Ecole Supérieure Privée d'Ingénierie
Et de Technologies

Engineer Summer internship

Subject :

Social Network Sentiment Analysis : Esprit

Developed by **Samar Jberi**

Supervised by **Mohammed Tlili**

August 6th to September 30th, 2021

Academic Year: 2021/2022

Summary

Table of figures

General Introduction

In order to apply the different skills we have acquired during the past couple of years, students conduct a summer internship after their second year of engineering studies.

Supported by the Private Higher School of Engineering and Technologies Esprit, I was offered the opportunity to work on the subject entitled "Social Network analysis" that took place from 06/08/2020 to 30/09/2021, under the supervision and help of Mr. Mohammed Tlili.

During this internship I had the chance to discover new technologies and acquire new skills.

Dived into the data and created a neat deliverable out of it.

In this report, I will list the different tasks performed as well as the steps I followed to develop my project.

In the first chapter, I will start by a presentation of the subject, study of the existing, and proposed solutions.

The second chapter will deal with the text mining part of my project, while the third chapter will deal with data visualization.

Finally, a conclusion will summarize the work done and the results obtained.

I. General project context:

In this chapter I will start by presenting the host organization, followed by a deep analysis of the project context including a study of the existing, the problematic at hand and the proposed solution. Lastly, I will specify the adopted methodology.

1. Presentation of the host organization:

Figure 1: Esprit Logo

Founded in 2003, Esprit has managed to become an internationally recognized establishment.

In addition to being accredited by the Tunisian Ministry of Higher Education and Scientific Research, the quality of Esprit education has become well renown over the years.

Following the lead of bigger higher education facilities such as MIT, Esprit builds the pedagogy of engineering training around professional scenarios. Esprit aims towards excellence.

And for this exact reason Esprit cares deeply about its public image, collecting and taking into consideration all the feedback in order to continuously improve its programs and keep up with the expectations to offer the best college experience possible for its students and employees.

2. Project context:

2.1 Study of the existing:

Nowadays, people tend to share so much of themselves and their feelings on internet and social networks.

This usually leads to communicate large amount of data.

these data can be very useful if it was correctly analyzed.

And since Esprit is one of the biggest private schools, there are lots of unexploited data that can make a huge difference in the University growth.

2.2 Problematic:

The educational system has developed to include198 public higher education institution versus **63** private institutions according to a report conducted by the British council called "Education in North Africa" in 2014.

the numbers of students enrolled in higher education has experienced a significant boost from over **2.6%** in 1974 to **35.2%** by 2015.

With such tight competition within the private sector, each private institute aims to step up in order to maintain its success and notoriety

The question asked, how can these data help in the growth of Esprit against other competitors in the private sector?

2.3 Proposed solutions:

Based on identified needs and my knowledge about the subject since I'm an Esprit student myself, I was assigned the project "Social Network analysis Esprit" in pursuance of finding solutions for the issues mentioned in the problematic.

Since every private facility relies heavily on its public image and, I realized that it is an essential starting point to consult the public reviews online about Esprit, study its foundation and followers and evaluate the interactions that take place on the social networks in order to identify and solve the most frequent problems.

To serve this purpose, I had to go through several steps:

- Scrapping the most relevant information (followers, employees, comments, interactions…) about Esprit on social networks (Facebook, LinkedIn)
- Shaping and cleaning the data, establishing relationships between the different data.
- Text Mining to detect the nature of the content and the most frequent searched terms.
- Using Data Visualization tools to classify posts based on different criteria to make the data more legible and have a general grasp on the image of Esprit.

2. 4. Adopted methodology - CRISP-DM

on this project I used CRISP-DM methodology which stands for cross-industry process for data mining.

it provides a structured approach to planning a data mining project.

it's composed essentially of 6 stages:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

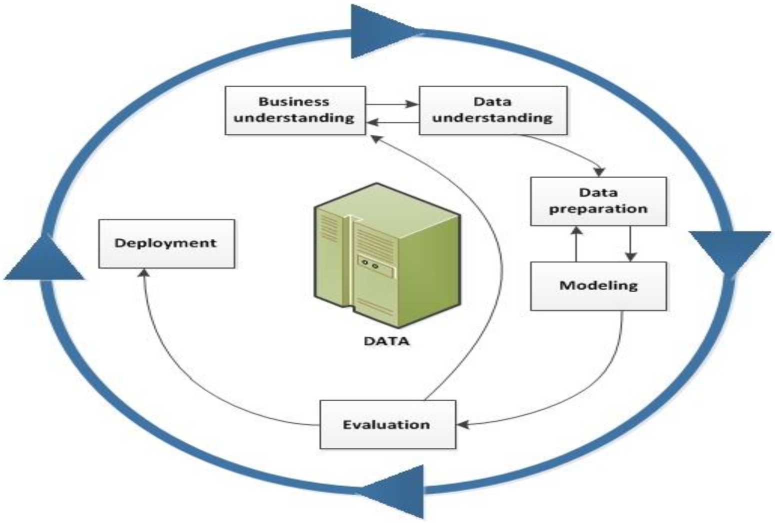


Figure 2: CRISP-DM Methodology

Conclusion

In this chapter, I got to explain the project’s context starting from the study of the existing to the proposed solution. Then proceeded to present my methodology. In the next chapter, I will be demonstrating the implementation of my data.

3. Implementation:

In this chapter I’m going to give a detailed description of the practical work I did. I’m going to start by listing the different tools I used, followed by the data model, screenshots of the integration work on power BI, screenshots of the Text Mining and Data Visualization work.

3.1. Used tools

3.1.1.Microsoft Power BI

Power BI is a business analytics service by Microsoft.

it comes with interactive visuals and business intelligence capabilities.

it is known for being super user friendly. Power BI provides cloud-based business intelligence services, known as "Power BI Services", along with a desktop-based interface, called "Power BI Desktop".

3.1.2.Python

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

4.Data Warehouse

4.1.Data Warehouse modeling approach

Designing a Data Warehouse is an essential part of business development. In this segment, I’m going to take an in depth look at the data warehouse modeling approach that I chose:

Bottom-Up approach by Ralph Kimball:

I used Ralph Kimball approach because it matched my needs. Since it is a bottom-up approach. i started by creating data marts going to my final data warehouse which made the ideal choice.

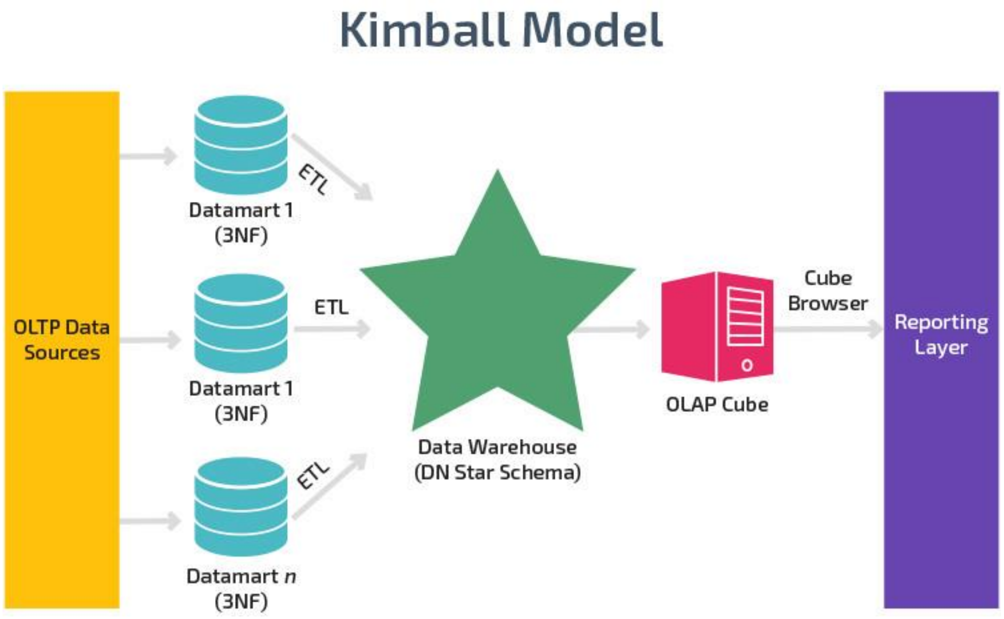


Figure 3: Bottom-Up approach by Ralph Kimball

II. Data Scraping:

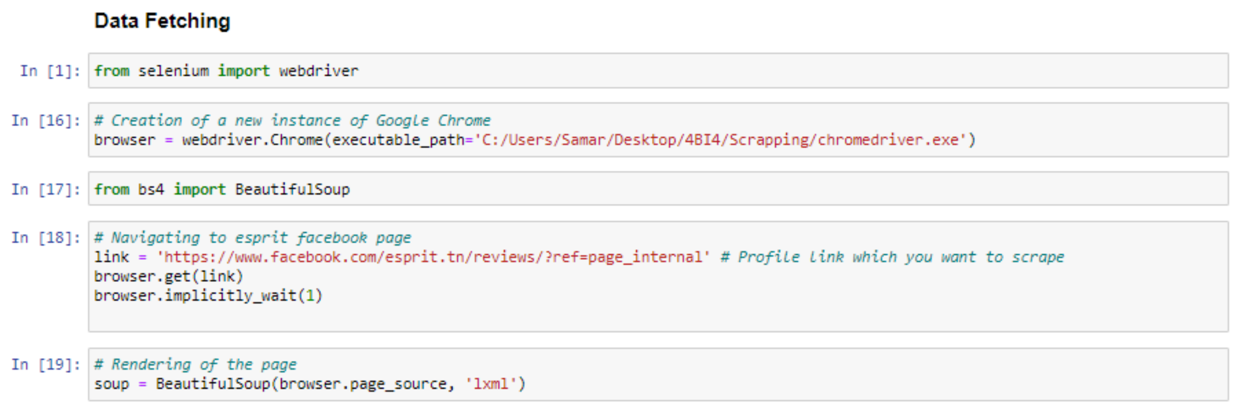
In order to gather maximum of information about Esprit online, I did some web scraping using python, relying essentially on two important libraries: selenium and Beautifulsoup.

Beautiful Soup is a Python HTML and XML document parsing library created by Leonard Richardson. It produces a syntax tree which can be used to search for elements or modify them.

Selenium is a powerful tool for controlling web browsers through programs and performing browser automation. It is functional for all browsers, works on all major OS and its scripts are written in various languages i.e. Python, Java, C#.

l will be working with Python.

