



Exploring Fan Sentiments: Unsupervised Learning and Sentiment Analysis in Formula 1



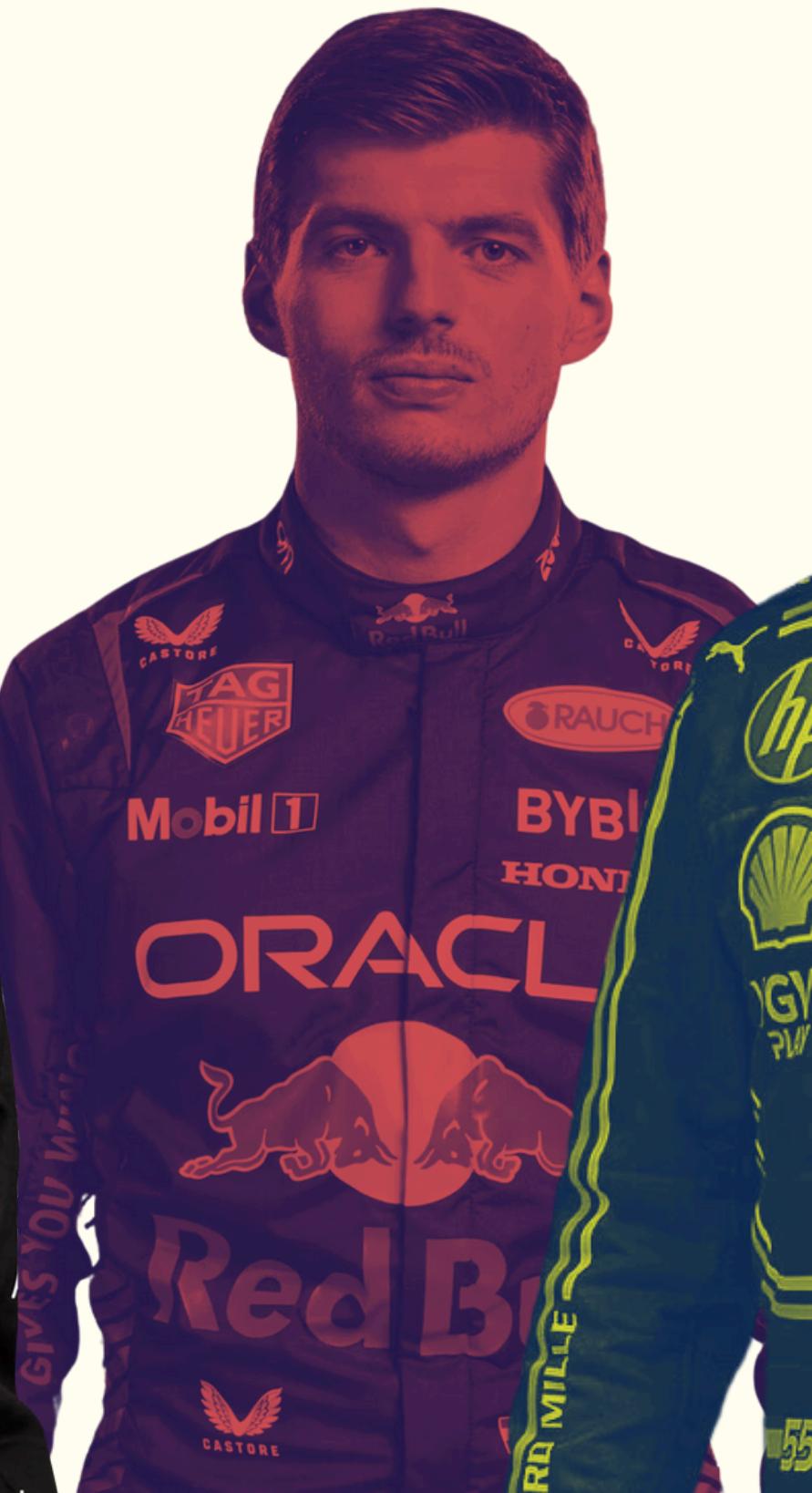
UPDATE

Lewis Hamilton to leave Mercedes and join Ferrari for 2025 Formula 1 season

Lewis Hamilton will join Ferrari next year and replace Carlos Sainz in a shock move; Hamilton signed a two-year contract with Mercedes which expires at the end of 2025 but he will spend just one more season with the team; **watch every F1 race in 2024 live on Sky Sports F1**

By Nigel Chiu and Craig Slater

⌚ Friday 2 February 2024 00:18, UK



Andrea Kimi Antonelli



- 2023 Formula Regional Europe Champion
- 2023 Formula Regional Middle East Champion
- 2022 Italian F4 Champion
- 2022 ADAC F4 Champion
- 2021 European Karting Champion
- 2020 European Karting Champion

Typical Road to Formula 1



Formula Regional





SR = Sprint Race	
FR = Feature Race	
Driver	Points
1 I. HADJAR	165
2 G. BORTOLETO	154.5
3 Z. MALONEY	135
4 P. ARON	124
5 J. CRAWFORD	105
6 A. ANTONELLI	99



Formula Regional

What is Sentiment Analysis?

Sentiment analysis is a technique used to determine the emotional tone behind words. It assesses text to categorize the sentiment as positive, negative, or neutral.

IMPORTANCE

Helps understand how people feel about a specific topic or event.

Application in Formula 1:

- Understand fan reactions to major news, such as:
 - Lewis Hamilton's move to Ferrari.
 - Speculation on who will replace him at Mercedes.
- Analyze thousands of comments from social media, forums, and news articles.
- Gauge if fans are excited, upset, or unsure about changes.

