

```
In [1]: !pip install spacy==3.0.6

import spacy
from spacy.cli.download import download as spacy_download

spacy_download("en_core_web_sm")
nlp = spacy.load("en_core_web_sm")
```

```
Requirement already satisfied: spacy==3.0.6 in /Users/alexdoyle/opt/anacond
a3/lib/python3.9/site-packages (3.0.6)
Requirement already satisfied: packaging>=20.0 in /Users/alexdoyle/opt/anac
onda3/lib/python3.9/site-packages (from spacy==3.0.6) (24.0)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in /Users/alexdoyle/opt/a
naconda3/lib/python3.9/site-packages (from spacy==3.0.6) (0.7.9)
Requirement already satisfied: srsly<3.0.0,>=2.4.1 in /Users/alexdoyle/opt/
anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (2.4.6)
Requirement already satisfied: pydantic<1.8.0,>=1.7.1 in /Users/alexdoyle/o
pt/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (1.7.4)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /Users/alexdoyle
/opt/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (1.0.9)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /Users/alexdoyle/opt/
anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (2.0.7)
Requirement already satisfied: numpy>=1.15.0 in /Users/alexdoyle/opt/anacon
da3/lib/python3.9/site-packages (from spacy==3.0.6) (1.21.5)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /Users/alexdoyle/op
t/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (3.0.8)
Requirement already satisfied: setuptools in /Users/alexdoyle/opt/anaconda
3/lib/python3.9/site-packages (from spacy==3.0.6) (63.4.1)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/alexdoyle/
opt/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (2.31.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.4 in /Users/alexdoyle
/opt/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (3.0.12)
Requirement already satisfied: wasabi<1.1.0,>=0.8.1 in /Users/alexdoyle/op
t/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (0.10.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.3 in /Users/alexdoyle/
opt/anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (2.0.8)
Requirement already satisfied: typer<0.4.0,>=0.3.0 in /Users/alexdoyle/opt/
anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (0.3.2)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /Users/alexdoyle/opt/
anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (4.64.1)
Requirement already satisfied: pathy>=0.3.5 in /Users/alexdoyle/opt/anacond
a3/lib/python3.9/site-packages (from spacy==3.0.6) (0.10.1)
Requirement already satisfied: jinja2 in /Users/alexdoyle/opt/anaconda3/li
b/python3.9/site-packages (from spacy==3.0.6) (2.11.3)
Requirement already satisfied: thinc<8.1.0,>=8.0.3 in /Users/alexdoyle/opt/
anaconda3/lib/python3.9/site-packages (from spacy==3.0.6) (8.0.17)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /Users/alexdoyle
/opt/anaconda3/lib/python3.9/site-packages (from pathy>=0.3.5->spacy==3.0.
6) (5.2.1)
Requirement already satisfied: certifi>=2017.4.17 in /Users/alexdoyle/opt/a
naconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy==
3.0.6) (2022.9.24)
Requirement already satisfied: idna<4,>=2.5 in /Users/alexdoyle/opt/anacond
a3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy==3.0.6)
(2.8)
Requirement already satisfied: charset-normalizer<4,>=2 in /Users/alexdoyle
/opt/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->
spacy==3.0.6) (2.0.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /Users/alexdoyle/opt/a
naconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy==
3.0.6) (1.24.3)
Requirement already satisfied: click<7.2.0,>=7.1.1 in /Users/alexdoyle/opt/a
naconda3/lib/python3.9/site-packages (from typer<0.4.0,>=0.3.0->spacy==3.
0.6) (7.1.2)
Requirement already satisfied: MarkupSafe>=0.23 in /Users/alexdoyle/opt/ana
conda3/lib/python3.9/site-packages (from jinja2->spacy==3.0.6) (2.0.1)
Collecting en-core-web-sm==3.0.0
  Downloading https://github.com/explosion/spacy-models/releases/download/e
n_core_web_sm-3.0.0/en_core_web_sm-3.0.0-py3-none-any.whl (13.7 MB)
  13.7/13.7 MB 9.6 MB/s eta 0:0
  0:00
Requirement already satisfied: spacy<3.1.0,>=3.0.0 in /Users/alexdoyle/opt/
```

```
anaconda3/lib/python3.9/site-packages (from en-core-web-sm==3.0.0) (3.0.6)
Requirement already satisfied: pydantic<1.8.0,>=1.7.1 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (1.7.4)
Requirement already satisfied: pathy>=0.3.5 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (0.10.1)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.31.0)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (0.7.9)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (3.0.8)
Requirement already satisfied: wasabi<1.1.0,>=0.8.1 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (0.10.1)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.0.7)
Requirement already satisfied: srslly<3.0.0,>=2.4.1 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.4.6)
Requirement already satisfied: catalogue<2.1.0,>=2.0.3 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.0.8)
Requirement already satisfied: jinja2 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.11.3)
Requirement already satisfied: typer<0.4.0,>=0.3.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (0.3.2)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (4.64.1)
Requirement already satisfied: numpy>=1.15.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (1.21.5)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.4 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (3.0.12)
Requirement already satisfied: setuptools in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (63.4.1)
Requirement already satisfied: packaging>=20.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (24.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (1.0.9)
Requirement already satisfied: thinc<8.1.0,>=8.0.3 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (8.0.17)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from pathy>=0.3.5->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (5.2.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (1.24.3)
Requirement already satisfied: certifi>=2017.4.17 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2022.9.24)
```

Requirement already satisfied: charset-normalizer<4,>=2 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.8)

Requirement already satisfied: click<7.2.0,>=7.1.1 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from typer<0.4.0,>=0.3.0->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (7.1.2)

Requirement already satisfied: MarkupSafe>=0.23 in /Users/alexdoyle/opt/anaconda3/lib/python3.9/site-packages (from jinja2->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (2.0.1)

✓ Download and installation successful

You can now load the package via spacy.load('en\_core\_web\_sm')

In [2]:

```
import os
import pandas as pd

folder_path = '/Users/alexdoyle/Desktop/LIS/Capstone/capstone_nlp/capstone_files'
files = [file for file in os.listdir(folder_path) if os.path.isfile(os.path.join(folder_path, file))]

data = []

for file in files:
    file_path = os.path.join(folder_path, file)
    encodings_to_try = ['utf-8', 'latin-1', 'ISO-8859-1']

    for encoding in encodings_to_try:
        try:
            with open(file_path, 'r', encoding=encoding) as f:
                content = f.read()
                # Process content if needed
                data.append({'filename': file, 'content': content})
                break # Stop trying encodings if successful
        except UnicodeDecodeError:
            pass # Try the next encoding if decoding fails

df = pd.DataFrame(data)
```

In [3]:

df

Out [3]:

	filename	content
0	cvctybLyEqE.mp4.srt	1\n00:00:00,177 --> 00:00:06,044\n>Welcome to...
1	oCl6MCoOi1Q.mp4.srt	1\n00:00:00,484 --> 00:00:02,841\nHello and w...
2	_kDc62-upoA.mp4.srt	1\n00:00:00,389 --> 00:00:03,488\n-The league...
3	41kO2TvmNbg.mp4.srt	1\n00:00:00,500 --> 00:00:03,566\nWelcome to ...
4	DjNA-MIdCbY.mp4.srt	1\n00:00:00,000 --> 00:00:04,064\n[music]\n\n...
...	...	...
1118	dJMNNQHkz0M.mp4.srt	1\n00:00:02,068 --> 00:00:04,994\nMatch begin...
1119	xK5R8h7zln0.mp4.srt	1\n00:00:08,885 --> 00:00:11,055\n[applause]\n...
1120	qG1IIIj9_fU.mp4.srt	1\n00:00:06,913 --> 00:00:12,072\n-Played by ...
1121	GsY4w7sjh9A.mp4.srt	1\n00:00:00,171 --> 00:00:04,057\n[music]\n\n...
1122	R0XfelUhG_Y.mp4.srt	1\n00:00:00,429 --> 00:00:03,888\nHello, and ...

1123 rows × 2 columns

```
In [4]: df['content']
```

```
Out[4]: 0      1\n00:00:00,177 --> 00:00:06,044\n-Welcome to...
1      1\n00:00:00,484 --> 00:00:02,841\nHello and w...
2      1\n00:00:00,389 --> 00:00:03,488\n-The league...
3      1\n00:00:00,500 --> 00:00:03,566\nWelcome to ...
4      1\n00:00:00,000 --> 00:00:04,064\n[music]\n\n...
...
1118    1\n00:00:02,068 --> 00:00:04,994\nMatch begin...
1119    1\n00:00:08,885 --> 00:00:11,055\n[applause]...
1120    1\n00:00:06,913 --> 00:00:12,072\n-Played by ...
1121    1\n00:00:00,171 --> 00:00:04,057\n[music]\n\n...
1122    1\n00:00:00,429 --> 00:00:03,888\nHello, and ...

Name: content, Length: 1123, dtype: object
```

```
In [5]: import pandas as pd
```

```
# Set max_colwidth to None to display entire content
pd.set_option('display.max_colwidth', None)

# Define a function to remove non-letter characters
def remove_non_letters(text):
    return ''.join(char for char in text if char.isalpha() or char.isspace())

# Apply the function to the 'content' column
df['cleaned_content'] = df['content'].apply(remove_non_letters)

# Display the DataFrame with cleaned content
df[['filename', 'cleaned_content']]
```

Out[5]:

filename	cleaned_content
0 cvctybLyEqE.mp4.srt	<p>\n \nWelcome to a rainy Stadio Olimpico\n Lazio against Juventus\n \n\n\n \nshould be a real atmosphere this \n\n\n \nCan Lazio become the first team \nthis Serie A season to beat Juventus\n \n\n\n \nI dont think theyre going to stop\n their irrepressible march to the title \n\n\n \nparticularly after Napoli could only\n \ndraw against Milan last night\n\n \n Marco Guida gets us underway\n\n \nReferee got in the way there \n\n\n \nHere goes Dybala \nLucas once more wins it back \n\n\n \nNow Immobile opens up for Correa \n\n\n \nSzczęsny got behind it on an \nawkward surface and held it well\n\n \nJuve have a tricky game \nagainst Atalanta in the Cup \n\n\n \nTricky moments here as \nwell as it falls for Luiz Alberto\n\n \n Szczęsny with the save\nand Lucas with a miss hit \n\n\n \nHeres Correa \n\n\n \nmisskick there from Matuidi \nMilinkovićSavić\n\n \nLook at the space for Parolo\nFabulous save from Szczęsny \n\n\n \nwho has kept Juve in it here\nThats the pick of them\n\n \n\nThe effort levels never\n drop when hes out there \n\n\n \nIt might bounce here for Immobile \n\n\n \nRugani does so well to recover\n and get it behind\n\n \nGreat attacking instinct from Immobile \n\n\n \n\nMuch to discuss much to find out \nNo changes at half time\n\n \nThey need something\n to show for their efforts \n\n\n \nIn from Luiz Alberto and they have it \n\n\n \nDid it ricochet off MilinkovićSavić \n\n\n \nIs it an owngoal\n No one at the Olmpico could care\n\n \nLazio in front Juve in trouble\n in the rain in Rome\n\n \nNow Correa \n\n\n \nOh Immobile is in with a chance to make it two\n\n \nForward by Alex Sandro \n\n\n \nOh great skill from Bernardeschi\n For Dybala spilled big chance \n\n\n \nJoão Cancelo with an equalizer for Juventus \n\n\n \nBack from injury to save them \n\n\n \nThe substitute makes it \n\n\n \nHold on to this or even better it \nand they might just do that \n\n\n \nRonaldo in for Bernardeschi \n\n\n \nJoão Cancelo went over on\nthe challenge from Lulić \n\n\n \nHes given a penalty \n\n\n \nMarco Guida has given the penalty \n\n\n \nThey were under massive pressure \nCan they turn it round They can\n\n \nIts what champions do on the ropes\n\n \nSergej MilinkovićSavić over the bar \n\n\n \nJuventus can celebrate\n can punch the air\n\n \nJuve win it by two goals to one\n\n</p>
1 oCI6MCoOi1Q.mp4.srt	<p>\n \nHello and welcome\n to the Stadio Bentegodi\n\n \n\n\n \nnwhere this afternoon\nChievo Verona and Frosinone\n\n \n\n\n \ngo headtohead\nin round of Serie A\n\n \n\n\n \nlm Marco Palmieri\nGreat to be with you\n\n \n\n\n \nIts a mustwin match for both sides\n\n \n\n\n \nWere away at the Bentegodi\nbetween Chievo and Frosinone\n\n \n\n\n \n\nThe corners are coming\n\n \n\n\n \ncan they take advantage\n\n \n\n\n \n\nThe big men in the area\nas Birsa steps up to take the corner\n\n \n\n\n \nIt is indeed\n\n \n\n\n \nSergio Pellissier\nwith a header from the corner\n\n \n\n\n \nand the captain does it again\n\n\n \nfor Chievo\n\n \n\n\n \n\nMaking sure\n\n \n\n\n \nthat there werent any irregularities\nin the box\n\n \n\n\n \nmaybe a push\n\n \n\n\n \nIt will be a onfield review\n\n \n\n\n \n\nThe goal has been disallowed\n\n \n\n\n \nIts still\n\n \nhere at the Bentegodi\n\n \n\n\n \nPellissier is furious\n\n \n\n\n \nA push\n\n \n\n\n \nlm assuming\n\n \nthere was a push by Pellissier\n\n \n\n\n \nas he lept up into the air\n\n \n\n\n \nAs Rocchi gets set to start\nthe second half\n\n \n\n\n \nand were away\n\n \n\n\n \nAs Hetemaj gets it to Birsa\n\n \n\n\n \nHe skips past Maiello\n\n \n\n\n \nNow Đorđević\n\n \n\n\n \nthe first shot on target\n\n \n\n\n \nbut Sportiello denies him\nwith a quick save\n\n \n\n\n \nnot necessarily a difficult one\n\n \n\n\n \nbut one that had to be made\n\n \n\n\n \nas the ball found its way\ninto the Serb striker\n\n \n\n\n \nCorner taken short\n\n \n\n\n \nAs the ball at the far post\n\n \n\n\n \nfor Chibsah\n\n \n\n\n \nBrilliant save from Sorrentino\n\n \n\n\n \n\nThe Frosinone players thought\nthat itd crossed the line\n\n \n\n\n \n\nThe watch on referee Rocchi\n\n \ndid not signal a goal\n\n \n\n\n \nand play continues\n\n \n\n\n \nAs now Depaoli is in space\n\n \n\n\n \nand hes tripped\n\n \non the edge of the area here\n\n \n\n\n \nThere will be a foul and a yellow\n\n \n\n\n \nIts a second yellow\nCapuano is off\n\n \n\n\n \nFrosinone reduced to men\n\n \n\n\n \nHetemaj sees the overlap</p>

filename	cleaned_content
	<p>now\n\n\n \nPlay inside Pellissier\n\n\n \nOn his left foot\nhe slipped over as he took the shot\n\n\n \nGreat chance for Chievo\n\n\n \nas they continue to attack now\n\n\n \nTomović spreads it across\nnto Radovanović\n\n\n \nNow Bani plays it in the box\n\n\n \nbut Depaoli cannot reach it\n\n\n \nand the ball dribbles back\nnto the keeper\n\n\n \nWhat a chance for Pellissier here\n\n\n \nThe shot comes in\nand thats a wonderful goal\n\n\n \nGiaccherini gives Chievo the lead\nnon minutes\n\n\n \nwith a superb freekick\n\n\n \nSportiello rooted to the spot\n\n\n \nAn excellent goal\nand thoroughly deserved\n\n\n \nThere it is the final whistle\n\n\n \nA simply perfect\nGiaccherini freekick\n\n\n \nhas given Chievo\ntheir first win of the season\n\n\n \nin the last match of the year\n\n\n \nTheyve defeated a man Frosinone\nby a one goal to nil\n\n\n</p>
2 _kDc62-upoA.mp4.srt	<p>\n \nThe leagues top scorers\nagainst Seria As weakest frontline\n\n\n \nAtalanta are bidding to make \nthe best year in club history\n\n\n \nSpal are four points from safety\n\n\n \nand running out of time\nin their efforts to beat the drop\n\n\n \nDi Francesco in the channel\n\n\n \nable to wander inside\n\n\n \nGabriel Strefezza lining it up\n\n\n \nlts a good save from Sportiello\n\n\n \n\nThe game becoming stretched\njust like Atalanta want it to be\n\n\n \nZapata running at Thiago Cionek\n\n\n \nDuván Zapatas cross\nnts in for Ilicic\n\n\n \nA difficult finish\nmade to look all so easy\n\n\n \nZapata still seeking his first goal\nnsince his return to action\n\n\n \nThat was a telling contribution\n\n\n \nIlicic has scored\nagainst Fiorentina in midweek\n\n\n \ncontinues his\nwonderful start to \n\n\n \nWell in by Gosens\n\n\n \nThis is Pasalic\n\n\n \nplaying slightly deeper this evening\n\n\n \nNow Zapata\n\n\n \nplenty of room for him to get moving\n\n\n \nUp against Vicari\n\n\n \nDuván Zapata with the shot\n\n\n \nlts off the frame of the goal\n\n\n \nThis is Valoti\n\n\n \nThiago Cionek joining in\n\n\n \nThe cross is only halfcleared\n\n\n \nStrafezza\n\n\n \nCionek with the ball in\n\n\n \nlts off the post\n\n\n \nSo close to yet another goal\nagainst his former club\n\n\n \nThats a nice touch\nfrom Di Francesco\n\n\n \nReca is very quick\n\n\n \nHes trying to get\naway from Djimsiti\n\n\n \nThis is Recas cross\n\n\n \nand its put in by Petagna\n\n\n \ncouple of former\nAtalanta players combine\n\n\n \nReca on loan\n\n\n \nPetagna once again\nscores against La Dea\n\n\n \nNo one could catch up with the ball\n\n\n \nDjimsiti\n\n\n \nZapata a foul on him Vicari\n\n\n \nDabo\n\n\n \nValoti\n\n\n \nHes kept it\nMatti Valoti\n\n\n \nSpal have another\n\n\n \nTheyve come from behind here\nlead in Bergamo\n\n\n \nWhat a superb individual strike\n\n\n \nPetagna first and now Valoti\n\n\n \nHe took his goal splendidly well\n\n\n \nLaunched long by Ilicic\n\n\n \nMuriel gives chase hes got there\n\n\n \nThis is Zapata\n\n\n \nDuvan Zapata its deflected\n\n\n \nTerrific save from Berisha\n\n\n \nThat was creeping in\n\n\n \nAnother look\nat that long ball from Ilicic\n\n\n \nMuriel did well\n\n\n \nZapata struck it firmly\n\n\n \nSpal win in Bergamo\n\n\n \nWhat a turn around for Semplici\n\n\n \nonly a fourth victory of the season\nfor Seria As bottom side\n\n\n \nThree massive points in their quest\nnto retain their topflight status\n\n\n \nAtalanta offcolor\n\n\n \nand out of sorts they stay fifth\n\n\n \ncrowd cheering\n\n\n \ncrowd cheering\n\n\n</p>
3 41kO2TvmNbg.mp4.srt	<p>\n \nWelcome to the Ennio Tardini stadium \n\n\n \nThis week \n\n\n \nlts the meeting of the two sides \n\n\n \nwho are the bane of the City of Milan \n\n\n \nin the last match week \n\n\n \nParma against Cagliari \n\n\n \nLast weekend \n\n\n \nParma recorded a surprise victory \n\n\n \nat San Siro against Inter \n\n\n \nwhile Cagliari got an impressive draw \n\n\n \nagainst Milan in Sardinia \n\n\n \nWere underway \n\n\n \nStolen and now Gervinho \n\n\n \ncan show us that famous sprinting power \n\n\n \nStill going \n\n\n \nGervinho still going \n\n\n \ngets the shot off \n\n\n \njust wide of the post \n\n\n \ntrademark run from Gervinho\n\n\n</p>

filename	cleaned_content
	<p>\nLook where he picks the ball up in his own half \n\n\n\nThen he goes on an angled run inside \n\n\n\n gets past Klavan \n\n\n\n sees off Ioniță \n\n\n\n and a low shot \n\n\n\n You can see what he was going for there \n\n\n\n \nDimarco \n\n\n\n \nInglese is in there again \n\n\n\n \nInglese loops it over the keeper \n\n\n\n \nCragno \n\n\n\n \nThis time \n\n\n\n \nhas the goal been given \n\n\n\n \nIt looks like it has \n\n\n\n \nDenied by the offside flag a few minutes ago\ncan't find the file name</p>

4	DjNA-MIdCbY.mp4.srt	<p>\n \nmusic\n\n\n\nChilwell\n\n\n\nOh and its fired in in \n\n\n\n Typical Oxlade Chamberlain style\n\n\n\n He is a goal threat\n\n\n\n He really didnt catch it well\n\n\n\n Its another lovely ball\n\n\n\n Its not the easiest of chances\n\n\n\n When you strike it that purely\n\n\n\n theres no chance for the keeper\n\n\n\n Theyre stretched defensively here\n\n\n\n Harry Kane just makes that point\n\n\n\n Only needed a touch\n\n\n\n Floated in by Chilwell\n\n\n\n namongst a clutch of players\n\n\n\n but he manages just to\n\n\n\n keep his opponents at arms length\n\n\n\n Chilwell to take the corner\n\n\n\n Thats a free header Thats another goal\n\n\n\n Simple for Kane\n\n\n\n England incidentally are not\n\n\n\n showing their array of set pieces\n\n\n\n they want to sail to the finals\n\n\n\n On the play\n\n\n\n nobody knows what those Montenegro\n\n\n\n defenders were doing here\n\n\n\n to allow Kane a free run\n\n\n\n Theyve got to get a bit more\n\n\n\n physical contact and so easy\n\n\n\n Rashford gets in on the act\n\n\n\n Well thats four\n\n\n\n Oxlade Chamberlain with a really good ball\n\n\n\n Maguire cant believe he hasnt scored there\n\n\n\n Rashford tidies up\n\n\n\n Sancho\n\n\n\n Heres the hattrick\n\n\n\n Another Harry hattrick\n\n\n\n Five for England\n\n\n\n This landmark evening\n\n\n\n could not be going better\n\n\n\n It wasnt the easiest ball to control\n\n\n\n first time from AlexanderArnold\n\n\n\n just bouncing in front of him laugh\n\n\n\n Hes such a confident frame of mind\n\n\n\n that he just decides to take a touch\n\n\n\n It proves to be the right decision\n\n\n\n Rashfords in behind\n\n\n\n Wow how has it not gone in\n\n\n\n In the end it has\n\n\n\n It looked like the defender rushing back\n\n\n\n I see he got the final touch\n\n\n\n Rashfords done this several times tonight\n\n\n\n just easing past his man\n\n\n\n An instinctive reaction by Sofranac\n\n\n\n He might get a tap in this time\n\n\n\n Tammy Abraham opens his England account\n\n\n\n nit is seven with seven minutes to go\n\n\n\n Yes a big moment for him and a classic\n\n\n\n center forwards goal this one\n\n\n\n Making that dash across your man\n\n\n\n sticking out</p>
---	---------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

filename	cleaned_content
	<p>a leg\n\n\n \nWonderful stuff from Abraham\n\n\n \nIn an ideal world\n\n\n \nhe could get on the score sheet\n\n\n \nget him some belief at this level\n\n\n \nIt finishes England seven\n\n\n \nMontenegro nil\n\n\n \nThey qualify for the finals\n\n\n \nmusic\n\n</p> <p>...</p> <p>...</p> <p>...</p>
1118 dJMNNQHkz0M.mp4.srt	<p>\n \nMatch begins with Torino very much all won \n\n\n \nfive consecutive clean sheets victories\n over Inter and last week Atalanta \n\n\n \nToday begins a trio of fixtures\n against the relegation side \n\n\n \nChievo are expected to be beaten\n\n\n \nThe first whistle goes \n\n\n \nIt is Chievo in the yellow\n who kick us off \n\n\n \nThe cross is in\n Whipped in to De Silvestri \n\n\n \n\nThe shot from range\n Ansaldi converts\n\n\n \ncheering\n\n\n \ntwo goals for him this season\n the Argentinian\n\n\n \ncheering\n\n\n \nNow a round of applause and a pause\n for the memory of Davide Astori \n\n\n \nthe former Fiorentina player\n who so tragically lost his life last year\n\n\n \napplause\n\n\n \ncheering\n\n\n \nThis is better now\n\n\n \nAnsaldi Meité who wasnt wise enough\n to trick Diousse\n\n\n \ncheering\n\n\n \nMeggiorini Đorđević to his right\n He has a go at goal \n\n\n \nA goal that he used to shoot at himself\n when wearing the colors of Torino \n\n\n \nPlayed for them between \n\n\n \ncheering\n\n\n \nWere back underway\n Torino kick us off\n\n\n \ncheering\n\n\n \nDepaoli shifting inside\n Trying to make the angle for the cross\n\n\n \nTook a deflection\n Should score Sirigu denies \n\n\n \nĐorđević with the best chance\n of the game\n\n\n \ncheering\n\n\n \nHe keeps that clean sheet record \n\n\n \nit was an excellent reaction\n save close to him \n\n\n \nĐorđević with the second attempt\n Simply had to score\n\n\n \ncheering\n\n\n \nwhistle blows\n\n\n \nNow this could be the moment\n Zaza \n\n\n \nSorrentino wonderful save\n\n\n \ncheering\n\n\n \nBelotti with the strike\n and Belotti with the missile \n\n\n \nTorino take the lead \n\n\n \nWhat a vital goal\n Its the captains goal\n\n\n \ncheering\n\n\n \nNumber eight of the season\n Surely the best of the lots \n\n\n \nA cracker from Andrea Belotti\n\n\n \ncheering\n\n\n \nNow looking to just finish off\n this match \n\n\n \nTomás Rincón Tomás Rincón\n\n\n \ncheering\n\n\n \nThats a thing of beauty\n\n\n \ncheering\n\n\n \nHe comes off the bench \n\n\n \nHe was supposed to put solidity\n into the midfield \n\n\n \nIn the end hes slid\n a beautiful strike into the top corner \n\n\n \nTorino two\n Chievo nil in injury time\n\n\n \ncheering\n\n\n \nZaza looks for number three\n and in the net it goes \n\n\n \ntwo goals in injury time \n\n\n \nIts Torino three Chievo nil \n\n\n \nZaza scores home and away\n against Chievo this season \n\n\n \nHes had to wait since the end of September\n for that second goal\n\n\n \nThe final whistle goes \n\n\n \nit finishes here Torino three\n Chievo nil\n\n\n \ncheering</p>
1119 xK5R8h7zln0.mp4.srt	<p>\n \napplause\nHere is the handball Michael Oliver\n\n\n \nwaves it away\n\n\n \nHere is VAR trying to give us\nan alternative version of it\n\n\n \nThen goes on with Mane\n\n\n \nStones cant stop the cross\n\n\n \nSalah is coming in\n\n\n \nFabinho has a go and what a goal\n\n\n \napplause\nWhat a goal\n\n\n \nwhat a start\n\n\n \nfor Liverpool\n\n\n \nit flew in\n\n\n \nWell theres a VAR check\n\n\n \nWhattd he do though\n\n\n \nHere is the handball\n\n\n \nThere\n\n\n \nIts going to be a goal\n\n\n \nHes adamant it was a penalty\n\n\n \nPoor from Bravo\nbut its a fantastic shot\n\n\n \nfrom Fabinho\n\n\n \nHeres Robertson\n\n\n \nSalah coming in\n\n\n \napplause\nLiverpool\n\n\n \nland another counter punch\n\n\n \nMo Salah drive it in a header\n\n\n \nWhat a finish\n\n\n \nMartin Angelino\n\n\n \nAguero\n\n\n \nstill going and still not scoring\n\n\n \ncan see here Aguero\n\n\n \na little nutmeg from Fabinho\n\n\n \nand then you expect him to do better\n\n\n \nYou expect him to set it\nout into that far corner\n\n\n \nMartin Its a throwin again\n\n\n \nAll Liverpool body</p>

filename	cleaned_content
1120 qG1IIIj9_fU.mp4.srt	<p>is back there\n\n\n \nMartin Angelino\ntrying to get through\n\n\n\n\nThats so so close Its a corner\n\n\n \nSo close\n\n\n \nMartin Henderson\n\n\n \ndone very well to get the cross in\n\n\n \nSo well for Sadio Mane\n\n\n \ncame in at the back post\n\n\n \nand squeezed it off the goalkeeper\n\n\n \n\n\n \nAgain Angelino\n\n\n \nand theyve got one\nThey do always score\n\n\n \nBernardo Silva swept it in\n\n\n \nNot over yet at Anfield\n\n\n \n\nHeres Sterling\n\n\n \nhas never given up the ghost\n\n\n \nShouts for handball hits\nAlexanderArnold again\n\n\n \nand then Henderson comes\nand grabs it\n\n\n \nGabriel Jesus is hurt\n\n\n \nHes nodding his head\n\n\n \nMaybe thats the anger\nfrom the first half\n\n\n \nMartin Angelino again\n\n\n \n\n\n \nJust too high for Bernardo\n\n\n</p>
1121 GsY4w7sjh9A.mp4.srt	<p>\n \nmusic\n\n\n \nWell the contrast is stark\n\n\n \nSeven of this starting for wales \nhave single figures in terms of caps\n\n\n \nincluding this fella David Brooks\n\n\n \nFirst chance for him to run\n\n\n \nncouldnt quite play it into \nGareth Bale in time\n\n\n \nbut theres good pressure there\n\n\n \nOh first start for Wales \nfirst goal for Wales\n\n\n \nDan James inside the first five minutes\n\n\n \nThat is not far from\nthe stuff of dreams\n\n\n \nGreat work by David Brooks \nwho runs with the ball so well\n\n\n \nand then this young exuberant \nWales side win the ball back\n\n\n \nGreat strike \non his competitive debut for Wales\n\n\n \nAbsolutely incredible moment \nfor him and his family\n\n\n \nand a huge start for Ryan Giggs\n\n\n \nAllen out to James here\n\n\n \nGood cross from Ben Davies \nOh side netting\n\n\n \nDavid Brooks who played such \na key role is setting up the first\n\n\n \nvery nearly claiming a second\n\n\n \nGreat overlap from Ben Davies\n\n\n \nIts a really good cross\n\n\n \nand its good contact \nas well from David Brooks\n\n\n \nBrooks now\n\n\n \nThats a lovely touch\n\nto take him beyond Rusnák\n\n\n \nStill at it Brooks\n\n\n \nThis young man just loves playing\nin the shirt of his country\n\n\n \nRecently crowned Player of the Year\nand that was a lovely piece of play\n\n\n \nBale rises high but couldnt \nget enough on the header\n\n\n \nto divert it goalwards\n\n\n \nDubravka commits himself\n\n\n \nIf Gareth Bale could just get \nabove the ball a little bit more\n\n\n \nit would have been a certain goal for Wales\n\n\n \nGoalkeeper caught in no mans land\n\n\n \nMak here for Slovakia \nhes sent him through\n\n\n \nOut comes</p>

filename	cleaned_content
	<p>Hennessey\n\n\n \nHennessey made the block \nand then from the follow up\n\n\n \nit went into the side netting\n\n\n \nFirst time Wales \nhave really been cut apart\n\n\n \nlts a good run\n\n\n \nWayne Hennessey came out of his goal \nso quickly and shut down the angle\n\n\n \nRusnák as I said \nhe makes that run from\n\n\n \nthe righthand side across \nthe defense\n\n\n \nlDoes it so well tries to just \nlift it over Wayne Hennessey\n\n\n \nBig strong aggressive arm \ncame out towards him\n\n\n \nRusnák\n\n\n \nDuda Hamšík\n\n\n \nTidy turn but saved from Robert Maks \neffort by Wayne Hennessey\n\n\n \nTerrific football from \nthe Slovakians once again\n\n\n \nManoeuvre the ball \nfrom left to right really well\n\n\n \nthen back across the pitch again\n\n\n \nlts a good slide \nlts a good save by Wayne\n\n\n \nJoe Allen\n\n\n \nBen Davies to Gareth Bale\n\n\n \nLovely whip on that ball\n\n\n \nAll the way through to Wilson\n\n\n \nlIt might come for Tyler Roberts\n\n\n \nHe was offside when it was hit\nThe flags up on the far side\n\n\n \nMagnificent cross from Gareth Bale \nPuts it in such a dangerous area\n\n\n \njust a little bit too high \nfor Tyler Roberts\n\n\n \nNice first touch by Harry Wilson\n\n\n \nRusnák is full of running here for Slovakia\n\n\n \nTime to pick his pass\n\n\n \nGot the header hard \nand that shot across there\n\n\n \nMatt Smith just didnt quite go with the runner\n\n\n \nGood cross in\n\n\n \nlts a wonderful header \nisnt it by Ďuriš\n\n\n \nlIt was Rusnák with the cross\n\n\n \nSmart save by Hennessey\n\n\n \nand then a huge chance \lto get Slovakia back into the game\n\n\n \nUnfortunately for the visitors \nit goes begging\n\n\n</p>
1122	<p>R0XfelUhG_Y.mp4.srt</p> <p>\n \nHello and welcome to \nthe Stadio Olimpico in\n\n\n \nRome for round of \nthe Serie A season\n\n\n \nlts Lazio versus Sassuolo\n\n\n \nlm Patrick Terrone your \ncommentator for this game\n\n\n \nlIt is the home side Lazio\n\n\n \nwho get in the first half underway\n\n\n \nwhistle\n\n\n \nlts Whipped in and \nthat's a good header\n\n\n \nand a good save by Consigli\n\n\n \nwho watched it all the way\n\n\n \nLazio\n\n\n \nwith an early chance then\n\n\n \nlts been a little bit \ndisjointed that\n\n\n \nOh that flick Caicedo\n\n\n \nOh almost in the back of the net\n\n\n \na rebound\n\n\n \noff Immobile\n\n\n \nlwhistle\n\n\n \nlts Sassuolo\n\n\n \nwho get the second half underway\n\n\n \nlLulic\n\n\n \nsquaring the ball to Patrick \nwho comes forward\n\n\n \nlts now out wide to Marusic\n\n\n \nPatrick with a chance to cross\n\n\n \nhand ball and that's a penalty kick\n\n\n \nsays the referee\n\n\n \nWell the cross came from Patrick\n\n\n \nand hes going to go over \nand have a look so\n\n\n \nnot an easy decision\n\n\n \nfor the ref to make\n\n\n \nReferee has confirmed \nlts a penalty kick\n\n\n \nlImmobile then up against\n\n\n \nlthe Consigli here he goes\n\n\n \nlImmobile has given Lazio the lead\n\n\n \nnonenil to the home side\n\n\n \nlCiro Immobile\n\n\n \ngoal number for the striker\n\n\n \nAn easy finish sending Consigli\n\n\n \nlthe wrong way\n\n\n \nlDemiral\n\n\n \nlDuncan\n\n\n \nlts nicely forward Boga\n\n\n \nCan he twist and turn here\n\n\n \nHere he goes \nhes got the pace\n\n\n \nlnot a bad effort \nlLirola trying to place it in\n\n\n \nlts there Rogerio\n\n\n \nlts tapped it in\n\n\n \nand Sassuolo right back in it\n\n\n \nlfirst chance theyve had in \nthe entire match\n\n\n \nand its a goal\n\n\n \nlIt is of all people\n\n\n \nlRogerio the Brazilian\n\n\n \nlCorrea\n\n\n \noh that's a nice ball Immobile\n\n\n \noh what a chance\n\n\n \nlts give his side \nthe lead once again\n\n\n \nlCorrea\n\n\n \nlleaves it for Luiz Alberto\n\n\n \njust that's a wonderful ball \nlImmobile\n\n\n \nlWhat a miss\n\n\n \nlCiro Immobile all alone\n\n\n \nAll he had to do\n\n\n \nlwas put it in the back of the net\n\n\n \nand hes missed\n\n\n \nlLuis Alberto\n\n\n \nlnow to Radu Radu cross\n\n\n \nl\n\ntowards Milinković Savić\n\n\n \nlts saved off the line by Consigli\n\n\n \nlGreat header\n\n\n \nlfrom the substitute\n\n\n \nlbut equally\n\n\n \nlngood save\n\n\n \nlNow\n\n\n \nl\n\n\n \nlspace\n\n\n \nl\n\n\n \nlBerardi\n\n\n \nlSassuolo\n\n\n \nlhave taken the lead\n\n\n \nl\n\n\n \nlBerardi\n\n\n \nlfinishing with a plum\n\n\n \nlWhipped in\n\n\n</p>

filename	cleaned_content
	\nflicked out by\n\n\n \nSensi whos the smallest \nman on the pitch\n\n\n \nNow Patrick out too\n\n\n \nImmobile Immobile\n\n\n \n\nacross the face of goal its there\n\n\n \nThe equalizer stands\n\n\n \nLulic\n\n\n \nfrom a meter out taps it in\n\n\n \nThis time Immobile\n\n\n \nprovides the assist\n\n\n \nThe game has finished\n\n\n \nLazio two Sassuolo two\n\n\n \ncrowd cheering\n\n\n

1123 rows × 2 columns

```
In [6]: # Tokenise and lowercase
df['cleaned_content'] = df['cleaned_content'].apply(lambda x: ' '.join(word

# Cleaned data without lowering case
df['original_case'] = df['cleaned_content'].apply(lambda x: ' '.join(word fo

# Display the DataFrame with the new columns
print(df[['filename', 'cleaned_content', 'original_case']])
```

```
filename \
0    cvctybLyEqE.mp4.srt
1    oCl6MCo0i1Q.mp4.srt
2    _kDc62-upoA.mp4.srt
3    41k02TvmNbg.mp4.srt
4    DjNA-MIdCbY.mp4.srt
...
1118   dJMNNQHkz0M.mp4.srt
1119   xK5R8h7zIn0.mp4.srt
1120   qG1lIIj9_fU.mp4.srt
1121   GsY4w7sjh9A.mp4.srt
1122   R0XfelUhG_Y.mp4.srt
```

```
cleaned_content \
0
```

welcome to a rainy stadio olimpico lazio against juventus should be a real atmosphere this can lazio become the first team this serie a season to beat juventus i dont think theyre going to stop their irrepressible march to the title particularly after napolli could only draw against milan last night marco guida gets us underway referee got in the way there here goes dybala lucas once more wins it back now immobile opens up for correa szczęsny got behind it on an awkward surface and held it well juve have a tricky game against atalanta in the cup tricky moments here as well as it falls for luiz alberto szczęsny with the save and lucas with a miss hit heres correa miss kick there from matuidi milinkovićsavić look at the space for parolo fabulous save from szczęsny who has kept juve in it here thats the pick of them the effort levels never drop when hes out there it might bounce here for immobile rugani does so well to recover and get it behind great attacking instinct from immobile much to discuss much to find out no changes at half time they need something to show for their efforts in from luiz alberto and they have it did it ricochet off milinkovićsavić is it an own goal no one at the olympico could care lazio in front juve in trouble in the rain in rome now correa oh immobile is in with a chance to make it two forward by alex sandro oh great skill from bernardeschi for dybala spilled big chance joão cancelo with an equalizer for juventus back from injury to save them the substitute makes it hold on to this or even better it and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a penalty marco guida has given the penalty they were under massive pressure can they turn it round they can its what champions do on the ropes sergej milinkovićsavić over the bar juventus can celebrate can punch the air juve win it by two goals to one

1  
hello and welcome to the stadio bentegodi where this afternoon chievo verona and frosinone go headtohead in round of serie a im marco palmieri great to be with you its a mustwin match for both sides were away at the bentegodi between chievo and frosinone the corners are coming can they take advantage the big men in the area as birsa steps up to take the corner it is indeed sergio pellissier with a header from the corner and the captain does it again for chievo making sure that there werent any irregularities in the box maybe a push it will be a onfield review the goal has been disallowed its still here at the bentegodi pellissier is furious a push im assuming there was a push by pellissier as he lept up into the air as rotti gets set to start the second half and were away as hetemaj gets it to birsa he skips past maimo now đorđević the first shot on target but sportiello denies him with a quick save not necessarily a difficult one but one that had to be made as the ball found its way to the serb striker corner taken short as the ball at the far post for chiesa brilliant save from sorrentino the frosinone players thought that itd crossed the line the watch on referee rotti did not signal a goal and play continues as now depaoli is in space and hes tripped on the edge of the area here there will be a foul and a yellow its a second yellow capuano is off frosinone reduced to men hetemaj sees the overlap now play inside pellissier on his left foot he slipped over as he took the shot great chance for chievo as they continue to attack now tomović spreads it a

cross to radovanović now bani plays it in the box but depaoli cannot reach it and the ball dribbles back to the keeper what a chance for pellissier here the shot comes in and thats a wonderful goal giaccherini gives chievo the lead on minutes with a superb freekick sportiello rooted to the spot an excellent goal and thoroughly deserved there it is the final whistle a simply perfect giaccherini freekick has given chievo their first win of the season in the last match of the year theyve defeated a man frosinone by a one goal to nill

2

the leagues top scorers against seria as weakest frontline atalanta are bidding to make the best year in club history spal are four points from safety and running out of time in their efforts to beat the drop di francesco in the channel able to wander inside gabriel strefezza lining it up its a good save from sportiello the game becoming stretched just like atalanta want it to be zapata running at thiago cionek duván zapatas cross its in for ilicic a difficult finish made to look all so easy zapata still seeking his first goal since his return to action that was a telling contribution ilicic has scored against fiorentina in midweek continues his wonderful start to well in by gosens this is pasalic playing slightly deeper this evening now zapat a plenty of room for him to get moving up against vicari duván zapata with the shot its off the frame of the goal this is valoti thiago cionek joining in the cross is only halfcleared strafezza cionek with the ball in sportielo stays at home petagnas header is off the post so close to yet another goal against his former club thats a nice touch from di francesco reca is very quick hes trying to get away from djimsiti this is recas cross and its put in by petagna a couple of former atalanta players combine reca on loan petagna once again scores against la dea no one could catch up with the ball djimsiti zapata a foul on him vicari dabo valoti hes kept it mattia valoti spal have another theyve come from behind here to lead in bergamo what a superb individual strike petagna first and now valoti he took his goal splendidly well launched long by ilicic muriel gives chase hes got there this is zapata duvan zapata its deflected terrific save from berisha that was creeping in another look at that long ball from ilicic muriel did well zapata struck it firmly spal win in bergamo what a turn around for semplici only a fourth victory of the season for seria as bottom side three massive points in their quest to retain their topflight status atalanta offcolor and out of sorts they stay fifth crowd cheering crowd cheering

3

welcome to the ennio tardini stadium this week its the meeting of the two sides who are the bane of the city of milan in the last match week parma against cagliari last weekend parma recorded a surprise victory at san siro against inter while cagliari got an impressive draw against milan in sardinia were underway stolen and now gervinho can show us that famous sprinting power still going gervinho still going gets the shot off just wide of the post trademark run from gervinho look where he picks the ball up in his own half then he goes on an angled run inside gets past klavan sees off ionă and a low shot you can see what he was going for there dimarco inglese is in there again inglese loops it over the keeper cragno this time has the goal been given it looks like it has denied by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma the lead excellent run inglese persistent got lucky there from the deflection off cragno his shot cannoned off the goalkeeper back into his face and into the net theres a couple of checks referee calvarese the goal has been given parma do go into the lead luca rigoni di gaudio will he go for the cross he plays it back its a shot from antonino barillà the ball is still alive gervinho gervinho pulls it back into the area its luca rigoni a familiar sight the raiding midfielder the second half starts marco sau goes down referee says play on now now gervinho on one of those runs gervinho still going gervinho past another whats the shot like its a fantastic shot gervinho what a goal a standing ovation from the tardini danger at one end gervinho picked the ball up and then showed his genius a rare lack of control from srna but tenacity getting the ball back plays it in an attempted backflip by marco sau had there been a foul in the buildup won back by parma the break is on into the area ceravolo gets it back will he make it three great save by cragno refer

ee juan paolo calvarese blows the fulltime whistle parma cagliari

4

music chilwell oh and its fired in in typical oxlade chamberlain style he is a goal threat he really didnt catch it well its another lovely ball its not the easiest of chances when you strike it that purely theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in by chilwell amongst a clutch of players but he manages just to keep his opponents at arms length chilwell to take the corner thats a free header thats another goal simple for kane england incidentally are not showing their array of set pieces they want to sail to the finals on the play nobody knows what those montenegro defenders were doing here to allow kane a free run theyve got to get a bit more physical contact and so easy rashford gets in on the act well thats four oxlade chamberlain with a really good ball maguire cant believe he hasnt scored there rashford tidies up sancho heres the hattrick another harry hattrick five for england this landmark evening could not be going better it wasnt the easiest ball to control first time from alexanderarnold just bouncing in front of him laugh hes such a confident frame of mind that he just decides to take a touch it proves to be the right decision rashfords in behind wow how has it not gone in in the end it has it looked like the defender rushing back instead he got the final touch rashfords done this several times tonight just easing past his man an instinctive reaction by šofranac he might get a tap in this time tammy abraham opens his england account it is seven with seven minutes to go yes a big moment for him and a classic center forwards goal this one making that dash across your man sticking out a leg wonderful stuff from abraham in an ideal world he could get on the score sheet get him some belief at this level it finishes england seven montenegro nil they qualify for the finals music

...

...

1118

match begins with torino very much all won five consecutive clean sheets victories over inter and last week atalanta today begins a trio of fixtures against the relegation side chievo are expected to be beaten the first whistle goes it is chievo in the yellow who kick us off the cross is in whipped in to de silvestri the shot from range ansaldi converts cheering two goals for him this season the argentinian cheering now a round of applause and a pause for the memory of davide astori the former fiorentina player who so tragically lost his life last year applause cheering this is better now ansaldi meité who wasnt wise enough to trick diousse cheering meggiolini đorđević to his right he has a go at goal a goal that he used to shoot at himself when wearing the colors of torino played for them between cheering were back underway torino kick us off cheering depaoli shifting inside trying to make the angle for the cross took a deflection should score sirigu denies đorđević with the best chance of the game cheering he keeps that clean sheet record it was an excellent reaction save close to him đorđević with the second attempt simply had to score cheering whistle blows now this could be the moment zaza sorrentino wonderful save cheering belotti with the strike and belotti with the missile torino take the lead what a vital goal its the captains goal cheering number eight of the season surely the best of the lots a cracker from andrea belotti cheering now looking to just finish off this match tomás rincón tomás rincón cheering thats a thing of beauty cheering he comes off the bench he was supposed to put solidity into the midfield in the end hes slid a beautiful strike into the top corner torino two chievo nil in injury time cheering zaza looks for number three and in the net it goes two goals in injury time its torino three chievo nil zaza scores home and away against chievo this season hes had to wait since the end of september for that second goal the final whistle goes it finishes here torino three chievo nil cheering

1119

applause here is the handball michael oliver waves it away here is var trying to give us an alternative version of it then goes on with mane stones cannot stop the cross salah is coming in fabinho has a go and what a goal applause what a goal what a start for liverpool it flew in well theres a var che

ck whatd he do though here is the handball there its going to be a goal hes adamant it was a penalty poor from bravo but its a fantastic shot from fabi nho heres robertson salah coming in applause liverpool land another counter punch mo salah drive it in a header what a finish martin angelino aguero st ill going and still not scoring i can see here aguero a little nutmeg from fabinho and then you expect him to do better you expect him to set it out i nto that far corner martin its a throwin again all liverpool body is back t here martin angelino trying to get through thats so so close its a corner s o close martin henderson done very well to get the cross in so well for sad io mane came in at the back post and squeezed it off the goalkeeper again a ngelino and theyve got one they do always score bernardo silva swept it in not over yet at anfield heres sterling has never given up the ghost shouts for handball hits alexanderarnold again and then henderson comes and grabs it gabriel jesus is hurt hes nodding his head maybe thats the anger from th e first half martin angelino again just too high for bernardo

1120

played by firmino to andy robertson just beyond the stretch of mane but con trolled there by mo salah to mane lallana goes wide alexanderarnold then sa lah poses an early threat to wolves a lot of bodies central on the midfield today we need to see that a lot more heres henderson firmino finds alexande rarnold by but not too far its in to sadio mane whos on the turn and rui pa tricio scrambling across his goal the ball going on the top onto mane crowd cheering there was a touch on the arm and its not going to count hes certai nly not offside whistle blows the goal stands sadio mane has made the break through here for liverpool won by moutinho jonny trying to break free of an dy robertson and manages to pull it back its gone all the way across and ne to stuns anfield with an equalizer right on the brink of halftime meanwhile theres frustration from the first but that will be proven to be right were still waiting for confirmation on this one brace yourselves the decision ha s come in wolves are going to be fuming about this the goal has been disall owed that was over the top of the bar from wijnaldum van dijk in trouble he s lost that here to diogo jota he produces the save certainly this time its the goalkeeper who saves virgil van dijk alexanderarnold ought to be carefu l jimenez

1121 music well the contrast is stark seven of this starting for wales hav e single figures in terms of caps including this fellas david brooks first c hance for him to run couldnt quite play it into gareth bale in time but the res good pressure there oh first start for wales first goal for wales dan j ames inside the first five minutes that is not far from the stuff of dreams great work by david brooks who runs with the ball so well and then this you ng exuberant wales side win the ball back great strike on his competitive d ebut for wales absolutely incredible moment for him and his family and a hu ge start for ryan giggs allen out to james here good cross from ben davies oh side netting david brooks who played such a key role is setting up the f irst very nearly claiming a second great overlap from ben davies its a real ly good cross and its good contact as well from david brooks brooks now tha ts a lovely touch to take him beyond rusnák still at it brooks this young m an just loves playing in the shirt of his country recently crowned player o f the year and that was a lovely piece of play bale rises high but couldnt get enough on the header to divert it goalwards dubravka commits himself if gareth bale could just get above the ball a little bit more it would been a certain goal for wales goalkeeper caught in no mans land mak here for slova kia hes sent him through out comes hennessey hennessey made the block and t hen from the follow up it went into the side netting first time wales have really been cut apart its a good run wayne hennessey came out of his goal s o quickly and shut down the angle rusnák as i said he makes that run from t he righthand side across the defense does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda hamšík tidy turn but saved from robert maks effort by wayne hennessey terrific football from the slovakians once again manoeuvre the ball from le ft to right really well then back across the pitch again its a good slide i ts a good save by wayne joe allen ben davies to gareth bale lovely whip on that ball all the way through to wilson it might come for tyler roberts he was offside when it was hit the flags up on the far side magnificent cross

from gareth bale puts it in such a dangerous area just a little bit too high for tyler roberts nice first touch by harry wilson rusnák is full of running here for slovakia time to pick his pass got the header hard and that shot across there matt smith just didnt quite go with the runner good cross in its a wonderful header isnt it by ďuriš it was rusnák with the cross smart save by hennessey and then a huge chance to get slovakia back into the game unfortunately for the visitors it goes begging

1122

hello and welcome to the stadio olimpico in rome for round of the serie a season its lazio versus sassuolo im patrick terrone your commentator for this game it is the home side lazio who get in the first half underway whistle its whipped in and thats a good header and a good save by consigli who watched it all the way lazio with an early chance then its been a little bit disjointed that oh that flick caicedo oh almost in the back of the net a rebound off immobile whistle its sassuolo who get the second half underway lulin squaring the ball to patrick who comes forward its now out wide to marusic patrick with a chance to cross hand ball and thats a penalty kick says the referee well the cross came from patrick and hes going to go over and have a look so not an easy decision for the ref to make referee has confirmed its a penalty kick immobile then up against the consigli here he goes immobile has given lazio the lead onenil to the home side ciro immobile goal number for the striker an easy finish sending consigli the wrong way demiral Duncan its nicely forward boga can he twist and turn here here he goes hes got the pace not a bad effort lirola trying to place it in its there rogerio tapped it in and sassuolo right back in it first chance theyve had in the entire match and its a goal it is of all people rogerio the brazilian correa oh thats a nice ball immobile oh what a chance to give his side the lead once again correa leaves it for luiz alberto just it thats a wonderful ball immobile what a miss ciro immobile all alone all he had to do was put it in the back of the net and hes missed luis alberto now to radu radu cross towards milinkovićsavić its saved off the line by consigli great header from the substitute but equally good save now space and berardi through here clearer through berardi sassuolo have taken the lead on the break berardi finishing with a plum whipped in flicked out by sensi whos the smallest man on the pitch now patrick out too immobile immobile across the face of goal its there the equalizer stands lulic from a meter out taps it in this time immobile provides the assist the game has finished lazio two sassuolo two crowd cheering

original\_case

0

welcome to a rainy stadio olimpico lazio against juventus should be a real atmosphere this can lazio become the first team this serie a season to beat juventus i dont think theyre going to stop their irrepressible march to the title particularly after napoli could only draw against milan last night mario guida gets us underway referee got in the way here goes dybala lucas once more wins it back now immobile opens up for correa szczęsny got behind it on an awkward surface and held it well juve have a tricky game against atalanta in the cup tricky moments here as well as it falls for luiz alberto szczęsny with the save and lucas with a miss hit heres correa misskick there from matuidi milinkovićsavić look at the space for parolo fabulous save from szczęsny who has kept juve in it here thats the pick of them the effort levels never drop when hes out there it might bounce here for immobile rugani does so well to recover and get it behind great attacking instinct from immobile much to discuss much to find out no changes at half time they need something to show for their efforts in from luiz alberto and they have it did it ricochet off milinkovićsavić is it an owngoal no one at the olimpico could care lazio in front juve in trouble in the rain in rome now correa oh immobile is in with a chance to make it two forward by alex sandro oh great skill from bernardeschi for dybala spilled big chance joão cancelo with an equalizer for juventus back from injury to save them the substitute makes it hold on to this or even better it and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulic

hes given a penalty marco guida has given the penalty they were under massive pressure can they turn it round they can its what champions do on the ropes sergej milinković savić over the bar juventus can celebrate can punch the air juve win it by two goals to one

1

hello and welcome to the stadio bentegodi where this afternoon chievo verona and frosinone go headtohead in round of serie a im marco palmieri great to be with you its a mustwin match for both sides were away at the bentegodi between chievo and frosinone the corners are coming can they take advantage the big men in the area as birsa steps up to take the corner it is indeed sergio pellissier with a header from the corner and the captain does it again for chievo making sure that there werent any irregularities in the box maybe a push it will be a onfield review the goal has been disallowed its still here at the bentegodi pellissier is furious a push im assuming there was a push by pellissier as he lept up into the air as rocchi gets set to start the second half and were away as hetemaj gets it to birsa he skips past maello now đorđević the first shot on target but sportiello denies him with a quick save not necessarily a difficult one but one that had to be made as the ball found its way to the serb striker corner taken short as the ball at the far post for chibsah brilliant save from sorrentino the frosinone players thought that itd crossed the line the watch on referee rocchi did not signal a goal and play continues as now depaoli is in space and hes tripped on the edge of the area here there will be a foul and a yellow its a second yellow capuano is off frosinone reduced to men hetemaj sees the overlap now play inside pellissier on his left foot he slipped over as he took the shot great chance for chievo as they continue to attack now tomović spreads it across to radovanović now bani plays it in the box but depaoli cannot reach it and the ball dribbles back to the keeper what a chance for pellissier here the shot comes in and thats a wonderful goal giaccherini gives chievo the lead on minutes with a superb freekick sportiello rooted to the spot an excellent goal and thoroughly deserved there it is the final whistle a simply perfect giaccherini freekick has given chievo their first win of the season in the last match of the year theyve defeated a man frosinone by a one goal to nill

2

the leagues top scorers against seria as weakest frontline atalanta are bidding to make the best year in club history spal are four points from safety and running out of time in their efforts to beat the drop di francesco in the channel able to wander inside gabriel strefezza lining it up its a good save from sportiello the game becoming stretched just like atalanta want it to be zapata running at thiago cionek duván zapatas cross its in for ilicic a difficult finish made to look all so easy zapata still seeking his first goal since his return to action that was a telling contribution ilicic has scored against fiorentina in midweek continues his wonderful start to well in by gosens this is pasalic playing slightly deeper this evening now zapat a plenty of room for him to get moving up against vicari duván zapata with the shot its off the frame of the goal this is valoti thiago cionek joining in the cross is only halfcleared strafezza cionek with the ball in sportiello stays at home petagnas header is off the post so close to yet another goal against his former club thats a nice touch from di francesco reca is very quick hes trying to get away from djimsiti this is recas cross and its put in by petagna a couple of former atalanta players combine reca on loan petagna once again scores against la dea no one could catch up with the ball djimsiti zapata a foul on him vicari dabo valoti hes kept it mattia valoti spal have another theyve come from behind here to lead in bergamo what a superb individual strike petagna first and now valoti he took his goal splendidly well launched long by ilicic muriel gives chase hes got there this is zapata duvan zapata its deflected terrific save from berisha that was creeping in another look at that long ball from ilicic muriel did well zapata struck it firmly spal win in bergamo what a turn around for semplici only a fourth victory of the season for seria as bottom side three massive points in their quest to retain their topflight status atalanta offcolor and out of sorts they stay fifth crowd cheering crowd cheering crowd cheering

3

welcome to the ennio tardini stadium this week its the meeting of the two sides who are the bane of the city of milan in the last match week parma against cagliari last weekend parma recorded a surprise victory at san siro against inter while cagliari got an impressive draw against milan in sardinia were underway stolen and now gervinho can show us that famous sprinting power still going gervinho still going gets the shot off just wide of the post trademark run from gervinho look where he picks the ball up in his own half then he goes on an angled run inside gets past klavan sees off ionita and a low shot you can see what he was going for there dimarco inglese is in there again inglese loops it over the keeper cragno this time has the goal been given it looks like it has denied by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma the lead excellent run inglese persistent got lucky there from the deflection off cragno his shot cannoned off the goalkeeper back into his face and into the net theres a couple of checks referee calvarese the goal has been given parma do go into the lead luca rigoni di gaudio will he go for the cross he plays it back its a shot from antonino barilla the ball is still alive gervinho gervinho pulls it back into the area its luca rigoni a familiar sight the raiding midfielder the second half starts marco sau goes down referee says play on now now gervinho on one of those runs gervinho still going gervinho past another whats the shot like its a fantastic shot gervinho what a goal a standing ovation from the tardini danger at one end gervinho picked the ball up and then showed his genius a rare lack of control from srna but tenacity getting the ball back plays it in an attempted backflip by marco sau had there been a foul in the buildup won back by parma the break is on into the area ceravolo gets it back will he make it three great save by cragno referee juan paolo calvarese blows the fulltime whistle parma cagliari

4

music chilwell oh and its fired in in typical oxlade chamberlain style he is a goal threat he really didnt catch it well its another lovely ball its not the easiest of chances when you strike it that purely theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in by chilwell amongst a clutch of players but he manages just to keep his opponents at arms length chilwell to take the corner thats a free header thats another goal simple for kane england incidentally are not showing their array of set pieces they want to sail to the finals on the play nobody knows what those montenegro defenders were doing here to allow kane a free run theyve got to get a bit more physical contact and so easy rashford gets in on the act well thats four oxlade chamberlain with a really good ball maguire cant believe he hasnt scored there rashford tidies up sancho heres the hattrick another harry hattrick five for england this landmark evening could not be going better it wasnt the easiest ball to control first time from alexanderarnold just bouncing in front of him laugh hes such a confident frame of mind that he just decides to take a touch it proves to be the right decision rashfords in behind wow how has it not gone in in the end it has it looked like the defender rushing back i see he got the final touch rashfords done this several times tonight just easing past his man an instinctive reaction by softranac he might get a tap in this time tammy abraham opens his england account it is seven with seven minutes to go yes a big moment for him and a classic center forwards goal this one making that dash across your man sticking out a leg wonderful stuff from abraham in an ideal world he could get on the score sheet get him some belief at this level it finishes england seven montenegro nil they qualify for the finals music

...

