King Saud University

College of Computer and Information SciencesDepartment

of Software Engineering

SWE 486 – Cloud Computing and Big Data

Sentiment Analysis

PHASE 1

#	NAME	ID
1	Lina Alkhodhiri	441201109
2	Noura Alsaif	439200832
3	Alanoud Alomar	441201113

GROUP #	4
SUPERVISOR	L. Hailah Almazrua

Submission Date: *August 14, 2022*

Table of contents

1. Project Description....1
   1.1 Introduction....1
   1.2 Goal....1
   1.3 Initial Hypothesis....1
   1.4 Objective....2
   1.5 Analysis Plan....2
2. Data Exploration....3
3. Data Issues....5
   Issue#1: Punctuation....5
   Issue#2: Text Normalization....6
   Issue#3: English Characters....7
4. Tool Used....8
5. Reference....9

Table of figures

Figure 1....3
Figure 2....3
Figure 3....4
Figure 4....4
Figure 5....4
Figure 6....5
Figure 7....5
Figure 8....5
Figure 9....6
Figure 10....6
Figure 11....7
Figure 12....7

1. PROJECT DESCRIPTION

1.1 INTRODUCTION

In our world today, technology has become everywhere around us. It has more progress. The data grows quickly and has a large size, so storing, managing, and processing this data is very complex. As a result, we have the so-called 'huge data and cloud computing', indicating that huge data is a large amount of data of different types. Therefore, in this course, we learned about the concept of data analysis, which is the process of examining, purifying, converting, and modeling data to reveal useful information on a specific topic [1].

To obtain more experience in data analysis curricula, in our project, we aim to acquire knowledge about people's impressions and their satisfaction with the films they see by extracting and analyzing the available data on *Elcinema.com* on our topic and producing goals that we look forward to achieving.

1.2 GOAL

Nowadays, the industry market and the production of movies are developing very quickly. Films have become important in influencing society issues. *Elcinema.com*, a site that provides the largest Arabic cinematic content in which you can follow the latest news of the art world, and you can also book tickets online and watch movies and exclusive meetings with the stars of the Arab world where you can provide your evaluation and your observations for the movies that you have seen and read movie reviews for other site pioneers [2]. So, our main goal is to use our data analysis skills to explore and study the opinions of the site's pioneers to achieve the project goals.

1.3 INITIAL HYPOTHESIS

By the end of the project, we will either approve or disapprove the following hypothesis:

- People tend to watch movies of comedy nature.

If the hypothesis is approved, *Elcinema* will give an insight into the best comedy movies. On the contrary, if we cannot agree to the hypothesis, this may indicate that we must know what the site's pioneers want more.

- People tend to hate very long movies.

If the hypothesis is approved, the *Elcinema* will help to focus more on the number of movies and display movies that attract viewers. On the contrary, if we cannot agree to the hypothesis, this means that *Elcinema* should display films of all kinds and may not pay attention to the duration of the movie.

1.4 OBJECTIVE

Our objective is to understand how to implement data analysis techniques to produce a result that helps us know the behavior and interests of the viewers, and what they want to see more and what they like and what they do not like to give *Elcinema.com* some recommendations based on it, and at the end of this project, we hope that we will learn and apply data analysis techniques, starting From collecting data using Python libraries, to clean it, process it and even finally provide these data.

1.5 ANALYSIS PLAN

In our project we will implement the Data Analytics Lifecycle which defines the analytic process and best practices from discovery to project completion [3]. We will use "Google Collab" to write our code to clean the Dataset in addition to building the model and training it [9]. Data Analytics Lifecycle consists of 6 phases: Discovery, Data Preparation, Model Planning, Model Building, Communicate Results and Operationalize.

Step 1: Discovery

In this step, we got ready data on people's reviews from Elcinema.com.

Step 2: Data Preparation

In this step, we will clean the data by using preprocessing methods such as, ETL (Extraction, Transformation, Loading) and using Python libraries:

- Pandas: is a fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language [4].

- Re: A regular expression is a special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern [5].

- String: Python String module contains some constants, utility function, and classes for string manipulation [6].

- Nltk: (Natural Language Toolkit) is an open-source Python library for Natural Language Processing [7].
    - Nltk tokenizer: Tokenizers divide strings into lists of substrings. we used tokenizers to splits a string using a regular expression, which matches either the tokens or the separators between tokens [8].

