Name: Chirag Rao KV

reg:240962180

PC 66

LAB 9

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('/home/chirag/Documents/DAV/sentiment.csv')

data
```

```
      ID                                                Text
Sentiment  \
0      0   Enjoying a beautiful day at the park!         ...   Positive

1      1   Traffic was terrible this morning.            ...   Negative

2      2   Just finished an amazing workout! □           ...   Positive

3      3   Excited about the upcoming weekend getaway!   ...   Positive

4      4   Trying out a new recipe for dinner tonight.   ...   Neutral

..   ...                                                ...           .
..
727  728  Collaborating on a science project that receiv...
Happy
728  729  Attending a surprise birthday party organized ...
Happy
729  730  Successfully fundraising for a school charity ...
Happy
730  731  Participating in a multicultural festival, cel...
Happy
731  732  Organizing a virtual talent show during challe...
Happy

             Timestamp                                    User
Platform  \
0    15-01-2023 12:30                          User123
Twitter
1    15-01-2023 08:45                          CommuterX
```

```
Twitter
2    15-01-2023 15:45                            FitnessFan
Instagram
3    15-01-2023 18:20                            AdventureX
Facebook
4    15-01-2023 19:55                            ChefCook
Instagram
..                ...                                   ...
...
727  18-08-2017 18:20       ScienceProjectSuccessHighSchool
Facebook
728  22-06-2018 14:15            BirthdayPartyJoyHighSchool
Instagram
729  05-04-2019 17:30   CharityFundraisingTriumphHighSchool
Twitter
730  29-02-2020 20:45    MulticulturalFestivalJoyHighSchool
Facebook
731  15-11-2020 15:15    VirtualTalentShowSuccessHighSchool
Instagram

                                          Hashtags  Retweets  Likes  \
0        #Nature #Park                                      15     30
1        #Traffic #Morning                                   5     10
2        #Fitness #Workout                                  20     40
3        #Travel #Adventure                                  8     15
4        #Cooking #Food                                     12     25
..                                                ...       ...    ...
727         #ScienceFairWinner #HighSchoolScience           20     39
728    #SurpriseCelebration #HighSchoolFriendship           25     48
729      #CommunityGiving #HighSchoolPhilanthropy           22     42
730         #CulturalCelebration #HighSchoolUnity           21     43
731   #VirtualEntertainment #HighSchoolPositivity           24     47

          Country  Year  Month  Day  Hour
0       USA        2023      1   15    12
1       Canada     2023      1   15     8
2     USA          2023      1   15    15
3       UK         2023      1   15    18
4      Australia   2023      1   15    19
..             ...   ...    ...  ...   ...
727            UK  2017      8   18    18
728           USA  2018      6   22    14
729        Canada  2019      4    5    17
730            UK  2020      2   29    20
731           USA  2020     11   15    15

[732 rows x 14 columns]
```

```
# 1. In the sentiment feature and text, trim white spaces and convert
to lowercase.
```

```
data['Text'] = data['Text'].str.strip().str.lower()
data['Sentiment'] = data['Sentiment'].str.strip().str.lower()
data['Country'] = data['Country'].str.strip()

data
```

```
      ID                                                Text Sentiment
\
0      0              enjoying a beautiful day at the park!  positive

1      1                 traffic was terrible this morning.  negative

2      2                just finished an amazing workout! □  positive

3      3        excited about the upcoming weekend getaway!  positive

4      4        trying out a new recipe for dinner tonight.   neutral

..   ...                                                ...       ...