Tools for Sentiment Analysis: VADER and TextBlob

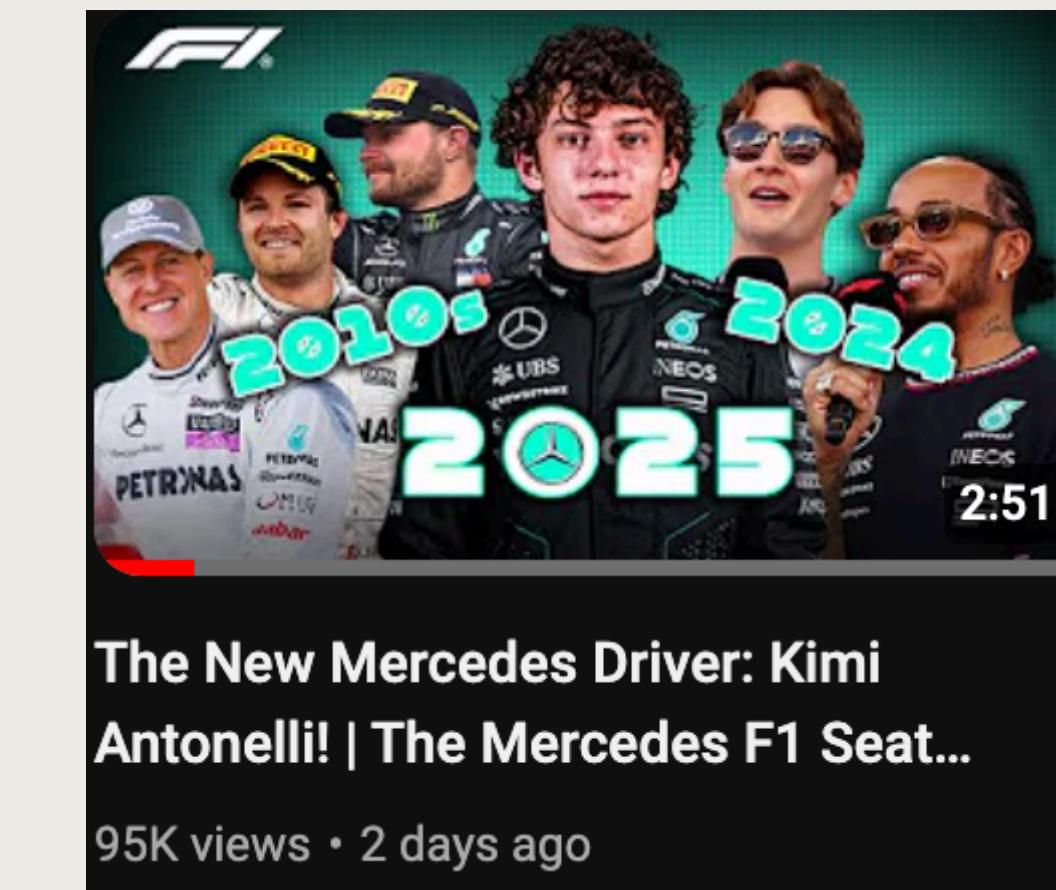
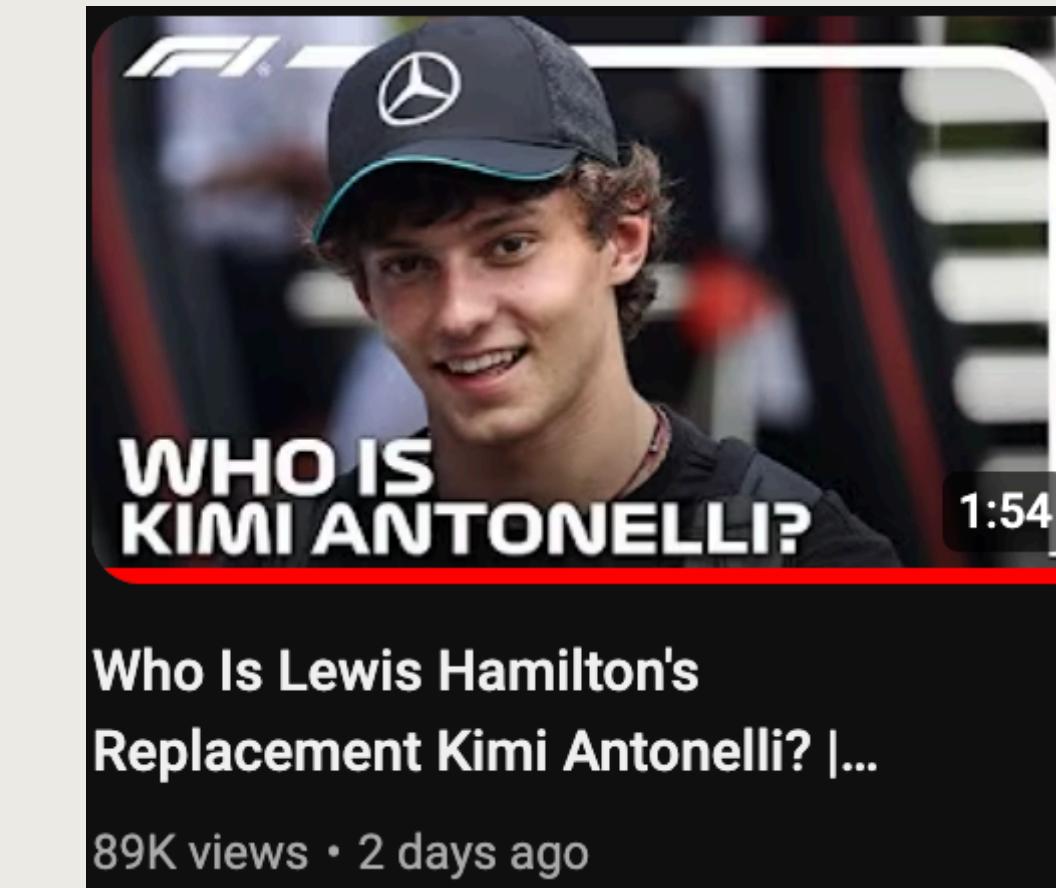
VADER (VALENCE AWARE DICTIONARY AND SENTIMENT REASONER):

- Specially designed to understand the tone of text in social media.
- **Capabilities:**
- Detects whether sentiment is positive, negative, or neutral.
- Measures the strength of the sentiment (intensity).
- Best For: Analyzing comments and opinions online, where text can be informal or include slang.

TEXTBLOB:

- Provides a simple API for common natural language processing (NLP) tasks.
- **Capabilities:**
- Determines if the overall tone of a text is positive, negative, or neutral.
- More generalized and straightforward.
- Limitations: May not always catch the nuances of informal or social media language as effectively as VADER.

How to extract comments from YouTube videos?





Setting Up the YouTube Data API

- A service provided by Google that allows developers to interact with YouTube's data.
- Can retrieve video information, playlists, channel details, and, importantly, comments.