Step 3: Model Planning

In this step, we will define the methods and techniques that are based on our hypotheses and make sure that they meet the goals of our business by understanding data and relationships between variables.

Step 4: Model Building

In this step, we will implement the models specified in step 3, develop the dataset, and divide it into two parts:

1.Training Data: develop model in the Dataset
2.Test data: After completing the training, we will test the testing dataset to ensure that the model works well and as we want.

Step 5: Communicate Results

In this step, and after validating our building model we will communicate and document the key findings and determine if we succeeded or failed in our objectives.

2. Data Exploration

Data exploration is an important step to make pre -processing of data. Because it tells us a lot of information about the data we want to process and make us understand the nature of this data. Such as the data size, the types of information it contains and other important features that we can use to form appropriate cleaning processes.

• View the total number of data (rows) using len().

```
[21]
     len(data_df)

     1524
```

Figure 1. Exploration action (number of data)

• View the names of the columns included in the data to clarify the types of information included in the data using the columns.

```
[22] data_df.columns

     Index(['Text', 'classification'], dtype='object')
```

Figure 2. Exploration action (Columns names)

• View some statistical data using describe().

```
[41] data_df.describe()
```

	Text	classification
count	1524	1519
unique	1522	3
top	الطفلة جني أنقذت "سامي أكسيد الكربون" من الملل يحظي هذا الفيلم بميزة أنه الفيلم الوحيد في السوق حالياً الذي تستطيع أن تميزه بأنه "عائلي" وتستطيع أن تصطحب إليه عائلتك وأطفالك لو عندك أطفال ويستمتعون به بعيداً عن فيلمي "سرخة نملة" و"الفاجومي" اللذان يحملان جرعة سياسية كبيرة، أما "سامي أكسيد الكربون" فهو خفيف ويصلح ليكون الفيلم الذي تذهب إليه عائلة يوم الجمعة ليشاهدوه أثناء تناول الفشار والتخبيط برجليهم علي كراسي الناس اللي قدامهم، الفيلم يبدأ بداية تشعرك بأنه سيكون نسخة من آخر أفلام هاني رمزي "الرجل الغامض بسلامته" بعرض هوس بطل الفيلم "سامي" الذي يجسد شخصيته هاني رمزي بالسنات، في "الرجل الغامض بسلامته" كان البطل مهووس بالسنات فيما يتم تجاهله تماماً، أما "سامي" فهو كابتن طيار مهووس بالسنات والسنات مهووسات به، وتم عرض ذلك بشكل مبالغ فيه، ضمن مجموعة من المبالغات التي تم عرضها والخيوط الدرامية الكثيرة الضعيفة ولكنها "تعدي" إذا قررنا ألا نحمل الفيلم أكبر من حجمه وإذا عرفنا مقدماً أنه فيلم تصطحب إليه الأطفال عشان ياكلوا فشار ويخبطوا علي كراسي اللي قدامهم –وبي مش حاجة وحشة بالمناسبة–، درة بطلة الفيلم جاء أدائها مصطنعاً ولم تناسبها الشخصية علي الإطلاق خاصة وأن الشخصية يبدو أنها ضعيفة علي الورق في الأساس وبها الكثير من التفاصيل المقحمة عليها فجأت في النهاية مهلهلة التفاصيل ومبالغ في ملامحها، فهي شخصية ثورية إذ لم يخلو الفيلم من الثورة أيضاً ففي بداية الفيلم يظهر إدوارد، الذي يجسد شخصية صديق البطل ومساعد الطيار سامي وهو يقول "الكابتن يريد هدوء في الطيارة" علي طريقة "الشعب يريد إسقاط النظام"، وتم إستغلال بعض ملامح الثورة الأخري في الفيلم مثل مقولة "أنتوا كدة خلاص أسقطوا النظام" عندما تعرض سامي للضرب بعد معاكسته لفتاة، يبدو أن الثورة ستلازم جميع الأفلام في الفترة القادمة حتي الكوميدية، يسير الفيلم مملاً هي فعلاً طفلة تلقائية وخفيفة الدم وعشان نكون واضحين، "E.T" ويشكل يكاد يكون مشابه تماماً لفيلم هاني رمزي السابق حتي تظهر فيه الطفلة "جنا" حتي يتغير الفيلم تماماً، ستحبها وتحب الفيلم من أجلها فلديها حضور غريب ولا كأنها درو باري مور في "شالت الفيلم"، وفي مشاهدها مع هاني رمزي ستجده هو نفسه يقوم بالإضحاك بشكل أكبر وكأنها فتحت أمامه مجالاً أكبر خاصة وأنه ممثل كوميدي صارخ عندما تتاح له الفرصة ولا يحبس نفسه في ورق دمه تقيل وبتتذكر أدواره في "وجهة نظر" و"تخاريف" مثلاً، يتغير الفيلم تماماً بعد ظهور جنا وتبدأ خيوطه الدرامية الضعيفة ومبالغاته في أخذ شكل مقبول ولطيف وكل ذلك بسبب القبول الكبير الذي تحظي به الطفلة جنا، وسيؤكد ذلك ظهور الممثلة تتيانا لأك ستقول حينها لا حول ولا قوة إلا بالله فعلاً القبول من عند ربنا، في ظل محاولاتها المضنية لإضافة أي ملامح للشخصية وحركة القطة الكارتونية التي كانت تفعلها، يسير الفيلم علي هذه الشاكلة حتي تحدث بعض الأحداث المتوقعة والكليشهات المعروفة مثل أن يقتل المؤلف أم الطفلة في الفيلم ويصيبها بمرض مميت فتذهب الطفلة إلي أبيها سامي الذي يتحول بالطبع من شخص مهووس بالسنات لأب مثالي ثم تساعده الطفلة علي الزواج من حبيبته التي تنسي كل شئ من أجله ويتزوجه ويعيش الجميع سعداء، في نهاية سعيدة مثالية لأي فيلم عربي، الفيلم "الطيف" وقابل للمشاهدة إذا وضعته في حجمه وإذا دخلته من أجل أطفالك أو عشان تقسح إبن أختك أو ولاد صاحبك أو حتي عشان عندك وقت فاضي ومايز تاكل فشار وأنت بتتفرج علي حاجة ظريفة، هتضحك شوية وهتستمتع بأداء الطفلة جنا اللي فعلاً تدخل القلب، وأي محاولة لتحليل الفيلم سينمائياً سينمائياً فعلاً يعني" ستكون محاولة عبثية، فالفيلم في النهاية فيلم ظريف وليس من الغريب أن يحظي بإقبال في ظل ما نعانيه من ظروف وحاجة الناس لفيلم خفيف "يفصلها" من حالة الشد العصبي الدائمة".	Positive
freq	2	950