727  728  collaborating on a science project that receiv...     happy

728  729  attending a surprise birthday party organized ...     happy

729  730  successfully fundraising for a school charity ...     happy

730  731  participating in a multicultural festival, cel...     happy

731  732  organizing a virtual talent show during challe...     happy


            Timestamp                                    User
Platform  \
0    15-01-2023 12:30                          User123
Twitter
1    15-01-2023 08:45                          CommuterX
Twitter
2    15-01-2023 15:45                          FitnessFan
Instagram
3    15-01-2023 18:20                          AdventureX
Facebook
4    15-01-2023 19:55                          ChefCook
Instagram
..                ...                                ...
...
727  18-08-2017 18:20       ScienceProjectSuccessHighSchool
Facebook
728  22-06-2018 14:15            BirthdayPartyJoyHighSchool
Instagram
729  05-04-2019 17:30   CharityFundraisingTriumphHighSchool
Twitter
```

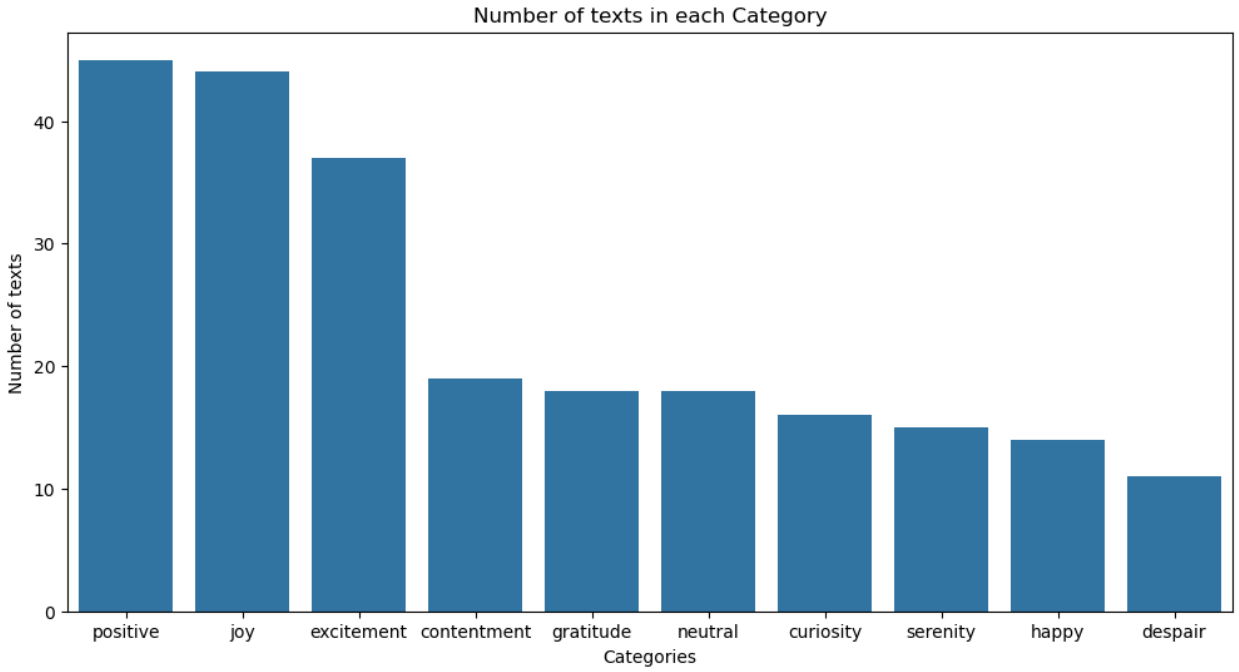
```
730  29-02-2020 20:45    MulticulturalFestivalJoyHighSchool
Facebook
731  15-11-2020 15:15    VirtualTalentShowSuccessHighSchool
Instagram

                                          Hashtags  Retweets  Likes  \
0        #Nature #Park                                      15     30
1        #Traffic #Morning                                   5     10
2        #Fitness #Workout                                  20     40
3        #Travel #Adventure                                  8     15
4        #Cooking #Food                                     12     25
..                                                ...       ...    ...
727         #ScienceFairWinner #HighSchoolScience           20     39
728    #SurpriseCelebration #HighSchoolFriendship           25     48
729      #CommunityGiving #HighSchoolPhilanthropy           22     42
730         #CulturalCelebration #HighSchoolUnity           21     43
731   #VirtualEntertainment #HighSchoolPositivity           24     47

       Country  Year  Month  Day  Hour
0          USA  2023      1   15    12
1       Canada  2023      1   15     8
2          USA  2023      1   15    15
3           UK  2023      1   15    18
4    Australia  2023      1   15    19
..         ...   ...    ...  ...   ...
727         UK  2017      8   18    18
728        USA  2018      6   22    14
729     Canada  2019      4    5    17
730         UK  2020      2   29    20
731        USA  2020     11   15    15

[732 rows x 14 columns]
```

```python
# 2. What are the sentiment categories? Tabulate and visualize the
number of texts for each category.
print(f"Sentiment categories: \n")
print(data['Sentiment'].unique())
```

```
Sentiment categories:

['positive' 'negative' 'neutral' 'anger' 'fear' 'sadness' 'disgust'
 'happiness' 'joy' 'love' 'amusement' 'enjoyment' 'admiration'
'affection'
 'awe' 'disappointed' 'surprise' 'acceptance' 'adoration'
'anticipation'
 'bitter' 'calmness' 'confusion' 'excitement' 'kind' 'pride' 'shame'
 'elation' 'euphoria' 'contentment' 'serenity' 'gratitude' 'hope'
 'empowerment' 'compassion' 'tenderness' 'arousal' 'enthusiasm'
 'fulfillment' 'reverence' 'despair' 'grief' 'loneliness' 'jealousy'
 'resentment' 'frustration' 'boredom' 'anxiety' 'intimidation'
```

