**SOCIAL MEDIA COURSEWORK  
Margot robbıe**

|  |
| --- |
| *Hasan E. Guray* |
|  |
| MSc Data Science and Analytics student |
| University of Westminster |
| w1948912@my.westminster.ac.uk |

ABSTRACT

This research paper presents an exploratory data analysis and visualization study using Power BI to investigate the profound impact of the COVID-19 pandemic on the global video streaming market. The study analyses data obtained from Statista, exploring key metrics such as subscription trends, user penetration rate, and market revenues during and after the COVID-19 outbreak. Through 15 different graphs, the research delves into shifts in user behaviour, industry trends, and financial implications for streaming platforms, providing valuable insights for industry stakeholders to adapt their strategies and maintain a competitive edge in the transformed digital entertainment landscape.

**Keywords**: COVID-19, Video Streaming Market, Exploratory Data Analysis, Data Visualization, Power BI, User Behaviour, Subscription Trends, Market Revenues.

# PERSON SELECTED

Margot Elise Robbie, the celebrated Australian actress and producer, was born on July 2, 1990, in Australia. After completing drama studies at Somerset College, she made her professional acting debut at 17 and relocated to Melbourne to pursue her passion. In the late 2000s, she gained recognition for her performances in Australian independent films and further acclaim for her roles in the soap opera "Neighbors" (2008–2011), earning two Logie Award nominations. Moving to the United States, she found success in the TV series "Pan Am" (2011–2012) and achieved her breakthrough in 2013 with prominent roles in "About Time" and "The Wolf of Wall Street." Margot continued her ascent with roles in various films, including "Focus" (2015), "Z for Zachariah" (2015), "The Big Short" (2015), "The Legend of Tarzan" (2016), and "Suicide Squad" (2016), where she portrayed Harley Quinn. Her exceptional performance as Tonya Harding in "I, Tonya" earned her Oscar and Golden Globe nominations in 2018. Aside from her acting success, Margot Robbie's influence was widely acknowledged, with Time magazine listing her among the "World's 100 Most Influential People" in 2017, and Forbes recognizing her as one of the highest-paid actresses in 2019. She also showcased her versatility and impact as a producer through her company, LuckyChap Entertainment.

Recently, Margot's popularity surged due to her role in the 2023 American fantasy comedy film "Barbie." Directed by Greta Gerwig and co-written with Noah Baumbach, the movie presents Margot Robbie as Barbie and Ryan Gosling as Ken, marking the first live-action adaptation of the iconic fashion dolls by Mattel. The plot revolves around a journey of self-discovery triggered by an existential crisis. The film boasts an ensemble cast, including America Ferrera, Kate McKinnon, Issa Rae, Rhea Perlman, and Will Ferrell. "Barbie" premiered on July 9, 2023, in Los Angeles and had a successful theatrical release, grossing over $578 million worldwide, becoming the eighth highest-grossing film of 2023.

# Data COLLECTION

Günümüzde teknoloğinin etkisiyle bilgi ve düşünceler çok hızlı yayılabiliyor, bunda da sosyal medyanın çok büyük bir katkısı var. Birçok sosyal medya platformu kullanıcılarına ücretsiz bir şekilde fikirlerini paylaşıp diğer kullanıcılarla etkileşime girebileceği sanal ortamı oluşturuyor. Margot Robbie hakkında detaylı bilgi edinebilmek ve özellikle son süreçte yaptığı işlerin insanlar açısından nasıl değerlendirildiğini görebilmek için oluşturulacak verisetinde trafiği olabildiğince yüksek bir platform ve API desteğini ücretsiz sunan bir platform olan Reddit'ten yararlanıldı. 2023 Haziran verilerine göre Reddit’in dünya çapında en yüksek trafiğe sahip 10. Websitesi olması tercih edilmesinde rol oynadı. (<https://www.semrush.com/website/top/> accessed by 29 July 2023) Özellikle son zamanlarda yükselen popularitesinin etkisini inceleyebilmek için filmin çıkış tarihi ve sonrasındaki bir haftalık sürece odaklanılarak 21-27 Temmuz 2023 tarih aralığı tercih edildi. Python’ın pandas, praw ve datetime kütüphaneleri aracılığıyla ve Reddit API desteğiyle 1633 satırdan oluşan comment veriseti oluşturuldu (Figure 1).