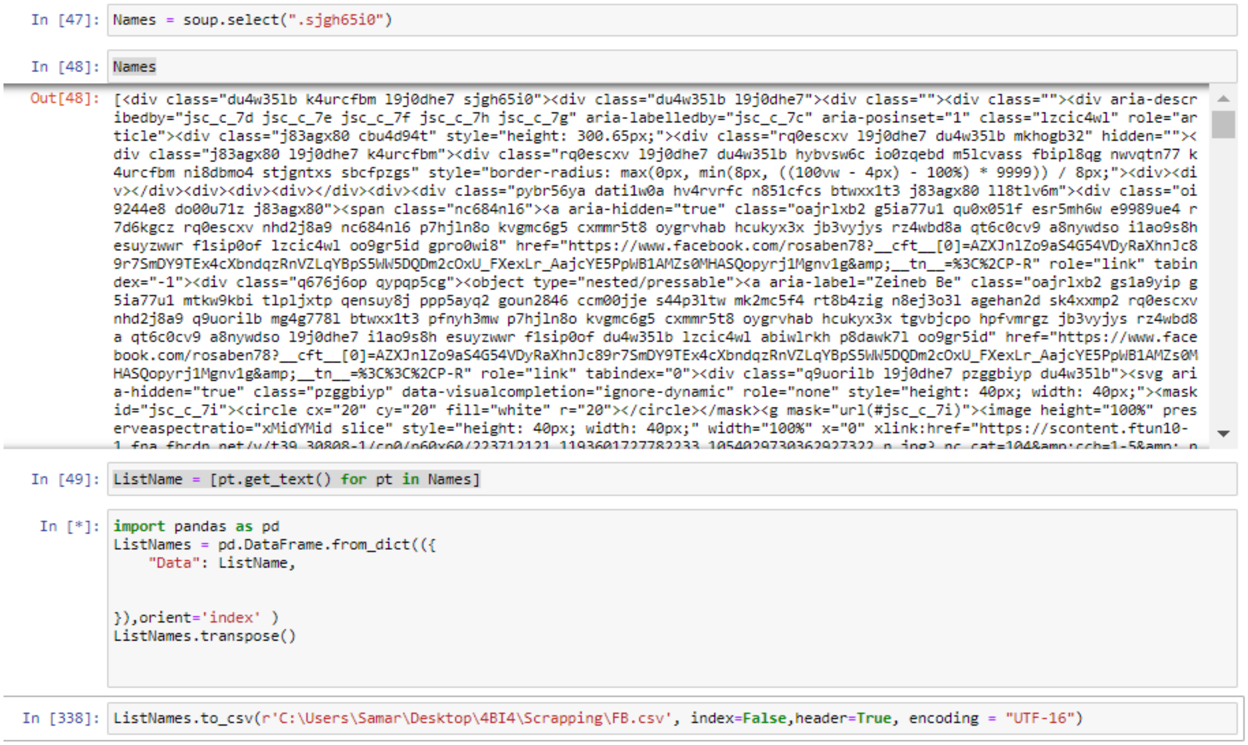
In this first part I started by importing the web driver and connected to the web page then fetched the XML content using beautifulSoup.



Data Fetching

```
In [1]: from selenium import webdriver

In [16]: # Creation of a new instance of Google Chrome
         browser = webdriver.Chrome(executable_path='C:/Users/Samar/Desktop/4BI4/Scrapping/chromedriver.exe')

In [17]: from bs4 import BeautifulSoup

In [18]: # Navigating to esprit facebook page
         link = 'https://www.facebook.com/esprit.tn/reviews/?ref=page_internal' # Profile link which you want to scrape
         browser.get(link)
         browser.implicitly_wait(1)

In [19]: # Rendering of the page
         soup = BeautifulSoup(browser.page_source, 'lxml')
```

Figure 5: Connecting to the web Page



```
In [47]: Names = soup.select(".sjgh65i0")

In [48]: Names

Out[48]: [<div class="du4w351b k4urcfbm l9j0dhe7 sjgh65i0"><div class="du4w351b l9j0dhe7"><div class=""><div class=""><div aria-descr
         ibedby="jsc_c_7d jsc_c_7e jsc_c_7f jsc_c_7h jsc_c_7g" aria-labelledby="jsc_c_7c" aria-posinset="1" class="lzcic4wl" role="ar
         ticle"><div class="j83agx80 cbu4d94t" style="height: 300.65px;"><div class="rq0escxv l9j0dhe7 du4w351b mkhogb32" hidden=""><
         div class="j83agx80 l9j0dhe7 k4urcfbm"><div class="rq0escxv l9j0dhe7 du4w351b hybvsw6c io0zqebd m5lcvass fbipl8qg nwvqtn77 k
         4urcfbm ni8dbmo4 stjgntxs sbcfpzgs" style="border-radius: max(0px, min(8px, ((100vw - 4px) - 100%) * 9999)) / 8px;"><div><di
         v></div><div><div><div></div><div><div class="pybr56ya dati1w0a hv4rvrfc n851cfcs btwxx1t3 j83agx80 ll8tlv6m"><div class="oi
         9244e8 do00u71z j83agx80"><span class="nc684nl6"><a aria-hidden="true" class="oajrlxb2 g5ia77u1 qu0x051f esr5mh6w e9989ue4 r
         7d6kgcz rq0escxv nhd2j8a9 nc684nl6 p7hjln8o kvgmc6g5 cxmmr5t8 oygrvhab hcukyx3x jb3vyjys rz4wbd8a qt6c0cv9 a8nywdso i1ao9s8h
         esuyzwwr f1sip0of lzcic4wl oo9gr5id gpro0wi8" href="https://www.facebook.com/rosaben78?__cft__[0]=AZXJnlZo9aS4G54VDyRaXhnJc8
         9r7SmDY9TEx4cXbndqzRnVZLqYBpS5WW5DQDm2cOxU_FXexLr_AajcYE5PpWB1AMZs0MHASQopyrj1Mgnv1g&amp;__tn__=%3C%2CP-R" role="link" tabin
         dex="-1"><div class="q676j6op qypqp5cg"><object type="nested/pressable"><a aria-label="Zeineb Be" class="oajrlxb2 gs1a9yip g
         5ia77u1 mtkw9kbi tlpljxtp qensuy8j ppp5ayq2 goun2846 ccm00jje s44p3ltw mk2mc5f4 rt8b4zig n8ej3o3l agehan2d sk4xxmp2 rq0escxv
         nhd2j8a9 q9uorilb mg4g778l btwxx1t3 pfnyh3mw p7hjln8o kvgmc6g5 cxmmr5t8 oygrvhab hcukyx3x tgvbjcpo hpfvmrgz jb3vyjys rz4wbd8
         a qt6c0cv9 a8nywdso l9j0dhe7 i1ao9s8h esuyzwwr f1sip0of du4w351b lzcic4wl abiwlrkh p8dawk7l oo9gr5id" href="https://www.face
         book.com/rosaben78?__cft__[0]=AZXJnlZo9aS4G54VDyRaXhnJc89r7SmDY9TEx4cXbndqzRnVZLqYBpS5WW5DQDm2cOxU_FXexLr_AajcYE5PpWB1AMZs0M
         HASQopyrj1Mgnv1g&amp;__tn__=%3C%3C%2CP-R" role="link" tabindex="0"><div class="q9uorilb l9j0dhe7 pzggbiyp du4w351b"><svg ari
         a-hidden="true" class="pzggbiyp" data-visualcompletion="ignore-dynamic" role="none" style="height: 40px; width: 40px;"><mask
         id="jsc_c_7i"><circle cx="20" cy="20" fill="white" r="20"></circle></mask><g mask="url(#jsc_c_7i)"><image height="100%" pres
         erveaspectratio="xMidYMid slice" style="height: 40px; width: 40px;" width="100%" x="0" xlink:href="https://scontent.ftun10-
         1.fna.fbcdn.net/v/t39.30808-1/cp0/p60x60/223712121_1193601727782233_1054029730362927322_n.jpg?_nc_cat=104&amp;ccb=1-5&amp;_n

In [49]: ListName = [pt.get_text() for pt in Names]

In [*]: import pandas as pd
        ListNames = pd.DataFrame.from_dict(({
            "Data": ListName,


        }),orient='index' )
        ListNames.transpose()

In [338]: ListNames.to_csv(r'C:\Users\Samar\Desktop\4BI4\Scrapping\FB.csv', index=False,header=True, encoding = "UTF-16")
```

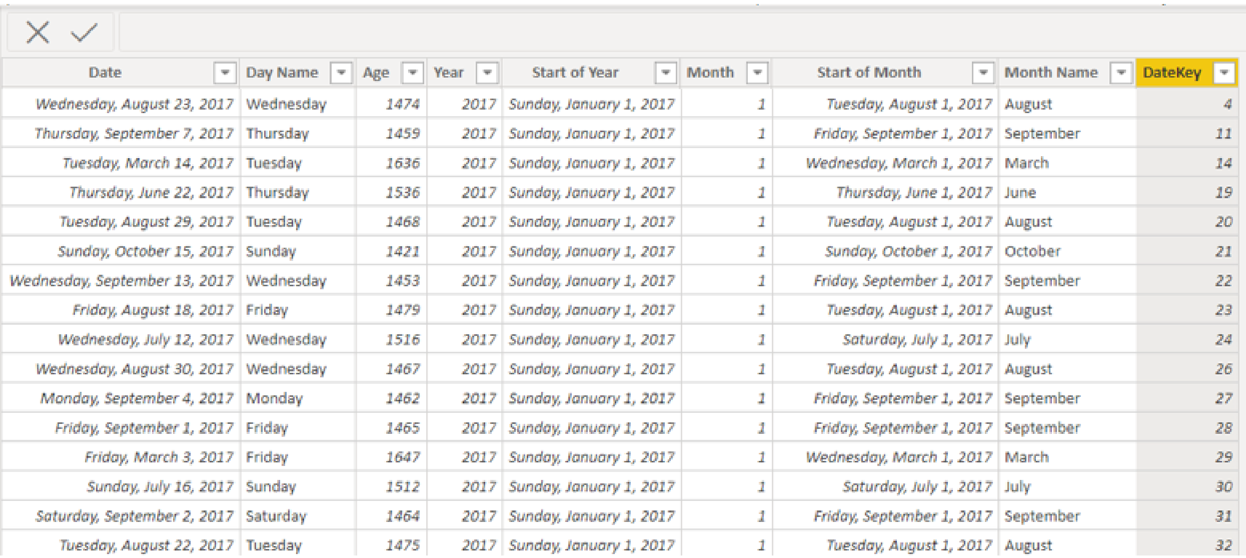
Figure 4: Data picking and storing in csv file

1. Data integration using power BI

Data integration is the process of combining data from different sources into a single, unified view. Integration begins with the ingestion process and includes steps such as cleansing and transformation. Data integration ultimately enables analytics tools to produce effective, actionable business intelligence.

The scraped data of the chapter before lacks structure and meaning , that’s when Power Bi comes in handy. Applying the different necessary tools lead to the creation of the tables figuring in the in the following screenshots.

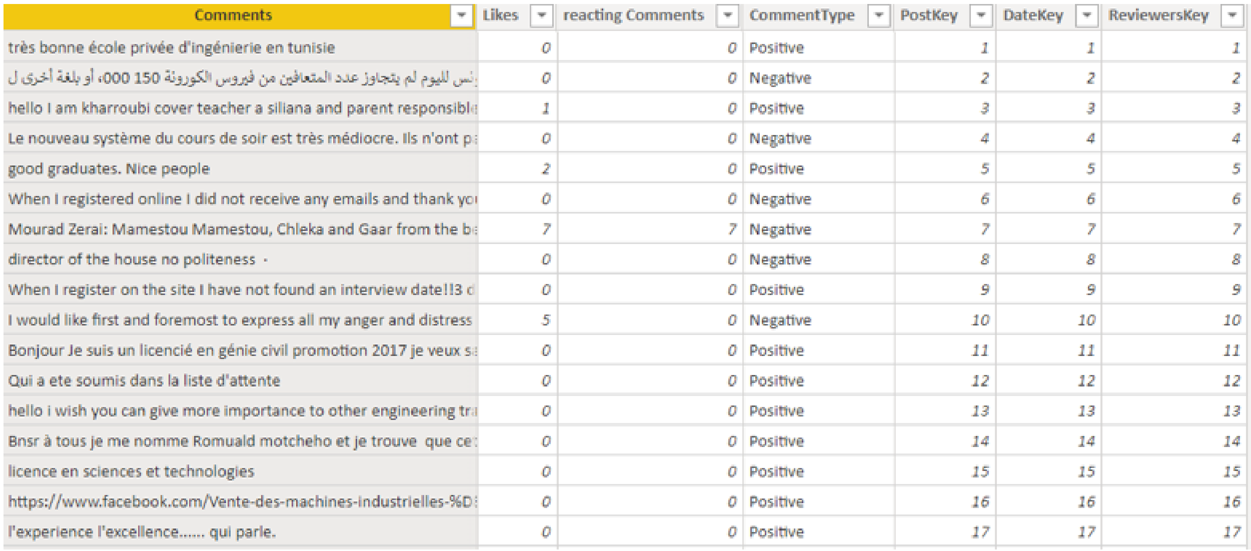
This figure demonstrates a detailed calendar of the posts updates which make it very flexible and useful with the data visualization.