Figure 3. Exploration action (Statistical summary)

• View the count of non-null values contained in each column using count().

```
[42] data_df.count()
```

```
Text              1524
classification    1519
clean text        1524
tokens            1524
dtype: int64
```

Figure 4. Exploration action (columns non-null count)

• View each column with the corresponding non-null values count and its data type using info(). There is also a brief description of the datatypes included in the dataset.

```
[43] data_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1524 entries, 0 to 1523
Data columns (total 4 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Text            1524 non-null   object
 1   classification  1519 non-null   object
 2   clean text      1524 non-null   object
 3   tokens          1524 non-null   object
dtypes: object(4)
memory usage: 47.8+ KB
```

Figure 5. Exploration action (columns datatype and non-null values count)

3. Data Issues

Issue#1: Punctuation

Libraries used: pandas, and string

After examining the dataset, we noticed that the text has many unnecessary punctuations, so we decided to delete this punctuation. Figure 6 shows the list of punctuation that will be removed from the text in addition to the pre-specified punctuation that is provided with the string library using String.Puncturation constant [6].

```
arabic_punctuations = '''`÷×_—“…”!|+¦~{}',.؟":/،_][%^&*()_<>؛'''
english_punctuations = string.punctuation
punctuations_list = arabic_punctuations + english_punctuations
```

Figure 6. punctuation List

Figure 7 view the punctuation removal function that uses Re (Regular Expression) library to remove punctuation from a given text.

```
#this funcation take txet as input and return the text after removing punctuations
def remove_punctuations(text):
    translator = str.maketrans('', '', punctuations_list)
    return text.translate(translator)
```

Figure 7. remove punctuation function

Figure 8 view random sample before and after removing punctuation from the reviews.

```
data_df.sample(3)
```