**Figure 1** *This is a screenshot of Python code to scrape Reddit comments about Margot Robbie.*

Kullanıcıların birbirileriyle olan diyaloglarıyla verisetinde karışıklılık yaratmamak adına sadece top level comments’e odaklanıldı. Ayrıca bazı commentlerin system tarafından otomatik oluşturulduğu ve bazılarının ise diğer kullanıcılara hitaben yazılması sebebiyle o commetler exclude edildi. Sonucunda comment’in kendisini içinde barındıran “Comment Body”, commentin hangi kullanıcı tarafından yazıldığını içeren “Author”, commente yapılan etkileşimleri içeren “Upvotes” ve “Awards” ve son olarak da commentin tam olarak ne zaman platformda paylaşıldığı bilgisini içeren “Date” kolonlarıyla veriseti oluşturuldu (Figure 2).

A screenshot of a computer

Description automatically generated

**Figure 2** *This is a screenshot of Pandas dataframe, consists Reddit comments about Margot Robbie.*

Elbette ki veriseti raw data ve doğrudan kullanıcıların comment’lerini içerdiği için exploratory data analysis öncesinde bazı data cleaning ve preprocessing işlemlerinin yapılması gerekir. Bunun için veriseti incelendiğinde öncelikle “Date” kolonunun type’ı datetime olarak değiştirildi (Figure 3).

A screenshot of a computer

Description automatically generated

**Figure 3** *This is a screenshot of Pandas dataframe’s column details, consists Reddit comments about Margot Robbie.*

Topic model analysis ve sentiment analysis işlemlerinin daha başarılı olabilmesi için raw commentlerin bazı işlemlerden geçirilerek analizlere hazır hale getirilmesi gerekir. Commentin contractions’lardan kurtarılması için Python’ın contractions kütüphanesinden yararlanılarak comment’ler standart bir hale getirildi (Figure 4).

A screenshot of a computer

Description automatically generated

**Figure 4** *This is a screenshot of Python formula to fix contractions.*

Ardından price, size, volume ve weight içeren kelimelerin standartlaştırılması için Python’ın re kütüphanesinden yararlanıldı (Figure 5). Böylece birbirinden farklı rakamların analizleri farklı etkilemesinin önüne geçildi.

A screenshot of a computer program

Description automatically generated

**Figure 5** *This is a screenshot of Python formula to substitute patterns.*

Internet ortamında kullancıları yönlendirmek için birçok link commentler içerisinde yer almaktadır. Bu bazen düşünceleri açıklamak ya da desteklemek için kullanılırken, bazen de mizahi yorum katmak aracılığıyla kullanılır. Analizlerin daha başarılı olabilmesi için bu linklerin standartlaştırılması gerektiğinden, yine re kütüphanesi araclığıyla linkler tek bir başlık altında toplanır (Figure 6).

A screenshot of a computer

Description automatically generated

**Figure 6** *This is a screenshot of Python formula to replace http links.*

Her ne kadar Reddit platformunda surf edildiğinde karşılaşılmasa da, scraping ile edilen verisetinde line geçişlerini sembolize etmek için bazı commentlerde “\n” ifadesi yer almaktadır. Herhangi bir anlamı olmayan bu ifadelerin ortadan kaldırılması gerektiği için Python’ın replace formülünden yaralanılır (Figure 7).