```
 'helplessness' 'envy' 'regret' 'curiosity' 'indifference' 'numbness'
 'melancholy' 'nostalgia' 'ambivalence' 'determination' 'zest'
'hopeful'
 'proud' 'grateful' 'empathetic' 'compassionate' 'playful' 'free-
spirited'
 'inspired' 'confident' 'bitterness' 'yearning' 'fearful'
'apprehensive'
 'overwhelmed' 'jealous' 'devastated' 'frustrated' 'envious'
'dismissive'
 'thrill' 'bittersweet' 'overjoyed' 'inspiration' 'motivation'
 'contemplation' 'joyfulreunion' 'satisfaction' 'blessed' 'reflection'
 'appreciation' 'confidence' 'accomplishment' 'wonderment' 'optimism'
 'enchantment' 'intrigue' 'playfuljoy' 'mindfulness' 'dreamchaser'
 'elegance' 'whimsy' 'pensive' 'harmony' 'creativity' 'radiance'
'wonder'
 'rejuvenation' 'coziness' 'adventure' 'melodic' 'festivejoy'
 'innerjourney' 'freedom' 'dazzle' 'adrenaline' 'artisticburst'
 'culinaryodyssey' 'resilience' 'immersion' 'spark' 'marvel'
'heartbreak'
 'betrayal' 'suffering' 'emotionalstorm' 'isolation' 'disappointment'
 'lostlove' 'exhaustion' 'sorrow' 'darkness' 'desperation' 'ruins'
 'desolation' 'loss' 'heartache' 'solitude' 'positivity' 'kindness'
 'friendship' 'success' 'exploration' 'amazement' 'romance'
'captivation'
 'tranquility' 'grandeur' 'emotion' 'energy' 'celebration' 'charm'
 'ecstasy' 'colorful' 'hypnotic' 'connection' 'iconic' 'journey'
 'engagement' 'touched' 'suspense' 'triumph' 'heartwarming' 'obstacle'
 'sympathy' 'pressure' 'renewed effort' 'miscalculation' 'challenge'
 'solace' 'breakthrough' 'joy in baking' 'envisioning history'
 'imagination' 'vibrancy' 'mesmerizing' 'culinary adventure'
 'winter magic' 'thrilling journey' "nature's beauty" 'celestial
wonder'
 'creative inspiration' 'runway creativity' "ocean's freedom"
 'whispers of the past' 'relief' 'embarrassed' 'mischievous' 'sad'
'hate'
 'bad' 'happy']


texts = data.groupby('Sentiment')
['Text'].count().sort_values(ascending = False).head(10)
print(texts)
plt.figure(figsize=(12,6))
sns.barplot(x = texts.index, y= texts.values)
plt.title('Number of texts in each Category')
plt.xlabel('Categories', fontsize = 10)
plt.ylabel('Number of texts', fontsize = 10)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)
plt.show()
```

```
Sentiment
positive       45
joy            44
excitement     37
contentment    19
gratitude      18
neutral        18
curiosity      16
serenity       15
happy          14
despair        11
Name: Text, dtype: int64
```



Observation:

1. Most Frequent Categories:

positive, joy and excitement have high occurance

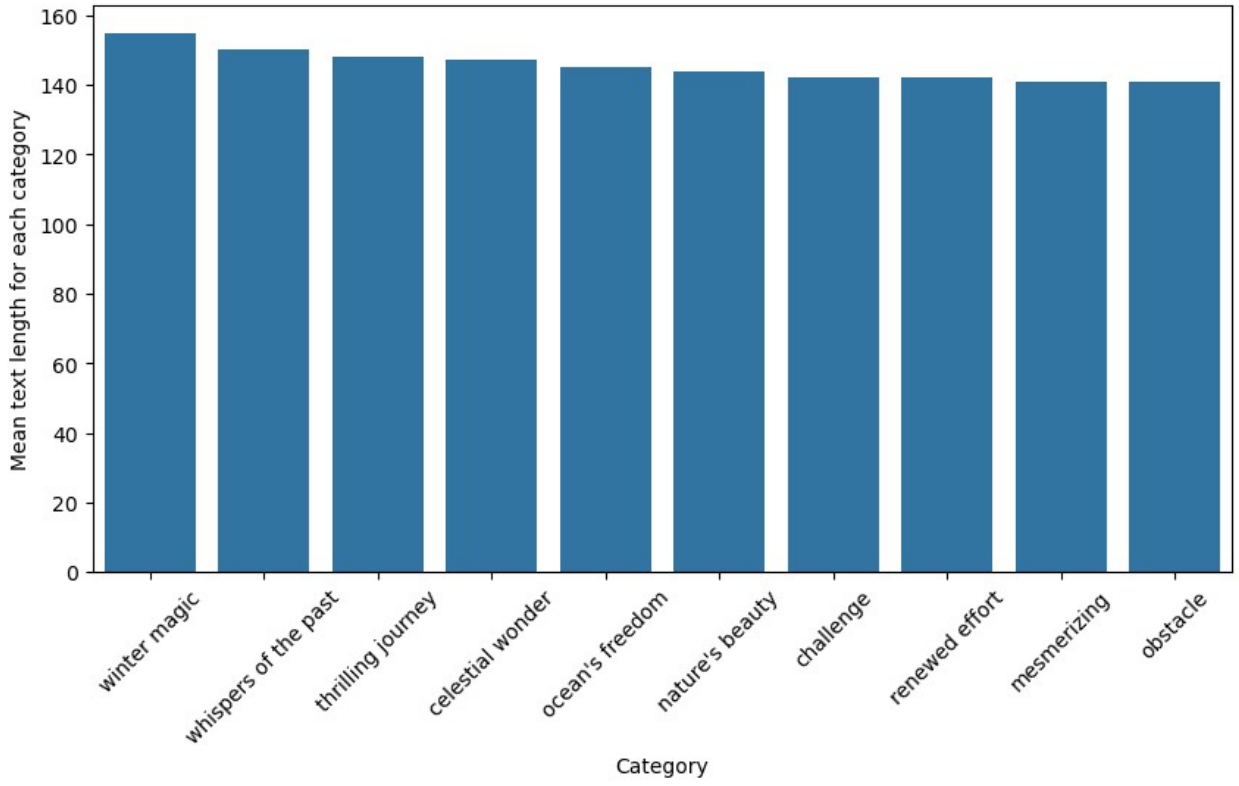
1. Moderate Categories:

contentment, gratitude and neutral have medium occurance.

1. Less Frequent Categories:

others have less occurance.