	Text	classification	clean text
412	حلمى ما بين الدب باندا والقرد ميسى نجح احمد حلمى مع مسلسل ميسى وتجسيد صوتى للقرد ولاقى استحسان الكثير من الجمهور ولكنه فشل فى فيلمه (صنع فى مصر) بتجسيده لدمية دب باندا حيث تصيبه لعنة حينما تتمنى اخته الصغيرة (نور عثمان) ان يتحول اخيها الى دمية وان يصبح الدبدوب اخيها وبطريقه تفتقد الى اى حبكة درامية يتم التحول ويحاول علاء (احمد حلمى) ان يعود الى شخصيته الادمية التى كانت شخصيه فاشلة وكسولة وينجح فى النهايه ولكن بعد ان تتغير شخصيته ليصبح شاب يعُتَمد عليه اما بالنسبه للاداء التمثيلى فنجد الطفلة (نور عثمان) التى لفتت الانظار فى برنامج اراب جود ايدل بادائها المتميز بالاستعراض والغناء لكن لم يتم الاستفادة من مواهبها وكذلك الحال مع الممثل المتألق فى الفترة الاخيرة (بيومى فؤاد) و النجمة (دلال عبد العزيز) فادوارهما هشة لا تليق بهما .. اما ياسمين رئيس ذات الوجه الملائكى والتى تألقت فى (فتاة المصنع) فلم تأت بجديد اما (ادوار) فعلى الرغم من صغر الدور فانه جاء تغييراً له على الاقل كشكلاً	Positive	حلمي ما بين الدب باندا والقرد ميسي نجح احمد حلمي مع مسلسل ميسي وتجسيد صوتي للقرد ولاقي استحسان الكثير من الجمهور ولكنه فشل في فيلمه صنع في مصر بتجسيده لدميه دب باندا حيث تصيبه لعنه حينما تتمني اخته الصغيره نور عثمان ان يتحول اخيها الي دميه وان يصبح الدبدوب اخيها وبطريقه تفتقد الي اي حبكه دراميه يتم التحول ويحاول علاء احمد حلمي ان يعود الي شخصيته الادميه التي كانت شخصيه فاشله وكسوله وينجح في النهايه ولكن بعد ان تتغير شخصيته ليصبح شاب يعُتَمد عليه اما بالنسبه للاداء التمثيلي فنجد الطفله نور عثمان التي لفتت الانظار في برنامج اراب جود ايدل بادائها المتميز بالاستعراض والغناء لكن لم يتم الاستفاده من مواهبها وكذلك الحال مع الممثل المتالق في الفتره الاخيره بيومي فواد و النجمه دلال عبد العزيز فادوارهما هشه لا تليق بهما اما ياسمين رئيس ذات الوجه الملائكي والتي تالقت في فتاه المصنع فلم تات بجديد اما ادوار فعلي الرغم من صغر الدور فانه جاء تغييراً له علي الاقل كشكلاً
	شمشون (عنتر) و لبلب عن فيلم شمشون (عنتر) ولبلب : ********** ياترى اتفرجت كام مرة على فيلم عنتر و لبلب اكيد كتير ماهي الفضائيات عمالة تزيد و تعيد و ياترى هل انت قوي الملاحظة و لا بتتفرج و انت بتشرب حاجة ساقعة و بتاكل فاكهة و يمكن كمان بتذاكر لعيالك لحد ماييجي المدرس الخصوصي ما علينا المهم لو انت ركزت شويه هتلاقي كل من نطق اسم عنتر بالفيلم		شمشون عنتر و لبلب عن فيلم شمشون عنتر ولبلب ياتري اتفرجت كام مره علي فيلم عنتر و لبلب اكيد كتير ماهي الفضائيات عماله تزيد و تعيد و ياتري هل انت قوي الملاحظه و لا بتتفرج و انت بتشرب حاجه ساقعه و بتاكل فاكهه و يمكن كمان بتذاكر لعيالك لحد ماييجي المدرس الخصوصي ما علينا المهم لو انت ركزت شويه هتلاقي كل من نطق اسم عنتر بالفيلم الصوت مش مناسب مع حركه الشفايف دي اولا ثانيا هتلاقي الصوت مش بتاع الممثل نفسه يعني لو عبد الوار ث عسر المعلم رضوان بيقول عنتر مش هتلاقيه صوت عبد الوارث عسر

Figure 8. dataset before and after removing punctuation from the reviews

Issue#2: Text Normalization

Libraries used: pandas, and re

The Arabic text has many shapes for one letter. We will unify the appearance of these letters