...

1118

match begins with torino very much all won five consecutive clean sheets victories over inter and last week atalanta today begins a trio of fixtures against the relegation side chievo are expected to be beaten the first whistle goes it is chievo in the yellow who kick us off the cross is in whipped in to de silvestri the shot from range ansaldi converts cheering two goals for him this season the argentinian cheering now a round of applause and a pause for the memory of davide astori the former fiorentina player who so t

ragically lost his life last year applause cheering this is better now ansa ldi meíté who wasnt wise enough to trick diousse cheering meggiolini dördević to his right he has a go at goal a goal that he used to shoot at himself when wearing the colors of torino played for them between cheering were back underway torino kick us off cheering depaoli shifting inside trying to make the angle for the cross took a deflection should score sirigu denies dördević with the best chance of the game cheering he keeps that clean sheet record it was an excellent reaction save close to him dördević with the second attempt simply had to score cheering whistle blows now this could be the moment zaza sorrentino wonderful save cheering belotti with the strike and belotti with the missile torino take the lead what a vital goal its the captains goal cheering number eight of the season surely the best of the lots a cracker from andrea belotti cheering now looking to just finish off this match tomás rincón tomás rincón cheering thats a thing of beauty cheering he comes off the bench he was supposed to put solidity into the midfield in the end hes slid a beautiful strike into the top corner torino two chievo nil in injury time cheering zaza looks for number three and in the net it goes two goals in injury time its torino three chievo nil zaza scores home and away against chievo this season hes had to wait since the end of september for that second goal the final whistle goes it finishes here torino three chievo nil cheering

1119

applause here is the handball michael oliver waves it away here is var trying to give us an alternative version of it then goes on with mane stones cant stop the cross salah is coming in fabinho has a go and what a goal applause what a goal what a start for liverpool it flew in well theres a var check whatd he do though here is the handball there its going to be a goal hes adamant it was a penalty poor from bravo but its a fantastic shot from fabinho heres robertson salah coming in applause liverpool land another counter punch mo salah drive it in a header what a finish martin angelino aguero still going and still not scoring i can see here aguero a little nutmeg from fabinho and then you expect him to do better you expect him to set it out into that far corner martin its a throwin again all liverpool body is back there martin angelino trying to get through thats so so close its a corner so close martin henderson done very well to get the cross in so well for sadio mane came in at the back post and squeezed it off the goalkeeper again angelino and theyve got one they do always score bernardo silva swept it in not over yet at anfield heres sterling has never given up the ghost shouts for handball hits alexanderarnold again and then henderson comes and grabs it gabriel jesus is hurt hes nodding his head maybe thats the anger from the first half martin angelino again just too high for bernardo

1120

played by firmino to andy robertson just beyond the stretch of mane but controlled there by mo salah to mane lallana goes wide alexanderarnold then salah poses an early threat to wolves a lot of bodies central on the midfield today we need to see that a lot more heres henderson firmino finds alexanderarnold by but not too far its in to sadio mane whos on the turn and rui patrício scrambling across his goal the ball going on the top onto mane crowd cheering there was a touch on the arm and its not going to count hes certainly not offside whistle blows the goal stands sadio mane has made the breakthrough here for liverpool won by moutinho jonny trying to break free of andy robertson and manages to pull it back its gone all the way across and net stuns anfield with an equalizer right on the brink of halftime meanwhile theres frustration from the first but that will be proven to be right were still waiting for confirmation on this one brace yourselves the decision has come in wolves are going to be fuming about this the goal has been disallowed that was over the top of the bar from wijnaldum van dijk in trouble hes lost that here to diogo jota he produces the save certainly this time its the goalkeeper who saves virgil van dijk alexanderarnold ought to be careful jimenez

1121 music well the contrast is stark seven of this starting for wales have single figures in terms of caps including this fella david brooks first chance for him to run couldnt quite play it into gareth bale in time but theres good pressure there oh first start for wales first goal for wales dan j

ames inside the first five minutes that is not far from the stuff of dreams great work by david brooks who runs with the ball so well and then this young exuberant wales side win the ball back great strike on his competitive debut for wales absolutely incredible moment for him and his family and a huge start for ryan giggs allen out to james here good cross from ben davies oh side netting david brooks who played such a key role is setting up the first very nearly claiming a second great overlap from ben davies its a really good cross and its good contact as well from david brooks brooks now that's a lovely touch to take him beyond rusnák still at it brooks this young man just loves playing in the shirt of his country recently crowned player of the year and that was a lovely piece of play bale rises high but couldnt get enough on the header to divert it goalwards dubravka commits himself if gareth bale could just get above the ball a little bit more it would been a certain goal for wales goalkeeper caught in no mans land mak here for slovakia hes sent him through out comes hennessey hennessey made the block and then from the follow up it went into the side netting first time wales have really been cut apart its a good run wayne hennessey came out of his goal so quickly and shut down the angle rusnák as i said he makes that run from the righthand side across the defense does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda hamšík tidy turn but saved from robert maks effort by wayne hennessey terrific football from the slovakians once again manoeuvre the ball from left to right really well then back across the pitch again its a good slide its a good save by wayne joe allen ben davies to gareth bale lovely whip on that ball all the way through to wilson it might come for tyler roberts he was offside when it was hit the flags up on the far side magnificent cross from gareth bale puts it in such a dangerous area just a little bit too high for tyler roberts nice first touch by harry wilson rusnák is full of running here for slovakia time to pick his pass got the header hard and that shot across there matt smith just didnt quite go with the runner good cross in its a wonderful header isnt it by duriš it was rusnák with the cross smart save by hennessey and then a huge chance to get slovakia back into the game unfortunately for the visitors it goes begging

1122

hello and welcome to the stadio olimpico in rome for round of the serie a season its lazio versus sassuolo im patrick terrone your commentator for this game it is the home side lazio who get in the first half underway whistle its whipped in and thats a good header and a good save by consigli who watched it all the way lazio with an early chance then its been a little bit disjointed that oh that flick caicedo oh almost in the back of the net a rebound off immobile whistle its sassuolo who get the second half underway lulin squaring the ball to patrick who comes forward its now out wide to marusic patrick with a chance to cross hand ball and thats a penalty kick says the referee well the cross came from patrick and hes going to go over and have a look so not an easy decision for the ref to make referee has confirmed its a penalty kick immobile then up against the consigli here he goes immobile has given lazio the lead onenil to the home side ciro immobile goal number for the striker an easy finish sending consigli the wrong way demiral Duncan its nicely forward boga can he twist and turn here here he goes hes got the pace not a bad effort lirola trying to place it in its there rogerio tapped it in and sassuolo right back in it first chance theyve had in the entire match and its a goal it is of all people rogerio the brazilian correa oh thats a nice ball immobile oh what a chance to give his side the lead once again correa leaves it for luiz alberto just it thats a wonderful ball immobile what a miss ciro immobile all alone all he had to do was put it in the back of the net and hes missed luis alberto now to radu radu cross towards milinkovićsavić its saved off the line by consigli great header from the substitute but equally good save now space and berardi through here clearer through berardi sassuolo have taken the lead on the break berardi finishing with a plum whipped in flicked out by sensi whos the smallest man on the pitch now patrick out too immobile immobile across the face of goal its there the equalizer stands lulic from a meter out taps it in this time immobile provides the assist the game has finished lazio two sassuolo two crowd cheering

```
[1123 rows x 3 columns]
```

In [7]: df

Out[7]:

	filename	content	cleaned_content
0	cvctybLyEqE.mp4.srt	<p>1\n00:00:00,177 --&gt; 00:00:06,044\n-Welcome to a rainy Stadio Olimpico.\n Lazio against Juventus \n\n2\n00:00:06,288 --&gt; 00:00:09,244\nshould be a real atmosphere this. \n\n3\n00:00:09,466 --&gt; 00:00:14,688\nCan Lazio become the first team inthis Serie A season to beat Juventus?\n\n4\n00:00:14,680 --&gt; 00:00:18,644\nI don't think they're going to stop\ntheir irrepressible march to the title, \n\n5\n00:00:18,640 --&gt; 00:00:22,666\nparticularly after Napoli could only \ndraw against Milan last night.\n\n6\n00:00:23,666 --&gt; 00:00:26,133\nMarco Guida gets us underway.\n\n7\n00:00:32,355 --&gt; 00:00:34,111\nReferee got in the way there.\n\n8\n00:00:34,800 --&gt; 00:00:38,177\nHere goes Dybala, \nLucas once more wins it back.\n\n9\n00:00:41,288 --&gt; 00:00:44,933\nNow Immobile opens up for Correa.\n\n10\n00:00:45,622 --&gt; 00:00:50,444\nSzczęsny got behind it on an \nawkward surface and held it well.\n\n11\n00:00:56,244 --&gt; 00:00:58,933\nJuve have a tricky game \nagainst Atalanta in the Cup.\n\n12\n00:01:02,088 --&gt; 00:01:06,066\nTricky moments here as \nwell as it falls for Luiz Alberto.\n\n13\n00:01:06,266 --&gt; 00:01:09,844\n Szczęsny with the save\nand Lucas with a miss hit. \n\n14\n00:01:13,822 --&gt; 00:01:15,088\nHere's Correa,\n\n15\n00:01:17,866 --&gt; 00:01:22,488\nmiss kick there from Matuidi. \nMilinković-Savić.\n\n16\n00:01:24,488 --&gt; 00:01:28,577\nLook at the space for Parolo.\nFabulous save from Szczęsny \n\n17\n00:01:29,022 --&gt; 00:01:31,977\nwho has kept Juve in it here.\nThat's the pick of them.\n\n18\n00:01:44,000 --&gt; 00:01:46,688\nThe effort levels never\ndrop when he's out there. \n\n19\n00:01:47,244 --&gt; 00:01:49,688\nIt might bounce here for Immobile.\n\n20\n00:01:50,444 --&gt; 00:01:54,200\nRugani does so well to recover\nand get it behind.\n\n21\n00:01:55,066 --&gt; 00:01:57,622\nGreat attacking instinct from Immobile.\n\n22\n00:02:00,111 --&gt; 00:02:04,711\nMuch to discuss, much to find out.\nNo changes at half time.\n\n23\n00:02:08,977 --&gt; 00:02:11,066\nThey need something\nto show for their efforts. \n\n24\n00:02:11,060 --&gt; 00:02:13,755\nIn from Luiz Alberto, and they have it.\n\n25\n00:02:16,311 --&gt; 00:02:18,933\nDid it ricochet off Milinković-Savić? \n\n26\n00:02:19,133 --&gt; 00:02:22,577\nIs it an own-goal?\nNo one at the Olmpico could care.\n\n27\n00:02:22,955 --&gt; 00:02:28,488\nLazio in front, Juve in trouble\nin the rain in Rome.\n\n28\n00:02:32,688 --&gt; 00:02:33,933\nNow Correa.\n\n29\n00:02:35,044 --&gt; 00:02:39,888\nOh, Immobile is in with a chance to make it two.\n\n30\n00:02:42,822 --&gt; 00:02:44,600\nForward by Alex Sandro.</p>	welcome to a rainy stadio olimpico lazio against juventus should be a real atmosphere this can lazio become the first team this serie a season to beat juventus i dont think theyre going to stop their irrepressible march to the title particularly after napoli could only draw against milan last night marco guida gets us underway referee got in the way there here goes dybala lucas once more wins it back now immobile opens up for correa szczęsny got behind it on an awkward surface and held it well juve have a tricky game against atalanta in the cup tricky moments here as well as it falls for luiz alberto szczęsny with the save and lucas with a miss hit heres correa misskick there from matuidi milinkovićsavić look at the space for parolo fabulous save from szczęsny who has kept juve in it here thats the pick of them the effort levels never drop when hes out there it might bounce here for immobile rugani does so well to recover and get it behind great attacking instinct from immobile much to discuss much to find out no changes at half time they

filename	content	cleaned_content
	<p>\n\n31\n00:02:46,488 --&gt; 00:02:52,822\nOh, great skill from Bernardeschi.\nFor Dybala, spilled, big chance. \n\n32\n00:02:52,820 --&gt; 00:02:58,444\nJoão Cancelo with an equalizer for Juventus. \n\n33\n00:03:01,822 --&gt; 00:03:04,066\nBack from injury to save them. \n\n34\n00:03:05,000 --&gt; 00:03:07,755\nThe substitute makes it 1-1.\n\n35\n00:03:10,755 --&gt; 00:03:13,711\nHold on to this or even better it \nand they might just do that. \n\n36\n00:03:13,710 --&gt; 00:03:15,800\nRonaldo in for Bernardeschi. \n\n37\n00:03:16,311 --&gt; 00:03:19,400\nJoão Cancelo went over on\nthe challenge from Lulić. \n\n38\n00:03:19,600 --&gt; 00:03:20,977\nHe's given a penalty. \n\n39\n00:03:21,911 --&gt; 00:03:24,844\nMarco Guida has given the penalty. \n\n40\n00:03:25,555 --&gt; 00:03:30,155\nThey were under massive pressure. \nCan they turn it round? They can.\n\n41\n00:03:36,088 --&gt; 00:03:40,844\nIt's what champions do on the ropes.\n\n42\n00:03:43,844 --&gt; 00:03:46,822\nSergej Milinković-Savić, over the bar. \n\n43\n00:03:46,820 --&gt; 00:03:50,622\nJuventus can celebrate,\n can punch the air.\n\n44\n00:03:50,777 --&gt; 00:03:54,222\nJuve win it by two goals to one.\n\n</p>	need something to show for their efforts in from luiz alberto and they have it did it ricochet off milinkovićsavić is it an owngoal no one at the olmpico could care lazio in front juve in trouble in the rain in rome now correa oh immobile is in with a chance to make it two forward by alex sandro oh great skill from bernardeschi for dybala spilled big chance joão cancelo with an equalizer for juventus back from injury to save them the substitute makes it hold on to this or even better it and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a penalty marco guida has given the penalty they were under massive pressure can they turn it round they can its what champions do on the ropes sergej milinkovićsavić over the bar juventus can celebrate can punch the air juve win it by two goals to one

	filename	content	cleaned_content
1	oCI6MCoOj1Q.mp4.srt	<p>1 n00:00:00,484 --&gt; 00:00:02,841 nHello and welcome nto the Stadio Bentegodi. n n2 n00:00:03,009 --&gt; 00:00:06,176 nwhere this afternoon nChievo Verona and Frosinone n n3 n00:00:06,344 --&gt; 00:00:09,462 ngo head-to-head nin round 19 of Serie A. n n4 n00:00:09,630 --&gt; 00:00:11,730 nl'm Marco Palmieri. nGreat to be with you. n n5 n00:00:11,908 --&gt; 00:00:14,463 nlt's a must-win match for both sides. n n6 n00:00:14,726 --&gt; 00:00:18,136 nWe're away at the Bentegodi nbetween Chievo and Frosinone. n n7 n00:00:18,574 --&gt; 00:00:19,774 nThe corners are coming, n n8 n00:00:20,159 --&gt; 00:00:21,409 ncan they take advantage? n n9 n00:00:21,945 --&gt; 00:00:24,995 nThe big men in the area, nas Birsa steps up to take the corner. n n10 h00:00:25,739 --&gt; 00:00:26,794 nlt is indeed. n n11 h00:00:27,921 --&gt; 00:00:32,238 nSergio Pellissier nwith a header from the corner n n12 h00:00:34,619 --&gt; 00:00:37,761 nand the captain does it again nfor Chievo. n n13 h00:00:45,111 --&gt; 00:00:46,111 nMaking sure n n14 h00:00:46,279 --&gt; 00:00:48,679 nthat there weren't any irregularities nin the box, n n15 h00:00:49,400 --&gt; 00:00:51,253 nmaybe a push. n n16 h00:00:52,539 --&gt; 00:00:54,699 nlt will be a on-field review. n n17 h00:01:04,585 --&gt; 00:01:06,290 nThe goal has been disallowed. n n18 h00:01:06,703 --&gt; 00:01:09,344 nlt's still 0-0 nhere at the Bentegodi. n n19 h00:01:11,495 --&gt; 00:01:13,184 nPellissier is furious. n n20 h00:01:14,073 --&gt; 00:01:15,073 nA push, n n21 h00:01:15,241 --&gt; 00:01:17,341 nl'm assuming nthere was a push by Pellissier n n22 h00:01:17,509 --&gt; 00:01:19,036 nas he lept up into the air. n n23 h00:01:19,397 --&gt; 00:01:21,941 nAs Rocchi gets set to start nthe second half n n24 h00:01:22,109 --&gt; 00:01:23,109 nand we're away. n n25 h00:01:26,478 --&gt; 00:01:28,702 nAs Hetemaj gets it to Birsa. n n26 h00:01:32,067 --&gt; 00:01:34,061 nHe skips past Maiello. n n27 h00:01:34,252 --&gt; 00:01:35,530 nNow, Đorđević, n n28 h00:01:35,698 --&gt; 00:01:37,118 nthe first shot on target, n n29 h00:01:38,515 --&gt; 00:01:41,476 nbut Sportiello denies him nwith a quick save, n n30 h00:01:42,449 --&gt; 00:01:43,999 nnot necessarily a difficult</p>	<p>hello and welcome to the stadio bentegodi where this afternoon chievo verona and frosinone go headtohead in round of serie a im marco palmieri great to be with you its a mustwin match for both sides were away at the bentegodi between chievo and frosinone the corners are coming can they take advantage the big men in the area as birsa steps up to take the corner it is indeed sergio pellissier with a header from the corner and the captain does it again for chievo making sure that there werent any irregularities in the box maybe a push it will be a onfield review the goal has been disallowed its still here at the bentegodi pellissier is furious a push im assuming there was a push by pellissier as he lept up into the air as rocchi gets set to start the second half and were away as hetemaj gets it to birsa he skips past maiello now đorđević the first shot on target but sportiello denies him with a quick save not necessarily a</p>

filename	content	cleaned_content
	<p>one\n\n31\n00:01:44,167 --&gt; 00:01:45,917\nbut one that had to be made\n\n32\n00:01:46,553 -&gt; 00:01:49,746\nas the ball found its way\nonto the Serb striker.\n\n33\n00:01:56,255 --&gt; 00:01:57,442\nCorner taken short.\n\n34\n00:01:58,268 --&gt; 00:02:00,318\nAs the ball at the far post\nfor Chibsah.\n\n35\n00:02:00,792 --&gt; 00:02:02,529\nBrilliant save from Sorrentino.\n\n36\n00:02:04,704 --&gt; 00:02:07,504\nThe Frosinone players thought\nthat it'd crossed the line.\n\n37\n00:02:08,228 --&gt; 00:02:12,538\nThe watch on referee Rocchi\nndid not signal a goal.\n\n38\n00:02:12,706 --&gt; 00:02:14,012\nand play continues.\n\n39\n00:02:17,279 --&gt; 00:02:19,187\nAs now Depaoli is in space\n\n40\n00:02:22,155 --&gt; 00:02:24,505\nand he's tripped\non the edge of the area here.\n\n41\n00:02:25,095 --&gt; 00:02:26,845\nThere will be a foul and a yellow.\n\n42\n00:02:27,956 --&gt; 00:02:30,192\nlt's a second yellow,\nCapuano is off.\n\n43\n00:02:32,798 --&gt; 00:02:34,966\nFrosinone reduced to 10 men.\n\n44\n00:02:43,927 --&gt; 00:02:45,892\nHetenmaj sees the overlap now.\n\n45\n00:02:46,798 --&gt; 00:02:48,532\nPlay inside, Pellissier.\n\n46\n00:02:48,890 --&gt; 00:02:51,932\nOn his left foot,\nhe slipped over as he took the shot.\n\n47\n00:02:54,193 --&gt; 00:02:55,393\nGreat chance for Chievo.\n\n48\n00:02:57,253 --&gt; 00:02:58,803\nas they continue to attack now.\n\n49\n00:02:58,971 --&gt; 00:03:02,219\nTomović, spreads it across\nRadovanović.\n\n50\n00:03:02,831 --&gt; 00:03:04,344\nNow Bani plays it in the box\n\n51\n00:03:04,512 --&gt; 00:03:06,021\nbut Depaoli cannot reach it\n\n52\n00:03:06,506 --&gt; 00:03:08,506\nand the ball dribbles back\nonto the keeper.\n\n53\n00:03:10,892 --&gt; 00:03:12,757\nWhat a chance for Pellissier here.\n\n54\n00:03:18,695 --&gt; 00:03:21,506\nThe shot comes in\nand that's a wonderful goal.\n\n55\n00:03:21,674 --&gt; 00:03:25,926\nGiaccherini gives Chievo the lead\n\n56\n00:03:26,125 --&gt; 00:03:27,756\nwith a superb free-kick.\n\n57\n00:03:28,167 --&gt; 00:03:30,209\nSportiello rooted to the spot.\n\n58\n00:03:32,392 --&gt; 00:03:35,565\nAn excellent goal\nand thoroughly deserved.\n\n59\n00:03:38,248 --&gt; 00:03:40,175\nThere it is, the final whistle.\n\n60\n00:03:40,343 --&gt; 00:03:42,836\nA simply perfect\nGiaccherini free-kick.\n\n61\n00:03:43,122 --&gt; 00:03:45,550\nhas given Chievo\ntheir first win of the season.\n\n62\n00:03:45,780 --&gt; 00:03:47,392\nin the last match of the year.\n\n63\n00:03:47,647 --&gt; 00:03:51,108\nThey've defeated a 10-man Frosinone\nby a one goal to nil.\n</p>	<p>difficult one but one that had to be made as the ball found its way to the serb striker corner taken short as the ball at the far post for chibsah brilliant save from sorrentino the frosinone players thought that itd crossed the line the watch on referee rocchi did not signal a goal and play continues as now depaoli is in space and hes tripped on the edge of the area here there will be a foul and a yellow its a second yellow capuano is off frosinone reduced to men hetemaj sees the overlap now play inside pellissier on his left foot he slipped over as he took the shot great chance for chievo as they continue to attack now tomović spreads it across to radovanović now bani plays it in the box but depaoli cannot reach it and the ball dribbles back to the keeper what a chance for pellissier here the shot comes in and that's a wonderful goal giaccherini gives chievo the lead on minutes with a superb freekick sportiello rooted to the spot an excellent goal and thoroughly deserved there it is the final whistle a simply perfect giaccherini freekick has</p>

filename	content	cleaned_content
		given chievo their first win of the season in the last match of the year theyve defeated a man frosinone by a one goal to nill
2 _kDc62-upoA.mp4.srt	<p>1\n00:00:00,389 --&gt; 00:00:03,488\n-The league's top scorers,\nagainst Seria A's weakest frontline.\n\n00:00:03,862 --&gt; 00:00:07,576\nAtalanta are bidding to make 2020\nthe best year in club history.\n\n00:00:07,847 --&gt; 00:00:09,927\nSpal are four points from safety\n\n00:00:10,205 --&gt; 00:00:12,883\nand running out of time\nin their efforts to beat the drop.\n\n00:00:16,606 --&gt; 00:00:17,964\nDi Francesco in the channel,\n\n00:00:19,401 --&gt; 00:00:20,643\nable to wander inside.\n\n00:00:22,965 --&gt; 00:00:25,139\nGabriel Strefezza lining it up.\n\n00:00:25,307 --&gt; 00:00:27,307\nIt's a good save from Sportiello.\n\n00:00:34,144 --&gt; 00:00:37,996\nThe game becoming stretched\njust like Atalanta want it to be.\n\n00:00:39,648 --&gt; 00:00:41,506\nZapata running at Thiago Cionek.\n\n00:00:43,006 --&gt; 00:00:45,935\nDuván Zapata's cross,\nit's in for Ilicic.\n\n00:00:48,166 --&gt; 00:00:51,321\nA difficult finish\nmade to look all so easy.\n\n00:00:52,973 --&gt; 00:00:56,670\nZapata still seeking his first goal\nsince his return to action.\n\n00:00:56,955 --&gt; 00:00:58,648\nThat was a telling contribution.\n\n00:00:59,990 --&gt; 00:01:03,751\nIlicic has scored\nagainst Fiorentina in midweek.\n\n00:01:04,792 --&gt; 00:01:07,815\ncontinues his wonderful start to 2020.\n\n00:01:13,191 --&gt; 00:01:14,586\nWell in by Gosens.\n\n00:01:15,643 --&gt; 00:01:16,784\nThis is</p>	the leagues top scorers against seria as weakest frontline atalanta are bidding to make the best year in club history spal are four points from safety and running out of time in their efforts to beat the drop di francesco in the channel able to wander inside gabriel strefezza lining it up its a good save from sportiello the game becoming stretched just like atalanta want it to be zapata running at thiago cionek duván zapatas cross its in for ilicic a difficult finish made to look all so easy zapata still seeking his first goal since his return to action that was a telling contribution ilicic has scored

filename	content	cleaned_content
	<p>Pasalic\n n 19\n00:01:17,293 --&gt; 00:01:18,973\nplaying slightly deeper this evening.\n n 20\n00:01:19,373 --&gt; 00:01:20,407\nNow Zapata,\n n 21\n00:01:21,115 --&gt; 00:01:22,965\nplenty of room for him to get moving.\n n 22\n00:01:23,871 --&gt; 00:01:25,010\nUp against Vicari,\n n 23\n00:01:25,232 --&gt; 00:01:26,797\nDuván Zapata with the shot.\n n 24\n00:01:27,312 --&gt; 00:01:28,590\nIt's off the frame of the goal.\n n 25\n00:01:40,444 --&gt; 00:01:41,461\nThis is Valoti.\n n 26\n00:01:41,827 --&gt; 00:01:43,213\nThiago Cionek joining in.\n n 27\n00:01:43,920 --&gt; 00:01:45,111\nThe cross is only half-cleared.\n n 28\n00:01:45,279 --&gt; 00:01:46,279\nStrafezza.\n n 29\n00:01:48,239 --&gt; 00:01:49,239\nCionek with the ball in.\n n 30\n00:01:49,407 --&gt; 00:01:50,407\nSportiello stays at home.\n n 31\n00:01:50,575 --&gt; 00:01:53,137\nPetagna's header is off the post.\n n 32\n00:01:54,066 --&gt; 00:01:57,295\nSo close to yet another goal\nagainst his former club.\n n 33\n00:02:07,298 --&gt; 00:02:08,924\nThat's a nice touch\nfrom Di Francesco.\n n 34\n00:02:09,092 --&gt; 00:02:10,148\nReca is very quick.\n n 35\n00:02:10,316 --&gt; 00:02:12,005\nHe's trying to get\naway from Djimsiti.\n n 36\n00:02:12,158 --&gt; 00:02:13,467\nThis is Reca's cross,\n n 37\n00:02:13,635 --&gt; 00:02:15,429\nand it's put in by Petagna.\n n 38\n00:02:17,302 --&gt; 00:02:20,080\nA couple of former\nplayers combine,\n n 39\n00:02:20,248 --&gt; 00:02:21,279\nReca on loan.\n n 40\n00:02:21,699 --&gt; 00:02:24,644\nPetagna once again\nscores against La Dea.\n n 41\n00:02:25,303 --&gt; 00:02:27,328\nNo one could catch up with the ball.\n n 42\n00:02:32,296 --&gt; 00:02:33,296\nDjimsiti.\n n 43\n00:02:36,408 --&gt; 00:02:38,837\nZapata, a foul on him Vicari.\n n 44\n00:02:40,020 --&gt; 00:02:41,020\nDabo,\n n 45\n00:02:42,440 --&gt; 00:02:43,440\nValoti.\n n 46\n00:02:44,846 --&gt; 00:02:47,100\nHe's kept it.\nMattia Valoti.\n n 47\n00:02:47,306 --&gt; 00:02:48,520\nSpal have another.\n n 48\n00:02:49,482 --&gt; 00:02:52,570\nThey've come from behind here\ninto lead in Bergamo.\n n 49\n00:02:53,894 --&gt; 00:02:56,284\nWhat a superb individual strike.\n n 50\n00:02:57,522 --&gt; 00:02:59,690\nPetagna first and now Valoti.\n n 51\n00:03:00,556 --&gt; 00:03:02,921\nHe took his goal splendidly well.\n n 52\n00:03:09,356 --&gt; 00:03:11,040\nLaunched long by Ilicic.\n n 53\n00:03:12,230 --&gt; 00:03:14,114\nMuriel gives chase, he's got there.\n n 54\n00:03:14,282 --&gt; 00:03:15,282\nThis is</p>	<p>against fiorentina in midweek continues his wonderful start to i well in by gosens this is pasalic playing slightly deeper this evening now zapata plenty of room for him to get moving up against vicari duván zapata with the shot its off the frame of the goal this is valoti thiago cionek joining in the cross is only halfcleared \\ strafezza cionek with the ball in sportiello stays at home petagnas header is off the post so close to yet another goal against his former club thats   a nice touch from di francesco reca is very quick hes trying to get away from djimsiti this is recas cross and its put in by petagna a couple of former atalanta players combine reca on loan petagna once again scores against la dea no one could catch up with the ball   djimsiti zapata a foul on him vicari dabo valoti hes kept it mattia valoti spal have another theyve come from behind here to lead in bergamo what a superb individual strike petagna first and now valoti he took his goal splendidly well launched long by ilicic muriel gives chase hes got there this is zapata duvan zapata its deflected terrific  </p>

filename	content	cleaned_content
	Zapata.\n\n55\n00:03:15,450 --> 00:03:17,217\nDuvan Zapata, it's deflected.\n\n56\n00:03:17,589 --> 00:03:19,653\nTerrific save from Berisha.\n\n57\n00:03:20,019 --> 00:03:21,153\nThat was creeping in.\n\n58\n00:03:23,401 --> 00:03:25,590\nAnother look\nat that long ball from Ilicic.\n\n59\n00:03:25,758 --> 00:03:26,758\nMuriel did well.\n\n60\n00:03:28,632 --> 00:03:30,378\nZapata struck it firmly.\n\n61\n00:03:34,600 --> 00:03:36,981\nSpal win in Bergamo.\n\n62\n00:03:37,149 --> 00:03:39,276\nWhat a turn around for Semplici,\n\n63\n00:03:39,538 --> 00:03:43,103\nonly a fourth victory of the season\nfor Seria A's bottom side.\n\n64\n00:03:43,524 --> 00:03:48,548\nThree massive points in their quest\nto retain their topflight status.\n\n65\n00:03:49,018 --> 00:03:50,454\nAtalanta off-color\n\n66\n00:03:50,637 --> 00:03:52,637\nand out of sorts they stay fifth.\n\n67\n00:03:55,074 --> 00:04:02,074\n[crowd cheering]\n\n68\n00:04:02,242 --> 00:04:09,242\n[crowd cheering]\n\n69\n00:04:09,410 --> 00:04:13,126\n[crowd cheering]\n\n	save from berisha that was creeping in another look at that long ball from ilicic muriel did well zapata struck it firmly spal win in bergamo what a turn around for semplici only a fourth victory of the season for seria as bottom side three massive points in their quest to retain their topflight status atalanta offcolor and out of sorts they stay fifth crowd cheering crowd cheering crowd cheering

3	41kO2TvmNbg.mp4.srt	1\n00:00:00,500 --> 00:00:03,566\nWelcome to the Ennio Tardini stadium.\n\n2\n00:00:04,133 --> 00:00:04,733\nThis week, \n\n3\n00:00:04,733 --> 00:00:06,599\nit's the meeting of the two sides\n\n4\n00:00:06,600 --> 00:00:09,166\nwho are the bane of the City of Milan\n\n5\n00:00:09,166 --> 00:00:10,799\nin the last match week: \n\n6\n00:00:10,800 --> 00:00:13,133\nParma against Cagliari.\n\n7\n00:00:13,133 --> 00:00:14,433\nLast weekend, \n\n8\n00:00:14,433 --> 00:00:17,233\nParma recorded a surprise victory\n\n9\n00:00:17,233 --> 00:00:19,799\nat San Siro against Inter,\n\n10\n00:00:19,800 --> 00:00:23,500\nwhile Cagliari got an impressive draw\n\n11\n00:00:23,500 --> 00:00:25,566\nagainst Milan in Sardinia.\n\n12\n00:00:26,300 --> 00:00:27,333\nWe're underway. \n\n13\n00:00:30,500 --> 00:00:32,933\nStolen and now Gervinho\n\n14\n00:00:34,266 --> 00:00:37,132\nshow us that famous sprinting power.\n\n15\n00:00:37,133 --> 00:00:37,866\nStill	welcome to the ennio tardini stadium this week its the meeting of the two sides who are the bane of the city of milan in the last match week parma against cagliari last weekend parma recorded a surprise victory at san siro against inter while cagliari got an impressive draw against milan in sardinia were underway stolen and now gervinho can show us that famous sprinting power still
---	---------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

filename	content	cleaned_content
	<p>going. \n\n16\n00:00:37,866 --&gt; 00:00:39,066\nGervinho still going, \n\n17\n00:00:39,066 --&gt; 00:00:40,166\ngets the shot off \n\n18\n00:00:40,833 --&gt; 00:00:42,633\njust wide of the post, \n\n19\n00:00:43,466 --&gt; 00:00:45,932\ntrademark run from Gervinho.\n\n20\n00:00:47,900 --&gt; 00:00:50,200\nLook where he picks the ball up in his own half. \n\n21\n00:00:50,900 --&gt; 00:00:53,166\nThen he goes on an angled run inside, \n\n22\n00:00:54,266 --&gt; 00:00:55,832\ngets past Klavan, \n\n23\n00:00:55,833 --&gt; 00:00:57,266\nsees off Ionită \n\n24\n00:00:57,266 --&gt; 00:00:58,366\nand a low shot. \n\n25\n00:00:58,366 --&gt; 00:01:00,032\nYou can see what he was going for there. \n\n26\n00:01:01,966 --&gt; 00:01:02,732\nDimarco, \n\n27\n00:01:06,933 --&gt; 00:01:07,966\nInglese is in there again. \n\n28\n00:01:07,966 --&gt; 00:01:10,332\nInglese loops it over the keeper, \n\n29\n00:01:10,333 --&gt; 00:01:11,366\nCragno. \n\n30\n00:01:11,366 --&gt; 00:01:12,632\nThis time, \n\n31\n00:01:12,633 --&gt; 00:01:13,999\nhas the goal been given? \n\n32\n00:01:14,000 --&gt; 00:01:15,333\nIt looks like it has. \n\n33\n00:01:16,266 --&gt; 00:01:19,132\nDenied by the offside flag a few minutes ago.\n\n34\n00:01:19,133 --&gt; 00:01:23,199\nThis time, Roberto Inglese is off the mark \n\n35\n00:01:23,200 --&gt; 00:01:24,133\nin the match, \n\n36\n00:01:24,866 --&gt; 00:01:26,499\ngiving Parma the lead.\n\n37\n00:01:30,233 --&gt; 00:01:31,233\nExcellent run. \n\n38\n00:01:32,633 --&gt; 00:01:34,233\nInglese persistent, \n\n39\n00:01:34,866 --&gt; 00:01:37,599\nnot lucky there from the deflection off Cragno. \n\n40\n00:01:39,333 --&gt; 00:01:41,499\nHis shot cannoned off the goalkeeper, \n\n41\n00:01:41,500 --&gt; 00:01:43,933\nback into his face and into the net. \n\n42\n00:01:45,200 --&gt; 00:01:46,266\nThere's a couple of checks, \n\n43\n00:01:46,266 --&gt; 00:01:47,799\nReferee Calvarese. \n\n44\n00:01:49,633 --&gt; 00:01:51,166\nThe goal has been given. \n\n45\n00:01:51,166 --&gt; 00:01:53,332\nParma do go into the lead. \n\n46\n00:01:55,600 --&gt; 00:01:56,566\nLuca Rigoni, \n\n47\n00:01:57,133 --&gt; 00:01:57,899\nDi Gaudio. \n\n48\n00:01:59,966 --&gt; 00:02:00,732\nWill he go for the cross? \n\n49\n00:02:00,733 --&gt; 00:02:01,833\nHe plays it back. \n\n50\n00:02:01,833 --&gt; 00:02:03,033\nIt's a shot from \n\n51\n00:02:05,300 --&gt; 00:02:06,766\nAntonino Barillà. \n\n52\n00:02:06,766 --&gt; 00:02:08,032\nThe ball is still alive. \n\n53\n00:02:08,033 --&gt; 00:02:08,899\nGervinho \n\n54\n00:02:09,666 --&gt; 00:02:12,499\nGervinho pulls it back into the area.\n\n55\n00:02:12,500 --&gt; 00:02:13,900\nIt's Luca Rigoni, \n\n56\n00:02:13,900 --&gt; 00:02:15,200\nna</p>	<p>power still going gervinho still going gets the shot off just wide of the post trademark run from gervinho look where he picks the ball up in his own half then he goes on an angled run inside gets past klavan sees off ionită and a low shot you can see what he was going for there dimarco inglese is in there again inglese loops it over the keeper cragno this time has the goal been given it looks like it has denied by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma the lead excellent run inglese persistent got lucky there from the deflection off cragno his shot cannoned off the goalkeeper back into his face and into the net theres a couple of checks referee calvarese the goal has been given parma do go into the lead luca rigoni di gaudio will he go for the cross he plays it back its a shot from antonino barillà the ball is still alive gervinho gervinho pulls it back into the area its luca rigoni a familiar sight the raiding midfielder the second half starts marco sau goes down referee says play on now now</p>

filename	content	cleaned_content
DjNA-MIdCbY.mp4.srt	familiar sight, \n\n57\n00:02:16,633 --> 00:02:17,966\nthe raiding midfielder.\n\n58\n00:02:21,400 --> 00:02:22,966\nThe second half starts. \n\n59\n00:02:24,300 --> 00:02:25,433\nMarco Sau goes down.\n\n60\n00:02:25,433 --> 00:02:26,866\nReferee says, "Play on now." \n\n61\n00:02:26,866 --> 00:02:29,999\nNow Gervinho on one of those runs.\n\n62\n00:02:30,000 --> 00:02:32,466\nGervinho still going.\n\n63\n00:02:33,166 --> 00:02:35,366\nGervinho past another.\n\n64\n00:02:35,366 --> 00:02:36,699\nWhat's the shot like?\n\n65\n00:02:36,700 --> 00:02:39,666\nIt's a fantastic shot. \n\n66\n00:02:39,666 --> 00:02:41,099\nGervinho. \n\n67\n00:02:41,100 --> 00:02:42,666\nWhat a goal.\n\n68\n00:02:47,600 --> 00:02:50,566\nA standing ovation from the Tardini.\n\n69\n00:02:51,800 --> 00:02:53,233\nDanger at one end.\n\n70\n00:02:54,200 --> 00:02:56,100\nGervinho picked the ball up\n\n71\n00:02:56,100 --> 00:02:59,566\nand then showed his genius.\n\n72\n00:03:07,066 --> 00:03:09,266\nA rare lack of control from Srna\n\n73\n00:03:10,766 --> 00:03:12,699\nbut tenacity getting the ball back,\n\n74\n00:03:13,433 --> 00:03:14,433\nplays it in. \n\n75\n00:03:14,433 --> 00:03:16,966\nAn attempted backflip by Marco Sau.\n\n76\n00:03:19,666 --> 00:03:21,466\nHad there been a foul in the build-up? \n\n77\n00:03:31,033 --> 00:03:32,533\nWon back by Parma.\n\n78\n00:03:32,533 --> 00:03:35,033\nThe break is on into the area. \n\n79\n00:03:35,033 --> 00:03:36,866\nCeravolo gets it back.\n\n80\n00:03:36,866 --> 00:03:38,066\nWill he make it three? \n\n81\n00:03:38,666 --> 00:03:40,799\nGreat save by Cragno.\n\n82\n00:03:46,933 --> 00:03:49,599\nReferee Juan Paolo Calvarese\n\n83\n00:03:49,600 --> 00:03:51,533\nblows the full-time whistle. \n\n84\n00:03:51,533 --> 00:03:53,966\nParma 2-0 Cagliari.\n\n	gervinho on one of those runs gervinho still going gervinho past another whats the shot like its a fantastic shot gervinho what a goal a standing ovation from the tardini danger at one end gervinho picked the ball up and then showed his genius a rare lack of control from srna but tenacity getting the ball back plays it in an attempted backflip by marco sau had there been a foul in the buildup won back by parma the break is on into the area ceravolo gets it back will he make it three great save by cragno referee juan paolo calvarese blows the fulltime whistle parma cagliari

4 DjNA-MIdCbY.mp4.srt

1\n00:00:00,000 --> 00:00:04,064\n[music]\n\n2\n00:00:08,104 --> 00:00:09,480\nChilwell.\n\n3\n00:00:10,840 --> 00:00:12,696\nOh, and it's fired in in.\n\n4\n00:00:12,776 --> 00:00:15,360\nTypical Oxlade Chamberlain style.\n\n5\n00:00:16,256 --> 00:00:18,360\nHe is a goal threat.\n\n6\n00:00:19,104 --> 00:00:20,776\nHe really didn't catch it well.\n\n7\n00:00:20,816 --> 00:00:22,904\nIt's another lovely ball.\n\n8\n00:00:24,840 --> 00:00:26,920\nIt's not the easiest of chances.\n\n9\n00:00:27,576 --> 00:00:29,136\nWhen you strike it that

music chilwell oh and its fired in in typical oxlade chamberlain style he is a goal threat he really didnt catch it well its another lovely ball its not the easiest of chances when you strike it that purely theres no chance for the

filename	content	cleaned_content
	<p>purely,\n\n10\n00:00:29,160 --&gt; 00:00:31,368\nthere's no chance for the keeper.\n\n11\n00:00:33,448 --&gt; 00:00:36,104\nThey're stretched defensively here,\n\n12\n00:00:36,264 --&gt; 00:00:39,624\nHarry Kane just makes that point.\n\n13\n00:00:40,712 --&gt; 00:00:42,592\nOnly needed a touch.\n\n14\n00:00:43,232 --&gt; 00:00:45,480\nFloated in by Chilwell.\n\n15\n00:00:45,632 --&gt; 00:00:47,056\namongst a clutch of players,\n\n16\n00:00:47,112 --&gt; 00:00:48,408\nbut he manages just to\n\n17\n00:00:49,100 --&gt; 00:00:51,874\nkeep his opponents at arm's length.\n\n18\n00:00:52,666 --&gt; 00:00:54,874\nChilwell to take the corner.\n\n19\n00:00:57,874 --&gt; 00:01:00,888\nThat's a free header. That's another goal.\n\n20\n00:01:01,137 --&gt; 00:01:02,488\nSimple for Kane.\n\n21\n00:01:02,835 --&gt; 00:01:04,702\nEngland incidentally are hot\n\n22\n00:01:05,057 --&gt; 00:01:07,057\nshowing their array of set pieces,\n\n23\n00:01:07,128 --&gt; 00:01:08,960\nthey want to sail to the finals.\n\n24\n00:01:10,168 --&gt; 00:01:11,048\nOn the play,\n\n25\n00:01:11,050 --&gt; 00:01:13,857\nnobody knows what those Montenegro\n\n26\n00:01:14,204 --&gt; 00:01:16,151\ndefenders were doing here\n\n27\n00:01:16,240 --&gt; 00:01:18,311\nallow Kane a free run.\n\n28\n00:01:18,360 --&gt; 00:01:19,352\nThey've got to get a bit more\n\n29\n00:01:19,396 --&gt; 00:01:22,283\nphysical contact and so easy.\n\n30\n00:01:28,622 --&gt; 00:01:32,915\nRashford gets in on the act.\n\n31\n00:01:36,833 --&gt; 00:01:38,333\nWell that's four.\n\n32\n00:01:39,860 --&gt; 00:01:42,990\nOxlade Chamberlain with a really good ball.\n\n33\n00:01:43,030 --&gt; 00:01:45,712\nMaguire can't believe he hasn't scored there.\n\n34\n00:01:45,844 --&gt; 00:01:48,712\nRashford tidies up.\n\n35\n00:01:54,627 --&gt; 00:01:55,526\nSancho.\n\n36\n00:01:58,603 --&gt; 00:02:00,295\nHere's the hat-trick.\n\n37\n00:02:01,784 --&gt; 00:02:03,495\nAnother Harry hat-trick.\n\n38\n00:02:03,846 --&gt; 00:02:05,200\nFive for England.\n\n39\n00:02:06,092 --&gt; 00:02:07,821\nThis landmark evening\n\n40\n00:02:07,889 --&gt; 00:02:09,538\ncould not be going better.\n\n41\n00:02:10,030 --&gt; 00:02:12,941\nIt wasn't the easiest ball to control.\n\n42\n00:02:13,040 --&gt; 00:02:15,825\nfirst time from Alexander-Arnold.\n\n43\n00:02:15,905 --&gt; 00:02:18,080\njust bouncing in front of him.\n[laugh]\n\n44\n00:02:18,080 --&gt; 00:02:20,821\nHe's such a confident frame of mind.\n\n45\n00:02:21,825 --&gt; 00:02:23,820\nthat he just decides to take a touch.\n\n46\n00:02:25,629 --&gt;</p>	<p>keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in by chilwell amongst a clutch of players but he manages just to keep his opponents at arms length chilwell to take the corner thats a free header thats another goal simple for kane england incidentally are not showing their array of set pieces they want to sail to the finals on the play nobody knows what those montenegro defenders were doing here to allow kane a free run theyve got to get a bit more physical contact and so easy rashford gets in on the act well thats four oxlade chamberlain with a really good ball maguire cant believe he hasnt scored there rashford tidies up sancho heres the hattrick another harrtack five for england this landmark evening could not be going better it wasnt the easiest ball to control first time from alexanderarnold just bouncing in front of him laugh hes such a confident frame of mind that he just decides to take a touch it proves to be the right decision rashfords in behind wow how has it not gone in</p>

filename	content	cleaned_content
	<p>00:02:27,905\nlt proves to be the right decision.\n\n47 h00:02:33,672 --&gt;      00:02:35,512 Rashford's in behind.\n\n48 h00:02:36,887 --&gt;      00:02:38,298 nWow, how has it not gone in?\n\n49 h00:02:38,400 --&gt; 00:02:39,789 nIn the end, it has.\n\n50 h00:02:42,429 --&gt;      00:02:45,963 nlt looked like the defender rushing back.\n\n51 h00:02:46,429 --&gt;      00:02:48,210 nl see he got the final touch.\n\n52 h00:02:49,330 --&gt;      00:02:51,687 Rashford's done this several times tonight\n\n53 h00:02:51,752 --&gt;      00:02:53,374 njust easing past his man.\n\n54 h00:02:55,512 --&gt;      00:02:59,469 nAn instinctive reaction by Šofranac.\n\n55 h00:03:06,501 --&gt;      00:03:08,152 nHe might get a tap in this time.\n\n56 h00:03:08,283 --&gt;      00:03:11,447 nTammy Abraham opens his England account.\n\n57 h00:03:11,636 --&gt;      00:03:14,447 nit is seven with seven minutes to go.\n\n58 h00:03:14,698 --&gt;      00:03:18,000 nYes, a big moment for him and a classic\n\n59 h00:03:18,058 --&gt;      00:03:19,560 ncenter forward's goal this one.\n\n60 h00:03:19,610 --&gt;      00:03:21,229 nMaking that dash across your man.\n\n61 h00:03:21,323 --&gt;      00:03:22,800 nsticking out a leg.\n\n62 h00:03:23,650 --&gt;      00:03:25,432 nWonderful stuff from Abraham.\n\n63 h00:03:25,898 --&gt;      00:03:27,512 nIn an ideal world\n\n64 h00:03:28,050 --&gt;      00:03:29,738 nhe could get on the score sheet.\n\n65 h00:03:29,898 --&gt;      00:03:31,781 nget him some belief at this level.\n\n66 h00:03:32,676 --&gt;      00:03:34,781 nlt finishes England seven.\n\n67 h00:03:34,780 --&gt;      00:03:36,080 nMontenegro nil.\n\n68 h00:03:36,174 --&gt;      00:03:38,160 nThey qualify for the finals.\n\n69 h00:03:41,665 --&gt;      00:03:45,883 n[music]\n\n</p>	<p>in the end it has it looked like the defender rushing back i see he got the final touch rashfords done this several times tonight just easing past his man an instinctive reaction by Šofranac he might get a tap in this time tammy abraham opens his england account it is seven with seven minutes to go yes a big moment for him and a classic center forwards goal this one j making that dash across your man sticking out a leg wonderful stuff from abraham in an ideal world he could get on the score sheet get him some belief at this level it finishes england seven montenegro nil they qualify for the finals music</p>