Date	Day Name	Age	Year	Start of Year	Month	Start of Month	Month Name	DateKey
Wednesday, August 23, 2017	Wednesday	1474	2017	Sunday, January 1, 2017	1	Tuesday, August 1, 2017	August	4
Thursday, September 7, 2017	Thursday	1459	2017	Sunday, January 1, 2017	1	Friday, September 1, 2017	September	11
Tuesday, March 14, 2017	Tuesday	1636	2017	Sunday, January 1, 2017	1	Wednesday, March 1, 2017	March	14
Thursday, June 22, 2017	Thursday	1536	2017	Sunday, January 1, 2017	1	Thursday, June 1, 2017	June	19
Tuesday, August 29, 2017	Tuesday	1468	2017	Sunday, January 1, 2017	1	Tuesday, August 1, 2017	August	20
Sunday, October 15, 2017	Sunday	1421	2017	Sunday, January 1, 2017	1	Sunday, October 1, 2017	October	21
Wednesday, September 13, 2017	Wednesday	1453	2017	Sunday, January 1, 2017	1	Friday, September 1, 2017	September	22
Friday, August 18, 2017	Friday	1479	2017	Sunday, January 1, 2017	1	Tuesday, August 1, 2017	August	23
Wednesday, July 12, 2017	Wednesday	1516	2017	Sunday, January 1, 2017	1	Saturday, July 1, 2017	July	24
Wednesday, August 30, 2017	Wednesday	1467	2017	Sunday, January 1, 2017	1	Tuesday, August 1, 2017	August	26
Monday, September 4, 2017	Monday	1462	2017	Sunday, January 1, 2017	1	Friday, September 1, 2017	September	27
Friday, September 1, 2017	Friday	1465	2017	Sunday, January 1, 2017	1	Friday, September 1, 2017	September	28
Friday, March 3, 2017	Friday	1647	2017	Sunday, January 1, 2017	1	Wednesday, March 1, 2017	March	29
Sunday, July 16, 2017	Sunday	1512	2017	Sunday, January 1, 2017	1	Saturday, July 1, 2017	July	30
Saturday, September 2, 2017	Saturday	1464	2017	Sunday, January 1, 2017	1	Friday, September 1, 2017	September	31
Tuesday, August 22, 2017	Tuesday	1475	2017	Sunday, January 1, 2017	1	Tuesday, August 1, 2017	August	32

Figure 6: Calender Table

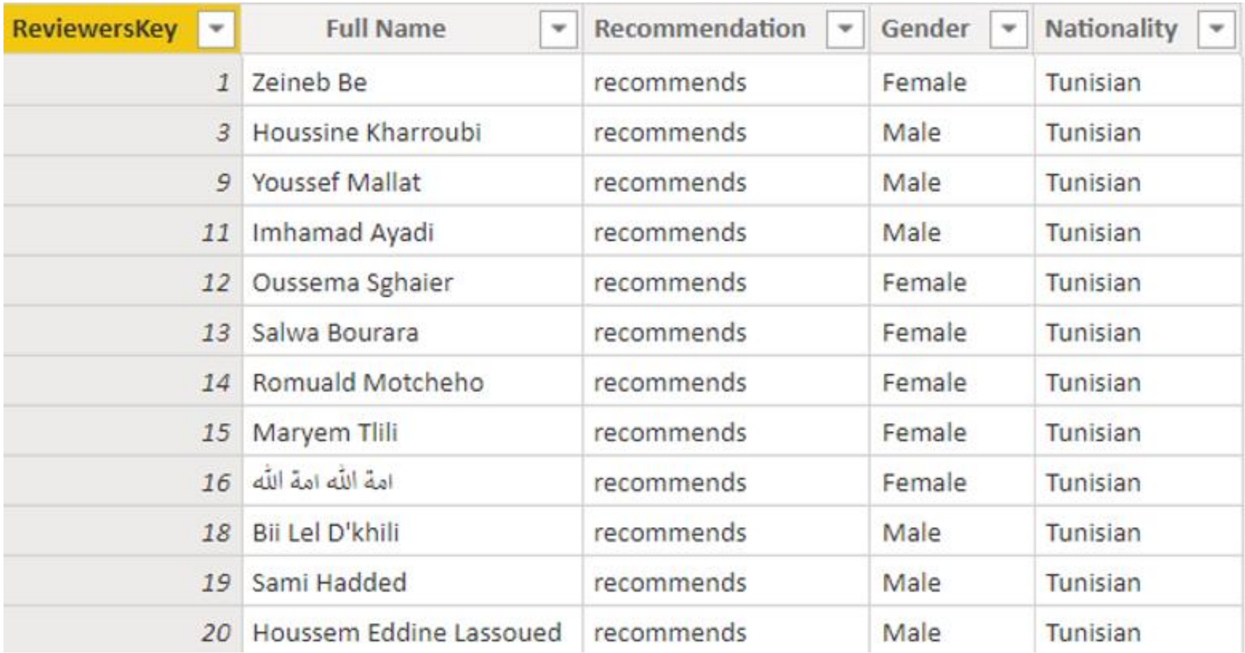
This figure indicates posts information, starting with the comments, number of likes, number of reacting comments, type of comment (whether positive or negative), along with the foreign keys of the lookup tables in which our Data table is associated.



Comments	Likes	reacting Comments	CommentType	PostKey	DateKey	ReviewersKey
très bonne école privée d'ingénierie en tunisie	0	0	Positive	1	1	1
نس لليوم لم يتجاوز عدد المتعافين من فيروس الكورونة 150 000، أو بلغة أخرى ل	0	0	Negative	2	2	2
hello I am kharroubi cover teacher a siliana and parent responsibl	1	0	Positive	3	3	3
Le nouveau système du cours de soir est très médiocre. Ils n'ont p	0	0	Negative	4	4	4
good graduates. Nice people	2	0	Positive	5	5	5
When I registered online I did not receive any emails and thank yo	0	0	Negative	6	6	6
Mourad Zerai: Mamestou Mamestou, Chleka and Gaar from the b	7	7	Negative	7	7	7
director of the house no politeness ·	0	0	Negative	8	8	8
When I register on the site I have not found an interview date!!3 c	0	0	Positive	9	9	9
I would like first and foremost to express all my anger and distress	5	0	Negative	10	10	10
Bonjour Je suis un licencié en génie civil promotion 2017 je veux s	0	0	Positive	11	11	11
Qui a ete soumis dans la liste d'attente	0	0	Positive	12	12	12
hello i wish you can give more importance to other engineering tr	0	0	Positive	13	13	13
Bnsr à tous je me nomme Romuald motcheho et je trouve que ce	0	0	Positive	14	14	14
licence en sciences et technologies	0	0	Positive	15	15	15
https://www.facebook.com/Vente-des-machines-industrielles-%D	0	0	Positive	16	16	16
l'experience l'excellence...... qui parle.	0	0	Positive	17	17	17

Figure 7: Posts_Fact_table

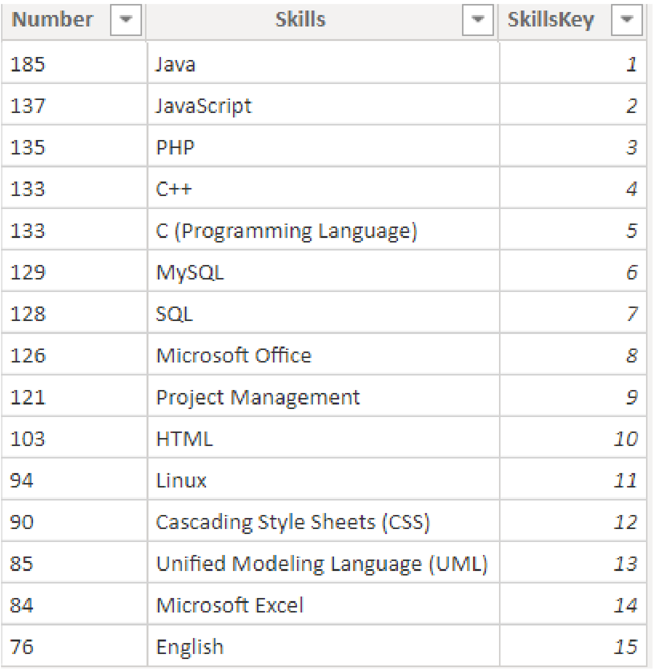
This figure shows the information about the Reviewers members, starting by their full names, recommendation (whether they recommend esprit or not), their gender and their nationality.



ReviewersKey	Full Name	Recommendation	Gender	Nationality
1	Zeineb Be	recommends	Female	Tunisian
3	Houssine Kharroubi	recommends	Male	Tunisian
9	Youssef Mallat	recommends	Male	Tunisian
11	Imhamad Ayadi	recommends	Male	Tunisian
12	Oussema Sghaier	recommends	Female	Tunisian
13	Salwa Bourara	recommends	Female	Tunisian
14	Romuald Motcheho	recommends	Female	Tunisian
15	Maryem Tlili	recommends	Female	Tunisian
16	امة الله امة الله	recommends	Female	Tunisian
18	Bii Lel D'khili	recommends	Male	Tunisian
19	Sami Hadded	recommends	Male	Tunisian
20	Houssem Eddine Lassoued	recommends	Male	Tunisian

Figure 8: Revieweres LookUp

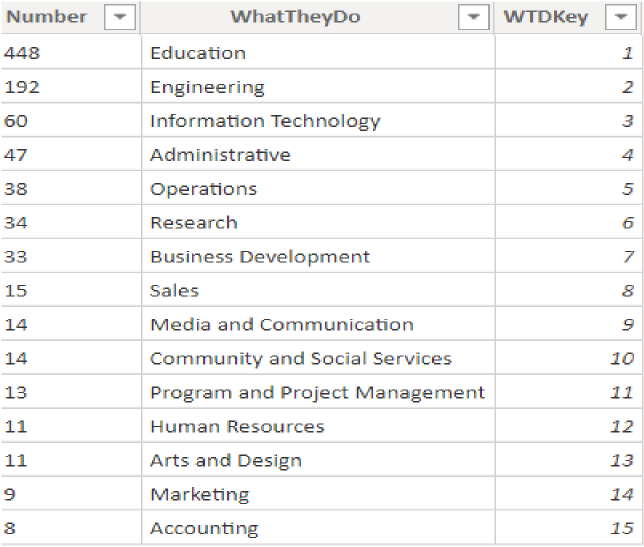
This figure demonstrates the number of esprit employees matched to their skill according to LinkedIn



Number	Skills	SkillsKey
185	Java	1
137	JavaScript	2
135	PHP	3
133	C++	4
133	C (Programming Language)	5
129	MySQL	6
128	SQL	7
126	Microsoft Office	8
121	Project Management	9
103	HTML	10
94	Linux	11
90	Cascading Style Sheets (CSS)	12
85	Unified Modeling Language (UML)	13
84	Microsoft Excel	14
76	English	15

Figure 9: skills LookUp_Table

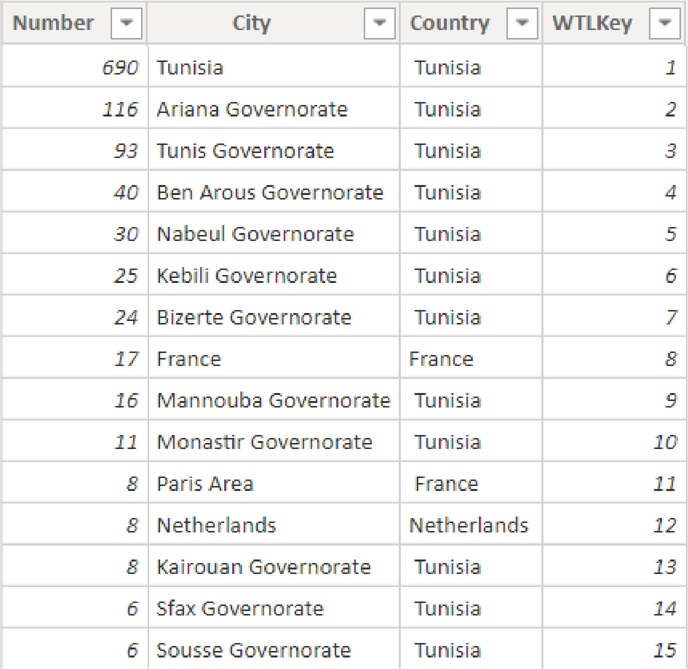
This figure demonstrates the number of Esprit employees matched to their occupation according to LinkedIn.