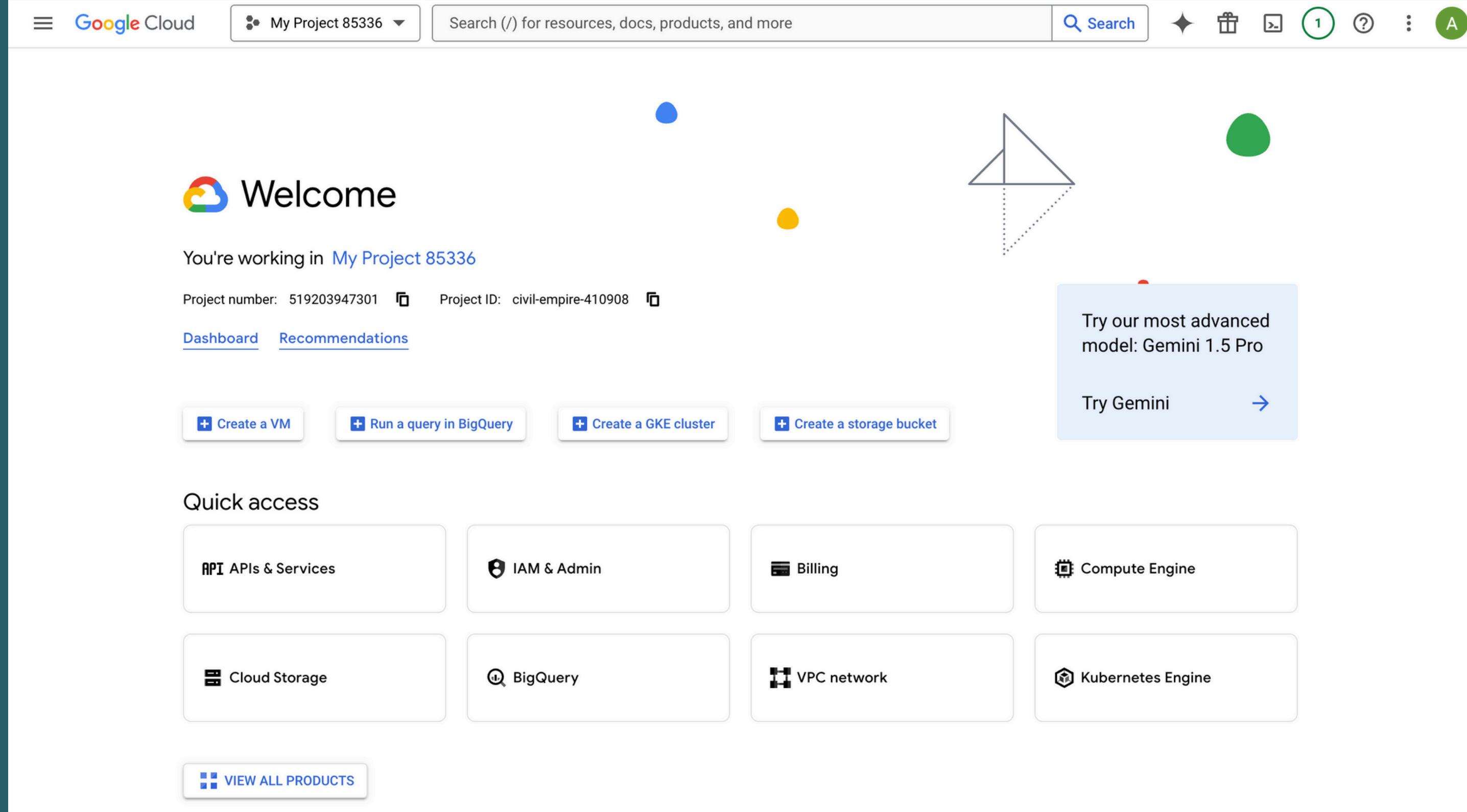
Requirements for Access:

- API Key: Needed for authenticating requests to the YouTube Data API.
- Google Cloud Platform Account: Set up an account and enable the YouTube Data API.
- Python Environment: Use libraries like googleapiclient for API interaction.

Step-by-Step Setup:

- Create a Project in Google Cloud Platform.
- Enable the YouTube Data API in the project.
- Generate an API Key to use in your application.

Setting Up the YouTube Data API



The image shows the Google Cloud Platform (GCP) dashboard. At the top, there's a navigation bar with the Google Cloud logo, a dropdown for 'My Project 85336', a search bar, and various icons for account management and notifications. A green notification bubble with the number '1' is visible.

The main area features a 'Welcome' section with a colorful cloud icon. It displays the project name 'My Project 85336', project number '519203947301', and project ID 'civil-empire-410908'. Below this, there are two navigation links: 'Dashboard' and 'Recommendations'. A call-to-action button for 'Try Gemini' is present, along with a note about the 'Gemini 1.5 Pro' model.

The 'Quick access' section contains eight boxes, each with an icon and a label: 'API APIs & Services', 'IAM & Admin', 'Billing', 'Compute Engine', 'Cloud Storage', 'BigQuery', 'VPC network', and 'Kubernetes Engine'. At the bottom left, there's a 'VIEW ALL PRODUCTS' button.



Extracted Data From API

	video_id	author	published_at	updated_at	like_count	text
0	rcpw6vxmVVg	@randallvargas4457	2024-09-01T05:26:21Z	2024-09-01T05:26:21Z	0	This is great for the sport! What a year! YES!
1	rcpw6vxmVVg	@jeff69cpl	2024-09-01T01:26:18Z	2024-09-01T01:26:18Z	0	Insane that Mercedes has been in Formula 1 sin...
2	rcpw6vxmVVg	@MaxiF.	2024-09-01T00:01:37Z	2024-09-01T00:01:37Z	0	Schumacher retires in 2012 after begun in '...
3	rcpw6vxmVVg	@wockie6342	2024-08-31T21:10:21Z	2024-08-31T21:10:21Z	0	In 2023, you would be called crazy if you said...
4	rcpw6vxmVVg	@Deif311	2024-08-31T18:18:12Z	2024-08-31T18:18:12Z	0	Lewis hamilton masih bisa di andalkan di bandi...
...
793	AT-mHHW8m8M	@egj1975	2024-08-31T12:04:17Z	2024-08-31T12:04:17Z	0	In 25 (the year 2025). He is only 18 years old 😊
794	AT-mHHW8m8M	@tommysommers9295	2024-08-31T11:58:29Z	2024-08-31T11:58:29Z	2	wow 22 seconds and 1 view
795	AT-mHHW8m8M	@vanditdostonkadost	2024-08-31T11:58:26Z	2024-08-31T11:58:26Z	0	Not my Kimi
796	AT-mHHW8m8M	@ggschatten	2024-08-31T11:58:23Z	2024-08-31T11:58:23Z	0	sheesh 3rd comment
797	AT-mHHW8m8M	@AmMachy	2024-08-31T11:58:17Z	2024-08-31T11:58:17Z	0	1st comment