1118 dJMNNQHkz0M.mp4.srt

1|n00:00:02,068 --> 00:00:04,994|nMatch begins with Torino very much all won \n\n2|h00:00:05,085 --> 00:00:09,691|nfive consecutive clean sheets victories\n over Inter and last week, Atalanta. \n\n3|h00:00:09,845 --> 00:00:13,011|nToday begins a trio of fixtures\n against the relegation side. \n\n4|h00:00:13,188

match begins with torino very much all won five consecutive clean sheets victories over inter and last

filename	content	cleaned_content
	--> 00:00:15,600\nChievo are expected to be beaten.\n\n5\n00:00:16,365 --> 00:00:17,720\nThe first whistle goes.\n\n6\n00:00:17,800 --> 00:00:20,605\nIt is Chievo in the yellow\n who kick us off.\n\n7\n00:00:22,125 --> 00:00:24,908\nThe cross is in.\n Whipped in to De Silvestri.\n\n8\n00:00:25,451 --> 00:00:28,308\nThe shot from range.\n Ansaldi converts.\n\n9\n00:00:28,388 --> 00:00:33,485\n[cheering]\n\n10\n00:00:33,577 --> 00:00:35,908\nTwo goals for him this season,\n the Argentinian.\n\n11\n00:00:36,005 --> 00:00:39,194\n[cheering]\n\n12\n00:00:39,490 --> 00:00:44,880\nNow, a round of applause and a pause\n for the memory of Davide Astori,\n\n13\n00:00:45,040 --> 00:00:51,000\nthe former Fiorentina player\n who so tragically lost his life last year.\n\n14\n00:00:51,091 --> 00:00:56,108\n[applause]\n\n15\n00:00:56,217 --> 00:00:58,697\n[cheering]\n\n16\n00:00:58,737 --> 00:00:59,765\nThis is better now.\n\n17\n00:00:59,828 --> 00:01:04,908\nAnsaldi, Meíté who wasn't wise enough\n to trick Diousse.\n\n18\n00:01:05,011 --> 00:01:07,611\n[cheering]\n\n19\n00:01:07,708 -> 00:01:11,022\nMeggiorini, Đorđević to his right.\n He has a go at goal.\n\n20\n00:01:12,148 --> 00:01:17,400\nA goal that he used to shoot at himself\n when wearing the colors of Torino.\n\n21\n00:01:18,422 --> 00:01:22,102\nPlayed for them between 2012, 2014.\n\n22\n00:01:22,200 --> 00:01:25,908\n[cheering]\n\n23\n00:01:26,262 --> 00:01:28,708\nWe're back underway.\n Torino kick us off.\n\n24\n00:01:28,800 --> 00:01:32,714\n[cheering]\n\n25\n00:01:32,811 --> 00:01:35,897\nDepaoli shifting inside.\n Trying to make the angle for the cross.\n\n26\n00:01:35,960 --> 00:01:38,897\nTook a deflection.\n Should score. Sirigu denies.\n\n27\n00:01:39,651 --> 00:01:42,708\nĐorđević with the best chance\n of the game.\n\n28\n00:01:42,817 --> 00:01:43,994\n[cheering]\n\n29\n00:01:44,194 --> 00:01:46,211\nHe keeps that clean sheet record.\n\n30\n00:01:47,360 --> 00:01:50,308\nIt was an excellent reaction\n save close to him.\n\n31\n00:01:51,137 --> 00:01:54,497\nĐorđević with the second attempt.\n Simply had to score.\n\n32\n00:01:54,594 --> 00:01:57,897\n[cheering]\n\n33\n00:01:57,977 --> 00:01:59,188\n[whistle blows]\n\n34\n00:01:59,662 --> 00:02:02,022\nNow, this could be the moment.\n Zaza.\n\n35\n00:02:02,377 --> 00:02:06,611\nSorrentino, wonderful save.\n\n36\n00:02:06,725 --> 00:02:19,120\n[cheering]\n\n37\n00:02:19,234 --> 00:02:22,800\nBelotti with the strike,\n and Belotti with the missile.\n\n38\n00:02:23,268 --> 00:02:25,000\nTorino take the lead.\n\n39\n00:02:25,154 --> 00:02:28,520\nWhat a vital goal.\n It's the captain's	week atlanta today begins a trio of fixtures against the relegation side chievo are expected to be beaten the first whistle goes it is chievo in the yellow who kick us off the cross is in whipped in to de silvestri the shot from range ansaldi converts cheering two goals for him this season the argentinian cheering now a round of applause and a pause for the memory of davide astori the former fiorentina player who so tragically lost his life last year applause cheering this is better now ansaldi meité who wasnt wise enough to trick diousse cheering meggiorini đorđević to his right he has a go at goal a goal that he used to shoot at himself when wearing the colors of torino played for them between cheering were back underway torino kick us off cheering depaoli shifting inside trying to make the angle for the cross took a deflection should score sirigu denies đorđević with the best chance of the game cheering he keeps that clean sheet record it was an excellent reaction save close to him t đorđević with the second attempt

filename	content	cleaned_content
	<p>goal.\n\n40\n00:02:28,611 --&gt; 00:02:31,017\n[cheering]\n\n41\n00:02:31,097 --&gt; 00:02:35,325\nNumber eight of the season.\n Surely, the best of the lots.</p> <p>\n\n42\n00:02:35,428 --&gt; 00:02:38,400\nA cracker from Andrea Belotti.\n\n43\n00:02:38,508 --&gt; 00:02:44,502\n[cheering]\n\n44\n00:02:44,805 --&gt; 00:02:47,297\nNow, looking to just finish off\n this match. \n\n45\n00:02:47,594 --&gt; 00:02:50,302\nTomás Rincón. Tomás Rincón.\n\n46\n00:02:50,377 --&gt; 00:02:51,205\n[cheering]\n\n47\n00:02:51,297 --&gt; 00:02:53,205\nThat's a thing of beauty.\n\n48\n00:02:53,308 --&gt; 00:02:56,937\n[cheering]\n\n49\n00:02:57,097 --&gt; 00:02:58,565\nHe comes off the bench. \n\n50\n00:02:58,651 --&gt; 00:03:01,708\nHe was supposed to put solidity\n into the midfield. \n\n51\n00:03:01,794 --&gt; 00:03:07,314\nIn the end, he's slid\n a beautiful strike into the top corner. \n\n52\n00:03:07,680 --&gt; 00:03:10,605\nTorino two,\n Chievo nil in injury time.\n\n53\n00:03:10,708 --&gt; 00:03:20,600\n[cheering]\n\n54\n00:03:20,725 --&gt; 00:03:25,005\nZaza, looks for number three\n and in the net it goes. \n\n55\n00:03:25,691 --&gt; 00:03:28,720\nTwo goals in injury time. \n\n56\n00:03:28,931 --&gt; 00:03:31,691\nIt's Torino three, Chievo nil. \n\n57\n00:03:31,788 --&gt; 00:03:36,022\nZaza scores home and away\n against Chievo this season. \n\n58\n00:03:36,651 --&gt; 00:03:40,280\nHe's had to wait since the end of September\n for that second goal.\n\n59\n00:03:40,508 --&gt; 00:03:42,897\nThe final whistle goes. \n\n60\n00:03:42,960 --&gt; 00:03:46,177\nIt finishes here Torino three,\n Chievo nil.\n\n61\n00:03:55,348 --&gt; 00:04:14,410\n[cheering]\n\n</p>	simply had to score cheering whistle blows now this could be the moment zaza sorrentino wonderful save cheering belotti with the strike and belotti with the missile torino take the lead what a vital goal its the captains goal cheering number eight of the season surely the best of the lots a cracker from andrea belotti cheering now looking to just finish off this match tomás rincón tomás rincón cheering thats a thing of beauty cheering he comes off the bench he was supposed to put solidity into the midfield in the end hes slid a beautiful strike into the top corner torino two chievo nil in injury time cheering zaza looks for number three and in the net it goes two goals in injury time its torino three chievo nil zaza scores home and away against chievo this season hes had to wait since the end of september for that second goal the final whistle goes it finishes here torino three chievo nil cheering

filename	content	cleaned_content
1119 xK5R8h7zIn0.mp4.srt	<p>1\n00:00:08,885 --&gt;  00:00:11,055\n[applause]\n-Here is the handball, Michael Oliver\n\n2\n00:00:11,223 --&gt;  00:00:12,303\nwaves it away.\n\n3\n00:00:12,471 --&gt;  00:00:16,510\nHere is VAR trying to give us\nan alternative version of it.\n\n4\n00:00:19,134 --&gt;  00:00:20,650\nThen goes on with Mane.\n\n5\n00:00:24,237 --&gt;  00:00:25,603\nStones can't stop the cross.\n\n6\n00:00:25,937 --&gt;  00:00:27,032\nSalah is coming in.\n\n7\n00:00:29,412 --&gt;  00:00:30,755\nFabinho has a go and what a goal.\n\n8\n00:00:30,938 --&gt;  00:00:32,532\n[applause]\n-What a goal,\n\n9\n00:00:32,700 --&gt;  00:00:33,861\nwhat a start\n\n10\n00:00:35,605 --&gt;  00:00:37,796\nfor Liverpool.\n\n11\n00:00:37,964 --&gt;  00:00:39,368\nIt flew in.\n\n12\n00:00:42,583 --&gt;  00:00:43,813\nWell, there's a VAR check.\n\n13\n00:00:44,241 --&gt;  00:00:45,258\nWhat'd he do though?\n\n14\n00:00:48,337 --&gt;  00:00:49,432\nHere is the handball.\n\n15\n00:00:49,940 --&gt;  00:00:50,940\nThere.\n\n16\n00:00:52,529 --&gt;  00:00:53,656\nIt's going to be a goal.\n\n17\n00:00:53,824 --&gt;  00:00:55,507\nHe's adamant it was a penalty.\n\n18\n00:00:56,960 --&gt;  00:00:58,841\nPoor from Bravo\nbut it's a fantastic shot\n\n19\n00:00:59,009 --&gt;  00:01:00,056\nfrom Fabinho.\n\n20\n00:01:07,501 --&gt;  00:01:08,573\n-Here's Robertson,\n\n21\n00:01:08,914 --&gt;  00:01:10,390\nSalah coming in.\n\n22\n00:01:10,558 --&gt;  00:01:12,236\n[applause]\n-Liverpool\n\n23\n00:01:12,404 --&gt;  00:01:14,491\nand another counter punch.\n\n24\n00:01:16,515 --&gt;  00:01:19,020\nMo Salah drive it in a header.\n\n25\n00:01:19,188 --&gt;  00:01:20,322\n-What a finish.\n\n26\n00:01:23,894 --&gt;  00:01:24,902\nMartin Angelino,\n\n27\n00:01:26,919 --&gt;  00:01:27,979\nAguero,\n\n28\n00:01:28,587 --&gt;  00:01:31,583\nstill going, and still not scoring.\n\n29\n00:01:35,017 --&gt;  00:01:36,326\n-I can see here Aguero,\n\n30\n00:01:36,604 --&gt;  00:01:37,674\na little nutmeg from Fabinho,\n\n31\n00:01:37,842 --&gt;  00:01:39,954\nand then you expect him to do better.\n\n32\n00:01:40,122 --&gt;  00:01:42,415\nYou expect him to set it out into that far corner.\n\n33\n00:01:43,391 --&gt;</p>	applause here is the handball michael oliver waves it away here is var trying to give us an alternative version of it then goes on with mane stones cant stop the cross salah is coming in fabinho has a go and what a goal applause what a goal what a start for liverpool it flew in well theres a var check whtad he do though here is the handball there its going to be a goal hes adamant it was a penalty poor from bravo but its a fantastic shot from fabinho heres robertson salah coming in applause liverpool land another counter punch mo salah drive it in a header what a finish martin angelino aguero still going and still not scoring i can see here aguero a little nutmeg from fabinho and then you expect him to do better you expect him to set it out into that far corner martin its a throwin again all liverpool body is back there martin angelino trying to get through thats so close its a corner so close martin henderson done very well to get the cross in

	filename	content	cleaned_content
1120	qG1IIIj9_fU.mp4.srt	<p>00:01:44,534\nMartin: It's a throw-in again.\n\n34\n00:01:44,955 --&gt; 00:01:46,939\n-All Liverpool body is back there.\n\n35\n00:01:48,711 --&gt; 00:01:50,695\nMartin: Angelino\ntrying to get through.\n\n36\n00:01:51,434 --&gt; 00:01:53,942\nThat's so, so close. It's a corner.\n\n37\n00:01:54,656 --&gt; 00:01:55,957\n-So close.\n\n38\n00:02:00,474 --&gt; 00:02:01,490\nMartin: Henderson,\n\n39\n00:02:01,965 --&gt; 00:02:02,965\nndone very well to get the cross in.\n\n40\n00:02:03,133 --&gt; 00:02:05,521\nSo well for Sadio Mane,\n\n41\n00:02:05,737 --&gt; 00:02:07,157\nncame in at the back post\n\n42\n00:02:08,483 --&gt; 00:02:10,689\nand squeezed it off the goalkeeper.\n\n43\n00:02:18,537 --&gt; 00:02:20,194\n[?].\n\n44\n00:02:20,362 --&gt; 00:02:21,696\nAgain, Angelino\n\n45\n00:02:23,498 --&gt; 00:02:25,966\nand they've got one.\nThey do always score.\n\n46\n00:02:26,435 --&gt; 00:02:28,086\nBernardo Silva swept it in.\n\n47\n00:02:28,673 --&gt; 00:02:30,324\nNot over yet at Anfield.\n\n48\n00:02:30,603 --&gt; 00:02:31,690\nHere's Sterling,\n\n49\n00:02:32,635 --&gt; 00:02:34,143\nhas never given up the ghost.\n\n50\n00:02:34,461 --&gt; 00:02:36,762\nShouts for handball hits\n\nAlexander-Arnold again,\n\n51\n00:02:36,930 --&gt; 00:02:39,643\nand then Henderson comes\nand grabs it.\n\n52\n00:02:39,811 --&gt; 00:02:41,382\nGabriel Jesus is hurt.\n\n53\n00:02:42,065 --&gt; 00:02:43,255\nHe's nodding his head.\n\n54\n00:02:44,263 --&gt; 00:02:46,096\nMaybe that's the anger\nfrom the first half.\n\n55\n00:02:47,057 --&gt; 00:02:48,231\nMartin: Angelino again.\n\n56\n00:02:49,627 --&gt; 00:02:52,175\n[?].\n\n57\n00:02:56,007 --&gt; 00:02:57,611\nJust too high for Bernardo.\n\n</p>	so well for sadio mane came in at the back post and squeezed it off the goalkeeper again angelino and theyve got one they do always score bernardo silva swept it in not over yet at anfield heres sterling has never given up the ghost shouts for handball hits alexanderarnold again and then henderson comes and grabs it gabriel jesus is hurt hes nodding his head maybe thats the anger from the first half martin angelino again just too high for bernardo

filename	content	cleaned_content
	<p>goal.\n\n12\n00:00:47,494 --&gt; 00:00:48,880\nThe ball going on the top onto Mane.\n\n13\n00:00:49,047 --&gt; 00:00:51,974\n[crowd cheering]\n\n14\n00:00:52,141 --&gt; 00:00:56,923\n-There was a touch on the arm,\nhand it's not going to count.\n\n15\n00:00:59,838 --&gt; 00:01:01,239\n-He's certainly not offside.\n\n16\n00:01:02,572 --&gt; 00:01:04,177\n[whistle blows]\n\n17\n00:01:08,061 --&gt; 00:01:09,217\n- The goal stands.\n\n18\n00:01:10,269 --&gt; 00:01:13,611\n-Sadio Mane has made\nthe breakthrough here for Liverpool.\n\n19\n00:01:14,242 --&gt; 00:01:15,359\n-Won by Moutinho.\n\n20\n00:01:15,726 --&gt; 00:01:17,926\n-Jonny trying to break free\nof Andy Robertson,\n\n21\n00:01:18,192 --&gt; 00:01:20,288\n-and manages to pull it back,\nnit's gone all the way across\n\n22\n00:01:20,455 --&gt; 00:01:21,561\n-and Neto\n\n23\n00:01:24,078 --&gt; 00:01:28,546\n-stuns Anfield with an equalizer\nright on the brink\n\n24\n00:01:28,713 --&gt; 00:01:30,451\nnof half-time.\n\n25\n00:01:31,806 --&gt; 00:01:35,678\n-Meanwhile, there's frustration\nfrom the first [?]\n\n26\n00:01:35,845 --&gt; 00:01:37,430\nbut that will be proven to be right.\n\n27\n00:01:37,597 --&gt; 00:01:39,998\n-We're still waiting\nfor confirmation on this one.\n\n28\n00:01:45,320 --&gt; 00:01:46,793\n-Brace yourselves.\n\n29\n00:01:47,513 --&gt; 00:01:48,763\n-The decision has come in.\n\n30\n00:01:49,423 --&gt; 00:01:53,096\n-Wolves are going\nonto be fuming about this,\n\n31\n00:01:53,263 --&gt; 00:01:55,590\n-the goal has been disallowed.\n\n32\n00:02:01,060 --&gt; 00:02:04,649\n-That was over the top\nof the bar from Wijnaldum.\n\n33\n00:02:11,793 --&gt; 00:02:15,182\n-van Dijk in trouble,\nhe's lost that here to Diogo Jota.\n\n34\n00:02:15,424 --&gt; 00:02:17,433\n-He produces the save.\n\n35\n00:02:20,520 --&gt; 00:02:22,046\n-Certainly, this time,\nit's the goalkeeper who saves\n\n36\n00:02:22,213 --&gt; 00:02:23,213\n-Virgil van Dijk.\n\n37\n00:02:28,430 --&gt; 00:02:30,780\n-Alexander-Arnold,\nought to be careful, Jimenez.\n\n</p>	<p>the turn and rui patricio scrambling across his goal the ball going on the top onto mane crowd cheering there was a touch on the arm and its not going to count hes certainly not offside whistle blows the goal stands sadio mane has made the breakthrough here for liverpool won by moutinho jonny trying to break free of andy robertson and manages to pull it back its gone all the way across and neto stuns anfield with an equalizer right on the brink of halftime meanwhile theres frustration from the first but that will be proven to be right were still waiting for confirmation on this one brace yourselves the decision has come in wolves are going to be fuming about this the goal has been disallowed that was over the top of the bar from wijnaldum van dijk in trouble hes lost that here to diogo jota he produces the save certainly this time its the goalkeeper who saves virgil van dijk alexanderarnold ought to be careful jimenez</p>

filename	content	cleaned_content
1121 GsY4w7sjh9A.mp4.srt	<p>1 n00:00:00,171 --&gt;  00:00:04,057 n[music] n n2 h00:00:05,657 --&gt;  00:00:06,800 nWell, the contrast is stark. n n3 h00:00:06,942 --&gt;  00:00:12,342 nSeven of this starting 11 for wales  nhave single figures in terms of caps, n n4 h00:00:12,828 --&gt;  00:00:14,600 nincluding this fella, David Brooks. n n5 h00:00:14,971 --&gt;  00:00:16,228 nFirst chance for him to run, n n6 h00:00:16,514 --&gt;  00:00:18,714 ncouldn't quite play it into  nGareth Bale in time, n n7 h00:00:18,800 --&gt;  00:00:20,114 nbut there's good pressure there. n n8 h00:00:20,257 --&gt;  00:00:26,571 nOh, first start for Wales,  nfirst goal for Wales. n n9 h00:00:26,971 --&gt;  00:00:30,057 nDan James inside the first five minutes. n n10 h00:00:31,828 --&gt;  00:00:35,000 nThat is not far from n the stuff of dreams. n n11 h00:00:37,571 --&gt;  00:00:40,828 nGreat work by David Brooks  nwho runs with the ball so well, n n12 h00:00:41,771 --&gt;  00:00:44,942 nand then this young exuberant  nWales side win the ball back. n n13 h00:00:45,200 --&gt;  00:00:48,800 nGreat strike  non his competitive debut for Wales. n n14 h00:00:49,571 --&gt;  00:00:53,171 nAbsolutely incredible moment  nfor him and his family, n n15 h00:00:53,657 --&gt;  &gt; 00:00:55,742 nand a huge start for Ryan Giggs. n n16 h00:00:56,971 --&gt;  00:00:58,457 nAllen out to James here. n n17 h00:01:01,257 --&gt;  00:01:04,828 nGood cross from Ben Davies.  nOh, side netting. n n18 h00:01:05,371 --&gt;  00:01:08,571 nDavid Brooks who played such  na key role is setting up the first n n19 h00:01:08,880 --&gt;  00:01:10,742 nvery nearly claiming a second. n n20 h00:01:11,571 --&gt;  00:01:13,371 nGreat overlap from Ben Davies. n n21 h00:01:14,085 --&gt;  00:01:14,828 nIt's a really good cross n n22 h00:01:14,914 --&gt;  00:01:17,057 nand it's good contact  nas well from David Brooks. n n23 h00:01:20,885 --&gt;  00:01:21,857 nBrooks now. n n24 h00:01:23,000 --&gt;  00:01:24,771 nThat's a lovely touch n to take him beyond Rusnák. n n25 h00:01:24,885 --&gt;  00:01:25,914 nStill at it, Brooks. n n26 h00:01:30,828 --&gt;  00:01:33,885 nThis young man just loves playing n in the shirt of his country. n n27 h00:01:34,828 --&gt;  00:01:39,200 nRecently crowned Player of the Year, n and that was a lovely piece of play. n n28 h00:01:44,800 --&gt;  00:01:47,714 nBale rises high but couldn't  nget enough on the header n n29 h00:01:48,171 --&gt;</p>	<p>music well the contrast is stark seven of this starting for wales have single figures in terms of caps including this fella david brooks first chance for him to run couldnt quite play it into gareth bale in time but theres good pressure there oh first start for wales first goal for wales dan james inside the first five minutes that is not far from the stuff of dreams great work by david brooks who runs with the ball so well and then this young exuberant wales side win the ball back great strike on his competitive debut for wales absolutely incredible moment for him and his family and a huge start for ryan giggs allen out to james here good cross from ben davies oh side netting david brooks who played such a key role is setting up the first very nearly claiming a second great overlap from ben davies its a really good cross and its good contact as well from david brooks brooks now thats a lovely touch to take him beyond rusnák still at it brooks this young man just loves playing in the shirt of his</p>

filename	content	cleaned_content
	<p>00:01:49,714 nto divert it goalwards.\n\n30 n00:01:51,085 --&gt; 00:01:52,914 nDubravka commits himself.\n\n31 n00:01:53,057 --&gt; 00:01:55,971 nlf Gareth Bale could just get above the ball a little bit more,\n\n32 n00:01:56,800 --&gt; 00:01:58,685 nit would been a certain goal for Wales.\n\n33 n00:01:59,542 --&gt; 00:02:01,542 nGoalkeeper caught in no man's land.\n\n34 n00:02:08,542 --&gt; 00:02:10,428 nMak here for Slovakia, \nhe's sent him through.\n\n35 n00:02:10,485 --&gt; 00:02:11,628 nOut comes Hennessey.\n\n36 n00:02:13,485 --&gt; 00:02:15,885 nHennessey made the block, &amp; then from the follow up,\n\n37 n00:02:15,971 --&gt; 00:02:17,714 nit went into the side netting.\n\n38 n00:02:18,000 --&gt; 00:02:20,828 nFirst time Wales \nhave really been cut apart.\n\n39 n00:02:21,542 --&gt; 00:02:22,771 nlIt's a good run.\n\n40 n00:02:23,542 --&gt; 00:02:27,228 nWayne Hennessey came out of his goal \nso quickly and shut down the angle.\n\n41 n00:02:27,820 --&gt; 00:02:29,542 nRusnák, as I said, \nhe makes that run from\n\n42 n00:02:30,171 --&gt; 00:02:31,628 nthe right-hand side across \nthe defense.\n\n43 n00:02:31,714 --&gt; 00:02:34,971 nDoes it so well, tries to just \nlift it over Wayne Hennessey.\n\n44 n00:02:35,457 --&gt; 00:02:39,742 nBig, strong, aggressive arm &amp; came out towards him.\n\n45 n00:02:41,600 - -&gt; 00:02:42,771 nRusnák.\n\n46 n00:02:45,771 --&gt; 00:02:47,828 nDuda. Hamšík.\n\n47 n00:02:48,942 --&gt; 00:02:53,371 nTidy turn but saved from Robert Mak's \nneffort by Wayne Hennessey.\n\n48 n00:02:54,371 --&gt; 00:02:56,628 nTerrific football from \nthe Slovakians once again.\n\n49 n00:02:57,228 --&gt; 00:02:59,200 nManoeuvre the ball \nfrom left to right really well,\n\n50 n00:02:59,542 --&gt; 00:03:01,142 nthen back across the pitch again.\n\n51 n00:03:01,457 --&gt; 00:03:03,400 nlIt's a good slide. \nlIt's a good save by Wayne.\n\n52 n00:03:04,285 --&gt; 00:03:06,085 nJoe Allen.\n\n53 n00:03:08,428 --&gt; 00:03:10,485 nBen Davies to Gareth Bale.\n\n54 n00:03:12,628 --&gt; 00:03:14,857 nLovely whip on that ball.\n\n55 n00:03:15,628 --&gt; 00:03:17,085 nAll the way through to Wilson.\n\n56 n00:03:17,628 --&gt; 00:03:19,000 nlIt might come for Tyler Roberts.\n\n57 n00:03:19,514 --&gt; 00:03:22,342 nHe was offside when it was hit.\n\nThe flag's up on the far side.\n\n58 n00:03:23,314 --&gt; 00:03:26,571 nMagnificent cross from Gareth Bale. \nPut it in such a dangerous area,\n\n59 n00:03:26,657 --&gt; 00:03:29,371 njust a little bit too high \nfor Tyler Roberts.\n\n60 n00:03:30,571 --&gt; 00:03:32,285 nNice first touch by Harry Wilson.\n\n61 n00:03:35,485 --&gt; 00:03:37,771 nRusnák is full of running here for Slovakia.\n\n62 n00:03:38,085 --&gt;</p>	<p>country recently crowned player of the year and that was a lovely piece of play bale rises high but couldnt get enough on the header to divert it goalwards dubravka commits himself if gareth bale could just get above the ball a little bit more it would been a certain goal for wales goalkeeper caught in no mans land mak here for slovakia hes sent him through out comes hennessey henessey made the block and then from the follow up it went into the side netting first time wales have really been cut apart its a good run wayne henessey came out of his goal so quickly and shut down the angle rusnák as i said he makes that run from the righthand side across the defense does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda hamšík tidy turn but saved from robert maks effort by wayne henessey terrific football from the slovakians once again manoeuvre the ball from left to right really well then back across the pitch again its a good slide its a good save by wayne joe allen ben davies to gareth bale lovely</p>

filename	content	cleaned_content
	<p>00:03:39,628\nTime to pick his pass.\n\n63\n00:03:39,714 --&gt; 00:03:42,257\nGot the header hard \nand that shot across there.\n\n64\n00:03:45,342 --&gt; 00:03:47,485\nMatt Smith just didn't quite go with the runner.\n\n65\n00:03:48,457 --&gt; 00:03:49,571\nGood cross in.\n\n66\n00:03:50,142 --&gt; 00:03:53,314\nIt's a wonderful header, \nisn't it, by Ďuriš.\n\n67\n00:03:55,171 --&gt; 00:03:56,942\nIt was Rusnák with the cross.\n\n68\n00:03:57,228 --&gt; 00:03:58,942\nSmart save by Hennessey.\n\n69\n00:03:59,657 --&gt; 00:04:03,342\nand then a huge chance \nto get Slovakia back into the game.\n\n70\n00:04:04,571 --&gt; 00:04:06,742\nUnfortunately for the visitors, \nit goes begging.\n\n</p>	whip on that ball all the way through to wilson it might come for tyler roberts he was offside when it was hit the flags up on the far side magnificent cross from gareth bale puts it in such a dangerous area just a little bit too high for tyler roberts nice first touch by harry wilson rusnák is full of running here for slovakia time to pick his pass got the header hard and that shot across there matt smith just didnt quite go with the runner good cross in its a wonderful header isnt it by ďuriš it was rusnák with the cross smart save by hennessey and then a huge chance to get slovakia back into the game unfortunately for the visitors it goes begging

	filename	content	cleaned_content
1122	R0XfelUhG_Y.mp4.srt	1\n00:00:00,429 --> 00:00:03,888\nHello, and welcome to \nthe Stadio Olimpico in\n\n2\n00:00:04,044 --> 00:00:07,878\nRome for round 31 of \nthe Serie A season,\n\n3\n00:00:07,982 --> 00:00:10,259\nit's Lazio versus Sassuolo.\n\n4\n00:00:10,453 --> 00:00:13,367\nI'm Patrick Terrone your \ncommentator for this game.\n\n5\n00:00:13,967 --> 00:00:15,721\nIt is the home side, Lazio,\n\n6\n00:00:15,746 --> 00:00:17,404\nwho get in the first half underway.\n\n7\n00:00:17,691 --> 00:00:18,818\n[whistle]\n\n8\n00:00:22,077 --> 00:00:23,823\nIt's Whipped in and \nthat's a good header\n\n9\n00:00:24,000 --> 00:00:25,785\nand a good save by Consigli\n\n10\n00:00:25,945 --> 00:00:27,438\nwho watched it all the way,\n\n11\n00:00:28,021 --> 00:00:29,021\nLazio\n\n12\n00:00:29,203 --> 00:00:31,243\nwith an early chance then.\n\n13\n00:00:35,091 --> 00:00:37,052\nIt's been a little bit \nndisjointed that.\n\n14\n00:00:37,610 --> 00:00:39,467\nOh, that flick Caicedo.\n\n15\n00:00:39,737 --> 00:00:41,284\nOh, almost in the back of the net,\n\n16\n00:00:42,104 --> 00:00:43,159\na rebound\n\n17\n00:00:43,756 --> 00:00:44,978\noff Immobile.\n\n18\n00:00:46,738 --> 00:00:48,230\n[whistle]\n\n19\n00:00:48,349 --> 00:00:49,802\nwho get the second half underway.\n\n20\n00:00:51,013 --> 00:00:52,021\nLulic,\n\n21\n00:00:53,187 --> 00:00:55,941\nsquaring the ball to Patrick \nwho comes forward.\n\n22\n00:00:56,532 --> 00:00:58,366\nIt's now out wide to Marusic.\n\n23\n00:00:59,219 --> 00:01:00,791\nPatrick with a chance to cross,\n\n24\n00:01:01,437 --> 00:01:03,460\nhand ball and that's a penalty kick,\n\n25\n00:01:03,595 --> 00:01:04,603\nsays the referee.\n\n26\n00:01:07,100 --> 00:01:09,180\nWell, the cross came from Patrick\n\n27\n00:01:11,685 --> 00:01:13,764\nand he's going to go over \nand have a look, so\n\n28\n00:01:14,693 --> 00:01:16,026\nnot an easy decision\n\n29\n00:01:16,370 --> 00:01:17,385\nfor the ref to make.\n\n30\n00:01:21,743 --> 00:01:24,513\nReferee has confirmed \nit's a penalty kick.\n\n31\n00:01:25,221 --> 00:01:26,650\nImmobile then up against\n\n32\n00:01:26,835 --> 00:01:28,644\nthe Consigli, here he goes.\n\n33\n00:01:28,958 --> 00:01:31,832\nImmobile has given Lazio the lead,\n\n34\n00:01:32,597 --> 00:01:34,549\nnone-nil to the home side.\n\n35\n00:01:34,811 -->	hello and welcome to the stadio olimpico in rome for round of the serie a season its lazio versus sassuolo im patrick terrone your commentator for this game it is the home side lazio who get in the first half underway whistle its whipped in and thats a good header and a good save by consigli who watched it all the way lazio with an early chance then its been a little bit disjointed that oh that flick caicedo oh almost in the back of the net a rebound off immobile whistle its sassuolo who get the second half underway lulic squaring the ball to patrick who comes forward its now out wide to marusic patrick with a chance to cross hand ball and thats a penalty kick says the referee well the cross came from patrick and hes going to go over and have a look so not an easy decision for the ref to make referee has confirmed its a penalty kick immobile then up against the consigli here he goes immobile has given lazio the lead onenil to the home side ciro immobile goal number for the striker an easy finish

filename	content	cleaned_content
	00:01:36,144\nCiro Immobile,\n\n36\n00:01:36,782 --> 00:01:39,131\ngoal number 14 for the striker.\n\n37\n00:01:39,649 --> 00:01:41,816\nAn easy finish sending Consigli.\n\n38\n00:01:42,112 --> 00:01:43,342\nthe wrong way.\n\n39\n00:01:43,853 --> 00:01:44,861\nDemiral,\n\n40\n00:01:45,942 --> > 00:01:46,950\nDuncan,\n\n41\n00:01:47,259 --> 00:01:48,435\nit's nicely forward Boga.\n\n42\n00:01:49,222 --> 00:01:50,667\nCan he twist and turn here? \n\n43\n00:01:53,746 --> 00:01:55,412\nHere he goes, \nhe's got the pace,\n\n44\n00:01:55,599 --> 00:01:57,956\nnot a bad effort. \nLirola trying to place it in.\n\n45\n00:01:58,149 --> 00:01:59,934\nit's there, Rogerio,\n\n46\n00:02:00,166 --> 00:02:01,198\n tapped it in\n\n47\n00:02:02,319 --> 00:02:04,017\nand Sassuolo right back in it,\n\n48\n00:02:04,194 --> 00:02:06,821\nfirst chance they've had in \nthe entire match\n\n49\n00:02:07,004 --> 00:02:08,004\nand it's a goal.\n\n50\n00:02:08,728 --> 00:02:10,157\nit is, of all people,\n\n51\n00:02:10,814 --> 00:02:12,369\nRogerio the Brazilian.\n\n52\n00:02:20,022 --> 00:02:21,038\nCorrea,\n\n53\n00:02:21,229 --> > 00:02:22,951\noh, that's a nice ball, Immobile,\n\n54\n00:02:23,642 --> 00:02:24,737\noh, what a chance\n\n55\n00:02:25,274 --> 00:02:27,235\n to give his side \nthe lead once again.\n\n56\n00:02:29,246 --> 00:02:30,278\nCorrea\n\n57\n00:02:31,453 --> 00:02:32,984\nleaves it for Luiz Alberto\n\n58\n00:02:33,272 --> 00:02:35,765\njust [?] it that's a wonderful ball, \nImmobile.\n\n59\n00:02:36,283 --> 00:02:37,323\nWhat a miss,\n\n60\n00:02:37,516 --> 00:02:39,595\nCiro Immobile all alone.\n\n61\n00:02:39,817 --> 00:02:41,016\nAll he had to do\n\n62\n00:02:41,357 --> 00:02:42,730\nwas put it in the back of the net\n\n63\n00:02:43,184 --> 00:02:44,264\nand he's missed.\n\n64\n00:02:44,830 --> 00:02:45,869\nLuis Alberto\n\n65\n00:02:46,524 --> 00:02:48,651\nnow to Radu, Radu's cross\n\n66\n00:02:48,916 --> 00:02:50,360\nTowards Milinković- Savić,\n\n67\n00:02:50,733 --> 00:02:53,487\nit's saved off the line by Consigli.\n\n68\n00:02:54,413 --> 00:02:55,445\nGreat header\n\n69\n00:02:56,247 --> 00:02:57,461\nfrom the substitute\n\n70\n00:02:57,922 --> 00:02:58,969\nbut equally\n\n71\n00:02:59,363 --> 00:03:00,379\ngood save.\n\n72\n00:03:03,018 --> 00:03:04,025\nNow,\n\n73\n00:03:05,216 -	sending consigli the wrong way demiral duncan its nicely forward boga can he twist and turn here here he goes hes got the pace not a bad effort lirola trying to place it in its there rogerio tapped it in and sassuolo right back in it first chance theyve had in the entire match and its a goal it is of all people rogerio the brazilian correa oh thats a nice ball immobile oh what a chance to give his side the lead once again correa leaves it for luiz alberto just it thats a wonderful ball immobile what a miss ciro immobile all alone all he had to do was put it in the back of the net and hes missed luis alberto now to radu radus cross towards milinkovićsavić its saved off the line by consigli great header from the substitute but equally good save now space and berardi through here clear through berardi sassuolo have taken the lead on the break berardi finishing with a plum whipped in flicked out by sensi whos the smallest man on the pitch now patrick out too immobile immobile across the face of goal its there the equalizer stands lulic from a meter

filename	content	cleaned_content
	<p>-&gt; 00:03:06,224\inspace\n\n74\n00:03:06,458 -&gt; 00:03:08,839\inand Berardi through here, \nclear through,\n\n75\n00:03:09,016 --&gt; 00:03:10,841\nBerardi.\n\n76\n00:03:11,072 --&gt; 00:03:12,500\nSassuolo\n\n77\n00:03:14,371 --&gt; 00:03:15,935\nhave taken the lead\n\n78\n00:03:16,387 --&gt; 00:03:17,498\nnon the break.\n\n79\n00:03:19,463 --&gt; 00:03:20,574\nBerardi\n\n80\n00:03:21,058 --&gt; 00:03:22,423\nfinishing with a plum.\n\n81\n00:03:27,797 --&gt; 00:03:28,813\nWhipped in,\n\n82\n00:03:29,170 --&gt; 00:03:30,266\nflicked out by\n\n83\n00:03:30,539 --&gt; 00:03:32,507\nSensi who's the smallest \nman on the pitch.\n\n84\n00:03:32,643 --&gt; 00:03:33,944\nNow, Patrick out too,\n\n85\n00:03:34,127 --&gt; 00:03:35,444\nImmobile, Immobile,\n\n86\n00:03:35,686 --&gt; 00:03:37,463\nacross the face of goal, it's there.\n\n87\n00:03:37,895 --&gt; 00:03:39,094\nThe equalizer stands,\n\n88\n00:03:39,451 --&gt; 00:03:40,459\nLulic\n\n89\n00:03:41,058 --&gt; 00:03:43,121\nfrom a meter out taps it in.\n\n90\n00:03:43,838 --&gt; 00:03:45,378\nThis time Immobile\n\n91\n00:03:46,028 --&gt; 00:03:47,267\nprovides the assist.\n\n92\n00:03:51,187 --&gt; 00:03:52,536\nThe game has finished,\n\n93\n00:03:52,973 --&gt; 00:03:54,997\nLazio two, Sassuolo two.\n\n94\n00:03:58,741 --&gt; 00:04:00,471\n[crowd cheering]\n\n</p>	out taps it in this time immobile provides the assist the game i has finished lazio two sassuolo two crowd cheering

1123 rows x 4 columns

```
In [8]: import pandas as pd

# Load player names from the first CSV file
players_df1 = pd.read_csv('/Users/alexdoyle/Desktop/LIS/Capstone/capstone_n'
player_names1 = players_df1['full_name'].unique().tolist()

# Load player names from the second CSV file
players_df2 = pd.read_csv('/Users/alexdoyle/Desktop/LIS/Capstone/capstone_n'
player_names2 = players_df2['full_name'].unique().tolist()

# Combine player names from both CSV files into a single list
all_player_names = player_names1 + player_names2

# Function to find and return matching player names in a given text
def find_player_names(text, player_names):
    found_names = []
    for name in player_names:
        if name.lower() in text.lower(): # Convert both to lowercase to ens
            found_names.append(name)
    return found_names

# Apply the function to each transcription and store the results in a new co
df['found_player_names'] = df['cleaned_content'].apply(lambda text: find_pla
```

```
# Display the DataFrame with the found player names
df[['filename', 'found_player_names']]
```

Out[8]:

	filename	found_player_names
0	cvctybLyEqE.mp4.srt	[João Cancelo]
1	oCI6MCoOi1Q.mp4.srt	[]
2	_kDc62-upoA.mp4.srt	[]
3	41kO2TvmNbg.mp4.srt	[Roberto]
4	DjNA-MIdCbY.mp4.srt	[Harry Kane, Harry Kane, Tammy Abraham]
...	...	...
1118	dJMNNQHkz0M.mp4.srt	[]
1119	xK5R8h7zln0.mp4.srt	[Bernardo, Bernardo Silva, Fabinho, Gabriel Jesus, Bernardo, Bernardo Silva, Fabinho, Gabriel Jesus]
1120	qG1IIIj9_fU.mp4.srt	[Jonny, Virgil van Dijk, Jonny, Virgil van Dijk]
1121	GsY4w7sjh9A.mp4.srt	[Ben Davies, David Brooks, Wayne Hennessey, Ben Davies, David Brooks, Harry Wilson, Wayne Hennessey]
1122	R0XfelUhG_Y.mp4.srt	[]

1123 rows × 2 columns

In [9]:

```
import os
import pandas as pd

# Initialise an empty set to store unique player names
unique_found_names = set()

# Define a function to find player names in a CSV file
def find_player_names_in_csv(csv_file_path):
    # Load player names from the CSV file
    players_df = pd.read_csv(csv_file_path)
    player_names = players_df['full_name'].unique().tolist()
    return player_names

# List of file paths for CSV files
csv_file_paths = [
    '/Users/alexdoyle/Desktop/LIS/Capstone/capstone_nlp/player_info/england',
    '/Users/alexdoyle/Desktop/LIS/Capstone/capstone_nlp/player_info/england'
]

# Iterate through CSV files
for csv_file_path in csv_file_paths:
    # Find player names in the current CSV file
    player_names = find_player_names_in_csv(csv_file_path)

    # Add new found names to the master set of unique names
    unique_found_names.update(player_names)

# Now, unique_found_names contains all unique player names found across all
print(unique_found_names)
```

{'Jay Rodriguez', 'André Schürrle', 'David McGoldrick', 'Joe Daniel Tupper', 'Max Thompson', 'Lewis Jack Gibson', 'Rodri', 'Víctor Camarasa', 'Joshua King', 'Tom Cleverley', 'Manolo Gabbiadini', 'Jermain Defoe', 'Loris Karius', 'Laurent Depoitre', 'Jonny Evans', 'Souleymane Bamba', 'Johann Berg Gümünds\xadson', 'Stuart Armstrong', 'Loïc Damour', 'Lucas Digne', 'Emi Bueno', 'Davy Pröpper', 'Kazaiah Sterling', 'Tom Cairney', 'Bernd Leno', 'Angelo Ogbonna', 'Mohamed Diamé', 'Christopher Schindler', 'Joel Coleman', 'Scott McTominay', 'Darren Randolph', 'Simon Moore', 'Wayne Hennessey', 'Jean-Michel Seri', 'Idrissa Gueye', 'David Martin', 'Stephan Lichtsteiner', 'Mohamed Salah', 'Isaac Hayden', 'Vincent Kompany', 'James McCarthy', 'Valentino Lazaro', 'João Virgínia', 'Emile Smith Rowe', 'Vincent Janssen', 'Rúben Neves', 'Anwar El Ghazi', 'Jan Bednarek', 'Anthony Gordon', 'Dimitriou Zohorè', 'Kepa Arrizabalaga', 'Felipe Anderson', 'Ruben Loftus-Cheek', 'Max Meyer', 'Jefferson Lerma', 'İlkay Gündoğan', 'Shane Duffy', 'Billy Gilmour', 'Matt Macey', 'Emerson Palmieri', 'Dwight Gayle', 'Alfie Mawson', 'Christian Kabasele', 'Luka Milivojević', 'Scott Dann', 'Wesley Moraes Ferreira da Silva', 'Ralf Fährmann', 'Georges-Kevin N\'Koudou Mbida', 'Marc Pugh', 'Angel Gomes', 'Alexis Mac Allister', 'Jazz Richards', 'Moussa Djeneb', 'Jeffrey Schlupp', 'Maya Yoshida', 'Ahmed Eissa El Mohamady Abdel Fattah', 'Charlie Daniels', 'Brandon Austin', 'Federico Fernández', 'Cedric Soares', 'Jannik Vestergaard', 'Jose Izquierdo', 'Gaëtan Bong Songo', 'Roberto', 'Jack Simpson', 'Ederson', 'Ramadan Sobhi', 'Jonjoe Kenny', 'Marcus Rashford', 'Danny Drinkwater', 'Maxime Le Marchand', 'Marco Stiepermann', 'Heung-Min Son', 'Shane Long', 'Teemu Pukki', 'Kyle Walker-Peters', 'Aboubakar Kamara', 'Harry Winks', 'Wilfred Ndidi', 'Ben Wilmot', 'Rui Patrício', 'Łukasz Fabiański', 'Xherdan Shaqiri', 'Dele Alli', 'Junior Stanislas', 'Joe Bryan', 'Steve Cook', 'Craig Dawson', 'Abdoulaye Doucouré', 'Fred', 'James Tarkowski', 'Javier Manquillo', 'Leandro Bacuna', 'Ryan Fredericks', 'Cenk Tosun', 'Moussa Sissoko', 'Michael McGovern', 'Onel Lázaro Hernández Mayea', 'Sam Surridge', 'Marc Albrighton', 'Sullay Kaikai', 'Ben Osborn', 'Neal Maupay', 'Antonio Rüdiger', 'Jamie Sterry', 'Stefano Okaka', 'Caoimhin Kelleher', 'Nicolás Pépé', 'Danny Simpson', 'Morgan Schneiderlin', 'Todd Cantwell', 'Indiana Vassilev', 'Gabriel Jesus', 'John Egan', 'Bernardo Silva', 'Curtis Jones', 'Ethan Ampadu', 'Mesut Özil', 'Yerry Mina', 'Kieran Dowell', 'Ondrej Duda', 'Gonzalo Higuaín', 'David Silva', 'Aaron Cresswell', 'Andriy Yarmolenko', 'Luca de la Torre', 'Wesley Hoedt', 'Fabricio Martín Agosto Ramírez', 'Shinji Okazaki', 'Gary Cahill', 'Eric Bailly', 'Jarrad Branthwaite', 'Ryan Giles', 'Nemanja Matić', 'Maarten Stekelenburg', 'Henrikh Mkhitaryan', 'Yves Bissouma', 'Matt Butcher', 'Antonio Barreca', 'Michael Obafemi', 'Dan Burn', 'Aaron Mooy', 'Lys Mousset', 'Jayson Molumby', 'Rafael Euclides Soares Camacho', 'Ismaïla Sarr', 'Taylor Perry', 'Fabinho', 'Joe Ralls', 'Odion Jude Ighalo', 'Gedson Carvalho Fernandes', 'Enda Stevens', 'Scott Carson', 'Adama Diakhaby', 'Richard Stearman', 'Diogo José Teixeira da Silva', 'Will Norris', 'Maximillian James Aarons', 'Aaron Ramsey', 'Simon Mignolet', 'Adam Smith', 'Étienne Capoue', 'Kyle Taylor', 'Ki Sung-yueng', 'Kieron Freeman', 'Arian Mariappa', 'Jack Wilshere', 'Alexandre Nascimento Costa Silva', 'Benjamin Mendy', 'Kevin McDonald', 'Matt Ritchie', 'Oriol Romeu', 'Nacho Monreal', 'Willy Caballero', 'Abdelhamid Sabiri', 'Nathaniel Clyne', 'João Pedro Junqueira de Jesus', 'Patrick Cutrone', 'Alfie Whiteman', 'Michel Vorm', 'Harry Kane', 'Grant Hanley', 'Paulo Gazzaniga', 'Bailey Peacock-Farrell', 'John Stones', 'Kevin Danso', 'Arijanet Muric', 'Nabil Bentaleb', 'Pepe Reina', 'Diogo Dalot', 'Bruno Saltor Grau', 'Dennis Srbeny', 'Timothy Fosu-Mensah', 'Ben White', 'Fabian Delph', 'Theo Walcott', 'Andrew Robertson', 'James Garner', 'Luciano Vietto', 'Nick Pope', 'Anthony Knockaert', 'Luke Dyer', 'Marc Guehi', 'Bernardo', 'Philippe Sandler', 'Gonçalo Cardoso', 'Adam Masina', 'Dejan Lovren', 'Bernardo Rusa', 'Armando Broja', 'Jack Rodwell', 'Matteo Guendouzi', 'Aaron Wan-Bissaka', 'Kevin Long', 'Terence Kongolo', 'Allan Saint-Maximin', 'Giovani Lo Celso', 'Álvaro Morata', 'Nathaniel Chalobah', 'Fabián Balbuena', 'Daniel Podence', 'Lucas Perez', 'Connor Wickham', 'Christoph Zimmermann', 'Michael Keane', 'Ayoze Perez', 'Steven Alzate', 'Ryan Bennett', 'Philip Billing', 'Alfie Jones', 'Max Kilman', 'Welfried Zaha', 'Max Harrison Sanders', 'Yan Valery', 'Callum Slattery', 'Jamie Vardy', 'Che Adams', 'Alexandre Lacazette', 'Oliver McBurnie', 'Marcos Alonso'}

o', 'Davinson Sánchez', 'Luke Amos', 'Mathias Jørgensen', 'Moise Kean', 'Sean Morrison', 'Lazar Marković', 'Zeze Steven Sessegnon', 'Olivier Giroud', 'Mace Lewin Goodridge', 'Lee Grant', 'Oumar Niassé', 'Simon Francis', 'André Gray', 'Mark Travers', 'Morgan Gibbs-White', 'Neil Leonard Dula Etheridge', 'Youri Tielemans', 'Lewis Cass', 'Joe Powell', 'Anthony Driscoll-Glenndon', 'James Milner', 'Adalberto Peñaranda Maestre', 'Joel Matip', 'Martín Montoya', 'Mateo Kovačić', 'Jason Steele', "N'Golo Kanté", 'Edward Nketiah', 'Asmir Begović', 'Jonas Lössl', 'Timothy Eyoma', 'Pablo Mari', 'Aaron Lennon', 'Jack Grealish', 'Taylor Harwood-Bellis', 'Dennis Praet', 'Jonathan Hogg', 'Steven Defour', 'Nathan Holland', 'Gylfi Sigurdsson', 'Kyle Walker', 'Ben Hamer', 'Isaac Success', 'Matty Daly', 'José Ángel Esmoris Tasende', 'Danilo', 'Aymeric Laporte', 'Craig Cathcart', 'Danny Ward', 'Romain Saïss', 'Adrián', 'Erik Lamela', 'Danny Ings', 'Sam Vokes', 'Emerson Hyndman', 'James Justin', 'Daniel Adshead', 'Cameron John', 'Steve Mounié', 'Ciaran Clark', 'Konstantinos Mavropanos', 'Serge Aurier', 'George Marsh', 'André Gomes', 'Jon Gorenc Stanković', 'Faustino Anjorin', 'Thomas David Allan', 'Phil Jones', 'Tom Trybull', 'Andreas Pereira', 'Cole Palmer', 'Bruno Jordão', 'Lucas Moura', 'Roberto Pereyra', 'Aaron Ramsdale', 'Luke Shaw', 'Oskar Buur', 'Akin Famewo', 'Matthew Smith', 'Miguel Almirón', 'Glenn Murray', 'Jean-Philippe Gbamin', 'Tyrick Mitchell', 'Dan Nlundulu', 'Filip Benković', 'Alex Iwobi', 'Stephen Ward', 'Callum Robinson', 'Brian Murphy', 'Panagiotis Retsos', 'Jesse Lingard', 'Roberto Firmino', 'Ayotomiwa Dele-Bashiru', 'James Ward-Prowse', 'Sergio Romero', 'Kelland Watts', 'Matt Targett', 'Jeff Hendrick', 'Jorginho', 'Dean Henderson', 'Matthew Connolly', 'Luke Freeman', 'Ricardo Pereira', 'Fernandinho', 'Tyreke Johnson', 'Julián Speroni', 'Muhammed Bešić', 'Oliver Skipp', 'John Lundstram', 'Nnamdi Ofoborh', 'Neco Williams', 'Christian Eriksen', 'Mason Greenwood', 'Moussa Dembele', 'Nikola Tavares', 'Matias Ezequiel Schelotto', 'Leonardo Campana', 'Ibrahima Cissé', 'Brandon Williams', 'Hélder Wander Sousa Azevedo Costa', 'John Ruddy', 'Manuel Lanzini', 'Fernando Llorente Torres', 'Kortney Hause', 'Philip Heise', 'Leon Balogun', 'Cesc Fabregas', 'Mahmoud Trezeguet', 'Lucas Torreira', 'Tudor Băluță', 'Christian Fuchs', 'Alireza Jahanbakhsh', 'Pablo Zabaleta', 'Billy Sharp', 'Harry Arter', 'Ross Barkley', 'Ben Johnson', 'Jordan Pickford', 'Jordon Ibe', 'Robert Kenedy Nunes do Nascimento', 'Sead Kolašinac', 'Joel Ward', 'Joseph Willock', 'Nya Kirby', 'Richarlison', 'Andreas Christensen', 'Kieran Trippier', 'Ben Davies', 'Alex Smithies', 'Danny Rose', 'Peter Crouch', 'Victor Moses', 'Ali Koiki', 'Kurt Zouma', 'Bakary Sako', 'Conor Coady', 'Dominic Solanke', 'Bright Enobakhare', 'James Chester', 'Alex Pritchard', 'Aaron Rowe', 'Dwight McNeil', 'Jason Puncheon', 'Ben Gibson', 'Adam Traoré', 'Pierre-Emerick Aubameyang', 'Mark Noble', 'Tyreece John-Jules', 'Rondon', 'Mason Holgate', 'João Cancelo', 'José Ignacio Peleteiro Ramallo', 'Beram Kayal', 'Liam Gibson', 'Jacob Maddox', 'Lee Peltier', 'Marko Arnautović', 'Sean Longstaff', 'Tyias Browning', 'Markus Suttner', 'Tanguy Ndombele', 'Josh Brownhill', 'Will Ferry', 'Victor Wanyama', 'Dani Ceballos', 'Hamza Choudhury', 'Leonardo Bonatini Lohner Maia', 'Bruno Fernandes', 'Joe Mumbongo', 'Brandon Leroy Pierrick Keutcha', 'Greg Cunningham', 'Adrien Sebastian Perruchet Silva', 'Nathaniel Mendez-Laing', 'Mohamed Elyounoussi', 'Rhu-endly Martina', 'Christian Benteke', 'Ivan Ricardo Neves Abreu Cavaleiro', 'Tim Krul', 'Ryan Bertrand', 'Shkodran Mustafi', 'Tahith Chong', 'Angus Gunn', 'Jack Stacey', 'Declan Rice', 'Steven Bergwijn', 'David Brooks', 'Leander Dendoncker', 'Aleksandar Mitrović', 'Bjorn Engels', 'Jeremy Ngakia', 'John Fleck', 'Darnell Johnson', 'Jamal Lewis', 'Gary Madine', 'Riyad Mahrez', 'Albian Ajeti', 'Cyrus Christie', 'Victor Lindelöf', 'Ki-Jana Hoever', 'Neil Taylor', 'Marvelous Nakamba', 'Tyrone Mings', 'Reece James', 'David de Gea', 'Arthur Masuaku', 'Collin Quaner', 'Grady Diangana', 'Lewis Dunk', 'Erik Pieters', 'Dennis Cirkin', 'Nathan Tella', 'Joe Gomez', 'Leighton Baines', 'Bernard Anício Caldeira Duarte', 'Jan Vertonghen', 'Kieran Tierney', 'Alfie Lewis', 'Daniel Amartey', 'Mario Lemina', 'Chris Wood', 'Ellis Simms', 'Pierre-Emile Højbjerg', 'Tammy Abraham', 'Daniel Williams', 'Andrew Surman', 'Matt Doherty', 'Charlie Austin', 'Jonathan Walters', 'Jack Stephens', 'Josh Murphy', 'Issa Diop', 'Romelu Lukaku', 'Eric García', 'Georginio Wijnaldum', 'Conor Coventry', 'Ruben Vinagre', 'Hector Bellerin', 'Gabriel Martinelli', 'Tariq Lamptey', 'Josip Drmic', 'Erik Durm', 'Matty James', 'Jordan Ayew', 'Sergio Rico', 'Florin Andone', 'Jonny', 'Jed Steer', 'Kadeem

