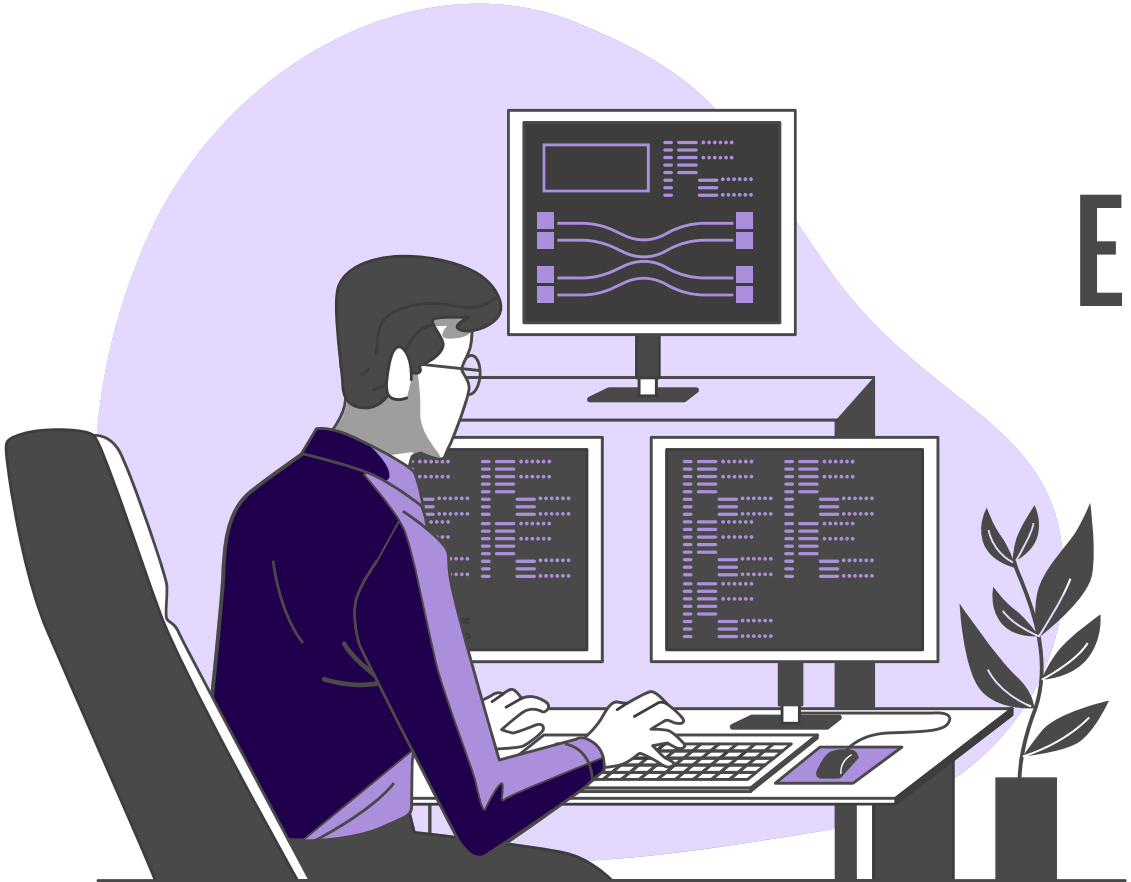
# ASSIGNMENT 2

| | | | |
|---|---|---|---|
| Qualification | BTEC Level 5 HND Diploma in Computing | | |
| Unit number and title | Unit 14: Business Intelligence | | |
| Submission date | 7/1/2022 | Date Received 1st submission | |
| Re-submission Date | | Date Received 2nd submission | |
| Student Name | Ngo Tuan Anh | Student ID | GCH190543 |
| Class | GCH0803 | Assessor name | Doan Trung Tung |
| Student declaration<br><br>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: | | |



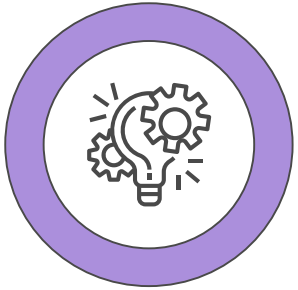
# E-commerce fashion

Nguyen Dang Khoa
Ngo Tuan Anh
Luong Ngoc Thai
Nguyen Vu Thai
Le Quoc Thai

# What is BI?

BI (Business Intelligence) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge. BI has a direct influence on the strategic, tactical, and operational business choices made by a company. BI encourages fact-based decision-making based on previous data rather than guesswork and intuition.

# 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 decides its competitive position in various markets.
By analyzing these trends year by year, month by month and day by day helps management to offer discounts on room rentals.

# BI techniques

## Querying

a request for specific data or information from a database

## Reporting

BI-first reports efficiently collect and present information to support management, planning, and decision-making.

## Predictive Techniques

Based on the historical data, we may also endeavor to predict future trends or outcomes.

## Statistical Analysis

Statistical analysis is used for devising and analyzing the results from data mining.

## Data Visualization

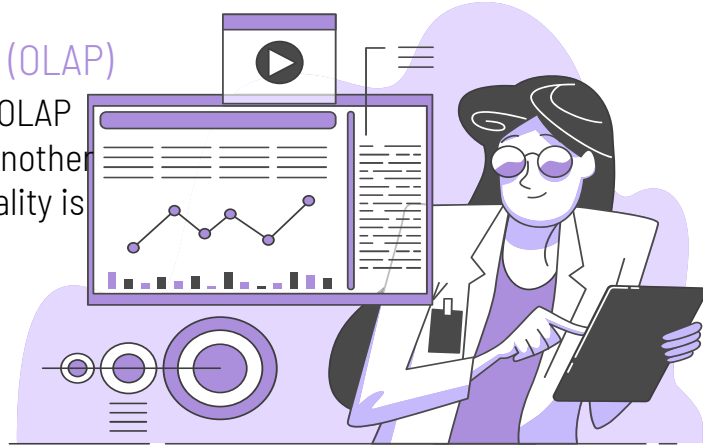
the visualization of data in charts is a convenient way to immediately understand how to interpret the data.

## Data Mining

Data mining is a computer supported method to reveal previously unknown or unnoticed relations among data entities.