Number	WhatTheyDo	WTDKey
448	Education	1
192	Engineering	2
60	Information Technology	3
47	Administrative	4
38	Operations	5
34	Research	6
33	Business Development	7
15	Sales	8
14	Media and Communication	9
14	Community and Social Services	10
13	Program and Project Management	11
11	Human Resources	12
11	Arts and Design	13
9	Marketing	14
8	Accounting	15

Figure 10: What They_Do LookUp table

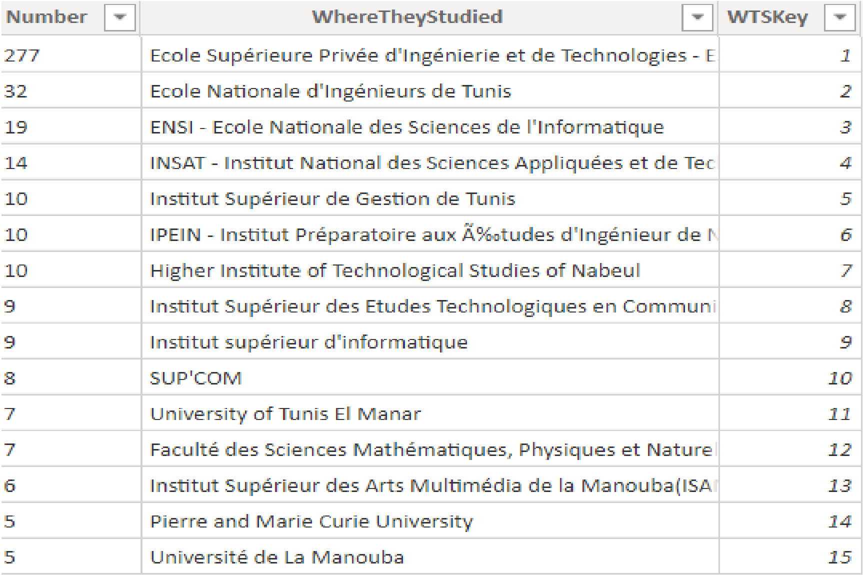
This figure demonstrates the number of Esprit employees matched to where they live according to LinkedIn.



Number	City	Country	WTLKey
690	Tunisia	Tunisia	1
116	Ariana Governorate	Tunisia	2
93	Tunis Governorate	Tunisia	3
40	Ben Arous Governorate	Tunisia	4
30	Nabeul Governorate	Tunisia	5
25	Kebili Governorate	Tunisia	6
24	Bizerte Governorate	Tunisia	7
17	France	France	8
16	Mannouba Governorate	Tunisia	9
11	Monastir Governorate	Tunisia	10
8	Paris Area	France	11
8	Netherlands	Netherlands	12
8	Kairouan Governorate	Tunisia	13
6	Sfax Governorate	Tunisia	14
6	Sousse Governorate	Tunisia	15

Figure 11: Where_They_live lookUp Table

This figure demonstrates the number of Esprit employees matched to where they studied according to LinkedIn.



Number	WhereTheyStudied	WTSKey
277	Ecole Supérieure Privée d'Ingénierie et de Technologies - E	1
32	Ecole Nationale d'Ingénieurs de Tunis	2
19	ENSI - Ecole Nationale des Sciences de l'Informatique	3
14	INSAT - Institut National des Sciences Appliquées et de Tec	4
10	Institut Supérieur de Gestion de Tunis	5
10	IPEIN - Institut Préparatoire aux Ã‰tudes d'Ingénieur de N	6
10	Higher Institute of Technological Studies of Nabeul	7
9	Institut Supérieur des Etudes Technologiques en Communi	8
9	Institut supérieur d'informatique	9
8	SUP'COM	10
7	University of Tunis El Manar	11
7	Faculté des Sciences Mathématiques, Physiques et Nature	12
6	Institut Supérieur des Arts Multimédia de la Manouba(ISA	13
5	Pierre and Marie Curie University	14
5	Université de La Manouba	15

Figure 12: Where_They_Studied LookUp Table

1.2 Data warehouse model:

On this part, I had to figure out how link the tables together in way it can serve my purpose.

Since Data was gathered from two completely different sources (LinkedIn and Facebook) and represents two different ideas. I had to create two different Data warehouses following the star model.