Harris', 'Joe Bennett', 'Ravel Morrison', 'Sergio Aguero', 'Jonathan Kodjia', "Pape N'Diaye Souaré", 'Thomas Doyle', 'Douglas Luiz', 'Ignacio Pussetto', 'Jordan Thomas', 'Rob Elliot', 'Daniel James', 'Daryl Janmaat', 'Trent Alexander-Arnold', 'Caglar Söyüncü', 'George Baldock', 'Elias Kachunga', 'Jetro Willems', 'Christian Atsu Twasam', 'Danny Welbeck', 'Troy Parrott', 'Stefan Marius Johansen', 'Ainsley Maitland-Niles', 'Alex Oxlade-Chamberlain', 'Kevin De Bruyne', 'Alexis Sanchez', 'Axel Tuanzebe', 'Archie Mair', 'Andy Carroll', 'Harvey Barnes', 'Adam Webster', 'Harvey White', 'Naby Keita', 'Jesús Vallejo', 'Livre Kalinić', 'Bernard Patrick Ashley-Seal', 'James Daly', 'Oliver Norwood', 'Lewis Cook', 'Carlos Sánchez', 'Karlan Ahearne-Grant', 'Rachid Ghezzal', 'Neeskens Kebano', 'Cheikhou Kouyaté', 'Marouane Fellaini', 'Michail Antonio', 'Demeaco Duhaney', 'Joshua Sims', 'Jack Cork', 'Nathan Aké', 'Phil Jagielka', 'Mario Vrančić', 'Kelechi Iheanacho', 'Séamus Coleman', 'Robbie Brady', 'Brandon Pierrick', 'Ashley Barnes', 'Lukas Rupp', 'Alex McCarthy', 'Martin Dúbravka', 'Harry Lewis', 'João Moutinho', 'Araujo Danjuma', 'Marc Navarro Cecílano', 'Matteo Darmian', 'Jamaal Lascelles', 'Mathew Ryan', 'Jack Robinson', 'Kenny Mclean', 'Sofiane Boufal', 'Sokratis Papastathopoulos', 'Florent Hadergjonaj', 'Yoshinori Muto', 'Tom Heaton', 'Florian Lejeune', 'Jarrod Bowen', 'Virgil van Dijk', 'Carl Jenkinson', 'Michael Verrips', 'Tom Davies', 'Leandro Trossard', 'James Tomkins', 'Jaíro Riedewald', 'Callum Wilson', 'Ashley Westwood', 'Sadio Mané', 'Alex Dobre', 'Anthony Martial', 'Martin Kelly', 'DeAndre Yedlin', 'Josh Benson', 'Djibril Sidibé', 'Nicolas Otamendi', 'Isaac Mbenza', 'Ashley Young', 'Josh Knight', 'David Junior Hoilett', 'Artur Boruc', 'Harry Maguire', 'Mason Mount', 'Dale Stephens', 'Timm Klose', 'Pedro Rodriguez', 'Japhet Tanganga', 'André-Frank Zambo Anguissa', 'Jack O'Connell', 'Sandro Ramirez', 'Kasper Schmeichel', 'Alexandre Jankevitz', 'Charlie Taylor', 'Jimmy Dunne', 'Diego Rico', 'Bobby Reid', 'Robert Snodgrass', 'Joelinton', 'Adam Uche Idah', 'Antonio Valencia', 'Fraser Forster', 'Conor Hourihane', 'Nampalys Mendy', 'Alisson Becker', 'Granit Xhaka', 'Mohamed Elneny', 'Borja Bastón', 'Fabian Schär', 'James McArthur', 'Patrick van Aanholt', 'David Luiz', 'Divock Origi', 'Vicente Iborra De La Fuente', 'Jake Eastwood', 'Tom Smith', 'John-Kymani Gordon', 'Joe Hart', 'Sam Byram', 'David Button', 'Denis Suárez', 'Ezri Konsa', 'Fikayo Tomori', 'Viktor Gyökeres', 'Kiko Femenía', 'Jacob Murphy', 'Ryan Fraser', 'Davide Zappacosta', 'Emiliano Martínez', 'Pedro Neto', 'Alexander Sørloth', 'Ally Samatta', 'Willy Boly', 'Heurelho da Silva Gomes', 'Rajiv van La Parra', 'Aron Einar Gunnarsson', 'John McGinn', 'Sam Gallagher', 'George Hirst', 'Alberto Moreno', 'Christain Saydee', 'Reiss Nelson', 'Callum Hudson-Odoi', 'Chris Mepham', 'Alexander Baner Tettey', 'Sander Berge', 'Gerard Deulofeu', 'Pedro Obiang', 'Stephen Henderson', 'Sam Woods', 'Christian Pulisic', 'José Holebas', 'Samir Nasri', 'Dominic Calvert-Lewin', 'Paul Dummett', 'Raheem Sterling', 'Ryan Sessegnon', 'Jürgen Locadia', 'Rob Holding', 'Michy Batshuayi', 'Bruno Écuélé Manga', 'Phil Bardsley', 'Keinan Davis', 'Claudio Bravo', 'Floyd Ayité', 'Tim Ream', 'Vicente Guaita', 'Beni Bainingime', 'Frédéric Guilbert', 'Calum Chambers', 'Laurent Koscielny', 'Ryan Schofield', 'Nathan Redmond', 'Joselu', 'Gavin Kilkenny', 'Juninho Bacuna', 'Raúl Jiménez', 'Fousseni Diabaté', 'Aaron Anthony Connolly', 'Solly March', 'Callum Roberts', 'Rhian Brewster', 'Callum Paterson', 'Juan Foyth', 'Kayne Ramsey', 'Andros Townsend', 'James Maddison', 'Ben Godfrey', 'Tomáš Souček', 'Dan Gosling', 'Domingos Quina', 'Eric Dier', 'Petr Čech', 'Pablo Fornals', 'Pascal Groß', 'Karl Darlow', 'Patrick Roberts', 'Richairo Zivkovic', 'William Smallbone', 'Ben Chilwell', 'Håvard Nordtveit', 'Chris Löwe', 'Lloyd Kelly', 'Ben Mee', 'Ben Foster', 'Juan Mata', 'Mamadou Sakho', 'Miguel Ángel Britos Cabrera', 'Ryan Guno Babel', 'Henri Lansbury', 'Faustino Marcos Alberto Rojo', 'Chris Basham', 'Troy Deeney', 'Adam Lallana', 'Denis Odoi', 'Ibrahim Amadou', 'Hugo Lloris', 'Leon Clarke', 'Andrew Lonergan', 'Ademola Lookman', 'Matthew Lowton', 'Rhys Healey', 'Sebastian Prödl', 'Cesar Azpilicueta', 'Sébastien Haller', 'Jonjo Shelvey', 'Willian', 'Emil Kraft', 'Zechariah Medley', 'Wes Morgan', 'Demarai Gray', 'Harry Wilson', 'Paul Pogba', 'Matěj Vydra', 'Eden Hazard', 'Daniel Sturridge', 'Takumi Minamino', 'Ørjan Nyland', 'Matthew Longstaff', 'Ander Herrera', 'Jake Vokins', 'Winston Reid', 'Chris Smalling', 'Javier Hernandez', 'Steven Davis', 'Will Hughes', 'Moritz Leitner', 'Leroy Sané', 'Ken Sema', 'Nathaniel Phillips', 'Toby Alderweireld', 'Oleksandr Zinchenko', 'Marcus Bettinelli'}

```
In [10]: len(unique_found_names)
```

```
Out[10]: 800
```

```
In [11]: # Convert the set to a list and sort it
sorted_unique_names = sorted(list(unique_found_names))

# Print the sorted list of unique player names
print(sorted_unique_names)
```

['Aaron Anthony Connolly', 'Aaron Cresswell', 'Aaron Lennon', 'Aaron Mooy', 'Aaron Ramsdale', 'Aaron Ramsey', 'Aaron Rowe', 'Aaron Wan-Bissaka', 'Abdelhamid Sabiri', 'Abdoulaye Doucouré', 'Aboubakar Kamara', 'Adalberto Peñaranda Maestre', 'Adam Lallana', 'Adam Masina', 'Adam Smith', 'Adam Uche Idah', 'Adam Webster', 'Adama Diakhaby', 'Adama Traoré', 'Ademola Lookman', 'Adrián Mariappa', 'Adrien Sebastian Perruchet Silva', 'Adrián', 'Ahmed Eissa El Mohamady Abdel Fattah', 'Ainsley Maitland-Niles', 'Akin Famewo', 'Alberto Moreno', 'Albian Ajeti', 'Aleksandar Mitrović', 'Alex Dobre', 'Alex Iwobi', 'Alex McCarthy', 'Alex Oxlade-Chamberlain', 'Alex Pritchard', 'Alex Smithies', 'Alexander Banor Tettey', 'Alexander Sørloth', 'Alexandre Jankevitz', 'Alexandre Lacazette', 'Alexandre Nascimento Costa Silva', 'Alexis Mac Allister', 'Alexis Sanchez', 'Alfie Jones', 'Alfie Lewis', 'Alfie Mawson', 'Alfie Whiteman', 'Ali Koiki', 'Alireza Jahanbakhsh', 'Alisson Becker', 'Allan Saint-Maximin', 'Ally Samatta', 'Alvaro Morata', 'Ander Herrera', 'Andre Gray', 'Andreas Christensen', 'Andreas Pereira', 'Andrew Lonergan', 'Andrew Robertson', 'Andrew Surman', 'Andriy Yarmolenko', 'Andros Townsend', 'André Gomes', 'André Schürrle', 'André-Frank Zambo Anguissa', 'Andy Carroll', 'Angel Gomes', 'Angelo Ogbonna', 'Angus Gunn', 'Anthony Driscoll-Glennon', 'Anthony Gordon', 'Anthony Knockaert', 'Anthony Martial', 'Antonio Barreca', 'Antonio Rüdiger', 'Antonio Valencia', 'Anwar El Ghazi', 'Archie Mair', 'Arijanet Muric', 'Armando Broja', 'Arnaud Djoum', 'Aron Einar Gunnarsson', 'Arthur Masuaku', 'Artur Boruc', 'Ashley Barnes', 'Ashley Westwood', 'Ashley Young', 'Asmir Begović', 'Axel Tuanzebe', 'Aymeric Laporte', 'Ayotomiwa Délé-Bashiru', 'Ayoze Perez', 'Bailey Peacock-Farrell', 'Bakary Sako', 'Ben Chilwell', 'Ben Davies', 'Ben Foster', 'Ben Gibson', 'Ben Godfrey', 'Ben Hamer', 'Ben Johnson', 'Ben Mee', 'Ben Osborn', 'Ben White', 'Ben Wilmot', 'Beni Baningime', 'Benjamin Mendy', 'Beram Kayal', 'Bernard Anício Caldeira Duarte', 'Bernard Patrick Ashley-Seal', 'Bernardo', 'Bernardo Rusa', 'Bernardo Silva', 'Bernd Leno', 'Billy Gilmour', 'Billy Sharp', 'Bjorn Engels', 'Bobbi Reid', 'Borja Bastón', 'Brandon Austin', 'Brandon Leroy Pierrick Keutch', 'Brandon Pierrick', 'Brandon Williams', 'Brian Murphy', 'Bright Enobakhare', 'Bruno Fernandes', 'Bruno Jordão', 'Bruno Saltor Grau', 'Bruno Écuél Manga', 'Caglar Söyüncü', 'Callum Hudson-Odoi', 'Callum Paterson', 'Callum Roberts', 'Callum Robinson', 'Callum Slattery', 'Callum Wilson', 'Calum Chambers', 'Cameron John', 'Caoimhin Kelleher', 'Carl Jenkinson', 'Carlos Sánchez', 'Cedric Soares', 'Cenk Tosun', 'Cesar Azpilicueta', 'Cesc Fabregas', 'Charlie Austin', 'Charlie Daniels', 'Charlie Taylor', 'Che Adams', 'Chéikh Kouyaté', 'Chris Basham', 'Chris Löwe', 'Chris Mepham', 'Chris Smalling', 'Chris Wood', 'Christain Saydee', 'Christian Atsu Twasam', 'Christian Benteke', 'Christian Eriksen', 'Christian Fuchs', 'Christian Kabasele', 'Christian Pulisic', 'Christoph Zimmermann', 'Christopher Schindler', 'Ciaran Clark', 'Claudio Bravo', 'Cole Palmer', 'Collin Quaner', 'Connor Wickham', 'Conor Coady', 'Conor Coventry', 'Conor Hourihane', 'Craig Cathcart', 'Craig Dawson', 'Curtis Jones', 'Cyrus Christie', 'Dale Stephens', 'Dan Burn', 'Dan Gosling', 'Dan Nlundulu', 'Dani Ceballos', 'Daniel Adshead', 'Daniel A martey', 'Daniel James', 'Daniel Podence', 'Daniel Sturridge', 'Daniel Williams', 'Danilo', 'Danny Drinkwater', 'Danny Ings', 'Danny Rose', 'Danny Simpson', 'Danny Ward', 'Danny Welbeck', 'Darnell Johnson', 'Darren Randolph', 'Daryl Janmaat', 'David Brooks', 'David Button', 'David Junior Hoilett', 'David Luiz', 'David Martin', 'David McGoldrick', 'David Silva', 'David de Gea', 'Davide Zappacosta', 'Davinson Sánchez', 'Davy Propper', 'DeAndre Yedlin', 'Dean Henderson', 'Declan Rice', 'Dejan Lovren', 'Dele Alli', 'Demarai Gray', 'Demeaco Duhaney', 'Denis Odoi', 'Denis Suárez', 'Dennis Cirkin', 'Dennis Praet', 'Dennis Srbeny', 'Diego Rico', 'Dimitri Foulquier', 'Diogo Da lot', 'Diogo José Teixeira da Silva', 'Divock Origi', 'Djibril Sidibé', 'Dominos Quina', 'Dominic Calvert-Lewin', 'Dominic Solanke', 'Douglas Luiz', 'Dwight Gayle', 'Dwight McNeil', 'Eden Hazard', 'Ederson', 'Edward Nketiah', 'Elias Kachunga', 'Ellis Simms', 'Emerson Hyndman', 'Emerson Palmieri', 'Emi Buendía', 'Emil Kraft', 'Emile Smith Rowe', 'Emiliano Martínez', 'Enda Stevens', 'Eric Bailly', 'Eric Dier', 'Eric García', 'Erik Durm', 'Erik Lamela', 'Erik Pieters', 'Ethan Ampadu', 'Ezri Konsa', 'Fabian Delph', 'Fabian Schär', 'Fabinho', 'Fabián Balbuena', 'Fabricio Martín Agosto Ramírez', 'Faustino Anjorin', 'Faustino Marcos Alberto Rojo', 'Federico Fernández', 'Felipe Anderson', 'Fernandinho', 'Fernando Llorente Torres', 'Fikayo Tomor

i', 'Filip Benković', 'Florent Hadergjonaj', 'Florian Lejeune', 'Florin And one', 'Floyd Ayité', 'Fousseni Diabaté', 'Fraser Forster', 'Fred', 'Freddie Woodman', 'Frédéric Guilbert', 'Gabriel Jesus', 'Gabriel Martinelli', 'Gary Cahill', 'Gary Madine', 'Gavin Kilkenny', 'Gaëtan Bong Songo', 'Gedson Carvalho Fernandes', 'George Baldock', 'George Hirst', 'George Marsh', 'Georges -Kevin N'Koudou Mbida', 'Georginio Wijnaldum', 'Gerard Deulofeu', 'Giovani Lo Celso', 'Glenn Murray', 'Gonzalo Higuain', 'Gonçalo Cardoso', 'Grady Diaz', 'Granit Xhaka', 'Grant Hanley', 'Greg Cunningham', 'Gylfi Sigurðsson', 'Hamza Choudhury', 'Harry Arter', 'Harry Kane', 'Harry Lewis', 'Harry Maguire', 'Harry Wilson', 'Harry Winks', 'Harvey Barnes', 'Harvey White', 'Hector Bellerin', 'Henri Lansbury', 'Henrikh Mkhitaryan', 'Heung-Min Son', 'Heurelho da Silva Gomes', 'Hugo Lloris', 'Håvard Nordtveit', 'Hélder Wandrer Sousa Azevedo Costa', 'Ibrahim Amadou', 'Ibrahima Cissé', 'Idrissa Gueye', 'Ignacio Pussetto', 'Indiana Vassilev', 'Isaac Hayden', 'Isaac Mbenza', 'Isaac Success', 'Ismaïla Sarr', 'Issa Diop', 'Ivan Ricardo Neves Abreu Cavaleiro', 'Jack Cork', 'Jack Grealish', 'Jack O'Connell', 'Jack Robinson', 'Jack Rodwell', 'Jack Simpson', 'Jack Stacey', 'Jack Stephens', 'Jack Wilshere', 'Jacob Maddox', 'Jacob Murphy', 'Jake Eastwood', 'Jake Vokins', 'Jamaal Lascelles', 'Jamal Lewis', 'James Chester', 'James Daly', 'James Garner', 'James Justin', 'James Maddison', 'James McArthur', 'James McCarthy', 'James Milner', 'James Tarkowski', 'James Tomkins', 'James Ward-Prowse', 'Jamie Sterry', 'Jamie Vardy', 'Jan Bednarek', 'Jan Vertonghen', 'Jannik Vestergaard', 'Japhet Tanganga', 'Jarrad Branthwaite', 'Jarrod Bowen', 'Jason Puncheon', 'Jason Steele', 'Javier Hernandez', 'Javier Manquillo', 'Jay Rodriguez', 'Jayson Molumby', 'Jazz Richards', 'Jaïro Riedewald', 'Jean Michau Øebl Seri', 'Jean-Philippe Gbamin', 'Jed Steer', 'Jeff Hendrick', 'Jefferson Lerma', 'Jeffrey Schlupp', 'Jeremy Ngakia', 'Jermain Defoe', 'Jesse Lingard', 'Jesús Vallejo', 'Jetro Willems', 'Jimmy Dunne', 'Joe Bennett', 'Joe Bryan', 'Joe Daniel Tupper', 'Joe Gomez', 'Joe Hart', 'Joe Powell', 'Joe Ralls', 'Joel Coleman', 'Joel Matip', 'Joel Mumbongo', 'Joel Ward', 'Joelinton', 'Johann Berg Guðmunds\xadson', 'John Egan', 'John Fleck', 'John Lundstram', 'John McGinn', 'John Ruddy', 'John Stones', 'John-Kymani Gordon', 'Jon Gorenc Stanković', 'Jonas Lössl', 'Jonathan Hogg', 'Jonathan Kodjia', 'Jonathan Walters', 'Jonjo Shelvey', 'Jonjoe Kenny', 'Jonny', 'Jonny Evans', 'Jordan Ayew', 'Jordan Henderson', 'Jordan Pickford', 'Jordan Thomas', 'Jordon Ibe', 'Jorginho', 'Jose Izquierdo', 'Joselu', 'Joseph Willock', 'Josh Benson', 'Josh Brownhill', 'Josh Knight', 'Josh Murphy', 'Joshua King', 'Joshua Sims', 'Josip Drmic', 'José Holebas', 'José Ignacio Peleteiro Ramallo', 'José Ángel Esmoris Tasende', 'João Cancelo', 'João Moutinho', 'João Pedro Junqueira de Jesus', 'João Virgínia', 'Juan Foyth', 'Juan Mata', 'Julián Speroni', 'Juninho Bacuna', 'Junior Stanislas', 'Jürgen Locadia', 'Kadeem Harris', 'Karl Darlow', 'Karlan Ahearne-Grant', 'Kasper Schmeichel', 'Kayne Ramsay', 'Kazaiah Sterling', 'Keinan Davis', 'Kelechi Iheanacho', 'Kelland Watt', 'Ken Sema', 'Kenneth Dahrup Zohorè', 'Kenny McLean', 'Kepa Arrizabalaga', 'Kevin Danso', 'Kevin De Bruyne', 'Kevin Long', 'Kevin McDonald', 'Ki Sung-yueng', 'Ki-Jana Hoever', 'Kieran Dowell', 'Kieran Tierney', 'Kieran Trippier', 'Kieron Freeman', 'Kiko Femenía', 'Konstantinos Mavropanos', 'Kortney Hause', 'Kurt Zouma', 'Kyle Taylor', 'Kyle Walker', 'Kyle Walker-Peters', 'Laurent Depoitre', 'Laurent Koscielny', 'Lazar Marković', 'Leander Dendoncker', 'Leandro Bacuna', 'Leandro Trossard', 'Lee Grant', 'Lee Peltier', 'Leighton Baines', 'Leon Balogun', 'Leon Clarke', 'Leonardo Bonatini Lohner Maia', 'Leonardo Campana', 'Leroy Sané', 'Lewis Cass', 'Lewis Cook', 'Lewis Dunk', 'Lewis Jack Gibson', 'Liam Gibson', 'Lloyd Kelly', 'Loris Karius', 'Lovre Kalinić', 'Loïc Damour', 'Luca de la Torre', 'Lucas Digne', 'Lucas Moura', 'Lucas Perez', 'Lucas Torreira', 'Luciano Vietto', 'Luka Milivojević', 'Lukas Rupp', 'Luke Amos', 'Luke Dreher', 'Luke Freeman', 'Luke Shaw', 'Lys Mousset', 'Maarten Stekelenburg', 'Mace Lewin Goodridge', 'Mahmoud Trezeguet', 'Mamadou Sakho', 'Manolo Gabbiadini', 'Manuel Lanzini', 'Marc Albrighton', 'Marc Guehi', 'Marc Navarro Ceciliano', 'Marc Pugh', 'Marco Stepien', 'Marcos Alonso', 'Marcus Bettinelli', 'Marcus Rashford', 'Mario Lemina', 'Mario Vrančić', 'Mark Noble', 'Mark Travers', 'Marko Arnautović', 'Markus Suttner', 'Marouane Fellaini', 'Martin Dúbravka', 'Martin Kelly', 'Martín Montoya', 'Marvelous Nakamba', 'Mason Greenwood', 'Mason Holgate', 'Mason Mount', 'Mateo Kovacic', 'Mathew Ryan', 'Mathias Jørgensen', 'Matias Ez

equiel Schelotto', 'Matt Butcher', 'Matt Doherty', 'Matt Macey', 'Matt Ritchie', 'Matt Targett', 'Matteo Darmian', 'Matteo Guendouzi', 'Matthew Connolly', 'Matthew Longstaff', 'Matthew Lowton', 'Matthew Smith', 'Matty Daly', 'Matty James', 'Matěj Vydra', 'Max Harrison Sanders', 'Max Kilman', 'Max Meyer', 'Max Thompson', 'Maxime Le Marchand', 'Maximillian James Aarons', 'Maya Yoshida', 'Mesut Özil', 'Michael Keane', 'Michael McGovern', 'Michael Obafemi', 'Michael Verrips', 'Michail Antonio', 'Michel Vorm', 'Michy Batshuayi', 'Miguel Almirón', 'Miguel Ángel Britos Cabrera', 'Mohamed Diamé', 'Mohamed Elneny', 'Mohamed Elyounoussi', 'Mohamed Salah', 'Moise Kean', 'Morgan Gibbs-White', 'Morgan Schneiderlin', 'Moritz Leitner', 'Moussa Dembele', 'Moussa Djeneb', 'Moussa Sissoko', 'Muhammed Bešić', "N'Golo Kanté", 'Nabil Bentaleb', 'Naby Keïta', 'Nacho Monreal', 'Nampalys Mendy', 'Nathan Aké', 'Nathan Holland', 'Nathan Redmond', 'Nathan Tella', 'Nathaniel Chalobah', 'Nathaniel Clyne', 'Nathaniel Mendez-Laing', 'Nathaniel Phillips', 'Neal Maupay', 'Neco Williams', 'Neeskens Kebano', 'Neil Leonard Dula Etheridge', 'Neil Taylor', 'Nemanja Matić', 'Nick Pope', 'Nicolas Otamendi', 'Nicolas Pépé', 'Nikola Tavares', 'Nnamdi Ofori', 'Nya Kirby', 'Odion Jude Ighalo', 'Oleksandr Zinchenko', 'Oliver McBurnie', 'Oliver Norwood', 'Oliver Skipp', 'Olivier Giroud', 'Ondrej Duda', 'Onel Lázaro Hernández Mayea', 'Oriol Romeu', 'Oskar Buur', 'Oumar Niasse', 'Pablo Fornals', 'Pablo Mari', 'Pablo Zabaleta', 'Panagiotis Retsos', "Pape N'Diaye Souaré", 'Pascal Groß', 'Patrick Cutrone', 'Patrick Roberts', 'Patrick van Aanholt', 'Paul Dummett', 'Paul Pogba', 'Paulo Gazzaniga', 'Pedro Neto', 'Pedro Obiang', 'Pedro Rodriguez', 'Pepe Reina', 'Peter Crouch', 'Petr Čech', 'Phil Bardsley', 'Phil Foden', 'Phil Jagielka', 'Phil Jones', 'Philip Billing', 'Philip Heise', 'Philippe Sandler', 'Pierre-Emerick Aubameyang', 'Pierre-Emile Højbjerg', 'Rachid Ghezzal', 'Rafael Euclides Soares Camacho', 'Raheem Sterling', 'Rajiv van La Parra', 'Ralf Fährmann', 'Ramadan Sobhi', 'Ravel Morrison', 'Raúl Jiménez', 'Reece James', 'Reiss Nelson', 'Rhian Brewster', 'Rhu-endly Martina', 'Rhys Healey', 'Ricardo Pereira', 'Richairo Zivkovic', 'Richard Stearman', 'Richarlison', 'Riyad Mahrez', 'Rob Elliot', 'Rob Holding', 'Robbie Brady', 'Robert Kenedy Nunes do Nascimento', 'Robert Snodgrass', 'Roberto', 'Roberto Firmino', 'Roberto Pereyra', 'Rodri', 'Romain Saïss', 'Romelu Lukaku', 'Rondon', 'Ross Barkley', 'Ruben Loftus-Cheek', 'Ruben Vinagre', 'Rui Patrício', 'Ryan Bennett', 'Ryan Bertrand', 'Ryan Fraser', 'Ryan Fredericks', 'Ryan Giles', 'Ryan Guno Babel', 'Ryan Schofield', 'Ryan Sessegnon', 'Rúben Neves', 'Sadio Mané', 'Sam Byram', 'Sam Gallagher', 'Sam Surridge', 'Sam Vokes', 'Sam Woods', 'Samir Nasri', 'Sander Berge', 'Sandro Ramirez', 'Scott Carson', 'Scott Dann', 'Scott McTominay', 'Sead Kolašinac', 'Sean Longstaff', 'Sean Morrison', 'Sebastian Prödl', 'Serge Aurier', 'Sergio Aguero', 'Sergio Rico', 'Sergio Romero', 'Shane Duffy', 'Shane Long', 'Shinji Okazaki', 'Shkodran Mustafi', 'Simon Francis', 'Simon Mignolet', 'Simon Moore', 'Sofiane Boufal', 'Sokratis Papastathopoulos', 'Solly March', 'Souleymane Bamba', 'Stefan Marius Johansen', 'Stefano Okaka', 'Stephan Lichtsteiner', 'Stephen Henderson', 'Stephen Ward', 'Steve Cook', 'Steve Mounié', 'Steven Alzate', 'Steven Bergwijn', 'Steven Davis', 'Steven Defour', 'Stuart Armstrong', 'Sullay Kaikai', 'Séamus Coleman', 'Sébastien Haller', 'Tahith Chong', 'Takumi Minamino', 'Tammy Abraham', 'Tanguy Ndombele', 'Tariq Lamptey', 'Taylor Harwood-Bellis', 'Taylor Perry', 'Teemu Pukki', 'Terence Kongolo', 'Theo Walcott', 'Thomas David Allan', 'Thomas Doyle', 'Tim Krul', 'Tim Ream', 'Timm Klose', 'Timothy Eyoma', 'Timothy Fosu-Mensah', 'Toby Alderweireld', 'Todd Cantwell', 'Tom Cairney', 'Tom Cleverley', 'Tom Davies', 'Tom Heaton', 'Tom Smit', 'Tom Trybull', 'Tomáš Souček', 'Trent Alexander-Arnold', 'Troy Deeney', 'Troy Parrott', 'Tudor Băluță', 'Tyias Browning', 'Tyreece John-Jules', 'Treke Johnson', 'Tyrick Mitchell', 'Tyrone Mings', 'Valentino Lazaro', 'Vicente Guaita', 'Vicente Iborra De La Fuente', 'Victor Lindelöf', 'Victor Moses', 'Victor Wanyama', 'Víktor Gyökeres', 'Vincent Janssen', 'Vincent Kompany', 'Virgil van Dijk', 'Víctor Camarasa', 'Wayne Hennessey', 'Wes Morgan', 'Wesley Hoedt', 'Wesley Moraes Ferreira da Silva', 'Wilfred Ndidi', 'Wilfried Zaha', 'Will Ferry', 'Will Hughes', 'Will Norris', 'William Smallbone', 'Willian', 'Willy Boly', 'Willy Caballero', 'Winston Reid', 'Xherdan Shaqiri', 'Yan Valery', 'Yerry Mina', 'Yoshinori Muto', 'Youri Tielemans', 'Yves Bissouma', 'Zechariah Medley', 'Zeze Steven Sessegnon', 'Étienne Capoue', 'Ørjan Nyland', 'İlkay Gündoğan', 'Łukasz Fabiański']

In [ ]:

The above code was setting up the codebook and the transcript for analysis. The next section will attempt to assign the skin tone labels (which have been curated using the skin tone classifier, and stored within a spreadsheet alongside each player name) to the players identified within this text. The aim here is that I can establish the frequency of words (associated with power, pace, intelligence, leadership, communication, tactical awareness, etc...) with each skin tone category. Determining if a lighter or darker skin tone is spoken more in terms of physical or cognitive attributes. I would like to be able to establish if the commentary talks about these attributes in a positive or negative manner, however I am still working on a way that will accurately do so.

In [ ]:

```
In [12]: # Load the Excel file containing player names and skin tone labels
df_skin_tones = pd.read_csv('/Users/alexdoyle/Desktop/player_names.csv')
```

```
# Check the contents of the DataFrame
print(df_skin_tones.head())
```

	Player Names	Player Images
0	Aaron Anthony Connolly	<a href="https://cdn.sofifa.net/players/237/286/18_60.png">https://cdn.sofifa.net/players/237/286/18_60.png</a>
1	Aaron Cresswell	<a href="https://cdn.sofifa.net/players/189/615/18_60.png">https://cdn.sofifa.net/players/189/615/18_60.png</a>
2	Aaron Lennon	NaN
3	Aaron Mooy	<a href="https://cdn.sofifa.net/players/194/958/18_60.png">https://cdn.sofifa.net/players/194/958/18_60.png</a>
4	Aaron Ramsdale	<a href="https://cdn.sofifa.net/players/233/934/18_60.png">https://cdn.sofifa.net/players/233/934/18_60.png</a>

```
In [13]: # Print the column names
print(df_skin_tones.columns)
```

```
Index(['Player Names', 'Player Images'], dtype='object')
```

```
In [14]: # Load the Excel file again
df_skin_tones = pd.read_csv('/Users/alexdoyle/Desktop/player_names.csv')
```

```
# Adjust to use the actual column names
df_skin_tones = df_skin_tones[['Player Names', 'Player Images']].dropna()
```

```
# Normalise player names by lowercasing and stripping any extra whitespace
df_skin_tones['Player Names'] = df_skin_tones['Player Names'].str.lower().str.strip()
```

```
In [15]: # Print the column names of df to see if 'found_player_names' is present
print(df.columns)
```

```
Index(['filename', 'content', 'cleaned_content', 'original_case',
       'found_player_names'],
      dtype='object')
```

```
In [16]: # Ensure sorted_unique_names contains a list of unique player names
sorted_unique_names = sorted(list(unique_found_names))
```

```
# Function to find and return matching player names in a given text
def find_player_names(text, player_names):
    found_names = []
    for name in player_names:
        if name.lower() in text.lower(): # Convert both to lowercase to ensure case insensitivity
            found_names.append(name)
    return found_names
```

```
# Apply the function to each transcription and store the results in a new column
df['found_player_names'] = df['cleaned_content'].apply(lambda text: find_player_names(text, sorted_unique_names))
```

```
# Check if 'found_player_names' column is created
print(df[['filename', 'found_player_names']].head(10))
```

df

```
filename \
0 cvctybLyEqE.mp4.srt
1 oCl6MCo0i1Q.mp4.srt
2 _kDc62-upoA.mp4.srt
3 41k02TvmNbg.mp4.srt
4 DjNA-MIdCbY.mp4.srt
5 H0rLlWM30IU.mp4.srt
6 wa-f96ESXwA.mp4.srt
7 7lmXk6g1ito.mp4.srt
8 psMNMM2CpgQ.mp4.srt
9 FZkS3hBabXg.mp4.srt

found_pla
yer_names
0 [João
Cancelo]
1 []
2 []
3 [Roberto]
4 [Harry Kane, Tammy
Abraham]
5 []
6 []
7 [Alex McCarthy, Bernardo, Bernardo Silva, Danny Ings, Kyle Walker, Sergi
o Aguero]
8 []
9 [Rodri]
```

Out[16]:

	filename	content	cleaned_content
0	cvctybLyEqE.mp4.srt	<p>1\n00:00:00,177 --&gt; 00:00:06,044\n-Welcome to a rainy Stadio Olimpico.\n Lazio against Juventus \n\n2\n00:00:06,288 --&gt; 00:00:09,244\nshould be a real atmosphere this. \n\n3\n00:00:09,466 --&gt; 00:00:14,688\nCan Lazio become the first team inthis Serie A season to beat Juventus?\n\n4\n00:00:14,680 --&gt; 00:00:18,644\nI don't think they're going to stop\ntheir irrepressible march to the title, \n\n5\n00:00:18,640 --&gt; 00:00:22,666\nparticularly after Napoli could only \ndraw against Milan last night.\n\n6\n00:00:23,666 --&gt; 00:00:26,133\nMarco Guida gets us underway.\n\n7\n00:00:32,355 --&gt; 00:00:34,111\nReferee got in the way there.\n\n8\n00:00:34,800 --&gt; 00:00:38,177\nHere goes Dybala, \nLucas once more wins it back.\n\n9\n00:00:41,288 --&gt; 00:00:44,933\nNow Immobile opens up for Correa.\n\n10\n00:00:45,622 --&gt; 00:00:50,444\nSzczęsny got behind it on an \nawkward surface and held it well.\n\n11\n00:00:56,244 --&gt; 00:00:58,933\nJuve have a tricky game \nagainst Atalanta in the Cup.\n\n12\n00:01:02,088 --&gt; 00:01:06,066\nTricky moments here as \nwell as it falls for Luiz Alberto.\n\n13\n00:01:06,266 --&gt; 00:01:09,844\n Szczęsny with the save\nand Lucas with a miss hit. \n\n14\n00:01:13,822 --&gt; 00:01:15,088\nHere's Correa,\n\n15\n00:01:17,866 --&gt; 00:01:22,488\nmiss kick there from Matuidi. \nMilinković-Savić.\n\n16\n00:01:24,488 --&gt; 00:01:28,577\nLook at the space for Parolo.\nFabulous save from Szczęsny \n\n17\n00:01:29,022 --&gt; 00:01:31,977\nwho has kept Juve in it here.\nThat's the pick of them.\n\n18\n00:01:44,000 --&gt; 00:01:46,688\nThe effort levels never\ndrop when he's out there. \n\n19\n00:01:47,244 --&gt; 00:01:49,688\nIt might bounce here for Immobile.\n\n20\n00:01:50,444 --&gt; 00:01:54,200\nRugani does so well to recover\nand get it behind.\n\n21\n00:01:55,066 --&gt; 00:01:57,622\nGreat attacking instinct from Immobile.\n\n22\n00:02:00,111 --&gt; 00:02:04,711\nMuch to discuss, much to find out.\nNo changes at half time.\n\n23\n00:02:08,977 --&gt; 00:02:11,066\nThey need something\nto show for their efforts. \n\n24\n00:02:11,060 --&gt; 00:02:13,755\nIn from Luiz Alberto, and they have it.\n\n25\n00:02:16,311 --&gt; 00:02:18,933\nDid it ricochet off Milinković-Savić? \n\n26\n00:02:19,133 --&gt; 00:02:22,577\nIs it an own-goal?\nNo one at the Olmpico could care.\n\n27\n00:02:22,955 --&gt; 00:02:28,488\nLazio in front, Juve in trouble\nin the rain in Rome.\n\n28\n00:02:32,688 --&gt; 00:02:33,933\nNow Correa.\n\n29\n00:02:35,044 --&gt; 00:02:39,888\nOh, Immobile is in with a chance to make it two.\n\n30\n00:02:42,822 --&gt; 00:02:44,600\nForward by Alex Sandro.</p>	welcome to a rainy stadio olimpico lazio against juventus should be a real atmosphere this can lazio become the first team this serie a season to beat juventus i dont think theyre going to stop their irrepressible march to the title particularly after napoli could only draw against milan last night marco guida gets us underway referee got in the way there here goes dybala lucas once more wins it back now immobile opens up for correa szczęsny got behind it on an awkward surface and held it well juve have a tricky game against atalanta in the cup tricky moments here as well as it falls for luiz alberto szczęsny with the save and lucas with a miss hit heres correa misskick there from matuidi milinkovićsavić look at the space for parolo fabulous save from szczęsny who has kept juve in it here thats the pick of them the effort levels never drop when hes out there it might bounce here for immobile rugani does so well to recover and get it behind great attacking instinct from immobile much to discuss much to find out no changes at half time they

filename	content	cleaned_content
	<p>\n\n31\n00:02:46,488 --&gt; 00:02:52,822\nOh, great skill from Bernardeschi.\nFor Dybala, spilled, big chance. \n\n32\n00:02:52,820 --&gt; 00:02:58,444\nJoão Cancelo with an equalizer for Juventus. \n\n33\n00:03:01,822 --&gt; 00:03:04,066\nBack from injury to save them. \n\n34\n00:03:05,000 --&gt; 00:03:07,755\nThe substitute makes it 1-1.\n\n35\n00:03:10,755 --&gt; 00:03:13,711\nHold on to this or even better it \nand they might just do that. \n\n36\n00:03:13,710 --&gt; 00:03:15,800\nRonaldo in for Bernardeschi. \n\n37\n00:03:16,311 --&gt; 00:03:19,400\nJoão Cancelo went over on\nthe challenge from Lulić. \n\n38\n00:03:19,600 --&gt; 00:03:20,977\nHe's given a penalty. \n\n39\n00:03:21,911 --&gt; 00:03:24,844\nMarco Guida has given the penalty. \n\n40\n00:03:25,555 --&gt; 00:03:30,155\nThey were under massive pressure. \nCan they turn it round? They can.\n\n41\n00:03:36,088 --&gt; 00:03:40,844\nIt's what champions do on the ropes.\n\n42\n00:03:43,844 --&gt; 00:03:46,822\nSergej Milinković-Savić, over the bar. \n\n43\n00:03:46,820 --&gt; 00:03:50,622\nJuventus can celebrate,\n can punch the air.\n\n44\n00:03:50,777 --&gt; 00:03:54,222\nJuve win it by two goals to one.\n\n</p>	need something I to show for their i efforts in from i luiz alberto and they have it did it ricochet off milinkovićsavić is it an owngoal no one at the olmpico could care lazio in front juve in trouble in the rain in rome now correa oh immobile is in with a chance to make it two forward by alex sandro oh great skill from bernardeschi for dybala spilled big chance joão cancelo with an equalizer for juventus back from injury to save them the substitute makes it hold on to this or even better it and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a penalty marco guida has given the penalty they were under massive pressure can they turn it round they can its what champions do on the ropes sergej milinkovićsavić over the bar juventus can celebrate can punch the air juve win it by two goals to one

	filename	content	cleaned_content
1	oCI6MCoOj1Q.mp4.srt	<p>1 n00:00:00,484 --&gt; 00:00:02,841 nHello and welcome nto the Stadio Bentegodi. n n2 n00:00:03,009 --&gt; 00:00:06,176 nwhere this afternoon nChievo Verona and Frosinone n n3 n00:00:06,344 --&gt; 00:00:09,462 ngo head-to-head nin round 19 of Serie A. n n4 n00:00:09,630 --&gt; 00:00:11,730 nl'm Marco Palmieri. nGreat to be with you. n n5 n00:00:11,908 --&gt; 00:00:14,463 nlt's a must-win match for both sides. n n6 n00:00:14,726 --&gt; 00:00:18,136 nWe're away at the Bentegodi nbetween Chievo and Frosinone. n n7 n00:00:18,574 --&gt; 00:00:19,774 nThe corners are coming, n n8 n00:00:20,159 --&gt; 00:00:21,409 ncan they take advantage? n n9 n00:00:21,945 --&gt; 00:00:24,995 nThe big men in the area, nas Birsa steps up to take the corner. n n10 n00:00:25,739 --&gt; 00:00:26,794 nlt is indeed. n n11 n00:00:27,921 --&gt; 00:00:32,238 nSergio Pellissier nwith a header from the corner n n12 n00:00:34,619 --&gt; 00:00:37,761 nand the captain does it again nfor Chievo. n n13 n00:00:45,111 --&gt; 00:00:46,111 nMaking sure n n14 n00:00:46,279 --&gt; 00:00:48,679 nthat there weren't any irregularities nin the box, n n15 n00:00:49,400 --&gt; 00:00:51,253 nmaybe a push. n n16 n00:00:52,539 --&gt; 00:00:54,699 nlt will be a on-field review. n n17 n00:01:04,585 --&gt; 00:01:06,290 nThe goal has been disallowed. n n18 n00:01:06,703 --&gt; 00:01:09,344 nlt's still 0-0 nhere at the Bentegodi. n n19 n00:01:11,495 --&gt; 00:01:13,184 nPellissier is furious. n n20 n00:01:14,073 --&gt; 00:01:15,073 nA push, n n21 n00:01:15,241 --&gt; 00:01:17,341 nl'm assuming nthere was a push by Pellissier n n22 n00:01:17,509 --&gt; 00:01:19,036 nas he lept up into the air. n n23 n00:01:19,397 --&gt; 00:01:21,941 nAs Rocchi gets set to start nthe second half n n24 n00:01:22,109 --&gt; 00:01:23,109 nand we're away. n n25 n00:01:26,478 --&gt; 00:01:28,702 nAs Hetemaj gets it to Birsa. n n26 n00:01:32,067 --&gt; 00:01:34,061 nHe skips past Maiello. n n27 n00:01:34,252 --&gt; 00:01:35,530 nNow, Đorđević, n n28 n00:01:35,698 --&gt; 00:01:37,118 nthe first shot on target, n n29 n00:01:38,515 --&gt; 00:01:41,476 nbut Sportiello denies him nwith a quick save, n n30 n00:01:42,449 --&gt; 00:01:43,999 nnot necessarily a difficult</p>	<p>hello and welcome to the stadio bentegodi where this afternoon chievo verona and frosinone go headtohead in round of serie a im marco palmieri great to be with you its a mustwin match for both sides were away at the bentegodi between chievo and frosinone the corners are coming can they take advantage the big men in the area as birsa steps up to take the corner it is indeed sergio pellissier with a header from the corner and the captain does it again for chievo making sure that there werent any irregularities in the box maybe a push it will be a onfield review the goal has been disallowed its still here at the bentegodi pellissier is furious a push im assuming there was a push by pellissier as he lept up into the air as rocchi gets set to start the second half and were away as hetemaj gets it to birsa he skips past maiello now đorđević the first shot on target but sportiello denies him with a quick save not necessarily a</p>

filename	content	cleaned_content
	<p>one\n\n31\n00:01:44,167 --&gt; 00:01:45,917\nbut one that had to be made\n\n32\n00:01:46,553 -&gt; 00:01:49,746\nas the ball found its way\nonto the Serb striker.\n\n33\n00:01:56,255 --&gt; 00:01:57,442\nCorner taken short.\n\n34\n00:01:58,268 --&gt; 00:02:00,318\nAs the ball at the far post\nfor Chibsah.\n\n35\n00:02:00,792 --&gt; 00:02:02,529\nBrilliant save from Sorrentino.\n\n36\n00:02:04,704 --&gt; 00:02:07,504\nThe Frosinone players thought\nthat it'd crossed the line.\n\n37\n00:02:08,228 --&gt; 00:02:12,538\nThe watch on referee Rocchi\nndid not signal a goal.\n\n38\n00:02:12,706 --&gt; 00:02:14,012\nand play continues.\n\n39\n00:02:17,279 --&gt; 00:02:19,187\nAs now Depaoli is in space\n\n40\n00:02:22,155 --&gt; 00:02:24,505\nand he's tripped\non the edge of the area here.\n\n41\n00:02:25,095 --&gt; 00:02:26,845\nThere will be a foul and a yellow.\n\n42\n00:02:27,956 --&gt; 00:02:30,192\nlt's a second yellow,\nCapuano is off.\n\n43\n00:02:32,798 --&gt; 00:02:34,966\nFrosinone reduced to 10 men.\n\n44\n00:02:43,927 --&gt; 00:02:45,892\nHetemaj sees the overlap now.\n\n45\n00:02:46,798 --&gt; 00:02:48,532\nPlay inside, Pellissier.\n\n46\n00:02:48,890 --&gt; 00:02:51,932\nOn his left foot,\nhe slipped over as he took the shot.\n\n47\n00:02:54,193 --&gt; 00:02:55,393\nGreat chance for Chievo.\n\n48\n00:02:57,253 --&gt; 00:02:58,803\nas they continue to attack now.\n\n49\n00:02:58,971 --&gt; 00:03:02,219\nTomović, spreads it across\nRadovanović.\n\n50\n00:03:02,831 --&gt; 00:03:04,344\nNow Bani plays it in the box\n\n51\n00:03:04,512 --&gt; 00:03:06,021\nbut Depaoli cannot reach it\n\n52\n00:03:06,506 --&gt; 00:03:08,506\nand the ball dribbles back\nonto the keeper.\n\n53\n00:03:10,892 --&gt; 00:03:12,757\nWhat a chance for Pellissier here.\n\n54\n00:03:18,695 --&gt; 00:03:21,506\nThe shot comes in\nand that's a wonderful goal.\n\n55\n00:03:21,674 --&gt; 00:03:25,926\nGiaccherini gives Chievo the lead\n\n56\n00:03:26,125 --&gt; 00:03:27,756\nwith a superb free-kick.\n\n57\n00:03:28,167 --&gt; 00:03:30,209\nSportiello rooted to the spot.\n\n58\n00:03:32,392 --&gt; 00:03:35,565\nAn excellent goal\nand thoroughly deserved.\n\n59\n00:03:38,248 --&gt; 00:03:40,175\nThere it is, the final whistle.\n\n60\n00:03:40,343 --&gt; 00:03:42,836\nA simply perfect\nGiaccherini free-kick.\n\n61\n00:03:43,122 --&gt; 00:03:45,550\nhas given Chievo\ntheir first win of the season.\n\n62\n00:03:45,780 --&gt; 00:03:47,392\nin the last match of the year.\n\n63\n00:03:47,647 --&gt; 00:03:51,108\nThey've defeated a 10-man Frosinone\nby a one goal to nil.\n</p>	<p>difficult one but one that had to be made as the ball found its way to the serb striker corner taken short as the ball at the far post for chibsah brilliant save from sorrentino the frosinone players thought that itd crossed the line the watch on referee rocchi did not signal a goal and play continues as now depaoli is in space and hes tripped on the edge of the area here there will be a foul and a yellow its a second yellow capuano is off frosinone reduced to men hetemaj sees the overlap now play inside pellissier on his left foot he slipped over as he took the shot great chance for chievo as they continue to attack now tomović spreads it across to radovanović now bani plays it in the box but depaoli cannot reach it and the ball dribbles back to the keeper what a chance for pellissier here the shot comes in and that's a wonderful goal giaccherini gives chievo the lead on minutes with a superb freekick sportiello rooted to the spot an excellent goal and thoroughly deserved there it is the final whistle a simply perfect giaccherini freekick has</p>

filename	content	cleaned_content
		given chievo their first win of the season in the last match of the year theyve defeated a man frosinone by a one goal to nill

2	_kDc62-upoA.mp4.srt	<p>1\n00:00:00,389 --&gt; 00:00:03,488\n-The league's top scorers,\nagainst Seria A's weakest frontline.\n\n00:00:03,862 --&gt; 00:00:07,576\nAtalanta are bidding to make 2020\nthe best year in club history.\n\n00:00:07,847 --&gt; 00:00:09,927\nSpal are four points from safety\n\n00:00:10,205 --&gt; 00:00:12,883\nand running out of time\nin their efforts to beat the drop.\n\n00:00:16,606 --&gt; 00:00:17,964\nDi Francesco in the channel,\n\n00:00:19,401 --&gt; 00:00:20,643\nable to wander inside.\n\n00:00:22,965 --&gt; 00:00:25,139\nGabriel Strefezza lining it up.\n\n00:00:25,307 --&gt; 00:00:27,307\nIt's a good save from Sportiello.\n\n00:00:34,144 --&gt; 00:00:37,996\nThe game becoming stretched\njust like Atalanta want it to be.\n\n00:00:39,648 --&gt; 00:00:41,506\nZapata running at Thiago Cionek.\n\n00:00:43,006 --&gt; 00:00:45,935\nDuván Zapata's cross,\nit's in for Ilicic.\n\n00:00:48,166 --&gt; 00:00:51,321\nA difficult finish\nmade to look all so easy.\n\n00:00:52,973 --&gt; 00:00:56,670\nZapata still seeking his first goal\nsince his return to action.\n\n00:00:56,955 --&gt; 00:00:58,648\nThat was a telling contribution.\n\n00:00:59,990 --&gt; 00:01:03,751\nIlicic has scored\nagainst Fiorentina in midweek.\n\n00:01:04,792 --&gt; 00:01:07,815\ncontinues his wonderful start to 2020.\n\n00:01:13,191 --&gt; 00:01:14,586\nWell in by Gosens.\n\n00:01:15,643 --&gt; 00:01:16,784\nThis is</p>	<p>the leagues top scorers against seria as weakest frontline atalanta are bidding to make the best year in club history spal are four points from safety and running out of time in their efforts to beat the drop di francesco in the channel able to wander inside gabriel strefezza lining it up its a good save from sportiello the game becoming stretched just like atalanta want it to be zapata running at thiago cionek duván zapatas cross its in for ilicic a difficult finish made to look all so easy zapata still seeking his first goal since his return to action that was a telling contribution ilicic has scored</p>
---	---------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

filename	content	cleaned_content
	Pasalic\n\n19\n00:01:17,293 --> 00:01:18,973\nplaying slightly deeper this evening.\n\n20\n00:01:19,373 --> 00:01:20,407\nNow Zapata,\n\n21\n00:01:21,115 --> 00:01:22,965\nplenty of room for him to get moving.\n\n22\n00:01:23,871 --> 00:01:25,010\nUp against Vicari,\n\n23\n00:01:25,232 --> 00:01:26,797\nDuván Zapata with the shot.\n\n24\n00:01:27,312 --> 00:01:28,590\nIt's off the frame of the goal.\n\n25\n00:01:40,444 --> 00:01:41,461\nThis is Valoti.\n\n26\n00:01:41,827 --> 00:01:43,213\nThiago Cionek joining in.\n\n27\n00:01:43,920 --> 00:01:45,111\nThe cross is only half-cleared.\n\n28\n00:01:45,279 --> 00:01:46,279\nStrafezza.\n\n29\n00:01:48,239 --> 00:01:49,239\nCionek with the ball in.\n\n30\n00:01:49,407 --> 00:01:50,407\nSportiello stays at home.\n\n31\n00:01:50,575 --> 00:01:53,137\nPetagna's header is off the post.\n\n32\n00:01:54,066 --> 00:01:57,295\nSo close to yet another goal\nagainst his former club.\n\n33\n00:02:07,298 --> 00:02:08,924\nThat's a nice touch\nfrom Di Francesco.\n\n34\n00:02:09,092 --> 00:02:10,148\nReca is very quick.\n\n35\n00:02:10,316 --> 00:02:12,005\nHe's trying to get\naway from Djimsiti.\n\n36\n00:02:12,158 --> 00:02:13,467\nThis is Reca's cross,\n\n37\n00:02:13,635 --> 00:02:15,429\nand it's put in by Petagna.\n\n38\n00:02:17,302 --> 00:02:20,080\nA couple of former\nAtalanta players combine,\n\n39\n00:02:20,248 --> 00:02:21,279\nReca on loan.\n\n40\n00:02:21,699 --> 00:02:24,644\nPetagna once again\nscores against La Dea.\n\n41\n00:02:25,303 --> 00:02:27,328\nNo one could catch up with the ball.\n\n42\n00:02:32,296 --> 00:02:33,296\nDjimsiti.\n\n43\n00:02:36,408 -> 00:02:38,837\nZapata, a foul on him Vicari.\n\n44\n00:02:40,020 --> 00:02:41,020\nDabo,\n\n45\n00:02:42,440 --> 00:02:43,440\nValoti.\n\n46\n00:02:44,846 --> 00:02:47,100\nHe's kept it.\nMattia Valoti.\n\n47\n00:02:47,306 --> 00:02:48,520\nSpal have another.\n\n48\n00:02:49,482 --> 00:02:52,570\nThey've come from behind here\nto lead in Bergamo.\n\n49\n00:02:53,894 --> 00:02:56,284\nWhat a superb individual strike.\n\n50\n00:02:57,522 --> 00:02:59,690\nPetagna first and now Valoti.\n\n51\n00:03:00,556 --> 00:03:02,921\nHe took his goal splendidly well.\n\n52\n00:03:09,356 --> 00:03:11,040\nLaunched long by Ilicic.\n\n53\n00:03:12,230 --> 00:03:14,114\nMuriel gives chase, he's got there.\n\n54\n00:03:14,282 --> 00:03:15,282\nThis is	against fiorentina in midweek continues his wonderful start to i well in by gosens this is pasalic playing slightly deeper this evening now zapata plenty of room for him to get moving up against vicari duván zapata with the shot its off the frame of the goal this is valoti thiago cionek joining in the cross is only halfcleared \\\' strafezza cionek with the ball in sportiello stays at home petagnas header is off the post so close to yet another goal against his former club thats \\\' a nice touch from di francesco reca is very quick hes trying to get away from djimsiti this is recas cross and its put in by petagna a couple of former atalanta players combine reca on loan petagna once again scores against la dea no one could catch up with the ball \\\' djimsiti zapata a foul on him vicari dabo valoti hes kept it mattia valoti spal have another theyve come from behind here to lead in bergamo what a superb individual strike petagna first and now valoti he took his goal splendidly well launched long by ilicic muriel gives chase hes got there this is zapata duvan zapata its deflected terrific \

filename	content	cleaned_content
	Zapata.\n\n55\n00:03:15,450 --> 00:03:17,217\nDuvan Zapata, it's deflected.\n\n56\n00:03:17,589 --> 00:03:19,653\nTerrific save from Berisha.\n\n57\n00:03:20,019 --> 00:03:21,153\nThat was creeping in.\n\n58\n00:03:23,401 --> 00:03:25,590\nAnother look\nat that long ball from Ilicic.\n\n59\n00:03:25,758 --> 00:03:26,758\nMuriel did well.\n\n60\n00:03:28,632 --> 00:03:30,378\nZapata struck it firmly.\n\n61\n00:03:34,600 --> 00:03:36,981\nSpal win in Bergamo.\n\n62\n00:03:37,149 --> 00:03:39,276\nWhat a turn around for Semplici,\n\n63\n00:03:39,538 --> 00:03:43,103\nonly a fourth victory of the season\nfor Seria A's bottom side.\n\n64\n00:03:43,524 --> 00:03:48,548\nThree massive points in their quest\nto retain their topflight status.\n\n65\n00:03:49,018 --> 00:03:50,454\nAtalanta off-color\n\n66\n00:03:50,637 --> 00:03:52,637\nand out of sorts they stay fifth.\n\n67\n00:03:55,074 --> 00:04:02,074\n[crowd cheering]\n\n68\n00:04:02,242 --> 00:04:09,242\n[crowd cheering]\n\n69\n00:04:09,410 --> 00:04:13,126\n[crowd cheering]\n\n	save from berisha that was creeping in another look at that long ball from ilicic muriel did well zapata struck it firmly spal win in bergamo what a turn around for semplici only a fourth victory of the season for seria as bottom side three massive points in their quest to retain their topflight status atalanta offcolor and out of sorts they stay fifth crowd cheering crowd cheering crowd cheering