798 rows × 6 columns

Text Preprocessing

1. CONVERTING TEXT TO LOWERCASE

2. EXPANDING CONTRACTIONS

3. REMOVE NON ALPHABETIC (EXCEPT NUMBERS)

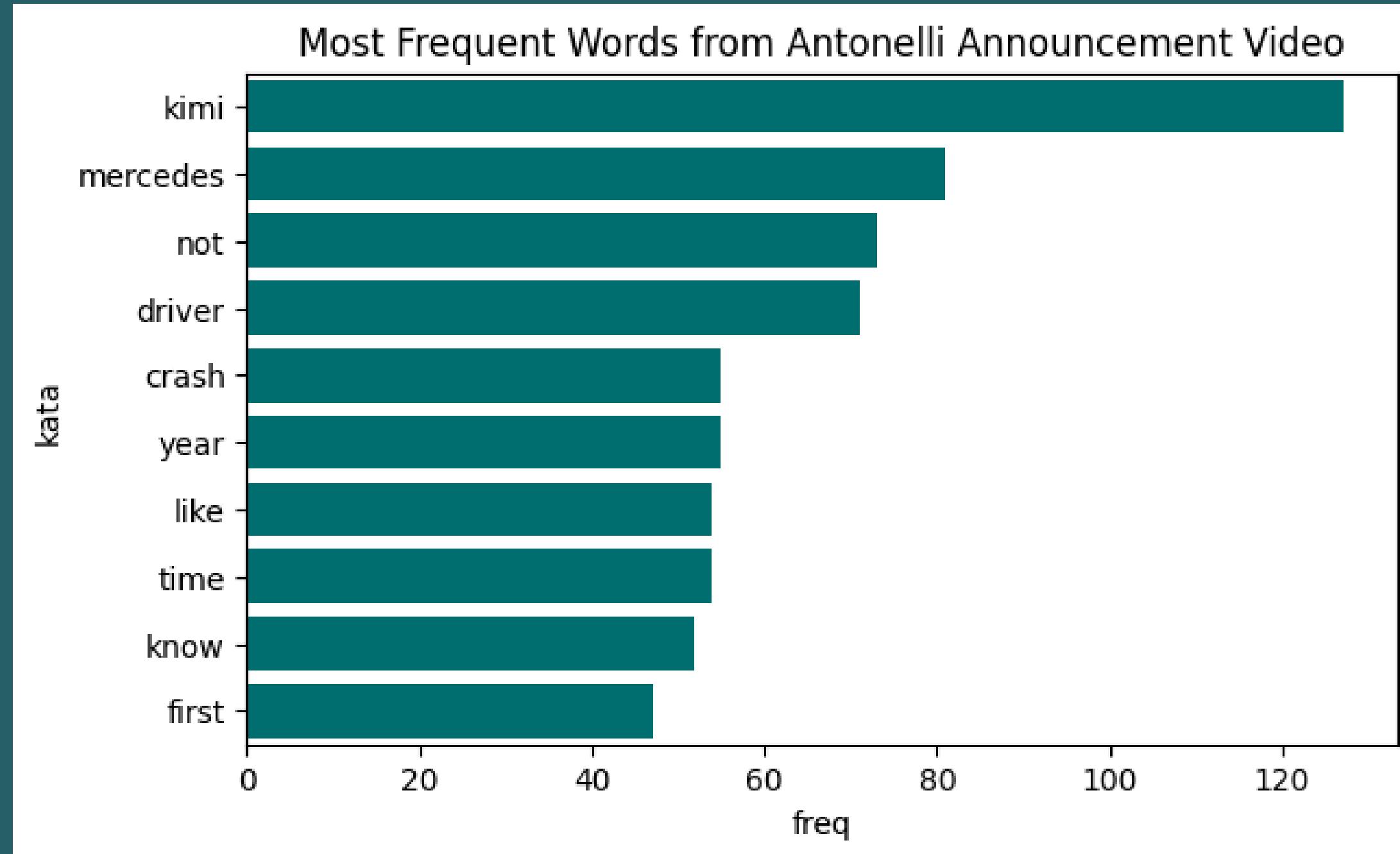
4. REMOVE SINGLE CHARACTERS

5. REMOVING EXTRA WHITESPACES

6. REMOVING STOPWORDS

7. LEMMATIZATION

Analyzing Word Frequencies



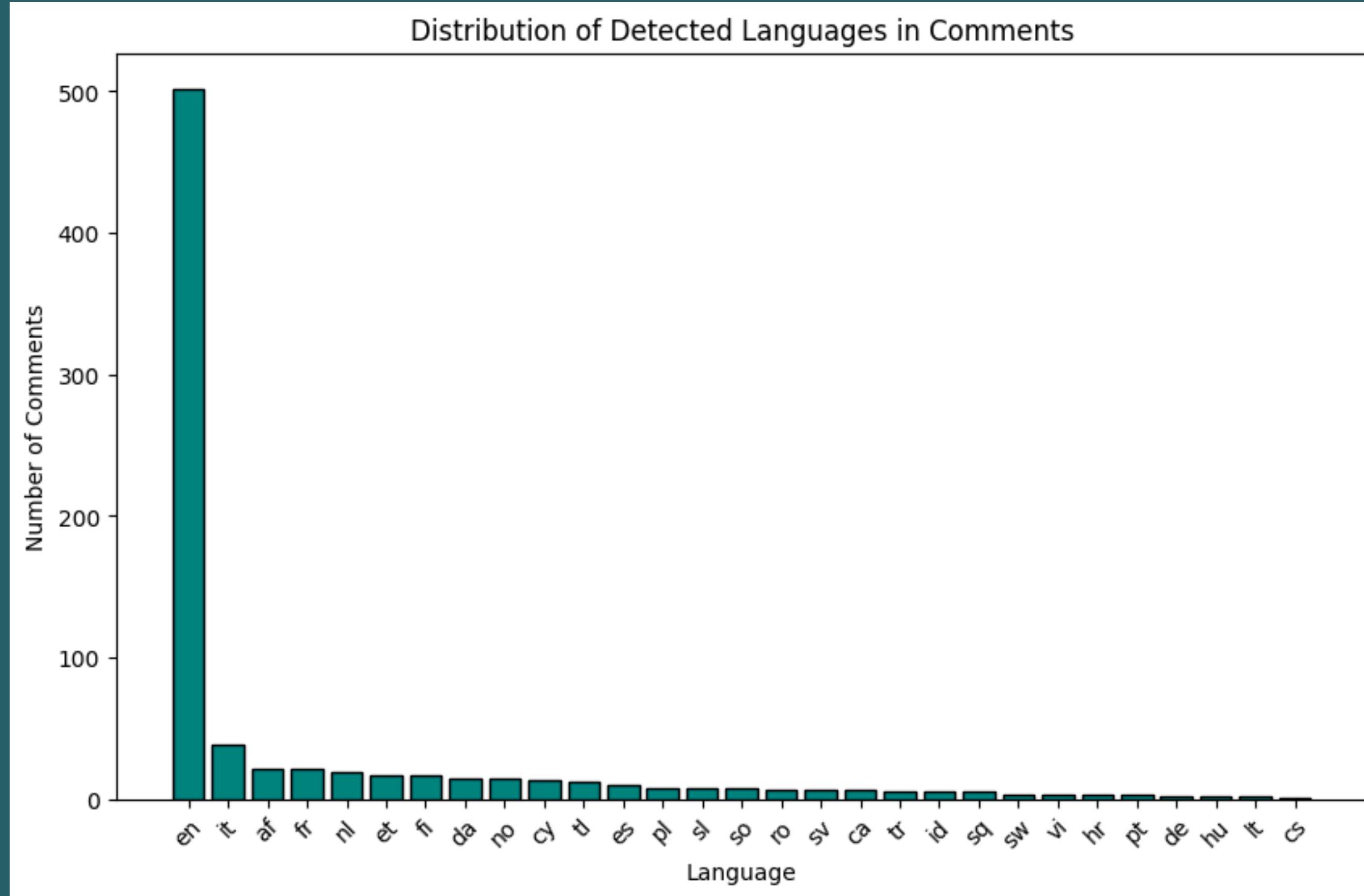
"Top words from fan comments on Kimi Antonelli's Mercedes announcement video show key themes like 'kimi,' 'mercedes,' and 'driver,' reflecting fan focus on his readiness and impact on the team."

