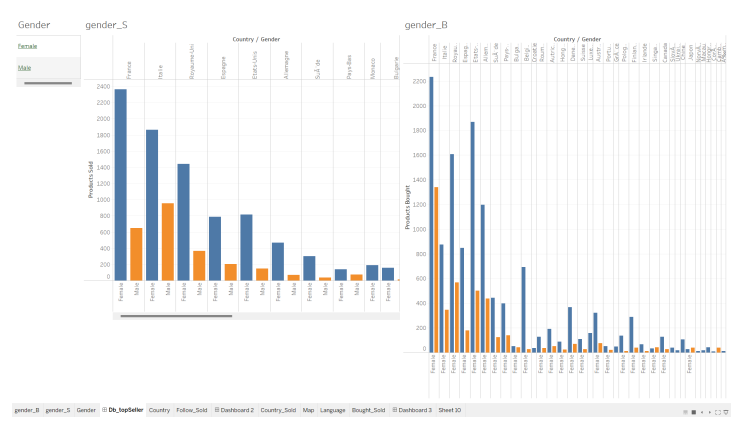


Figure 19.Dashboard 1

Gender chart

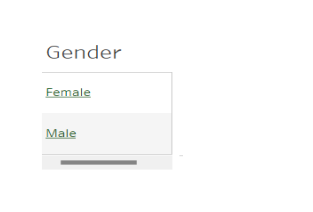


Figure 20.gender chart

Gender\_Bought Chart:

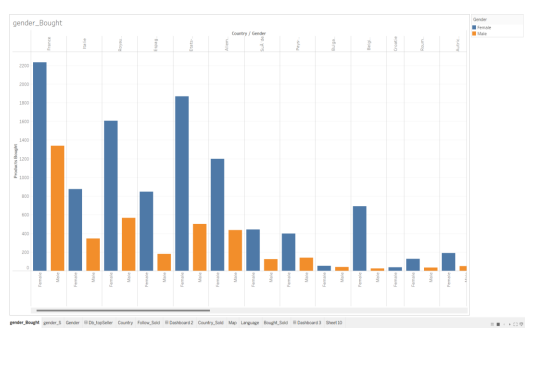


Figure 21.Gender\_bought chart

Gender\_Sold Chart

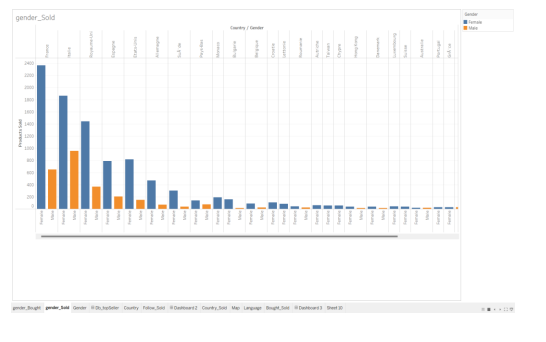


Figure 22.Gender\_sold chart

Gender Sold, bought Chart:

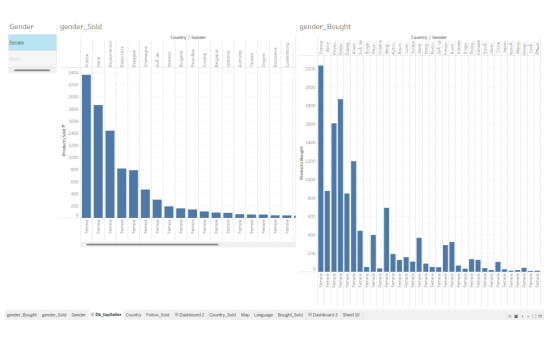


Figure 23.Dashboard 1

Dashboard 2:

This dashboard have socialNBFollowers and productsSold, gender, country, productsBought, socialNBFollows

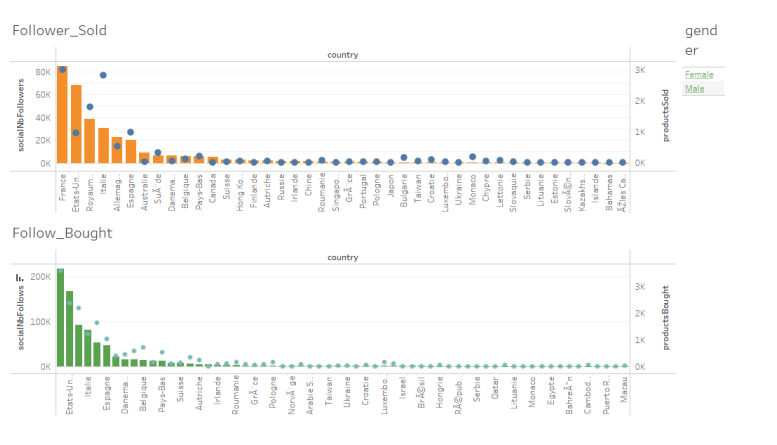


Figure 24.Dashboard 2

Followers\_Sold Chart:

