

6rmqfxpgw

May 3, 2024

```
[63]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

```
[64]: DATA= pd.read_csv('app_reviews.csv')
```

```
[65]: DATA.head(5)
```

```
[65]:
```

	name	userName	\
0	WhiteHatJr	All in one Captions	
1	WhiteHatJr	githin george	
2	WhiteHatJr	Anita Choudhury	
3	WhiteHatJr	Snehil gupta	
4	WhiteHatJr	UTKALIKA PRUSTI	

	content	score	thumbsUpCount	\
0	I was very unhappy with this app because I can...	1	1234	
1	Come on! Man it's not normal for kids to devel...	1	213	
2	This app is good for coding but I think this o...	1	156	
3	This app is good for coding but I think this o...	1	82	
4	It is good app . But it's app use only who has...	1	150	

	reviewCreatedVersion	at	\
0	1.5.1	2020-09-18 09:21:07	
1	1.5.2	2020-09-30 14:34:38	
2	1.5.1	2020-09-17 17:50:57	
3	1.5.2	2020-10-09 21:10:06	
4	1.5.1	2020-09-21 05:27:50	

	replyContent	repliedAt	\
0	Hi! Thank you for reviewing. Kindly note that ...	2020-09-18 11:58:46	
1	Hi, Thank you for reviewing and sharing your p...	2020-09-30 16:01:39	
2	Hi! We apologize for the inconvenience. We wou...	2020-09-17 17:59:10	
3	Hi, Thank you for reviewing. Please be apprise...	2020-10-10 10:35:40	

4 Hi! Thank you for reviewing. Kindly note that ... 2020-09-21 10:46:15

	sortOrder	appId
0	most_relevant	com.whitehatjr
1	most_relevant	com.whitehatjr
2	most_relevant	com.whitehatjr
3	most_relevant	com.whitehatjr
4	most_relevant	com.whitehatjr

```
[66]: for col in DATA.columns:
      print(col)
```

```
name
userName
content
score
thumbsUpCount
reviewCreatedVersion
at
replyContent
repliedAt
sortOrder
appId
```

```
[67]: # Then, replace NaN values in thumbsUpCount with a default value (e.g., -1)
DATA['thumbsUpCount'].fillna(-1, inplace=True)

bins = [0, 10, 20, 30, 40, 50, 100, 500, 1000, 1234]
labels = ['0-10', '11-20', '21-30', '31-40', '41-50', '51-100', '101-500',
↪ '501-1000', '1001-1234']
DATA['thumbsUpCategory'] = pd.cut(DATA['thumbsUpCount'], bins=bins,
↪ labels=labels, right=False)

print(DATA)
```

	name	userName \
0	WhiteHatJr	All in one Captions
1	WhiteHatJr	githin george
2	WhiteHatJr	Anita Choudhury
3	WhiteHatJr	Snehil gupta
4	WhiteHatJr	UTKALIKA PRUSTI
...	...	...
13109	TestBook	vivek anand Mishra
13110	TestBook	Raushan singh Rathod
13111	TestBook	IPL 2020 live
13112	TestBook	Saurabh Rathod
13113	TestBook	Daanish King

	content	score \
0	I was very unhappy with this app because I can...	1
1	Come on! Man it's not normal for kids to devel...	1
2	This app is good for coding but I think this o...	1
3	This app is good for coding but I think this o...	1
4	It is good app . But it's app use only who has...	1
...	...	...
13109	Good app.	5
13110	Good	5
13111	This is awesome app for ssc all exams because ...	5
13112	It is very useful for practicing	5
13113	Best I love this	5

	thumbsUpCount	reviewCreatedVersion	at \
0	1234	1.5.1	2020-09-18 09:21:07
1	213	1.5.2	2020-09-30 14:34:38
2	156	1.5.1	2020-09-17 17:50:57
3	82	1.5.2	2020-10-09 21:10:06
4	150	1.5.1	2020-09-21 05:27:50
...	...	...	...
13109	0	5.21.3	2020-10-09 13:51:08
13110	0	NaN	2020-10-09 13:45:22
13111	0	5.21.3	2020-10-09 12:46:16
13112	0	5.21.3	2020-10-09 12:15:46
13113	0	NaN	2020-10-09 12:09:14

	replyContent	repliedAt \
0	Hi! Thank you for reviewing. Kindly note that ...	2020-09-18 11:58:46
1	Hi, Thank you for reviewing and sharing your p...	2020-09-30 16:01:39
2	Hi! We apologize for the inconvenience. We wou...	2020-09-17 17:59:10
3	Hi, Thank you for reviewing. Please be apprise...	2020-10-10 10:35:40
4	Hi! Thank you for reviewing. Kindly note that ...	2020-09-21 10:46:15
...	...	...
13109	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:53
13110	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:52
13111	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:47
13112	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:48
13113	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:42

	sortOrder	appId	thumbsUpCategory
0	most_relevant	com.whitehatjr	NaN
1	most_relevant	com.whitehatjr	101-500
2	most_relevant	com.whitehatjr	101-500
3	most_relevant	com.whitehatjr	51-100
4	most_relevant	com.whitehatjr	101-500
...	...	...	...
13109	newest	com.testbook.tbapp	0-10

13110	newest	com.testbook.tbapp	0-10
13111	newest	com.testbook.tbapp	0-10
13112	newest	com.testbook.tbapp	0-10
13113	newest	com.testbook.tbapp	0-10

[13114 rows x 12 columns]

[68]: `print(DATA)`

	name	userName	\
0	WhiteHatJr	All in one Captions	
1	WhiteHatJr	githin george	
2	WhiteHatJr	Anita Choudhury	
3	WhiteHatJr	Snehil gupta	
4	WhiteHatJr	UTKALIKA PRUSTI	
...	...	...	
13109	TestBook	vivek anand Mishra	
13110	TestBook	Raushan singh Rathod	
13111	TestBook	IPL 2020 live	
13112	TestBook	Saurabh Rathod	
13113	TestBook	Daanish King	

	content	score	\
0	I was very unhappy with this app because I can...	1	
1	Come on! Man it's not normal for kids to devel...	1	
2	This app is good for coding but I think this o...	1	
3	This app is good for coding but I think this o...	1	
4	It is good app . But it's app use only who has...	1	
...	...	...	
13109	Good app.	5	
13110	Good	5	
13111	This is awesome app for ssc all exams because ...	5	
13112	It is very useful for practicing	5	
13113	Best I love this	5	

	thumbsUpCount	reviewCreatedVersion	at	\
0	1234	1.5.1	2020-09-18 09:21:07	
1	213	1.5.2	2020-09-30 14:34:38	
2	156	1.5.1	2020-09-17 17:50:57	
3	82	1.5.2	2020-10-09 21:10:06	
4	150	1.5.1	2020-09-21 05:27:50	
...	...	...	...	
13109	0	5.21.3	2020-10-09 13:51:08	
13110	0	NaN	2020-10-09 13:45:22	
13111	0	5.21.3	2020-10-09 12:46:16	
13112	0	5.21.3	2020-10-09 12:15:46	
13113	0	NaN	2020-10-09 12:09:14	

		replyContent	repliedAt \
0	Hi! Thank you for reviewing. Kindly note that ...		2020-09-18 11:58:46
1	Hi, Thank you for reviewing and sharing your p...		2020-09-30 16:01:39
2	Hi! We apologize for the inconvenience. We wou...		2020-09-17 17:59:10
3	Hi, Thank you for reviewing. Please be apprise...		2020-10-10 10:35:40
4	Hi! Thank you for reviewing. Kindly note that ...		2020-09-21 10:46:15
...		...	...
13109	Hello, Thanks for your review. We are glad tha...		2020-10-11 06:52:53
13110	Hello, Thanks for your kind words and review. ...		2020-10-11 06:52:52
13111	Hello, Thanks for your review. We are glad tha...		2020-10-11 06:52:47
13112	Hello, Thanks for your kind words and review. ...		2020-10-11 06:52:48
13113	Hello, Thanks for your review. We are glad tha...		2020-10-11 06:52:42

	sortOrder	appId	thumbsUpCategory
0	most_relevant	com.whitehatjr	NaN
1	most_relevant	com.whitehatjr	101-500
2	most_relevant	com.whitehatjr	101-500
3	most_relevant	com.whitehatjr	51-100
4	most_relevant	com.whitehatjr	101-500
...	...	...	...
13109	newest	com.testbook.tbapp	0-10
13110	newest	com.testbook.tbapp	0-10
13111	newest	com.testbook.tbapp	0-10
13112	newest	com.testbook.tbapp	0-10
13113	newest	com.testbook.tbapp	0-10

[13114 rows x 12 columns]

```
[69]: DATA = DATA.drop(columns=['thumbsUpCount'])
DATA.tail(5)
```

```
[69]:
```

	name	userName \
13109	TestBook	vivek anand Mishra
13110	TestBook	Raushan singh Rathod
13111	TestBook	IPL 2020 live
13112	TestBook	Saurabh Rathod
13113	TestBook	Daanish King

	content	score \
13109	Good app.	5
13110	Good	5
13111	This is awesome app for ssc all exams because ...	5
13112	It is very useful for practicing	5
13113	Best I love this	5

	reviewCreatedVersion	at \
13109	5.21.3	2020-10-09 13:51:08

```

13110      NaN  2020-10-09 13:45:22
13111      5.21.3  2020-10-09 12:46:16
13112      5.21.3  2020-10-09 12:15:46
13113      NaN  2020-10-09 12:09:14

```

```

                                replyContent      repliedAt \
13109  Hello, Thanks for your review. We are glad tha...  2020-10-11 06:52:53
13110  Hello, Thanks for your kind words and review. ...  2020-10-11 06:52:52
13111  Hello, Thanks for your review. We are glad tha...  2020-10-11 06:52:47
13112  Hello, Thanks for your kind words and review. ...  2020-10-11 06:52:48
13113  Hello, Thanks for your review. We are glad tha...  2020-10-11 06:52:42

```

```

      sortOrder      appId thumbsUpCategory
13109    newest  com.testbook.tbapp          0-10
13110    newest  com.testbook.tbapp          0-10
13111    newest  com.testbook.tbapp          0-10
13112    newest  com.testbook.tbapp          0-10
13113    newest  com.testbook.tbapp          0-10

```

```
[70]: DATA=DATA.reset_index()
```

```
[71]: DATA
```

```

[71]:      index      name      userName \
0         0  WhiteHatJr  All in one Captions
1         1  WhiteHatJr      githin george
2         2  WhiteHatJr  Anita Choudhury
3         3  WhiteHatJr      Snehil gupta
4         4  WhiteHatJr  UTKALIKA PRUSTI
...      ...      ...      ...
13109  13109    TestBook  vivek anand Mishra
13110  13110    TestBook  Raushan singh Rathod
13111  13111    TestBook      IPL 2020 live
13112  13112    TestBook  Saurabh Rathod
13113  13113    TestBook    Daanish King

```