```
# 3. Tabulate and Visualize the text length for each sentiment
category.
data['Text_length'] = data['Text'].apply(len)
text_lengths = data.groupby('Sentiment')
['Text_length'].mean().sort_values(ascending = False).head(10)
plt.figure(figsize = (10,5))
sns.barplot(x=text_lengths.index, y= text_lengths.values)
plt.xlabel('Category')
plt.ylabel('Mean text length for each category')
plt.xticks(rotation = 45)
plt.show()
```



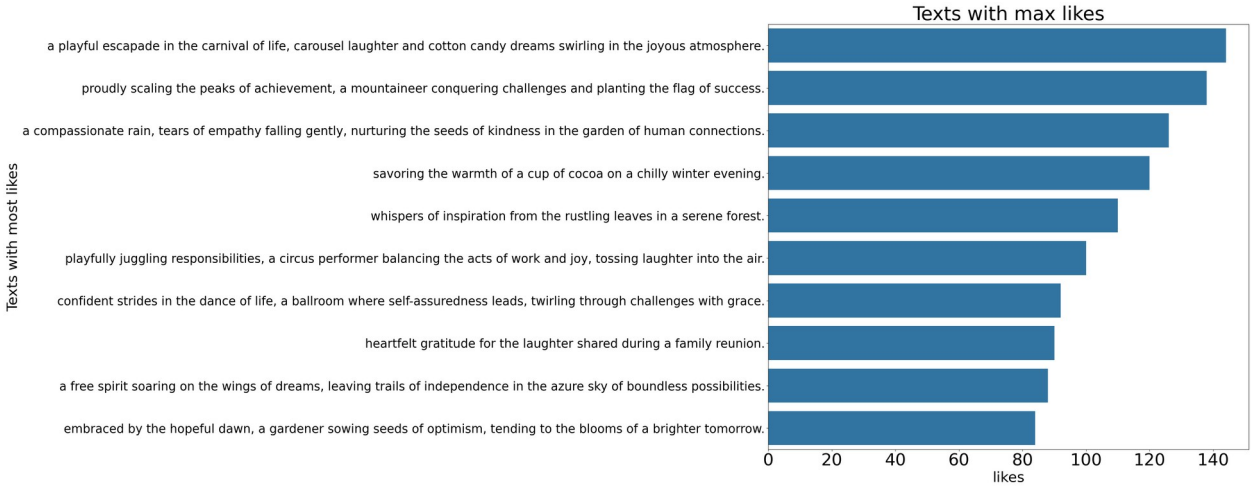
Observations:

1. Winter magic has highest mean text length.
2. Mesmerizing has lowest mean text length.

rest have similar, represented in order.

```
# 4. Find and display the text(s) with the maximum number of likes.
likes = data.groupby('Text')['Likes'].sum().sort_values(ascending =
False).head(10)
plt.figure(figsize = (18,16))
```