Figure 9 view unify character's function that uses sub function in regex library to substitute the any occurrences of Arabic characters in the right with the one in the left in the given text [5].

```
def normalize_arabic(text):

    text = re.sub("[إأآا]", "ا", text)
    text = re.sub("ي" ,"ى", text)
    text = re.sub("ه" ,"ة", text)
    text = re.sub("ك" ,"گ", text)
    text = re.sub("و" ,"ؤ", text)

    return text
```

Figure 9. Unify Characters Function

Figure 10 view random sample before and after removing punctuation from the reviews. For example, the word 'أجنبى' in the first line appeared after the text normalization app with the letter "ي" instead of the letter "ى", and the letter "ا" instead of the letter "أ".

```
data_df.sample(3)
```

	Text	classification	clean text
1115	عندما يتحد الضعفاء عندما نشاهد أى فيلم أجنبى نجد طرفين طرف يمثل الخير تكمن قوته فى الارادة والعزيمة وطرف يمثل الشر يكمن فى الخوف ودائما يكون الطرف الشرير هو الطرف الأقوى فى البداية لأن الخوف من السهل جدا ان يستثار ويظهره أى شخص فى العالم فلا يوجد انسان على وجه الأرض لايخاف ولايمتلك نقطة ضعف تمكن من هزيمته بسهولة بداية الفيلم تقليدية جدا شاهدنا مثلها من قبل فى عدة أفلام مثلا فيلم الشبح لأحمد عز وهى ان يستقيط شخص فاقد الذاكرة فى مكان لايعرف ماذا أحضره الى هنا وماذا حدث وماذا جرى له ؟ يكتشف بطلنا (جاك فروست) الأربعة الأقوياء حراس الكون وهى شخصيات خيالية معروفة لاى طفل فى العالم (حورية الأسنان و بابا نويل ورجل الأحلام وأخيرا أرنب عيد الفصح)ولكل شخصية مكانها المعين وتمتلك قوتها الخاصة اضافة الى اتباعها المساعدين فى الأعمال المنوسبة الى كل حارس وكما يوجد الخير يوجد الشر والذى يدعى (بيتش بلاك)أو الملعب الأسود ومن اسمه فهو يمثل الخوف ويمتلك هو الاخر جنوده من الفرسان الذى يمثلون الكوابيس لأى طفل ويسببون له الخوف وكلما زاد الخوف كلما ازدادت طاقة بيتش بلاك السلبية والتى تمكنه من مواجهة الحراس يحاول (جاك الفروست)ان يكتشف لماذا تم اختياره كحارس من الحراس ويحاول سنتا كلوز (بابا نويل)ان يساعده ان يكتشف الطاقة الايجابية التى بداخله ليتمكن من هزيمة خوفه هزيمة بيتش بلاك	Neutral	عندما يتحد الضعفاء عندما نشاهد اي فيلم اجنبي نجد طرفين طرف يمثل الخير تكمن قوته في الاراده والعزيمه وطرف يمثل الشر يكمن في الخوف ودائما يكون الطرف الشرير هو الطرف الاقوي في البدايه لان الخوف من السهل جدا ان يستثار ويظهره اي شخص في العالم فلا يوجد انسان علي وجه الارض لايخاف ولايمتلك نقطه ضعف تمكن من هزيمته بسهوله بدايه الفيلم تقليديه جدا شاهدنا مثلها من قبل في عده افلام مثلا فيلم الشبح لاحمد عز وهي ان يستقيظ شخص فاقد الذاكره في مكان لايعرف ماذا احضره الي هنا وماذا حدث وماذا جري له يكتشف بطلنا جاك فروست الاربعه الاقوياء حراس الكون وهي شخصيات خياليه معروفه لاي طفل في العالم حوريه الاسنان و بابا نويل ورجل الاحلام واخيرا ارنب عيد الفصح ولكل شخصيه مكانها المعين وتمتلك قوتها الخاصه اضافه الي اتباعها المساعدين في الاعمال المنوسبه الي كل حارس وكما يوجد الخير يوجد الشر والذي يدعي بيتش بلاكاو الملعب الاسود ومن اسمه فهو يمثل الخوف ويمتلك هو الاخر جنوده من الفرسان الذي يمثلون الكوابيس لاي طفل ويسببون له الخوف وكلما زاد الخوف كلما ازدادت طاقه بيتش بلاك السلبيه والتي تمكنه من مواجهه الحراس يحاول جاك الفروست ان يكتشف لماذا تم اختياره كحارس من الحراس ويحاول سنتا كلوز بابا نويل ان يساعده ان يكتشف الطاقه الايجابيه التي بداخله ليتمكن من هزيمه خوفه هزيمه بيتش بلاك

