In [17]:

```python
import nltk
nltk.download('punkt')
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
from nltk import stem
stemmer = stem.PorterStemmer()
from nltk import word_tokenize
nltk.download('stopwords')
from nltk.corpus import stopwords
stops = set(stopwords.words('english'))
import string
punct = list(string.punctuation)
from collections import Counter
import requests
import pandas as pd
import seaborn as sns
sns.set()
import matplotlib.pyplot as plt
!pip install PRAW
import numpy as np
import praw
import datetime
```

```
[nltk_data] Downloading package punkt to /Users/alexdoyle/nltk_dat
a...
[nltk_data]    Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]      /Users/alexdoyle/nltk_data...
[nltk_data]    Package stopwords is already up-to-date!

Requirement already satisfied: PRAW in /Users/alexdoyle/opt/anacond
a3/lib/python3.9/site-packages (7.7.0)
Requirement already satisfied: prawcore<3,>=2.1 in /Users/alexdoyl
e/opt/anaconda3/lib/python3.9/site-packages (from PRAW) (2.3.0)
Requirement already satisfied: update-checker>=0.18 in /Users/alexd
oyle/opt/anaconda3/lib/python3.9/site-packages (from PRAW) (0.18.0)
Requirement already satisfied: websocket-client>=0.54.0 in /Users/a
lexdoyle/opt/anaconda3/lib/python3.9/site-packages (from PRAW) (0.5
8.0)
Requirement already satisfied: requests<3.0,>=2.6.0 in /Users/alexd
oyle/opt/anaconda3/lib/python3.9/site-packages (from prawcore<3,>=
2.1->PRAW) (2.28.1)
Requirement already satisfied: six in /Users/alexdoyle/opt/anaconda
```

In [18]:

```python
df = pd.read_csv('reddit_comments.csv')
df
```

Out[18]:

	comment_id	post_id	author	body	score	subreddit	
							foo
							opinic
0	jc7soq4	11razey	thathock	If the player taking a corner doesn't clear th...	463	r/PremierLeague	you
							that w
							foo
							opinic
1	jc7jmbx	11razey	mesenanch	I'm being triggered by most these comments. Th...	169	r/PremierLeague	you
							that w
							foo
							opinic
2	jc7xyhn	11razey	Digess	the short pass corners are fucking annoying an	306	r/PremierLeague	you

In [9]:

```python
# creating an empty list to store all the comments as individual words
# (now the data has been cleaned) ready for sentiment analysis

words = []


for text in df['processed_body']:
    if isinstance(text, str):  # check if text is a string
        words += text.split()




# this is an existing excel file with VAD scores for corresponding words,
# in order to run this code this excel file exists within my jupyter file
vad = pd.read_excel('vad.xlsx', index_col = 0)

text = []
vad_score = []

for i in words:
    if i in vad.index:
        vad_score.append(vad.loc[i])
        # append the set of values that are in the index for that word
        text.append(i)
        # and then take the word and put it into the text list
    else:
        pass
    # and if a word doesnt satisfy this criteria, ignore it

# this creates a dataframe to store the comments and their VAD score alongside

vad_df = pd.DataFrame(vad_score, index = text)

vad_df
```

Out[9]:

	valence	arousal	dominance
player	0.476636	0.473413	0.587379
corner	0.570093	0.360206	0.582524
clear	0.672897	0.308748	0.621359
first	0.747664	0.596913	0.747573
defender	0.738318	0.617496	0.747573
...	...	...	...
match	0.719626	0.504288	0.714806
discussion	0.588785	0.504288	0.572816
fire	0.495327	0.833619	0.174757
put	0.485981	0.380789	0.632888

In [ ]:

In [11]:

```
pip install pingouin
```

```
Collecting pingouin
  Downloading pingouin-0.5.3-py3-none-any.whl (198 kB)
                                                    198.6/198.6 kB 1.7 MB/
s eta 0:00:0000:0100:01
Requirement already satisfied: seaborn>=0.11 in /Users/alexdoyle/op
t/anaconda3/lib/python3.9/site-packages (from pingouin) (0.11.2)
Collecting pandas-flavor>=0.2.0
  Downloading pandas_flavor-0.5.0-py3-none-any.whl (7.1 kB)
Requirement already satisfied: tabulate in /Users/alexdoyle/opt/ana
conda3/lib/python3.9/site-packages (from pingouin) (0.8.10)
Requirement already satisfied: scikit-learn in /Users/alexdoyle/op
t/anaconda3/lib/python3.9/site-packages (from pingouin) (1.0.2)
Requirement already satisfied: statsmodels>=0.13 in /Users/alexdoyl
e/opt/anaconda3/lib/python3.9/site-packages (from pingouin) (0.13.
2)
Collecting outdated
  Downloading outdated-0.2.2-py2.py3-none-any.whl (7.5 kB)
Requirement already satisfied: pandas>=1.0 in /Users/alexdoyle/opt/
anaconda3/lib/python3.9/site-packages (from pingouin) (1.4.4)
```

In [13]:

```python
import pandas as pd
import pingouin as pg
import numpy as np
```

In [14]:

```python
# Perform the ANOVA test
aov = pg.anova(dv='valence', between='group', data=pd.concat([vad.assign(group='vad

# Print the ANOVA results
print(aov)

# Extracting the F-value and p-value from the ANOVA results
f_value = aov['F'][0]
p_value = aov['p-unc'][0]

# Comparing the p-value to the significance level to determine statistical significa
if p_value < 0.05:
    print("The median valence scores are significantly different between the two gr
else:
    print("The median valence scores are not significantly different between the tw
```

```
   Source  ddof1  ddof2           F         p-unc       np2
0   group      1  15819  475.968531  5.467838e-104  0.02921
The median valence scores are significantly different between the two
groups.
```

In [15]:

```python
# Perform the ANOVA test
aov = pg.anova(dv='arousal', between='group', data=pd.concat([vad.assign(group='vac
# Print the ANOVA results
print(aov)

# Extracting the F-value and p-value from the ANOVA results
f_value = aov['F'][0]
p_value = aov['p-unc'][0]

# Comparing the p-value to the significance level to determine statistical signific
if p_value < 0.05:
    print("The median arousal scores are significantly different between the two gr
else:
    print("The median arousal scores are not significantly different between the tw
```

```
   Source  ddof1  ddof2         F      p-unc        np2
0   group      1  15819  3.840475  0.050047  0.000243
The median arousal scores are not significantly different between the
two groups.
```

In [16]:

```python
# Perform the ANOVA test
aov = pg.anova(dv='dominance', between='group', data=pd.concat([vad.assign(group='
# Print the ANOVA results
print(aov)

# Extracting the F-value and p-value from the ANOVA results
f_value = aov['F'][0]
p_value = aov['p-unc'][0]

# Comparing the p-value to the significance level to determine statistical signifi
if p_value < 0.05:
    print("The median dominance scores are significantly different between the two
else:
    print("The median dominance scores are not significantly different between the
```

```
   Source  ddof1  ddof2           F         p-unc        np2
0   group      1  15819  825.679608  4.813557e-177  0.049606
The median dominance scores are significantly different between the t
wo groups.
```

In [ ]:

In [24]:

```python
import random
```

In [25]:

```python
df['hashed_comment_id'] = [hash(i) for i in df['comment_id']]
df['hashed_post_id'] = [hash(i) for i in df['post_id']]
df['hashed_author'] = [hash(i) for i in df['author']]

scrambled_comments = [i.split(' ') for i in df['body']]

for i in scrambled_comments:
    random.shuffle(i)

df['scrambled_comments'] = scrambled_comments
```

In [26]:

```python
df
```

Out[26]:

	comment_id	post_id	author	body	score	subreddit	
0	jc7soq4	11razey	thathock	If the player taking a corner doesn't clear th...	463	r/PremierLeague	foc opinic you that w
1	jc7jmbx	11razey	mesenanch	I'm being triggered by most these comments. Th...	169	r/PremierLeague	foc opinic you that w
2	jc7xyhn	11razey	Digess	the short pass corners are fucking annoying an	306	r/PremierLeague	foc opinic you