Ahmed Ali Khan

Data Scrapping of Snapchat

In [66]:
```
from google_play_scraper import app

result = app(
    'com.snapchat.android',
    lang='en',
    country='pk'
)
print(result)
```

```
{'title': 'Snapchat', 'description': "Snapchat is a fast and fun way to s
hare the moment with your friends and family 👻\r\n\r\nSNAP \r\n• Snapch
at opens right to the Camera – just tap to take a photo, or press and hol
d for video.\r\n• Express yourself with Lenses, Filters, Bitmoji and mor
e! \r\n• Try out new Lenses daily created by the Snapchat community!\r\n
\r\nCHAT \r\n• Stay in touch with friends through live messaging, or shar
e your day with Group Stories.\r\n• Video Chat with up to 16 friends at o
nce – you can even use Lenses and Filters when chatting!\r\n• Express you
rself with Friendmojis – exclusive Bitmoji made just for you and a frien
d.\r\n\r\nSTORIES\r\n• Watch friends' Stories to see their day unfold.\r
\n• See Stories from the Snapchat community that are based on your intere
sts.\r\n• Discover breaking news and exclusive Original Shows.\r\n\r\nSPO
TLIGHT\r\n• Spotlight showcases the best of Snapchat!\r\n• Submit your ow
n Snaps or sit back, relax, and watch.\r\n• Pick your favorites and share
them with friends.\r\n\r\nMAP \r\n• Share your location with your best fr
iends or go off the grid with Ghost Mode.\r\n• See what your friends are
up to on your most personal map when they share their location with you.
\r\n• Explore live Stories from the community nearby or across the world!
\r\n\r\nMEMORIES \r\n• Save unlimited photos and videos of all your favor
```

In [67]:
```
result.keys()
```

Out[67]:
```
dict_keys(['title', 'description', 'descriptionHTML', 'summary', 'installs',
'minInstalls', 'realInstalls', 'score', 'ratings', 'reviews', 'histogram',
'price', 'free', 'currency', 'sale', 'saleTime', 'originalPrice', 'saleTex
t', 'offersIAP', 'inAppProductPrice', 'developer', 'developerId', 'developer
Email', 'developerWebsite', 'developerAddress', 'privacyPolicy', 'genre', 'g
enreId', 'categories', 'icon', 'headerImage', 'screenshots', 'video', 'video
Image', 'contentRating', 'contentRatingDescription', 'adSupported', 'contain
sAds', 'released', 'lastUpdatedOn', 'updated', 'version', 'comments', 'appI
d', 'url'])
```

Now Getting the data

In [68]:
```
result = search("Snapchat", lang="en", country="pk")
result
```

Out[68]:
```
[{'appId': 'com.snapchat.android',
  'icon': 'https://play-lh.googleusercontent.com/KxeSAjPTKliCErbivNiXrd6c
TwfbqUJcbSRPe_IBVK_YmwckfMRS1VIHz-5cgT09yMo',
  'screenshots': ['https://play-lh.googleusercontent.com/xKCYMMuIshGmxLVc
kXnGYsdorvBxF0oI58Yt82Vkj_cn3Dby52gdrt4Lmr7BTYiVww',
   'https://play-lh.googleusercontent.com/yoN8h1j4M0Axz1UK2-iyPOmlQmIHqZ1
tO8p7PjRicfmyPxj3-rztyB3YImG58zeMvOI',
   'https://play-lh.googleusercontent.com/3KtSEofVcRIQUWQnq814EYI7xo61uD6
Cg1tvWdYTPUXweFMjv7HkA9b9FVmt3zYDkSau',
   'https://play-lh.googleusercontent.com/CHV1Yg5KonaNU_HHXG_Iol6zn7bAQ0b
0o868LTPyh9bCLoZDu0GM8EJxBQ4zTcMfy90',
   'https://play-lh.googleusercontent.com/Mbz03v3wiyrsITveiF4UO9dj_XNo1Wp
Gw7ZoTSISZZmu3Nl1fu1uuDK3k-WwmaLJYWY',
   'https://play-lh.googleusercontent.com/OfIjocrhVT1G54fmxhNKm9mhNa2bQOO
ryyxiWc2Wk6AAcCZheJ7JTtNhusl9G_c5lQ',
   'https://play-lh.googleusercontent.com/Bj1VhBkftmSTR7VxeWV0IISO5bw9EQ8
Nk7hjMYHHmy5EgkLhr82SdGm4xJcQwB87d44',
   'https://play-lh.googleusercontent.com/RHW0l21p3QP3ZaU2uS7JU_s0-UNrjiv
C9VhiGBwB1hB0Q4ALY8oWH_6Hd8BTizCEXYE',
```