```

                                content  score \
0      I was very unhappy with this app because I can...      1
1      Come on! Man it's not normal for kids to devel...      1
2      This app is good for coding but I think this o...      1
3      This app is good for coding but I think this o...      1
4      It is good app . But it's app use only who has...      1
...      ...      ...      ...
13109                                Good app.      5
13110                                Good      5
13111  This is awesome app for ssc all exams because ...      5
13112                                It is very useful for practicing      5

```

13113 Best I love this 5

	reviewCreatedVersion	at \
0	1.5.1	2020-09-18 09:21:07
1	1.5.2	2020-09-30 14:34:38
2	1.5.1	2020-09-17 17:50:57
3	1.5.2	2020-10-09 21:10:06
4	1.5.1	2020-09-21 05:27:50
...	...	...
13109	5.21.3	2020-10-09 13:51:08
13110	NaN	2020-10-09 13:45:22
13111	5.21.3	2020-10-09 12:46:16
13112	5.21.3	2020-10-09 12:15:46
13113	NaN	2020-10-09 12:09:14

	replyContent	repliedAt \
0	Hi! Thank you for reviewing. Kindly note that ...	2020-09-18 11:58:46
1	Hi, Thank you for reviewing and sharing your p...	2020-09-30 16:01:39
2	Hi! We apologize for the inconvenience. We wou...	2020-09-17 17:59:10
3	Hi, Thank you for reviewing. Please be apprise...	2020-10-10 10:35:40
4	Hi! Thank you for reviewing. Kindly note that ...	2020-09-21 10:46:15
...	...	...
13109	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:53
13110	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:52
13111	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:47
13112	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:48
13113	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:42

	sortOrder	appId	thumbsUpCategory
0	most_relevant	com.whitehatjr	NaN
1	most_relevant	com.whitehatjr	101-500
2	most_relevant	com.whitehatjr	101-500
3	most_relevant	com.whitehatjr	51-100
4	most_relevant	com.whitehatjr	101-500
...	...	...	...
13109	newest	com.testbook.tbapp	0-10
13110	newest	com.testbook.tbapp	0-10
13111	newest	com.testbook.tbapp	0-10
13112	newest	com.testbook.tbapp	0-10
13113	newest	com.testbook.tbapp	0-10

[13114 rows x 12 columns]

```
[72]: DATA['reviewCreatedVersion'].fillna("Unknown", inplace=True)

print(DATA)
```

	index	name	userName \
0	0	WhiteHatJr	All in one Captions
1	1	WhiteHatJr	githin george
2	2	WhiteHatJr	Anita Choudhury
3	3	WhiteHatJr	Snehil gupta
4	4	WhiteHatJr	UTKALIKA PRUSTI
...	...	...	...
13109	13109	TestBook	vivek anand Mishra
13110	13110	TestBook	Raushan singh Rathod
13111	13111	TestBook	IPL 2020 live
13112	13112	TestBook	Saurabh Rathod
13113	13113	TestBook	Daanish King

	content	score \
0	I was very unhappy with this app because I can...	1
1	Come on! Man it's not normal for kids to devel...	1
2	This app is good for coding but I think this o...	1
3	This app is good for coding but I think this o...	1
4	It is good app . But it's app use only who has...	1
...	...	...
13109	Good app.	5
13110	Good	5
13111	This is awesome app for ssc all exams because ...	5
13112	It is very useful for practicing	5
13113	Best I love this	5

	reviewCreatedVersion	at \
0	1.5.1	2020-09-18 09:21:07
1	1.5.2	2020-09-30 14:34:38
2	1.5.1	2020-09-17 17:50:57
3	1.5.2	2020-10-09 21:10:06
4	1.5.1	2020-09-21 05:27:50
...	...	...
13109	5.21.3	2020-10-09 13:51:08
13110	Unknown	2020-10-09 13:45:22
13111	5.21.3	2020-10-09 12:46:16
13112	5.21.3	2020-10-09 12:15:46
13113	Unknown	2020-10-09 12:09:14

	replyContent	repliedAt \
0	Hi! Thank you for reviewing. Kindly note that ...	2020-09-18 11:58:46
1	Hi, Thank you for reviewing and sharing your p...	2020-09-30 16:01:39
2	Hi! We apologize for the inconvenience. We wou...	2020-09-17 17:59:10
3	Hi, Thank you for reviewing. Please be apprise...	2020-10-10 10:35:40
4	Hi! Thank you for reviewing. Kindly note that ...	2020-09-21 10:46:15
...	...	...
13109	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:53
13110	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:52



```

13111 Hello, Thanks for your review. We are glad tha... 2020-10-11 06:52:47
13112 Hello, Thanks for your kind words and review. ... 2020-10-11 06:52:48
13113 Hello, Thanks for your review. We are glad tha... 2020-10-11 06:52:42

```

```

      sortOrder      appId thumbsUpCategory
0  most_relevant  com.whitehatjr      NaN
1  most_relevant  com.whitehatjr    101-500
2  most_relevant  com.whitehatjr    101-500
3  most_relevant  com.whitehatjr     51-100
4  most_relevant  com.whitehatjr    101-500
...
13109      newest  com.testbook.tbapp      0-10
13110      newest  com.testbook.tbapp      0-10
13111      newest  com.testbook.tbapp      0-10
13112      newest  com.testbook.tbapp      0-10
13113      newest  com.testbook.tbapp      0-10

```

[13114 rows x 12 columns]

[73]: DATA

```

[73]:      index      name      userName \
0         0  WhiteHatJr  All in one Captions
1         1  WhiteHatJr      githin george
2         2  WhiteHatJr  Anita Choudhury
3         3  WhiteHatJr      Snehil gupta
4         4  WhiteHatJr  UTKALIKA PRUSTI
...
13109  13109    TestBook  vivek anand Mishra
13110  13110    TestBook  Raushan singh Rathod
13111  13111    TestBook      IPL 2020 live
13112  13112    TestBook  Saurabh Rathod
13113  13113    TestBook  Daanish King

```

```

      content  score \
0  I was very unhappy with this app because I can...      1
1  Come on! Man it's not normal for kids to devel...      1
2  This app is good for coding but I think this o...      1
3  This app is good for coding but I think this o...      1
4  It is good app . But it's app use only who has...      1
...
13109      Good app.      5
13110      Good      5
13111  This is awesome app for ssc all exams because ...      5
13112      It is very useful for practicing      5
13113      Best I love this      5

```

	reviewCreatedVersion	at	\
0	1.5.1	2020-09-18 09:21:07	
1	1.5.2	2020-09-30 14:34:38	
2	1.5.1	2020-09-17 17:50:57	
3	1.5.2	2020-10-09 21:10:06	
4	1.5.1	2020-09-21 05:27:50	
...	...	...	
13109	5.21.3	2020-10-09 13:51:08	
13110	Unknown	2020-10-09 13:45:22	
13111	5.21.3	2020-10-09 12:46:16	
13112	5.21.3	2020-10-09 12:15:46	
13113	Unknown	2020-10-09 12:09:14	

	replyContent	repliedAt	\
0	Hi! Thank you for reviewing. Kindly note that ...	2020-09-18 11:58:46	
1	Hi, Thank you for reviewing and sharing your p...	2020-09-30 16:01:39	
2	Hi! We apologize for the inconvenience. We wou...	2020-09-17 17:59:10	
3	Hi, Thank you for reviewing. Please be apprise...	2020-10-10 10:35:40	
4	Hi! Thank you for reviewing. Kindly note that ...	2020-09-21 10:46:15	
...	...	...	
13109	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:53	
13110	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:52	
13111	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:47	
13112	Hello, Thanks for your kind words and review. ...	2020-10-11 06:52:48	
13113	Hello, Thanks for your review. We are glad tha...	2020-10-11 06:52:42	

	sortOrder	appId	thumbsUpCategory
0	most_relevant	com.whitehatjr	NaN
1	most_relevant	com.whitehatjr	101-500
2	most_relevant	com.whitehatjr	101-500
3	most_relevant	com.whitehatjr	51-100
4	most_relevant	com.whitehatjr	101-500
...	...	...	...
13109	newest	com.testbook.tbapp	0-10
13110	newest	com.testbook.tbapp	0-10
13111	newest	com.testbook.tbapp	0-10
13112	newest	com.testbook.tbapp	0-10
13113	newest	com.testbook.tbapp	0-10

[13114 rows x 12 columns]

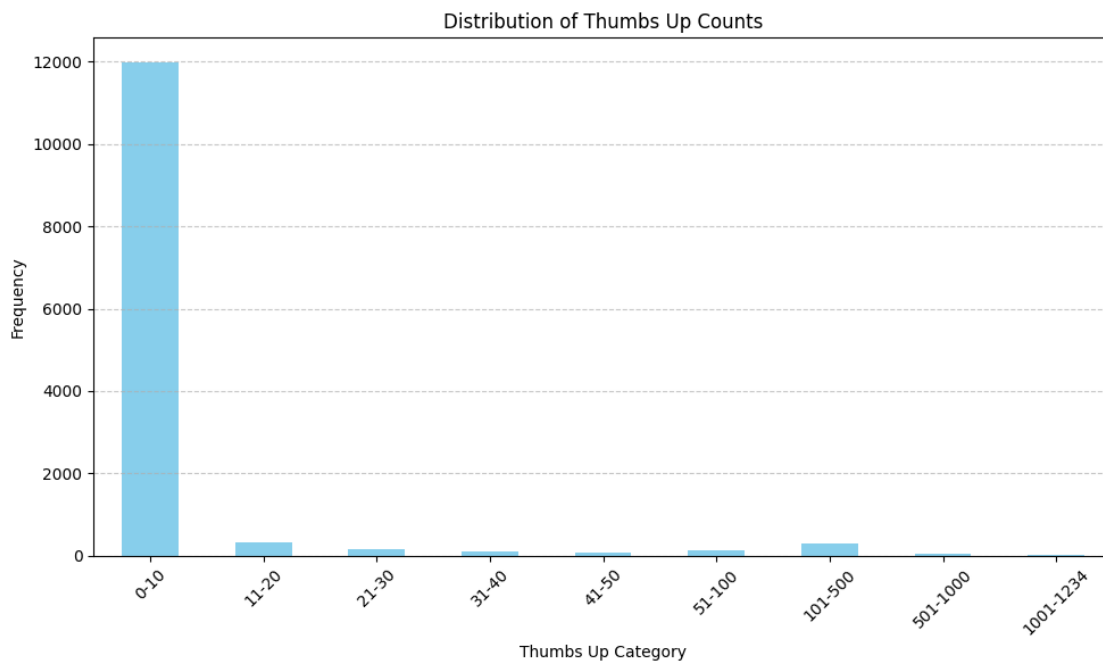
```
[74]: for col in DATA.columns:
      print(col)
```