```
sns.barplot(y= likes.index, x = likes.values)
plt.xticks(fontsize = 35)
plt.yticks(fontsize = 25)
plt.xlabel('likes', fontsize = 30)
plt.ylabel('Texts with most likes', fontsize = 30)
plt.title('Texts with max likes', fontsize = 40)
plt.show()
```



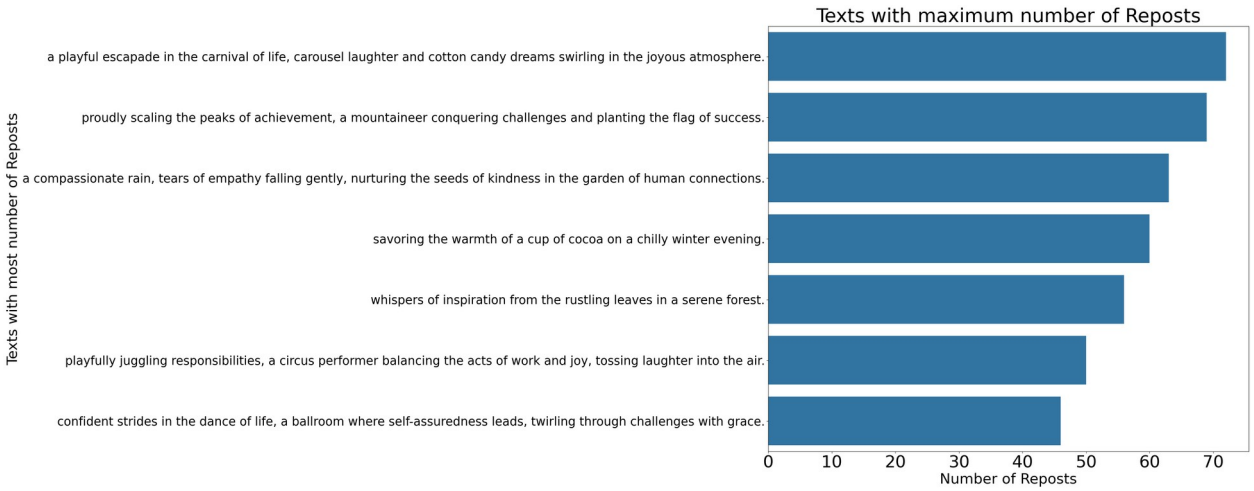
Observations:

1. a playful escapade. received the highest number of likes.

2.embraced by the hopeful dawn... received the least number of likes among the top-ranked texts.

1. the other are in descending order of likes.

```
# 5. Find and display the text(s) with the maximum reposts. what is
its sentiment?
posts = data.groupby('Text')['Retweets'].sum().sort_values(ascending =
False).head(7)
plt.figure(figsize = (18,16))
sns.barplot(y= posts.index, x = posts.values)
plt.xticks(fontsize = 35)
plt.yticks(fontsize = 25)
plt.xlabel('Number of Reposts', fontsize = 30)
plt.ylabel('Texts with most number of Reposts', fontsize = 30)
plt.title('Texts with maximum number of Reposts', fontsize = 40)
plt.show()
```



Observations:

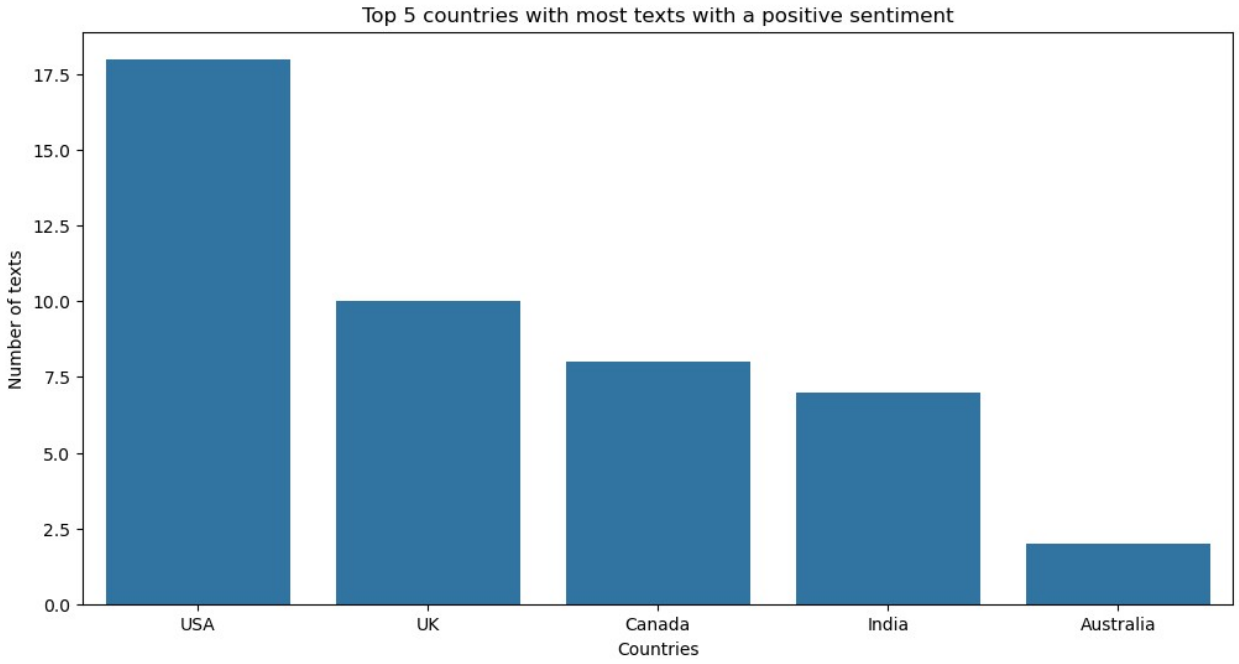
1. The top text has the highest number of reposts, exceeding 70.

rest are in descending order of reposts, this graph has only top 7