In [69]:
```
len(result)
```

Out[69]:
```
19
```

In [70]:
```
print("Title:", snapchat_info.get('title'))
print("Summary:", snapchat_info.get('summary'))
print("Installs:", snapchat_info.get('installs'))
print("Score:", snapchat_info.get('score'))
print("Ratings:", snapchat_info.get('ratings'))
print("Reviews:", snapchat_info.get('reviews'))
print("Genre:", snapchat_info.get('genre'))
```

```
Title: Snapchat
Summary: Share the moment!
Installs: 1,000,000,000+
Score: 4.055323
Ratings: 34074549
Reviews: 1753512
Genre: Communication
```

In [71]:
```
data = [{
    "Id": snapchat_info.get("appId"),
    "title": snapchat_info.get("title"),
    "summary": snapchat_info.get("summary"),
    "realInstalls": snapchat_info.get("installs"),
    "Average_rating": snapchat_info.get("score"),
    "Ratings_received": snapchat_info.get("ratings"),
    "Reviews_received": snapchat_info.get("reviews"),
    "price": snapchat_info.get("price"),
    "developer_name": snapchat_info.get("developer"),
    "genre": snapchat_info.get("genre")
}]
```

In [72]:
```
import pandas as pd

df = pd.DataFrame(data)
df
```

Out[72]:

	Id	title	summary	realInstalls	Average_rating	Ratings_received	R
0	com.snapchat.android	Snapchat	Share the moment!	1,000,000,000+	4.055323	34074549	

Functions for the above steps

In [73]:
```
import google_play_scraper
import pandas as pd

def retrieve_snapchat_info():
    snapchat_package_name = 'com.snapchat.android'
    snapchat_info = google_play_scraper.app(snapchat_package_name)

    data = [{
        "Id": snapchat_info.get("appId"),
        "title": snapchat_info.get("title"),
        "summary": snapchat_info.get("summary"),
        "number_of_installs": snapchat_info.get("installs"),
        "Average_rating": snapchat_info.get("score"),
        "Ratings_received": snapchat_info.get("ratings"),
        "Reviews_received": snapchat_info.get("reviews"),
        "Ratings_histogram": snapchat_info.get("histogram"),
        "price": snapchat_info.get("price"),
        "developer_name": snapchat_info.get("developer"),
        "genre": snapchat_info.get("genre")
    }]

    df = pd.DataFrame(data)
    return df

# Call the function
snapchat_df = retrieve_snapchat_info()
print(snapchat_df)
```

```
                    Id     title            summary number_of_installs  \
0  com.snapchat.android  Snapchat  Share the moment!     1,000,000,000+

   Average_rating  Ratings_received  Reviews_received  \
0        4.055319          34074467           1753511

                               Ratings_histogram  price developer_name  \
0  [5129201, 1316360, 2131377, 3460800, 22036706]      0       Snap Inc

          genre
0  Communication
```

Save the data into Csv

In [74]:
```
snapchat_df.to_csv("snapchat_info.csv", index=False)
```

In [75]:
```
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv("snapchat_info.csv")

# Filter the DataFrame for Snapchat
snapchat_df = df[df['title'] == 'Snapchat']

# Display the shape of the filtered DataFrame
print(snapchat_df.shape)
```

```
(1, 11)
```

Now Lets Scrap the data

In [76]:
```
snapchat_reviews, continuation_token = reviews(snapchat_package_name, sort=So
```

In [77]:
```
snapchat_reviews
```

Out[77]:
```
[{'reviewId': 'e15c144a-b89f-4b84-a0aa-144680c01dae',
  'userName': 'sharif ALASSil',
  'userImage': 'https://play-lh.googleusercontent.com/a/ACg8ocIPoMxh9AKzB
oQnQuFLbJzQhPeuZrjsjZMBAapJCyCa=mo',
  'content': 'Byoutiful',
  'score': 5,
  'thumbsUpCount': 3,
  'reviewCreatedVersion': None,
  'at': datetime.datetime(2024, 2, 23, 20, 36, 4),
  'replyContent': None,
  'repliedAt': None,
  'appVersion': None},
 {'reviewId': '8f6d725d-491c-4fba-9b08-8d360df1913f',
  'userName': 'Aziz Baloch',
  'userImage': 'https://play-lh.googleusercontent.com/a/ACg8ocIKpOrAqdspH
S-PqSsx9T1DxvF-zKyQVDsJ9-i3lX_z=mo',
  'content': 'Nice',
  'score': 5,
  'thumbsUpCount': 0,
```