```
index
name
userName
```

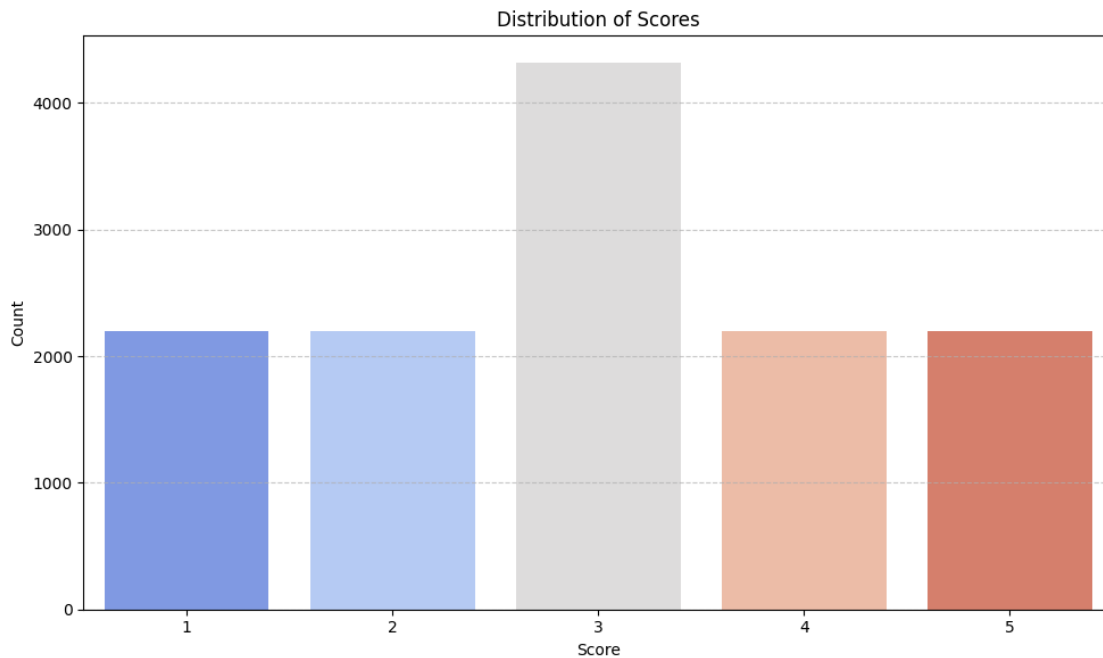
```
content
score
reviewCreatedVersion
at
replyContent
repliedAt
sortOrder
appId
thumbsUpCategory
```

```
[75]: # Assuming DATA is your DataFrame with 'thumbsUpCategory' column
# Count the frequency of each category
category_counts = DATA['thumbsUpCategory'].value_counts().sort_index()

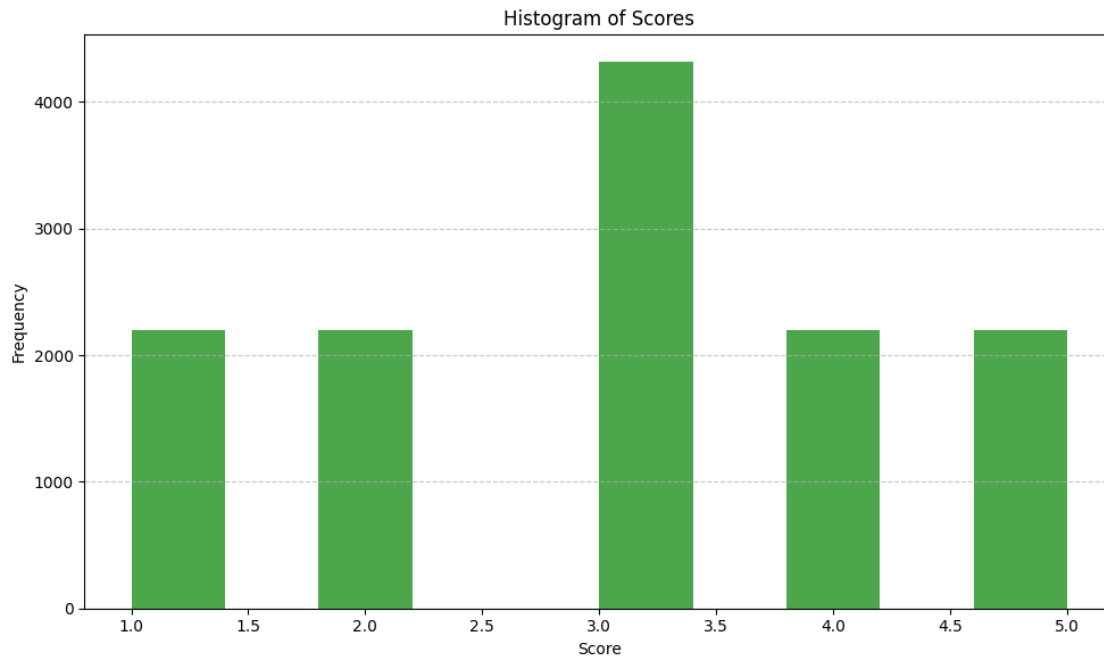
# Plotting
plt.figure(figsize=(10, 6))
category_counts.plot(kind='bar', color='skyblue')
plt.title('Distribution of Thumbs Up Counts')
plt.xlabel('Thumbs Up Category')
plt.ylabel('Frequency')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



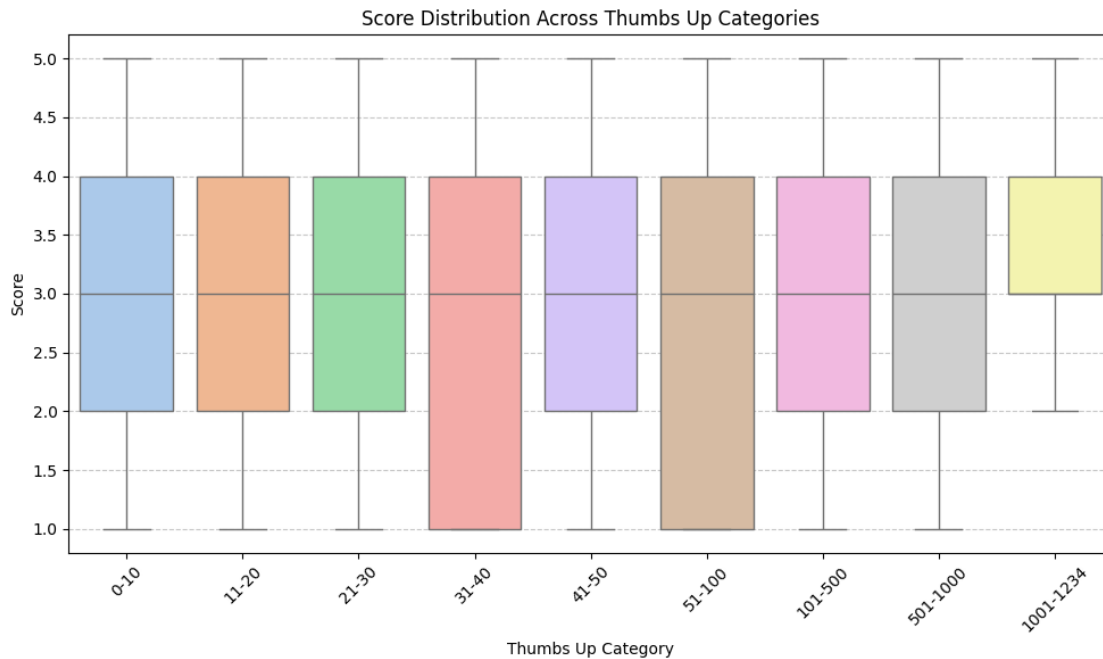
```
[78]: plt.figure(figsize=(10, 6))
sns.countplot(data=DATA, x='score', palette='coolwarm')
plt.title('Distribution of Scores')
plt.xlabel('Score')
plt.ylabel('Count')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



```
[79]: plt.figure(figsize=(10, 6))
plt.hist(DATA['score'], bins=10, color='green', alpha=0.7)
plt.title('Histogram of Scores')
plt.xlabel('Score')
plt.ylabel('Frequency')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