```
# 6. Tabulate and Visualize the top 5 countries that have posted the
most texts with a positive sentiment.
countries = data[data['Sentiment']=='positive'].groupby('Country')
['Sentiment'].count().sort_values(ascending = False).head(5)
countries
```

```
Country
USA          18
UK           10
Canada        8
India         7
Australia     2
Name: Sentiment, dtype: int64
```

```
plt.figure(figsize = (12,6))
sns.barplot(x = countries.index, y = countries.values)
plt.xlabel('Countries')
plt.ylabel('Number of texts')
plt.title('Top 5 countries with most texts with a positive sentiment')
plt.show()
```



Observations:

The USA has the highest number of texts with a positive sentiment (~18), nearly double that of the second-highest country, the UK (~10).

Australia has the lowest count – Despite being in the top 5, Australia has a much lower number of positive sentiment texts compared to the USA.

Balanced middle group – The UK, Canada, and India show a relatively closer range

```
# 7. Find and display all the text(s) with the word 'Travel' in their hashtag.
data_travel = data[data['Hashtags'].str.contains('Travel')]
data_travel
```

```
      ID                                               Text
Sentiment  \
3      3        excited about the upcoming weekend getaway!
positive
72    73  excitement building up for the upcoming vacation!
excitement
84    85  enjoying every moment of this trip—pure enjoym...
enjoyment
145  146  arousal of excitement before a much-awaited trip.
arousal
596  597  spontaneously booked a weekend getaway. advent...
excitement
610  611  embarked on a road trip to revisit cherished p...
```

```
joy
622  623  embarked on a solo travel adventure, discoveri...
excitement

             Timestamp               User     Platform  \
3    15-01-2023 18:20      AdventureX       Facebook
72   18-02-2023 18:45   VacationDreamer     Facebook
84   23-02-2023 19:45   TravelExplorer      Facebook
145  10-03-2018 09:45   TravelExcitement    Twitter
596  05-06-2023 10:00   AdventureSeeker    Instagram
610  19-06-2023 16:30    RoadTripSenior     Facebook
622  01-07-2023 14:00    SoloAdventurer    Instagram

                                          Hashtags  Retweets  Likes
Country  \
3          #Travel #Adventure                              8     15
UK
72       #Excitement #TravelPlans                         22     45
Canada
84     #Enjoyment #Travel                                 25     50
USA
145   #Arousal #TravelAdventure                           20     40
India
596                      #GetawayGoals #TeenTravel        18     35
UK
610                  #CherishedPlaces #SeniorTravel       22     45
Canada
622                #DiscoverNewPlaces #SeniorTravel       30     60
USA

     Year  Month  Day  Hour  Text_length
3    2023      1   15    18           43
72   2023      2   18    18           49
84   2023      2   23    19           50
145  2018      3   10     9           49
596  2023      6    5    10           83
610  2023      6   19    16          133
622  2023      7    1    14          117
```

```
# 8. Find and display the most popular (most number of likes) posts
(top 5) from the year 2020 which were of a negative sentiment.

data_2020_negative_posts = data[(data['Year']==2020) &
(data['Sentiment']=='negative')]
data_2020_negative_posts
```

```
Empty DataFrame
Columns: [ID, Text, Sentiment, Timestamp, User, Platform, Hashtags,
Retweets, Likes, Country, Year, Month, Day, Hour, Text_length]
Index: []
```