In [78]:
```
for review in snapchat_reviews:
    print(review.keys())
    break
```

```
dict_keys(['reviewId', 'userName', 'userImage', 'content', 'score', 'thumbsU
pCount', 'reviewCreatedVersion', 'at', 'replyContent', 'repliedAt', 'appVers
ion'])
```

In [79]:
```
def scrap_snapchat_reviews(output_csv_file):
    # Specify the package name for Snapchat
    snapchat_package_name = 'com.snapchat.android'

    # Fetch the reviews for Snapchat
    snapchat_reviews, _ = reviews(
        snapchat_package_name,
        sort=Sort.NEWEST
    )

    # Create a DataFrame for the reviews
    reviews_df = pd.DataFrame(snapchat_reviews)

    # Add 'Id' column with Snapchat's package name
    reviews_df['Id'] = snapchat_package_name

    # Save the reviews DataFrame to a CSV file
    reviews_df.to_csv(output_csv_file, index=False, encoding='utf-8')

    print(f"Snapchat reviews retrieval completed and saved to {output_csv_file

# Call the function with the desired output CSV file
scrap_snapchat_reviews("snapchat_reviews.csv")
```

```
Snapchat reviews retrieval completed and saved to snapchat_reviews.csv.
```

In [80]:
```
snapchat_reviews_df = pd.read_csv("snapchat_reviews.csv")

print(snapchat_reviews_df.head(2))
```

```
                                reviewId        userName  \
0  e15c144a-b89f-4b84-a0aa-144680c01dae  sharif ALASSil
1  8f6d725d-491c-4fba-9b08-8d360df1913f     Aziz Baloch

                                           userImage    content  score  \
0  https://play-lh.googleusercontent.com/a/ACg8oc... (https://play-lh.google
usercontent.com/a/ACg8oc...)  Byoutiful      5
1  https://play-lh.googleusercontent.com/a/ACg8oc... (https://play-lh.google
usercontent.com/a/ACg8oc...)       Nice      5

   thumbsUpCount reviewCreatedVersion                   at  replyContent  \
0              3                 NaN  2024-02-23 20:36:04           NaN
1              0                 NaN  2024-02-23 20:00:06           NaN

   repliedAt appVersion                    Id
0        NaN        NaN  com.snapchat.android
1        NaN        NaN  com.snapchat.android
```