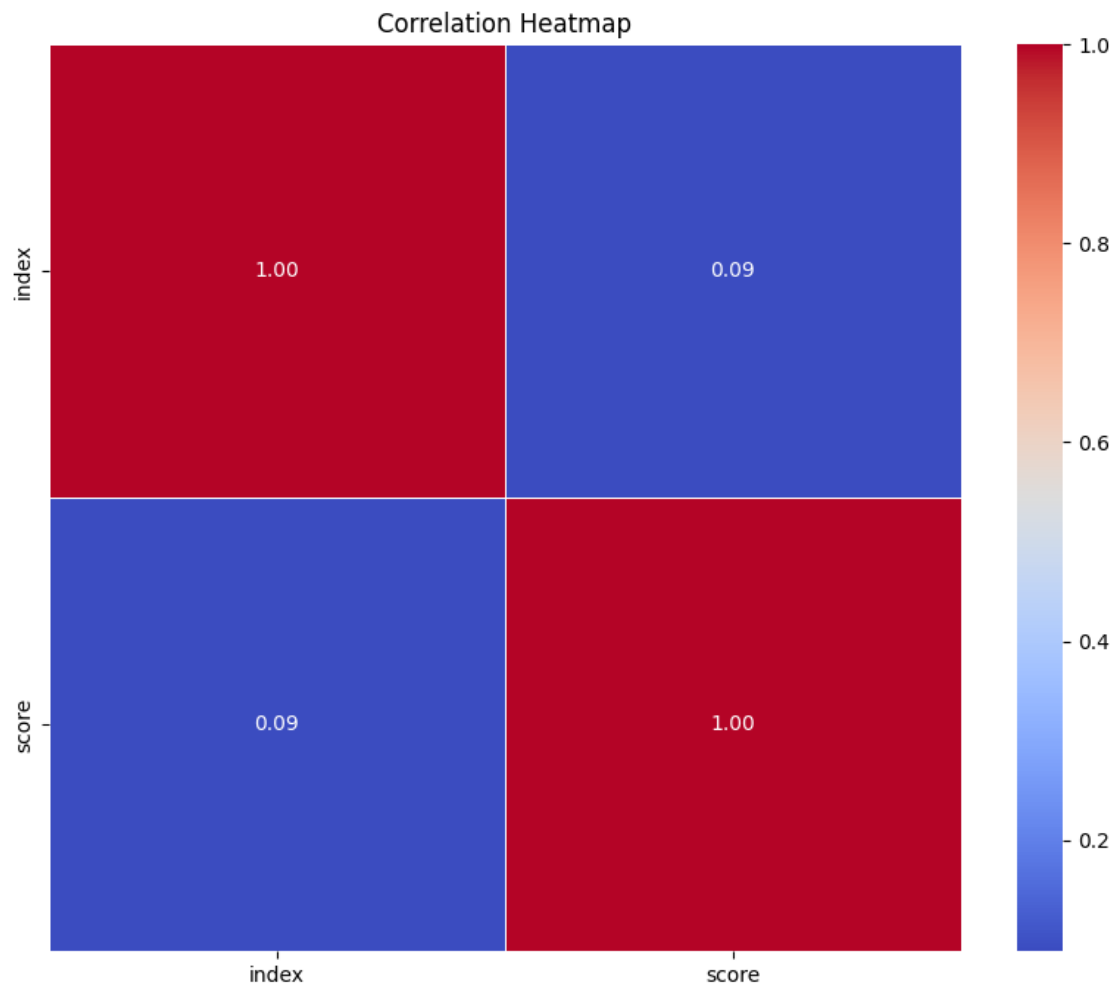
```
[80]: plt.figure(figsize=(10, 6))
sns.boxplot(data=DATA, x='thumbsUpCategory', y='score', palette='pastel')
plt.title('Score Distribution Across Thumbs Up Categories')
plt.xlabel('Thumbs Up Category')
plt.ylabel('Score')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



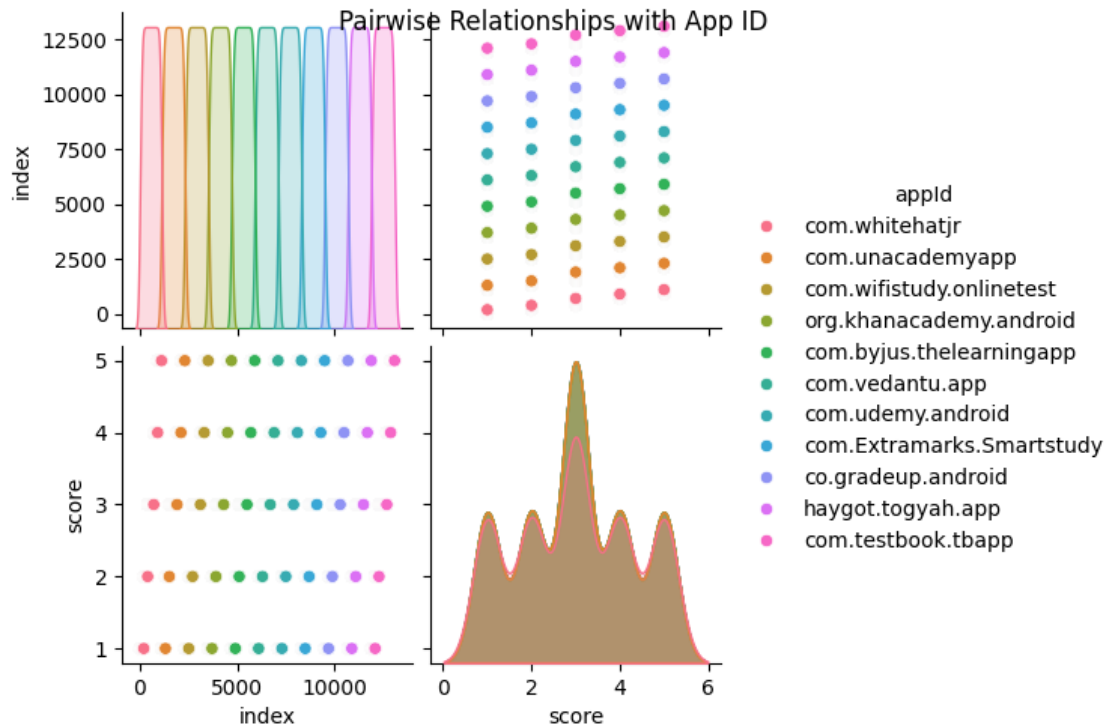
```
[82]: # Select only numeric columns
numeric_columns = DATA.select_dtypes(include=['number'])

# Compute the correlation matrix
correlation_matrix = numeric_columns.corr()

# Plot the correlation heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f',
            linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()
```

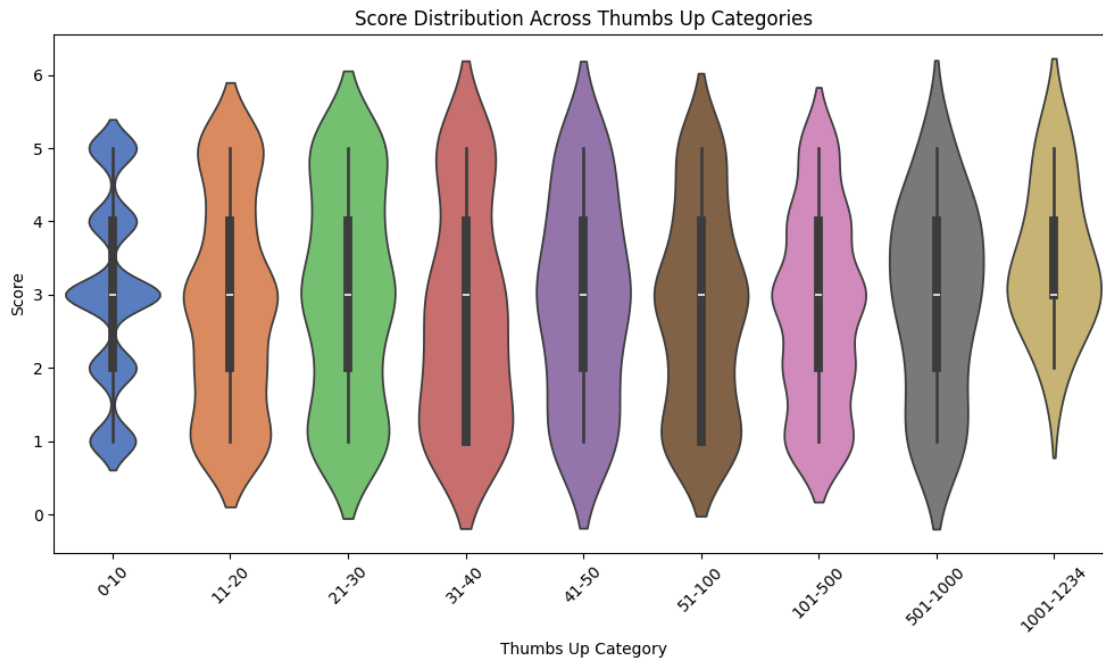


```
[86]: sns.pairplot(DATA, hue='appId', diag_kind='kde', height=2.5)
plt.suptitle('Pairwise Relationships with App ID')
plt.show()
```



```
[88]: plt.figure(figsize=(10, 6))
sns.violinplot(data=DATA, x='thumbsUpCategory', y='score', palette='muted')
plt.title('Score Distribution Across Thumbs Up Categories')
plt.xlabel('Thumbs Up Category')
plt.ylabel('Score')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```





```
[90]: from wordcloud import WordCloud

# Generate word cloud
wordcloud = WordCloud(width=800, height=400, background_color='white').
    generate(DATA['content'].iloc[0])

# Display the word cloud
plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.title('Word Cloud for Content')
plt.axis('off')
plt.show()
```

Word Cloud for Content



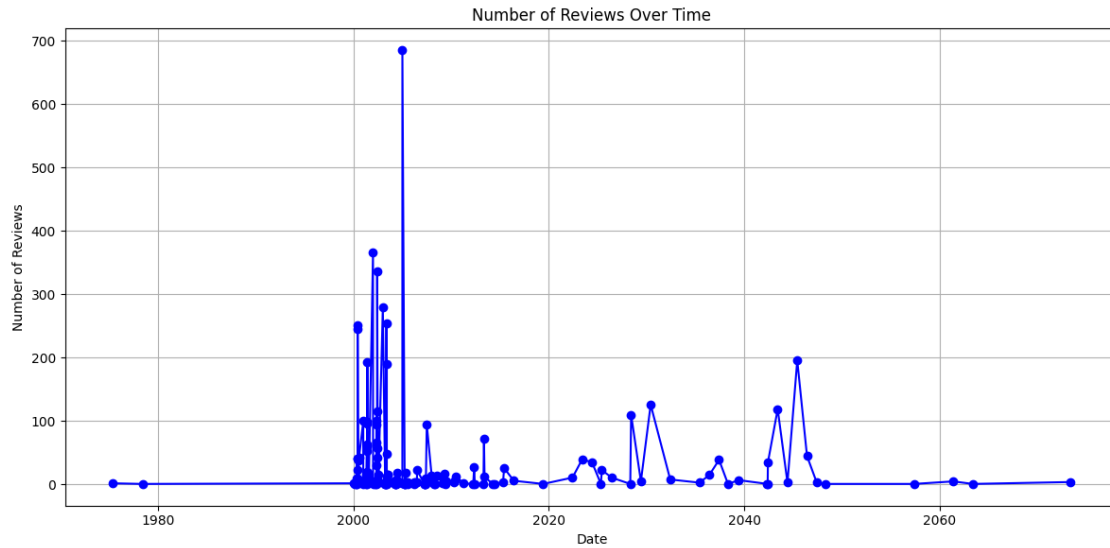
```
[95]: # Convert 'reviewCreatedVersion' to datetime format
DATA['reviewCreatedVersion'] = pd.to_datetime(DATA['reviewCreatedVersion'],
errors='coerce')

# Drop rows with NaT values (invalid dates)
DATA = DATA.dropna(subset=['reviewCreatedVersion'])

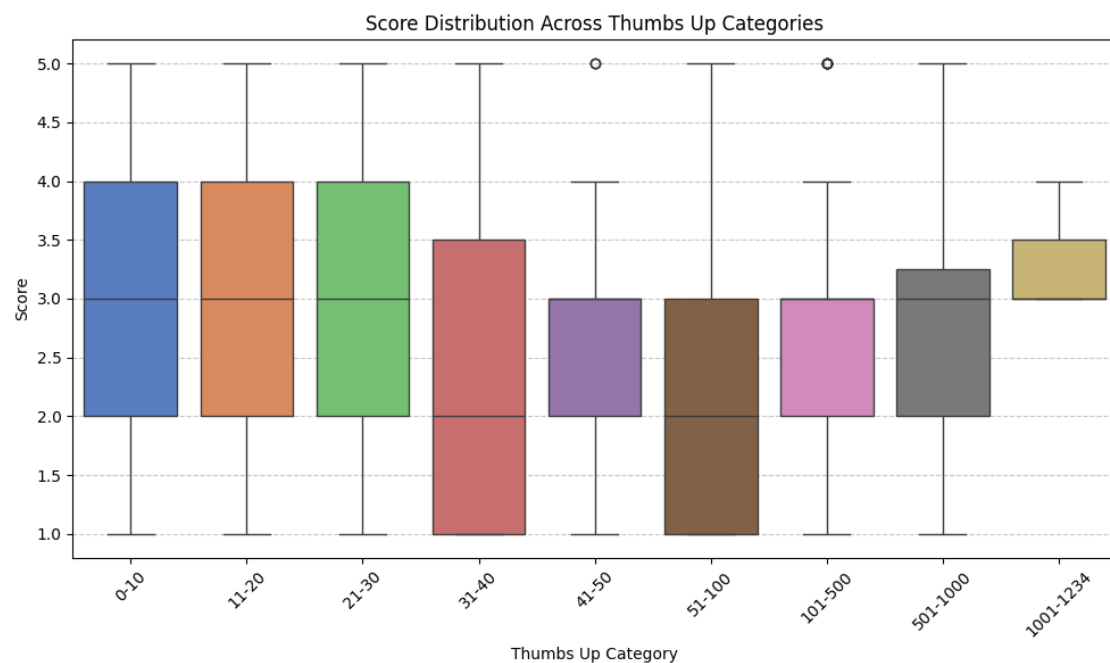
# Extract date from 'reviewCreatedVersion'
DATA['reviewDate'] = DATA['reviewCreatedVersion'].dt.date

# Count number of reviews per date
review_counts = DATA.groupby('reviewDate').size()

# Plot time series
plt.figure(figsize=(12, 6))
review_counts.plot(kind='line', marker='o', color='b')
plt.title('Number of Reviews Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Reviews')
plt.grid(True)
plt.tight_layout()
plt.show()
```