Creating Word Cloud



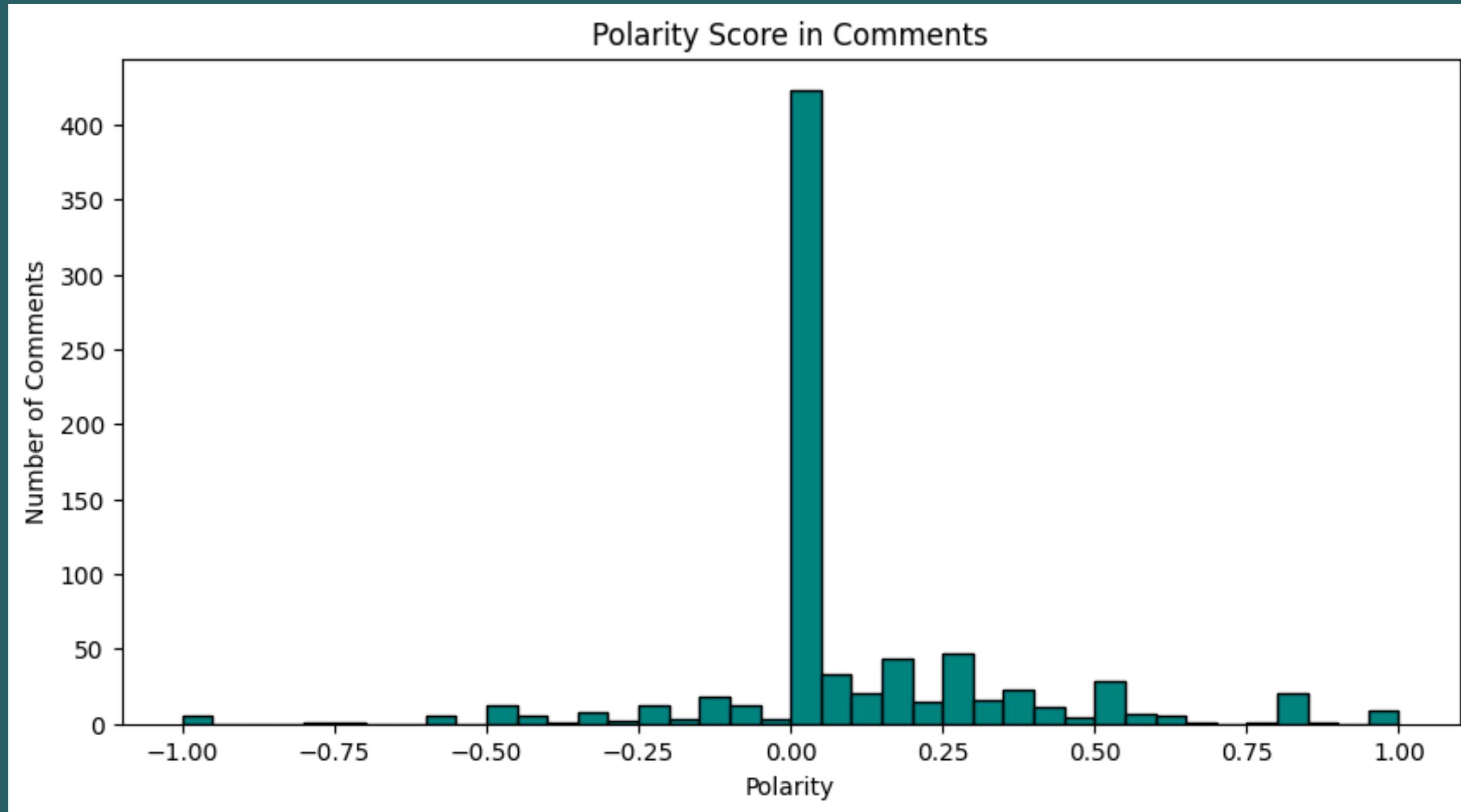
"The word cloud reveals key themes from comments, with words like 'kimi,' 'mercedes,' 'driver,' and 'time' reflecting discussions on Antonelli's readiness for F1 and his future at Mercedes."