## On-line Analytical Processing (OLAP)

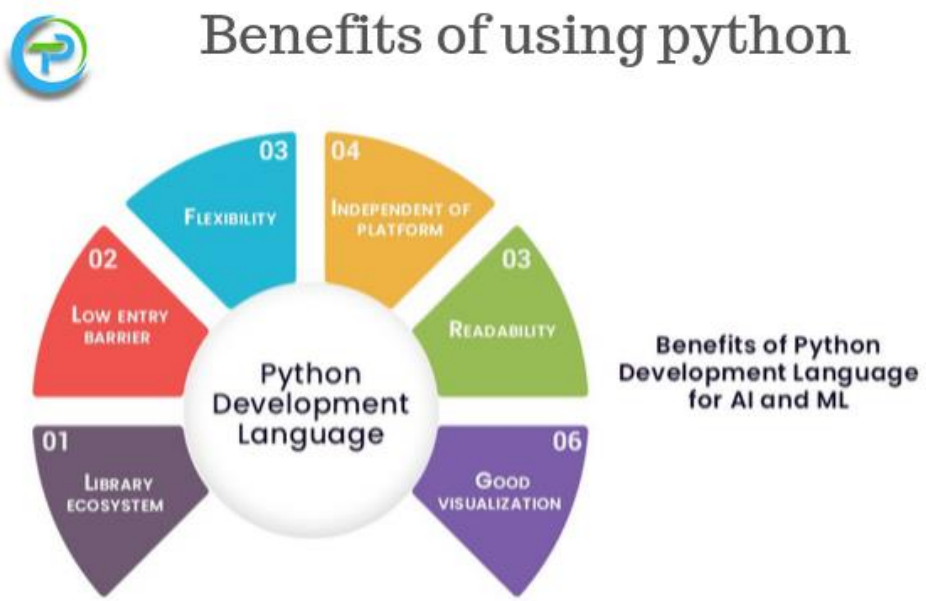
user can navigate to another OLAP cube to see the relations on another dimension(s). All the functionality is provided in real-time.



# BI tools (Python)

Guido Van Rossum designed Python, a sophisticated, high-level, object-oriented programming language. It's simple to pick up and is quickly becoming one of the finest beginning programming languages for beginners. Python employs automatic memory allocation and is totally dynamically typed. Python features sophisticated high-level data structures and an object-oriented programming paradigm that is both easy and effective. Python's command syntax is a significant benefit, since its clarity, ease of comprehension, and flexible typing make it a perfect language for scripting and application development across a wide range of sectors 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.



# BI tools (tableau)

Tableau is a software created by a company of the same name in Washington, USA. This software is popularly used in the field of Business Intelligence. This software deals with data visualization, allowing developers to visually represent the data and information they collect. In this way, it makes the process of studying numerical data and drawing conclusions easier and smoother. They come in the format of dashboards and sheets.

The three most prominent features of Tableau are:
Real-time data analysis
Combine data
Data Collaboration


- Tableau aids in the objective presentation of data. Simultaneously, this program assists users in making the most accurate trend forecasts. Algorithm-based analysis is used to do this. Tableau then builds a foundation to assist users in making the best decisions possible. The program aids in the retrieval of data in both raw and visualized formats.
- Tableau gives users a lot of help when it comes to showing data on a map. This program stores a variety of data, including postal codes, location names, contact information, and so on. In addition, visual representations of the maps will be provided, such as a flow map, a heat map, a secret map score, and so on.
- Tableau software can handle a variety of data sources, including cloud data, SQL and NoSQL data, files, and so on.

# Datasets

Enter a subtitle hThe 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,haslosApp, 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.

# Preprocess steps on datasets

# Preprocess steps on datasets

First of all we need to read the csv file.

```
#load file
import csv
f = open('data.csv')
data = []
for row in csv.reader(f, delimiter =','):
    data.append(row)
```

First I will import csv module, then use open() function to read file. After this, I will use reader() function to read all row in file and add to data

# Preprocess steps on datasets

This function is use to get the index of column that you want edit

```
2  def get_index(a):
3      current = 0
4      for d in data[0]:
5          if d != a:
6              current += 1
7          else:
8              break
9      return current
10
```

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

# Preprocess steps on datasets

This function is using to delete column by name

```
#delete column by name
import numpy as np
data = np.delete(data,(get_index('type')),1)
data = np.delete(data,(get_index('identifierHash')),1)
data = np.delete(data,(get_index('civilityGenderId')),1)
print(data)
```

Import the numpy library.Then using delete() function the to remove the column, the delete function will receive an array, the position to be deleted (Get it by the get_index function), and the last parameter equal to 1 is the collumn (which is the row if equal to 0).

# Preprocess steps on datasets

The purpose of this function is to get the data of a column.

```
# trả về giá trị của cột
def get_col(value, col, m= data):
    value = []
    for row in m:
        row = row[col]
        value.append(row)
    return value

```

This function takes 3 parameters, the third parameter is the default parameter with the data of the file that we get. In this function will run a for loop in the data array and then find out which column has the same position as the second argument passed. Then append that column to the given empty array.

# Preprocess steps on datasets

This function's purpose is to convert the words in the language column from abbreviations to full writing

```
#function of language
set_language = []
new_language = []
set_language = get_col(set_language,get_index('language'))
#Get unique value to create dict
set_language1= set(set_language[1:])
dict1 = {'es': 'Espagne' , 'de' : 'Germany' , 'fr' : 'France' , 'it' : 'Italy' , 'en' : 'English'}
for l in set_language[1:] :
    for v in dict1 : #get keys
        if l == v :
            l = dict1[v] #return value
            new_language.append(l)
for i in range(0,len(new_language)):
    data[i+1][get_index('language')] = new_language[i]
print(data[-1][get_index('language')])
```

Create two new empty arrays, set_language and new_language. Set_language will be equal to the language column in data. 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 word element of the language column with the key word in the dictionary. 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_language array in the language column of data.

# Preprocess steps on datasets

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

```
#function of gender
gender=[]
gender = get_col(gender,get_index('gender'))
new = []
for row in gender[1:]:
    if row == 'M':
        new.append('Male')
    else:
        new.append('Female')
for i in range(0,len(new)):
    data[i+1][get_index('gender')] = new[i]
```

Create an empty gender array and assign it equal to the gender column in the data using the get_col function. Create a new 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.