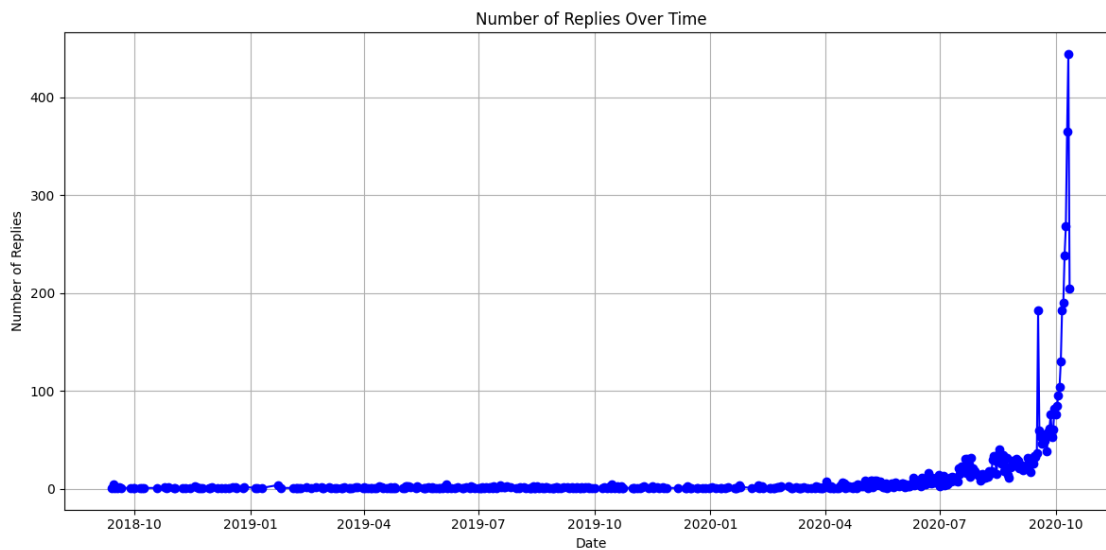
```
[97]: plt.figure(figsize=(10, 6))
sns.boxplot(data=DATA, x='thumbsUpCategory', y='score', palette='muted')
plt.title('Score Distribution Across Thumbs Up Categories')
plt.xlabel('Thumbs Up Category')
plt.ylabel('Score')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



```
[101]: # Extract date from 'at' and 'repliedAt' columns
DATA['reviewDate'] = DATA['at'].dt.date
DATA['replyDate'] = DATA['repliedAt'].dt.date

# Count number of replies per review date
reply_counts = DATA.groupby('reviewDate').size()

# Plot time series of reply counts
plt.figure(figsize=(12, 6))
reply_counts.plot(kind='line', marker='o', color='b')
plt.title('Number of Replies Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Replies')
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
[104]: # Calculate quartiles of response time
response_time_quartiles = DATA['responseTime'].quantile([0.25, 0.5, 0.75])

# Create function to categorize response time
def categorize_response_time(response_time):
    if response_time <= response_time_quartiles[0.25]:
        return 'fast'
    elif response_time <= response_time_quartiles[0.5]:
        return 'medium'
```

```

    else:
        return 'slow'

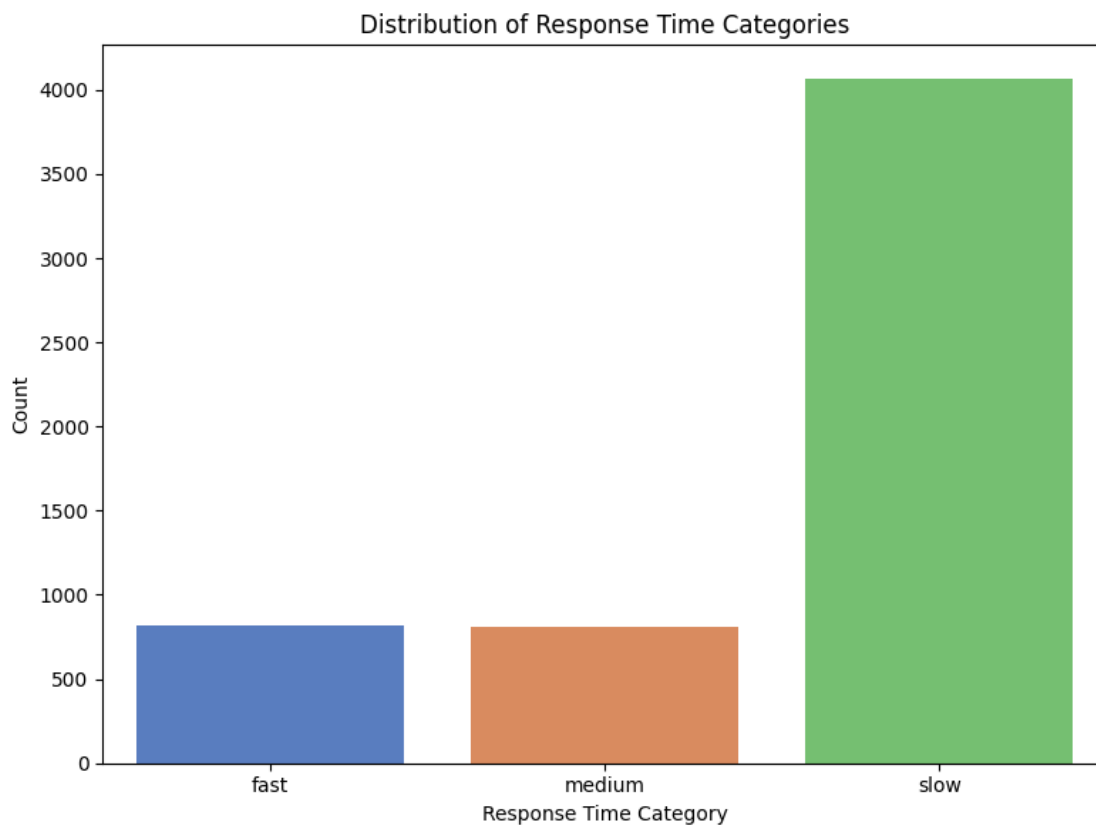
# Apply function to create response time category column
DATA['responseTimeCategory'] = DATA['responseTime'].
    ↪ apply(categorize_response_time)

```

```

[105]: # Plot count of response time categories
plt.figure(figsize=(8, 6))
sns.countplot(data=DATA, x='responseTimeCategory', palette='muted')
plt.title('Distribution of Response Time Categories')
plt.xlabel('Response Time Category')
plt.ylabel('Count')
plt.tight_layout()
plt.show()

```



```

[106]: from textblob import TextBlob

# Function to classify sentiment
def get_sentiment(text):
    analysis = TextBlob(text)

```

```

if analysis.sentiment.polarity > 0:
    return 'positive'
elif analysis.sentiment.polarity < 0:
    return 'negative'
else:
    return 'neutral'

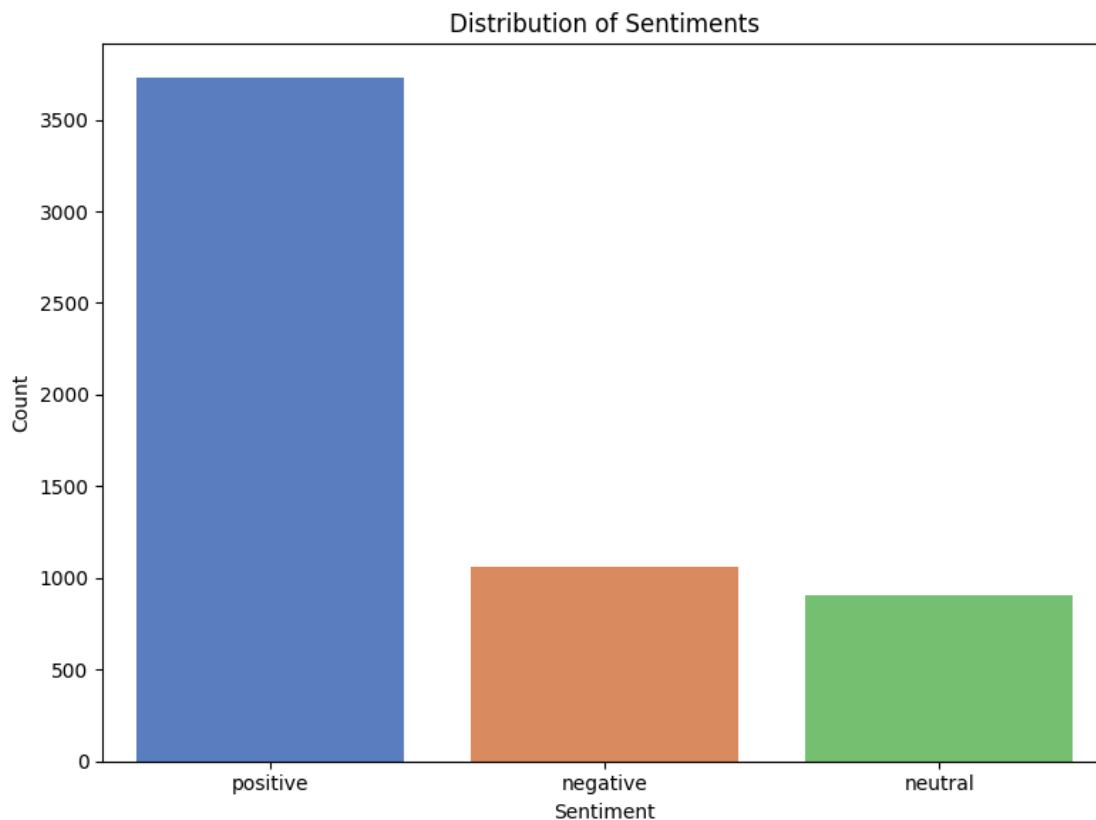
# Apply sentiment analysis to 'content' column
DATA['sentiment'] = DATA['content'].apply(get_sentiment)

```