Detecting and Visualizing Languages



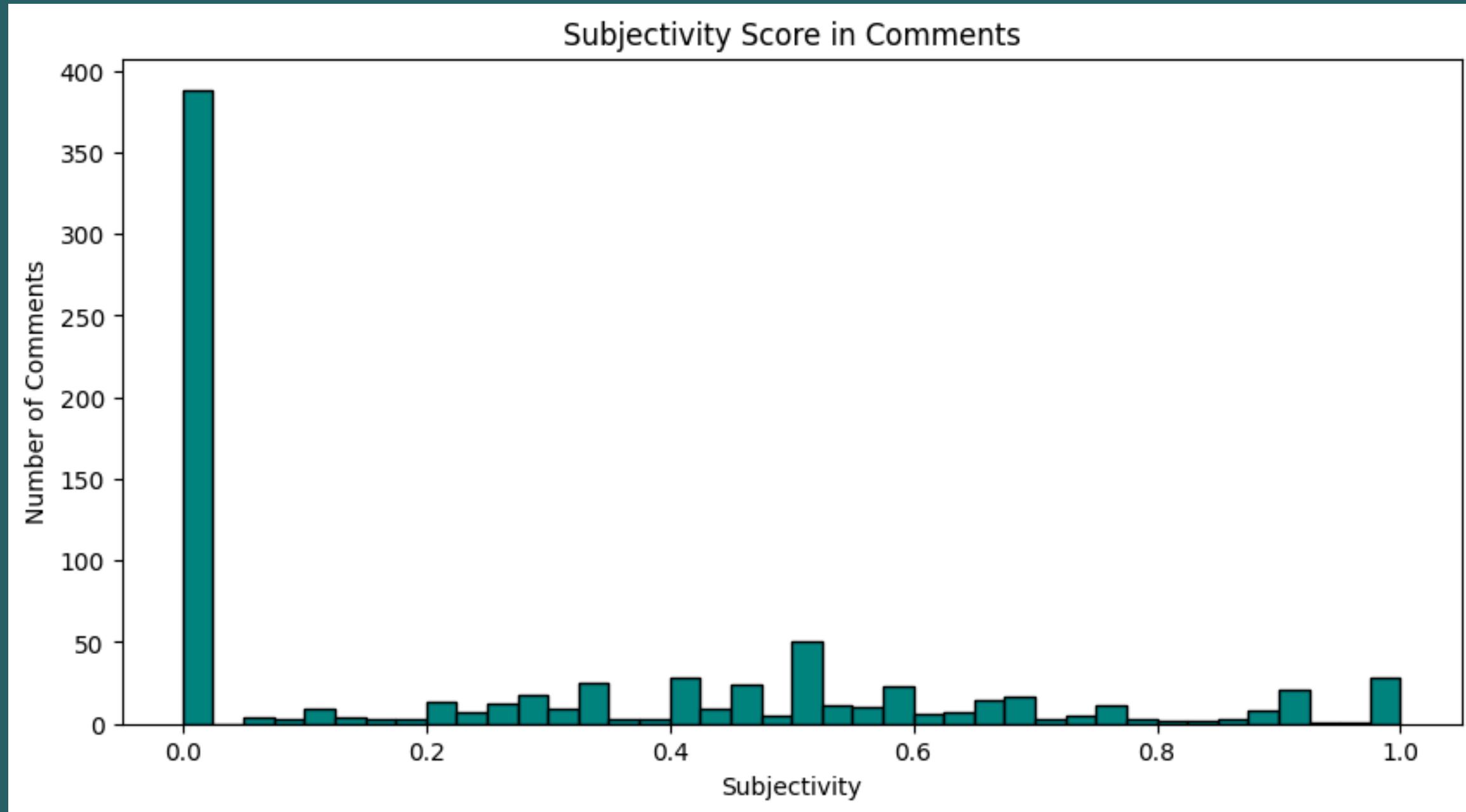
"The bar chart shows most comments were in English, with Italian as the second most common, reflecting local support for Kimi Antonelli."

Sentiment Analysis (Polarity Score)



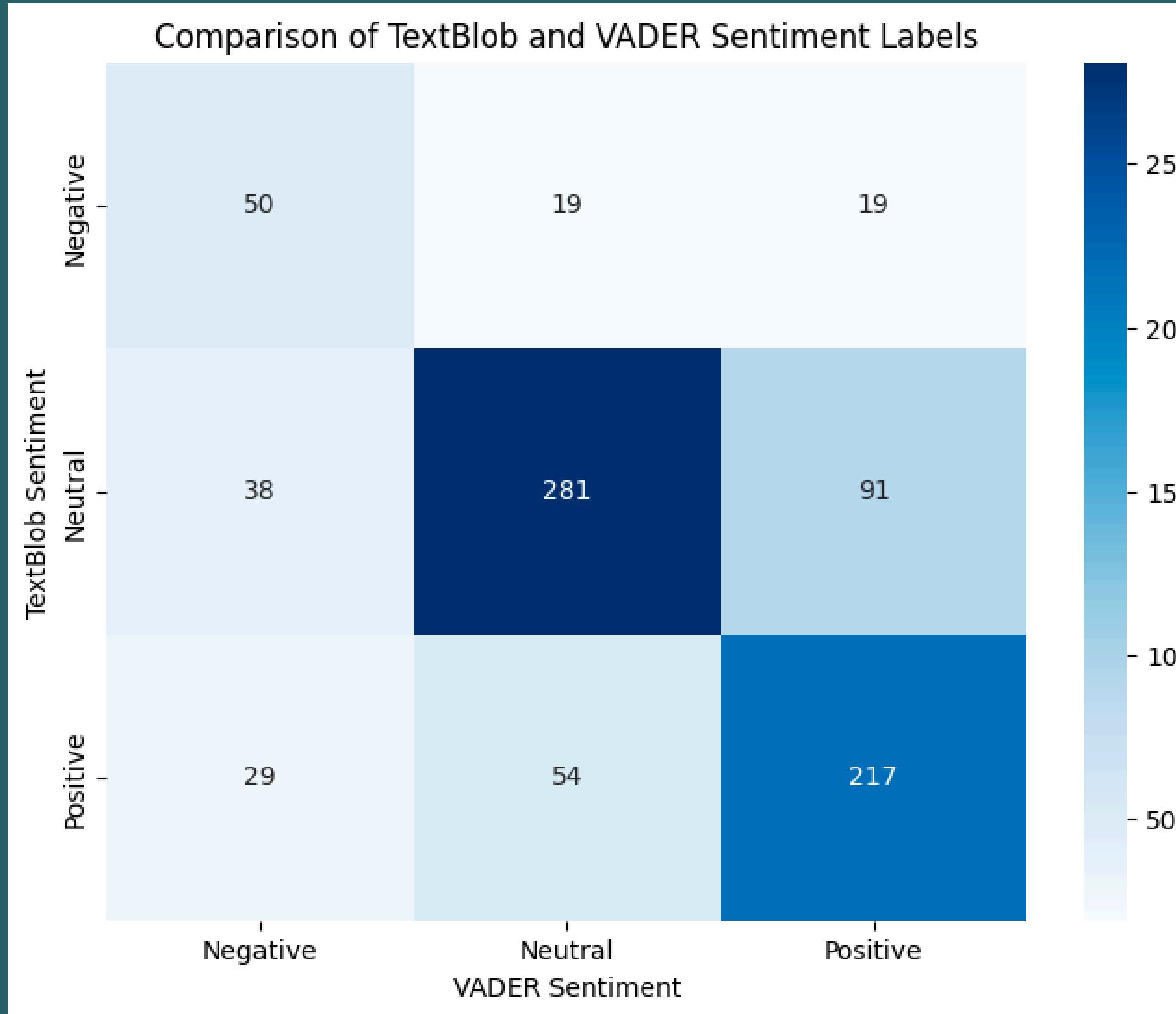
The histograms revealed that most comments are neutral in polarity, with a few being clearly positive or negative