# Preprocess steps on datasets

This function's purpose is to round numbers of type float

```
2 def float1(x):
3     m = x % 1
4     n = int(x) / 1
5     if m > 0.5 :
6         n = n + 1
7         return int(n)
8     else:
9         return int(n)
```

Take the x, x % 1 parameter to get the decimal part. int(x) / 1 is for getting the integer part. If the decimal part is greater than 5 then increment the integer part by 1, otherwise keep the integer part.

# Preprocess steps on datasets

Rounding senorityAsMonths column

```
# làm tròn cột seniorityAsMonths
seniorityAsMonths = []
a = 0
seniorityAsMonths = get_col(set_language,get_index('seniorityAsMonths'))
x = np.array(seniorityAsMonths[1:])
b = np.asarray(x, dtype= np.float64, order='C')
for i in b :
    i = float1(i)
    x[a] = i
    a = a + 1
for i in range(0,len(x)):
    data[i+1][get_index('seniorityAsMonths')] = x[i]
print(data[23][get_index('seniorityAsMonths')])
```

Generate a =0, 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 numpy and convert to float64 using the asarray 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 senorityAsMonths column to get the rounded numbers.

# Preprocess steps on datasets

Rounding senorityAsYears column

```
#làm tròn cột seniorityAsYears
seniorityAsYears = []
a = 0
seniorityAsYears = get_col(seniorityAsYears,get_index('seniorityAsYears'))
x = np.array(seniorityAsYears[1:])
b = np.asarray(x, dtype= np.float64, order='C')
for i in b :
    i = float1(i)
    x[a] = i
    a = a + 1
for i in range(0,len(x)):
    data[i+1][get_index('seniorityAsYears')] = x[i]
print(data[8][get_index('seniorityAsYears')])
```

Generate a =0, 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.

# Preprocess steps on datasets

Last step, output another file

```
with open('Asm2_Group1-Final.csv', 'w', newline='', encoding='utf-8') as f:
    w = csv.writer(f)
    for row in data:
        w.writerow(row)
```

Then create a new file with the name 'Asm2_Group1-Final.csv' using the open() function and write all the data's data to the new file using the for loop.

# Dashboard 1

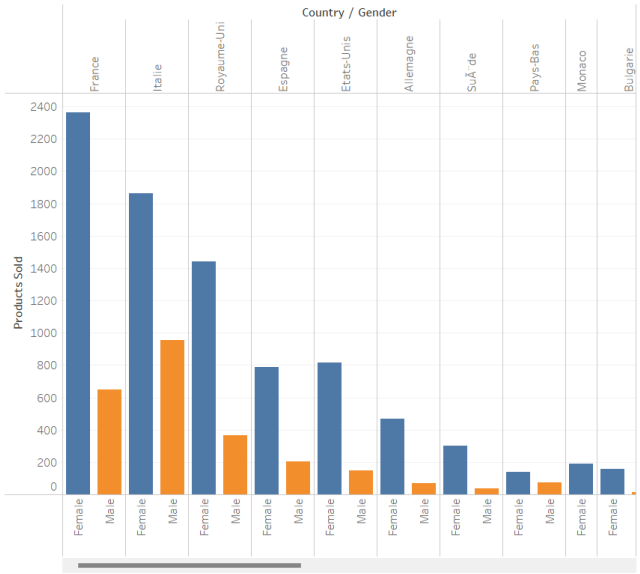
This dashboard displays products sold, products bought, gender and country. This table compares products sold and products bought for the country according to gender.