```

[107]: # Plot count of sentiment categories
plt.figure(figsize=(8, 6))
sns.countplot(data=DATA, x='sentiment', palette='muted')
plt.title('Distribution of Sentiments')
plt.xlabel('Sentiment')
plt.ylabel('Count')
plt.tight_layout()
plt.show()

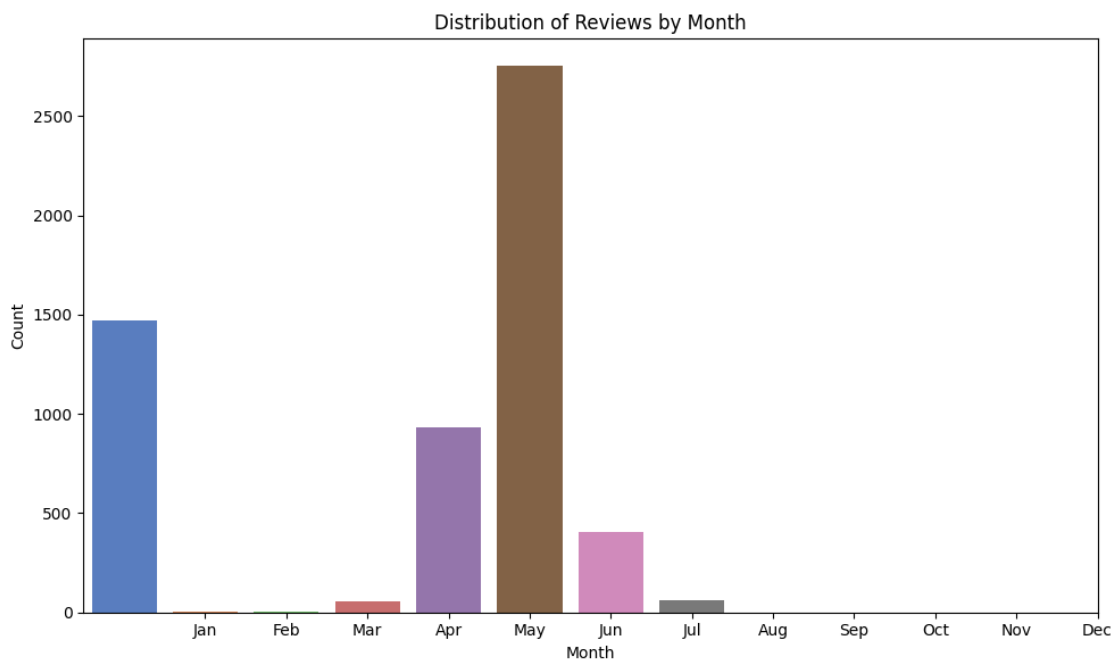
```



```
[108]: # Convert 'reviewCreatedVersion' to datetime format
DATA['reviewCreatedVersion'] = pd.to_datetime(DATA['reviewCreatedVersion'])

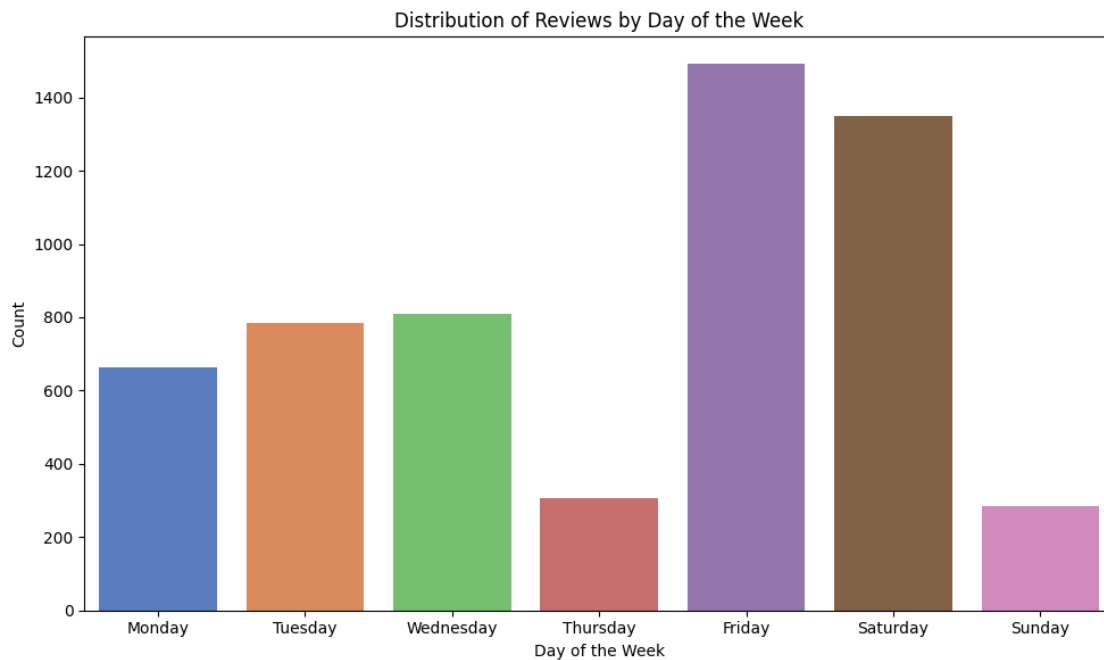
# Extract month and day of the week
DATA['month'] = DATA['reviewCreatedVersion'].dt.month
DATA['day_of_week'] = DATA['reviewCreatedVersion'].dt.day_name()
```

```
[109]: # Plot count of reviews by month
plt.figure(figsize=(10, 6))
sns.countplot(data=DATA, x='month', palette='muted')
plt.title('Distribution of Reviews by Month')
plt.xlabel('Month')
plt.ylabel('Count')
plt.xticks(ticks=range(1, 13), labels=['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])
plt.tight_layout()
plt.show()
```



```
[110]: # Plot count of reviews by day of the week
plt.figure(figsize=(10, 6))
sns.countplot(data=DATA, x='day_of_week', palette='muted', order=['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'])
plt.title('Distribution of Reviews by Day of the Week')
plt.xlabel('Day of the Week')
plt.ylabel('Count')
```

```
plt.tight_layout()
plt.show()
```



```
[115]: DATA2= pd.read_csv('app_info.csv')
DATA2
```

```
[115]:
```

	name	title \
0	WhiteHatJr	WhiteHat Jr: Book FREE Trial Class Now
1	Unacademy	Unacademy Learning App
2	WifiStudy	wifistudy - #1 Exam Preparation, Free Mock Tests
3	KhanAcademy	Khan Academy
4	Byjus	BYJU'S - The Learning App
5	Vedantu	Vedantu: LIVE Learning App   Class 1-12, JEE, ...
6	Udemy	Udemy - Online Courses
7	Extramarks	Extramarks - The Learning App
8	Gradeup	Exam Preparation App: Free Mock Test, Live Cla...
9	Toppr	Toppr - Learning App for Class 5 - 12
10	TestBook	Exam Preparation App: Free Mock Tests   Live C...

	minInstalls	score	ratings	reviews	size \
0	1000000	3.723704	12682	6482	32M
1	10000000	4.194891	556461	192168	Varies with device
2	1000000	4.309553	137081	61816	Varies with device
3	10000000	4.397173	133720	48253	Varies with device
4	50000000	4.440033	1436720	672082	87M



5	10000000	4.219771	232650	117145	Varies with device
6	10000000	4.365889	251093	87319	35M
7	5000000	4.174786	114000	62232	65M
8	10000000	4.554652	618801	211413	20M
9	10000000	4.005949	145027	66629	25M
10	5000000	4.396316	176892	68824	Varies with device

	androidVersion	genreId	released	updated	version \
0	5.0	EDUCATION	Sep 11, 2020	1601010278	1.5.2
1	4.3	EDUCATION	Feb 15, 2017	1602257855	5.27.1040
2	4.2	EDUCATION	Dec 11, 2014	1599554732	Varies with device
3	Varies	EDUCATION	Sep 22, 2015	1601414820	Varies with device
4	5.0	EDUCATION	Jul 6, 2015	1600688970	7.1.0.9005
5	5.0	EDUCATION	Mar 8, 2019	1601485384	1.7.5
6	6.0	EDUCATION	Jan 14, 2014	1602426634	6.9.1
7	4.4	EDUCATION	Sep 5, 2015	1602432069	6.4.3.11
8	5.0	EDUCATION	Sep 6, 2015	1601619074	9.53
9	4.2	EDUCATION	Aug 28, 2013	1602263809	6.5.45
10	Varies	EDUCATION	Jun 29, 2015	1602249153	Varies with device

	appId
0	com.whitehatjr
1	com.unacademyapp
2	com.wifistudy.onlinetest
3	org.khanacademy.android
4	com.byjus.thelearningapp
5	com.vedantu.app
6	com.udemy.android
7	com.Extramarks.Smartstudy
8	co.gradeup.android
9	haygot.togyah.app
10	com.testbook.tbapp

```
[116]: # Set up the visualization style
sns.set(style="whitegrid")

# Visualization 1: Distribution of App Ratings
plt.figure(figsize=(10, 6))
sns.histplot(DATA2['score'], kde=True, color='skyblue')
plt.title('Distribution of App Ratings')
plt.xlabel('Rating')
plt.ylabel('Frequency')
plt.show()

# Visualization 2: Number of Installs by App Genre
plt.figure(figsize=(12, 6))
```

```

sns.barplot(data=DATA2, x='genreId', y='minInstalls', estimator=sum, ci=None,
            palette='muted')
plt.title('Number of Installs by App Genre')
plt.xlabel('Genre')
plt.ylabel('Total Installs')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Visualization 3: App Size Distribution
plt.figure(figsize=(10, 6))
sns.histplot(DATA2['size'], kde=True, color='lightgreen')
plt.title('Distribution of App Sizes')
plt.xlabel('App Size')
plt.ylabel('Frequency')
plt.show()

# Visualization 4: Release Dates of Apps
DATA2['released'] = pd.to_datetime(DATA2['released'])
plt.figure(figsize=(12, 6))
sns.histplot(DATA2['released'], kde=False, color='salmon')
plt.title('Release Dates of Apps')
plt.xlabel('Release Date')
plt.ylabel('Number of Apps')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Visualization 5: Relationship between Ratings and Reviews
plt.figure(figsize=(10, 6))
sns.scatterplot(data=DATA2, x='ratings', y='reviews', color='orange', alpha=0.7)
plt.title('Relationship between Ratings and Reviews')
plt.xlabel('Ratings')
plt.ylabel('Number of Reviews')
plt.show()

# Visualization 6: Version-wise Ratings
plt.figure(figsize=(12, 6))
sns.boxplot(data=DATA2, x='version', y='score', palette='muted')
plt.title('Version-wise Ratings')
plt.xlabel('App Version')
plt.ylabel('Rating')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Visualization 7: Android Version Compatibility

```