A screenshot of a computer

Description automatically generated

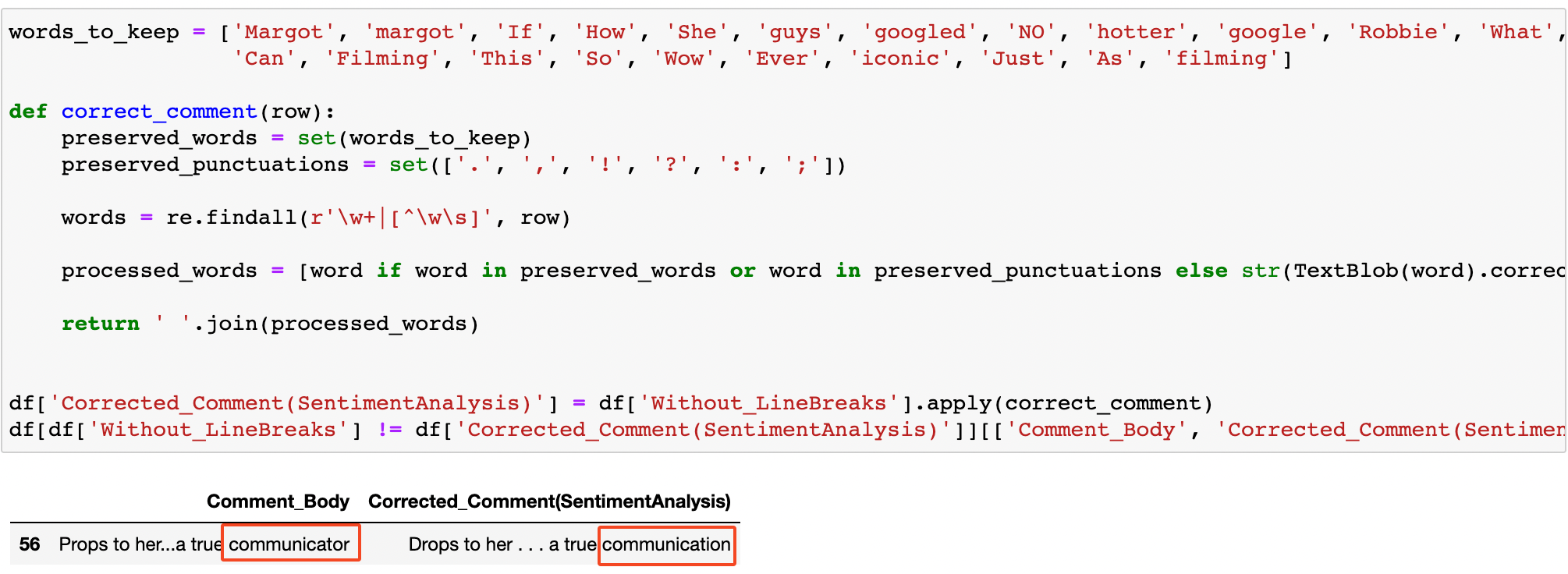
**Figure 7** *This is a screenshot of Python formula to remove line breaks.*

Hızlı yazılan commentler, sosyal media özelindeki kısaltmalar ve daha birçok sebepten dolayı yazım yanlışlarıyla karşılaşılmaktadır. Bu ifadeler eğer düzeltilmezse analizlerin yanlış sonuçlar çıkarmasına sebep olabilmektedir. Python’ın TextBlob kütüphanesinin desteğiyle, bu yazım yanlışlarını düzeltmek ve de aynı köke sahip kelimeleri standartlaştırmak mümkün, ancak tüm comment tek seferde işleme sokulduğunda bazı doğru kelimeleri yanlış hale döndürebiliyor (Figure 8). Ancak, yanlış çevirdiği kelimeler belirlenip, kelimeler tek tek TextBlob’dan geçirildiğinde daha doğru sonuçlar doğurdu (Figure 9).57

A screenshot of a computer

Description automatically generated

**Figure 8** *This is a screenshot of Python formula to correct the words.*



**Figure 9** *This is a screenshot of Python formula to correct the words.*

Topic model analysis’in daha başarılı gerçekleştirilebilmesi için olabildiğince standartlaştırma gerektiğinden, tüm harfleri lowercase yapmak da bu duruma katkı sağlar. Python’ın lower() formülü sayesinde, tüm comment’lerde yer alan tüm harfler lowercase’e çevrildi (Figure 10). Ancak büyük ya da küçük harf kullanım farkı farklı duygu durumlarını belirtmek ya da duygu durumunu vurgulamak için kullanıdığından sentiment analysis’i etkileyebilmektedir. Dolayısıyla bu ve bundan sonraki adımlar sentiment analysis için uygulanmayacaktır.