```
# 9. Find and display a list of Users and their posts whose texts are labeled ‘Happiness’ sentiment.
data_happy = data[data['Sentiment']=='happy']
data_happy
```

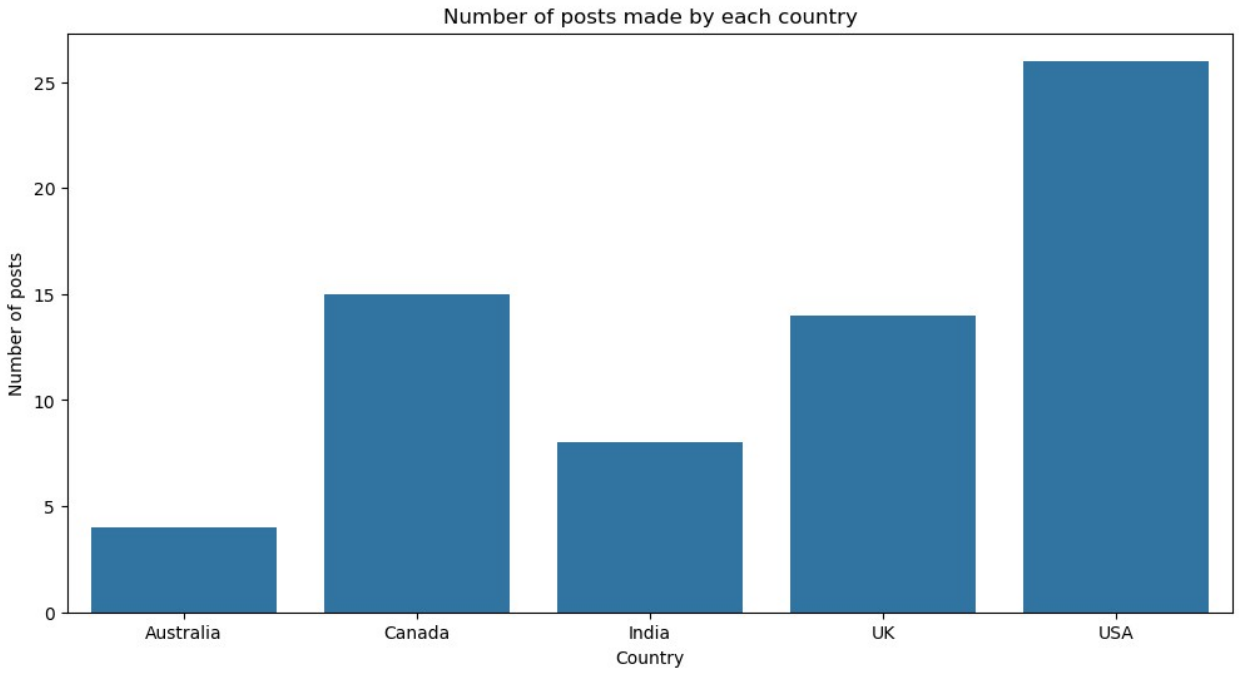
```
      ID                                               Text Sentiment
\
718  719  celebrating a friend's birthday with a surpris...     happy

719  720  successfully completing a challenging coding p...     happy

720  721  attending a school talent show to support clas...     happy

721  722  exploring a new hiking trail with friends over...     happy

722  723  winning a friendly sports competition against ...     happy

723  724  receiving a heartfelt letter from a pen pal in...     happy

724  725  creating a beautiful mural with fellow art ent...     happy

725  726  participating in a school-wide art exhibition....     happy

726  727  achieving a personal best in a track and field...     happy

727  728  collaborating on a science project that receiv...     happy

728  729  attending a surprise birthday party organized ...     happy

729  730  successfully fundraising for a school charity ...     happy

730  731  participating in a multicultural festival, cel...     happy

731  732  organizing a virtual talent show during challe...     happy


            Timestamp                                       User
Platform  \
718  16-10-2023 15:45            BirthdaySurpriseHighSchool
Instagram
719  17-10-2023 12:30               CodingSuccessHighSchool
Twitter
720  18-10-2023 16:45           TalentShowSupportHighSchool
Facebook
721  19-10-2023 18:20                 WeekendHikeHighSchool
Instagram
722  20-10-2023 14:15               SportsVictoryHighSchool
Twitter
723  21-10-2023 17:30            PenPalConnectionHighSchool
Facebook
724  22-10-2023 20:45               MuralCreationHighSchool
```

```
Instagram
725  10-11-2015 16:45            ArtExhibitionJoyHighSchool
Instagram
726  14-09-2016 12:30          TrackFieldSuccessHighSchool
Twitter
727  18-08-2017 18:20       ScienceProjectSuccessHighSchool
Facebook
728  22-06-2018 14:15            BirthdayPartyJoyHighSchool
Instagram
729  05-04-2019 17:30   CharityFundraisingTriumphHighSchool
Twitter
730  29-02-2020 20:45    MulticulturalFestivalJoyHighSchool
Facebook
731  15-11-2020 15:15    VirtualTalentShowSuccessHighSchool
Instagram

                                          Hashtags  Retweets  Likes
Country  \
718         #SurpriseParty #HighSchoolCelebration          27     52
UK
719                #CodingTriumph #HighSchoolTech          24     48
USA
720          #TalentShow #HighSchoolEntertainment          19     38
Canada
721          #NatureAdventures #HighSchoolFriends          21     42
UK
722           #SportsTriumph #HighSchoolAthletics          26     50
USA
723   #GlobalFriendship #HighSchoolCorrespondence          23     45
Canada
724       #ArtCollaboration #HighSchoolCreativity          22     43
UK
725     #ArtisticExpression #HighSchoolCreativity          23     44
USA
726        #AthleticAchievement #HighSchoolSports          26     51
Canada
727         #ScienceFairWinner #HighSchoolScience          20     39
UK
728    #SurpriseCelebration #HighSchoolFriendship          25     48
USA
729      #CommunityGiving #HighSchoolPhilanthropy          22     42
Canada
730         #CulturalCelebration #HighSchoolUnity          21     43
UK
731   #VirtualEntertainment #HighSchoolPositivity          24     47
USA