Sentiment Analysis (Subjectivity Score)



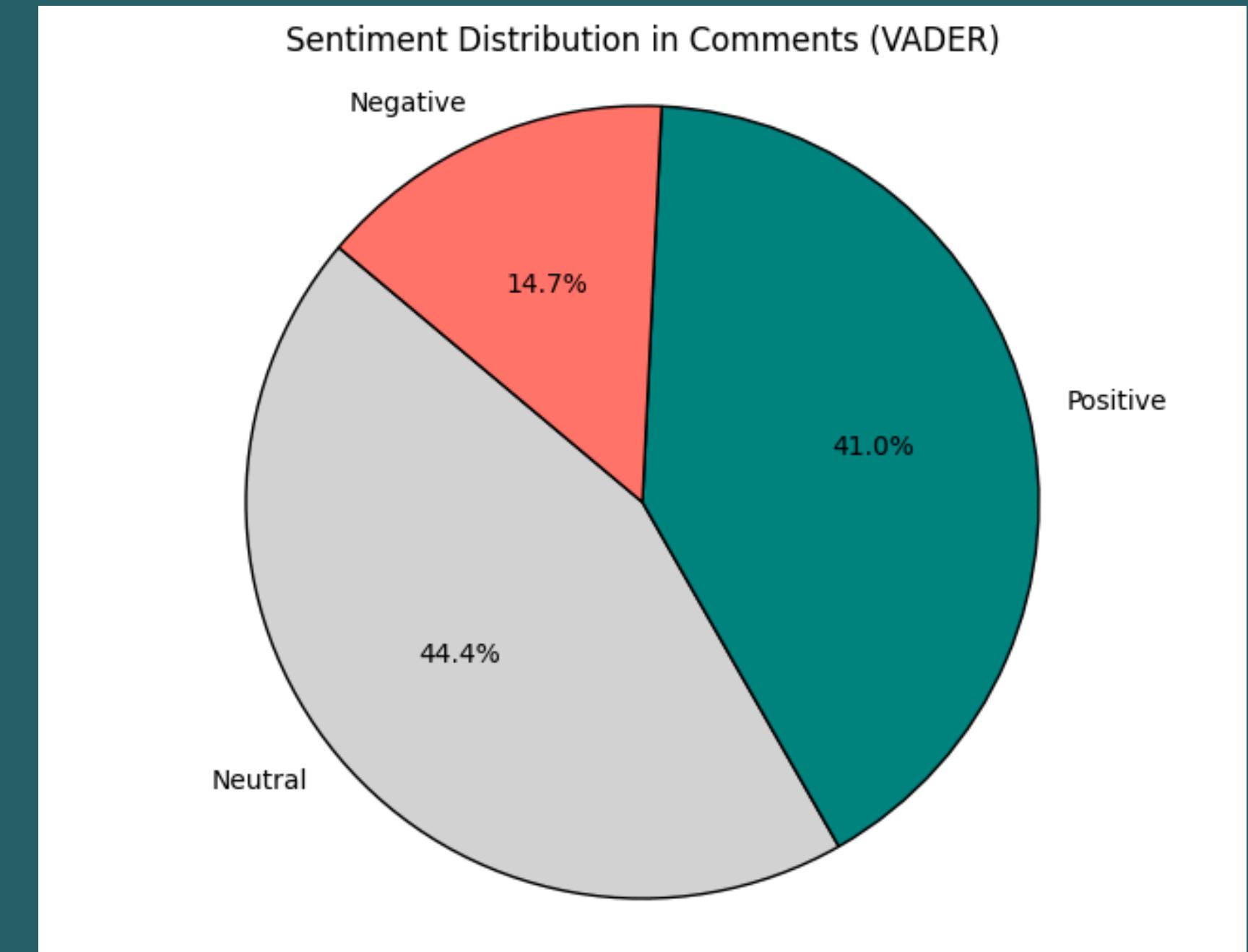
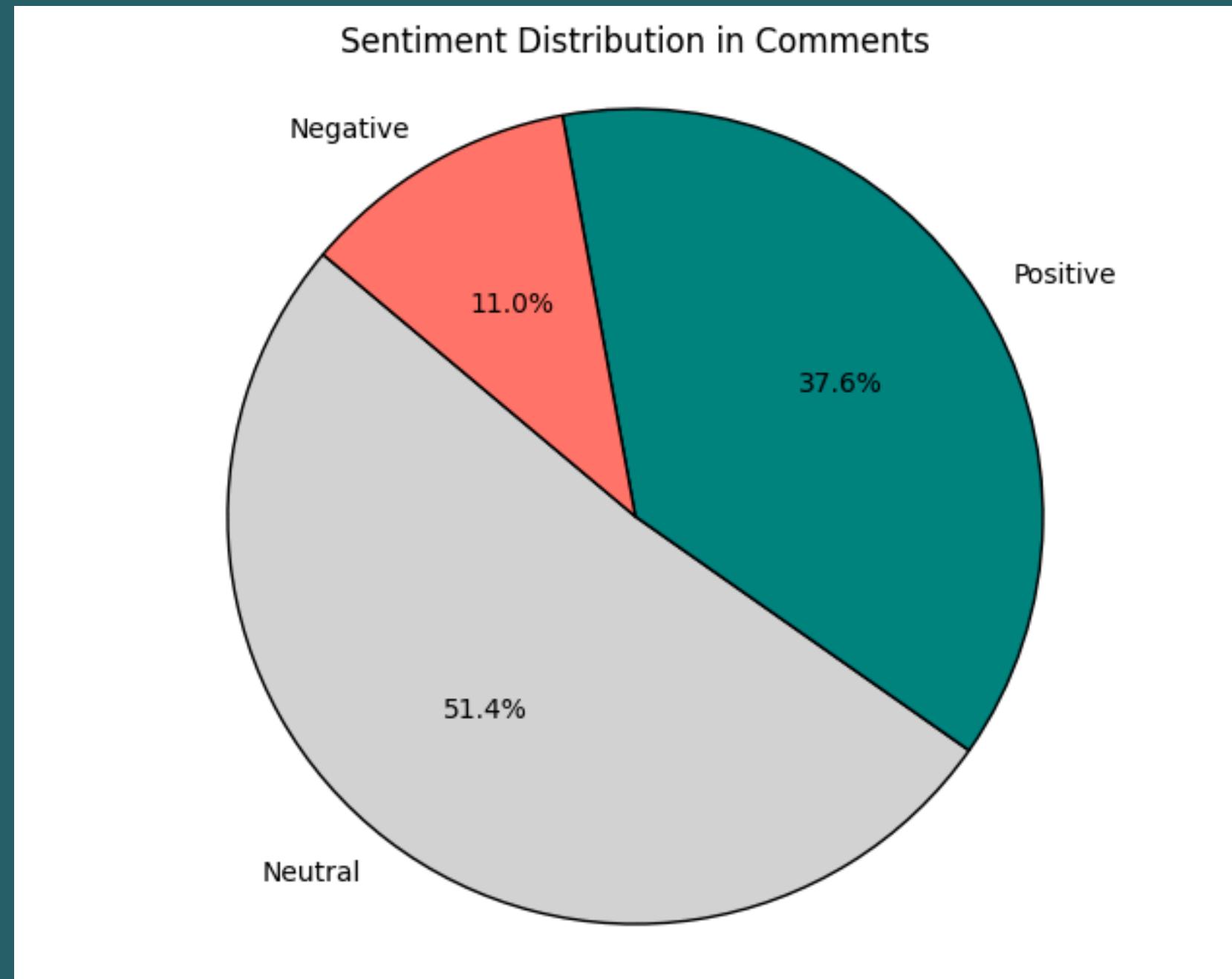
Similarly, most comments were found to be objective rather than subjective.