Gender

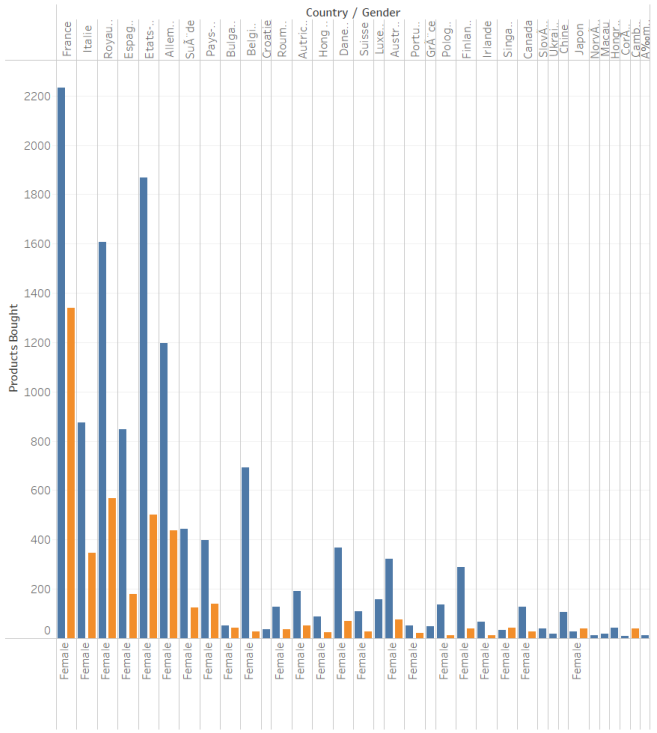
Female

Male

gender_S

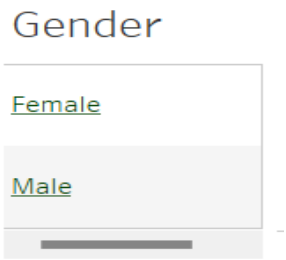



gender_B



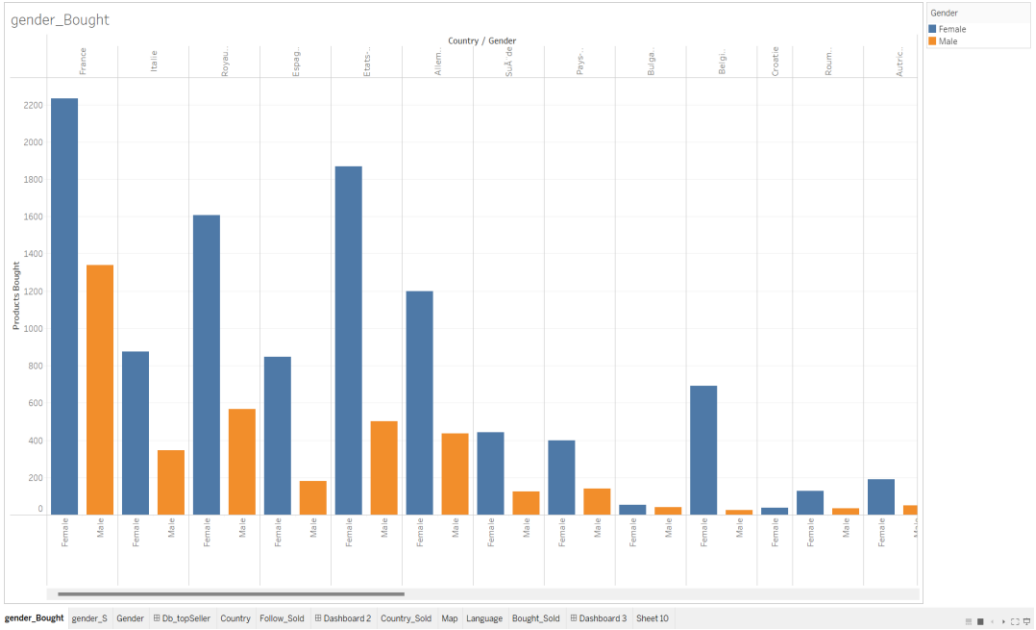

gender_B | gender_S | Gender | Db_topSeller | Country | Follow_Sold | Dashboard 2 | Country_Sold | Map | Language | Bought_Sold | Dashboard 3 | Sheet 10

# Gender chart



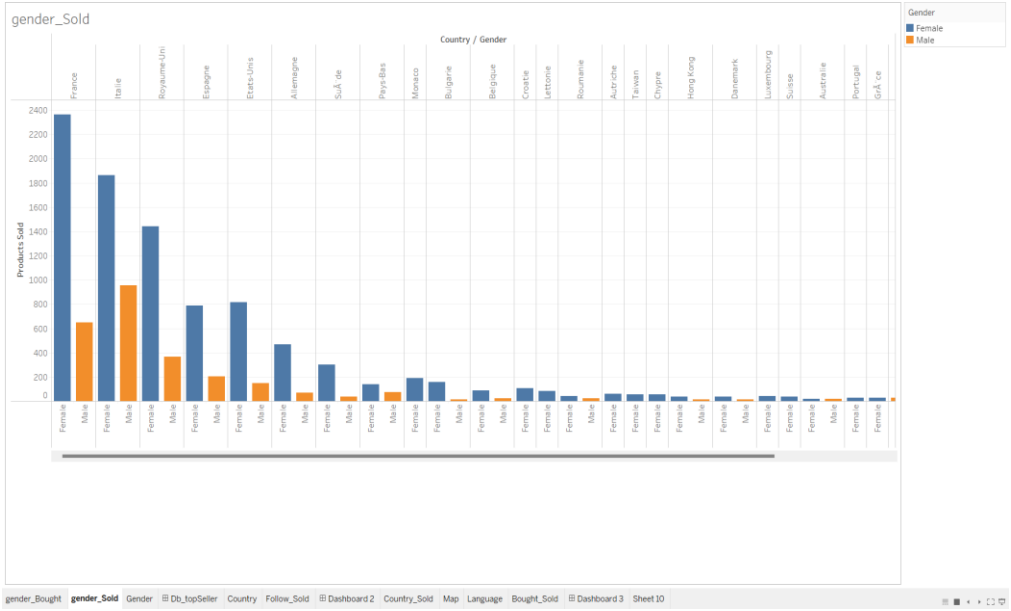

This chart is used to choose to compare female and male in the dashboards.

# Gender_ Bought Chart



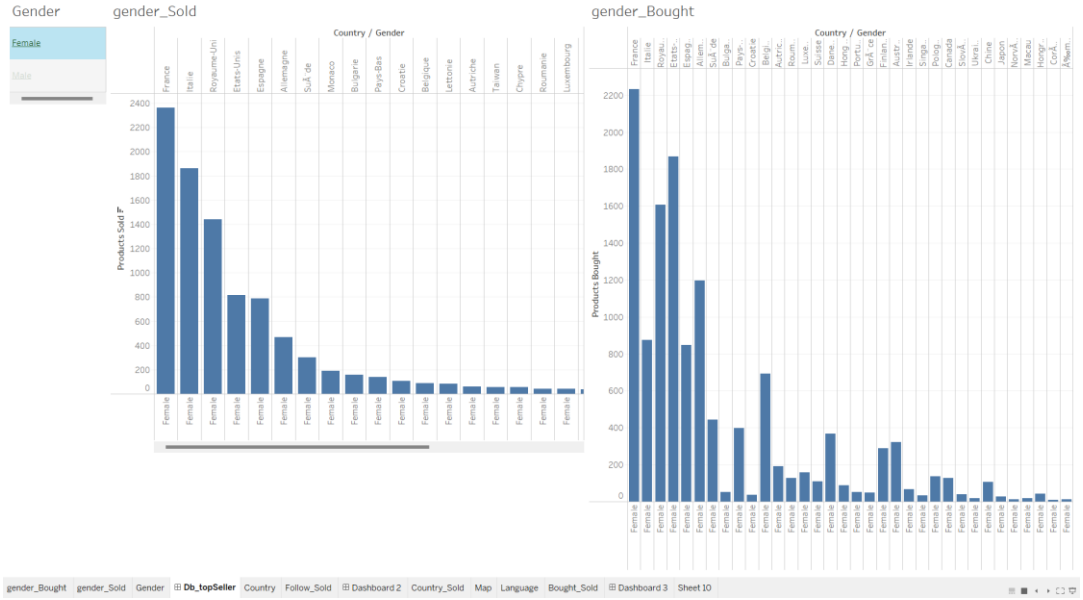

This chart helps to see the difference between products bought by each female or male in different countries.

# Gender_Sold Chart




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

# Dashboard Action




This is when moving the cursor to the gender, it will display general information in the products bought and products sold chart.