     Year  Month  Day  Hour  Text_length
718  2023     10   16    15           94
```

```
719  2023     10   17    12           95
720  2023     10   18    16           96
721  2023     10   19    18           93
722  2023     10   20    14           96
723  2023     10   21    17           92
724  2023     10   22    20           98
725  2015     11   10    16           98
726  2016      9   14    12          103
727  2017      8   18    18          109
728  2018      6   22    14          105
729  2019      4    5    17           98
730  2020      2   29    20          103
731  2020     11   15    15           96
```

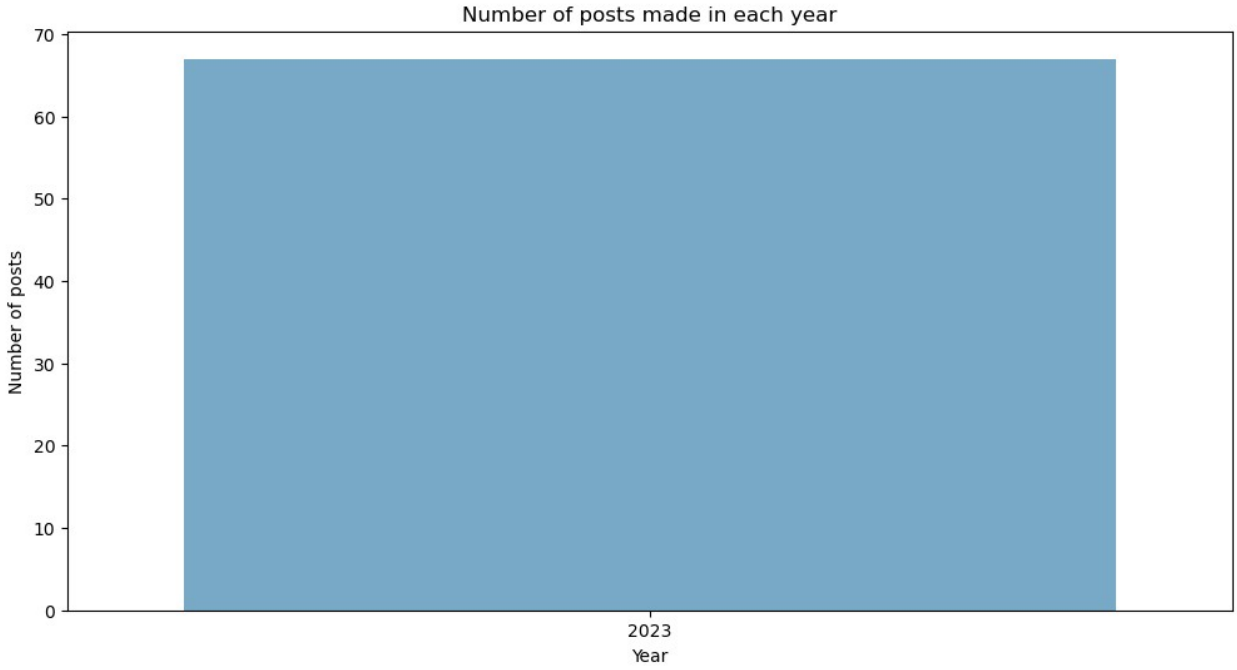
```
# 10. Tabulate and visualize the years and countrywide tweets of the
sentiments. Pick tweets with sentiments, 'Positive,' 'Negative,' and
'Neutral' only.
new_data = data[(data['Sentiment']=='positive') |
(data['Sentiment']=='negative') | (data['Sentiment']=='neutral')]
a = new_data.groupby('Country')['Text'].count()
plt.figure(figsize = (12,6))
sns.barplot(x =a.index, y = a.values)
plt.ylabel('Number of posts')
plt.title('Number of posts made by each country')
```

```
Text(0.5, 1.0, 'Number of posts made by each country')
```



```
b = new_data.groupby('Year')['Text'].count()
plt.figure(figsize = (12,6))
sns.barplot(x =b.index, y = b.values, palette = 'Blues')
plt.ylabel('Number of posts')
plt.title('Number of posts made in each year')
```

```
/tmp/ipykernel_106293/3135719151.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x =b.index, y = b.values, palette = 'Blues')

Text(0.5, 1.0, 'Number of posts made in each year')
```



In terms of country-wise distribution, the United States is the most active with approximately 27 posts,

while Canada and the United Kingdom each contribute around 14 to 15 posts. India shows lower engagement with about 8 posts,

and Australia has the least participation with roughly 4 posts. Additionally,

the year-wise distribution reveals that all posts in the dataset are from 2023,

totaling around 67 posts, with no entries from any other year.