3	41kO2TvmNbg.mp4.srt	1\n00:00:00,500 --> 00:00:03,566\nWelcome to the Ennio Tardini stadium.\n\n2\n00:00:04,133 --> 00:00:04,733\nThis week, \n\n3\n00:00:04,733 --> 00:00:06,599\nit's the meeting of the two sides\n\n4\n00:00:06,600 --> 00:00:09,166\nwho are the bane of the City of Milan\n\n5\n00:00:09,166 --> 00:00:10,799\nin the last match week: \n\n6\n00:00:10,800 --> 00:00:13,133\nParma against Cagliari.\n\n7\n00:00:13,133 --> 00:00:14,433\nLast weekend, \n\n8\n00:00:14,433 --> 00:00:17,233\nParma recorded a surprise victory \n\n9\n00:00:17,233 --> 00:00:19,799\nat San Siro against Inter,\n\n10\n00:00:19,800 --> 00:00:23,500\nwhile Cagliari got an impressive draw\n\n11\n00:00:23,500 --> 00:00:25,566\nagainst Milan in Sardinia.\n\n12\n00:00:26,300 --> 00:00:27,333\nWe're underway. \n\n13\n00:00:30,500 --> 00:00:32,933\nStolen and now Gervinho\n\n14\n00:00:34,266 --> 00:00:37,132\nshow us that famous sprinting power.\n\n15\n00:00:37,133 --> 00:00:37,866\nStill	welcome to the ennio tardini stadium this week its the meeting of the two sides who are the bane of the city of milan in the last match week parma against cagliari last weekend parma recorded a surprise victory at san siro against inter while cagliari got an impressive draw against milan in sardinia were underway stolen and now gervinho can show us that famous sprinting power still
---	---------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

filename	content	cleaned_content
	<p>going. \n\n16\n00:00:37,866 --&gt; 00:00:39,066\nGervinho still going, \n\n17\n00:00:39,066 --&gt; 00:00:40,166\ngets the shot off \n\n18\n00:00:40,833 --&gt; 00:00:42,633\njust wide of the post, \n\n19\n00:00:43,466 --&gt; 00:00:45,932\ntrademark run from Gervinho.\n\n20\n00:00:47,900 --&gt; 00:00:50,200\nLook where he picks the ball up in his own half. \n\n21\n00:00:50,900 --&gt; 00:00:53,166\nThen he goes on an angled run inside, \n\n22\n00:00:54,266 --&gt; 00:00:55,832\ngets past Klavan, \n\n23\n00:00:55,833 --&gt; 00:00:57,266\nsees off Ionită \n\n24\n00:00:57,266 --&gt; 00:00:58,366\nand a low shot. \n\n25\n00:00:58,366 --&gt; 00:01:00,032\nYou can see what he was going for there. \n\n26\n00:01:01,966 --&gt; 00:01:02,732\nDimarco, \n\n27\n00:01:06,933 --&gt; 00:01:07,966\nInglese is in there again. \n\n28\n00:01:07,966 --&gt; 00:01:10,332\nInglese loops it over the keeper, \n\n29\n00:01:10,333 --&gt; 00:01:11,366\nCragno. \n\n30\n00:01:11,366 --&gt; 00:01:12,632\nThis time, \n\n31\n00:01:12,633 --&gt; 00:01:13,999\nhas the goal been given? \n\n32\n00:01:14,000 --&gt; 00:01:15,333\nIt looks like it has. \n\n33\n00:01:16,266 --&gt; 00:01:19,132\nDenied by the offside flag a few minutes ago.\n\n34\n00:01:19,133 --&gt; 00:01:23,199\nThis time, Roberto Inglese is off the mark \n\n35\n00:01:23,200 --&gt; 00:01:24,133\nin the match, \n\n36\n00:01:24,866 --&gt; 00:01:26,499\ngiving Parma the lead.\n\n37\n00:01:30,233 --&gt; 00:01:31,233\nExcellent run. \n\n38\n00:01:32,633 --&gt; 00:01:34,233\nInglese persistent, \n\n39\n00:01:34,866 --&gt; 00:01:37,599\nnot lucky there from the deflection off Cragno. \n\n40\n00:01:39,333 --&gt; 00:01:41,499\nHis shot cannoned off the goalkeeper, \n\n41\n00:01:41,500 --&gt; 00:01:43,933\nback into his face and into the net. \n\n42\n00:01:45,200 --&gt; 00:01:46,266\nThere's a couple of checks, \n\n43\n00:01:46,266 --&gt; 00:01:47,799\nReferee Calvarese. \n\n44\n00:01:49,633 --&gt; 00:01:51,166\nThe goal has been given. \n\n45\n00:01:51,166 --&gt; 00:01:53,332\nParma do go into the lead. \n\n46\n00:01:55,600 --&gt; 00:01:56,566\nLuca Rigoni, \n\n47\n00:01:57,133 --&gt; 00:01:57,899\nDi Gaudio. \n\n48\n00:01:59,966 --&gt; 00:02:00,732\nWill he go for the cross? \n\n49\n00:02:00,733 --&gt; 00:02:01,833\nHe plays it back. \n\n50\n00:02:01,833 --&gt; 00:02:03,033\nIt's a shot from \n\n51\n00:02:05,300 --&gt; 00:02:06,766\nAntonino Barillà. \n\n52\n00:02:06,766 --&gt; 00:02:08,032\nThe ball is still alive. \n\n53\n00:02:08,033 --&gt; 00:02:08,899\nGervinho \n\n54\n00:02:09,666 --&gt; 00:02:12,499\nGervinho pulls it back into the area.\n\n55\n00:02:12,500 --&gt; 00:02:13,900\nIt's Luca Rigoni, \n\n56\n00:02:13,900 --&gt; 00:02:15,200\nna</p>	<p>power still going gervinho still going gets the shot off just wide of the post trademark run from gervinho look where he picks the ball up in his own half then he goes on an angled run inside gets past klavan sees off ionită and a low shot you can see what he was going for there dimarco inglese is in there again inglese loops it over the keeper cragno this time has the goal been given it looks like it has denied by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma the lead excellent run inglese persistent got lucky there from the deflection off cragno his shot cannoned off the goalkeeper back into his face and into the net theres a couple of checks referee calvarese the goal has been given parma do go into the lead luca rigoni di gaudio will he go for the cross he plays it back its a shot from antonino barillà the ball is still alive gervinho gervinho pulls it back into the area its luca rigoni a familiar sight the raiding midfielder the second half starts marco sau goes down referee says play on now now</p>

filename	content	cleaned_content
DjNA-MIdCbY.mp4.srt	familiar sight, \n\n57\n00:02:16,633 --> 00:02:17,966\nthe raiding midfielder.\n\n58\n00:02:21,400 --> 00:02:22,966\nThe second half starts. \n\n59\n00:02:24,300 --> 00:02:25,433\nMarco Sau goes down.\n\n60\n00:02:25,433 --> 00:02:26,866\nReferee says, "Play on now." \n\n61\n00:02:26,866 --> 00:02:29,999\nNow Gervinho on one of those runs.\n\n62\n00:02:30,000 --> 00:02:32,466\nGervinho still going.\n\n63\n00:02:33,166 --> 00:02:35,366\nGervinho past another.\n\n64\n00:02:35,366 --> 00:02:36,699\nWhat's the shot like?\n\n65\n00:02:36,700 --> 00:02:39,666\nIt's a fantastic shot. \n\n66\n00:02:39,666 --> 00:02:41,099\nGervinho. \n\n67\n00:02:41,100 --> 00:02:42,666\nWhat a goal.\n\n68\n00:02:47,600 --> 00:02:50,566\nA standing ovation from the Tardini.\n\n69\n00:02:51,800 --> 00:02:53,233\nDanger at one end.\n\n70\n00:02:54,200 --> 00:02:56,100\nGervinho picked the ball up\n\n71\n00:02:56,100 --> 00:02:59,566\nand then showed his genius.\n\n72\n00:03:07,066 --> 00:03:09,266\nA rare lack of control from Srna\n\n73\n00:03:10,766 --> 00:03:12,699\nbut tenacity getting the ball back,\n\n74\n00:03:13,433 --> 00:03:14,433\nplays it in. \n\n75\n00:03:14,433 --> 00:03:16,966\nAn attempted backflip by Marco Sau.\n\n76\n00:03:19,666 --> 00:03:21,466\nHad there been a foul in the build-up? \n\n77\n00:03:31,033 --> 00:03:32,533\nWon back by Parma.\n\n78\n00:03:32,533 --> 00:03:35,033\nThe break is on into the area. \n\n79\n00:03:35,033 --> 00:03:36,866\nCeravolo gets it back.\n\n80\n00:03:36,866 --> 00:03:38,066\nWill he make it three? \n\n81\n00:03:38,666 --> 00:03:40,799\nGreat save by Cragno.\n\n82\n00:03:46,933 --> 00:03:49,599\nReferee Juan Paolo Calvarese\n\n83\n00:03:49,600 --> 00:03:51,533\nblows the full-time whistle. \n\n84\n00:03:51,533 --> 00:03:53,966\nParma 2-0 Cagliari.\n\n	gervinho on one of those runs gervinho still going gervinho past another whats the shot like its a fantastic shot gervinho what a goal a standing ovation from the tardini danger at one end gervinho picked the ball up and then showed his genius a rare lack of control from srna but tenacity getting the ball back plays it in an attempted backflip by marco sau had there been a foul in the buildup won back by parma the break is on into the area ceravolo gets it back will he make it three great save by cragno referee juan paolo calvarese blows the fulltime whistle parma cagliari

4 DjNA-MIdCbY.mp4.srt

1\n00:00:00,000 --> 00:00:04,064\n[music]\n\n2\n00:00:08,104 --> 00:00:09,480\nChilwell.\n\n3\n00:00:10,840 --> 00:00:12,696\nOh, and it's fired in in.\n\n4\n00:00:12,776 --> 00:00:15,360\nTypical Oxlade Chamberlain style.\n\n5\n00:00:16,256 --> 00:00:18,360\nHe is a goal threat.\n\n6\n00:00:19,104 --> 00:00:20,776\nHe really didn't catch it well.\n\n7\n00:00:20,816 --> 00:00:22,904\nIt's another lovely ball.\n\n8\n00:00:24,840 --> 00:00:26,920\nIt's not the easiest of chances.\n\n9\n00:00:27,576 --> 00:00:29,136\nWhen you strike it that

music chilwell oh and its fired in in typical oxlade chamberlain style he is a goal threat he really didnt catch it well its another lovely ball its not the easiest of chances when you strike it that purely theres no chance for the

filename	content	cleaned_content
	<p>purely,\n\n10\n00:00:29,160 --&gt; 00:00:31,368\nthere's no chance for the keeper.\n\n11\n00:00:33,448 --&gt; 00:00:36,104\nThey're stretched defensively here,\n\n12\n00:00:36,264 --&gt; 00:00:39,624\nHarry Kane just makes that point.\n\n13\n00:00:40,712 --&gt; 00:00:42,592\nOnly needed a touch.\n\n14\n00:00:43,232 --&gt; 00:00:45,480\nFloated in by Chilwell.\n\n15\n00:00:45,632 --&gt; 00:00:47,056\namongst a clutch of players,\n\n16\n00:00:47,112 --&gt; 00:00:48,408\nbut he manages just to\n\n17\n00:00:49,100 --&gt; 00:00:51,874\nkeep his opponents at arm's length.\n\n18\n00:00:52,666 --&gt; 00:00:54,874\nChilwell to take the corner.\n\n19\n00:00:57,874 --&gt; 00:01:00,888\nThat's a free header. That's another goal.\n\n20\n00:01:01,137 --&gt; 00:01:02,488\nSimple for Kane.\n\n21\n00:01:02,835 --&gt; 00:01:04,702\nEngland incidentally are hot\n\n22\n00:01:05,057 --&gt; 00:01:07,057\nshowing their array of set pieces,\n\n23\n00:01:07,128 --&gt; 00:01:08,960\nthey want to sail to the finals.\n\n24\n00:01:10,168 --&gt; 00:01:11,048\nOn the play,\n\n25\n00:01:11,050 --&gt; 00:01:13,857\nnobody knows what those Montenegro\n\n26\n00:01:14,204 --&gt; 00:01:16,151\ndefenders were doing here\n\n27\n00:01:16,240 --&gt; 00:01:18,311\nallow Kane a free run.\n\n28\n00:01:18,360 --&gt; 00:01:19,352\nThey've got to get a bit more\n\n29\n00:01:19,396 --&gt; 00:01:22,283\nphysical contact and so easy.\n\n30\n00:01:28,622 --&gt; 00:01:32,915\nRashford gets in on the act.\n\n31\n00:01:36,833 --&gt; 00:01:38,333\nWell that's four.\n\n32\n00:01:39,860 --&gt; 00:01:42,990\nOxlade Chamberlain with a really good ball.\n\n33\n00:01:43,030 --&gt; 00:01:45,712\nMaguire can't believe he hasn't scored there.\n\n34\n00:01:45,844 --&gt; 00:01:48,712\nRashford tidies up.\n\n35\n00:01:54,627 --&gt; 00:01:55,526\nSancho.\n\n36\n00:01:58,603 --&gt; 00:02:00,295\nHere's the hat-trick.\n\n37\n00:02:01,784 --&gt; 00:02:03,495\nAnother Harry hat-trick.\n\n38\n00:02:03,846 --&gt; 00:02:05,200\nFive for England.\n\n39\n00:02:06,092 --&gt; 00:02:07,821\nThis landmark evening\n\n40\n00:02:07,889 --&gt; 00:02:09,538\ncould not be going better.\n\n41\n00:02:10,030 --&gt; 00:02:12,941\nIt wasn't the easiest ball to control.\n\n42\n00:02:13,040 --&gt; 00:02:15,825\nfirst time from Alexander-Arnold.\n\n43\n00:02:15,905 --&gt; 00:02:18,080\njust bouncing in front of him.\n[laugh]\n\n44\n00:02:18,080 --&gt; 00:02:20,821\nHe's such a confident frame of mind.\n\n45\n00:02:21,825 --&gt; 00:02:23,820\nthat he just decides to take a touch.\n\n46\n00:02:25,629 --&gt;</p>	<p>keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in by chilwell amongst a clutch of players but he manages just to keep his opponents at arms length chilwell to take the corner thats a free header thats another goal simple for kane england incidentally are not showing their array of set pieces they want to sail to the finals on the play nobody knows what those montenegro defenders were doing here to allow kane a free run theyve got to get a bit more physical contact and so easy rashford gets in on the act well thats four oxlade chamberlain with a really good ball maguire cant believe he hasnt scored there rashford tidies up sancho heres the hattrick another harrtack five for england this landmark evening could not be going better it wasnt the easiest ball to control first time from alexanderarnold just bouncing in front of him laugh hes such a confident frame of mind that he just decides to take a touch it proves to be the right decision rashfords in behind wow how has it not gone in</p>

filename	content	cleaned_content
	<p>00:02:27,905\nIt proves to be the right decision.\n\n47 h00:02:33,672 --&gt;      00:02:35,512 Rashford's in behind.\n\n48 h00:02:36,887 --&gt;      00:02:38,298 nWow, how has it not gone in?\n\n49 h00:02:38,400 --&gt; 00:02:39,789 nIn the end, it has.\n\n50 h00:02:42,429 --&gt;      00:02:45,963 nIt looked like the defender rushing back.\n\n51 h00:02:46,429 --&gt;      00:02:48,210 nI see he got the final touch.\n\n52 h00:02:49,330 --&gt;      00:02:51,687 Rashford's done this several times tonight\n\n53 h00:02:51,752 --&gt;      00:02:53,374 njust easing past his man.\n\n54 h00:02:55,512 --&gt;      00:02:59,469 nAn instinctive reaction by Šofranac.\n\n55 h00:03:06,501 --&gt;      00:03:08,152 nHe might get a tap in this time.\n\n56 h00:03:08,283 --&gt;      00:03:11,447 nTammy Abraham opens his England account.\n\n57 h00:03:11,636 --&gt;      00:03:14,447 nit is seven with seven minutes to go.\n\n58 h00:03:14,698 --&gt;      00:03:18,000 nYes, a big moment for him and a classic\n\n59 h00:03:18,058 --&gt;      00:03:19,560 ncenter forward's goal this one.\n\n60 h00:03:19,610 --&gt;      00:03:21,229 nMaking that dash across your man.\n\n61 h00:03:21,323 --&gt;      00:03:22,800 nsticking out a leg.\n\n62 h00:03:23,650 --&gt;      00:03:25,432 nWonderful stuff from Abraham.\n\n63 h00:03:25,898 --&gt;      00:03:27,512 nIn an ideal world\n\n64 h00:03:28,050 --&gt;      00:03:29,738 nhe could get on the score sheet.\n\n65 h00:03:29,898 --&gt;      00:03:31,781 nget him some belief at this level.\n\n66 h00:03:32,676 --&gt;      00:03:34,781 nIt finishes England seven.\n\n67 h00:03:34,780 --&gt;      00:03:36,080 nMontenegro nil.\n\n68 h00:03:36,174 --&gt;      00:03:38,160 nThey qualify for the finals.\n\n69 h00:03:41,665 --&gt;      00:03:45,883 n[music]\n\n</p>	<p>in the end it has it looked like the defender rushing back i see he got the final touch rashfords done this several times tonight just easing past his man an instinctive reaction by Šofranac he might get a tap in this time tammy abraham opens his england account it is seven with seven minutes to go yes a big moment for him and a classic center forwards goal this one j making that dash across your man sticking out a leg wonderful stuff from abraham in an ideal world he could get on the score sheet get him some belief at this level it finishes england seven montenegro nil they qualify for the finals music</p>

1118 dJMNNQHkz0M.mp4.srt

1|n00:00:02,068 --> 00:00:04,994|nMatch begins with Torino very much all won \n\n2|h00:00:05,085 --> 00:00:09,691|nfive consecutive clean sheets victories\n over Inter and last week, Atalanta. \n\n3|h00:00:09,845 --> 00:00:13,011|nToday begins a trio of fixtures\n against the relegation side. \n\n4|h00:00:13,188

match begins with torino very much all won five consecutive clean sheets victories over inter and last

filename	content	cleaned_content
	--> 00:00:15,600\nChievo are expected to be beaten.\n\n5\n00:00:16,365 --> 00:00:17,720\nThe first whistle goes.\n\n6\n00:00:17,800 --> 00:00:20,605\nIt is Chievo in the yellow\n who kick us off.\n\n7\n00:00:22,125 --> 00:00:24,908\nThe cross is in.\n Whipped in to De Silvestri.\n\n8\n00:00:25,451 --> 00:00:28,308\nThe shot from range.\n Ansaldi converts.\n\n9\n00:00:28,388 --> 00:00:33,485\n[cheering]\n\n10\n00:00:33,577 --> 00:00:35,908\nTwo goals for him this season,\n the Argentinian.\n\n11\n00:00:36,005 --> 00:00:39,194\n[cheering]\n\n12\n00:00:39,490 --> 00:00:44,880\nNow, a round of applause and a pause\n for the memory of Davide Astori,\n\n13\n00:00:45,040 --> 00:00:51,000\nthe former Fiorentina player\n who so tragically lost his life last year.\n\n14\n00:00:51,091 --> 00:00:56,108\n[applause]\n\n15\n00:00:56,217 --> 00:00:58,697\n[cheering]\n\n16\n00:00:58,737 --> 00:00:59,765\nThis is better now.\n\n17\n00:00:59,828 --> 00:01:04,908\nAnsaldi, Meíté who wasn't wise enough\n to trick Diousse.\n\n18\n00:01:05,011 --> 00:01:07,611\n[cheering]\n\n19\n00:01:07,708 -> 00:01:11,022\nMeggiolini, Đorđević to his right.\n He has a go at goal.\n\n20\n00:01:12,148 --> 00:01:17,400\nA goal that he used to shoot at himself\n when wearing the colors of Torino.\n\n21\n00:01:18,422 --> 00:01:22,102\nPlayed for them between 2012, 2014.\n\n22\n00:01:22,200 --> 00:01:25,908\n[cheering]\n\n23\n00:01:26,262 --> 00:01:28,708\nWe're back underway.\n Torino kick us off.\n\n24\n00:01:28,800 --> 00:01:32,714\n[cheering]\n\n25\n00:01:32,811 --> 00:01:35,897\nDepaoli shifting inside.\n Trying to make the angle for the cross.\n\n26\n00:01:35,960 --> 00:01:38,897\nTook a deflection.\n Should score. Sirigu denies.\n\n27\n00:01:39,651 --> 00:01:42,708\nĐorđević with the best chance\n of the game.\n\n28\n00:01:42,817 --> 00:01:43,994\n[cheering]\n\n29\n00:01:44,194 --> 00:01:46,211\nHe keeps that clean sheet record.\n\n30\n00:01:47,360 --> 00:01:50,308\nIt was an excellent reaction\n save close to him.\n\n31\n00:01:51,137 --> 00:01:54,497\nĐorđević with the second attempt.\n Simply had to score.\n\n32\n00:01:54,594 --> 00:01:57,897\n[cheering]\n\n33\n00:01:57,977 --> 00:01:59,188\n[whistle blows]\n\n34\n00:01:59,662 --> 00:02:02,022\nNow, this could be the moment.\n Zaza.\n\n35\n00:02:02,377 --> 00:02:06,611\nSorrentino, wonderful save.\n\n36\n00:02:06,725 --> 00:02:19,120\n[cheering]\n\n37\n00:02:19,234 --> 00:02:22,800\nBelotti with the strike,\n and Belotti with the missile.\n\n38\n00:02:23,268 --> 00:02:25,000\nTorino take the lead.\n\n39\n00:02:25,154 --> 00:02:28,520\nWhat a vital goal.\n It's the captain's	week atlanta today begins a trio of fixtures against the relegation side chievo are expected to be beaten the first whistle goes it is chievo in the yellow who kick us off the cross is in whipped in to de silvestri the shot from range ansaldi converts cheering two goals for him this season the argentinian cheering now a round of applause and a pause for the memory of davide astori the former fiorentina player who so tragically lost his life last year applause cheering this is better now ansaldi meité who wasnt wise enough to trick diousse cheering meggiolini đorđević to his right he has a go at goal a goal that he used to shoot at himself when wearing the colors of torino played for them between cheering were back underway torino kick us off cheering depaoli shifting inside trying to make the angle for the cross took a deflection should score sirigu denies đorđević with the best chance of the game cheering he keeps that clean sheet record it was an excellent reaction save close to him t đorđević with the second attempt

filename	content	cleaned_content
	<p>goal.\n\n40\n00:02:28,611 --&gt; 00:02:31,017\n[cheering]\n\n41\n00:02:31,097 --&gt; 00:02:35,325\nNumber eight of the season.\nSurely, the best of the lots.\n\n42\n00:02:35,428 --&gt; 00:02:38,400\nA cracker from Andrea Belotti.\n\n43\n00:02:38,508 --&gt; 00:02:44,502\n[cheering]\n\n44\n00:02:44,805 --&gt; 00:02:47,297\nNow, looking to just finish off\nthis match. \n\n45\n00:02:47,594 --&gt; 00:02:50,302\nTomás Rincón. Tomás Rincón.\n\n46\n00:02:50,377 --&gt; 00:02:51,205\n[cheering]\n\n47\n00:02:51,297 --&gt; 00:02:53,205\nThat's a thing of beauty.\n\n48\n00:02:53,308 --&gt; 00:02:56,937\n[cheering]\n\n49\n00:02:57,097 --&gt; 00:02:58,565\nHe comes off the bench.\n\n50\n00:02:58,651 --&gt; 00:03:01,708\nHe was supposed to put solidity\ninto the midfield.\n\n51\n00:03:01,794 --&gt; 00:03:07,314\nIn the end, he's slid\na beautiful strike into the top corner. \n\n52\n00:03:07,680 --&gt; 00:03:10,605\nTorino two,\nChievo nil in injury time.\n\n53\n00:03:10,708 --&gt; 00:03:20,600\n[cheering]\n\n54\n00:03:20,725 --&gt; 00:03:25,005\nZaza, looks for number three\nand in the net it goes.\n\n55\n00:03:25,691 --&gt; 00:03:28,720\nTwo goals in injury time. \n\n56\n00:03:28,931 --&gt; 00:03:31,691\nIt's Torino three, Chievo nil.\n\n57\n00:03:31,788 --&gt; 00:03:36,022\nZaza scores home and away\nagainst Chievo this season. \n\n58\n00:03:36,651 --&gt; 00:03:40,280\nHe's had to wait since the end\nof September\nfor that second goal.\n\n59\n00:03:40,508 --&gt; 00:03:42,897\nThe final whistle goes.\n\n60\n00:03:42,960 --&gt; 00:03:46,177\nIt finishes here Torino three,\nChievo nil.\n\n61\n00:03:55,348 --&gt; 00:04:14,410\n[cheering]\n</p>	simply had to score cheering whistle blows now this could be the moment zaza sorrentino wonderful save cheering belotti with the strike and belotti with the missile torino take the lead what a vital goal its the captains goal cheering number eight of the season surely the best of the lots a cracker from andrea belotti cheering now looking to just finish off this match tomás rincón tomás rincón cheering thats a thing of beauty cheering he comes off the bench he was supposed to put solidity into the midfield in the end hes slid a beautiful strike into the top corner torino two chievo nil in injury time cheering zaza looks for number three and in the net it goes two goals in injury time its torino three chievo nil zaza scores home and away against chievo this season hes had to wait since the end of september for that second goal the final whistle goes it finishes here torino three chievo nil cheering

filename	content	cleaned_content
1119 xK5R8h7zIn0.mp4.srt	<p>1\n00:00:08,885 --&gt;  00:00:11,055\n[applause]\n-Here is the handball, Michael Oliver\n\n2\n00:00:11,223 --&gt;  00:00:12,303\nwaves it away.\n\n3\n00:00:12,471 --&gt;  00:00:16,510\nHere is VAR trying to give us\nan alternative version of it.\n\n4\n00:00:19,134 --&gt;  00:00:20,650\nThen goes on with Mane.\n\n5\n00:00:24,237 --&gt;  00:00:25,603\nStones can't stop the cross.\n\n6\n00:00:25,937 --&gt;  00:00:27,032\nSalah is coming in.\n\n7\n00:00:29,412 --&gt;  00:00:30,755\nFabinho has a go and what a goal.\n\n8\n00:00:30,938 --&gt;  00:00:32,532\n[applause]\n-What a goal,\n\n9\n00:00:32,700 --&gt;  00:00:33,861\nwhat a start\n\n10\n00:00:35,605 --&gt;  00:00:37,796\nfor Liverpool.\n\n11\n00:00:37,964 --&gt;  00:00:39,368\nIt flew in.\n\n12\n00:00:42,583 --&gt;  00:00:43,813\nWell, there's a VAR check.\n\n13\n00:00:44,241 --&gt;  00:00:45,258\nWhat'd he do though?\n\n14\n00:00:48,337 --&gt;  00:00:49,432\nHere is the handball.\n\n15\n00:00:49,940 --&gt;  00:00:50,940\nThere.\n\n16\n00:00:52,529 --&gt;  00:00:53,656\nIt's going to be a goal.\n\n17\n00:00:53,824 --&gt;  00:00:55,507\nHe's adamant it was a penalty.\n\n18\n00:00:56,960 --&gt;  00:00:58,841\nPoor from Bravo\nbut it's a fantastic shot\n\n19\n00:00:59,009 --&gt;  00:01:00,056\nfrom Fabinho.\n\n20\n00:01:07,501 --&gt;  00:01:08,573\n-Here's Robertson,\n\n21\n00:01:08,914 --&gt;  00:01:10,390\nSalah coming in.\n\n22\n00:01:10,558 --&gt;  00:01:12,236\n[applause]\n-Liverpool\n\n23\n00:01:12,404 --&gt;  00:01:14,491\nand another counter punch.\n\n24\n00:01:16,515 --&gt;  00:01:19,020\nMo Salah drive it in a header.\n\n25\n00:01:19,188 --&gt;  00:01:20,322\n-What a finish.\n\n26\n00:01:23,894 --&gt;  00:01:24,902\nMartin Angelino,\n\n27\n00:01:26,919 --&gt;  00:01:27,979\nAguero,\n\n28\n00:01:28,587 --&gt;  00:01:31,583\nstill going, and still not scoring.\n\n29\n00:01:35,017 --&gt;  00:01:36,326\n-I can see here Aguero,\n\n30\n00:01:36,604 --&gt;  00:01:37,674\na little nutmeg from Fabinho,\n\n31\n00:01:37,842 --&gt;  00:01:39,954\nand then you expect him to do better.\n\n32\n00:01:40,122 --&gt;  00:01:42,415\nYou expect him to set it out into that far corner.\n\n33\n00:01:43,391 --&gt;</p>	applause here is the handball michael oliver waves it away here is var trying to give us an alternative version of it then goes on with mane stones cant stop the cross salah is coming in fabinho has a go and what a goal applause what a goal what a start for liverpool it flew in well theres a var check whtad he do though here is the handball there its going to be a goal hes adamant it was a penalty poor from bravo but its a fantastic shot from fabinho heres robertson salah coming in applause liverpool land another counter punch mo salah drive it in a header what a finish martin angelino aguero still going and still not scoring i can see here aguero a little nutmeg from fabinho and then you expect him to do better you expect him to set it out into that far corner martin its a throwin again all liverpool body is back there martin angelino trying to get through thats so close its a corner so close martin henderson done very well to get the cross in

	filename	content	cleaned_content
		<p>00:01:44,534\nMartin: It's a throw-in again.\n\n34\n00:01:44,955 --&gt; 00:01:46,939\n-All Liverpool body is back there.\n\n35\n00:01:48,711 --&gt; 00:01:50,695\nMartin: Angelino\ntrying to get through.\n\n36\n00:01:51,434 --&gt; 00:01:53,942\nThat's so, so close. It's a corner.\n\n37\n00:01:54,656 --&gt; 00:01:55,957\n-So close.\n\n38\n00:02:00,474 --&gt; 00:02:01,490\nMartin: Henderson,\n\n39\n00:02:01,965 --&gt; 00:02:02,965\nndone very well to get the cross in.\n\n40\n00:02:03,133 --&gt; 00:02:05,521\nSo well for Sadio Mane,\n\n41\n00:02:05,737 --&gt; 00:02:07,157\nncame in at the back post\n\n42\n00:02:08,483 --&gt; 00:02:10,689\nand squeezed it off the goalkeeper.\n\n43\n00:02:18,537 --&gt; 00:02:20,194\n[?].\n\n44\n00:02:20,362 --&gt; 00:02:21,696\nAgain, Angelino\n\n45\n00:02:23,498 --&gt; 00:02:25,966\nand they've got one.\nThey do always score.\n\n46\n00:02:26,435 --&gt; 00:02:28,086\nBernardo Silva swept it in.\n\n47\n00:02:28,673 --&gt; 00:02:30,324\nNot over yet at Anfield.\n\n48\n00:02:30,603 --&gt; 00:02:31,690\nHere's Sterling,\n\n49\n00:02:32,635 --&gt; 00:02:34,143\nhas never given up the ghost.\n\n50\n00:02:34,461 --&gt; 00:02:36,762\nShouts for handball hits\n\nAlexander-Arnold again,\n\n51\n00:02:36,930 --&gt; 00:02:39,643\nand then Henderson comes\nand grabs it.\n\n52\n00:02:39,811 --&gt; 00:02:41,382\nGabriel Jesus is hurt.\n\n53\n00:02:42,065 --&gt; 00:02:43,255\nHe's nodding his head.\n\n54\n00:02:44,263 --&gt; 00:02:46,096\nMaybe that's the anger\nfrom the first half.\n\n55\n00:02:47,057 --&gt; 00:02:48,231\nMartin: Angelino again.\n\n56\n00:02:49,627 --&gt; 00:02:52,175\n[?].\n\n57\n00:02:56,007 --&gt; 00:02:57,611\nJust too high for Bernardo.\n\n</p>	so well for sadio mane came in at the back post and squeezed it off the goalkeeper again angelino and theyve got one they do always score bernardo silva swept it in not over yet at anfield heres sterling has never given up the ghost shouts for handball hits alexanderarnold again and then henderson comes and grabs it gabriel jesus is hurt hes nodding his head maybe thats the anger from the first half martin angelino again just too high for bernardo
1120	qG1IIIj9_fU.mp4.srt	<p>1\n00:00:06,913 --&gt; 00:00:12,072\n-Played by Firmino to Andy Robertson,\njust beyond the stretch of Mane,\n\n2\n00:00:12,639 --&gt; 00:00:14,376\nbut controlled there by Mo Salah.\n\n3\n00:00:14,735 --&gt; 00:00:17,185\n[?] to Mane, Lallana goes wide.\n\n4\n00:00:17,352 --&gt; 00:00:21,934\nAlexander-Arnold [?] then Salah,\nposes an early threat to Wolves.\n\n5\n00:00:22,875 --&gt; 00:00:25,226\n-A lot of bodies central\non the mid-field today,\n\n6\n00:00:25,489 --&gt; 00:00:27,040\nwe need to see that a lot more.\n\n7\n00:00:28,027 --&gt; 00:00:29,027\nHere's Henderson.\n\n8\n00:00:31,892 --&gt; 00:00:33,865\nFirmino finds Alexander-Arnold.\n\n9\n00:00:34,352 --&gt; 00:00:36,886\n[?] by but not too far.\n\n10\n00:00:37,053 --&gt; 00:00:39,003\nit's in to Sadio Mane\nwho's on the turn,\n\n11\n00:00:39,970 --&gt; 00:00:42,222\nand Rui Patricio scrambling\nacross his</p>	played by firmino to andy robertson just beyond the stretch of mane but controlled there by mo salah to mane lallana goes wide alexanderarnold then salah poses an early threat to wolves a lot of bodies central on the midfield today we need to see that a lot more heres henderson firmino finds alexanderarnold by but not too far its in to sadio mane whos on

filename	content	cleaned_content
	<p>goal.\n\n12\n00:00:47,494 --&gt; 00:00:48,880\nThe ball going on the top onto Mane.\n\n13\n00:00:49,047 --&gt; 00:00:51,974\n[crowd cheering]\n\n14\n00:00:52,141 --&gt; 00:00:56,923\n-There was a touch on the arm,\nhand it's not going to count.\n\n15\n00:00:59,838 --&gt; 00:01:01,239\n-He's certainly not offside.\n\n16\n00:01:02,572 --&gt; 00:01:04,177\n[whistle blows]\n\n17\n00:01:08,061 --&gt; 00:01:09,217\n- The goal stands.\n\n18\n00:01:10,269 --&gt; 00:01:13,611\n-Sadio Mane has made\nthe breakthrough here for Liverpool.\n\n19\n00:01:14,242 --&gt; 00:01:15,359\n-Won by Moutinho.\n\n20\n00:01:15,726 --&gt; 00:01:17,926\n-Jonny trying to break free\nof Andy Robertson,\n\n21\n00:01:18,192 --&gt; 00:01:20,288\n-and manages to pull it back,\nnit's gone all the way across\n\n22\n00:01:20,455 --&gt; 00:01:21,561\n-and Neto\n\n23\n00:01:24,078 --&gt; 00:01:28,546\n-stuns Anfield with an equalizer\nright on the brink\n\n24\n00:01:28,713 --&gt; 00:01:30,451\nnof half-time.\n\n25\n00:01:31,806 --&gt; 00:01:35,678\n-Meanwhile, there's frustration\nfrom the first [?]\n\n26\n00:01:35,845 --&gt; 00:01:37,430\nbut that will be proven to be right.\n\n27\n00:01:37,597 --&gt; 00:01:39,998\n-We're still waiting\nfor confirmation on this one.\n\n28\n00:01:45,320 --&gt; 00:01:46,793\n-Brace yourselves.\n\n29\n00:01:47,513 --&gt; 00:01:48,763\n-The decision has come in.\n\n30\n00:01:49,423 --&gt; 00:01:53,096\n-Wolves are going\nonto be fuming about this,\n\n31\n00:01:53,263 --&gt; 00:01:55,590\n-the goal has been disallowed.\n\n32\n00:02:01,060 --&gt; 00:02:04,649\n-That was over the top\nof the bar from Wijnaldum.\n\n33\n00:02:11,793 --&gt; 00:02:15,182\n-van Dijk in trouble,\nhe's lost that here to Diogo Jota.\n\n34\n00:02:15,424 --&gt; 00:02:17,433\n-He produces the save.\n\n35\n00:02:20,520 --&gt; 00:02:22,046\n-Certainly, this time,\nit's the goalkeeper who saves\n\n36\n00:02:22,213 --&gt; 00:02:23,213\n-Virgil van Dijk.\n\n37\n00:02:28,430 --&gt; 00:02:30,780\n-Alexander-Arnold,\nought to be careful, Jimenez.\n\n</p>	<p>the turn and rui patricio scrambling across his goal the ball going on the top onto mane crowd cheering there was a touch on the arm and its not going to count hes certainly not offside whistle blows the goal stands sadio mane has made the breakthrough here for liverpool won by moutinho jonny trying to break free of andy robertson and manages to pull it back its gone all the way across and neto stuns anfield with an equalizer right on the brink of halftime meanwhile theres frustration from the first but that will be proven to be right were still waiting for confirmation on this one brace yourselves the decision has come in wolves are going to be fuming about this the goal has been disallowed that was over the top of the bar from wijnaldum van dijk in trouble hes lost that here to diogo jota he produces the save certainly this time its the goalkeeper who saves virgil van dijk alexanderarnold ought to be careful jimenez</p>

filename	content	cleaned_content
1121 GsY4w7sjh9A.mp4.srt	<p>1\n00:00:00,171 --&gt;  00:00:04,057\n[music]\n\n2\n00:00:05,657 --&gt;  00:00:06,800\nWell, the contrast is  stark.\n\n3\n00:00:06,942 --&gt;  00:00:12,342\nSeven of this starting 11 for  wales \nhave single figures in terms of  caps,\n\n4\n00:00:12,828 --&gt;  00:00:14,600\nincluding this fella, David  Brooks.\n\n5\n00:00:14,971 --&gt;  00:00:16,228\nFirst chance for him to  run,\n\n6\n00:00:16,514 --&gt;  00:00:18,714\ncouldn't quite play it into  \nGareth Bale in time,\n\n7\n00:00:18,800 --&gt;  00:00:20,114\nbut there's good pressure  there.\n\n8\n00:00:20,257 --&gt;  00:00:26,571\nOh, first start for Wales, \nfirst  goal for Wales.\n\n9\n00:00:26,971 --&gt;  00:00:30,057\nDan James inside the first five  minutes.\n\n10\n00:00:31,828 --&gt;  00:00:35,000\nThat is not far from\nthe stuff  of dreams.\n\n11\n00:00:37,571 --&gt;  00:00:40,828\nGreat work by David Brooks  \nwho runs with the ball so  well,\n\n12\n00:00:41,771 --&gt;  00:00:44,942\nand then this young exuberant  \nWales side win the ball  back.\n\n13\n00:00:45,200 --&gt;  00:00:48,800\nGreat strike \non his  competitive debut for  Wales.\n\n14\n00:00:49,571 --&gt;  00:00:53,171\nAbsolutely incredible moment  \nfor him and his family,\n\n15\n00:00:53,657 --&gt;  00:00:55,742\nand a huge start for Ryan  Giggs.\n\n16\n00:00:56,971 --&gt;  00:00:58,457\nAllen out to James  here.\n\n17\n00:01:01,257 --&gt;  00:01:04,828\nGood cross from Ben Davies.  \nOh, side netting.\n\n18\n00:01:05,371 --&gt;  00:01:08,571\nDavid Brooks who played such  \na key role is setting up the  first\n\n19\n00:01:08,880 --&gt;  00:01:10,742\nvery nearly claiming a  second.\n\n20\n00:01:11,571 --&gt;  00:01:13,371\nGreat overlap from Ben  Davies.\n\n21\n00:01:14,085 --&gt;  00:01:14,828\nIt's a really good  cross\n\n22\n00:01:14,914 --&gt;  00:01:17,057\nand it's good contact \nas well  from David Brooks.\n\n23\n00:01:20,885 --&gt;  00:01:21,857\nBrooks  now.\n\n24\n00:01:23,000 --&gt;  00:01:24,771\nThat's a lovely touch\n to take  him beyond Rusnák.\n\n25\n00:01:24,885 --&gt;  00:01:25,914\nStill at it,  Brooks.\n\n26\n00:01:30,828 --&gt;  00:01:33,885\nThis young man just loves  playing\nin the shirt of his  country.\n\n27\n00:01:34,828 --&gt;  00:01:39,200\nRecently crowned Player of the  Year,\n and that was a lovely piece of  play.\n\n28\n00:01:44,800 --&gt;  00:01:47,714\nBale rises high but couldn't \nget  enough on the header\n\n29\n00:01:48,171 --&gt;</p>	<p>music well the  contrast is stark  seven of this  starting for wales  have single  figures in terms  of caps including  this fella david  brooks first  chance for him to  run couldnt quite  play it into gareth  bale in time but  theres good  pressure there oh  first start for  wales first goal  for wales dan  james inside the  first five minutes  that is not far  from the stuff of  dreams great  work by david  brooks who runs  with the ball so  well and then this  young exuberant  wales side win  the ball back  great strike on  his competitive  debut for wales  absolutely  incredible  moment for him  and his family  and a huge start  for ryan giggs  allen out to james  here good cross  from ben davies  oh side netting  david brooks who  played such a key  role is setting up  the first very  nearly claiming a  second great  overlap from ben  davies its a really  good cross and  its good contact  as well from  david brooks  brooks now thats  a lovely touch to  take him beyond  rusnák still at it  brooks this young  man just loves  playing in the  shirt of his</p>

filename	content	cleaned_content
	<p>00:01:49,714 nto divert it goalwards.\n\n30 n00:01:51,085 --&gt; 00:01:52,914 nDubravka commits himself.\n\n31 n00:01:53,057 --&gt; 00:01:55,971 nlf Gareth Bale could just get above the ball a little bit more,\n\n32 n00:01:56,800 --&gt; 00:01:58,685 nit would been a certain goal for Wales.\n\n33 n00:01:59,542 --&gt; 00:02:01,542 nGoalkeeper caught in no man's land.\n\n34 n00:02:08,542 --&gt; 00:02:10,428 nMak here for Slovakia, \nhe's sent him through.\n\n35 n00:02:10,485 --&gt; 00:02:11,628 nOut comes Hennessey.\n\n36 n00:02:13,485 --&gt; 00:02:15,885 nHennessey made the block, &amp; then from the follow up,\n\n37 n00:02:15,971 --&gt; 00:02:17,714 nit went into the side netting.\n\n38 n00:02:18,000 --&gt; 00:02:20,828 nFirst time Wales \nhave really been cut apart.\n\n39 n00:02:21,542 --&gt; 00:02:22,771 nlIt's a good run.\n\n40 n00:02:23,542 --&gt; 00:02:27,228 nWayne Hennessey came out of his goal \nso quickly and shut down the angle.\n\n41 n00:02:27,820 --&gt; 00:02:29,542 nRusnák, as I said, \nhe makes that run from\n\n42 n00:02:30,171 --&gt; 00:02:31,628 nthe right-hand side across \nthe defense.\n\n43 n00:02:31,714 --&gt; 00:02:34,971 nDoes it so well, tries to just \nlift it over Wayne Hennessey.\n\n44 n00:02:35,457 --&gt; 00:02:39,742 nBig, strong, aggressive arm &amp; came out towards him.\n\n45 n00:02:41,600 - -&gt; 00:02:42,771 nRusnák.\n\n46 n00:02:45,771 --&gt; 00:02:47,828 nDuda. Hamšík.\n\n47 n00:02:48,942 --&gt; 00:02:53,371 nTidy turn but saved from Robert Mak's \nneffort by Wayne Hennessey.\n\n48 n00:02:54,371 --&gt; 00:02:56,628 nTerrific football from \nthe Slovakians once again.\n\n49 n00:02:57,228 --&gt; 00:02:59,200 nManoeuvre the ball \nfrom left to right really well,\n\n50 n00:02:59,542 --&gt; 00:03:01,142 nthen back across the pitch again.\n\n51 n00:03:01,457 --&gt; 00:03:03,400 nlIt's a good slide. \nlIt's a good save by Wayne.\n\n52 n00:03:04,285 --&gt; 00:03:06,085 nJoe Allen.\n\n53 n00:03:08,428 --&gt; 00:03:10,485 nBen Davies to Gareth Bale.\n\n54 n00:03:12,628 --&gt; 00:03:14,857 nLovely whip on that ball.\n\n55 n00:03:15,628 --&gt; 00:03:17,085 nAll the way through to Wilson.\n\n56 n00:03:17,628 --&gt; 00:03:19,000 nlIt might come for Tyler Roberts.\n\n57 n00:03:19,514 --&gt; 00:03:22,342 nHe was offside when it was hit.\n\nThe flag's up on the far side.\n\n58 n00:03:23,314 --&gt; 00:03:26,571 nMagnificent cross from Gareth Bale. \nPut it in such a dangerous area,\n\n59 n00:03:26,657 --&gt; 00:03:29,371 njust a little bit too high \nfor Tyler Roberts.\n\n60 n00:03:30,571 --&gt; 00:03:32,285 nNice first touch by Harry Wilson.\n\n61 n00:03:35,485 --&gt; 00:03:37,771 nRusnák is full of running here for Slovakia.\n\n62 n00:03:38,085 --&gt;</p>	<p>country recently crowned player of the year and that was a lovely piece of play bale rises high but couldnt get enough on the header to divert it goalwards dubravka commits himself if gareth bale could just get above the ball a little bit more it would been a certain goal for wales goalkeeper caught in no mans land mak here for slovakia hes sent him through out comes hennessey henessey made the block and then from the follow up it went into the side netting first time wales have really been cut apart its a good run wayne henessey came out of his goal so quickly and shut down the angle rusnák as i said he makes that run from the righthand side across the defense does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda hamšík tidy turn but saved from robert maks effort by wayne henessey terrific football from the slovakians once again manoeuvre the ball from left to right really well then back across the pitch again its a good slide its a good save by wayne joe allen ben davies to gareth bale lovely</p>

filename	content	cleaned_content
	<p>00:03:39,628\nTime to pick his pass.\n\n63\n00:03:39,714 --&gt; 00:03:42,257\nGot the header hard \nand that shot across there.\n\n64\n00:03:45,342 --&gt; 00:03:47,485\nMatt Smith just didn't quite go with the runner.\n\n65\n00:03:48,457 --&gt; 00:03:49,571\nGood cross in.\n\n66\n00:03:50,142 --&gt; 00:03:53,314\nIt's a wonderful header, \nisn't it, by Ďuriš.\n\n67\n00:03:55,171 --&gt; 00:03:56,942\nIt was Rusnák with the cross.\n\n68\n00:03:57,228 --&gt; 00:03:58,942\nSmart save by Hennessey.\n\n69\n00:03:59,657 --&gt; 00:04:03,342\nand then a huge chance \nto get Slovakia back into the game.\n\n70\n00:04:04,571 --&gt; 00:04:06,742\nUnfortunately for the visitors, \nit goes begging.\n\n</p>	whip on that ball all the way through to wilson it might come for tyler roberts he was offside when it was hit the flags up on the far side magnificent cross from gareth bale puts it in such a dangerous area just a little bit too high for tyler roberts nice first touch by harry wilson rusnák is full of running here for slovakia time to pick his pass got the header hard and that shot across there matt smith just didnt quite go with the runner good cross in its a wonderful header isnt it by ďuriš it was rusnák with the cross smart save by hennessey and then a huge chance to get slovakia back into the game unfortunately for the visitors it goes begging

	filename	content	cleaned_content
1122	R0XfelUhG_Y.mp4.srt	1\n00:00:00,429 --> 00:00:03,888\nHello, and welcome to \nthe Stadio Olimpico in\n\n2\n00:00:04,044 --> 00:00:07,878\nRome for round 31 of \nthe Serie A season,\n\n3\n00:00:07,982 --> 00:00:10,259\nit's Lazio versus Sassuolo.\n\n4\n00:00:10,453 --> 00:00:13,367\nI'm Patrick Terrone your \ncommentator for this game.\n\n5\n00:00:13,967 --> 00:00:15,721\nIt is the home side, Lazio,\n\n6\n00:00:15,746 --> 00:00:17,404\nwho get in the first half underway.\n\n7\n00:00:17,691 --> 00:00:18,818\n[whistle]\n\n8\n00:00:22,077 --> 00:00:23,823\nIt's Whipped in and \nthat's a good header\n\n9\n00:00:24,000 --> 00:00:25,785\nand a good save by Consigli\n\n10\n00:00:25,945 --> 00:00:27,438\nwho watched it all the way,\n\n11\n00:00:28,021 --> 00:00:29,021\nLazio\n\n12\n00:00:29,203 --> 00:00:31,243\nwith an early chance then.\n\n13\n00:00:35,091 --> 00:00:37,052\nIt's been a little bit \nndisjointed that.\n\n14\n00:00:37,610 --> 00:00:39,467\nOh, that flick Caicedo.\n\n15\n00:00:39,737 --> 00:00:41,284\nOh, almost in the back of the net,\n\n16\n00:00:42,104 --> 00:00:43,159\na rebound\n\n17\n00:00:43,756 --> 00:00:44,978\noff Immobile.\n\n18\n00:00:46,738 --> 00:00:48,230\n[whistle]\n\n19\n00:00:48,349 --> 00:00:49,802\nwho get the second half underway.\n\n20\n00:00:51,013 --> 00:00:52,021\nLulic,\n\n21\n00:00:53,187 --> 00:00:55,941\nsquaring the ball to Patrick\n\n22\n00:00:56,532 --> 00:00:58,366\nIt's now out wide to Marusic.\n\n23\n00:00:59,219 --> 00:01:00,791\nPatrick with a chance to cross,\n\n24\n00:01:01,437 --> 00:01:03,460\nhand ball and that's a penalty kick,\n\n25\n00:01:03,595 --> 00:01:04,603\nsays the referee.\n\n26\n00:01:07,100 --> 00:01:09,180\nWell, the cross came from Patrick\n\n27\n00:01:11,685 --> 00:01:13,764\nand he's going to go over \nand have a look, so\n\n28\n00:01:14,693 --> 00:01:16,026\nnot an easy decision\n\n29\n00:01:16,370 --> 00:01:17,385\nfor the ref to make.\n\n30\n00:01:21,743 --> 00:01:24,513\nReferee has confirmed \nit's a penalty kick.\n\n31\n00:01:25,221 --> 00:01:26,650\nImmobile then up against\n\n32\n00:01:26,835 --> 00:01:28,644\nthe Consigli, here he goes.\n\n33\n00:01:28,958 --> 00:01:31,832\nImmobile has given Lazio the lead,\n\n34\n00:01:32,597 --> 00:01:34,549\nnone-nil to the home side.\n\n35\n00:01:34,811 -->	hello and welcome to the stadio olimpico in rome for round of the serie a season its lazio versus sassuolo im patrick terrone your commentator for this game it is the home side lazio who get in the first half underway whistle its whipped in and thats a good header and a good save by consigli who watched it all the way lazio with an early chance then its been a little bit disjointed that oh that flick caicedo oh almost in the back of the net a rebound off immobile whistle its sassuolo who get the second half underway lulic squaring the ball to patrick who comes forward its now out wide to marusic patrick with a chance to cross hand ball and thats a penalty kick says the referee well the cross came from patrick and hes going to go over and have a look so not an easy decision for the ref to make referee has confirmed its a penalty kick immobile then up against the consigli here he goes immobile has given lazio the lead onenil to the home side ciro immobile goal number for the striker an easy finish

filename	content	cleaned_content
	<p>00:01:36,144\nCiro Immobile,\n\n36\n00:01:36,782 --&gt; 00:01:39,131\ngoal number 14 for the striker.\n\n37\n00:01:39,649 --&gt; 00:01:41,816\nAn easy finish sending Consigli.\n\n38\n00:01:42,112 --&gt; 00:01:43,342\nthe wrong way.\n\n39\n00:01:43,853 --&gt; 00:01:44,861\nDemiral,\n\n40\n00:01:45,942 --&gt; 00:01:46,950\nDuncan,\n\n41\n00:01:47,259 --&gt; 00:01:48,435\nit's nicely forward Boga.\n\n42\n00:01:49,222 --&gt; 00:01:50,667\nCan he twist and turn here?\n\n43\n00:01:53,746 --&gt; 00:01:55,412\nHere he goes, \nhe's got the pace,\n\n44\n00:01:55,599 --&gt; 00:01:57,956\nnot a bad effort. \nLirola trying to place it in.\n\n45\n00:01:58,149 --&gt; 00:01:59,934\nit's there, Rogerio,\n\n46\n00:02:00,166 --&gt; 00:02:01,198\n tapped it in\n\n47\n00:02:02,319 --&gt; 00:02:04,017\nand Sassuolo right back in it,\n\n48\n00:02:04,194 --&gt; 00:02:06,821\nfirst chance they've had in\nthe entire match\n\n49\n00:02:07,004 --&gt; 00:02:08,004\nand it's a goal.\n\n50\n00:02:08,728 --&gt; 00:02:10,157\nit is, of all people,\n\n51\n00:02:10,814 --&gt; 00:02:12,369\nRogerio the Brazilian.\n\n52\n00:02:20,022 --&gt; 00:02:21,038\nCorrea,\n\n53\n00:02:21,229 --&gt; 00:02:22,951\noh, that's a nice ball, Immobile,\n\n54\n00:02:23,642 --&gt; 00:02:24,737\noh, what a chance\n\n55\n00:02:25,274 --&gt; 00:02:27,235\nnot give his side\nthe lead once again.\n\n56\n00:02:29,246 --&gt; 00:02:30,278\nCorrea\n\n57\n00:02:31,453 --&gt; 00:02:32,984\nleaves it for Luiz Alberto\n\n58\n00:02:33,272 --&gt; 00:02:35,765\njust [?] it that's a wonderful ball,\n\n59\n00:02:36,283 --&gt; 00:02:37,323\nWhat a miss,\n\n60\n00:02:37,516 --&gt; 00:02:39,595\nCiro Immobile all alone.\n\n61\n00:02:39,817 --&gt; 00:02:41,016\nAll he had to do\n\n62\n00:02:41,357 --&gt; 00:02:42,730\nwas put it in the back of the net\n\n63\n00:02:43,184 --&gt; 00:02:44,264\nand he's missed.\n\n64\n00:02:44,830 --&gt; 00:02:45,869\nLuis Alberto\n\n65\n00:02:46,524 --&gt; 00:02:48,651\nnow to Radu, Radu's cross\n\n66\n00:02:48,916 --&gt; 00:02:50,360\ntowards Milinković-Savić,\n\n67\n00:02:50,733 --&gt; 00:02:53,487\nit's saved off the line by Consigli.\n\n68\n00:02:54,413 --&gt; 00:02:55,445\nGreat header\n\n69\n00:02:56,247 --&gt; 00:02:57,461\nfrom the substitute\n\n70\n00:02:57,922 --&gt; 00:02:58,969\nbut equally\n\n71\n00:02:59,363 --&gt; 00:03:00,379\ngood save.\n\n72\n00:03:03,018 --&gt; 00:03:04,025\nNow, \n\n73\n00:03:05,216 -</p>	<p>sending consigli the wrong way demiral duncan its nicely forward boga can he twist and turn here here he goes hes got the pace not a bad effort lirola trying to place it in its there rogerio tapped it in and sassuolo right back in it first chance theyve had in the entire match and its a goal it is of all people rogerio the brazilian correa oh thats a nice ball immobile oh what a chance to give his side the lead once again correa leaves it for luiz alberto just it thats a wonderful ball immobile what a miss ciro immobile all alone all he had to do was put it in the back of the net and hes missed luis alberto now to radu radus cross towards milinković-savić its saved off the line by consigli great header from the substitute but equally good save now space and berardi through here clear through berardi sassuolo have taken the lead on the break berardi finishing with a plum whipped in flicked out by sensi whos the smallest man on the pitch now patrick out too immobile immobile across the face of goal its there the equalizer stands lulic from a meter</p>

filename	content	cleaned_content
-> 00:03:06,224\inspace\ n\ n74\ n00:03:06,458 -> 00:03:08,839\inand Berardi through here, \ nclear through,\ n\ n75\ n00:03:09,016 --> 00:03:10,841\ nBerardi.\ n\ n76\ n00:03:11,072 --> 00:03:12,500\ nSassuolo\ n\ n77\ n00:03:14,371 --> 00:03:15,935\ nhave taken the lead\ n\ n78\ n00:03:16,387 --> 00:03:17,498\ non the break.\ n\ n79\ n00:03:19,463 --> 00:03:20,574\ nBerardi\ n\ n80\ n00:03:21,058 --> 00:03:22,423\ nfinishing with a plum.\ n\ n81\ n00:03:27,797 --> 00:03:28,813\ nWhipped in,\ n\ n82\ n00:03:29,170 --> 00:03:30,266\ nflicked out by\ n\ n83\ n00:03:30,539 --> 00:03:32,507\ nSensi who's the smallest \nman on the pitch.\ n\ n84\ n00:03:32,643 --> 00:03:33,944\ nNow, Patrick out too,\ n\ n85\ n00:03:34,127 --> 00:03:35,444\ nImmobile, Immobile,\ n\ n86\ n00:03:35,686 --> 00:03:37,463\ nacross the face of goal, it's there.\ n\ n87\ n00:03:37,895 --> 00:03:39,094\ nThe equalizer stands,\ n\ n88\ n00:03:39,451 --> 00:03:40,459\ nLulic\ n\ n89\ n00:03:41,058 --> 00:03:43,121\ nfrom a meter out taps it in.\ n\ n90\ n00:03:43,838 --> 00:03:45,378\ nThis time Immobile\ n\ n91\ n00:03:46,028 --> 00:03:47,267\ nprovides the assist.\ n\ n92\ n00:03:51,187 --> 00:03:52,536\ nThe game has finished,\ n\ n93\ n00:03:52,973 --> 00:03:54,997\ nLazio two, Sassuolo two.\ n\ n94\ n00:03:58,741 --> 00:04:00,471\ n[crowd cheering]\ n\ n	out taps it in this time immobile provides the assist the game i has finished lazio two sassuolo two crowd cheering	