```

plt.figure(figsize=(12, 6))
sns.boxplot(data=DATA2, x='androidVersion', y='score', palette='pastel')
plt.title('Android Version Compatibility')
plt.xlabel('Android Version')
plt.ylabel('Rating')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

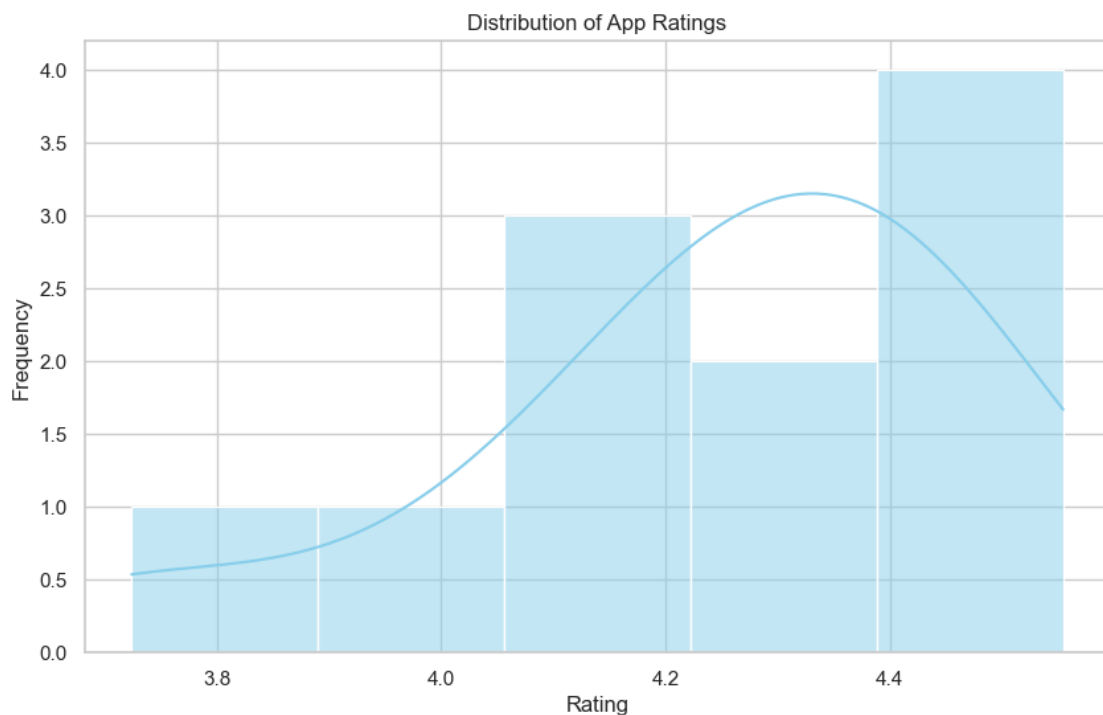
# Visualization 8: App Titles Word Cloud
from wordcloud import WordCloud

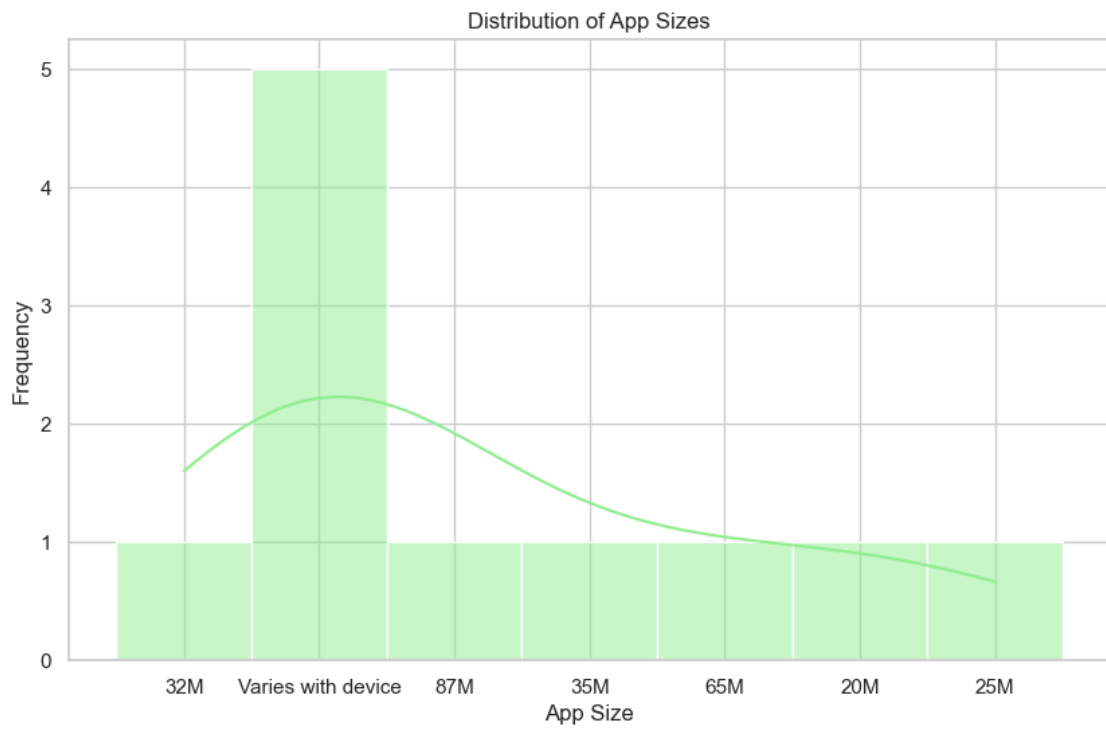
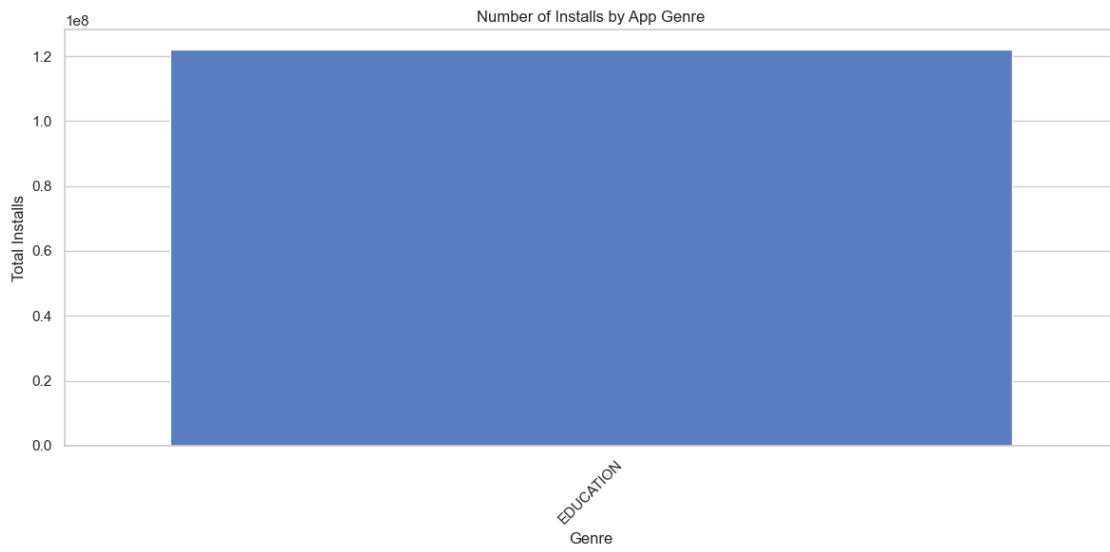
# Concatenate all app titles into a single string
app_titles_text = ' '.join(DATA2['title'])

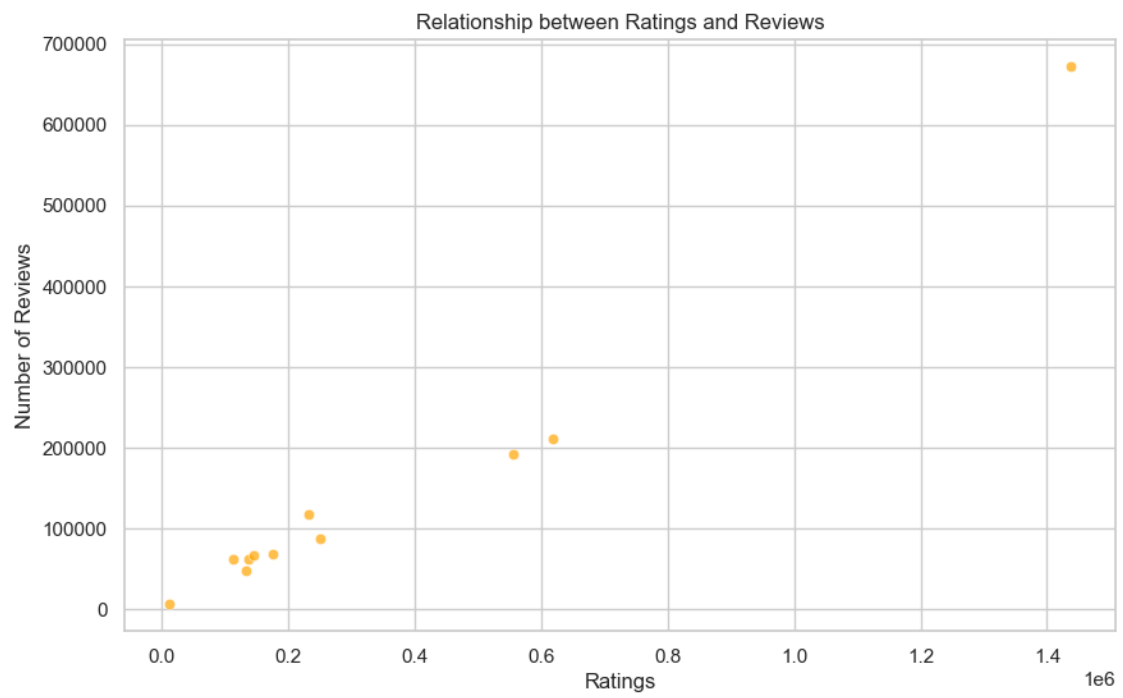
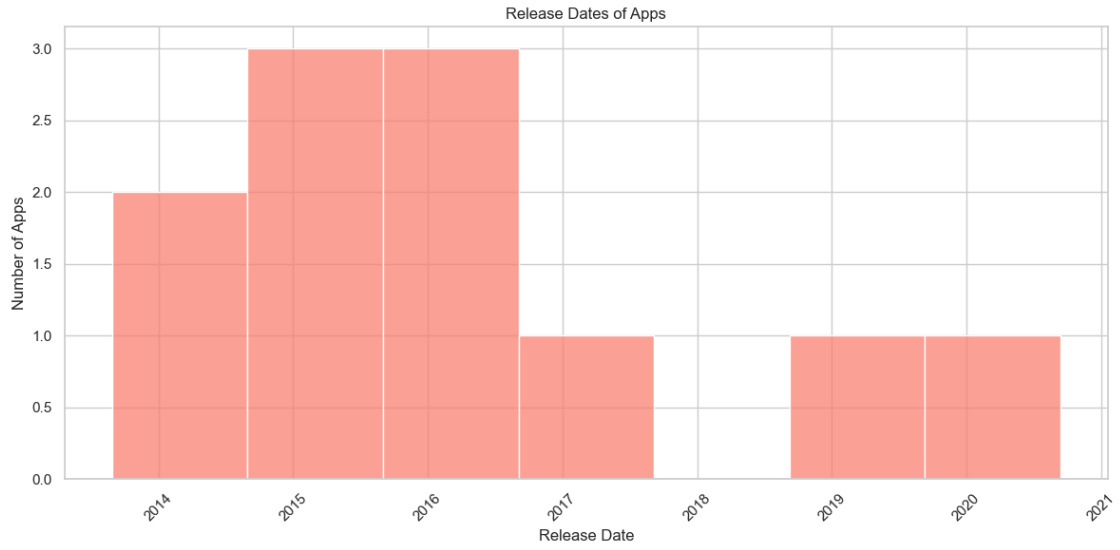
# Generate word cloud
wordcloud = WordCloud(width=800, height=400, background_color='white').
    generate(app_titles_text)