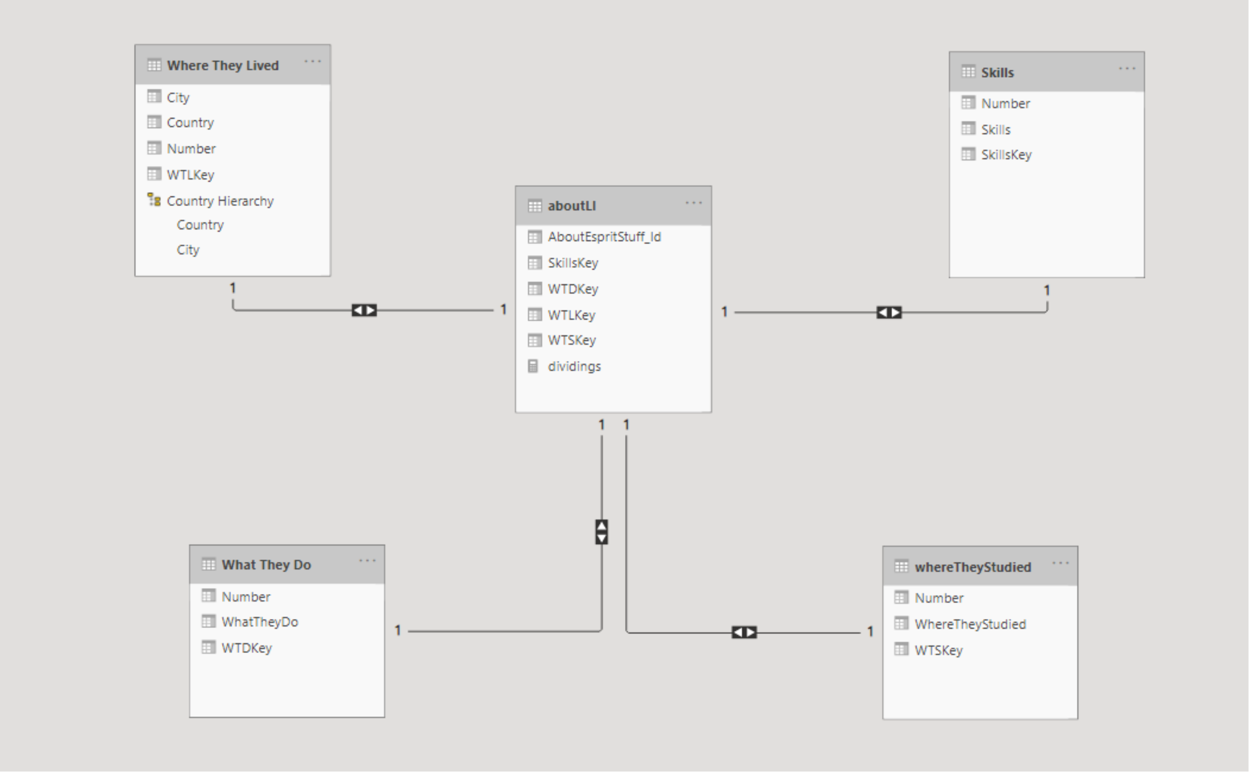


Figure 13: LinkedIn Warehouse

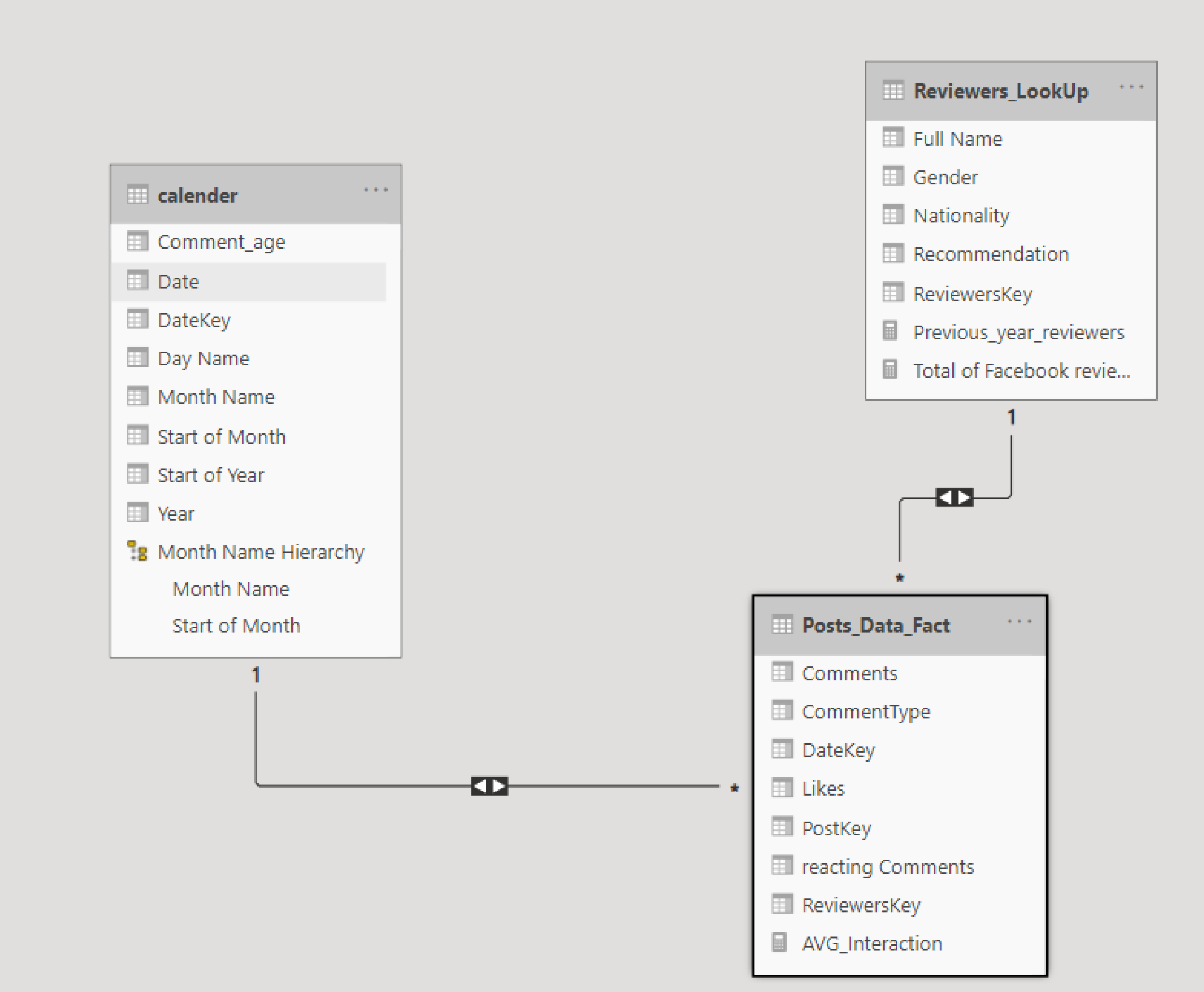


Figure 14: Facebook Warehouse

III.Text Mining

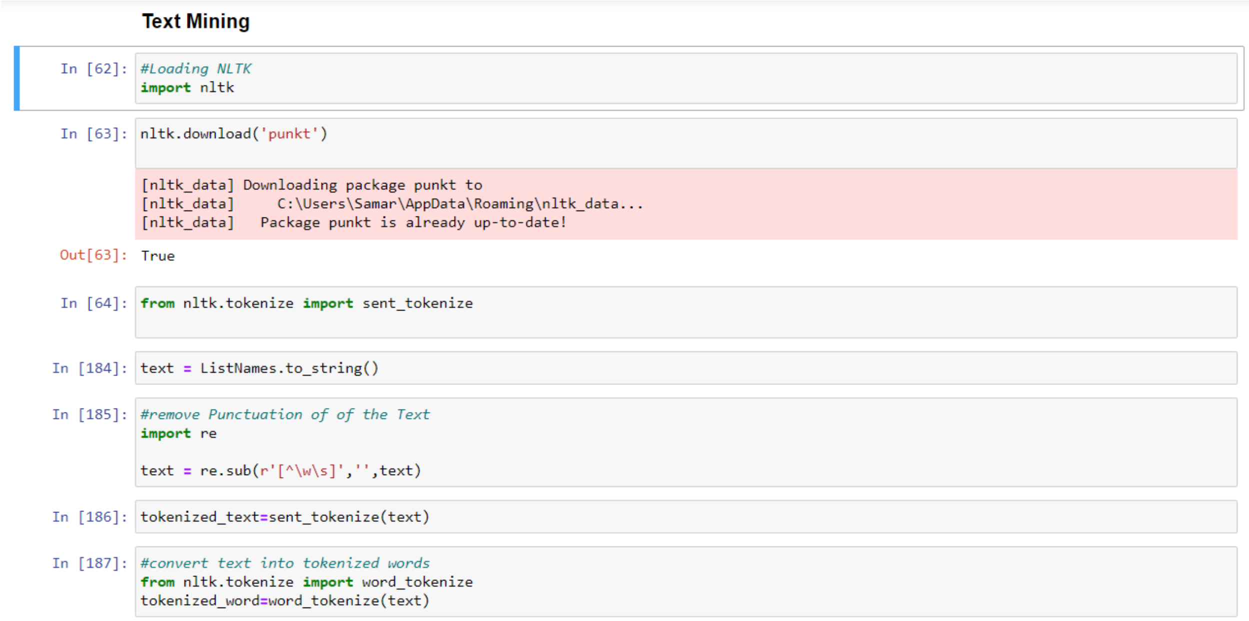
Text mining, also referred to as text data mining, like text analysis, is the process of deriving high-quality information from text.

Within the following, I’ll be demonstrating screenshots about my work of text Mining of esprit reviewers Facebook page.

At this first stage, I started by fetching the Data by scrapping them out of the reviewers’ page on Facebook.

The data fetched might contain irrelevant data, and this may parasitize with the data we want to work on. That’s why I cleaned the Data out of stop words and punctuations and applied the different methods such as stemming and Lemmatization.

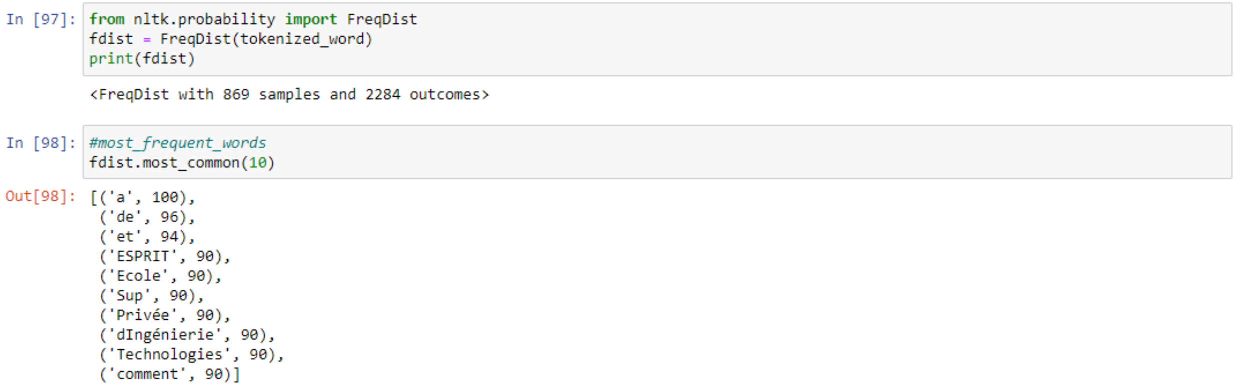
Also, I studied words frequency and highlighted them to have a clear vision about the most significant terms related to Esprit.



Text Mining

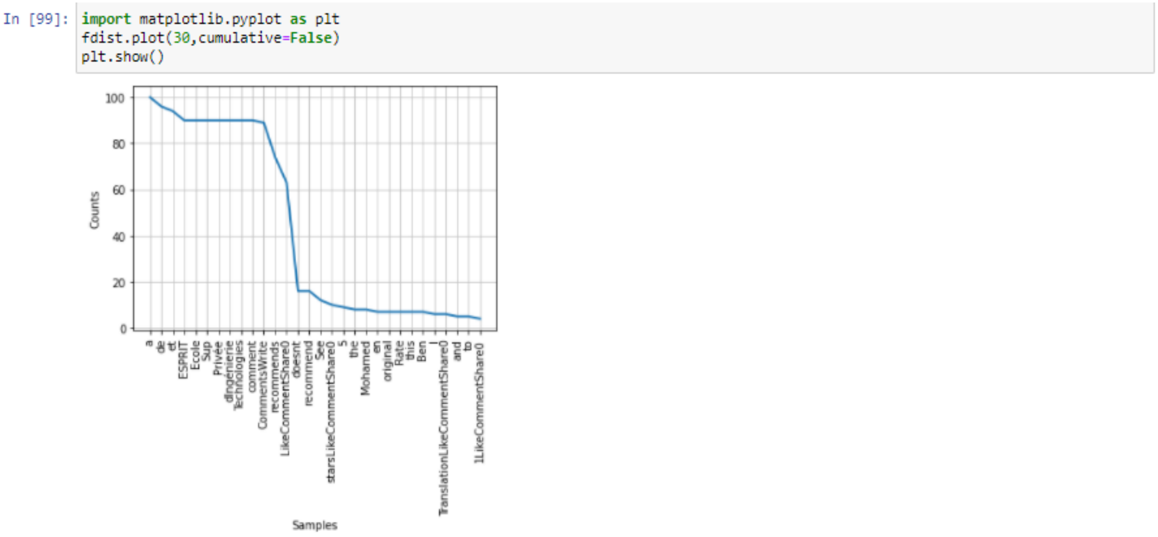
```
In [62]:  #Loading NLTK
          import nltk

In [63]:  nltk.download('punkt')

          [nltk_data] Downloading package punkt to
          [nltk_data]     C:\Users\Samar\AppData\Roaming\nltk_data...
          [nltk_data]   Package punkt is already up-to-date!

Out[63]:  True

In [64]:  from nltk.tokenize import sent_tokenize

In [184]: text = ListNames.to_string()

In [185]: #remove Punctuation of of the Text
          import re

          text = re.sub(r'[^\w\s]','',text)

In [186]: tokenized_text=sent_tokenize(text)

In [187]: #convert text into tokenized words
          from nltk.tokenize import word_tokenize
          tokenized_word=word_tokenize(text)
```

Figure 15: Cleaning data



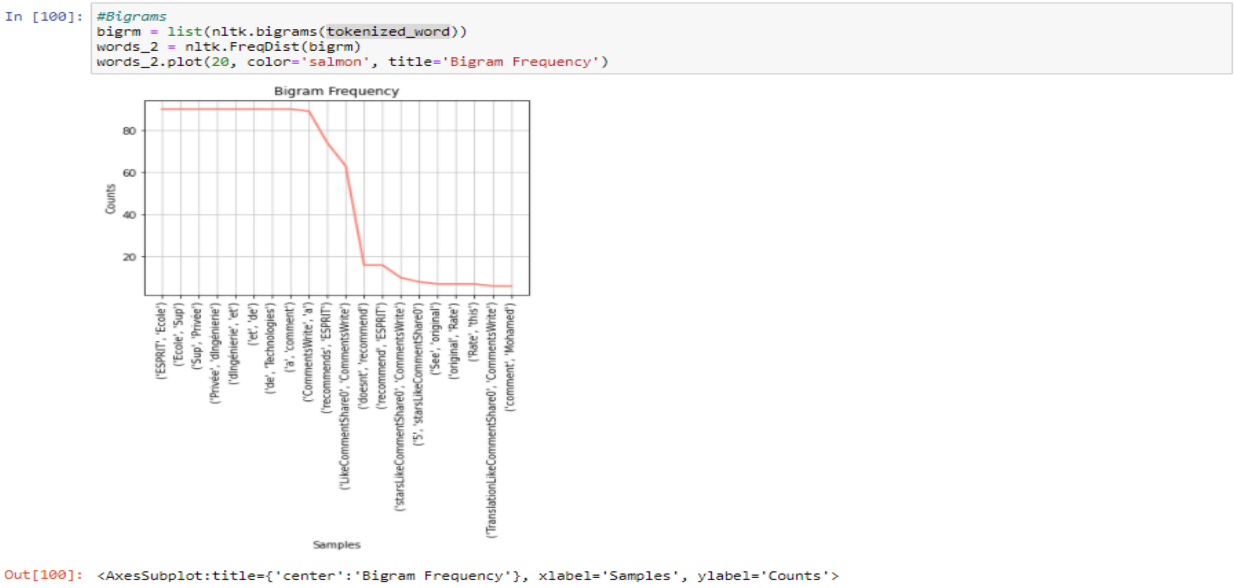
```
In [97]: from nltk.probability import FreqDist
         fdist = FreqDist(tokenized_word)
         print(fdist)

         <FreqDist with 869 samples and 2284 outcomes>

In [98]: #most_frequent_words
         fdist.most_common(10)

Out[98]: [('a', 100),
          ('de', 96),
          ('et', 94),
          ('ESPRIT', 90),
          ('Ecole', 90),
          ('Sup', 90),
          ('Privée', 90),
          ('dIngénierie', 90),
          ('Technologies', 90),
          ('comment', 90)]
```