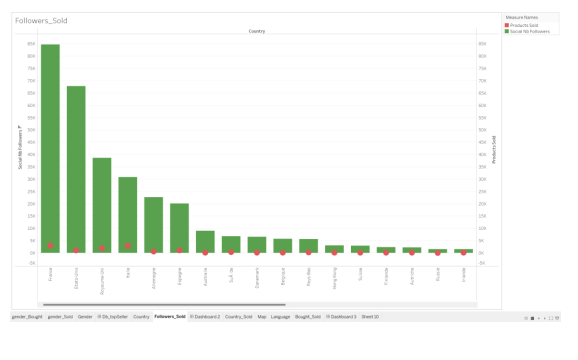


Figure 25.Follower\_sold chart

Follows\_Bought Chart

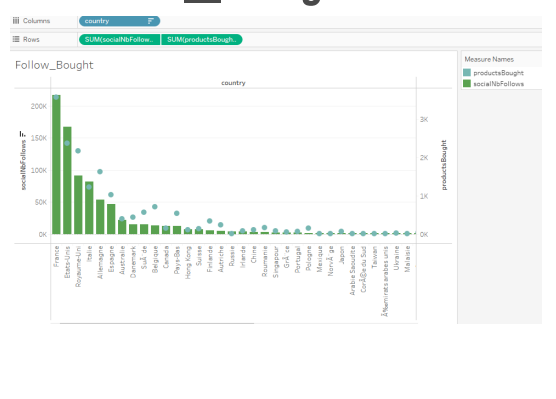


Figure 26.Followers\_bought chart

Gender Chart:

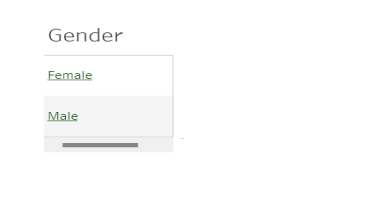


Figure 27.Gender chart

Dashboard Action:

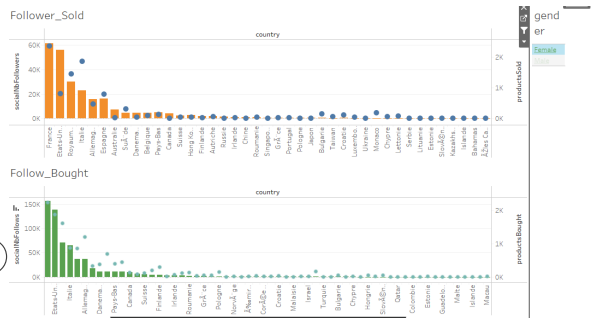


Figure 28.Action

Dashboard 3:

This dashboard shows the countries selling in what language, the number of items sold and purchased in that language, the number of items sold and purchased filtered by language and country.

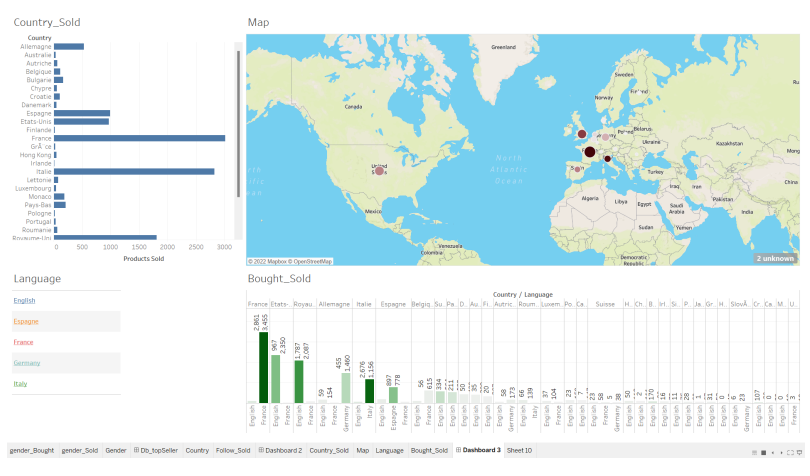


Figure 29.DashBoard 3

Map Chart:

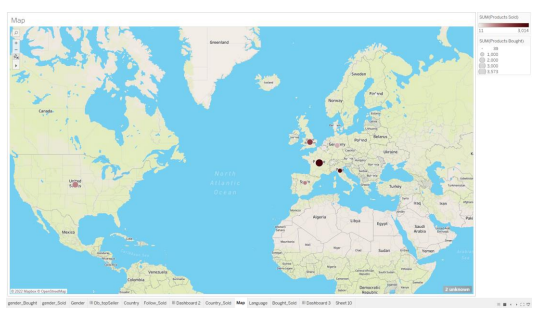


Figure 30. Map chart

Language Chart:

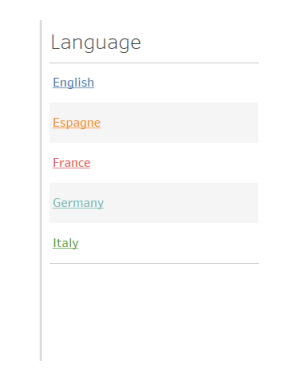


Figure 31.Language chart

Country\_Sold Chart

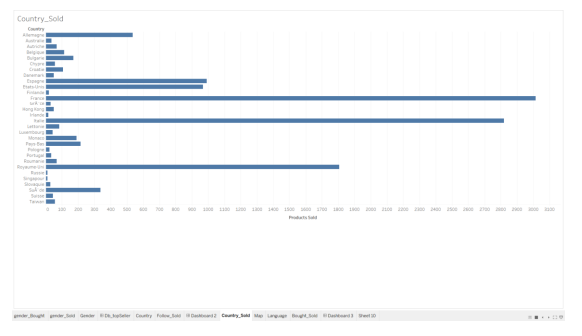


Figure 32.Contry\_sold chart

Bought\_Sold Chart

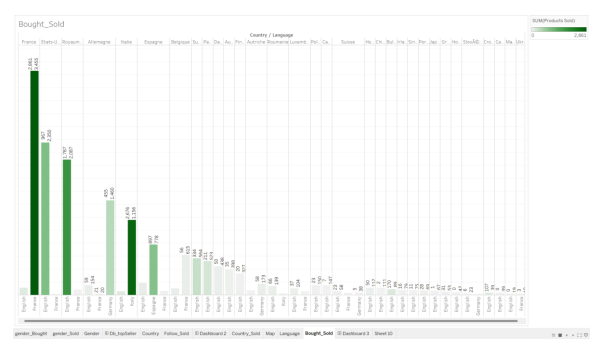


Figure 33.Bought\_sold chart

Dashboard Action:

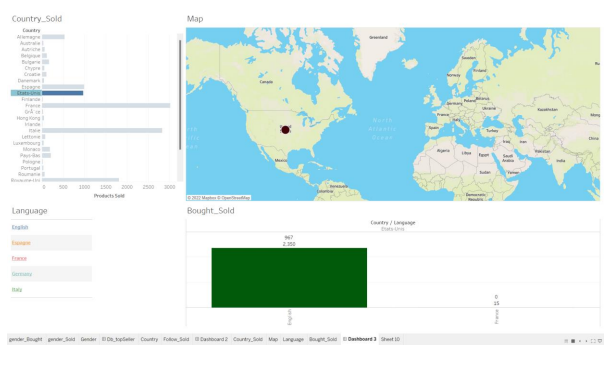


Figure 34.Action

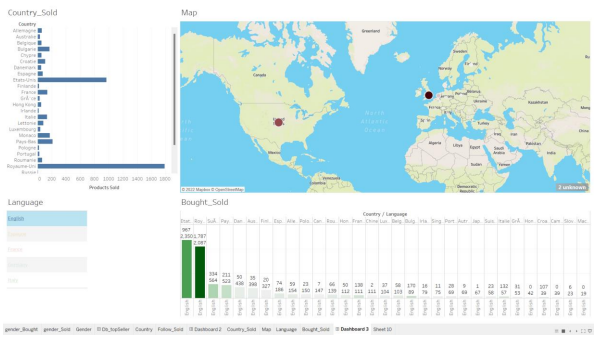


Figure 35.action

Point of view

For our company apply BI to collect the number of products sold by a country and the number of currently unsold products that has uploaded. From there, if the number of products sold is high and the number of unsold products is low, we will increase the number of items in that country. When the number of products sold is low but the number of unsold products is high, we will reduce the number of items in that country. Thanks to this BI tool we can assess the marketneeds of the countries. Thereby providing the essential needs of items to increase sales productivity and develop e-commerce.

# IV. Conclusion

This report I re-selected the e-commerce dataset with 24 columns covering a lot of information and data, after using python to remove unnecessary data and modify important data, I drew 3 dashboards from a lot of charts using tableau tools, and each of the dashboards drawn has important implications and effects for the industry. e-commerce. In the future I will apply multiple columns of data sets to draw more valuable dashboards for the e-commerce industry.

# Bibliography

dataintegration, 2022. *Business Intelligence*. [Online] Available at: <https://dataintegration.info/business-intelligence> [Accessed 2022].

Dux, N., 2021. *During the last twelve months, which products have you purchased from private individuals online?* [Online] Available at: <https://www.statista.com/statistics/427289/c2c-online-purchasing-penetration-product-type-france/> [Accessed 2022].

Mai, T., 2021. *Tableau là gì? Tính năng nổi bật của Tableau*. [Online] Available at: <https://tino.org/vi/tableau-la-gi/> [Accessed 2022].

Nguyễn, D., 2020. *Python là gì? Tại sao nên chọn Python?* [Online] Available at: <https://quantrimang.com/python-la-gi-tai-sao-nen-chon-python-140518> [Accessed 2022].

tableau, 2022. *Coca-Cola Bottling Company maximized operational efficiency*. [Online] Available at: <https://www.tableau.com/learn/articles/business-intelligence-examples> [Accessed 2022].

Taylor, D., 2022. *What is Business Intelligence? BI Definition, Meaning & Example*. [Online] Available at: <https://www.guru99.com/business-intelligence-definition-example.html> [Accessed 2022].