Now Lets Visualize it

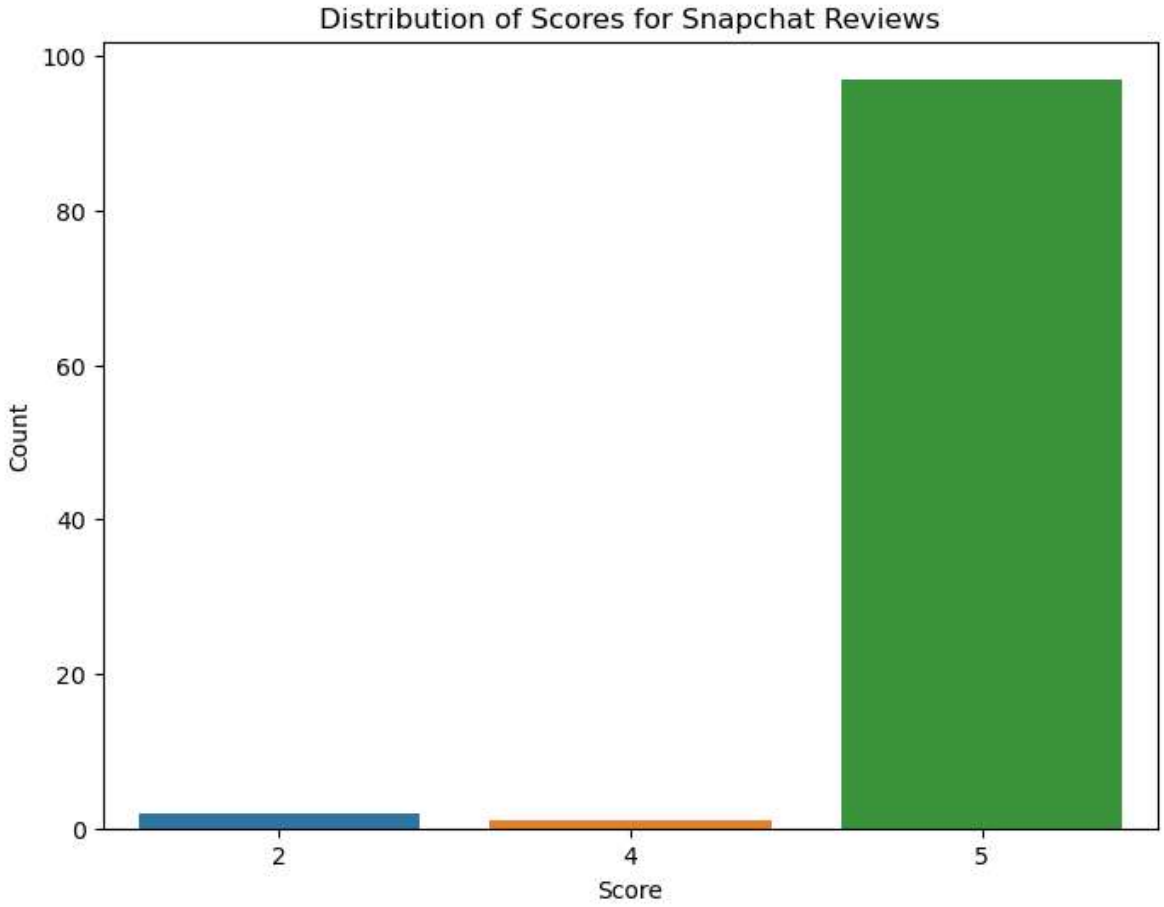
In [81]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

In [82]:
```
# Print the columns in your DataFrame
print(df_snapchat.columns)
```

```
Index(['reviewId', 'userName', 'userImage', 'content', 'score',
       'thumbsUpCount', 'reviewCreatedVersion', 'at', 'replyContent',
       'repliedAt', 'appVersion', 'Id'],
      dtype='object')
```

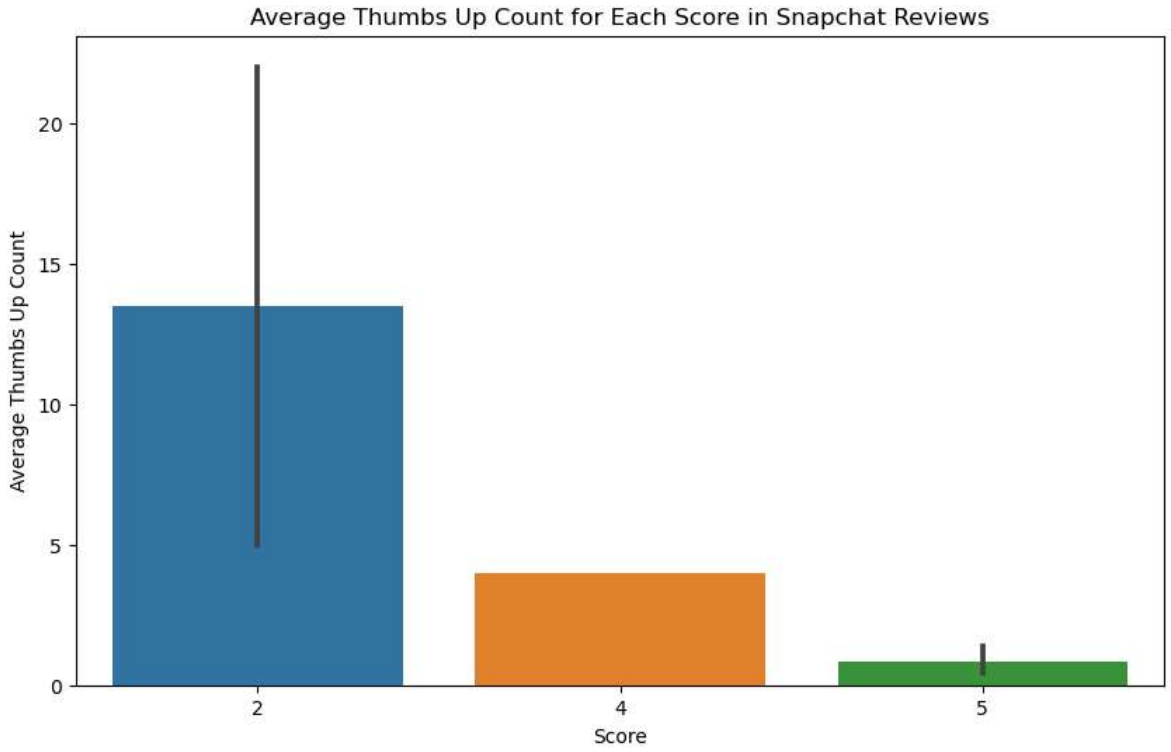
Countplot for the distribution of scores