Figure 16: Studying of frequency



```
In [99]: import matplotlib.pyplot as plt
         fdist.plot(30,cumulative=False)
         plt.show()
```

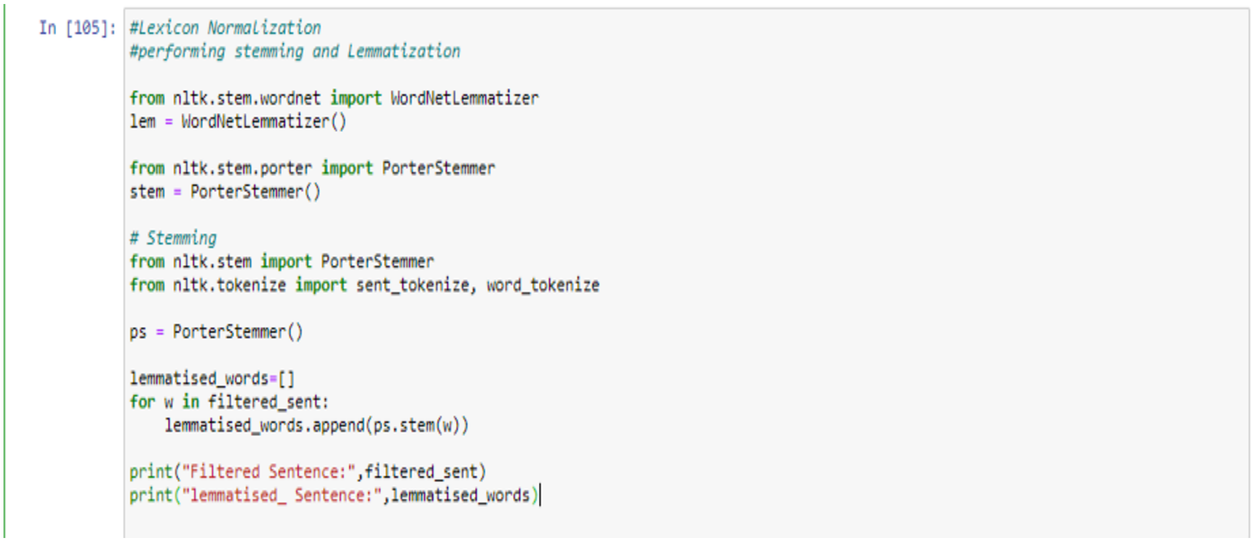
Figure 17: frequency plot of single terms



```
In [100]: #Bigrams
          bigrm = list(nltk.bigrams(tokenized_word))
          words_2 = nltk.FreqDist(bigrm)
          words_2.plot(20, color='salmon', title='Bigram Frequency')
```

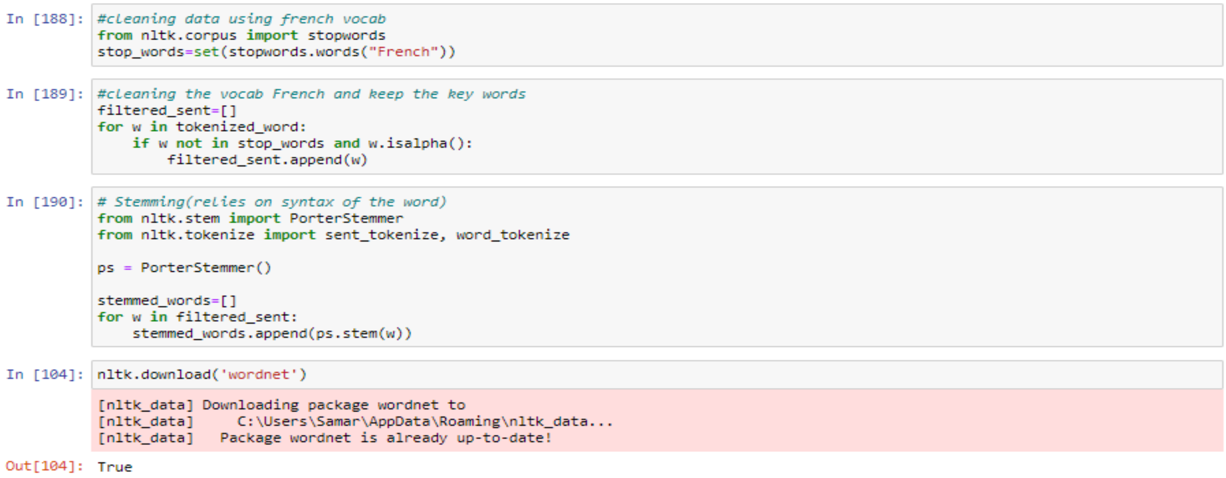
```
Out[100]: <AxesSubplot:title={'center':'Bigram Frequency'}, xlabel='Samples', ylabel='Counts'>
```

Figure 18: Frequency Plot of coupled term



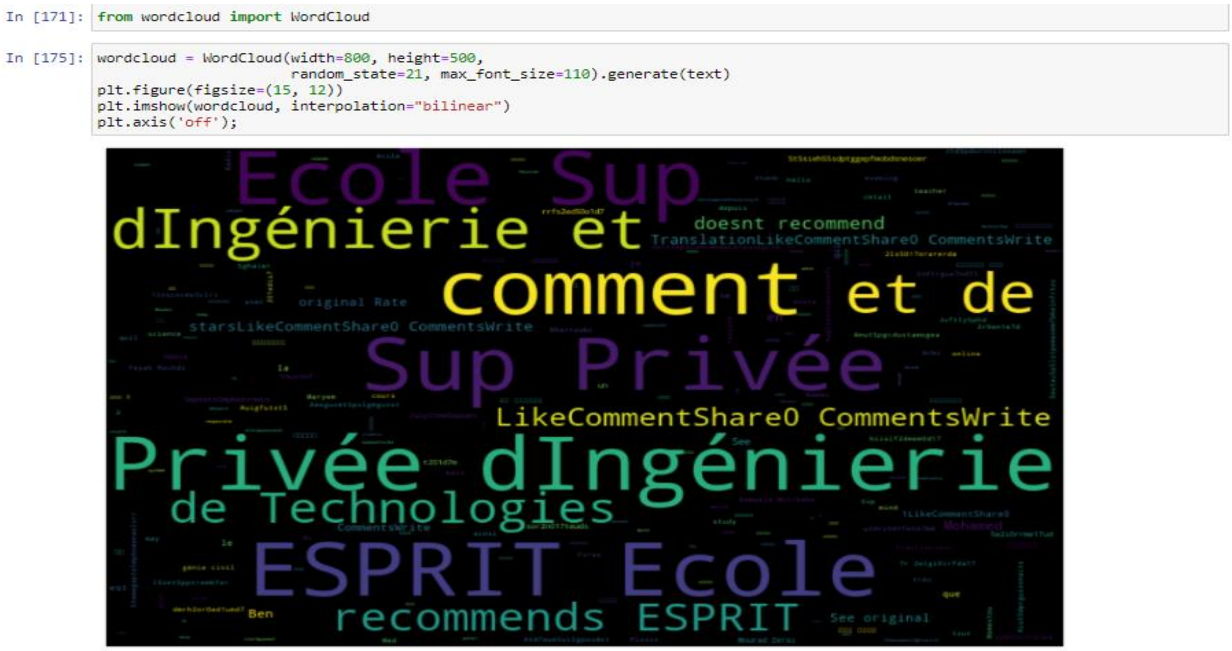
```
In [105]: #Lexicon Normalization
          #performing stemming and Lemmatization

          from nltk.stem.wordnet import WordNetLemmatizer
          lem = WordNetLemmatizer()

          from nltk.stem.porter import PorterStemmer
          stem = PorterStemmer()

          # Stemming
          from nltk.stem import PorterStemmer
          from nltk.tokenize import sent_tokenize, word_tokenize

          ps = PorterStemmer()

          lemmatised_words=[]
          for w in filtered_sent:
              lemmatised_words.append(ps.stem(w))

          print("Filtered Sentence:",filtered_sent)
          print("lemmatised_ Sentence:",lemmatised_words)
```

Figure19: Content filtering



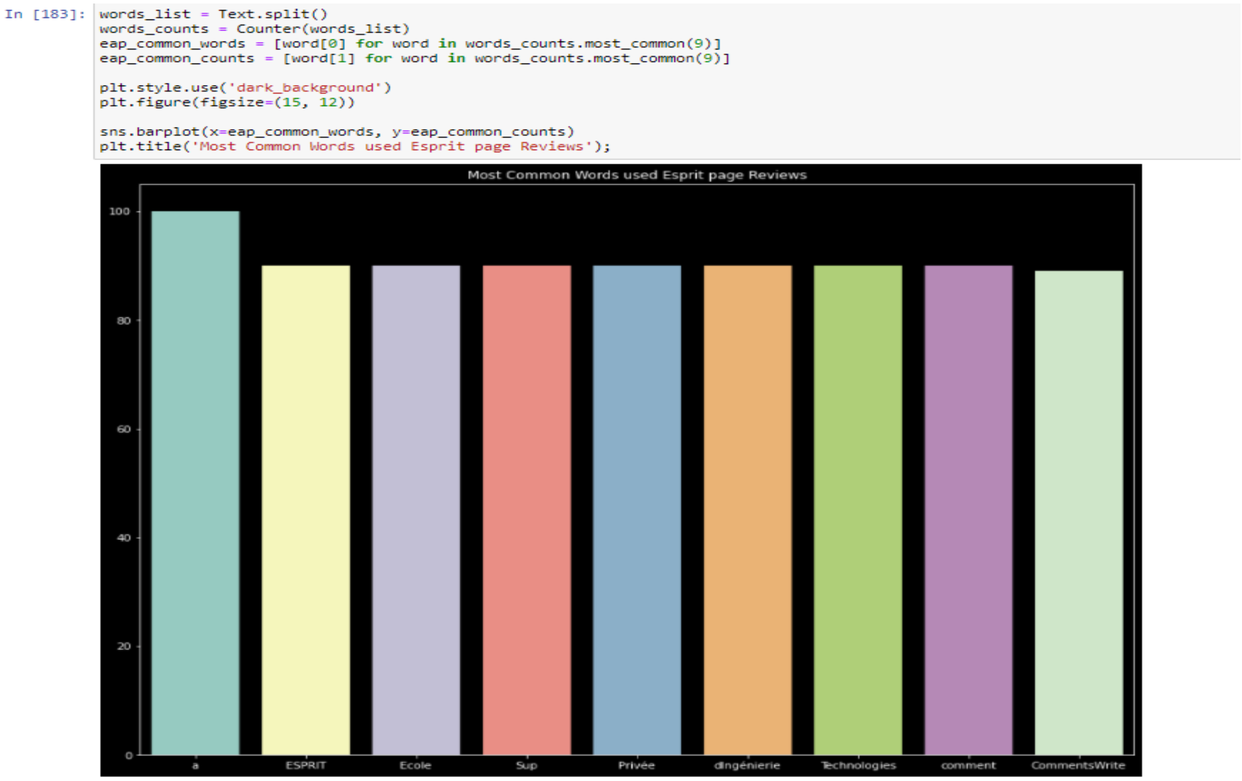
```
In [188]: #cleaning data using french vocab
          from nltk.corpus import stopwords
          stop_words=set(stopwords.words("French"))

In [189]: #cleaning the vocab French and keep the key words
          filtered_sent=[]
          for w in tokenized_word:
              if w not in stop_words and w.isalpha():
                  filtered_sent.append(w)

In [190]: # Stemming(relies on syntax of the word)
          from nltk.stem import PorterStemmer
          from nltk.tokenize import sent_tokenize, word_tokenize

          ps = PorterStemmer()

          stemmed_words=[]
          for w in filtered_sent:
              stemmed_words.append(ps.stem(w))

In [104]: nltk.download('wordnet')

          [nltk_data] Downloading package wordnet to
          [nltk_data]     C:\Users\Samar\AppData\Roaming\nltk_data...
          [nltk_data]   Package wordnet is already up-to-date!

Out[104]: True
```