1123 rows x 5 columns

```
In [17]: # Normalize player names in 'df_skin_tones' DataFrame
df_skin_tones['Player Names'] = df_skin_tones['Player Names'].str.lower().str.replace(' ', '_')

# Initialize an empty list to store the images for each player
player_images = []

# Iterate through each row in the DataFrame
for _, row in df.iterrows():
    found_names = row['found_player_names']
    images = []
    for name in found_names:
        # Iterate through each row in df_skin_tones to find matching player
        for _, skin_row in df_skin_tones.iterrows():
            if name.lower() == skin_row['Player Names'].lower():
                # Append the image URL to the images list
                images.append(skin_row['Player Images'])
    player_images.append(images)

# Add the player images as a new column
df['player_images'] = player_images

# Check the DataFrame with the new column
print(df[['filename', 'found_player_names', 'player_images']].head(10))
```

```

        filename \
0  cvctybLyEqE.mp4.srt
1  oCl6MCo0i1Q.mp4.srt
2  _kDc62-upoA.mp4.srt
3  41k02TvmNbg.mp4.srt
4  DjNA-MIdCbY.mp4.srt
5  H0rLlWM30IU.mp4.srt
6  wa-f96ESXwA.mp4.srt
7  7lmXk6g1ito.mp4.srt
8  psMNMM2CpgQ.mp4.srt
9  FZkS3hBabXg.mp4.srt

found_pla
yer_names \
0
Cancelo] [João
1
[]
2
[]
3
[Roberto]
4
Abraham] [Harry Kane, Tammy
5
[]
6
[]
7  [Alex McCarthy, Bernardo, Bernardo Silva, Danny Ings, Kyle Walker, Sergi
o Aguero]
8
[]
9
[Rodri]

player_images
0
[https://cdn.sofifa.net/players/210/514/18_60.png]
1
[]
2
[]
3
[https://cdn.sofifa.net/players/183/666/18_60.png]
4
[https://cdn.sofifa.net/players/202/126/18_60.png, https://cdn.sofifa.net/p
layers/231/352/18_60.png]
5
[]
6
[]
7  [https://cdn.sofifa.net/players/189/324/18_60.png, https://cdn.sofifa.ne
t/players/232/805/18_60.png, https://cdn.sofifa.net/players/218/667/18_60.p
ng, https://cdn.sofifa.net/players/195/859/18_60.png, https://cdn.sofifa.ne
t/players/188/377/18_60.png, https://cdn.sofifa.net/players/153/079/18_60.p
ng]
8
[]
9
[]
```

In [ ]:

The above code has identified the Premier League players mentioned within the transcripts I am analysing, and has provided their image pathways alongside their names.

In [ ]:

```
In [18]: # Load the CSV file containing player names and skin tone labels
df_skin_tones = pd.read_csv('/Users/alexdoyle/Desktop/PL_player_images/output.csv')

# Check the contents of the DataFrame
print(df_skin_tones.head())

      file image type  face id dominant 1 percent 1 dominant
2 \
0   Aaron Cresswell.jpg    color     1.0  #DCA183    0.66  #8968
50
1       Aaron Mooy.jpg    color     1.0  #B99684    0.56  #8A72
61
2     Aaron Ramsdale.jpg    color     1.0  #E2A388    0.57  #A16D
51
3      Aaron Ramsey.jpg    color     1.0  #E2BCA8    0.71  #B286
70
4  Aaron Wan-Bissaka.jpg    color     1.0  #5A2B1F    0.72  #7A4A
44

      percent 2 skin tone tone label  accuracy(0-100)
0      0.34  #BEA07E      CG      86.92
1      0.44  #9D7A54      CF      89.54
2      0.43  #9D7A54      CF      86.16
3      0.29  #E7C1B8      CI      91.61
4      0.28  #513B2E      CC      89.58
```

```
In [19]: # Print the column names to see the actual names
print(df_skin_tones.columns)
```

```
Index(['file', 'image type', 'face id', 'dominant 1', 'percent 1',
       'dominant 2', 'percent 2', 'skin tone', 'tone label',
       'accuracy(0-100)'],
      dtype='object')
```

In [ ]:

```
In [20]: # Assuming the correct columns are 'Player Names' and 'Skin Tone'
df_skin_tones['Player Names'] = df_skin_tones['file'].str.lower().str.strip()
```

```
In [21]: # Assuming the correct columns are 'Player Names' and 'Skin Tone'
df_skin_tones['Player Names'] = df_skin_tones['Player Names'].str.lower().str.strip()
```

In [ ]:

```
In [22]: # Function to extract context
def extract_context(text, name, window=10):
    words = text.lower().split()
    name_parts = name.lower().split()
    name_length = len(name_parts)

    contexts = []

    # Find all occurrences of the player name in the text
    for i in range(len(words) - name_length + 1):
        if words[i:i + name_length] == name_parts:
            start = max(0, i - window)
            end = min(len(words), i + name_length + window)
            context = ' '.join(words[start:end])
            contexts.append(context)

    return contexts
```

```
        contexts.append(context)

    return contexts

# Create a DataFrame to store player contexts
context_data = {'Player': [], 'Context': []}

# Iterate through the DataFrame and extract contexts
for _, row in df.iterrows():
    for name in row['found_player_names']:
        contexts = extract_context(row['cleaned_content'], name)
        for context in contexts:
            context_data['Player'].append(name)
            context_data['Context'].append(context)

# Convert to DataFrame
df_contexts = pd.DataFrame(context_data)
```

```
In [23]: # Normalize player names in df_skin_tones and remove '.jpg' extension
df_skin_tones['Player Names'] = df_skin_tones['file'].str.lower().str.replace('.jpg', '')

# Normalize player names in df_contexts
df_contexts['Player'] = df_contexts['Player'].str.lower().str.strip()

# Remove any prior occurrences of tone label columns
if 'tone label' in df_contexts.columns:
    df_contexts.drop(columns=['tone label'], inplace=True)

# Merge the DataFrames without creating duplicates
df_contexts = df_contexts.merge(
    df_skin_tones[['Player Names', 'tone label']],
    left_on='Player',
    right_on='Player Names',
    how='left'
)

# Drop the extra column used for merging
df_contexts.drop(columns=['Player Names'], inplace=True)

# Check the final DataFrame with the desired columns
print(df_contexts[['Player', 'Context', 'tone label']].head(15))
```

```

    Player \
0  joão cancelo
1  joão cancelo
2  roberto
3  harry kane
4  tammy abraham
5  alex mccarthy
6  alex mccarthy
7  alex mccarthy
8  bernardo
9  bernardo
10 bernardo
11 bernardo silva
12 bernardo silva
13 bernardo silva
14 danny ings

```

Context \

0 oh great skill from bernardeschi for dybala spilled big chance joão can celo with an equalizer for juventus back from injury to save  
 1 and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a  
 2 by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma  
 3 theres no chance for the keeper theyre stretched defensively here har ry kane just makes that point only needed a touch floated in  
 4 by Šofranac he might get a tap in this time tammy abra ham opens his england account it is seven with seven minutes  
 5 target the pace on the ball means he cant react alex mc carthy it goes through his legs hes tiptoeing his way around  
 6 legs hes tiptoeing his way around and its bernardo sil va alex mccarthy sees it up and over well hes not been asked  
 7 well hes not been asked to do an awful lot alex mccarth y bernardo silva gundogan de bruyne inches away from being a  
 8 instinct to get southampton the most unlikely of leads her e bernardo silva aguero is in and he is furious with himself  
 9 through his legs hes tiptoeing his way around and its bernardo silva alex mccarthy sees it up and over well hes  
 10 not been asked to do an awful lot alex mccarthy bernardo s ilva gundogan de bruyne inches away from being a gamechanger  
 11 instinct to get southampton the most unlikely of leads here bernar do silva aguero is in and he is furious with himself because  
 12 through his legs hes tiptoeing his way around and its b ernardo silva alex mccarthy sees it up and over well hes not  
 13 not been asked to do an awful lot alex mccarthy bernardo silva gundogan de bruyne inches away from being a gamechanger oh  
 14 romeu danny ings and redmond has no one to hit in the middle

	tone	label
0		CF
1		CF
2		CF
3		CE
4		CD
5		CF
6		CF
7		CF
8		CD
9		CD
10		CD
11		CE
12		CE

```
13      CE
14      CF
```

```
/var/folders/rb/2l3j3fzj6mng0r5wn3mnhtfm0000gn/T/ipykernel_78534/320831140.py:2: FutureWarning: The default value of regex will change from True to False in a future version.
  df_skin_tones['Player Names'] = df_skin_tones['file'].str.lower().str.replace('.jpg', '').str.strip()
```

In [ ]:

```
In [24]: import spacy
import pandas as pd

# Load Spacy model
nlp = spacy.load("en_core_web_sm")

# Define corrected physical traits list
physical_traits = {
    'acceleration', 'active', 'aggression', 'aggressive', 'agile', 'athletic',
    'beast', 'belter', 'body', 'brawny', 'brisk', 'broad', 'build', 'bulk',
    'bursting', 'capacity', 'clawed', 'coordination', 'daring', 'determined',
    'durable', 'dynamic', 'endurance', 'enduring', 'energetic', 'energy',
    'ferocity', 'fierce', 'firm', 'fit', 'fleet', 'flexible', 'force', 'ford',
    'header', 'healthy', 'heavy', 'hefty', 'height', 'intense', 'intimidating',
    'light', 'limber', 'lively', 'massive', 'mobility', 'muscle', 'muscular',
    'pace', 'physical', 'physicality', 'pounce', 'pounced', 'power', 'powerful',
    'prowled', 'quick', 'rapid', 'resilient', 'resistant', 'robust', 'shrugged',
    'speedy', 'sprightly', 'sprint', 'sprinted', 'spry', 'stability', 'stamina',
    'strength', 'strike', 'striking', 'strong', 'stumble', 'stumbled', 'sturdy',
    'supple', 'swift', 'tall', 'tenacious', 'tiptoeing', 'touch', 'tough',
    'weight', 'well built', 'wide'
}

# Define corrected cognitive traits list
cognitive_traits = {
    'adaptive', 'alert', 'ambitious', 'analytical', 'anticipation', 'articulate',
    'authoritative', 'aware', 'brave', 'bright', 'brightfloated', 'brilliant',
    'clever', 'commanding', 'communicate', 'communicative', 'compose', 'concentrated',
    'conscious', 'control', 'coordinated', 'creative', 'curl', 'cute', 'decisive',
    'diplomatic', 'directive', 'discerning', 'disciplined', 'driven', 'dynamic',
    'excite', 'excitement', 'expressive', 'feet', 'finish', 'finished', 'flexible',
    'gamechanger', 'good ball', 'great ball', 'imaginative', 'industrious',
    'initiative', 'innovative', 'insightful', 'inspiring', 'intelligent',
    'keen', 'lead', 'leader', 'leadership', 'link up', 'linking', 'logical',
    'manage', 'mediatorstrategic', 'mesmerising', 'motivated', 'motivational',
    'observant', 'organized', 'original', 'passion', 'perceptive', 'perfect',
    'persuasive', 'plan', 'planning', 'pride', 'proactive', 'problem-solver',
    'quick-witted', 'reliable', 'resolute', 'resourceful', 'rolls', 'scan',
    'silky', 'skill', 'smart', 'strategic', 'strategy', 'tactful', 'tactical',
    'tenacious', 'thoughtful', 'tidy', 'timed', 'timely pass', 'under pressure',
    'versatile', 'vision', 'visionary', 'wasted opportunity', 'wise'
}

# Function to lemmatize and filter context words
def lemmatize_and_filter(context, physical_traits, cognitive_traits):
    doc = nlp(context)
    lemmatized_words = [token.lemma_ for token in doc]
    descriptive_words = [word for word in lemmatized_words if word in physical_traits]
    physical_words = ' '.join([word for word in descriptive_words if word in physical_traits])
    cognitive_words = ' '.join([word for word in descriptive_words if word in cognitive_traits])
    all_descriptive_words = ' '.join(descriptive_words)
    return physical_words, cognitive_words, all_descriptive_words, len(physical_words)
```

```
# Apply the function to create new columns
df_contexts[['physical words', 'cognitive words', 'descriptive words', 'physical traits', 'tone label']] = df_contexts.apply(lambda row: lemmatize_and_filter(row['Context'], physical_traits, cognitive_words, descriptive_words), axis=1)

# Check the DataFrame to see the new columns
print(df_contexts[['Player', 'tone label', 'Context', 'physical words', 'cognitive words', 'descriptive words', 'physical traits']])
```

	Player	tone	label	\
0	joão cancelo	CF		
1	joão cancelo	CF		
2	roberto	CF		
3	harry kane	CE		
4	tammy abraham	CD		
5	alex mccarthy	CF		
6	alex mccarthy	CF		
7	alex mccarthy	CF		
8	bernardo	CD		
9	bernardo	CD		
10	bernardo	CD		
11	bernardo silva	CE		

## Context \

0 oh great skill from bernardeschi for dybala spilled big chance joão can celo with an equalizer for juventus back from injury to save  
 1 and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a  
 2 by the offside flag a few minutes ago this ti me roberto inglese is off the mark in the match giving parma  
 3 theres no chance for the keeper theyre stretched defensively here har ry kane just makes that point only needed a touch floated in  
 4 by šofranac he might get a tap in this time tammy abra ham opens his england account it is seven with seven minutes  
 5 target the pace on the ball means he cant react alex mc carthy it goes through his legs hes tiptoeing his way around  
 6 legs hes tiptoeing his way around and its bernardo sil va alex mccarthy sees it up and over well hes not been asked  
 7 well hes not been asked to do an awful lot alex mccarth y bernardo silva gundogan de bruyne inches away from being a  
 8 instinct to get southampton the most unlikely of leads her e bernardo silva aguero is in and he is furious with himself  
 9 through his legs hes tiptoeing his way around and i ts bernardo silva alex mccarthy sees it up and over well hes  
 10 not been asked to do an awful lot alex mccarthy bernardo s ilva gundogan de bruyne inches away from being a gamechanger  
 11 instinct to get southampton the most unlikely of leads here bernar do silva aguero is in and he is furious with himself because

	physical words	cognitive words	descriptive words	physical count	\
0			skill	skill	0
1					0
2					0
3	touch			touch	1
4					0
5	pace			pace	1
6					0
7					0
8			lead	lead	0
9					0
10		gamechanger		gamechanger	0
11			lead	lead	0

## cognitive count

0	1
1	0
2	0
3	0
4	0
5	0
6	0
7	0

8	1
9	0
10	1
11	1

```
In [25]: df_contexts
```

Out[25]:

	Player	Context	tone label	physical words	cognitive words	descriptive words	physical count	cognitive count
0	joão cancelo	oh great skill from bernardeschi for dybala spilled big chance joão cancelo with an equalizer for juventus back from injury to save	CF		skill	skill	0	1
1	joão cancelo	and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a	CF				0	0
2	roberto	by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma	CF				0	0
3	harry kane	theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in	CE	touch		touch	1	0
4	tammy abraham	by šofranac he might get a tap in this time tammy abraham opens his england account it is seven with seven minutes	CD				0	0
...	...	...	...	...	...	...	...	...
2071	david brooks	really good cross and its good contact as	CG	touch		touch	1	0

Player	Context	tone label	physical words	cognitive words	descriptive words	physical count	cognitive count
	well from david brooks brooks now thats a lovely touch to take him beyond						
2072	harry wilson	CG	touch	touch	1	0	
2073	wayne hennessey	CE			0	0	
2074	wayne hennessey	CE	strong aggressive	strong aggressive	2	0	
2075	wayne hennessey	CE	tidy	tidy	0	1	

2076 rows x 8 columns

In [26]: `import pandas as pd`

```
# Group by skin tone and aggregate counts for each trait type
grouped_counts = df_contexts.groupby('tone label').agg(
    physical_count=('physical count', 'sum'),
    cognitive_count=('cognitive count', 'sum'))
```

```
    ).reset_index()

    # Display the results
    print(grouped_counts)
```

	tone label	physical_count	cognitive_count
0	CC	17	12
1	CD	69	58
2	CE	104	89
3	CF	147	109
4	CG	64	44
5	CI	12	10

In [27]: `# Count the number of unique players per skin tone category`  
`players_per_tone = df_contexts.groupby('tone label')['Player'].nunique().reset_index()`

```
    # Rename the columns for clarity
    players_per_tone.columns = ['tone label', 'player count']

    # Display the results
    print(players_per_tone)
```

	tone label	player count
0	CC	12
1	CD	52
2	CE	79
3	CF	98
4	CG	53
5	CI	22

In [28]: `# Sum up the counts of players in the darker skin tone categories (CC, CD, CE)`  
`darker_skin_tone_count = players_per_tone[players_per_tone['tone label'].isin(['CC', 'CD', 'CE'])].sum().reset_index()`

```
    # Sum up the counts of players in the lighter skin tone categories (CF, CG, CI)
    lighter_skin_tone_count = players_per_tone[players_per_tone['tone label'].isin(['CF', 'CG', 'CI'])].sum().reset_index()
```

```
    # Print the results
    print("Darker Skin Tone (DST):", darker_skin_tone_count)
    print("Lighter Skin Tone (LST):", lighter_skin_tone_count)
```

Darker Skin Tone (DST): 143  
Lighter Skin Tone (LST): 173

In [29]: `# Aggregate counts of descriptive words and number of players by skin tone`  
`combined_stats = df_contexts.groupby('tone label').agg(`

```
    physical_count=('physical count', 'sum'),
    cognitive_count=('cognitive count', 'sum'),
    player_count=('Player', 'nunique')
).reset_index()
```

```
    # Rename columns if needed (for clarity)
    combined_stats.columns = ['Tone Label', 'Physical Count', 'Cognitive Count', 'Player Count']
```

```
    # Display the combined DataFrame
    print(combined_stats)
```

	Tone Label	Physical Count	Cognitive Count	Player Count
0	CC	17	12	12
1	CD	69	58	52
2	CE	104	89	79
3	CF	147	109	98
4	CG	64	44	53
5	CI	12	10	22

In [ ]:

```
In [30]: # Specify the file path where you want to save the Excel file
file_path = "/Users/alexdoyle/Desktop/df_contexts(updated).xlsx"

# Export the DataFrame to Excel
df_contexts.to_excel(file_path, index=False)

print("DataFrame exported to Excel successfully.")

DataFrame exported to Excel successfully.
```

In [ ]:

In [ ]:

```
In [31]: import pandas as pd

# Load the Excel file
file_path = '/Users/alexdoyle/Desktop/df_contexts(updated).xlsx'
df = pd.read_excel(file_path)

# Drop rows where 'descriptive words' column is empty
df_cleaned = df[df['descriptive words'].notna() & df['descriptive words'].str.strip().ne('')]

# Specify the file path where you want to save the cleaned Excel file
output_file_path = '/Users/alexdoyle/Desktop/df_contexts_cleaned.xlsx'

# Export the cleaned DataFrame to Excel
df_cleaned.to_excel(output_file_path, index=False)

print("Cleaned DataFrame exported to Excel successfully.")
```

Cleaned DataFrame exported to Excel successfully.

In [ ]:

In [ ]:

/Users/alexdoyle/Desktop/excel\_words\_identified/df\_contexts(updated).xlsx

```
In [33]: import pandas as pd

# Load the CSV file
file_path = '/Users/alexdoyle/Desktop/df_contexts_cleaned.csv'
df = pd.read_csv(file_path)

# Check the columns to ensure correct names
print("Column names in the DataFrame:")
print(df.columns)
```

Column names in the DataFrame:

```
Index(['Player', 'Context', 'tone label', 'descriptive words',
       'physical count', 'cognitive count', 'physical words',
       'cognitive words', 'Physical - Positive Sentiment',
       'Physical - Negative Sentiment', 'Cognitive - Positive Sentiment',
       'Cognitive - Negative Sentiment'],
      dtype='object')
```

In [34]: import pandas as pd

```
# Load the Excel file
file_path = '/Users/alexdoyle/Desktop/df_contexts_cleaned.csv'
df = pd.read_csv(file_path)

# Group by 'tone label' and sum the counts for each sentiment category
```

```

tone_sentiment_counts = df.groupby('tone label').agg({
    'Physical - Positive Sentiment': 'sum',
    'Physical - Negative Sentiment': 'sum',
    'Cognitive - Positive Sentiment': 'sum',
    'Cognitive - Negative Sentiment': 'sum'
})

print("Tally of positive and negative words for each skin tone:")
print(tone_sentiment_counts)

```

Tally of positive and negative words for each skin tone:

tone label	Physical - Positive Sentiment	Physical - Negative Sentiment
CC	14.0	1.0
CD	60.0	7.0
CE	86.0	14.0
CF	93.0	15.0
CG	46.0	7.0
CI	5.0	1.0

tone label	Cognitive - Positive Sentiment	Cognitive - Negative Sentiment
CC	7.0	0.0
CD	22.0	3.0
CE	34.0	5.0
CF	66.0	5.0
CG	29.0	1.0
CI	5.0	1.0

In [ ]:

```
In [46]: # Load the main data from CSV
main_df = pd.read_csv('/Users/alexdoyle/Desktop/df_contexts_cleaned.csv')
```

```
In [47]: main_df
```

Out[47]:

	Player	Context	tone label	descriptive words	physical count	cognitive count	physical words	cognitive words	
0	harry kane	theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in	CE	touch	1.0	0.0	touch	NaN	\$
1	alex mccarthy	target the pace on the ball means he cant react alex mccarthy it goes through his legs hes tiptoeing his way around	CF	tiptoeing	1.0	0.0	tiptoeing	NaN	
2	sergio aguero	expect aguero to do much better walker in towards aguero sergio aguero doing what comes so naturally to him strong ankle gets	CE	strong	1.0	0.0	strong	NaN	
3	danny ings	convincing header boufal up against schneiderlin its in from ings danny ings with the equalizer for southampton at long last it game	CF	header	1.0	0.0	header	NaN	
4	richarlison	wardprowse very few better at striking the dead ball sidibé richarlison brilliantly finished richarlison his first goal in eight	CF	finish	0.0	1.0	NaN	finish	

	Player	Context	tone label	descriptive words	physical count	cognitive count	physical words	cognitive words	
premier league									
398	wayne hennessey	does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda	CE	big strong aggressive	3.0	0.0	big strong aggressive	NaN	
399	wayne hennessey	hamšík tidy turn but saved from robert maks effort by wayne hennessey terrific football from the slovakians once again manoeuvre the ball	CE	terrific football	0.0	1.0	NaN	terrific football	
400	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
401	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
402	NaN	NaN	NaN	NaN	344.0	190.0	NaN	NaN	

403 rows × 12 columns

In [ ]:

In [ ]:

In [118...]: # actual skin tone count after manually sorting through the data

```
# define which labels are LST and DST
dark_tones = ['CC', 'CD', 'CE']
light_tones = ['CF', 'CG', 'CI']

# This code counts the frequency of each tone label
tone_counts = main_df['tone label'].value_counts()

# Dictionary to store the count
individual_counts = {tone: tone_counts.get(tone, 0) for tone in dark_tones}
tone_frequencies = {'DST': 0, 'LST': 0}

# Count the frequency for darker skin tones (DST)
for tone in dark_tones:
    tone_frequencies['DST'] += tone_counts.get(tone, 0)

# Count the frequency for lighter skin tones (LST)
```

```

for tone in light_tones:
    tone_frequencies['LST'] += tone_counts.get(tone, 0)

print("Individual Counts:", individual_counts)
print("Aggregated Counts:", tone_frequencies)

Individual Counts: {'CC': 14, 'CD': 69, 'CE': 110, 'CF': 137, 'CG': 60, 'CI': 10}
Aggregated Counts: {'DST': 193, 'LST': 207}

```

In [ ]:

In [ ]:

In [ ]:

```

In [36]: import cv2
import dlib
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import requests
from io import BytesIO

```

```

In [37]: # load in the data

df_player_images_raw = pd.read_csv('/Users/alexdoyle/Desktop/player_names.csv')

```

```

In [38]: # this code will ensure I have loaded in the pre-trained model for face landmarks

predictor_path = "/Users/alexdoyle/shape_predictor_68_face_landmarks.dat"
detector = dlib.get_frontal_face_detector()
predictor = dlib.shape_predictor(predictor_path)

```

```

In [39]: # this is a function that will get facial landmarks from an image

def get_landmarks(image):
    detections = detector(image, 1)
    if len(detections) > 0:
        return np.matrix([[p.x, p.y] for p in predictor(image, detections[0]).parts()])
    return None

```

```

In [40]: # function to download an image from a URL

def download_image(url):
    response = requests.get(url)
    image = np.asarray(bytearray(response.content), dtype="uint8")
    image = cv2.imdecode(image, cv2.IMREAD_COLOR)
    return image

```

```

In [41]: # function to calculate the average face from a list of image URLs

def get_average_face(image_urls):
    landmarks_list = []
    images = []

    for image_url in image_urls:
        image = download_image(image_url)
        if image is None:
            continue

```

```

        images.append(image)
        landmarks = get_landmarks(image)
        if landmarks is not None:
            landmarks_list.append(landmarks)

    if not landmarks_list:
        return None

    # Calculate the average landmarks
    avg_landmarks = np.mean(landmarks_list, axis=0)

    # Warp images to the average landmarks
    warped_images = []
    for image, landmarks in zip(images, landmarks_list):
        h, status = cv2.findHomography(landmarks, avg_landmarks)
        warped_image = cv2.warpPerspective(image, h, (image.shape[1], image.shape[0]))
        warped_images.append(warped_image)

    # Blend the warped images
    avg_face = np.mean(warped_images, axis=0).astype(np.uint8)

    return avg_face

```

In [48]:

```

# function to get image paths (URLs) for players described by a given word

def get_image_paths_for_word(word, df_player_images_raw, main_df):
    # Get players described by the given word
    players = main_df[main_df['descriptive words'].str.contains(word, case=False)]
    print("Players described by the word:", players.tolist())

    # Normalize player names in df_player_images_raw
    df_player_images_raw['Player Names'] = df_player_images_raw['Player Names'].str.lower()

    # Get image paths for these players
    image_paths = df_player_images_raw[df_player_images_raw['Player Names'].str.contains(word, case=False)]
    print("Image paths for these players:", image_paths)

    return image_paths

```

In [49]:

```

# Choose a descriptive word
word = 'power'

# Get the image paths of players described by the word
image_paths = get_image_paths_for_word(word, df_player_images_raw, main_df)

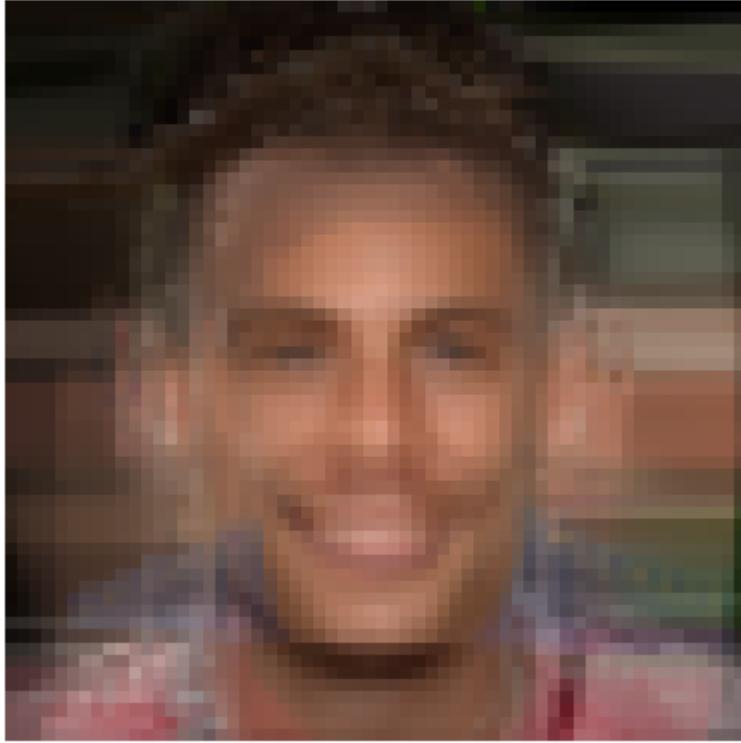
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

    # Save the average face to a file
    cv2.imwrite(f"average_face_{word}.jpg", average_face)
else:
    print(f"No faces found for the word: {word}")

```

Players described by the word: ['callum robinson', 'jordan henderson', 'tammy abraham', 'anthony martial', 'ken sema', 'james milner', 'lucas torreira', 'theo walcott']  
 Image paths for these players: ['https://cdn.sofifa.net/players/211/300/18\_60.png', 'https://cdn.sofifa.net/players/206/518/18\_60.png', 'https://cdn.sofifa.net/players/138/412/18\_60.png', 'https://cdn.sofifa.net/players/183/711/18\_60.png', 'https://cdn.sofifa.net/players/213/750/18\_60.png', 'https://cdn.sofifa.net/players/223/959/18\_60.png', 'https://cdn.sofifa.net/players/231/352/18\_60.png', 'https://cdn.sofifa.net/players/164/859/18\_60.png']



```
In [50]: # Choose a descriptive word
word = 'skill'

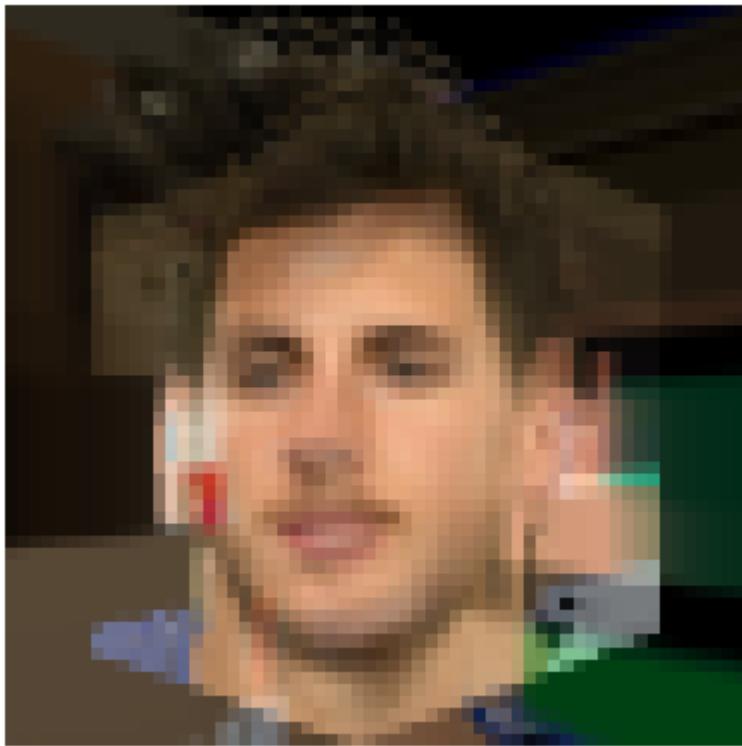
# Get the image paths of players described by the word
image_paths = get_image_paths_for_word(word, df_player_images_raw, main_df)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}.jpg", average_face)
else:
    print(f"No faces found for the word: {word}")
```

Players described by the word: ['eden hazard', 'mason mount']  
 Image paths for these players: ['https://cdn.sofifa.net/players/183/277/18\_60.png', 'https://cdn.sofifa.net/players/233/064/18\_60.png']



```
In [51]: # Choose a descriptive word
word = 'pace'

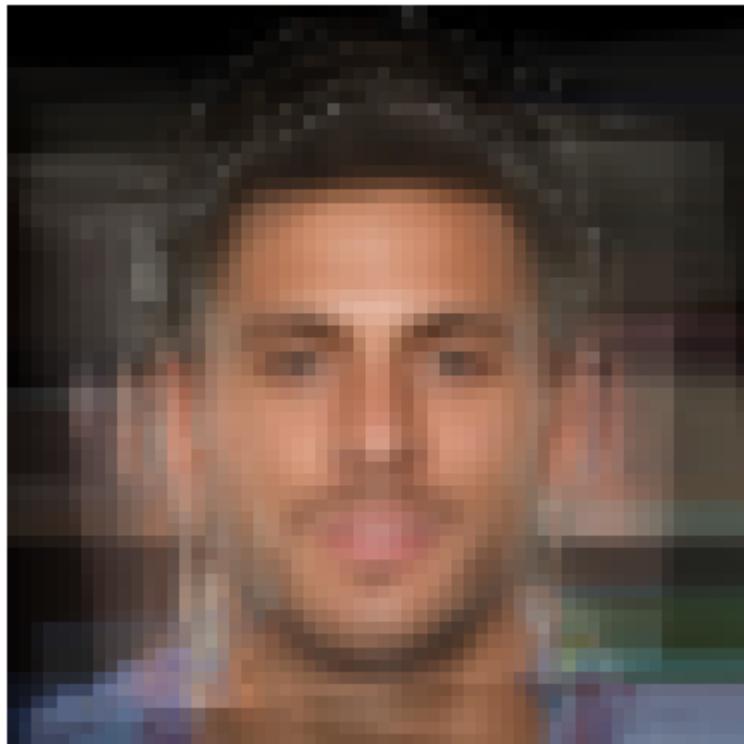
# Get the image paths of players described by the word
image_paths = get_image_paths_for_word(word, df_player_images_raw, main_df)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

    # Save the average face to a file
    cv2.imwrite(f"average_face_{word}.jpg", average_face)
else:
    print(f"No faces found for the word: {word}")
```

Players described by the word: ['eden hazard', 'lewis dunk', 'willian', 'ky le walker', 'jack cork', 'theo walcott', 'willian', 'nick pope']  
Image paths for these players: ['https://cdn.sofifa.net/players/183/277/18\_60.png', 'https://cdn.sofifa.net/players/169/706/18\_60.png', 'https://cdn.sofifa.net/players/188/377/18\_60.png', 'https://cdn.sofifa.net/players/199/915/18\_60.png', 'https://cdn.sofifa.net/players/203/841/18\_60.png', 'http s://cdn.sofifa.net/players/164/859/18\_60.png', 'https://cdn.sofifa.net/players/180/403/18\_60.png']



```
In [52]: # Choose a descriptive word
word = 'pounce'

# Get the image paths of players described by the word
image_paths = get_image_paths_for_word(word, df_player_images_raw, main_df)

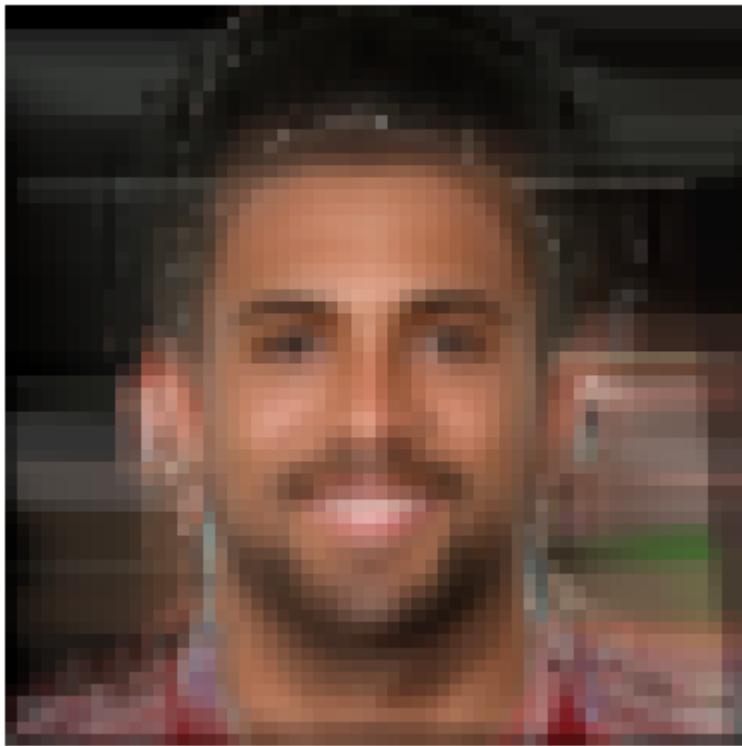
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}.jpg", average_face)
else:
    print(f"No faces found for the word: {word}")
```

Players described by the word: ['romelu lukaku', 'danny ings', 'olivier giroud', 'joão cancelo', 'callum wilson']

Image paths for these players: ['https://cdn.sofifa.net/players/196/978/18\_60.png', 'https://cdn.sofifa.net/players/195/859/18\_60.png', 'https://cdn.sofifa.net/players/210/514/18\_60.png', 'https://cdn.sofifa.net/players/178/509/18\_60.png', 'https://cdn.sofifa.net/players/192/505/18\_60.png']



```
In [53]: # Choose a descriptive word
word = 'intelligent'

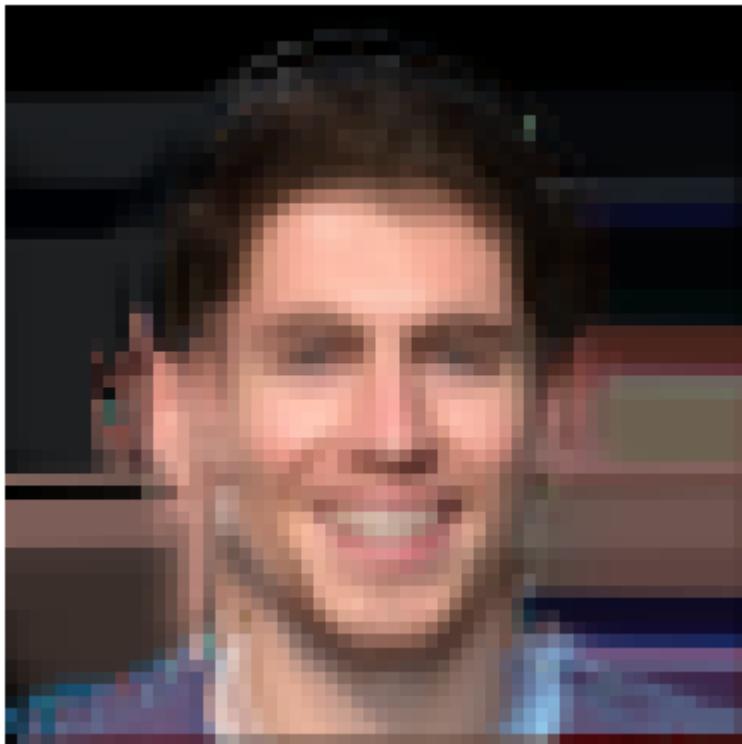
# Get the image paths of players described by the word
image_paths = get_image_paths_for_word(word, df_player_images_raw, main_df)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

    # Save the average face to a file
    cv2.imwrite(f"average_face_{word}.jpg", average_face)
else:
    print(f"No faces found for the word: {word}")
```

Players described by the word: ['jorginho', 'harry wilson']  
Image paths for these players: ['https://cdn.sofifa.net/players/220/710/18\_60.png', 'https://cdn.sofifa.net/players/205/498/18\_60.png']



```
In [54]: # Choose a descriptive word
word = 'strength'

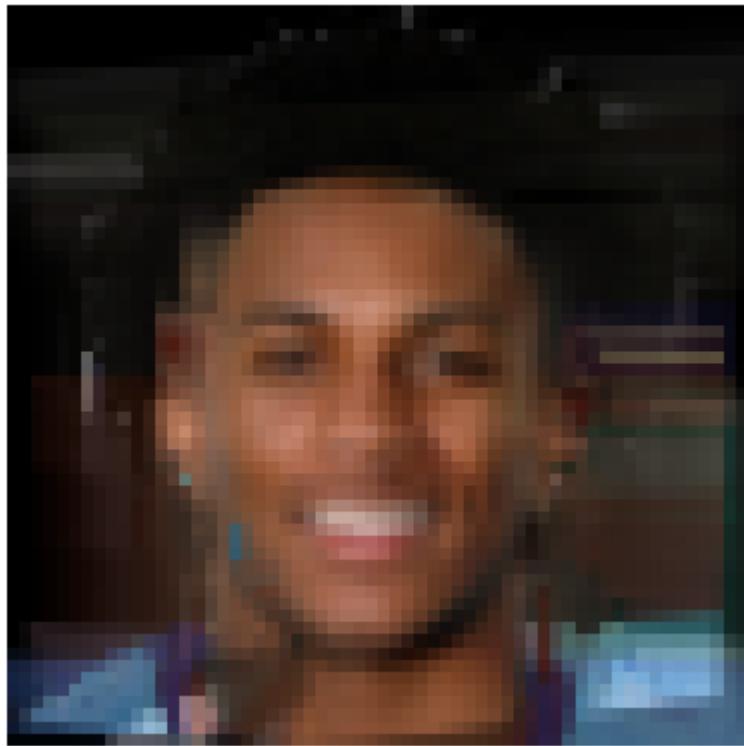
# Get the image paths of players described by the word
image_paths = get_image_paths_for_word(word, df_player_images_raw, main_df)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

    # Save the average face to a file
    cv2.imwrite(f"average_face_{word}.jpg", average_face)
else:
    print(f"No faces found for the word: {word}")

Players described by the word: ['willian', 'michail antonio', 'richarlison']
Image paths for these players: ['https://cdn.sofifa.net/players/190/717/18_60.png', 'https://cdn.sofifa.net/players/231/943/18_60.png', 'https://cdn.sofifa.net/players/180/403/18_60.png']
```



In [ ]:

```
In [55]: def get_landmarks(image):
    detections = detector(image, 1)
    if len(detections) > 0:
        return np.matrix([[p.x, p.y] for p in predictor(image, detections[0])])
    return None
```

```
In [56]: def download_image(url):
    response = requests.get(url)
    image = np.asarray(bytearray(response.content), dtype="uint8")
    image = cv2.imdecode(image, cv2.IMREAD_COLOR)
    return image
```

```
In [57]: def get_average_face(image_urls):
    landmarks_list = []
    images = []

    for image_url in image_urls:
        image = download_image(image_url)
        if image is None:
            continue
        images.append(image)
        landmarks = get_landmarks(image)
        if landmarks is not None:
            landmarks_list.append(landmarks)
```

```

if not landmarks_list:
    return None

# Calculate the average landmarks
avg_landmarks = np.mean(landmarks_list, axis=0)

# Warp images to the average landmarks
warped_images = []
for image, landmarks in zip(images, landmarks_list):
    h, status = cv2.findHomography(landmarks, avg_landmarks)
    warped_image = cv2.warpPerspective(image, h, (image.shape[1], image.shape[0]))
    warped_images.append(warped_image)

# Blend the warped images
avg_face = np.mean(warped_images, axis=0).astype(np.uint8)

return avg_face

```

In [58]:

```

def get_image_paths_for_words(words, df_player_images_raw, main_df):
    # Get players described by any of the given words
    players = main_df[main_df['descriptive words'].str.contains('|'.join(words))]
    print("Players described by the words:", players.tolist())

    # Normalize player names in df_player_images_raw
    df_player_images_raw['Player Names'] = df_player_images_raw['Player Names'].str.lower()

    # Get image paths for these players
    image_paths = df_player_images_raw[df_player_images_raw['Player Names'].isin(players['Player Names'])]
    print("Image paths for these players:", image_paths)

    return image_paths

```

In [59]:

```

# Choose descriptive words
words = ['strength', 'speed', 'pace', 'pounce', 'predator']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_df)

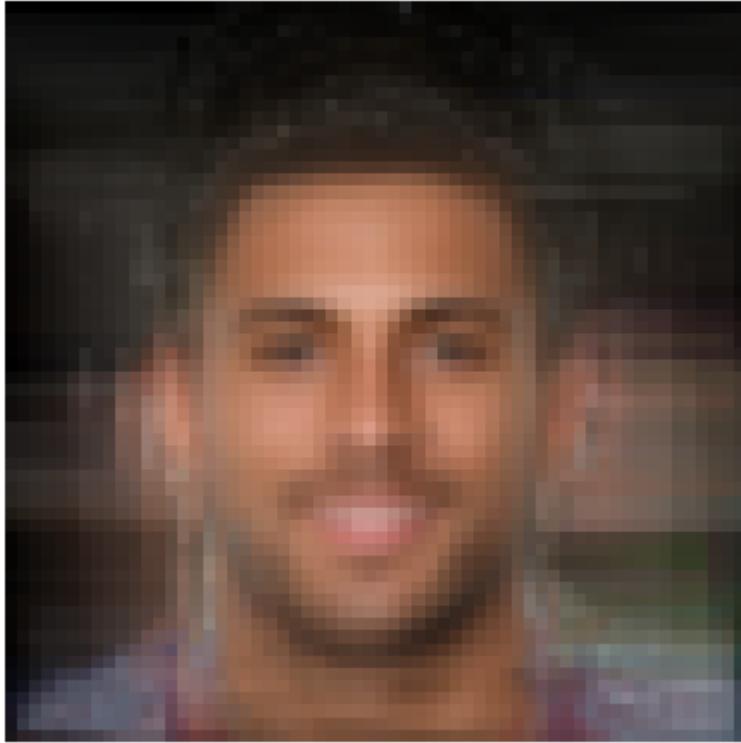
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

    # Save the average face to a file
    cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")

```

Players described by the words: ['eden hazard', 'romelu lukaku', 'danny ing s', 'olivier giroud', 'lewis dunk', 'willian', 'willian', 'kyle walker', 'm ichail antonio', 'jack cork', 'richarlison', 'joão cancelo', 'theo walcot t', 'willian', 'nick pope', 'callum wilson']  
 Image paths for these players: ['https://cdn.sofifa.net/players/196/978/18\_60.png', 'https://cdn.sofifa.net/players/195/859/18\_60.png', 'https://cdn.sofifa.net/players/183/277/18\_60.png', 'https://cdn.sofifa.net/players/169/706/18\_60.png', 'https://cdn.sofifa.net/players/210/514/18\_60.png', 'http s://cdn.sofifa.net/players/188/377/18\_60.png', 'https://cdn.sofifa.net/players/199/915/18\_60.png', 'https://cdn.sofifa.net/players/190/717/18\_60.png', 'https://cdn.sofifa.net/players/203/841/18\_60.png', 'https://cdn.sofifa.net/players/178/509/18\_60.png', 'https://cdn.sofifa.net/players/231/943/18\_60.png', 'https://cdn.sofifa.net/players/192/505/18\_60.png', 'https://cdn.sofifa.net/players/164/859/18\_60.png', 'https://cdn.sofifa.net/players/180/403/18\_60.png']



Despite there being less darker skinned players within my data set, they are overrepresented within this visual. With there being 7 white players out of 16 total that have been described by at least one of these words: - strength - speed - pace - pounce - predator

```
In [60]: # Choose descriptive words
words = ['strength', 'predator', 'stumble']

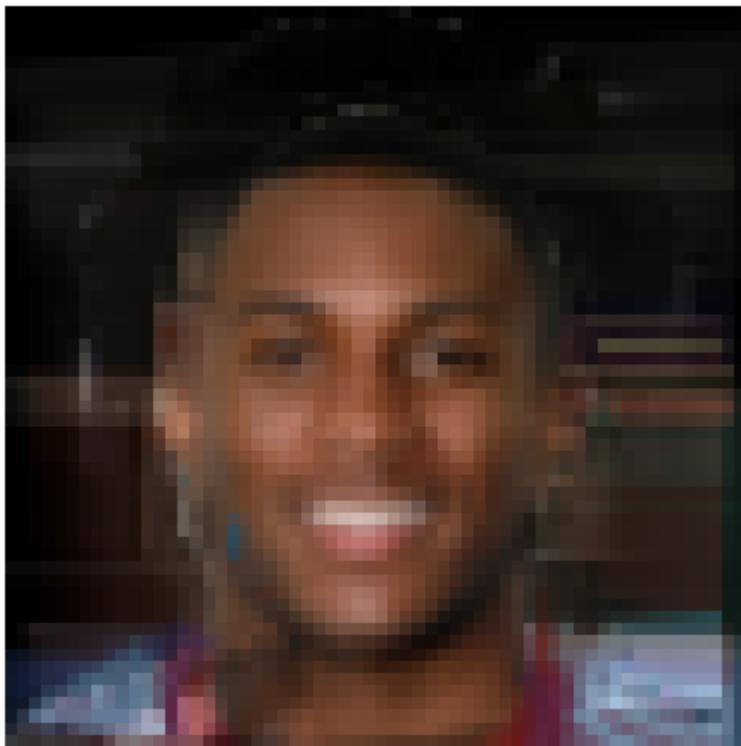
# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['romelu lukaku', 'michael antonio', 'michael antonio', 'willian', 'michael antonio', 'richarlison']  
Image paths for these players: ['https://cdn.sofifa.net/players/190/717/18\_60.png', 'https://cdn.sofifa.net/players/231/943/18\_60.png', 'https://cdn.sofifa.net/players/192/505/18\_60.png', 'https://cdn.sofifa.net/players/180/403/18\_60.png']



In [ ]:

```
In [61]: # Choose descriptive words
words = ['adaptive', 'alert', 'ambitious', 'analytical', 'anticipation', 'authoritative', 'aware', 'brave', 'bright', 'brightfloated', 'clever', 'commanding', 'communicate', 'communicative', 'compose', 'conscious', 'control', 'coordinated', 'creative', 'curl', 'cute', 'diplomatic', 'directive', 'discerning', 'disciplined', 'driven', 'excite', 'excitement', 'expressive', 'feet', 'finished', 'flexible', 'gamechanger', 'good ball', 'great ball', 'imaginative', 'industrial', 'initiative', 'innovative', 'insightful', 'inspiring', 'intelligent', 'judicious', 'keen', 'lead', 'leader', 'leadership', 'link up', 'lovely shape', 'manage', 'mediatorstrategic', 'mesmerising', 'motivate', 'neatly', 'negotiator', 'observant', 'organized', 'original', 'pass', 'perfectly', 'perspicacious', 'persuasive', 'plan', 'planning', 'prudent', 'pullback', 'purposeful', 'quick-witted', 'reliable', 'scan', 'sensible', 'sharp', 'shrewd', 'silky', 'skill', 'smart', 'tactical', 'talent', 'technical', 'technique', 'tenacious', 'thought', 'timely pass', 'tiptoeing', 'touch', 'under pressure', 'unselfish', 'wasted opportunity', 'wise']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

# Generate the average face
average_face = get_average_face(image_paths)

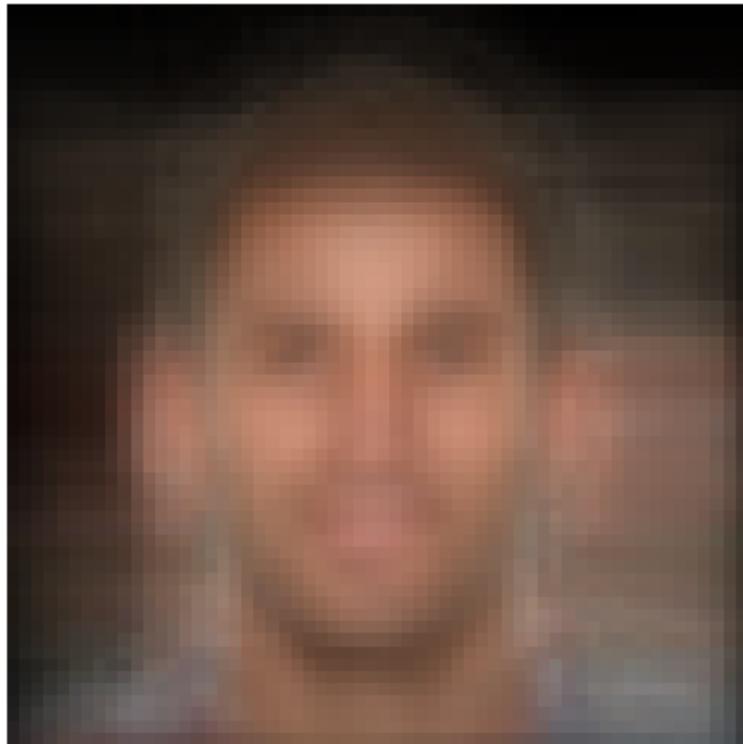
# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()
```

```
# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['harry kane', 'alex mccarthy', 'chris smalling', 'eden hazard', 'teemu pukki', 'harvey barnes', 'jamie vardy', 'jamie vardy', 'jack grealish', 'kortney hause', 'harvey barnes', 'james maddison', 'roberto', 'jonjo shelvey', 'patrick cutrone', 'marcus rashford', 'jack grealish', 'jesse lingard', 'jesse lingard', 'rui patrício', 'darren Randolph', 'willian', 'lewis dunk', 'brandon williams', 'jorginho', 'chris wood', 'ricardo pereira', 'kenny mclean', 'antonio barreca', 'willian', 'willian', 'callum wilson', 'jorginho', 'jorginho', 'fred', 'marcus rashford', 'sadio mané', 'alvaro morata', 'richarlison', 'jorginho', 'conor hourihane', 'felipe anderson', 'valentino lazaro', 'danny rose', 'joelinton', 'jordan Henderson', 'jonny', 'mario vrančić', 'roberto', 'ederson', 'fernandinho', 'kevin de bruyne', 'jorginho', 'willian', 'kadeem harris', 'steven davis', 'abdoulaye doucouré', 'ben foster', 'ben foster', 'roberto', 'david silva', 'teemu pukki', 'aaron ramsdale', 'willian', 'kieran dowell', 'pedro neto', 'jordan pickford', 'kevin de bruyne', 'fraser forster', 'mohamed elyounoussi', 'youri tielemans', 'james maddison', 'patrick cutrone', 'youri tielemans', 'dele alli', 'kevin de bruyne', 'david silva', 'danny rose', 'gabriel jesus', 'roberto firmino', 'olivier giroud', 'danny ings', 'danny ings', 'ben davies', 'eden hazard', 'ederson', 'tom heaton', 'tom heaton', 'joe gomez', 'willian', 'tammy abraham', 'lewis dunk', 'neil taylor', 'harry kane', 'moussa sissoko', 'ciaran clark', 'ben foster', 'ashley westwood', 'chris wood', 'jack cork', 'ben foster', 'ben foster', 'dele alli', 'kevin de bruyne', 'aaron mooy', 'jorginho', 'callum wilson', 'callum wilson', 'callum wilson', 'alex iwobi', 'phil foden', 'alex iwobi', 'theo walcott', 'chris basham', 'dean henderson', 'danny ings', 'josh murphy', 'chris basham', 'harry wilson', 'anthony knockaert', 'mark noble', 'michy batshuayi', 'fred', 'harry kane', 'charlie austin', 'danilo', 'david mcgoldrick', 'harvey barnes', 'jamie vardy', 'harry kane', 'deandre yedlin', 'ashley young', 'christian benteke', 'christian benteke', 'mason mount', 'raheem sterling', 'callum wilson', 'christian pulisic', 'gabriel jesus', 'danilo', 'david brooks', 'harry wilson']

Image paths for these players: ['https://cdn.sofifa.net/players/194/958/18\_60.png', 'https://cdn.sofifa.net/players/233/934/18\_60.png', 'https://cdn.sofifa.net/players/208/135/18\_60.png', 'https://cdn.sofifa.net/players/213/655/18\_60.png', 'https://cdn.sofifa.net/players/189/324/18\_60.png', 'https://cdn.sofifa.net/players/201/153/18\_60.png', 'https://cdn.sofifa.net/players/203/910/18\_60.png', 'https://cdn.sofifa.net/players/220/493/18\_60.png', 'https://cdn.sofifa.net/players/189/280/18\_60.png', 'https://cdn.sofifa.net/players/152/908/18\_60.png', 'https://cdn.sofifa.net/players/205/923/18\_60.png', 'https://cdn.sofifa.net/players/163/155/18\_60.png', 'https://cdn.sofifa.net/players/138/786/18\_60.png', 'https://cdn.sofifa.net/players/196/978/18\_60.png', 'https://cdn.sofifa.net/players/195/671/18\_60.png', 'https://cdn.sofifa.net/players/184/274/18\_60.png', 'https://cdn.sofifa.net/players/189/881/18\_60.png', 'https://cdn.sofifa.net/players/192/123/18\_60.png', 'https://cdn.sofifa.net/players/184/111/18\_60.png', 'https://cdn.sofifa.net/players/227/796/18\_60.png', 'https://cdn.sofifa.net/players/183/129/18\_60.png', 'https://cdn.sofifa.net/players/193/849/18\_60.png', 'https://cdn.sofifa.net/players/199/304/18\_60.png', 'https://cdn.sofifa.net/players/195/859/18\_60.png', 'https://cdn.sofifa.net/players/169/595/18\_60.png', 'https://cdn.sofifa.net/players/169/697/18\_60.png', 'https://cdn.sofifa.net/players/220/196/18\_60.png', 'https://cdn.sofifa.net/players/180/818/18\_60.png', 'https://cdn.sofifa.net/players/168/542/18\_60.png', 'https://cdn.sofifa.net/players/212/722/18\_60.png', 'https://cdn.sofifa.net/players/233/306/18\_60.png', 'https://cdn.sofifa.net/players/211/117/18\_60.png', 'https://cdn.sofifa.net/players/183/277/18\_60.png', 'https://cdn.sofifa.net/players/210/257/18\_60.png', 'https://cdn.sofifa.net/players/201/995/18\_60.png', 'https://cdn.sofifa.net/players/135/507/18\_60.png', 'https://cdn.sofifa.net/players/209/297/18\_60.png', 'https://cdn.sofifa.net/players/230/666/18\_60.png', 'https://cdn.sofifa.net/players/202/126/18\_60.png', 'https://cdn.sofifa.net/players/220/710/18\_60.png', 'https://cdn.sofifa.net/players/234/742/18\_60.png', 'https://cdn.sofifa.net/players/169/706/18\_60.png', 'https://cdn.sofifa.net/players/206/517/18\_60.png', 'https://cdn.sofifa.net/players/220/697/18\_60.png', 'https://cdn.sofifa.net/players/208/830/18\_60.png', 'https://cdn.sofifa.net/players/207/494/18\_60.png']

```
8_60.png', 'https://cdn.sofifa.net/players/225/100/18_60.png', 'https://cdn.sofifa.net/players/223/334/18_60.png', 'https://cdn.sofifa.net/players/189/165/18_60.png', 'https://cdn.sofifa.net/players/210/455/18_60.png', 'https://cdn.sofifa.net/players/204/935/18_60.png', 'https://cdn.sofifa.net/players/205/498/18_60.png', 'https://cdn.sofifa.net/players/206/085/18_60.png', 'https://cdn.sofifa.net/players/197/170/18_60.png', 'https://cdn.sofifa.net/players/201/368/18_60.png', 'https://cdn.sofifa.net/players/226/401/18_60.png', 'https://cdn.sofifa.net/players/210/635/18_60.png', 'https://cdn.sofifa.net/players/199/915/18_60.png', 'https://cdn.sofifa.net/players/231/677/18_60.png', 'https://cdn.sofifa.net/players/177/896/18_60.png', 'https://cdn.sofifa.net/players/233/064/18_60.png', 'https://cdn.sofifa.net/players/204/529/18_60.png', 'https://cdn.sofifa.net/players/203/362/18_60.png', 'https://cdn.sofifa.net/players/183/394/18_60.png', 'https://cdn.sofifa.net/players/173/735/18_60.png', 'https://cdn.sofifa.net/players/178/509/18_60.png', 'https://cdn.sofifa.net/players/237/715/18_60.png', 'https://cdn.sofifa.net/players/238/616/18_60.png', 'https://cdn.sofifa.net/players/237/692/18_60.png', 'https://cdn.sofifa.net/players/202/652/18_60.png', 'https://cdn.sofifa.net/players/210/243/18_60.png', 'https://cdn.sofifa.net/players/231/943/18_60.png', 'https://cdn.sofifa.net/players/183/666/18_60.png', 'https://cdn.sofifa.net/players/201/942/18_60.png', 'https://cdn.sofifa.net/players/178/005/18_60.png', 'https://cdn.sofifa.net/players/208/722/18_60.png', 'https://cdn.sofifa.net/players/151/508/18_60.png', 'https://cdn.sofifa.net/players/231/352/18_60.png', 'https://cdn.sofifa.net/players/190/362/18_60.png', 'https://cdn.sofifa.net/players/164/859/18_60.png', 'https://cdn.sofifa.net/players/163/264/18_60.png', 'https://cdn.sofifa.net/players/211/147/18_60.png', 'https://cdn.sofifa.net/players/180/403/18_60.png', 'https://cdn.sofifa.net/players/216/393/18_60.png']
```



In [274...]

```
# Choose descriptive words
words = ['acceleration', 'active', 'aggression', 'aggressive', 'agile', 'athletic', 'beast', 'belter', 'body', 'brawny', 'brisk', 'broad', 'build', 'bulk', 'bursting', 'capacity', 'clawed', 'cleverly', 'control', 'coordination', 'durable', 'dynamic', 'endurance', 'enduring', 'energetic', 'energy', 'energy', 'ferocity', 'fierce', 'finish', 'firm', 'fit', 'fleet', 'flexible', 'focused', 'header', 'healthy', 'heavy', 'hefty', 'height', 'intense', 'intimidating', 'light', 'limber', 'lively', 'massive', 'mobility', 'muscle', 'muscular', 'pace', 'physical', 'physicality', 'pounce', 'pounced', 'power', 'powerful', 'prowled', 'quick', 'rapid', 'resilient', 'resistant', 'robust', 'shrugged']
```

```
'speedy', 'sprightly', 'sprint', 'sprinted', 'spry', 'stability', 'stam·
'strength', 'strike', 'striking', 'strong', 'stumble', 'stumbled', 'stu·
'supple', 'swift', 'tall', 'tenacious', 'tiptoeing', 'touch', 'tough', ·
'weight', 'well built', 'wide'] ·

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d·

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

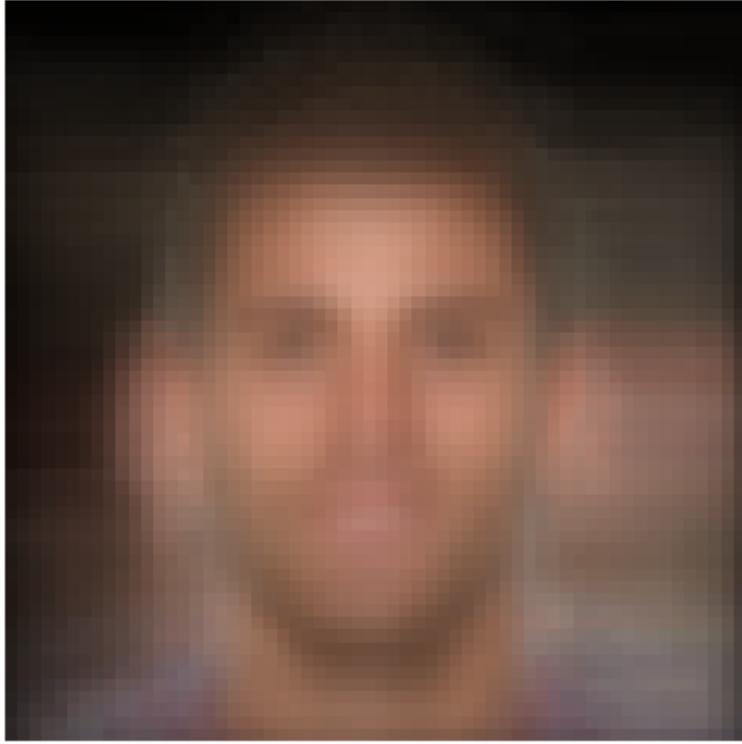
# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['harry kane', 'alex mccarthy', 'sergio aguero', 'danny ings', 'richarlison', 'jay rodriguez', 'eden hazard', 'danil o', 'harvey barnes', 'jack grealish', 'kortney hause', 'matt targett', 'ørjan nyland', 'romelu lukaku', 'jonjo shelvey', 'mark noble', 'michail antonio', 'michail antonio', 'patrick cutrone', 'danny ings', 'jack grealish', 'matt doherty', 'tammy abraham', 'richarlison', 'lys mousset', 'romelu lukaku', 'teemu pukki', 'teemu pukki', 'jamie vardy', 'ricardo pereira', 'callum paterson', 'kenny mclean', 'antonio barreca', 'cenk tosun', 'david martin', 'danilo', 'diego rico', 'diego rico', 'callum wilson', 'danny ings', 'rui p atrício', 'aaron mooy', 'olivier giroud', 'jordan pickford', 'marcus rashford', 'moussa sissoko', 'sadio mané', 'sadio mané', 'olivier giroud', 'james mcarthur', 'alvaro morata', 'lewis dunk', 'felipe anderson', 'che adams', 'christian eriksen', 'harry kane', 'nathan holland', 'manolo gabbiadini', 'nicolas pépé', 'lys mousset', 'jorginho', 'roberto', 'callum robinson', 'dele alli', 'jordan henderson', 'matt doherty', 'danny ings', 'eric dier', 'kenny mclean', 'tammy abraham', 'willian', 'tammy abraham', 'tammy abrahão', 'anthony martial', 'jordan henderson', 'roberto', 'roberto', 'roberto', 'fernandinho', 'kevin de bruyne', 'jorginho', 'tammy abraham', 'willian', 'kadeem harris', 'abdoulaye doucouré', 'roberto pereyra', 'roberto firmino', 'ben davies', 'declan rice', 'david silva', 'ashley young', 'juan mata', 'mark noble', 'chris wood', 'kadeem harris', 'dennis praet', 'jamie vardy', 'harry wilson', 'ross barkley', 'willian', 'kyle walker', 'ken sema', 'patrick cutrone', 'pedro neto', 'joão moutinho', 'mohamed elyounoussi', 'leroy sané', 'billy sharp', 'youri tielemans', 'patrick cutrone', 'ken sema', 'james milner', 'roberto firmino', 'gabriel jesus', 'ezri konsa', 'adrián', 'georginio wijnaldum', 'chris smalling', 'henrikh mkhitaryan', 'danny ings', 'michail antonio', 'shane long', 'ben davies', 'ederson', 'ederson', 'lucas torreira', 'roberto firmino', 'kepa arrizabalaga', 'manolo gabbiadini', 'neil taylor', 'manolo gabbiadini', 'manolo gabbiadini', 'dele alli', 'patrick roberts', 'harry kane', 'dennis praet', 'ciaran clark', 'joelinton', 'ben foster', 'gabriel jesus', 'aaron ramsey', 'chris wood', 'chris wood', 'jack cork', 'harry maguire', 'ben foster', 'dean henderson', 'gerard deulofeu', 'dele alli', 'richarlison', 'david silva', 'richarlison', 'chris wood', 'aaron mooy', 'raheem sterling', 'cesar azpilicueta', 'jorginho', 'jão cancelo', 'callum wilson', 'callum wilson', 'steve cook', 'theo walcott', 'christoph zimmermann', 'willian', 'jack grealish', 'danny ings', 'dele alli', 'andriy yarmolenko', 'nick pope', 'josh murphy', 'dean henderson', 'harry wilson', 'anthony knockaert', 'dele alli', 'harry wilson', 'michy batshuayi', 'brandon williams', 'willian', 'lewis cook', 'harry kane', 'serge aurier', 'charlie austin', 'bernardo silva', 'gabriel jesus', 'gabriel jesus', 'pedro neto', 'romain saïss', 'danilo', 'danilo', 'stefano okaka', 'david mcgoldrick', 'harvey barnes', 'jamie vardy', 'eric dier', 'harry kane', 'reiss nelson', 'willian', 'joelinton', 'jonny evans', 'fred', 'romelu lukaku', 'jonathan kodjia', 'willian', 'raheem sterling', 'brandon williams', 'callum wilson', 'callum wilson', 'dennis praet', 'christian pulisic', 'gabriel jesus', 'nathan redmond', 'stuart armstrong', 'david brooks', 'harry wilson', 'wayne hennessey']

Image paths for these players: ['[https://cdn.sofifa.net/players/194/958/18\\_60.png](https://cdn.sofifa.net/players/194/958/18_60.png)', '[https://cdn.sofifa.net/players/186/561/18\\_60.png](https://cdn.sofifa.net/players/186/561/18_60.png)', '[https://cdn.sofifa.net/players/208/135/18\\_60.png](https://cdn.sofifa.net/players/208/135/18_60.png)', '[https://cdn.sofifa.net/players/194/911/18\\_60.png](https://cdn.sofifa.net/players/194/911/18_60.png)', '[https://cdn.sofifa.net/players/189/324/18\\_60.png](https://cdn.sofifa.net/players/189/324/18_60.png)', '[https://cdn.sofifa.net/players/201/153/18\\_60.png](https://cdn.sofifa.net/players/201/153/18_60.png)', '[https://cdn.sofifa.net/players/194/794/18\\_60.png](https://cdn.sofifa.net/players/194/794/18_60.png)', '[https://cdn.sofifa.net/players/203/910/18\\_60.png](https://cdn.sofifa.net/players/203/910/18_60.png)', '[https://cdn.sofifa.net/players/211/300/18\\_60.png](https://cdn.sofifa.net/players/211/300/18_60.png)', '[https://cdn.sofifa.net/players/220/493/18\\_60.png](https://cdn.sofifa.net/players/220/493/18_60.png)', '[https://cdn.sofifa.net/players/152/908/18\\_60.png](https://cdn.sofifa.net/players/152/908/18_60.png)', '[https://cdn.sofifa.net/players/205/923/18\\_60.png](https://cdn.sofifa.net/players/205/923/18_60.png)', '[https://cdn.sofifa.net/players/163/155/18\\_60.png](https://cdn.sofifa.net/players/163/155/18_60.png)', '[https://cdn.sofifa.net/players/218/667/18\\_60.png](https://cdn.sofifa.net/players/218/667/18_60.png)', '[https://cdn.sofifa.net/players/163/050/18\\_60.png](https://cdn.sofifa.net/players/163/050/18_60.png)', '[https://cdn.sofifa.net/players/138/786/18\\_60.png](https://cdn.sofifa.net/players/138/786/18_60.png)', '[https://cdn.sofifa.net/players/210/724/18\\_60.png](https://cdn.sofifa.net/players/210/724/18_60.png)', '[https://cdn.sofifa.net/players/196/978/18\\_60.png](https://cdn.sofifa.net/players/196/978/18_60.png)', '[https://cdn.sofifa.net/players/192/991/18\\_60.png](https://cdn.sofifa.net/players/192/991/18_60.png)', '[https://cdn.sofifa.net/players/195/671/18\\_60.png](https://cdn.sofifa.net/players/195/671/18_60.png)', '[https://cdn.sofifa.net/players/226/093/18\\_60.png](https://cdn.sofifa.net/players/226/093/18_60.png)', '<https://cdn.sofifa.net/players/189/881/1>']

8\_60.png', 'https://cdn.sofifa.net/players/192/123/18\_60.png', 'https://cdn.sofifa.net/players/190/460/18\_60.png', 'https://cdn.sofifa.net/players/227/796/18\_60.png', 'https://cdn.sofifa.net/players/239/322/18\_60.png', 'https://cdn.sofifa.net/players/183/129/18\_60.png', 'https://cdn.sofifa.net/players/199/304/18\_60.png', 'https://cdn.sofifa.net/players/195/859/18\_60.png', 'https://cdn.sofifa.net/players/220/196/18\_60.png', 'https://cdn.sofifa.net/players/116/314/18\_60.png', 'https://cdn.sofifa.net/players/180/818/18\_60.png', 'https://cdn.sofifa.net/players/233/306/18\_60.png', 'https://cdn.sofifa.net/players/234/378/18\_60.png', 'https://cdn.sofifa.net/players/199/652/18\_60.png', 'https://cdn.sofifa.net/players/220/414/18\_60.png', 'https://cdn.sofifa.net/players/210/257/18\_60.png', 'https://cdn.sofifa.net/players/202/335/18\_60.png', 'https://cdn.sofifa.net/players/227/678/18\_60.png', 'https://cdn.sofifa.net/players/201/995/18\_60.png', 'https://cdn.sofifa.net/players/135/507/18\_60.png', 'https://cdn.sofifa.net/players/209/297/18\_60.png', 'https://cdn.sofifa.net/players/230/666/18\_60.png', 'https://cdn.sofifa.net/players/181/291/18\_60.png', 'https://cdn.sofifa.net/players/202/477/18\_60.png', 'https://cdn.sofifa.net/players/203/263/18\_60.png', 'https://cdn.sofifa.net/players/220/710/18\_60.png', 'https://cdn.sofifa.net/players/234/742/18\_60.png', 'https://cdn.sofifa.net/players/192/883/18\_60.png', 'https://cdn.sofifa.net/players/169/706/18\_60.png', 'https://cdn.sofifa.net/players/206/517/18\_60.png', 'https://cdn.sofifa.net/players/171/972/18\_60.png', 'https://cdn.sofifa.net/players/138/412/18\_60.png', 'https://cdn.sofifa.net/players/208/830/18\_60.png', 'https://cdn.sofifa.net/players/223/334/18\_60.png', 'https://cdn.sofifa.net/players/189/099/18\_60.png', 'https://cdn.sofifa.net/players/189/165/18\_60.png', 'https://cdn.sofifa.net/players/169/588/18\_60.png', 'https://cdn.sofifa.net/players/183/711/18\_60.png', 'https://cdn.sofifa.net/players/204/935/18\_60.png', 'https://cdn.sofifa.net/players/205/498/18\_60.png', 'https://cdn.sofifa.net/players/210/514/18\_60.png', 'https://cdn.sofifa.net/players/162/347/18\_60.png', 'https://cdn.sofifa.net/players/178/088/18\_60.png', 'https://cdn.sofifa.net/players/197/170/18\_60.png', 'https://cdn.sofifa.net/players/213/750/18\_60.png', 'https://cdn.sofifa.net/players/201/368/18\_60.png', 'https://cdn.sofifa.net/players/206/585/18\_60.png', 'https://cdn.sofifa.net/players/192/985/18\_60.png', 'https://cdn.sofifa.net/players/210/635/18\_60.png', 'https://cdn.sofifa.net/players/188/377/18\_60.png', 'https://cdn.sofifa.net/players/222/492/18\_60.png', 'https://cdn.sofifa.net/players/224/294/18\_60.png', 'https://cdn.sofifa.net/players/199/915/18\_60.png', 'https://cdn.sofifa.net/players/223/959/18\_60.png', 'https://cdn.sofifa.net/players/221/618/18\_60.png', 'https://cdn.sofifa.net/players/198/683/18\_60.png', 'https://cdn.sofifa.net/players/231/677/18\_60.png', 'https://cdn.sofifa.net/players/152/879/18\_60.png', 'https://cdn.sofifa.net/players/201/417/18\_60.png', 'https://cdn.sofifa.net/players/218/659/18\_60.png', 'https://cdn.sofifa.net/players/190/717/18\_60.png', 'https://cdn.sofifa.net/players/204/529/18\_60.png', 'https://cdn.sofifa.net/players/203/362/18\_60.png', 'https://cdn.sofifa.net/players/183/394/18\_60.png', 'https://cdn.sofifa.net/players/237/683/18\_60.png', 'https://cdn.sofifa.net/players/198/719/18\_60.png', 'https://cdn.sofifa.net/players/173/735/18\_60.png', 'https://cdn.sofifa.net/players/203/841/18\_60.png', 'https://cdn.sofifa.net/players/226/110/18\_60.png', 'https://cdn.sofifa.net/players/178/509/18\_60.png', 'https://cdn.sofifa.net/players/237/715/18\_60.png', 'https://cdn.sofifa.net/players/221/982/18\_60.png', 'https://cdn.sofifa.net/players/238/616/18\_60.png', 'https://cdn.sofifa.net/players/202/652/18\_60.png', 'https://cdn.sofifa.net/players/231/448/18\_60.png', 'https://cdn.sofifa.net/players/210/243/18\_60.png', 'https://cdn.sofifa.net/players/231/943/18\_60.png', 'https://cdn.sofifa.net/players/183/666/18\_60.png', 'https://cdn.sofifa.net/players/201/942/18\_60.png', 'https://cdn.sofifa.net/players/193/061/18\_60.png', 'https://cdn.sofifa.net/players/208/295/18\_60.png', 'https://cdn.sofifa.net/players/192/505/18\_60.png', 'https://cdn.sofifa.net/players/199/189/18\_60.png', 'https://cdn.sofifa.net/players/178/005/18\_60.png', 'https://cdn.sofifa.net/players/208/722/18\_60.png', 'https://cdn.sofifa.net/players/197/853/18\_60.png', 'https://cdn.sofifa.net/players/153/079/1

```
8_60.png', 'https://cdn.sofifa.net/players/169/216/18_60.png', 'https://cdn.sofifa.net/players/178/091/18_60.png', 'https://cdn.sofifa.net/players/193/011/18_60.png', 'https://cdn.sofifa.net/players/202/282/18_60.png', 'https://cdn.sofifa.net/players/231/352/18_60.png', 'https://cdn.sofifa.net/players/190/362/18_60.png', 'https://cdn.sofifa.net/players/164/859/18_60.png', 'https://cdn.sofifa.net/players/170/084/18_60.png', 'https://cdn.sofifa.net/players/180/403/18_60.png', 'https://cdn.sofifa.net/players/216/393/18_60.png', 'https://cdn.sofifa.net/players/185/427/18_60.png']
```



In [ ]:

In [ ]:

The code below is attempting to show the average face for players that have words to describe them in either a positive or negative sentiment

```
In [275...]: def get_image_paths_for_word_with_sentiment(word, sentiment, df_player_images):
    if sentiment == 'positive':
        physical_sentiment_col = 'Physical - Positive Sentiment'
        cognitive_sentiment_col = 'Cognitive - Positive Sentiment'
    elif sentiment == 'negative':
        physical_sentiment_col = 'Physical - Negative Sentiment'
        cognitive_sentiment_col = 'Cognitive - Negative Sentiment'
    else:
        raise ValueError("Sentiment must be 'positive' or 'negative'")

    # Filter rows where the word is in the physical words column and sentiment
    physical_filter = (main_df['physical words'].str.strip().str.lower() == word)
    cognitive_filter = (main_df['cognitive words'].str.strip().str.lower() == word)

    # Combine the two filters
    filtered_players = main_df[physical_filter | cognitive_filter]['Player']
    print("Players described by the word with", sentiment, "sentiment:", filtered_players)

    # Normalize player names in df_player_images_raw
    df_player_images_raw['Player Names'] = df_player_images_raw['Player Names'].str.strip()

    # Get image paths for these players
    image_paths = df_player_images_raw[df_player_images_raw['Player Names'].isin(filtered_players)]['image_path']
    print("Image paths for these players:", image_paths)
```

```
    return image_paths
```

```
In [256...]
# Choose a descriptive word and sentiment
word = 'control'
sentiment = 'negative'

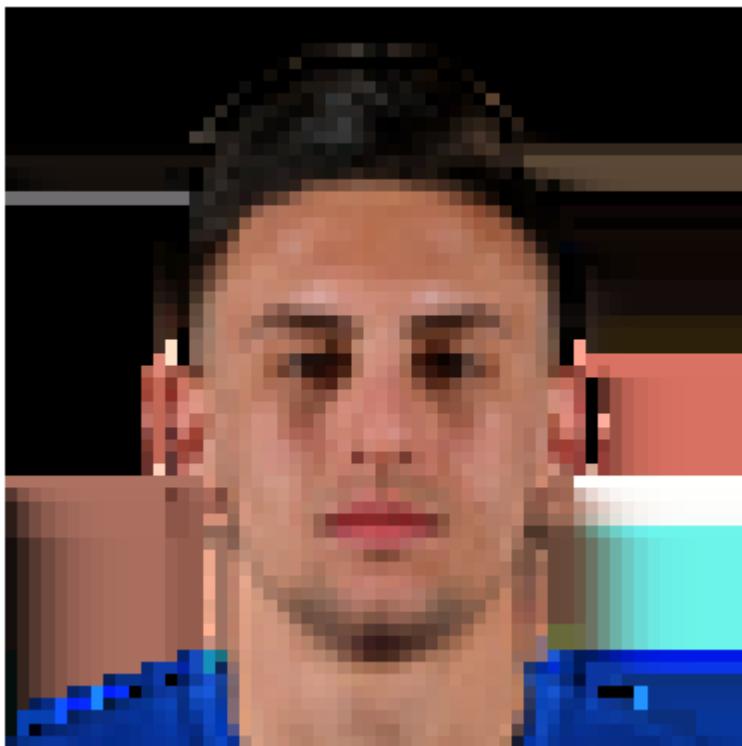
# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment")
```

Players described by the word with negative sentiment: ['antonio barreca']  
 Image paths for these players: ['https://cdn.sofifa.net/players/220/493/18\_60.png']



```
In [259...]
# Choose a descriptive word and sentiment
word = 'finish'
sentiment = 'negative'

# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
```

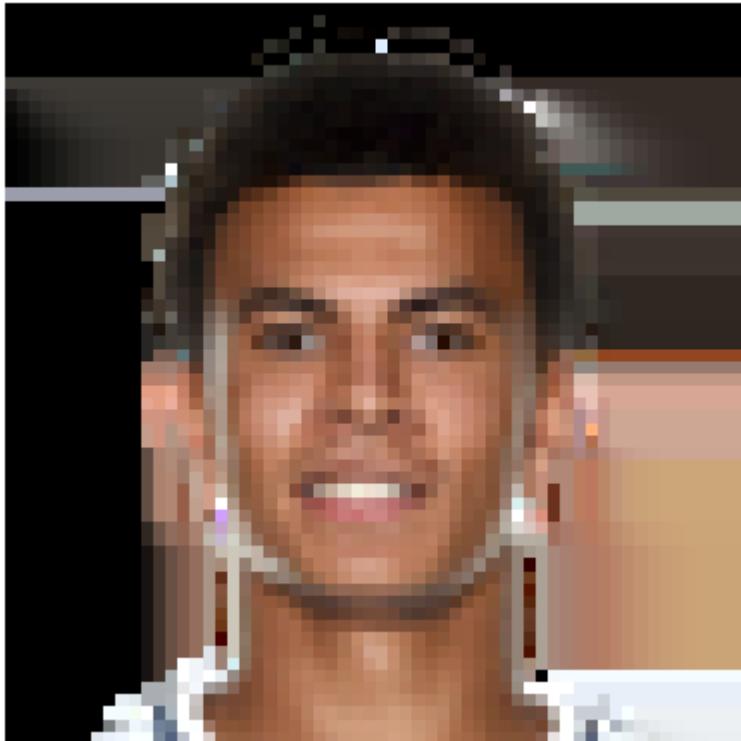
```

plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
plt.axis('off')
plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment"

```

Players described by the word with negative sentiment: ['dele alli']  
 Image paths for these players: ['https://cdn.sofifa.net/players/211/117/18\_60.png']



In [276...]

```

# Choose a descriptive word and sentiment
word = 'cleverly'
sentiment = 'positive'

# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment"

```

Players described by the word with positive sentiment: []  
 Image paths for these players: []  
 No faces found for the word 'cleverly' with positive sentiment

In [181...]

```

# Choose a descriptive word and sentiment
word = 'strike'

```

```
sentiment = 'negative'

# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment")
```

Players described by the word with negative sentiment: ['ricardo pereira']  
Image paths for these players: ['https://cdn.sofifa.net/players/210/243/18\_60.png']



In [ ]:

In [ ]:																																																																																																																																					
In [ ]:																																																																																																																																					
In [ ]:																																																																																																																																					
In [ ]:																																																																																																																																					
In [ ]:																																																																																																																																					
In [69]:	df_skin_tones																																																																																																																																				
Out[69]:	<table border="1"> <thead> <tr> <th></th><th>file</th><th>image type</th><th>face id</th><th>dominant 1</th><th>percent 1</th><th>dominant 2</th><th>percent 2</th><th>skin tone</th><th>tone label</th><th>ac</th></tr> </thead> <tbody> <tr> <td>0</td><td>Aaron Cresswell.jpg</td><td>color</td><td>1.0</td><td>#DCA183</td><td>0.66</td><td>#896850</td><td>0.34</td><td>#BEA07E</td><td>CG</td><td></td></tr> <tr> <td>1</td><td>Aaron Mooy.jpg</td><td>color</td><td>1.0</td><td>#B99684</td><td>0.56</td><td>#8A7261</td><td>0.44</td><td>#9D7A54</td><td>CF</td><td></td></tr> <tr> <td>2</td><td>Aaron Ramsdale.jpg</td><td>color</td><td>1.0</td><td>#E2A388</td><td>0.57</td><td>#A16D51</td><td>0.43</td><td>#9D7A54</td><td>CF</td><td></td></tr> <tr> <td>3</td><td>Aaron Ramsey.jpg</td><td>color</td><td>1.0</td><td>#E2BCA8</td><td>0.71</td><td>#B28670</td><td>0.29</td><td>#E7C1B8</td><td>CI</td><td></td></tr> <tr> <td>4</td><td>Aaron Wan-Bissaka.jpg</td><td>color</td><td>1.0</td><td>#5A2B1F</td><td>0.72</td><td>#7A4A44</td><td>0.28</td><td>#513B2E</td><td>CC</td><td></td></tr> <tr> <td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr> <td>687</td><td>Étienne Capoue.jpg</td><td>color</td><td>1.0</td><td>#BC745A</td><td>0.61</td><td>#74442F</td><td>0.39</td><td>#81654F</td><td>CE</td><td></td></tr> <tr> <td>688</td><td>Ørjan Nyland.jpg</td><td>color</td><td>1.0</td><td>#DDAC9B</td><td>0.60</td><td>#BB8773</td><td>0.40</td><td>#E7C1B8</td><td>CI</td><td></td></tr> <tr> <td>689</td><td>İlkay Gündoğan.jpg</td><td>color</td><td>1.0</td><td>#D48F6E</td><td>0.53</td><td>#784D3A</td><td>0.47</td><td>#81654F</td><td>CE</td><td></td></tr> <tr> <td>690</td><td>Łukasz Fabiański.jpg</td><td>color</td><td>1.0</td><td>#C18B72</td><td>0.62</td><td>#815944</td><td>0.38</td><td>#9D7A54</td><td>CF</td><td></td></tr> <tr> <td>691</td><td>Łukasz Fabiański.jpg</td><td>color</td><td>1.0</td><td>#C18B72</td><td>0.62</td><td>#815944</td><td>0.38</td><td>#9D7A54</td><td>CF</td><td></td></tr> </tbody> </table> <p>692 rows × 11 columns</p>		file	image type	face id	dominant 1	percent 1	dominant 2	percent 2	skin tone	tone label	ac	0	Aaron Cresswell.jpg	color	1.0	#DCA183	0.66	#896850	0.34	#BEA07E	CG		1	Aaron Mooy.jpg	color	1.0	#B99684	0.56	#8A7261	0.44	#9D7A54	CF		2	Aaron Ramsdale.jpg	color	1.0	#E2A388	0.57	#A16D51	0.43	#9D7A54	CF		3	Aaron Ramsey.jpg	color	1.0	#E2BCA8	0.71	#B28670	0.29	#E7C1B8	CI		4	Aaron Wan-Bissaka.jpg	color	1.0	#5A2B1F	0.72	#7A4A44	0.28	#513B2E	CC		...	...	...	...	...	...	...	...	...	...	...	687	Étienne Capoue.jpg	color	1.0	#BC745A	0.61	#74442F	0.39	#81654F	CE		688	Ørjan Nyland.jpg	color	1.0	#DDAC9B	0.60	#BB8773	0.40	#E7C1B8	CI		689	İlkay Gündoğan.jpg	color	1.0	#D48F6E	0.53	#784D3A	0.47	#81654F	CE		690	Łukasz Fabiański.jpg	color	1.0	#C18B72	0.62	#815944	0.38	#9D7A54	CF		691	Łukasz Fabiański.jpg	color	1.0	#C18B72	0.62	#815944	0.38	#9D7A54	CF	
	file	image type	face id	dominant 1	percent 1	dominant 2	percent 2	skin tone	tone label	ac																																																																																																																											
0	Aaron Cresswell.jpg	color	1.0	#DCA183	0.66	#896850	0.34	#BEA07E	CG																																																																																																																												
1	Aaron Mooy.jpg	color	1.0	#B99684	0.56	#8A7261	0.44	#9D7A54	CF																																																																																																																												
2	Aaron Ramsdale.jpg	color	1.0	#E2A388	0.57	#A16D51	0.43	#9D7A54	CF																																																																																																																												
3	Aaron Ramsey.jpg	color	1.0	#E2BCA8	0.71	#B28670	0.29	#E7C1B8	CI																																																																																																																												
4	Aaron Wan-Bissaka.jpg	color	1.0	#5A2B1F	0.72	#7A4A44	0.28	#513B2E	CC																																																																																																																												
...	...	...	...	...	...	...	...	...	...	...																																																																																																																											
687	Étienne Capoue.jpg	color	1.0	#BC745A	0.61	#74442F	0.39	#81654F	CE																																																																																																																												
688	Ørjan Nyland.jpg	color	1.0	#DDAC9B	0.60	#BB8773	0.40	#E7C1B8	CI																																																																																																																												
689	İlkay Gündoğan.jpg	color	1.0	#D48F6E	0.53	#784D3A	0.47	#81654F	CE																																																																																																																												
690	Łukasz Fabiański.jpg	color	1.0	#C18B72	0.62	#815944	0.38	#9D7A54	CF																																																																																																																												
691	Łukasz Fabiański.jpg	color	1.0	#C18B72	0.62	#815944	0.38	#9D7A54	CF																																																																																																																												

In [70]: main\_df

Out[70]:

	Player	Context	tone label	descriptive words	physical count	cognitive count	physical words	cognitive words	
0	harry kane	theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in	CE	touch	1.0	0.0	touch	NaN	\$
1	alex mccarthy	target the pace on the ball means he cant react alex mccarthy it goes through his legs hes tiptoeing his way around	CF	tiptoeing	1.0	0.0	tiptoeing	NaN	
2	sergio aguero	expect aguero to do much better walker in towards aguero sergio aguero doing what comes so naturally to him strong ankle gets	CE	strong	1.0	0.0	strong	NaN	
3	danny ings	convincing header boufal up against schneiderlin its in from ings danny ings with the equalizer for southampton at long last it game	CF	header	1.0	0.0	header	NaN	
4	richarlison	wardprowse very few better at striking the dead ball sidibé richarlison brilliantly finished richarlison his first goal in eight	CF	finish	0.0	1.0	NaN	finish	

	Player	Context	tone label	descriptive words	physical count	cognitive count	physical words	cognitive words	
premier league									
398	wayne hennessey	...  does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda	CE	big strong aggressive	3.0	0.0	big strong aggressive	NaN	
399	wayne hennessey	hamšík tidy turn but saved from robert maks effort by wayne hennessey terrific football from the slovakians once again manoeuvre the ball	CE	terrific football	0.0	1.0	NaN	terrific football	
400	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
401	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
402	NaN	NaN	NaN	NaN	344.0	190.0	NaN	NaN	

403 rows × 12 columns

In [ ]:

In [ ]:

In [216]:

```
import pandas as pd
from collections import Counter

# Combine all words in the 'descriptive words' column into a single list
all_descriptive_words = main_df['descriptive words'].dropna().astype(str).to

# Initialize a list to store all words
all_words = []

# Iterate over each phrase in the descriptive words list
for phrase in all_descriptive_words:
    # Split the phrase into words and extend the all_words list
    all_words.extend(phrase.split())

# Count the frequency of each word
word_counts = Counter(all_words)
```

```
# Get the most common words
most_common_words = word_counts.most_common(1000) # Change the number to get more or less words

# Print the most common words with their frequencies
print("Most frequently used words in the 'descriptive words' column:")
for word, count in most_common_words:
    print(f"{word}: {count}")
```

Most frequently used words in the 'descriptive words' column:

finish: 39

touch: 36

header: 35

brilliant: 30

strike: 23

strong: 16

wide: 16

save: 13

good: 10

the: 9

perfect: 8

quick: 8

ball: 8

perfectly: 7

pace: 7

power: 7

movement: 6

lead: 6

run: 6

curl: 5

pounce: 5

goal: 5

in: 5

control: 4

technique: 4

pullback: 4

it: 4

play: 4

lovely: 4

clever: 4

bright: 3

stumble: 3

reaction: 3

massive: 3

neatly: 3

place: 3

pressure: 3

great: 3

weight: 3

mistake: 3

body: 3

strength: 3

beaten: 3

him: 3

right: 3

timed: 2

belter: 2

skill: 2

fierce: 2

unselfish: 2

two: 2

reliable: 2

sharp: 2

brave: 2

big: 2

beat: 2

excellent: 2

intelligent: 2

react: 2

sprint: 2

blocked: 2

contact: 2

up: 2

terrific: 2  
danger: 2  
feet: 2  
found: 2  
of: 2  
past: 2  
error: 2  
covering: 2  
aware: 2  
little: 2  
back: 2  
poor: 2  
delivery: 2  
thinking: 2  
finding: 2  
and: 2  
on: 2  
composure: 2  
battle: 2  
quickly: 2  
force: 2  
unstoppable: 2  
shot: 2  
beautifully: 2  
poked: 2  
to: 2  
burst: 2  
challenges: 2  
smart: 2  
near: 2  
post: 2  
towering: 2  
tiptoeing: 1  
wrong: 1  
side: 1  
stutter: 1  
predatory: 1  
beast: 1  
focus: 1  
anticipation: 1  
shaking: 1  
off: 1  
slipped: 1  
balls: 1  
desire: 1  
taps: 1  
confident: 1  
shimmy: 1  
setup: 1  
stride: 1  
threatening: 1  
shrugged: 1  
reacts: 1  
difficulties: 1  
headed: 1  
purposeful: 1  
pushes: 1  
cleverly: 1  
build: 1  
is: 1  
confidence: 1  
intensity: 1  
involved: 1  
heavily: 1

comes: 1  
alive: 1  
into: 1  
area: 1  
calm: 1  
never: 1  
fazed: 1  
struck: 1  
battling: 1  
venom: 1  
overlap: 1  
well: 1  
pass: 1  
swift: 1  
strides: 1  
effort: 1  
fooled: 1  
keeper: 1  
too: 1  
straight: 1  
promising: 1  
leans: 1  
struggling: 1  
space: 1  
top: 1  
full: 1  
stretch: 1  
stop: 1  
fingertips: 1  
incisive: 1  
skipped: 1  
powerful: 1  
skips: 1  
wild: 1  
makes: 1  
them: 1  
pay: 1  
punching: 1  
block: 1  
value: 1  
energetic: 1  
penalty: 1  
cutting: 1  
shoots: 1  
just: 1  
missing: 1  
what: 1  
a: 1  
fired: 1  
finishing: 1  
positioning: 1  
winner: 1  
chip: 1  
athleticism: 1  
takes: 1  
charge: 1  
finds: 1  
bottom: 1  
corner: 1  
away: 1  
foul: 1  
won: 1  
fortunate: 1  
hurriedly: 1

ahead: 1  
decision: 1  
denying: 1  
smashing: 1  
long: 1  
throw: 1  
linkup: 1  
lean: 1  
slick: 1  
pure: 1  
running: 1  
from: 1  
deep: 1  
stike: 1  
easy: 1  
for: 1  
nice: 1  
move: 1  
direction: 1  
sprawl: 1  
playing: 1  
lay: 1  
genius: 1  
outstanding: 1  
truly: 1  
absolutely: 1  
spot: 1  
fires: 1  
strikes: 1  
gold: 1  
crucial: 1  
mentally: 1  
punched: 1  
placed: 1  
spins: 1  
scores: 1  
fit: 1  
persistence: 1  
bravery: 1  
arrowed: 1  
bouldered: 1  
leadership: 1  
quality: 1  
moment: 1  
again: 1  
stance: 1  
couldnt: 1  
quite: 1  
take: 1  
with: 1  
gets: 1  
there: 1  
first: 1  
defending: 1  
special: 1  
turn: 1  
stunner: 1  
exquisite: 1  
deaf: 1  
finished: 1  
alert: 1  
language: 1  
lets: 1  
fly: 1

sneaked: 1  
nifty: 1  
immaculate: 1  
smooth: 1  
made: 1  
amends: 1  
capable: 1  
executed: 1  
put: 1  
everton: 1  
sword: 1  
wise: 1  
muscle: 1  
slip: 1  
has: 1  
do: 1  
more: 1  
miscued: 1  
work: 1  
rate: 1  
picked: 1  
his: 1  
pocket: 1  
form: 1  
riding: 1  
athletic: 1  
time: 1  
majestic: 1  
magnificant: 1  
beauty: 1  
curling: 1  
leaning: 1  
lethal: 1  
emphatic: 1  
magnificent: 1  
wonderful: 1  
footwork: 1  
feast: 1  
couldn't: 1  
apply: 1  
telling: 1  
buildup: 1  
jet: 1  
heeled: 1  
acceleration: 1  
-: 1  
switched: 1  
punches: 1  
reads: 1  
script: 1  
presence: 1  
brushing: 1  
aside: 1  
feather: 1  
brings: 1  
down: 1  
drills: 1  
magic: 1  
anger: 1  
disguise: 1  
rare: 1  
break: 1  
free: 1

```
aggressive: 1
football: 1
```

In [281]:

```
# Choose descriptive words
words = ['predator']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['romelu lukaku']

Image paths for these players: ['https://cdn.sofifa.net/players/192/505/18\_60.png']



In [ ]:

In [ ]:

In [210]:

```
# Choose descriptive words
words = ['brave']

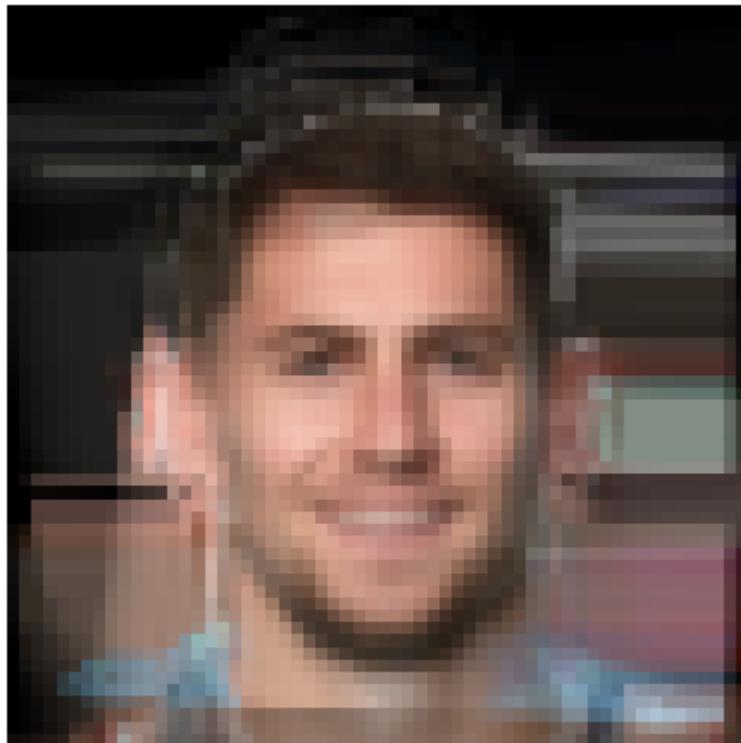
# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

# Generate the average face
average_face = get_average_face(image_paths)
```

```
# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['jorginho', 'ben davies', 'tom heaton']  
Image paths for these players: ['https://cdn.sofifa.net/players/205/923/18\_60.png', 'https://cdn.sofifa.net/players/205/498/18\_60.png', 'https://cdn.sofifa.net/players/163/264/18\_60.png']



In [ ]:

```
# Choose a descriptive word and sentiment
word = 'timed'
sentiment = 'positive'

# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment")
```

Players described by the word with positive sentiment: []  
 Image paths for these players: []  
 No faces found for the word 'timed' with positive sentiment

In [193...]