In [83]:
```
plt.figure(figsize=(8, 6))
sns.countplot(x='score', data=df_snapchat)
plt.title('Distribution of Scores for Snapchat Reviews')
plt.xlabel('Score')
plt.ylabel('Count')
plt.show()
```



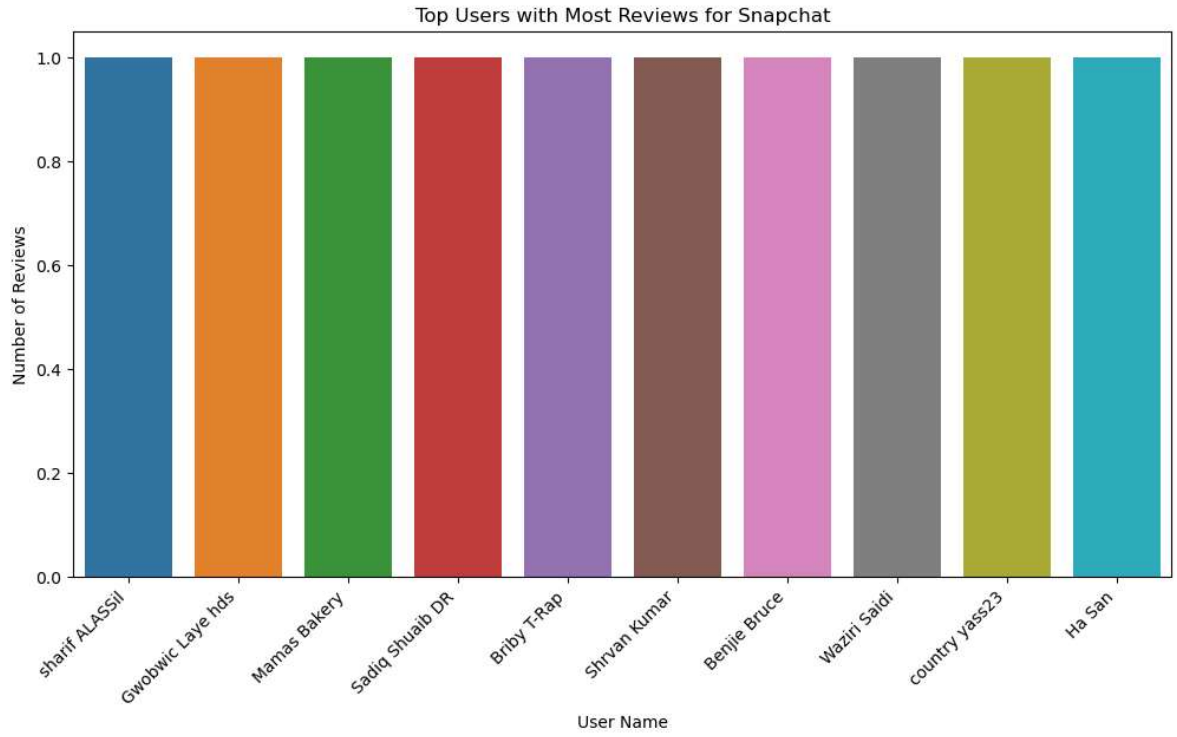
Barplot for average thumbs up count for each score

In [84]:
```
plt.figure(figsize=(10, 6))
sns.barplot(x='score', y='thumbsUpCount', data=df_snapchat)
plt.title('Average Thumbs Up Count for Each Score in Snapchat Reviews')
plt.xlabel('Score')
plt.ylabel('Average Thumbs Up Count')
plt.show()
```