Figure 10. Dataset before and after text normalization

Issue#3: English Characters

Libraries used: pandas, and re

One of the problems we have noticed in the dataset is that some reviews contain English characters, so to focus only on the Arabic text, we will remove all the English characters from the reviews, using the re (Regular expression) library in Figure 11, we remove any expression that matches the regular expression given.

```
def processPost(tweet):

    #remove english letters
    tweet= re.sub(r'[a-z]+'," ", tweet)
    tweet= re.sub(r'[A-Z]+'," ", tweet)


    return tweet
```

Figure 11. Remove English Characters Function

Figure 12 view a random sample before and after removing English characters from the reviews.

```
data_df["clean text"] =data_df["Text"].apply(lambda x: processPost(x))
```

```
data_df.sample(3)
```

	Text	classification	clean text
570	تغيير الماضي أفلام العودة للماضي، وأمنية الرجوع للوراء أملا في تحقيق الأمنيات؛ من أكثر الأفلام اللي ممكن تلاقي فيها نفسك، لإن ببساطة كل شخص فينا لو أتيحت له فرصة العودة للماضي وتغيير حياته، هيستغل الفرصة دي أفضل استغلال من الأفلام اللي بتتناول فكرة العودة للماضي أملا في تحقيق أمنية (Free Birds) ممكن ومعتقدتش إننا هنفوتها كدا. وفيلم ما. المخرج (جيمي هوارد)، والمؤلفان (كريج مازن وديفيد اي ستيرن) يقدمان توليفة بسيطة عن تلك التيمة من خلال قصة شيقة بتدور حول الديك الرومي (ريجي) اللي بيحاول السفر للماضي وتحديدا عام 1621 مستخدما آلة الزمن حتى يتمكن من تغيير عادة البشر في أكل الديوك الرومي ليلة عيد الميلاد..... ومن هنا هنلاحظ إن قصة الفيلم بتمتاز بالبساطة، وبتركز من خلال حوار شيق على الكثير من النصائح والرسائل الهامة على المستوى العام سواء للأطفال أو الكبار في إطار كوميدي خفيف، فالتعاون ومحاولة الاتحاد لتحقيق غاية ومصلحة جماعية تأتي في المقام الأول ثم المصلحة الفردية. محاولة نسيان الخلافات الشخصية وعدم الالتفات لها أملا في تحقيق تلك الأمنية الجماعية. وطبعا من المعروف أن البطل الأول والأساسي في أي فيلم رسوم متحركة بيتعمد على الجرافيك والألوان وتنفيذ المخاطر والمغامرات بحرفية وتقنية عالية، بالإضافة إلى الصوت واللي اعتقد أن أبطال العمل نجحوا بشكل جيد في توصيل تفاعلات وانفعالات الديوك مع بعضها البعض ومع البشر، والمطاردات اللي تعرضوا لها طوال أحداث الفيلم؛ فمثلا الممثل (اوين ويلسون) قدم دور الديك ريجي بطريقة مبهرة وقدر من خلال مستويات صوت مختلفة أنه يوصل سعي ريجي الدائم لإلغاء صنف الديوك من قائمة الطعام في تلك الليلة، وتعاونه مع في أول إسبوع له بصالات عرض الولايات (Free Birds) صديقه الديك جاك (وودى هاريلسون) لتحقيق ذلك. وقد حقق فيلم المتحدة الأمريكية حوالي 15 مليون دولار أمريكي، وقد ظل في المركز الثالث بالبوكس أوفيس الأمريكي لأكثر من أسبوعين من الأفلام اللي ممكن تسيب ابنك (Free Birds) متتالين ثم بدأ في التراجع للخلف حتى المركز الخامس. اعتقد أن فيلم ادامها وانت مطمئن	Positive	تغيير الماضي افلام العوده للماضي وامنيه الرجوع للوراء املا في تحقيق الامنيات من اكثر الافلام اللي ممكن تلاقي فيها نفسك لان ببساطه كل شخص فينا لو اتيحت له فرصه العوده للماضي وتغيير حياته هيستغل الفرصه دي افضل استغلال ممكن ومعتقدتش اننا هنفوتها كدا وفيلم من الافلام اللي بتتناول فكره العوده للماضي املا في تحقيق امنيه ما المخرج جيمي هوارد والمولفان كريج مازن وديفيد اي ستيرن يقدمان توليفه بسيطه عن تلك التيمه من خلال قصه شيقه بتدور حول الديك الرومي ريجي اللي بيحاول السفر للماضي وتحديدا عام مستخدما اله الزمن حتي يتمكن من تغيير عاده البشر في اكل الديوك الرومي ليله عيد الميلاد ومن هنا هنلاحظ ان قصه الفيلم بتمتاز بالبساطه ويتركز من خلال حوار شيق علي الكثير من النصائح والرسائل الهامه علي المستوي العام سواء للاطفال او الكبار في اطار كوميدي خفيف فالتعاون ومحاوله الاتحاد لتحقيق غايه ومصلحه جماعيه تاتي في المقام الاول ثم المصلحه الفرديه محاوله نسيان الخلافات الشخصيه وعدم الالتفات لها املا في تحقيق تلك الأمنيه الجماعيه وطبعا من المعروف ان البطل الاول والاساسي في اي فيلم رسوم متحركه بيتعمد علي الجرافيك والالوان وتنفيذ المخاطر والمغامرات بحرفيه وتقنيه عاليه بالاضافه الي الصوت واللي اعتقد ان ابطال العمل نجحوا بشكل جيد في توصيل تفاعلات وانفعالات الديوك مع بعضها البعض ومع البشر المثل اوين ويلسون قدم دور الديك ريجي بطريقه مبهره وقدر من خلال مستويات صوت مختلفه انه يوصل سعي ريجي الدائم لالغاء صنف الديوك من قائمه الطعام في تلك الليله وتعاونه مع صديقه الديك جاك وودي هاريلسون لتحقيق ذلك وقد حقق فيلم في اول اسبوع له بصالات عرض الولايات المتحده الامريكيه حوالي مليون دولار امريكي وقد ظل في المركز الثالث بالبوكس اوفيس الامريكي لأكثر من اسبوعين متتاليين ثم بدا في التراجع للخلف حتي المركز الخامس اعتقد ان فيلم من الافلام اللي ممكن تسيب ابنك ادامها وانت مطمئن