# Dashboard 2

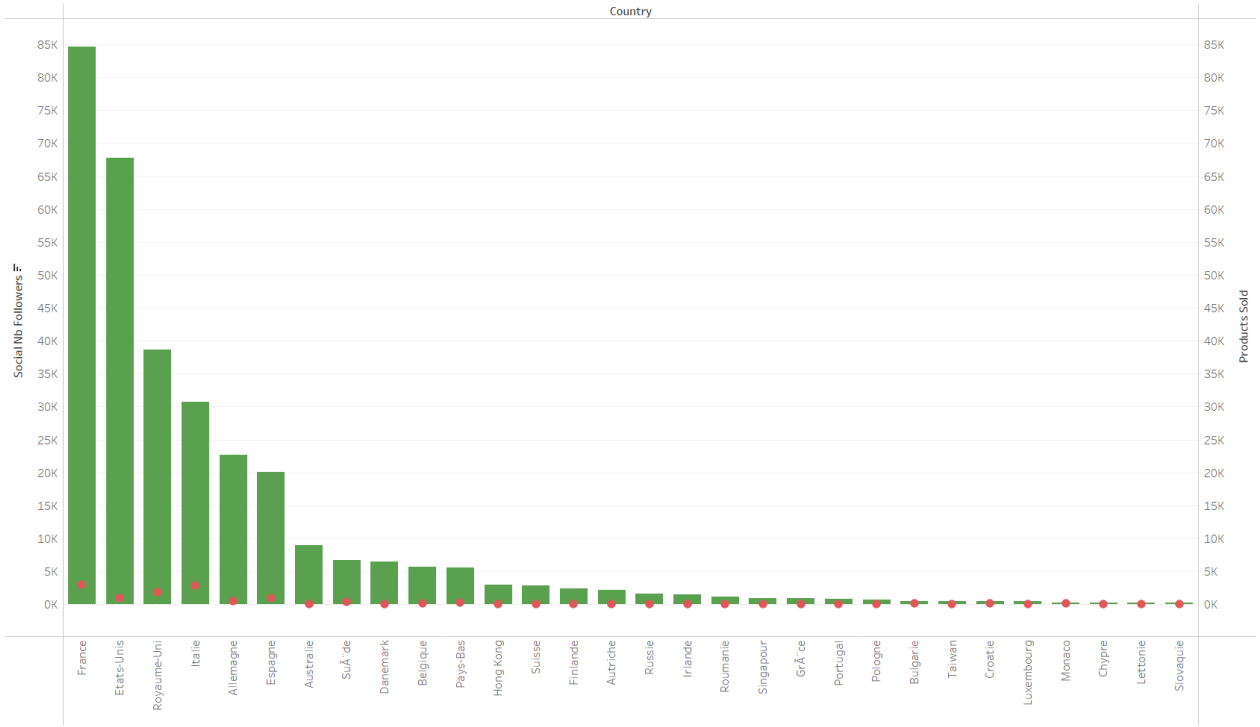
This dashboard socialNBFollowers and productsSold, gender, country. This dashboard compares the number of followers and the number of products sold, filtered by gender.