A screenshot of a chat

Description automatically generated

**Figure 10** *This is a screenshot of Python formula to convert to lowercase all of the letters.*

Topic model analysis sürecinin daha başarılı olmasını sağlayan bir diğer işlem tokenization’dur. Tokenization is a critical step in preparing text data for topic model analysis. It transforms unstructured text into a structured format that allows topic modeling algorithms to identify patterns and extract meaningful topics. By standardizing the representation of words and removing noise, tokenization improves the accuracy and interpretability of the topic modeling process. Python’ın nltk kütüphanesi ile tokenization işlemi gerçekleştirildi (Figure 11).

A screenshot of a computer

Description automatically generated

**Figure 11** *This is a screenshot of Python formula to convert to tokenize the words.*

Cümleleri bağlamak ve daha anlaşılır hale getirmek için tek başına anlamı olmayan bazı stopwords kullanılmaktadır. Removing stopwords in topic model analysis is necessary to reduce the influence of common, non-informative words that can obscure meaningful patterns. This process helps improve the quality and interpretability of the extracted topics by reducing noise, improving coherence, and focusing the model on more relevant and distinctive terms. Python’ın nltk kütüphanesi ile removing stopwords işlemi gerçekleştirildi (Figure 12).

A screen shot of a computer

Description automatically generated

**Figure 12** *This is a screenshot of Python formula to remove stopwords.*

Comment’lerin olmazsa olmazı emojiler ve noktalama işaretleridir. Removing emojis and punctuation in topic model analysis is essential to reduce noise, ensure consistency, and improve interpretability of the extracted topics. This preprocessing step helps to focus on meaningful words, reduce dimensionality, and create a cleaner and more standardized text representation for efficient and accurate topic modeling. Python’ın re ve string kütühaneleriyle emojiler ve punctations silindi (Figure 13).



**Figure 13** *This is a screenshot of Python formula to remove punctations and emojis.*

Lemmatization is a valuable preprocessing step for topic model analysis as it reduces words to their base forms, enhancing word normalization and consolidating inflected word variations. By improving interpretability, coherence, and reducing dimensionality, lemmatization helps produce more meaningful and efficient topic modeling results, capturing the core themes of the data more accurately. Python’ın nltk kütüphanesinin desteğiyle lemmatization işlemi gerçekleştirildi (Figure 14).

A screenshot of a computer

Description automatically generated

**Figure 14** *This is a screenshot of Python formula for lemmatisation.*

Aynı kullanıcıların aynı yorumlarının analiz işlemine girip karmaşıklık yaratmaması adına duplicates’in silinmesi gerekmektedir. Python’ın drop\_duplicates() formülü ile, duplicates sorununun önüne geçildi (Figure 15).



**Figure 15** *This is a screenshot of Python formula to remove duplicates.*

Yapılan bu cleaning and preprocessing işlemleri sonrasında bazı comment’ler tamamen silindi ve na values olarak verisetinde yer aldı. Python’ın dropna() formülüyle na olan veriler verisetinden kaldırıldı (Figure 16).

A close up of a sign

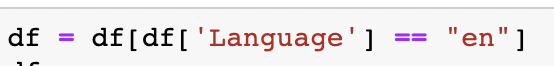
Description automatically generated

**Figure 16** *This is a screenshot of Python formula to remove na records.*

Commentlerin hangi dilde yazıldığı analizleri doğrudan etkilemektedir, çünkü analizler belirli bird il odağında gerçekleştirilmektedir. Most common language English olduğu için, English olmayan recordların kaldırılması daha başarılı sonuçlar doğurur, dolayısıyla Python’ın langdetect kütüphanesiyle tespit edilen language’a göre English olmayan comment’ler verisetinden kaldırılır ve böylece geriye 1105 satırlık veri kalır (Figure 17).

A screenshot of a computer program

Description automatically generated



**Figure 17** *This is a screenshot of Python formula to keep just English comments.*

Böylece veriseti için tüm preprocessing ve cleaning işlemi tamamlandı, verisetini daha iyi anlamak ve analizleri gerçekleştirmek için exploratory data analysis adımına geçildi.

# EXPLORATORY ANALYSIS

Present and evaluate findings of your initial

exploratory data analysis. Critically evaluate and reflect on the methods you used

and provide interpretation of the results. In this section you should include any

visualisations you feel are relevant to highlight important features of your dataset.