Figure 20: Word Stemming



```
In [171]: from wordcloud import WordCloud

In [175]: wordcloud = WordCloud(width=800, height=500,
                                random_state=21, max_font_size=110).generate(text)
          plt.figure(figsize=(15, 12))
          plt.imshow(wordcloud, interpolation="bilinear")
          plt.axis('off');
```

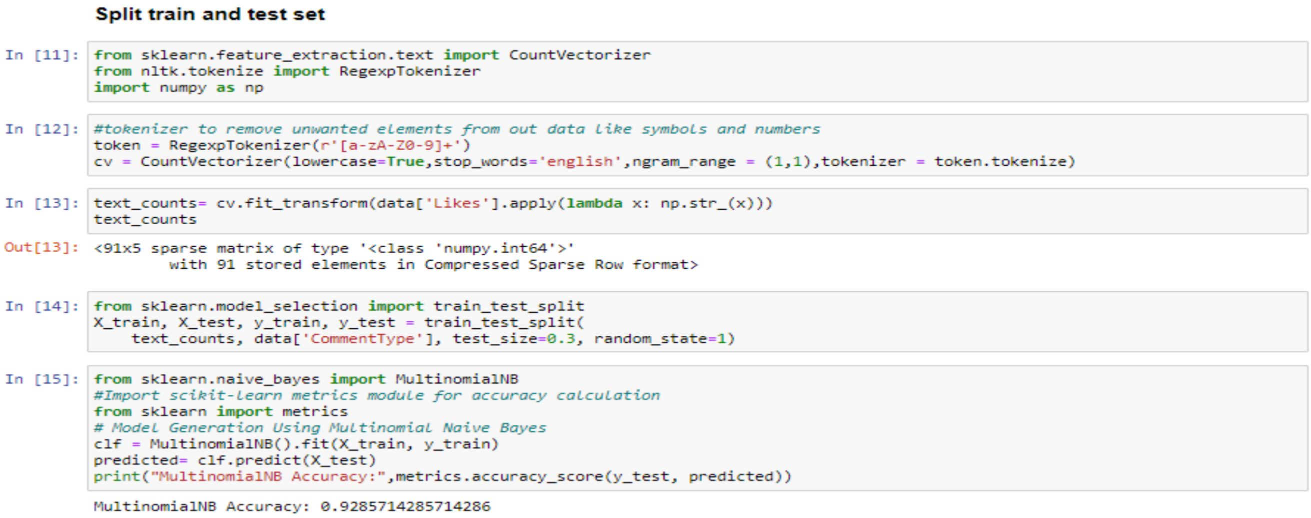
Figure 19: Words_Clouds



```
In [183]: words_list = Text.split()
          words_counts = Counter(words_list)
          eap_common_words = [word[0] for word in words_counts.most_common(9)]
          eap_common_counts = [word[1] for word in words_counts.most_common(9)]

          plt.style.use('dark_background')
          plt.figure(figsize=(15, 12))

          sns.barplot(x=eap_common_words, y=eap_common_counts);
          plt.title('Most Common Words used Esprit page Reviews');
```

Figure 21: Bar_Plot

As a final step of this chapter, I applied a supervised learning approach of classification called Naive Bayes to figure out the accuracy of my data since it is popular and used for sentiment classification.



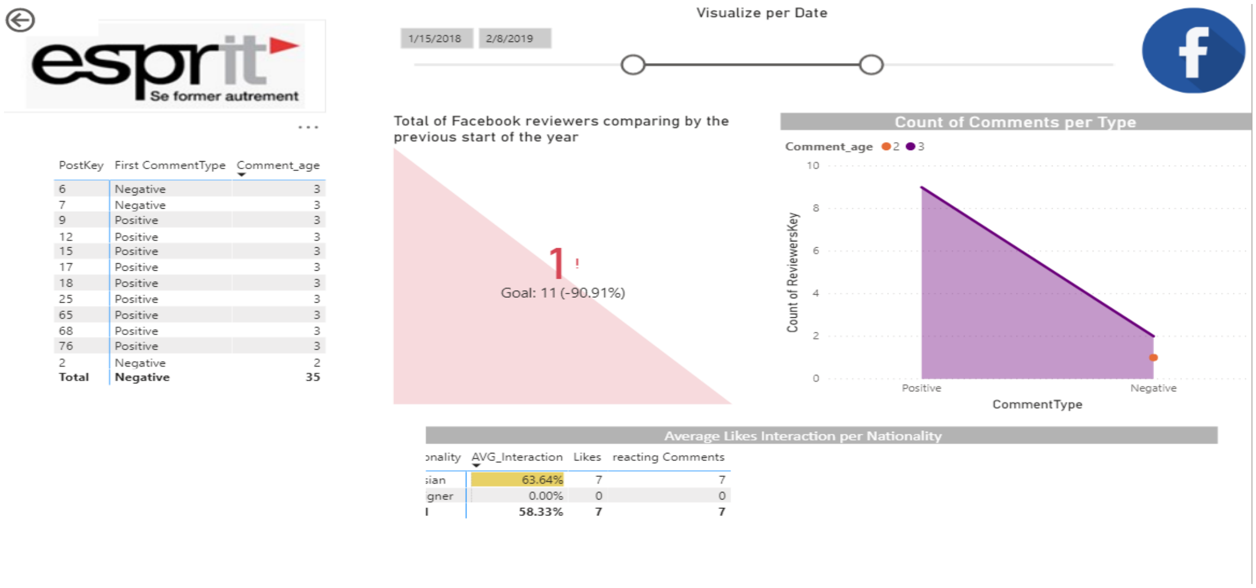
Split train and test set

```
In [11]: from sklearn.feature_extraction.text import CountVectorizer
         from nltk.tokenize import RegexpTokenizer
         import numpy as np

In [12]: #tokenizer to remove unwanted elements from out data like symbols and numbers
         token = RegexpTokenizer(r'[a-zA-Z0-9]+')
         cv = CountVectorizer(lowercase=True,stop_words='english',ngram_range = (1,1),tokenizer = token.tokenize)

In [13]: text_counts= cv.fit_transform(data['Likes'].apply(lambda x: np.str_(x)))
         text_counts

Out[13]: <91x5 sparse matrix of type '<class 'numpy.int64'>'
                 with 91 stored elements in Compressed Sparse Row format>

In [14]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(
             text_counts, data['CommentType'], test_size=0.3, random_state=1)

In [15]: from sklearn.naive_bayes import MultinomialNB
         #Import scikit-learn metrics module for accuracy calculation
         from sklearn import metrics
         # Model Generation Using Multinomial Naive Bayes
         clf = MultinomialNB().fit(X_train, y_train)
         predicted= clf.predict(X_test)
         print("MultinomialNB Accuracy:",metrics.accuracy_score(y_test, predicted))

         MultinomialNB Accuracy: 0.9285714285714286
```

Figure 23: Data_training

IV.Data Visualization with Power BI:

On this part, I will be demonstrating my final dashboard including the different reports for a neat and clear readability of my data and the relation between its different elements.

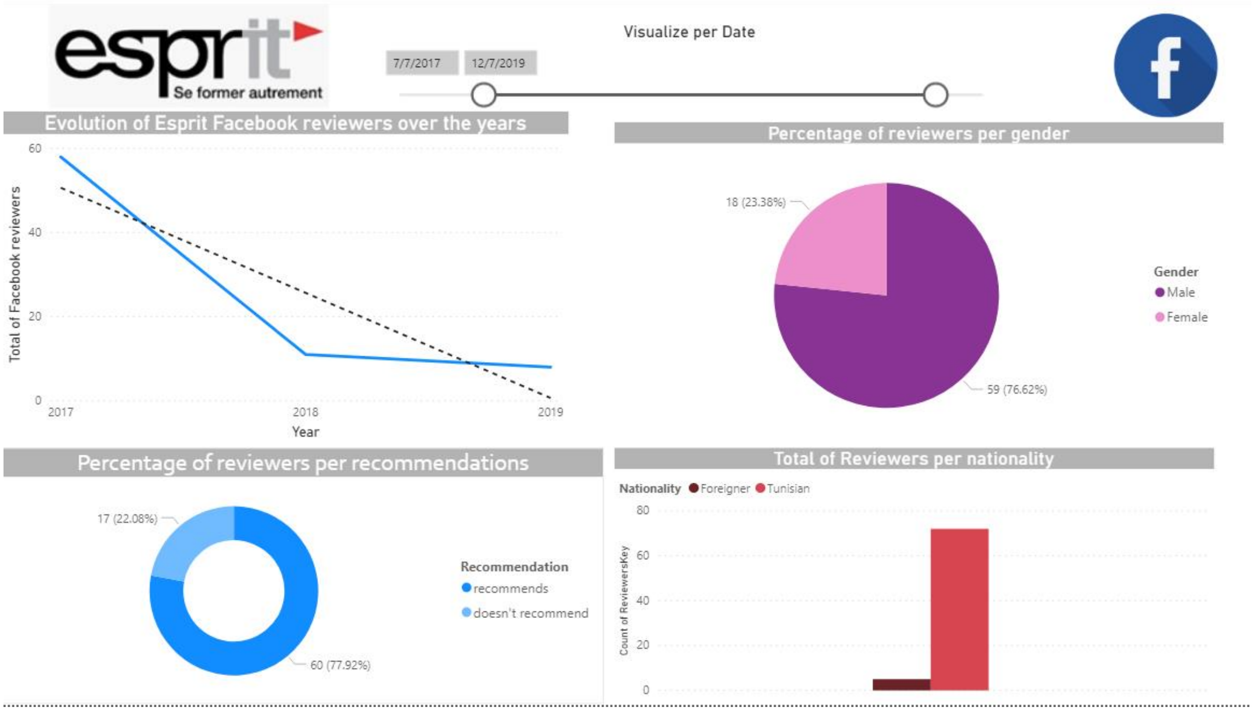


PostKey	First CommentType	Comment_age
6	Negative	3
7	Negative	3
9	Positive	3
12	Positive	3
15	Positive	3
17	Positive	3
18	Positive	3
25	Positive	3
65	Positive	3
68	Positive	3
76	Positive	3
2	Negative	2
Total	**Negative**	**35**

Average Likes Interaction per Nationality

onality	AVG_Interaction	Likes	reacting Comments
sian	63.64%	7	7
gner	0.00%	0	0
l	**58.33%**	**7**	**7**

Figure24: Facebook_dashboard_1



Visualize per Date
7/7/2017
12/7/2019
Evolution of Esprit Facebook reviewers over the years
60
40
20
0
Total of Facebook reviewers
2017
2018
2019
Year
Percentage of reviewers per gender
18 (23.38%)
59 (76.62%)
Gender
● Male
● Female
Percentage of reviewers per recommendations
17 (22.08%)
60 (77.92%)
Recommendation
● recommends
● doesn't recommend
Total of Reviewers per nationality
Nationality ● Foreigner ● Tunisian
80
60
40
20
0
Count of ReviewersKey