Comparison of TextBlob and VADER Sentiment Labels

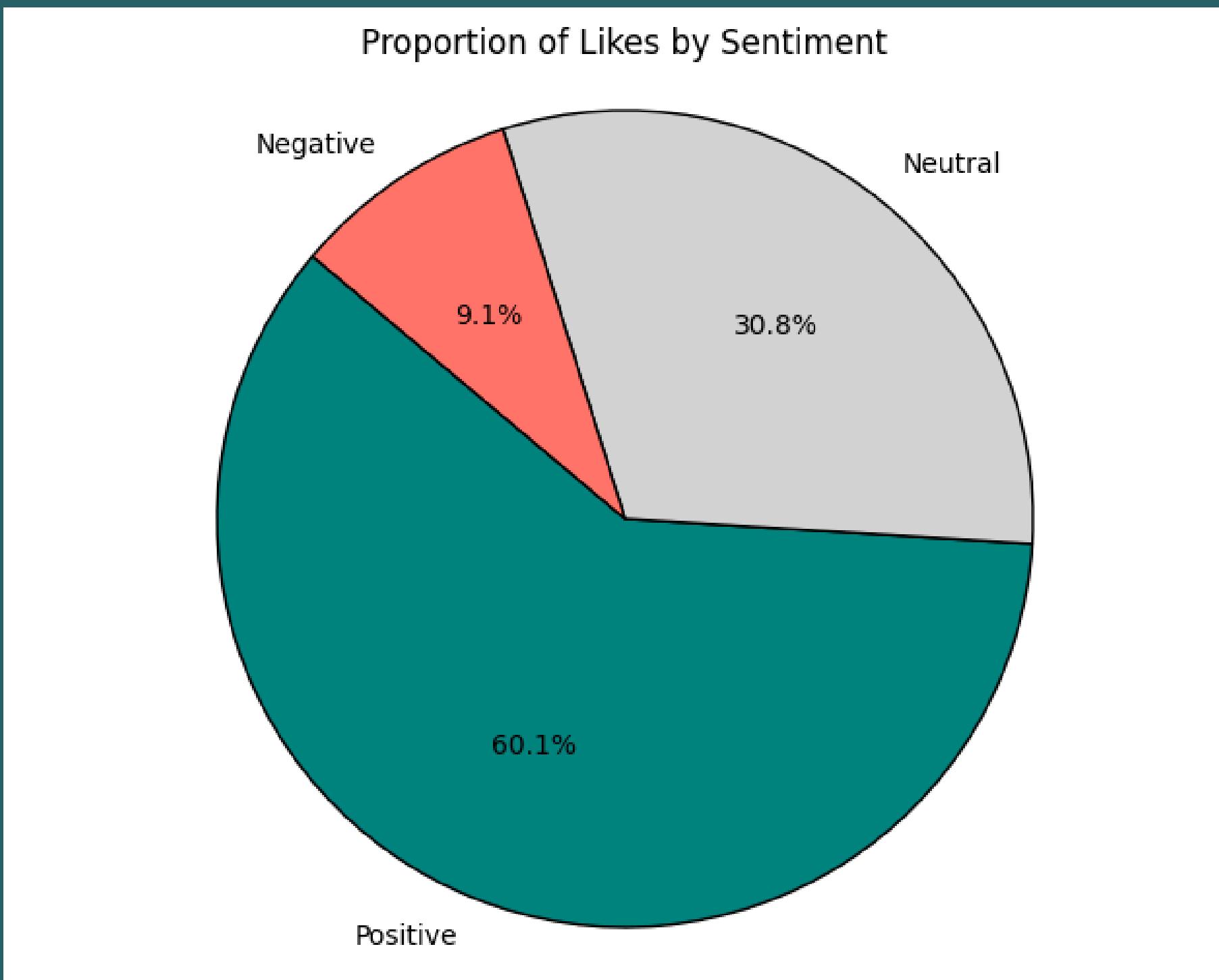


- **High Agreement in Neutral Sentiment:** TextBlob and VADER both classify a significant number of comments as neutral (281), showing consistency between the tools.
- **Differences in Positive and Negative Sentiment:** There are differences in how the tools label positive and negative comments. For example, both agree on 217 positive comments, but there are notable discrepancies in classifying others as neutral or negative.

Comparison of TextBlob and VADER Sentiment Labels



Analyzing Likes by Sentiment



Proportion of Likes by Sentiment:

Positive comments received the most likes (60.1%), indicating that optimistic or supportive sentiments resonate more with the audience. Neutral comments accounted for 30.8% of likes, while negative sentiments received the least engagement (9.1%). This shows a preference for positivity among viewers.

Word Cloud for Positive Sentiment



Word Cloud for Neutral Sentiment

Word Cloud for Negative Sentiment

Conclusion

- **Neutral Sentiments Dominate:** Most comments are neutral, showing a balanced fan reaction. Positive sentiments suggest support for Antonelli, while negative sentiments indicate some skepticism.
- **TextBlob vs. VADER:** Both tools agree on neutral sentiments but differ in positive/negative classifications, offering complementary insights.
- **Engagement Insights:** Positive comments get more likes, highlighting fan optimism and engagement.
- **Global Reach:** English and Italian dominate, reflecting Formula 1's global fanbase.
- **Fan Themes:** Word clouds show fans discussing excitement, skepticism, and team performance.