Gender

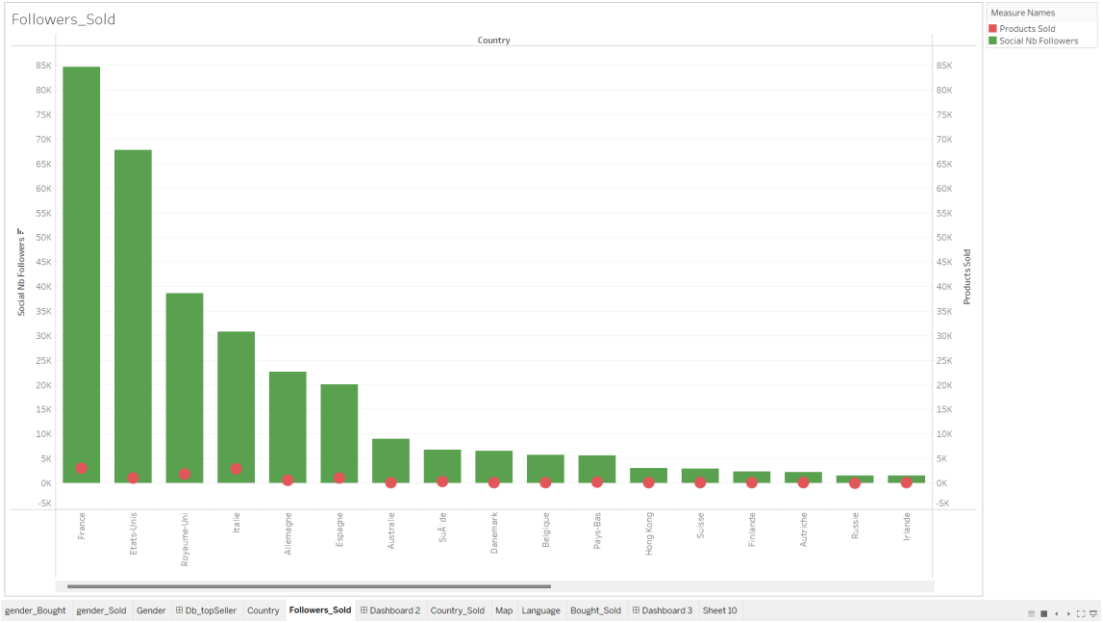
Female

Male

## Follow_Sold

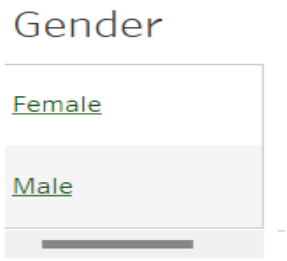


# Followers_Sold Chart



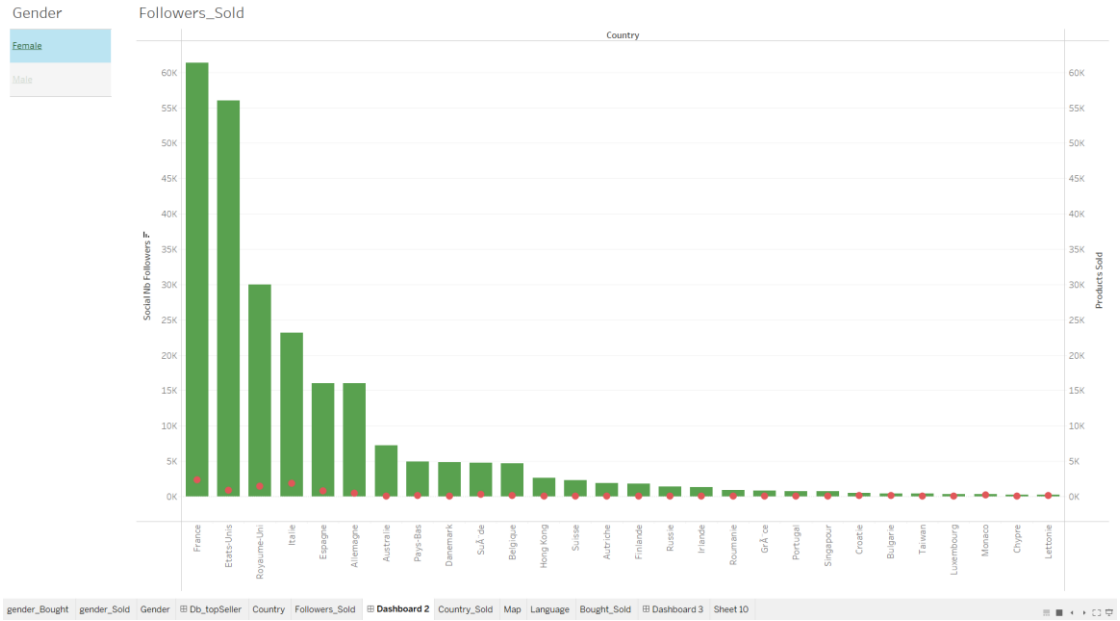

This chart helps to compare socialNBFollowers and productsSold of a country

# Gender Chart




This chart is used to filter by gender

# Dashboard Action

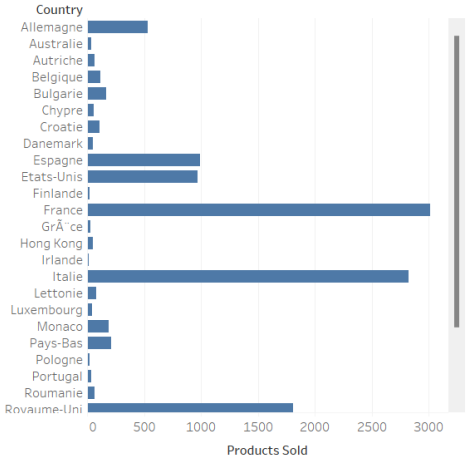



The action of this dashboard is to filter by male and female

# 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.

## Country_Sold

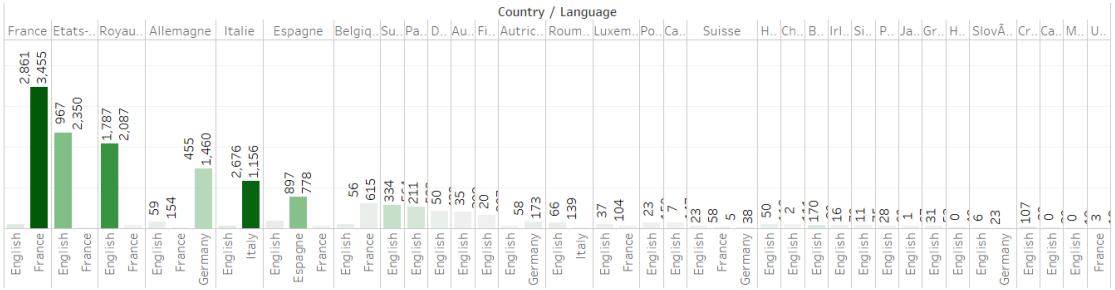



## Map


## Language

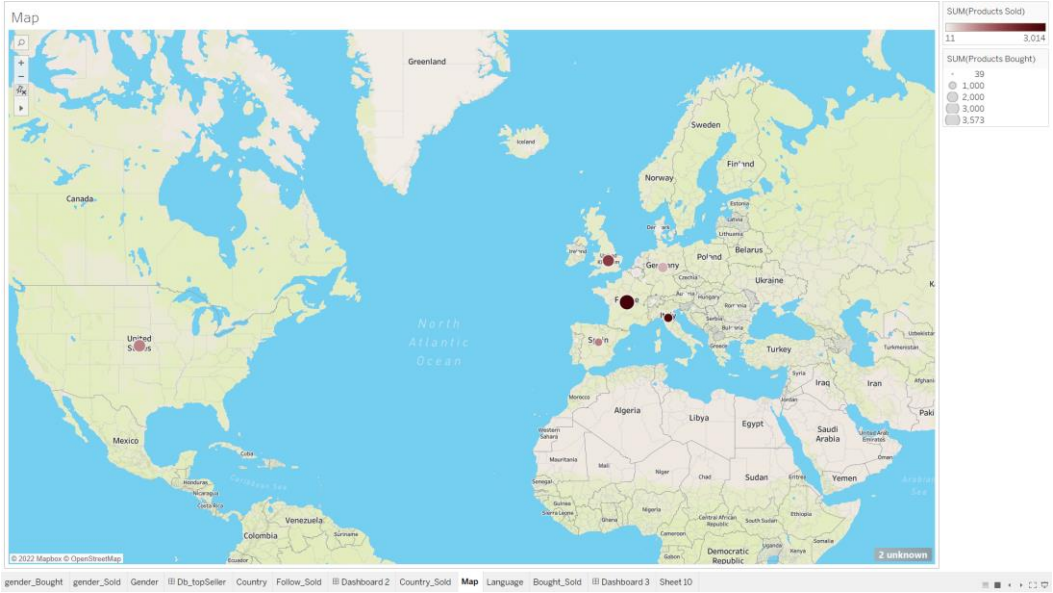
- English
- Espagne
- France
- Germany
- Italy

## Bought_Sold



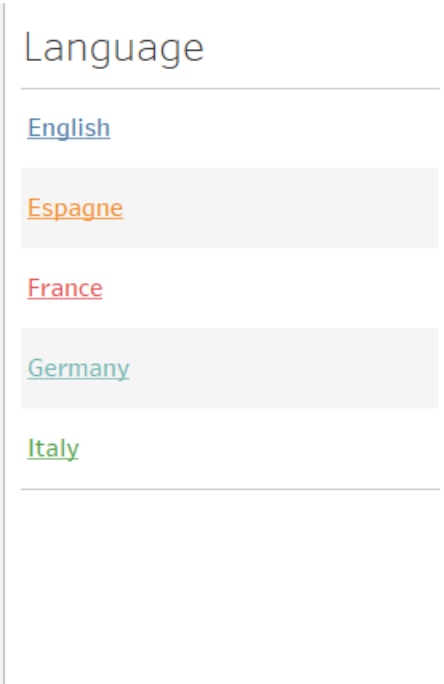
# Map Chart

This chart helps to show the quantity sold and bought by countries.