Verilerin toplanma aralığı için Barbie filminin çıkış tarihi itibariyle bir haftalık period tercih edilmişti. 21-27 Temmuz 2023 aralığında yer alan verisetinde her gün hemen hemen aynı sayıda comment sayısına ulaşılırken en yüksek comment 25 July 2023 tarihinde kaydedildi (Figure 18). Her ne kadar filmin çıkış tarihi 21 Temmuz 2023 olsa da, o gün yapılan commentlerin sayısı dominant bir oran oluşturmadı.

A graph of a person

Description automatically generated

**Figure 18** *This is a histogram of comments count related with Margot Robbie in Reddit.*

Atılan comment’lerin farklı kullanıcılardan olması analizlerin daha başarılı gerçekleşmesini sağlayacağı için belirli authors’ta yoğunlaşma olup olmadığı control edildi. En fazla comment sayısına sahip author’un sadece5 comment’i olması verisetinin farklı kişilerin duygu ve düşüncelerinden oluştuğunu kanıtladı (Figure 19).

A graph of a number of authors

Description automatically generated

**Figure 19** *This is a bar graph of comments count of the authors related with Margot Robbie in Reddit.*

Kullanıcıların birbirileriyle etkileşime girebileceği bir system olan upvotes ve awards sistemine göre en çok upvotes ve awards alan comments incelendi. Margot Robbie’nin bir hayranıyla işaret diliyle iletişime geçmesiyle ilgili olarak AncientOneders adlı kullanıcının yazdığı “I learned some sign language in GATE as a kid, and eventually I had a deaf women come through my gas station. I signed the letters for "receipt", asking her if she needed one, and her eyes lit up, and I got a big grin, never gonna forget that interaction.” commenti 4524 upvotes ile en çok upvotes alan comment oldu. SparklingLama adlı kullanıcının “I love how excited she looks” commenti ise en çok awards sayısı olan 1 award’a sahip commentlerden biri oldu.

Reddit global birr platform olduğu için sadece English commentler verisetinde yer almadı. Aynı zamanda diğer dillerle yazılan commentlerde yer aldı. Dil kırılımı incelendiğinde verisetinin 69.6%’sı olan 1105 comment English ilk sırada yer alırken, Afrikaans, Fransızca takip eden diller oldu (Figure 20). Comment’lerin ortalama uzunluğuna bakıldığında ise, yine English 88.70 karakterle ilk sırada yer alırken, 36.16 ile İtalyanca onu takip etti (Figure 21).

A white rectangular object with yellow text

Description automatically generated

**Figure 20** *This is a bar graph of comments count of the languages related with Margot Robbie in Reddit.*

A graph of a number of blue bars

Description automatically generated

**Figure 20** *This is a bar graph of average comment length of the languages related with Margot Robbie in Reddit.*

# TOPIC MODEL ANALYSIS

Describe the steps you took to develop and apply a

suitable topic model to your dataset. Critically evaluate the use of topic modelling to

generate insight and provide interpretation of the results.

# SENTIMENT ANALYSIS

Document the steps taken to identify the sentiment or

emotion expressed towards the person or their activities. Discuss the role of

sentiment analysis in social media modelling and provide a critical evaluation of your

findings.

# CONCLUSION

In this section you should summarise the main findings of your

analysis. You should provide a critical evaluation of your approach, identifying

aspects of your analysis which could be improved or developed further.

references

Undoubtedly, it has been the Barbie movie that has increased the popularity of Margot Robbie in recent days.In the 2023 American fantasy comedy film "Barbie," directed by Greta Gerwig and co-written with Noah Baumbach, Margot Robbie stars as Barbie alongside Ryan Gosling as Ken. The movie is based on the popular Barbie fashion dolls by Mattel and is the first live-action adaptation, departing from previous computer-animated releases. The plot follows Barbie and Ken as they embark on a journey of self-discovery triggered by an existential crisis. The star-studded ensemble cast includes America Ferrera, Kate McKinnon, Issa Rae, Rhea Perlman, and Will Ferrell. The film's development spanned several years and production took place in both England and Los Angeles. "Barbie" premiered on July 9, 2023, in Los Angeles and was released theatrically in the United States on July 21, 2023, by Warner Bros. Pictures. It received positive reviews from critics and achieved significant commercial success, grossing over $578 million worldwide, making it the eighth highest-grossing film of 2023.