Countplot for the top users with the most reviews

In [85]:
```
top_users = df_snapchat['userName'].value_counts().nlargest(10)
plt.figure(figsize=(12, 6))
sns.barplot(x=top_users.index, y=top_users.values)
plt.title('Top Users with Most Reviews for Snapchat')
plt.xlabel('User Name')
plt.ylabel('Number of Reviews')
plt.xticks(rotation=45, ha='right')
plt.show()
```

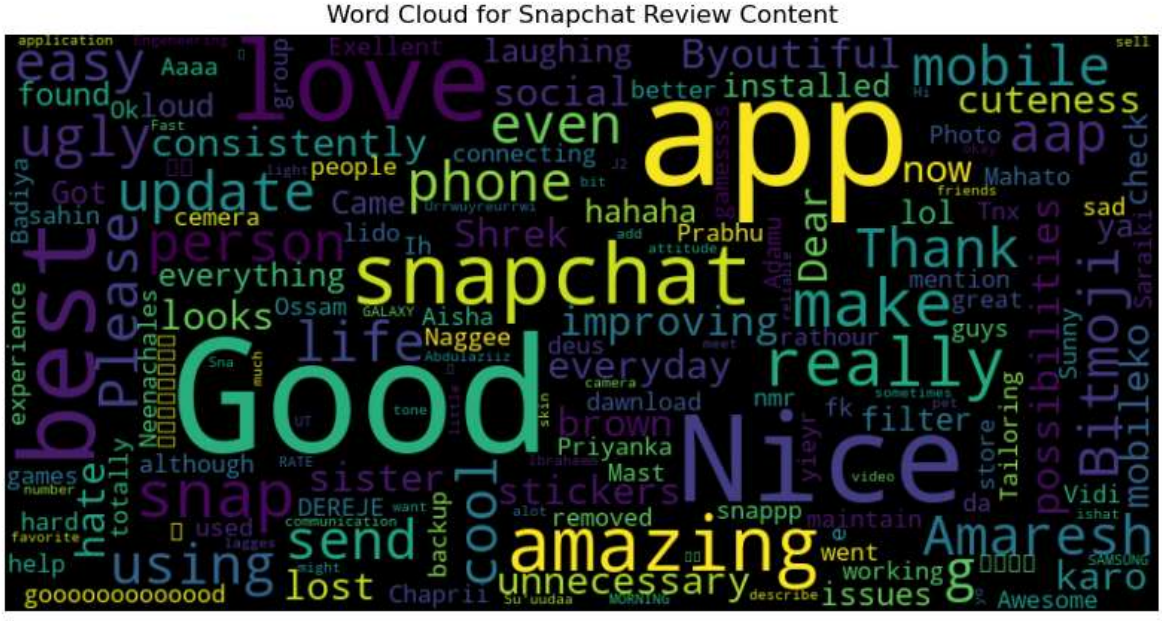


#Generate Word Cloud for review content

In [86]:
```
!pip install wordcloud
from wordcloud import WordCloud
```

```
Requirement already satisfied: wordcloud in c:\users\pmyls\anaconda3\lib\sit
e-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in c:\users\pmyls\anaconda3\lib
\site-packages (from wordcloud) (1.24.3)
Requirement already satisfied: pillow in c:\users\pmyls\anaconda3\lib\site-p
ackages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in c:\users\pmyls\anaconda3\lib\si
te-packages (from wordcloud) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\pmyls\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\pmyls\anaconda3\lib
\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\pmyls\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\pmyls\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\pmyls\anaconda3\l
ib\site-packages (from matplotlib->wordcloud) (23.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\pmyls\anaco
nda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\pmyls\anacon
da3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\pmyls\anaconda3\lib\site
-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
```

In [87]:
```
wordcloud = WordCloud(width=800, height=400, random_state=21, max_font_size=1
plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt.title('Word Cloud for Snapchat Review Content')
plt.show()
```



Ive Saved the data I Scrapped

In [88]:
```
# Specify the path to your desktop
desktop_path = "/path/to/your/desktop"  # Replace with the actual path to your

# Save the DataFrame to a CSV file on your desktop
df_snapchat.to_csv(f"C://Users//PMYLS//Desktop//snapchat_data.csv", index=Fals

print("Data saved to desktop successfully.")
```

```
Data saved to desktop successfully.
```

In [ ]:
```

```