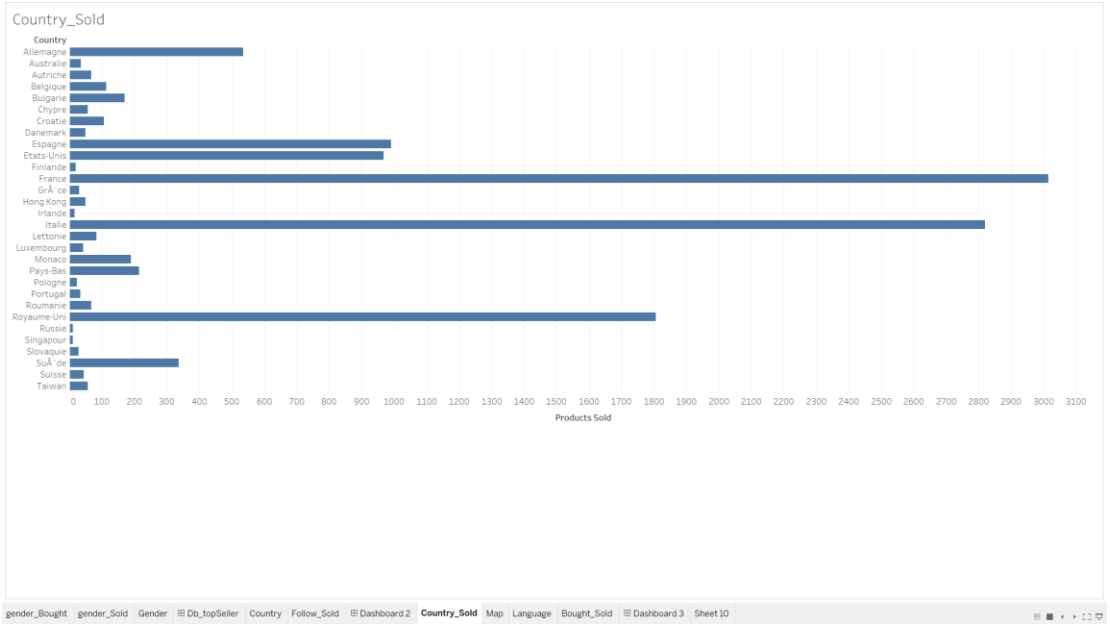


# Language Chart

This chart for language filtering

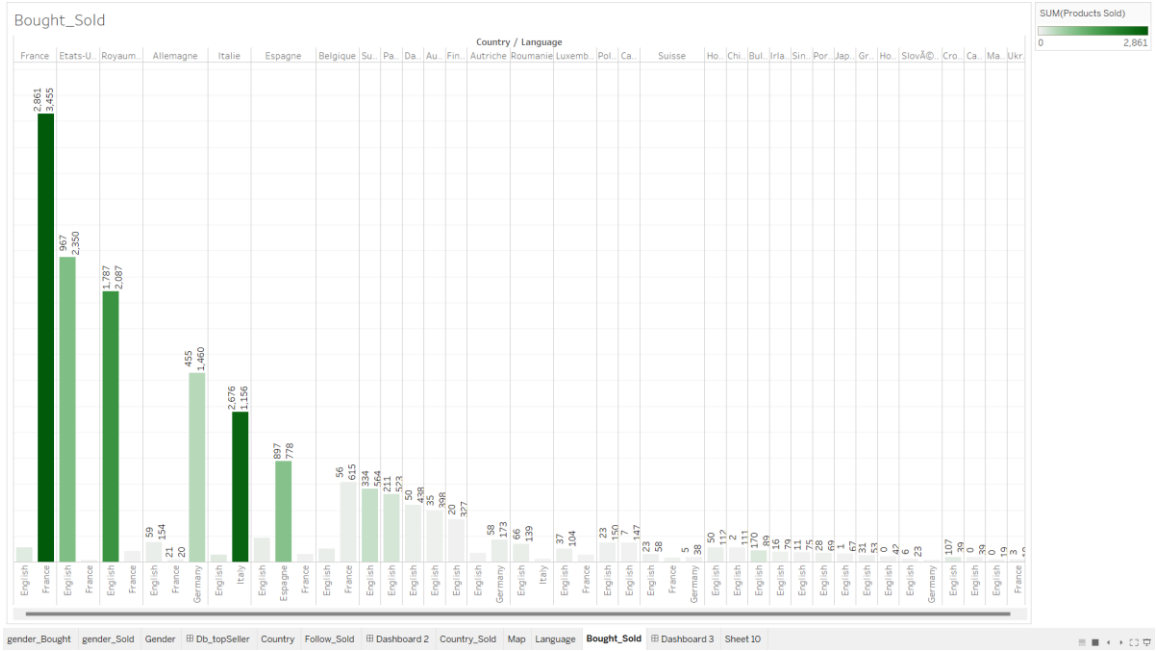


# Country_Sold Chart




This chart shows the number of sales through the country.