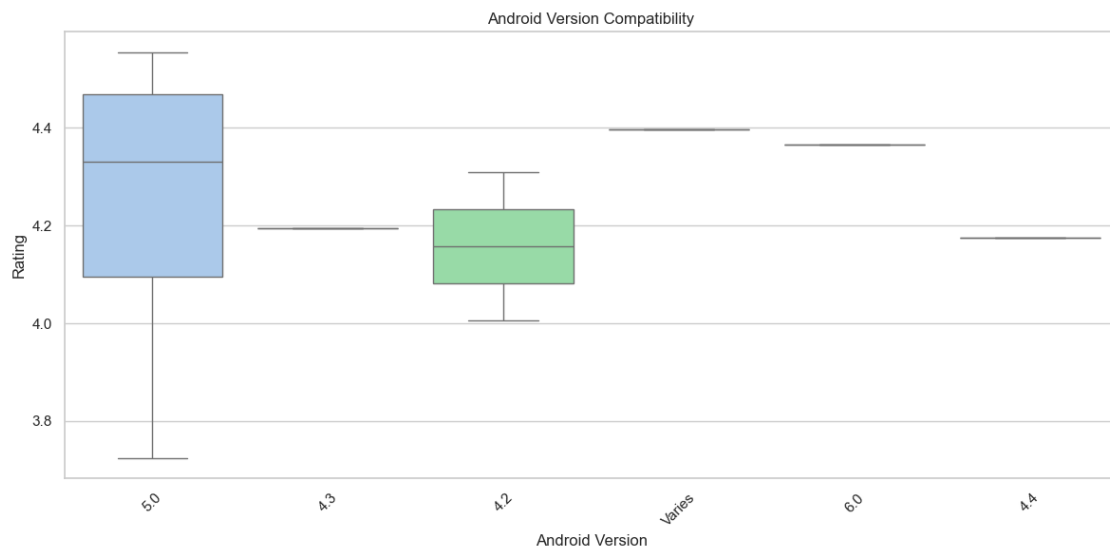
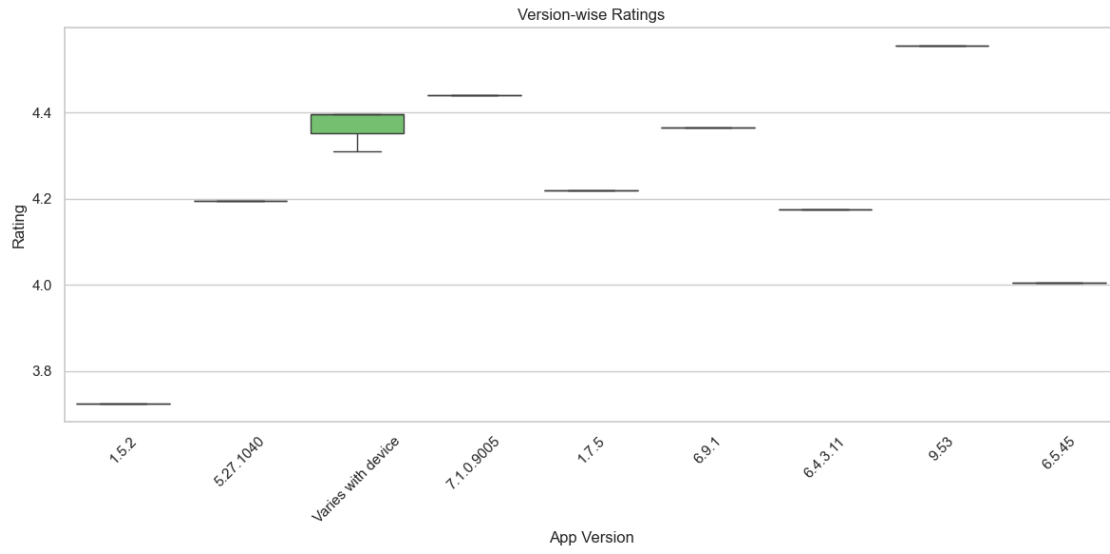
# Display the word cloud
plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.title('App Titles Word Cloud')
plt.axis('off')
plt.show()

```









WhiteHat Trial Toppr NEET Extramarks Khan Udem

# Learning

## Preparation

### App Class

### Exam

### Mock

### Test

### Now

### Courses

### Academy

### Jr

### JEE

### Unacademy

### Online

### Book

### Khan

### Udemy

### Wifistudy

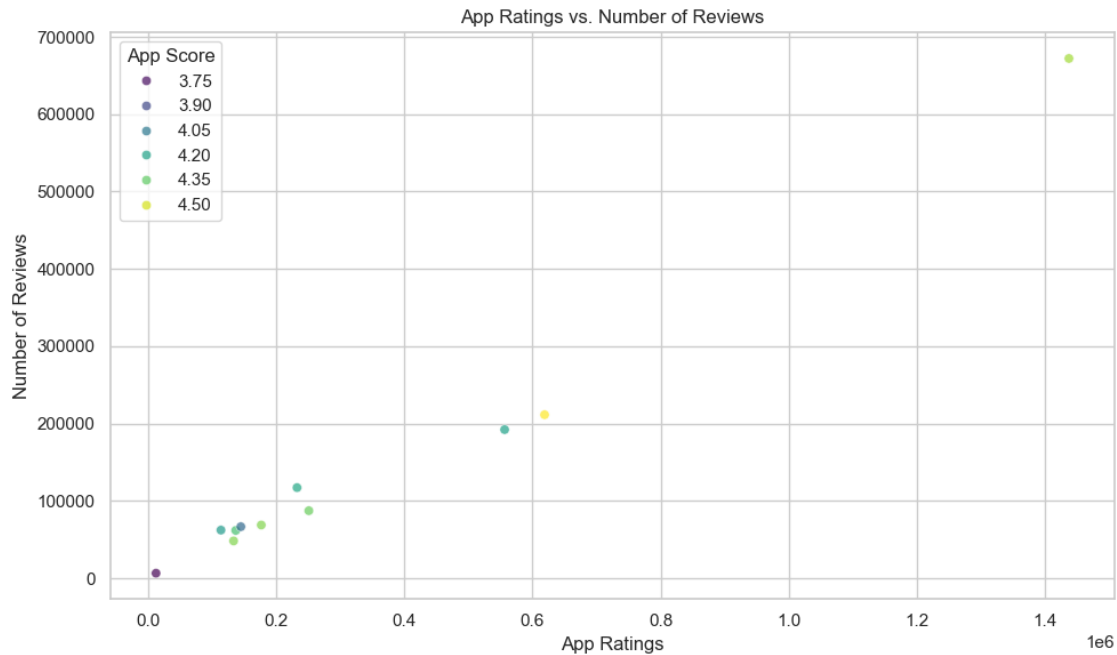
### Classes

BYJU Vedantu

```
# Visualization: Relationship between App Ratings and Number of Reviews from DATA
↳ DATA

plt.figure(figsize=(10, 6))
sns.scatterplot(data=DATA2, x='ratings', y='reviews', hue='score', ↳
↳ palette='viridis', alpha=0.7)

plt.title('App Ratings vs. Number of Reviews')
plt.xlabel('App Ratings')
plt.ylabel('Number of Reviews')
plt.legend(title='App Score')
plt.tight_layout()
plt.show()
```



```
[120]: # Visualization 1: App Ratings by Genre and Score from DATA2
plt.figure(figsize=(12, 6))
sns.barplot(data=DATA2, x='genreId', y='score', hue='ratings', palette='muted',
            ci=None)
plt.title('App Ratings by Genre and Score')
plt.xlabel('Genre')
plt.ylabel('Average Score')
plt.xticks(rotation=45)
plt.legend(title='App Ratings')
plt.tight_layout()
plt.show()

# Visualization 2: App Installs vs. Scores
plt.figure(figsize=(10, 6))
sns.scatterplot(data=DATA2, x='minInstalls', y='score', color='skyblue',
               alpha=0.7)
plt.title('App Installs vs. Scores')
plt.xlabel('Number of Installs')
plt.ylabel('Score')
plt.tight_layout()
plt.show()

# Visualization 3: App Ratings Over Time
DATA2['released'] = pd.to_datetime(DATA2['released'])
plt.figure(figsize=(12, 6))
```



```

sns.lineplot(data=DATA2, x='released', y='score', ci=None)
plt.title('App Ratings Over Time')
plt.xlabel('Release Date')
plt.ylabel('Average Score')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Visualization 4: Distribution of App Ratings by Android Version Compatibility
plt.figure(figsize=(10, 6))
sns.boxplot(data=DATA2, x='androidVersion', y='score', palette='pastel')
plt.title('App Ratings by Android Version Compatibility')
plt.xlabel('Android Version Compatibility')
plt.ylabel('Score')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Visualization 6: App Score Distribution Across Genres
plt.figure(figsize=(12, 6))
sns.boxplot(data=DATA2, x='genreId', y='score', palette='colorblind')
plt.title('App Score Distribution Across Genres')
plt.xlabel('Genre')
plt.ylabel('Score')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Visualization 7: App Ratings vs. Reviews Sentiment (Example)
# Assuming 'sentiment' column is available in DATA (from previous sentiment_
↳analysis)
plt.figure(figsize=(10, 6))
sns.boxplot(data=DATA, x='sentiment', y='score', palette='muted')
plt.title('App Ratings by Reviews Sentiment')
plt.xlabel('Sentiment')
plt.ylabel('Score')
plt.tight_layout()
plt.show()

```