Figure25: facebook_Dashboard_2

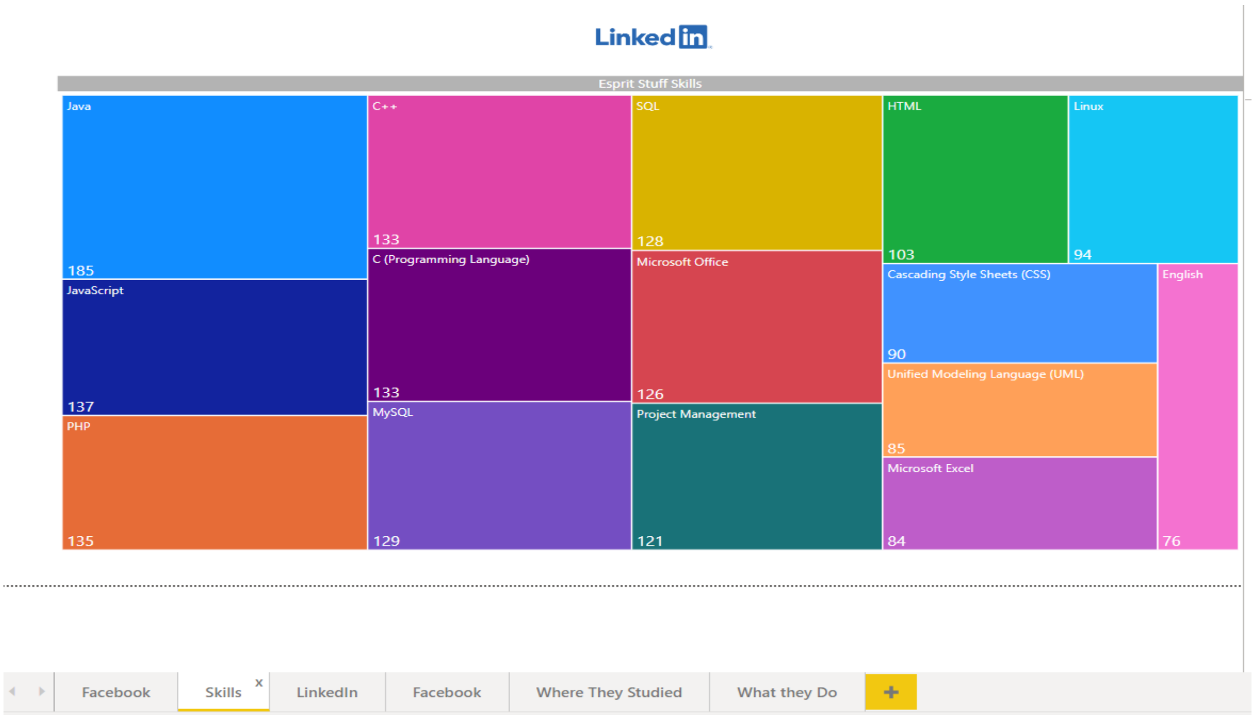


Figure 26:Skills

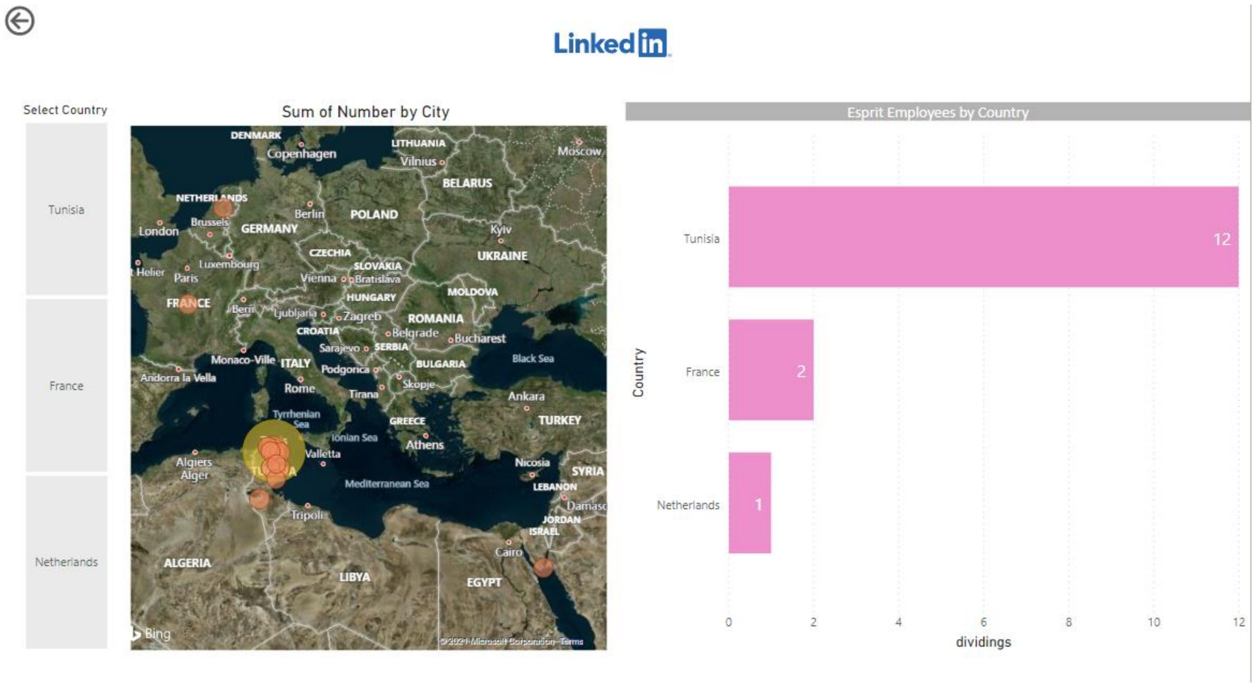


Figure 27:Where_They_Live

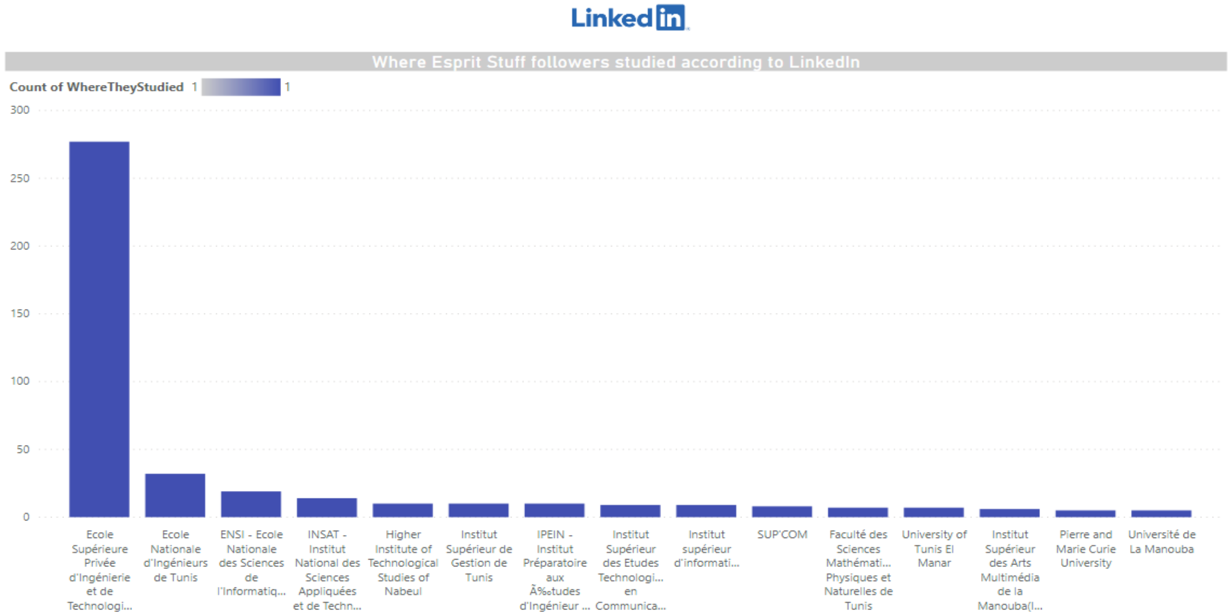


Figure28:Where_They_Studied

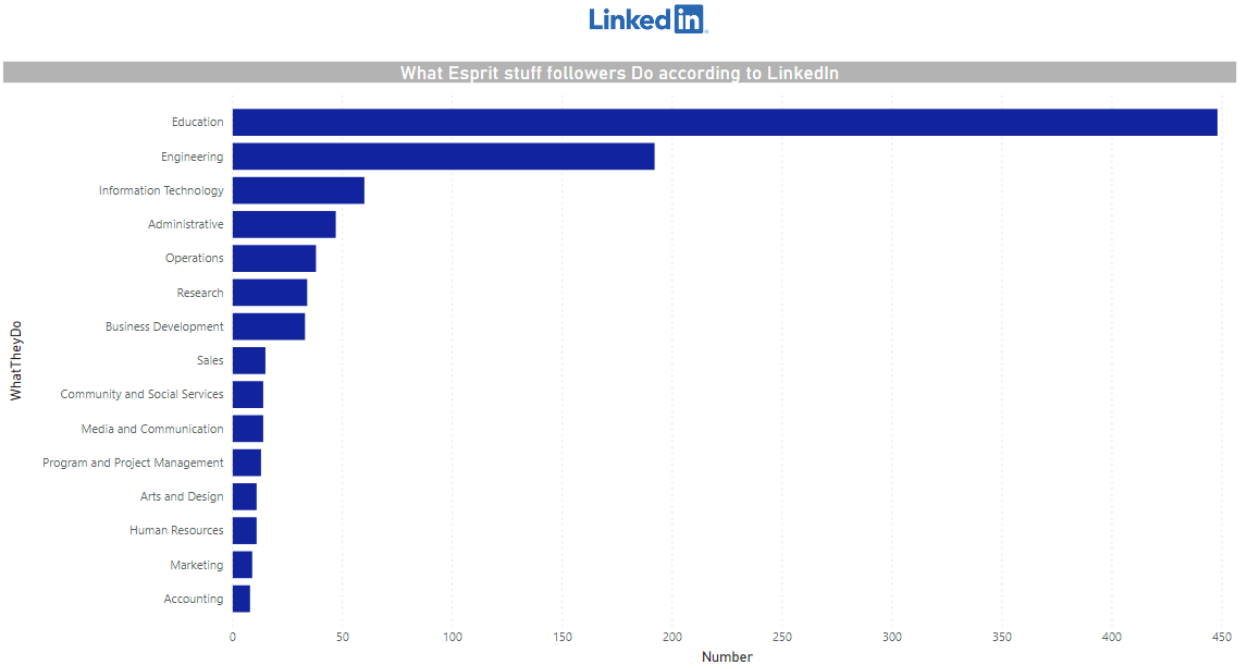


Figure 29:What_They_Do

V.conclusion :

Since data is massively growing over the years, learning the adequate tools, applying good analysis and visualizing data in a the best readable way is a shortcut towards successful entrepreneurial planning.

During this project, I covered the data scraping, text Mining going to data visualization. Power Bi and Python are two powerful tools that led into creating a dashboard so user friendly, for the optimum results and the best decision making that shall be affected by Esprit.

Using this project, Esprit can have a better idea about its public image, note its weaknesses and apply the necessary changes in order to remain the best destination for lots of students in Tunisia and otherwise.