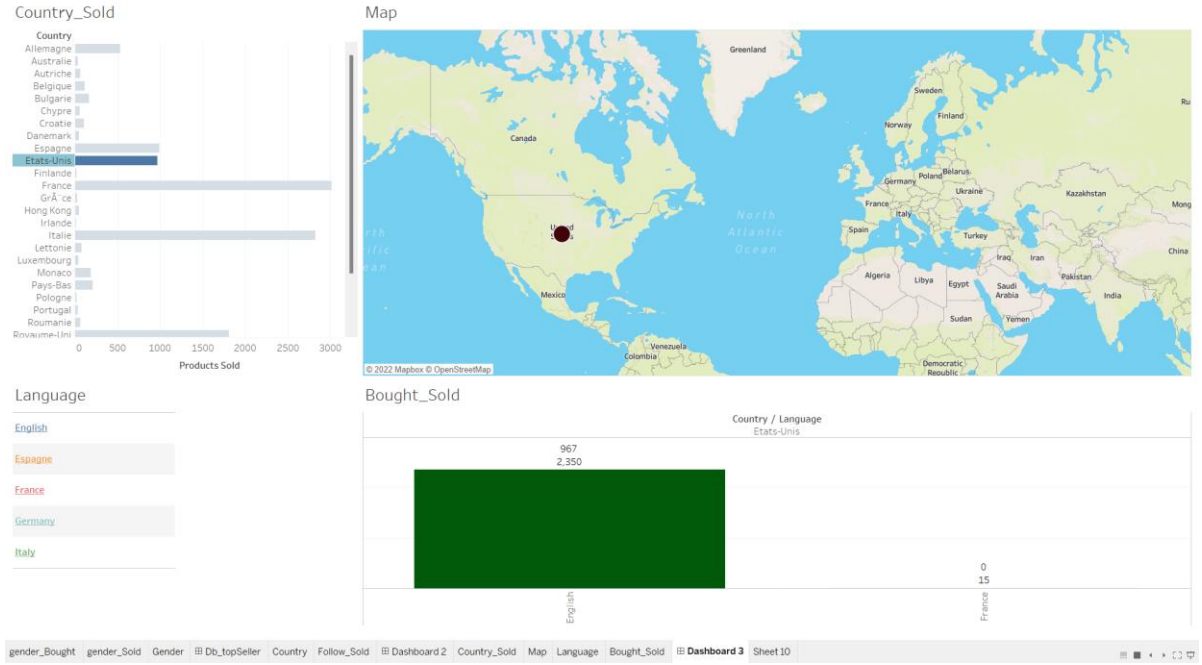
# Bought_Sold Chart




This chart shows what languages are sold in and how much has been sold and bought

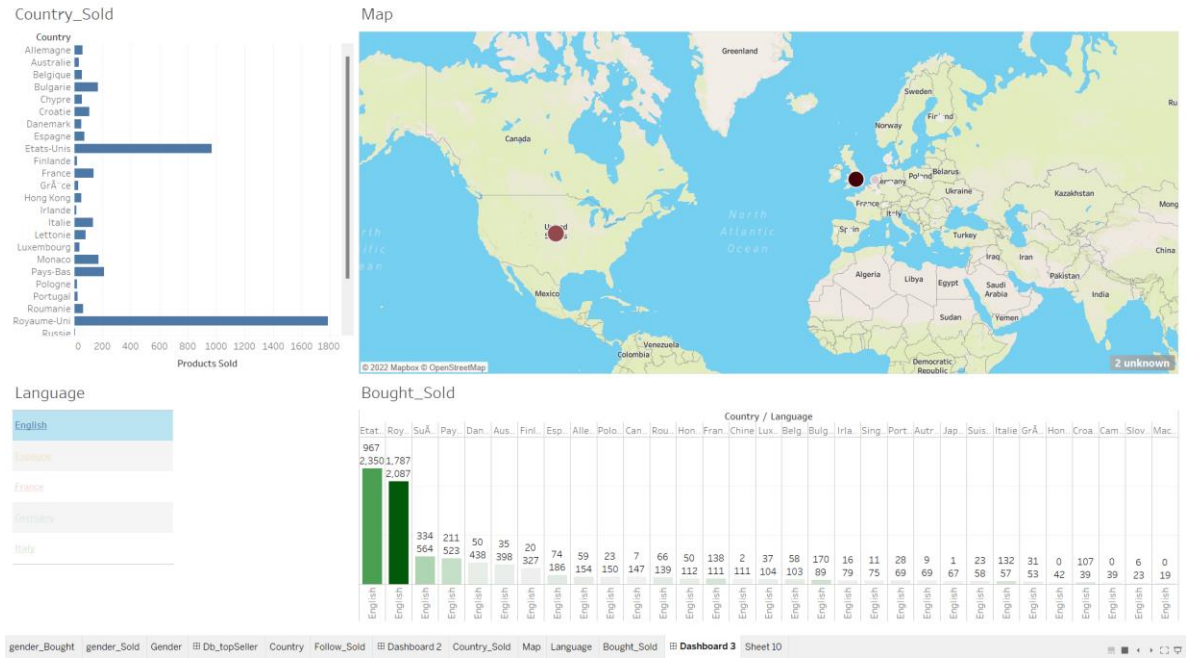
# Dashboard Action

This dashboard works by selecting a country, and showing what language it sold and bought in and how much it bought and sold

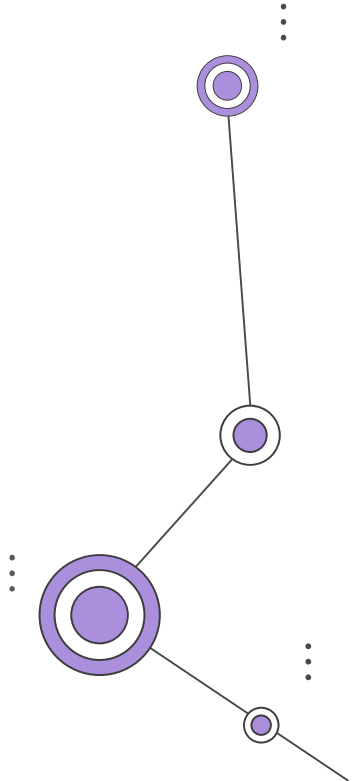


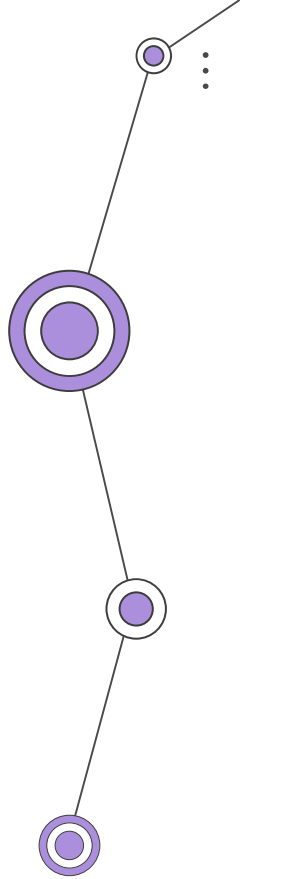
# Dashboard Action

The operation of this dashboard is to filter by countries of sale according to the selected language



# Thanks For Watching!





# Reference

https://www.guru99.com/business-intelligence-definition-example.html
https://dataintegration.info/business-intelligence
https://quantrimang.com/python-la-gi-tai-sao-nen-chon-python-140518
https://tino.org/vi/tableau-la-gi/
https://dataintegration.info/business-intelligence