In [14]:
```
import pandas as pd
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.corpus import stopwords
import string
import re
nltk.download('stopwords')
stemmer = nltk.SnowballStemmer("english")

data = pd.read_csv("tiktok_google_play_reviews.csv")
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\DELL\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
<ipython-input-14-b7d4c5cde063>:12: DtypeWarning:

Columns (8,9) have mixed types. Specify dtype option on import or set low_memory=Fals
e.
```

In [15]:
```
data.head()
```

Out[15]:

	reviewId	userName	userImage	content	score	thumbsUpCount
0	68ccaec8-1415-4301-a85e-3004679a3a83	Cassie Moore	https://play-lh.googleusercontent.com/a/ALm5wu...	No words	5	0
1	d84cbfd3-6aa3-485c-aaf9-c5dca27dc966	Kaleb Plummer	https://play-lh.googleusercontent.com/a-/ACNPE...	Great fun app so far!	5	0
2	96618aa1-31e5-4259-8649-89b75d962f00	Rylee Maher	https://play-lh.googleusercontent.com/a/ALm5wu...	The app would get a higher rating but I litera...	1	0
3	078c0bda-598b-474e-a04e-d7cb3e6f6301	Kittykatelyn Romilly	https://play-lh.googleusercontent.com/a-/ACNPE...	I WISH I COULD GIVE THIS A 100 PERCENT RATING ...	5	0
4	8e68c5cd-b12a-4206-a8da-6bfdbff44ae3	Loveness Malenga	https://play-lh.googleusercontent.com/a-/ACNPE...	Pictures and record	5	0

In [16]:
```
data = data[["content", "score"]]
data.head()
```

Out[16]:

	content	score
0	No words	5
1	Great fun app so far!	5
2	The app would get a higher rating but I litera...	1
3	I WISH I COULD GIVE THIS A 100 PERCENT RATING ...	5
4	Pictures and record	5

In [17]:
```
print(data.isnull().sum())
```

```
content    16
score       0
dtype: int64
```

In [18]:
```
data = data.dropna()
```

In [19]:
```
data.head()
```

Out[19]:

	content	score
0	No words	5
1	Great fun app so far!	5
2	The app would get a higher rating but I litera...	1
3	I WISH I COULD GIVE THIS A 100 PERCENT RATING ...	5
4	Pictures and record	5

In [20]:
```
data.isnull().sum()
```

Out[20]:
```
content    0
score      0
dtype: int64
```

In [21]:
```
stopword=set(stopwords.words('english'))
def clean(text):
    text = str(text).lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w*\d\w*', '', text)
    text = [word for word in text.split(' ') if word not in stopword]
    text=" ".join(text)
    text = [stemmer.stem(word) for word in text.split(' ')]
    text=" ".join(text)
    return text
```

In [24]:
```
clean("you are beauty")
```

Out[24]:
```
'beauti'
```

In [25]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 460271 entries, 0 to 460286
Data columns (total 2 columns):
 #   Column   Non-Null Count   Dtype 
---  ------   --------------   ----- 
 0   content  460271 non-null  object
 1   score    460271 non-null  int64 
dtypes: int64(1), object(1)
memory usage: 10.5+ MB
```

In [26]:
```
ratings = data["score"].value_counts()
numbers = ratings.index
quantity = ratings.values
import plotly.express as px
figure = px.pie(data,
             values=quantity,
             names=numbers,hole = 0.5)
figure.show()
```

In [27]:
```
text = " ".join(i for i in data.content)
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(text)
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
```

```
plt.axis("off")
plt.show()
```

In [28]:

```
nltk.download('vader_lexicon')
sentiments = SentimentIntensityAnalyzer()
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["content"]]
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["content"]]
data["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in data["content"]]
data = data[["content", "Positive", "Negative", "Neutral"]]
data.head()
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\DELL\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

Out[28]:

	content	Positive	Negative	Neutral
0	No words	0.000	0.688	0.312
1	Great fun app so far!	0.719	0.000	0.281
2	The app would get a higher rating but I litera...	0.000	0.073	0.927
3	I WISH I COULD GIVE THIS A 100 PERCENT RATING ...	0.561	0.000	0.439
4	Pictures and record	0.000	0.000	1.000

In [ ]:

In [29]:

```
positive =' '.join([i for i in data['content'][data['Positive'] > data["Negative"]]]
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(positi
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

In [30]:

```
negative =' '.join([i for i in data['content'][data['Negative'] > data["Positive"]]]
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(negati
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

In [31]:

```
x = sum(data["Positive"])
y = sum(data["Negative"])
z = sum(data["Neutral"])

def sentiment_score(a, b, c):
    if (a>b) and (a>c):
        print("Positive 😊 ")
    elif (b>a) and (b>c):
        print("Negative 😠 ")
    else:
        print("Neutral 😐 ")
sentiment_score(x, y, z)
```

```
Neutral 😐
```

In [32]:
```
print("Positive: ", x)
print("Negative: ", y)
print("Neutral: ", z)
```

```
Positive:  196353.47899996064
Negative:  16512.73600000027
Neutral:  244169.79300007297
```

In [ ]:
```

```