Figure 12 Dataset before and after removing English Characters

4. Tool Used

These are the tools and libraries we have used during this phase.

1. Pands

A Python package providing fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language [4].

2. Re

A regular expression is a special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern [5].

3. String

Python String module contains some constants, utility function, and classes for string manipulation [6].

4. NLTK

(Natural Language Toolkit) is an open-source Python library for Natural Language Processing [7].

5. Reference

[1] "What Is Data Analysis? Methods, Techniques, Types & How-To," *BI Blog |Data Visualization & Analytics Blog | datapine,* Mar. 09, 2022. https://www.datapine.com/blog/data-analysis-methods-and-techniques/ (accessed August.04, 2022).

[2] "What Is Elcinema Website" | *The Largest Arabic Movie Database* | elcinema, August. 04, 2022. https://elcinema.com/ (accessed August.04, 2022).

[3] W. Tian, Y. Zhao, Optimized Cloud Resource Management and Scheduling. Elsevier, 2014.

[4] "pandas – Python Data Analysis Library," *Pydata.org,* 2022. https://pandas.pydata.org/ (accessed August.05,2022).

[5] "re — Regular expression operations — Python 3.10.3 documentation," *Python.org,* 2022. https://docs.python.org/3/library/re.html (accessed August. 05, 2022).

[6] "string — Common string operations — Python 3.10.4 documentation," *Python.org,* 2022. https://docs.python.org/3/library/string.html (accessed August. 06, 2022).

[7] "NLTK - nltk package," nltk.org, 2022. [Online]. Available: https://www.nltk.org/api/nltk.html (accessed August. 06, 2022).

[8] "tokenizers — splits a string into substrings using a regular expression— Python 3.10.4 documentation," *Python.org,* 2022. https://www.nltk.org/_modules/nltk/tokenize/regexp.html (accessed August. 06, 2022).

[9] "Google colab" Colab.research, 2022. [Online]. https: //colab.research.google.com/ (accessed August. 05, 2022).