```
# Choose descriptive words
words = ['control']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['kortney hause', 'antonio barreca', 'abdou  
 laye doucouré', 'josh murphy']  
 Image paths for these players: ['https://cdn.sofifa.net/players/208/135/18\_60.png',  
 'https://cdn.sofifa.net/players/220/493/18\_60.png', 'https://cdn.sofifa.net/players/206/085/18\_60.png',  
 'https://cdn.sofifa.net/players/210/635/18\_60.png']



In [189...]

```
# Choose descriptive words
words = ['bright']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

# Generate the average face
average_face = get_average_face(image_paths)
```

```
# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['jamie vardy', 'christian pulisic']  
 Image paths for these players: ['https://cdn.sofifa.net/players/227/796/18\_60.png', 'https://cdn.sofifa.net/players/208/830/18\_60.png']



In [182]:

```
# Choose descriptive words
words = ['save']

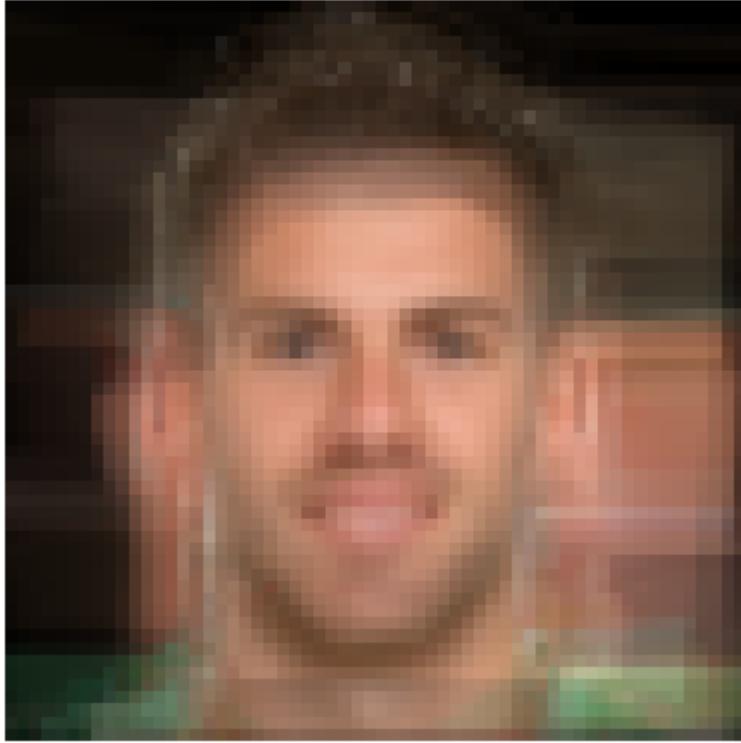
# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['simon mignolet', 'dean henderson', 'ben foster', 'ederson', 'roberto', 'adrián', 'jordan pickford', 'fraser forster', 'ben foster', 'dean henderson', 'fraser forster']  
 Image paths for these players: ['https://cdn.sofifa.net/players/194/911/18\_60.png', 'https://cdn.sofifa.net/players/163/155/18\_60.png', 'https://cdn.sofifa.net/players/233/306/18\_60.png', 'https://cdn.sofifa.net/players/210/257/18\_60.png', 'https://cdn.sofifa.net/players/172/203/18\_60.png', 'https://cdn.sofifa.net/players/204/935/18\_60.png', 'https://cdn.sofifa.net/players/183/666/18\_60.png', 'https://cdn.sofifa.net/players/173/426/18\_60.png']



```
In [177...]: # Choose descriptive words
words = ['leader']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['tom heaton']  
 Image paths for these players: ['https://cdn.sofifa.net/players/163/264/18\_60.png']



```
In [169...]: # Choose descriptive words
words = ['good']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

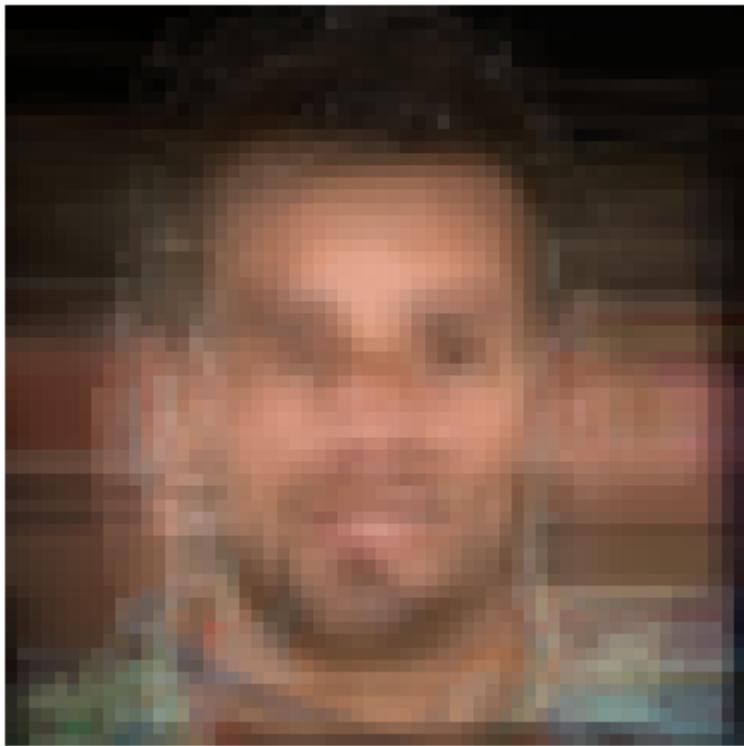
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['jorginho', 'simon mignolet', 'jonny', 'roberto', 'ashley young', 'cesar azpilicueta', 'tom cleverley', 'ben foster', 'brandon williams', 'danilo']

Image paths for these players: ['https://cdn.sofifa.net/players/152/908/18\_60.png', 'https://cdn.sofifa.net/players/163/155/18\_60.png', 'https://cdn.sofifa.net/players/138/786/18\_60.png', 'https://cdn.sofifa.net/players/184/432/18\_60.png', 'https://cdn.sofifa.net/players/199/304/18\_60.png', 'https://cdn.sofifa.net/players/210/455/18\_60.png', 'https://cdn.sofifa.net/players/205/498/18\_60.png', 'https://cdn.sofifa.net/players/183/666/18\_60.png', 'https://cdn.sofifa.net/players/173/426/18\_60.png', 'https://cdn.sofifa.net/players/189/725/18\_60.png']



```
In [ ]: # Choose descriptive words
words = ['quick']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

```
In [168...]: # Choose descriptive words
words = ['movement']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

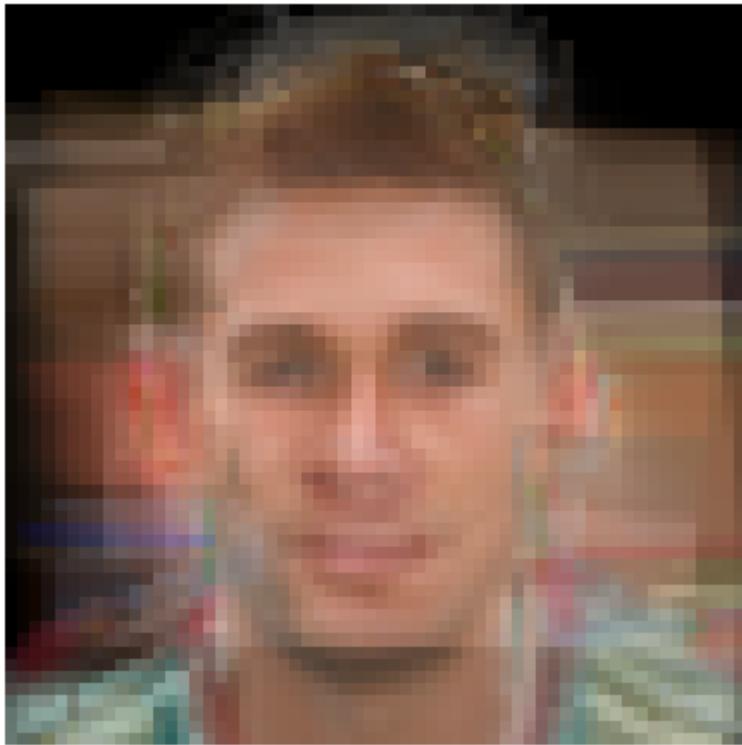
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
```

```
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['harvey barnes', 'james maddison', 'joelinton', 'teemu pukki', 'chris basham', 'harry kane']  
 Image paths for these players: ['https://cdn.sofifa.net/players/184/274/18\_60.png', 'https://cdn.sofifa.net/players/202/126/18\_60.png', 'https://cdn.sofifa.net/players/234/742/18\_60.png', 'https://cdn.sofifa.net/players/220/697/18\_60.png', 'https://cdn.sofifa.net/players/223/334/18\_60.png', 'https://cdn.sofifa.net/players/190/362/18\_60.png']



```
In [167]: # Choose descriptive words
words = ['quick']

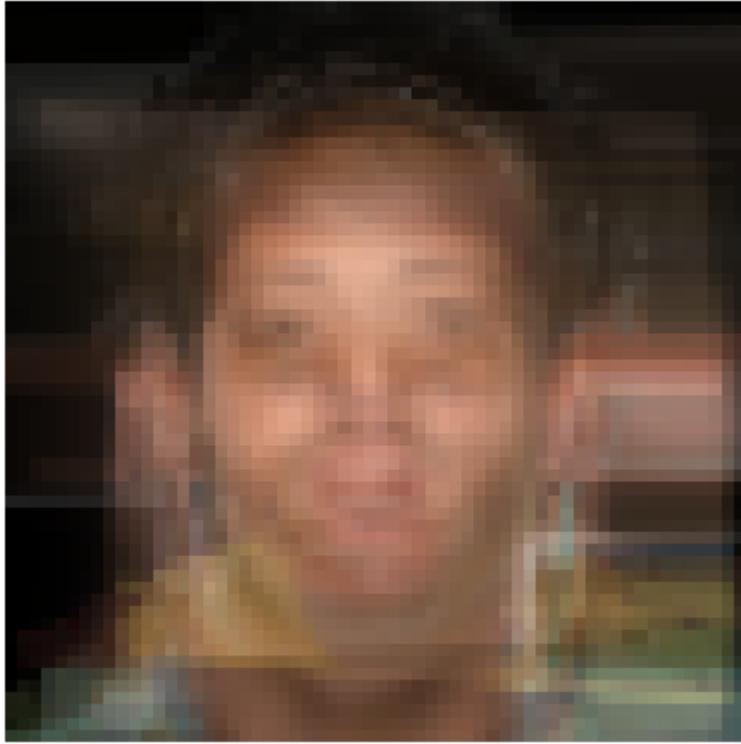
# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{' '.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['david martin', 'harry wilson', 'danny ing s', 'ederson', 'david silva', 'raheem sterling', 'brandon williams', 'gabriel jesus', 'reiss nelson']  
Image paths for these players: ['https://cdn.sofifa.net/players/138/786/18\_60.png', 'https://cdn.sofifa.net/players/195/859/18\_60.png', 'https://cdn.sofifa.net/players/116/314/18\_60.png', 'https://cdn.sofifa.net/players/168/542/18\_60.png', 'https://cdn.sofifa.net/players/210/257/18\_60.png', 'http s://cdn.sofifa.net/players/230/666/18\_60.png', 'https://cdn.sofifa.net/players/220/710/18\_60.png', 'https://cdn.sofifa.net/players/202/652/18\_60.png', 'https://cdn.sofifa.net/players/231/448/18\_60.png']



```
In [166]: # Choose descriptive words
words = ['perfect']

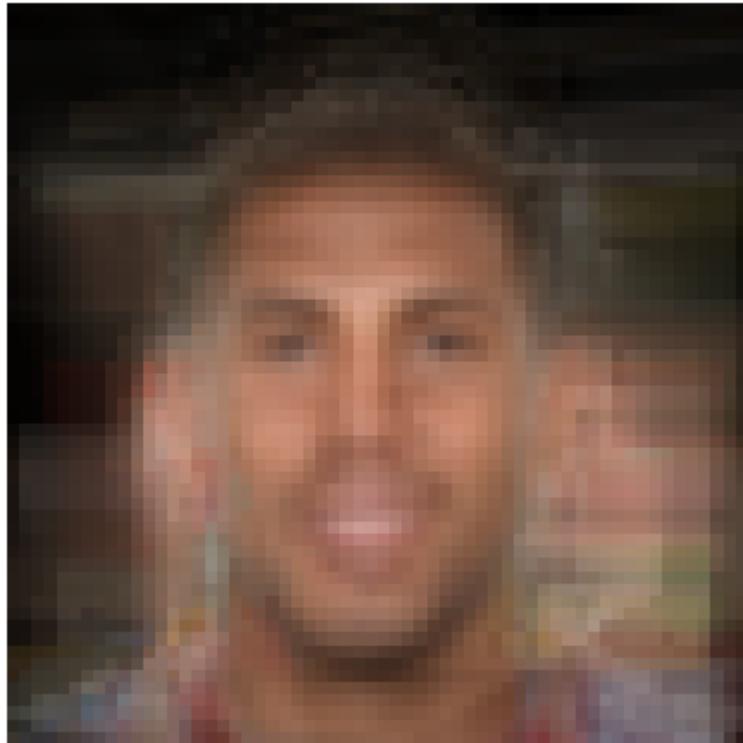
# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['chris smalling', 'patrick cutrone', 'willian', 'kenny mclean', 'conor hourihane', 'willian', 'kevin de bruyne', 'rob erto firmino', 'olivier giroud', 'tammy abraham', 'callum wilson', 'alex iwo bi', 'ashley young', 'christian benteke', 'mason mount']  
Image paths for these players: ['https://cdn.sofifa.net/players/213/655/18\_60.png', 'https://cdn.sofifa.net/players/152/908/18\_60.png', 'https://cdn.sofifa.net/players/196/978/18\_60.png', 'https://cdn.sofifa.net/players/189/81/18\_60.png', 'https://cdn.sofifa.net/players/184/111/18\_60.png', 'https://cdn.sofifa.net/players/193/849/18\_60.png', 'https://cdn.sofifa.net/players/201/368/18\_60.png', 'https://cdn.sofifa.net/players/233/064/18\_60.png', 'https://cdn.sofifa.net/players/178/509/18\_60.png', 'https://cdn.sofifa.net/players/237/715/18\_60.png', 'https://cdn.sofifa.net/players/201/942/18\_60.png', 'https://cdn.sofifa.net/players/231/352/18\_60.png', 'https://cdn.sofifa.net/players/180/403/18\_60.png']



```
In [160]: # Choose descriptive words
words = ['pounce']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

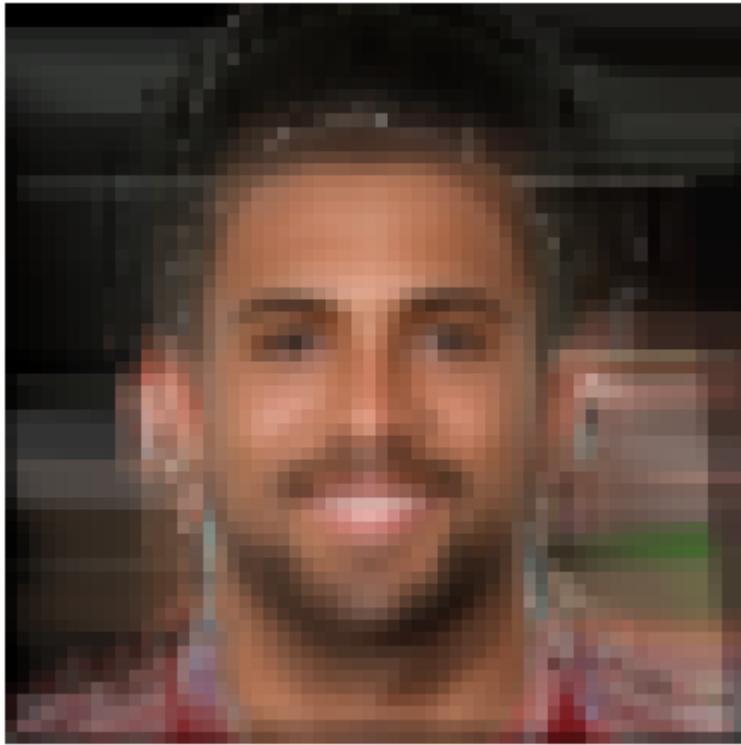
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['romelu lukaku', 'danny ings', 'olivier gi roud', 'joão cancelo', 'callum wilson']

Image paths for these players: ['https://cdn.sofifa.net/players/196/978/18\_60.png', 'https://cdn.sofifa.net/players/195/859/18\_60.png', 'https://cdn.sofifa.net/players/210/514/18\_60.png', 'https://cdn.sofifa.net/players/178/509/18\_60.png', 'https://cdn.sofifa.net/players/192/505/18\_60.png']



```
In [159]: # Choose descriptive words
words = ['control']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

    # Save the average face to a file
    cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['kortney hause', 'antonio barreca', 'abdou laye doucouré', 'josh murphy']

Image paths for these players: ['https://cdn.sofifa.net/players/208/135/18\_60.png', 'https://cdn.sofifa.net/players/220/493/18\_60.png', 'https://cdn.sofifa.net/players/206/085/18\_60.png', 'https://cdn.sofifa.net/players/210/635/18\_60.png']



```
In [158...]: # Choose descriptive words
words = ['technique']

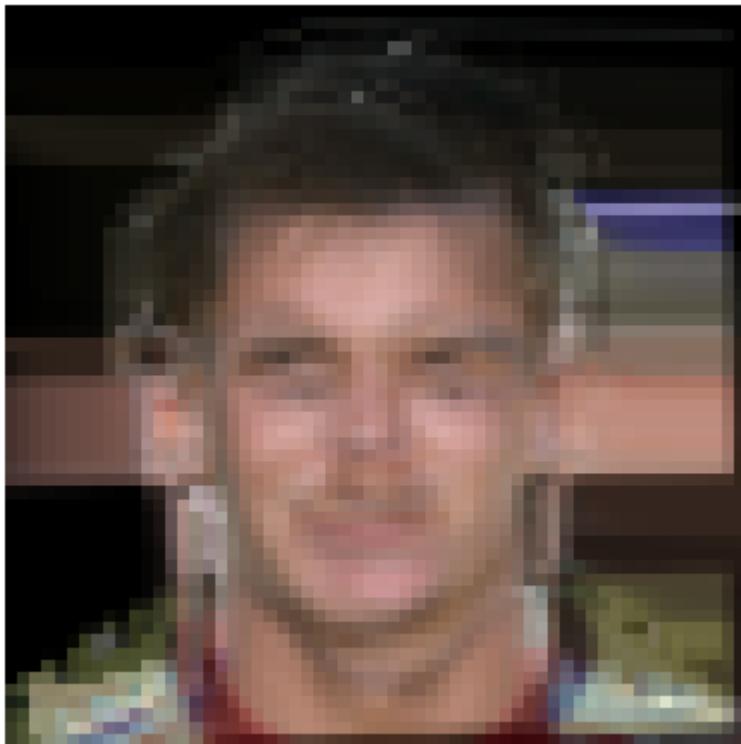
# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['jesse lingard', 'brandon williams', 'kieran dowell', 'theo walcott']  
Image paths for these players: ['https://cdn.sofifa.net/players/138/786/18\_60.png', 'https://cdn.sofifa.net/players/207/494/18\_60.png', 'https://cdn.sofifa.net/players/226/401/18\_60.png', 'https://cdn.sofifa.net/players/164/859/18\_60.png']



In [ ]:

In [ ]:

```
In [176...]: # Choose a descriptive word and sentiment
word = 'control'
sentiment = 'negative'

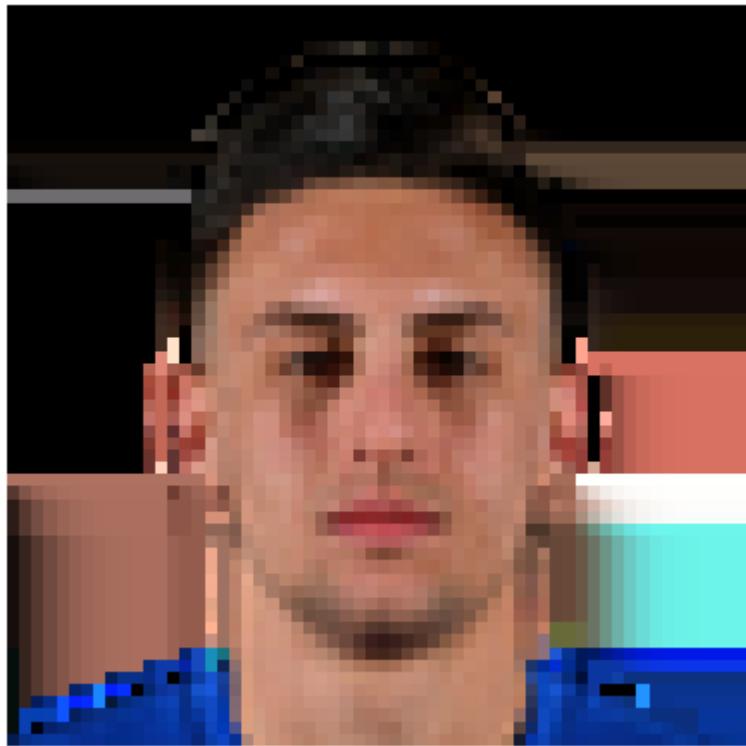
# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment")
```

Players described by the word with negative sentiment: ['antonio barreca']  
Image paths for these players: ['https://cdn.sofifa.net/players/220/493/18\_60.png']



In [ ]:

```
In [196]: # Choose a descriptive word and sentiment
word = 'clever'
sentiment = 'positive'

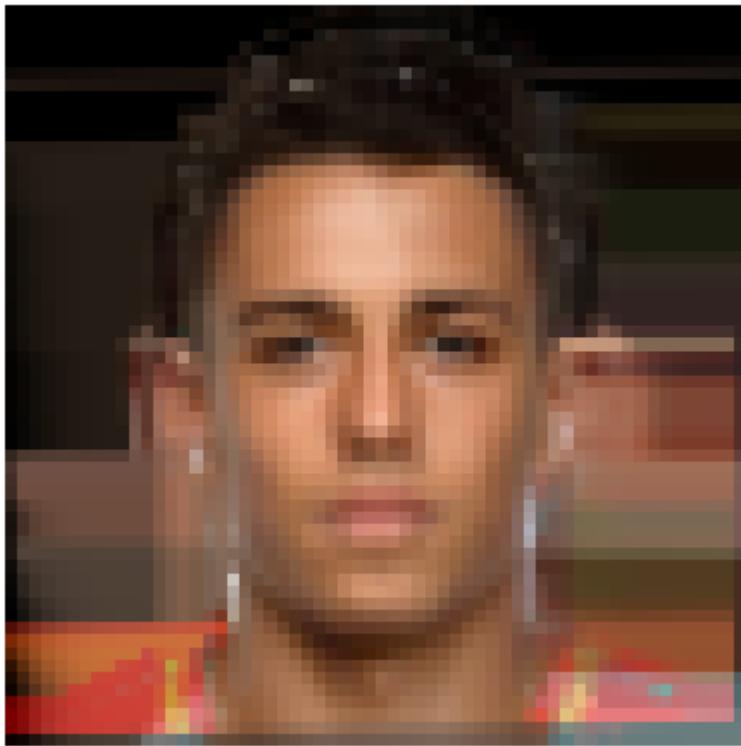
# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment")

Players described by the word with positive sentiment: ['youri tielemans', 'phil foden']
Image paths for these players: ['https://cdn.sofifa.net/players/237/692/18_60.png', 'https://cdn.sofifa.net/players/216/393/18_60.png']
```



```
In [197]: # Choose a descriptive word and sentiment
word = 'power'
sentiment = 'positive'

# Get the image paths of players described by the word with the specified sentiment
image_paths = get_image_paths_for_word_with_sentiment(word, sentiment, df_players)

# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{word}_{sentiment}.jpg", average_face)
else:
    print(f"No faces found for the word '{word}' with {sentiment} sentiment")
```

Players described by the word with positive sentiment: ['callum robinson']  
Image paths for these players: ['https://cdn.sofifa.net/players/206/518/18\_60.png']



```
In [155...]: # Choose descriptive words
words = ['power']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_d...

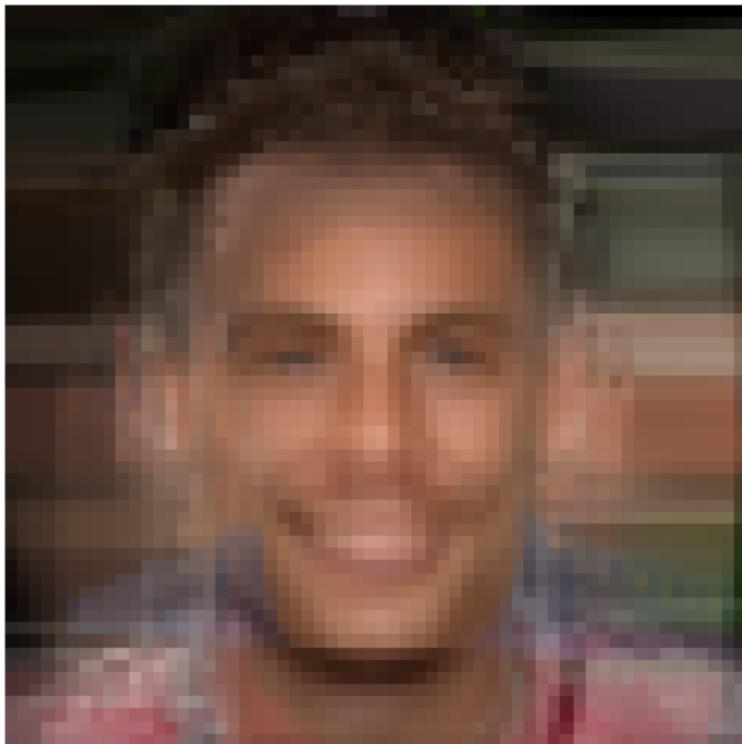
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['callum robinson', 'jordan henderson', 'tammy abraham', 'anthony martial', 'ken sema', 'james milner', 'lucas torreira', 'theo walcott']

Image paths for these players: ['https://cdn.sofifa.net/players/211/300/18\_60.png', 'https://cdn.sofifa.net/players/206/518/18\_60.png', 'https://cdn.sofifa.net/players/138/412/18\_60.png', 'https://cdn.sofifa.net/players/183/711/18\_60.png', 'https://cdn.sofifa.net/players/213/750/18\_60.png', 'https://cdn.sofifa.net/players/223/959/18\_60.png', 'https://cdn.sofifa.net/players/231/352/18\_60.png', 'https://cdn.sofifa.net/players/164/859/18\_60.png']



In [ ]:

In [ ]:

In [ ]:

In [ ]:

```
In [179]: # Choose descriptive words
words = ['power', 'power ']

# Get the image paths of players described by the words
image_paths = get_image_paths_for_words(words, df_player_images_raw, main_dir)

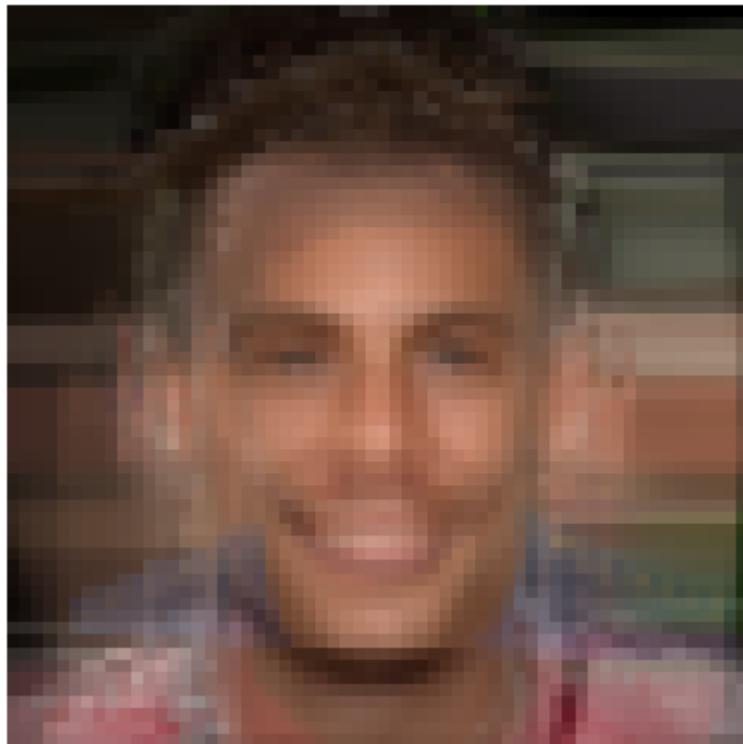
# Generate the average face
average_face = get_average_face(image_paths)

# Display the average face
if average_face is not None:
    plt.imshow(cv2.cvtColor(average_face, cv2.COLOR_BGR2RGB))
    plt.axis('off')
    plt.show()

# Save the average face to a file
cv2.imwrite(f"average_face_{'_'.join(words)}.jpg", average_face)
else:
    print(f"No faces found for the words: {', '.join(words)}")
```

Players described by the words: ['callum robinson', 'jordan henderson', 'tammie abraham', 'anthony martial', 'ken sema', 'james milner', 'lucas torreira', 'theo walcott']

Image paths for these players: ['https://cdn.sofifa.net/players/211/300/18\_60.png', 'https://cdn.sofifa.net/players/206/518/18\_60.png', 'https://cdn.sofifa.net/players/138/412/18\_60.png', 'https://cdn.sofifa.net/players/183/711/18\_60.png', 'https://cdn.sofifa.net/players/213/750/18\_60.png', 'https://cdn.sofifa.net/players/223/959/18\_60.png', 'https://cdn.sofifa.net/players/231/352/18\_60.png', 'https://cdn.sofifa.net/players/164/859/18\_60.png']



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

The below is an attempt to display the results and attribute the VAD scores to words used

In [ ]: `import pandas as pd`

```
# Load the VAD scores
vad_df = pd.read_csv('/Users/alexdoyle/Desktop/Warriner_rescale-Copy1.csv')
print(vad_df.head())
```

In [ ]: `import pandas as pd`

```
# Load the main data from CSV
main_df = pd.read_csv('/Users/alexdoyle/Desktop/df_contexts_cleaned.csv')
```

```
# Load the VAD scores
vad_df = pd.read_csv('/Users/alexdoyle/Desktop/Warriner_rescale-Copy1.csv')

# Check the initial data
print(main_df.head())
print(vad_df.head())
```

```
In [ ]: # Assuming your main data has a column with descriptive words separated by
# We'll split these into separate rows for easier merging

# Explode the descriptive words into separate rows
expanded_df = main_df.assign(descriptive_words=main_df['descriptive words'])
expanded_df['descriptive words'] = expanded_df['descriptive words'].str.lower()

# Print expanded DataFrame
print(expanded_df.head())
```

```
In [ ]: expanded_df
```

```
In [ ]: # Merge the expanded main data with the VAD scores
merged_df = expanded_df.merge(vad_df, left_on='descriptive words', right_on='word')
print(merged_df.head())
```

```
In [ ]: import seaborn as sns
import matplotlib.pyplot as plt

# Example of plotting Valence against skin tone
sns.boxplot(x='tone label', y='Valence', data=merged_df)
plt.title('Valence by Skin Tone')
plt.xlabel('Skin Tone')
plt.ylabel('Valence Score')
plt.show()
```

```
In [ ]: # Assuming 'vad_df' is your VAD DataFrame and 'expanded_df' is your main data
print("Columns in main DataFrame:", expanded_df.columns)
print("Columns in VAD DataFrame:", vad_df.columns)
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

In [ ]:

The code below is ensuring that my list of descriptive words has no duplicates

```
In [112]: # List of words with potential duplicates
words_list = [
    'strong', 'quick', 'fast', 'tall', 'muscular', 'agile', 'fit', 'powerful',
    'robust', 'flexible', 'athletic', 'energetic', 'tough', 'enduring', 'stable',
    'hefty', 'massive', 'bulky', 'slim', 'lean', 'durable', 'solid', 'physique',
    'explosive', 'forceful', 'vigorous', 'power', 'speed', 'strength', 'dexterity',
    'stability', 'stamina', 'energy', 'pace', 'mobility', 'grit', 'muscle',
    'heavy', 'light', 'weight', 'capacity', 'broad', 'narrow', 'wide', 'hardy',
    'steady', 'resistant', 'active', 'lively', 'spry', 'fit', 'healthy', 'handy',
    'limber', 'supple', 'fleet', 'swift', 'nippy', 'acceleration', 'endurance',
    'beast', 'stumbled', 'stumble', 'stuttered', 'stutter', 'predator', 'predatory',
    'bursting', 'burst', 'belter', 'header', 'strike', 'striking', 'shrugged',
    'clawed', 'prowled', 'leap', 'leapt', 'sprint', 'sprinted', 'ferocious',
    'intimidating', 'tenacious', 'determined', 'daring', 'resilient', 'tall',
    'fierce', 'vitality', 'sturdy', 'burst', 'well built', 'agile', 'resilient',
    'power', 'strong', 'weight', 'pace', 'brawny', 'explosive', 'quick', 'physical',
    'vigorous', 'fast', 'stamina', 'fit', 'muscle', 'burly', 'speed', 'nimble',
    'energy', 'endurance', 'solid', 'athleticism', 'muscular', 'flexible',
]
# Remove duplicates by converting the list to a set and back to a list
unique_words_list = list(set(words_list))

# Print the list of unique words
print(unique_words_list)

# Optionally sort the list if needed
sorted_unique_words_list = sorted(unique_words_list)
print(sorted_unique_words_list)
```

```
['endurance', 'muscle', 'massive', 'body', 'ferocious', 'burly', 'athletic', 'sprinted', 'sprint', 'height', 'belter', 'header', 'tough', 'nimble', 'vigorous', 'burst', 'vitality', 'well built', 'hefty', 'energetic', 'bursting', 'beast', 'force', 'dexterity', 'speedy', 'stumbled', 'stamina', 'steady', 'vital', 'flexible', 'fleet', 'clawed', 'striking', 'prowled', 'fit', 'bulky', 'forceful', 'build', 'stuttered', 'wide', 'predatory', 'pounced', 'limber', 'robust', 'supple', 'explosive', 'hardy', 'lean', 'pounce', 'dynamic', 'grit', 'shrugged', 'solid', 'leap', 'brisk', 'tenacious', 'leapt', 'capacity', 'quick', 'physicality', 'firm', 'strength', 'brawny', 'power', 'determined', 'light', 'sprightly', 'enduring', 'aggressive', 'mobility', 'stability', 'physical', 'athleticism', 'stumble', 'acceleration', 'aggression', 'balance', 'predator', 'tall', 'sturdy', 'weight', 'spry', 'heavy', 'resistant', 'dominant', 'resilient', 'healthy', 'lively', 'agile', 'slim', 'stature', 'stutter', 'swift', 'daring', 'nippy', 'fast', 'intense', 'muscular', 'powerful', 'coordination', 'fierce', 'active', 'intimidating', 'energy', 'narrow', 'hale', 'pace', 'broad', 'durable', 'speed', 'bulk', 'ferocity', 'strike', 'rapid', 'strong']  
['acceleration', 'active', 'aggression', 'aggressive', 'agile', 'athletic', 'athleticism', 'balance', 'beast', 'belter', 'body', 'brawny', 'brisk', 'broad', 'build', 'bulk', 'bulky', 'burly', 'burst', 'bursting', 'capacity', 'clawed', 'coordination', 'daring', 'determined', 'dexterity', 'dominant', 'durable', 'dynamic', 'endurance', 'enduring', 'energetic', 'energy', 'explosive', 'fast', 'ferocious', 'ferocity', 'fierce', 'firm', 'fit', 'fleet', 'flexible', 'force', 'forceful', 'grit', 'hale', 'hardy', 'header', 'health', 'heavy', 'hefty', 'height', 'intense', 'intimidating', 'lean', 'leap', 'leapt', 'light', 'limber', 'lively', 'massive', 'mobility', 'muscle', 'muscular', 'narrow', 'nimble', 'nippy', 'pace', 'physical', 'physicality', 'ponce', 'pounced', 'power', 'powerful', 'predator', 'predatory', 'prowled', 'quick', 'rapid', 'resilient', 'resistant', 'robust', 'shrugged', 'slim', 'solid', 'speed', 'speedy', 'sprightly', 'sprint', 'sprinted', 'spry', 'stability', 'stamina', 'stature', 'steady', 'strength', 'strike', 'striking', 'strong', 'stumble', 'stumbled', 'sturdy', 'stutter', 'stuttered', 'supple', 'swift', 'tall', 'tenacious', 'tough', 'vigorous', 'vital', 'vitality', 'weight', 'well built', 'wide']
```

In [113]: `len(sorted_unique_words_list)`

Out[113]: 115

In [114]: `# List of words with potential duplicates`

```
words_list = [  
    'intelligent', 'clever', 'smart', 'thoughtful', 'analytical', 'wise', 'sharp', 'bright', 'brilliant', 'canny', 'quick-witted', 'perceptive', 'perspicacious', 'inventive', 'creative', 'judicious', 'prudent', 'sensible', 'logical', 'planning', 'calculating', 'visionary', 'insightful', 'intuitive', 'resourceful', 'aware', 'conscious', 'focused', 'attentive', 'concentrated', 'determined', 'dynamic', 'adaptive', 'flexible', 'versatile', 'motivated', 'disciplined', 'leader', 'leadership', 'commanding', 'directive', 'authoritative', 'influential', 'purposeful', 'ambitious', 'enterprising', 'proactive', 'initiative', 'original', 'innovative', 'original', 'imaginative', 'inspiring', 'motivational', 'communicative', 'articulate', 'eloquent', 'persuasive', 'diplomatic', 'tactful', 'negotiable', 'strategic', 'adaptive', 'leader', 'silky', 'mesmerising', 'creative', 'discerning', 'inventive', 'clever', 'manage', 'problem-solver', 'alert', 'analytical', 'quick-witted', 'skill', 'communicate', 'passion', 'technique', 'tidy', 'excite', 'resourceful', 'intuitive', 'aware', 'talent', 'perceptive', 'cute', 'excitement', 'lead', 'scan', 'compose', 'intelligent', 'canny', 'floated', 'tiptoeing', 'gamechanger', 'finished', 'finish', 'perfectly', 'good ball', 'timely pass', 'lovely ball', 'wasted opportunity', 'curl', 'under pressure', 'unselfish', 'movement', 'focus', 'control', 'anticipation', 'pullback', 'rolls', 'confident', 'reliable', 'technique', 'brave', 'neat', ]  
  
# Remove duplicates by converting the list to a set and back to a list  
unique_words_list = list(set(words_list))
```

```
# Print the list of unique words
print(unique_words_list)

# Optionally sort the list if needed
sorted_unique_words_list = sorted(unique_words_list)
print(sorted_unique_words_list)

['decisive', 'motivational', 'confident', 'pullback', 'logical', 'finished', 'strategic', 'scan', 'excitement', 'observant', 'skill', 'excite', 'creative', 'innovative', 'perceptive', 'focus', 'planning', 'timely pass', 'persuasive', 'negotiator', 'shrewd', 'timed', 'reliable', 'tactful', 'intuitive', 'lovely ball', 'gamechanger', 'intelligent', 'organized', 'brightfloated', 'curl', 'tidy', 'diplomatic', 'lead', 'flexible', 'tiptoeing', 'industrious', 'leadership', 'ambitious', 'smart', 'control', 'calculating', 'movement', 'problem-solver', 'communicate', 'judicious', 'tactical', 'bright', 'eloquent', 'feet', 'discerning', 'focused', 'original', 'dynamic', 'pride', 'finish', 'neatly', 'desire', 'sensible', 'tenacious', 'communicative', 'insightful', 'wise', 'versatile', 'manage', 'plan', 'linking', 'determined', 'brilliant', 'anticipation', 'purposeful', 'perfect', 'conscious', 'mesmerising', 'brave', 'resolute', 'technique', 'resourceful', 'alert', 'vision', 'lovely shape', 'clever', 'influential', 'commanding', 'ingenious', 'inventive', 'perspicacious', 'motivated', 'analytical', 'mediatorstrategic', 'quick-witted', 'under pressure', 'initiative', 'attentive', 'inspiring', 'sharp', 'imaginative', 'astute', 'aware', 'cute', 'concentrated', 'great ball', 'directive', 'passion', 'proactive', 'adaptive', 'leader', 'technical', 'unselfish', 'silky', 'articulate', 'driven', 'disciplined', 'thoughtful', 'coordinated', 'good ball', 'wasted opportunity', 'enterprising', 'expressive', 'link up', 'keen', 'perfectly', 'touch', 'talent', 'canny', 'prudent', 'visionary', 'strategy', 'compose', 'authoritative', 'rolls']

['adaptive', 'alert', 'ambitious', 'analytical', 'anticipation', 'articulate', 'astute', 'attentive', 'authoritative', 'aware', 'brave', 'bright', 'brightfloated', 'brilliant', 'calculating', 'canny', 'clever', 'commanding', 'communicate', 'communicative', 'compose', 'concentrated', 'confident', 'conscious', 'control', 'coordinated', 'creative', 'curl', 'cute', 'decisive', 'desire', 'determined', 'diplomatic', 'directive', 'discerning', 'disciplined', 'driven', 'dynamic', 'eloquent', 'enterprising', 'excite', 'excitemen', 'expressive', 'feet', 'finish', 'finished', 'flexible', 'focus', 'focus ed', 'gamechanger', 'good ball', 'great ball', 'imaginative', 'industriou s', 'influential', 'ingenious', 'initiative', 'innovative', 'insightful', 'inspiring', 'intelligent', 'intuitive', 'inventive', 'judicious', 'keen', 'lead', 'leader', 'leadership', 'link up', 'linking', 'logical', 'lovely ba ll', 'lovely shape', 'manage', 'mediatorstrategic', 'mesmerising', 'motivated', 'motivational', 'movement', 'neatly', 'negotiator', 'observant', 'organized', 'original', 'passion', 'perceptive', 'perfect', 'perfectly', 'perspicacious', 'persuasive', 'plan', 'planning', 'pride', 'proactive', 'problem-solver', 'prudent', 'pullback', 'purposeful', 'quick-witted', 'reliable', 'resolute', 'resourceful', 'rolls', 'scan', 'sensible', 'sharp', 'shrewd', 'silky', 'skill', 'smart', 'strategic', 'strategy', 'tactful', 'tactical', 'talent', 'technical', 'technique', 'tenacious', 'thoughtful', 'tidy', 'timed', 'timely pass', 'tiptoeing', 'touch', 'under pressure', 'unselfish', 'versatile', 'vision', 'visionary', 'wasted opportunity', 'wise']
```

In [115]: len(sorted\_unique\_words\_list)

Out[115]: 131

In [ ]:

In [ ]:

In [ ]:

```
# create a spacy nlp objects  
# get the vector -
```

```
In [75]: text = http://www.mynameis.txt
```

In [74]: `text.vector`

```
Out[74]: array([ 4.43347991e-02,  3.69554579e-01, -3.65577877e-01,  8.15934390e-02,
   -7.40369201e-01, -4.67690527e-01, -1.55644804e-01, -6.62001595e-02,
   3.61878961e-01,  2.61597279e-02, -8.54376018e-01, -1.54692352e-01,
   6.36475068e-03, -1.60838082e-01, -3.77198398e-01,  2.49914810e-01,
   8.34477901e-01,  2.47324437e-01,  1.09429779e-02,  3.71164680e-02,
  -1.81302279e-01,  8.14927220e-01,  5.19608378e-01,  3.76807779e-01,
  1.13761231e-01,  4.56387609e-01,  1.00605178e+00, -4.22791094e-01,
  1.11593224e-01,  3.75818789e-01, -1.60430863e-01, -1.18916228e-01,
  -3.34408104e-01, -5.91866374e-01, -7.19094425e-02, -2.93491840e-01,
  -4.00043204e-02,  1.27167270e-01,  9.43757594e-02, -7.81842507e-03,
  -1.51136965e-01,  4.64446563e-03, -4.43247378e-01,  1.55954495e-01,
  -9.35475230e-01, -4.55588490e-01,  1.42921537e-01,  5.85912645e-01,
  -3.30756977e-02, -3.67043316e-01,  2.23920465e-01, -5.11161685e-01,
  -1.58291623e-01, -7.31517971e-01,  1.37946203e-01,  3.41797739e-01,
  6.57871366e-02,  1.16415523e-01,  6.71230435e-01, -1.64357603e-01,
  -2.43904263e-01, -7.19859719e-01,  9.85307544e-02, -1.83680326e-01,
  2.54221052e-01, -3.54287952e-01, -4.28178310e-01,  2.01481149e-01,
  6.01511180e-01,  7.48032928e-02,  6.13209784e-01, -1.66438773e-01,
  1.19335733e-01, -8.14892173e-01, -2.13663816e-01,  4.75782245e-01,
  -4.92138974e-02, -2.06017926e-01,  1.96190760e-01, -2.46845871e-01,
  -1.35785816e-02, -5.99628091e-01,  5.75006843e-01, -1.46245316e-01,
  -1.26129395e-04, -4.78758752e-01,  3.13951910e-01, -1.73229873e-02,
  1.65536791e-01, -4.28370237e-01,  1.81054980e-01,  3.54494691e-01,
  -6.26748800e-02,  5.17167211e-01,  4.32677686e-01, -1.45898044e-01],
 dtype=float32)
```

```
In [75]: # turn into a row on pandas dataframe
```

```
In [78]: main_df
```

Out[78]:

	Player	Context	tone label	descriptive words	physical count	cognitive count	physical words	cognitive words	
0	harry kane	theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in	CE	touch	1.0	0.0	touch	NaN	\$
1	alex mccarthy	target the pace on the ball means he cant react alex mccarthy it goes through his legs hes tiptoeing his way around	CF	tiptoeing	1.0	0.0	tiptoeing	NaN	
2	sergio aguero	expect aguero to do much better walker in towards aguero sergio aguero doing what comes so naturally to him strong ankle gets	CE	strong	1.0	0.0	strong	NaN	
3	danny ings	convincing header boufal up against schneiderlin its in from ings danny ings with the equalizer for southampton at long last it game	CF	header	1.0	0.0	header	NaN	
4	richarlison	wardprowse very few better at striking the dead ball sidibé richarlison brilliantly finished richarlison his first goal in eight	CF	finish	0.0	1.0	NaN	finish	

	Player	Context	tone label	descriptive words	physical count	cognitive count	physical words	cognitive words	
premier league									
398	wayne hennessey	does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda	CE	big strong aggressive	3.0	0.0	big strong aggressive	NaN	
399	wayne hennessey	hamšík tidy turn but saved from robert maks effort by wayne hennessey terrific football from the slovakians once again manoeuvre the ball	CE	terrific football	0.0	1.0	NaN	terrific football	
400	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
401	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
402	NaN	NaN	NaN	NaN	344.0	190.0	NaN	NaN	

403 rows × 12 columns

In [81]: `commentary = []`

```

for i in main_df['Context']:
    i = str(i)
    commentary.append(nlp(i).vector)

commentary = pd.DataFrame(commentary)

```

In [82]: `commentary`

Out[82]:

	0	1	2	3	4	5	6	7
0	0.032889	-0.014101	0.087074	-0.074806	0.087158	0.292692	-0.331217	0.114890
1	0.524425	-0.066856	0.188722	-0.167777	0.311309	0.026329	0.048551	0.077858
2	0.164791	-0.145766	-0.033419	-0.171619	0.181483	0.374052	0.315914	-0.017530
3	0.255077	-0.115331	0.323014	-0.258379	0.392981	0.008243	-0.232091	0.051816
4	0.372483	0.168694	0.306175	-0.126565	-0.024118	-0.126268	-0.391011	0.115768
...	...	...	...	...	...	...	...	...
398	0.217630	-0.081674	0.348199	-0.292235	-0.067460	0.117660	0.163864	-0.231157
399	0.288941	0.177326	0.135885	-0.129178	0.205182	0.125363	-0.140165	0.225419
400	-1.112455	0.776074	0.272335	-0.392310	0.117406	-0.304407	0.007586	-0.763419
401	-1.112455	0.776074	0.272335	-0.392310	0.117406	-0.304407	0.007586	-0.763419
402	-1.112455	0.776074	0.272335	-0.392310	0.117406	-0.304407	0.007586	-0.763419

403 rows × 96 columns

In [85]:

```
cat = []
for i in main_df['tone_label']:
    if i == 'CG' or i == 'CH' or i == 'CI' or i == 'CJ':
        cat.append(0)
    elif i == 'CC' or i == 'CD' or i == 'CE' or i == 'CF':
        cat.append(1)
```

In [86]:

cat













```
1,
1,
1,
0,
0,
0,
1,
1,
1,
1,
1,
1,
0,
0,
0,
1,
1]
```

In [87]: `commentary['category']=cat`

```
-----  

ValueError                                                 Traceback (most recent call last)
/var/folders/rb/2l3j3fzj6mng0r5wn3mnhtfm0000gn/T/ipykernel_78534/117992066
7.py in <module>
----> 1 commentary['category']=cat

~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py in __setitem__
(self, key, value)
    3653         else:
    3654             # set column
-> 3655                 self._set_item(key, value)
    3656
    3657     def _setitem_slice(self, key: slice, value):
```

```
~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py in __setitem__
(self, key, value)
    3830         ensure homogeneity.
    3831         """
-> 3832         value = self._sanitize_column(value)
    3833
    3834         if (
```

```
~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py in _sanitize_column
(self, value)
    4536
    4537         if is_list_like(value):
-> 4538             com.require_length_match(value, self.index)
    4539             return sanitize_array(value, self.index, copy=True, allow_2
d=True)
    4540
```

```
~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/common.py in require_length_match
(data, index)
    555     """
    556     if len(data) != len(index):
--> 557         raise ValueError(
    558             "Length of values "
    559             f"({len(data)}) "
```

```
ValueError: Length of values (400) does not match length of index (403)
```

In [ ]: `# x features document embeddings  
# y is category`

```
In [ ]: # what is document embedding??
```

```
In [ ]:
```

```
In [ ]: # logistic regression - is categorical  
# add new column on binary skin tone
```

```
In [ ]:
```

```
In [ ]:
```

The above code is a demonstration of how this works, it has failed because my dataframe has players without skin tones assigned to them in it. I now need to remove these players before attempting to do this again. My next attempt at a binary logistic regression will include more data - I will use all 800 players identified (rather than only the handful of these that have been used for the visualisation part of the project that required only players that had been described by a cognitive or physical trait).

```
In [ ]: ls
```

```
In [88]: df_lr = pd.read_csv('/Users/alexdoyle/Desktop/excel_words_identified/df_cont
```

```
In [89]: df_lr
```

Out[89]:

	Player	Context	tone label	physical words	cognitive words	physical count	cognitive count	descriptive words
0	joão cancelo	oh great skill from bernardeschi for dybala spilled big chance joão cancelo with an equalizer for juventus back from injury to save	CF	NaN	skill	0	1	skill
1	joão cancelo	and they might just do that ronaldo in for bernardeschi joão cancelo went over on the challenge from lulić hes given a	CF	NaN	NaN	0	0	NaN
2	roberto	by the offside flag a few minutes ago this time roberto inglese is off the mark in the match giving parma	CF	NaN	NaN	0	0	NaN
3	harry kane	theres no chance for the keeper theyre stretched defensively here harry kane just makes that point only needed a touch floated in	CE	touch	NaN	1	0	touch
4	tammy abraham	by šofranac he might get a tap in this time tammy abraham opens his england account it is seven with seven minutes	CD	NaN	NaN	0	0	NaN
...	...	...	...	...	...	...	...	...
1898	david brooks	really good cross and its good contact as	CG	touch	NaN	1	0	touch

	Player	Context	tone label	physical words	cognitive words	physical count	cognitive count	descriptive words
		well from david brooks brooks now thats a lovely touch to take him beyond						
1899	harry wilson	bit too high for tyler roberts nice first touch by harry wilson rusnák is full of running here for slovakia time to	CG	touch	NaN	1	0	touch
1900	wayne hennessey	wales have really been cut apart its a good run wayne hennessey came out of his goal so quickly and shut down	CE	NaN	NaN	0	0	NaN
1901	wayne hennessey	does it so well tries to just lift it over wayne hennessey big strong aggressive arm came out towards him rusnák duda	CE	strong aggressive	NaN	2	0	strong aggressive
1902	wayne hennessey	hamšík tidy turn but saved from robert maks effort by wayne hennessey terrific football from the slovakians once again manoeuvre the ball	CE	NaN	tidy	0	1	tidy

1903 rows x 8 columns

In [90]: commentary = []

```

for i in df_lr['Context']:
    i = str(i)
    commentary.append(nlp(i).vector)

```

```
commentary = pd.DataFrame(commentary)
```

In [91]: commentary

Out[91]:

	0	1	2	3	4	5	6
0	0.403887	-0.137225	-0.001783	-0.166812	0.272987	0.379075	-0.134555
1	0.534635	-0.171626	0.236206	0.026364	0.340959	0.218188	-0.103279
2	0.575738	0.249775	0.035228	-0.082993	0.308800	-0.283987	-0.308710
3	0.032889	-0.014101	0.087074	-0.074806	0.087158	0.292692	-0.331217
4	0.438001	0.303282	0.040907	-0.135130	0.198826	-0.029900	-0.130495
...	...	...	...	...	...	...	...
1898	0.299654	-0.095567	0.165226	-0.246856	-0.077081	0.137354	-0.126880
1899	0.064603	0.107657	0.410194	-0.242374	0.180408	0.353825	-0.041246
1900	-0.130037	0.070923	0.168779	-0.198302	-0.252212	0.226732	-0.010202
1901	0.217630	-0.081674	0.348199	-0.292235	-0.067460	0.117660	0.163864
1902	0.288941	0.177326	0.135885	-0.129178	0.205182	0.125363	-0.140165

1903 rows × 96 columns

```
In [108... st = []
for i in df_lr['tone_label']:
    if i == 'CF' or i == 'CG' or i == 'CI':
        st.append(0)
    elif i == 'CC' or i == 'CD' or i == 'CE':
        st.append(1)
```

In [109... commentary['skin tone']=cat

In [110... commentary

Out[110]:

	0	1	2	3	4	5	6
0	0.403887	-0.137225	-0.001783	-0.166812	0.272987	0.379075	-0.134555
1	0.534635	-0.171626	0.236206	0.026364	0.340959	0.218188	-0.103279
2	0.575738	0.249775	0.035228	-0.082993	0.308800	-0.283987	-0.308710
3	0.032889	-0.014101	0.087074	-0.074806	0.087158	0.292692	-0.331217
4	0.438001	0.303282	0.040907	-0.135130	0.198826	-0.029900	-0.130495
...	...	...	...	...	...	...	...
1898	0.299654	-0.095567	0.165226	-0.246856	-0.077081	0.137354	-0.126880
1899	0.064603	0.107657	0.410194	-0.242374	0.180408	0.353825	-0.041246
1900	-0.130037	0.070923	0.168779	-0.198302	-0.252212	0.226732	-0.010202
1901	0.217630	-0.081674	0.348199	-0.292235	-0.067460	0.117660	0.163864
1902	0.288941	0.177326	0.135885	-0.129178	0.205182	0.125363	-0.140165

1903 rows × 98 columns

In [ ]:

Before continuing I will check to see if I have an equal split of the data regarding lighter skin tone and darker skin tone players

```
In [111]: skin_tone_counts = commentary['skin tone'].value_counts()

print("Number of players with darker skin tone (0):", skin_tone_counts.get(0))
print("Number of players with lighter skin tone (1):", skin_tone_counts.get(1))

Number of players with darker skin tone (0): 986
Number of players with lighter skin tone (1): 917
```

In [ ]:

```
In [104]: X = commentary.drop('skin tone', axis=1)
y = commentary['skin tone']
```

```
In [105]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
In [106]: from sklearn.linear_model import LogisticRegression
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)
```

Out[106]: LogisticRegression(max\_iter=1000)

```
In [107]: from sklearn.metrics import accuracy_score, classification_report
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)
print(f"Accuracy: {accuracy}")
print(report)
```

Accuracy: 0.5984251968503937

Classification Report:

	precision	recall	f1-score	support
0	0.57	0.71	0.63	185
1	0.64	0.49	0.56	196
accuracy			0.60	381
macro avg	0.61	0.60	0.60	381
weighted avg	0.61	0.60	0.59	381

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

Chi-square test

```
In [130]: import numpy as np
from scipy.stats import chi2_contingency

# Observed frequencies
observed = np.array([[100, 7], [63, 8]])

# Perform chi-square test
chi2, p, dof, expected = chi2_contingency(observed)

# Print results
print("Chi-square statistic:", chi2)
print("p-value:", p)
print("Degrees of freedom:", dof)
print("Expected frequencies:\n", expected)
```

Chi-square statistic: 0.6985967176258632

p-value: 0.4032556195474325

Degrees of freedom: 1

Expected frequencies:

```
[[97.98314607  9.01685393]
 [65.01685393  5.98314607]]
```

```
In [122]: total = 178
```

```
In [128]: ob_tp_rw = (107/178)*(163/178)*total
ob_tp_rw
```

Out[128]: 97.98314606741573

In [ ]:

```
In [129]: import numpy as np
from scipy.stats import chi2_contingency

# Observed frequencies
observed = np.array([[7, 0], [22, 3], [34, 5], [66, 5], [29, 1], [5, 1]])

# Perform chi-square test
```

```
chi2, p, dof, expected = chi2_contingency(observed)

# Print results
print("Chi-square statistic:", chi2)
print("p-value:", p)
print("Degrees of freedom:", dof)
print("Expected frequencies:\n", expected)
```

```
Chi-square statistic: 3.7462717539572283
p-value: 0.5864933471199948
Degrees of freedom: 5
Expected frequencies:
[[ 6.41011236  0.58988764]
 [22.89325843  2.10674157]
 [35.71348315  3.28651685]
 [65.01685393  5.98314607]
 [27.47191011  2.52808989]
 [ 5.49438202  0.50561798]]
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: