

In []:

1.. How to read csv data or how to load data !

In [1]: *### Lets import all the necessary packages !*

In []:

```
In [2]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]: `comments = pd.read_csv(r"D:\data Youtube\UScomments.csv", on_bad_lines='skip')`

C:\Users\User\AppData\Local\Temp\ipykernel_3380\3559977166.py:1: DtypeWarning: Columns (2,3) have mixed types. Specify dtype option on import or set low_memory=False.

```
comments = pd.read_csv(r"D:\data Youtube\UScomments.csv", on_bad_lines='skip')
```

In [4]: *### above is a warning , u can ignore that ..*In [5]: `comments.head()`

Out[5]:

	video_id	comment_text	likes	replies
0	XpVt6Z1Gjjo	Logan Paul it's yo big day !!!!!	4	0
1	XpVt6Z1Gjjo	I've been following you from the start of your...	3	0
2	XpVt6Z1Gjjo	Say hi to Kong and maverick for me	3	0
3	XpVt6Z1Gjjo	MY FAN . attendance	3	0
4	XpVt6Z1Gjjo	trending 🤔	3	0

In [6]: *## Lets find out missing values in your data*
`comments.isnull().sum()`

```
Out[6]: video_id      0
comment_text    26
likes           0
replies         0
dtype: int64
```

In [7]: *## drop missing values as we have very few & Lets update dataframe as well..*
`comments.dropna(inplace=True)`

```
In [8]: comments.isnull().sum()
```

```
Out[8]: video_id      0
        comment_text  0
        likes        0
        replies      0
        dtype: int64
```

2.. Perform Sentiment Analysis

In short , Sentiment analysis is all about analyszing sentiments of Users

```
In [ ]:
```

```
In [9]: !pip install textblob
        ### Lets perform sentiment analysis using TextBlob which is a NLP library built
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: textblob in c:\users\user\appdata\roaming\python\python311\site-packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in d:\windows security\new folder\lib\site-packages (from textblob) (3.8.1)
Requirement already satisfied: click in d:\windows security\new folder\lib\site-packages (from nltk>=3.1->textblob) (8.0.4)
Requirement already satisfied: joblib in d:\windows security\new folder\lib\site-packages (from nltk>=3.1->textblob) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in d:\windows security\new folder\lib\site-packages (from nltk>=3.1->textblob) (2022.7.9)
Requirement already satisfied: tqdm in d:\windows security\new folder\lib\site-packages (from nltk>=3.1->textblob) (4.65.0)
Requirement already satisfied: colorama in d:\windows security\new folder\lib\site-packages (from click->nltk>=3.1->textblob) (0.4.6)
```

```
In [10]:
```

```
In [11]: from textblob import TextBlob
```

```
In [12]: comments.head(6)
```

```
Out[12]:
```

	video_id	comment_text	likes	replies
0	XpVt6Z1Gjjo	Logan Paul it's yo big day !!!!!	4	0
1	XpVt6Z1Gjjo	I've been following you from the start of your...	3	0
2	XpVt6Z1Gjjo	Say hi to Kong and maverick for me	3	0
3	XpVt6Z1Gjjo	MY FAN . attendance	3	0
4	XpVt6Z1Gjjo	trending 🤔	3	0
5	XpVt6Z1Gjjo	#1 on trending AYYYYEEEE	3	0

```
In [13]: TextBlob("Logan Paul it's yo big day !!!!!").sentiment.polarity
```

```
Out[13]: 0.0
```

```
In [14]: comments.shape
```

```
Out[14]: (691374, 4)
```

```
In [15]: ## for those of you who dont have good specifications , considering sample of d  
sample_df = comments[0:1000]
```

```
In [16]: sample_df.shape
```

```
Out[16]: (1000, 4)
```

```
In [17]: polarity = []  
  
for comment in comments['comment_text']:  
    try:  
        polarity.append(TextBlob(comment).sentiment.polarity)  
    except:  
        polarity.append(0)
```

```
In [18]: len(polarity)
```

```
Out[18]: 691374
```

```
In [19]: comments['polarity'] = polarity  
  
### Inserting polarity values into comments dataframe while defining feature na
```

```
In [20]: comments.head(5)
```

```
Out[20]:
```

	video_id	comment_text	likes	replies	polarity
0	XpVt6Z1Gjjo	Logan Paul it's yo big day !!!!!	4	0	0.0
1	XpVt6Z1Gjjo	I've been following you from the start of your...	3	0	0.0
2	XpVt6Z1Gjjo	Say hi to Kong and maverick for me	3	0	0.0
3	XpVt6Z1Gjjo	MY FAN . attendance	3	0	0.0
4	XpVt6Z1Gjjo	trending 🤔	3	0	0.0

```
In [ ]:
```

3.. Wordcloud Analysis of your data

In [21]: `### Lets perform EDA for the highly Positive sentences ie Polarity value will be`

In [22]: `filter1 = comments['polarity']==1`

In [23]: `comments_positive = comments[filter1]`

In []:

In [24]: `filter2 = comments['polarity']==-1`

In [25]: `comments_negative = comments[filter2]`

In [26]: `comments_positive.head(5)`

Out[26]:

	video_id	comment_text	likes	replies	polarity
64	XpVt6Z1Gjjo	yu are the best	1	0	1.0
156	cLdxuaxaQwc	Power is the disease. Care is the cure. Keep...	0	0	1.0
227	WYYvHb03Eog	YAS Can't wait to get it! I just need to sell ...	0	0	1.0
307	sjlHnJvXdQs	This is priceless	0	0	1.0
319	sjlHnJvXdQs	Summed up perfectly	0	0	1.0

In [27]: `!pip install wordcloud`

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: wordcloud in c:\users\user\appdata\roaming\python\python311\site-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in d:\windows security\new folder\lib\site-packages (from wordcloud) (1.24.3)
Requirement already satisfied: pillow in d:\windows security\new folder\lib\site-packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in d:\windows security\new folder\lib\site-packages (from wordcloud) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cycler>=0.10 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (23.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in d:\windows security\new folder\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in d:\windows security\new folder\lib\site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)

```
In [28]: from wordcloud import WordCloud , STOPWORDS
```

```
In [29]: set(STOPWORDS)
```

```
'all',
'an',
'and',
'any',
'are',
"aren't",
'as',
'at',
'be',
'because',
'been',
'before',
'being',
'below',
'between',
'both',
'but',
'by',
'can',
"can't",
...
```

```
In [30]: comments['comment_text']
```

```
Out[30]: 0          Logan Paul it's yo big day !!!!!
1      I've been following you from the start of your...
2          Say hi to Kong and maverick for me
3          MY FAN . attendance
4          trending 😊

...

691395          Лучшая
691396      qu'est ce que j'aimerais que tu viennes à Roan...
691397          Ven a mexico! 🍷 te amo LP
691398          Isliği yeter...
691399      Kocham tą piosenkę💕💕💕byłam zakochana po uszy ...
Name: comment_text, Length: 691374, dtype: object
```

```
In [31]: type(comments['comment_text'])
```

```
Out[31]: pandas.core.series.Series
```

```
In [32]: total_comments_positive = ' '.join(comments_positive['comment_text'])
```

```
In [33]: wordcloud = WordCloud(stopwords=set(STOPWORDS)).generate(total_comments_positive)
```



```
In [39]: !pip install emoji==2.2.0
```

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: emoji==2.2.0 in c:\users\user\appdata\roaming\python\python311\site-packages (2.2.0)

```
In [ ]:
```

```
In [40]: import emoji
```

```
In [41]: emoji.__version__
```

```
Out[41]: '2.2.0'
```

```
In [42]: comments['comment_text'].head(6)
```

```
Out[42]: 0          Logan Paul it's yo big day !!!!!
1  I've been following you from the start of your...
2          Say hi to Kong and maverick for me
3          MY FAN . attendance
4          trending 🤔
5          #1 on trending AYYEEEEEE
Name: comment_text, dtype: object
```

```
In [43]: comment = 'trending 🤔'
```

```
In [44]: [char for char in comment if char in emoji.EMOJI_DATA]
```

```
Out[44]: ['🤔']
```

```
In [ ]:
```

```
In [45]: emoji_list = []

for char in comment:
    if char in emoji.EMOJI_DATA:
        emoji_list.append(char)
```

```
In [46]: emoji_list
```

```
Out[46]: ['🤔']
```

```
In [47]: all_emojis_list = []

for comment in comments['comment_text'].dropna():
    for char in comment:
        if char in emoji.EMOJI_DATA:
            all_emojis_list.append(char)
```

```
In [48]: all_emojis_list[0:10]
```

```
Out[48]: ['!', '!', '!', '🤔', '🇧🇷', '👍', '👍', '👍', '👍', '👍']
```

In []:

In []:

In [49]: `from collections import Counter`

In [50]: `Counter(all_emojis_list).most_common(10)`

Out[50]:

```
[('😂', 36987),  
 ('😍', 33453),  
 ('❤️', 31119),  
 ('🔥', 8694),  
 ('🤖', 8398),  
 ('👉', 5719),  
 ('🤔', 5545),  
 ('👍', 5476),  
 ('❤️', 5359),  
 ('💕', 5147)]
```

In [51]: `Counter(all_emojis_list).most_common(10)[0]`

Out[51]: `('😂', 36987)`

In [52]: `Counter(all_emojis_list).most_common(10)[0][0]`

Out[52]: `'😂'`

In [53]: `Counter(all_emojis_list).most_common(10)[1][0]`

Out[53]: `'😍'`

In [54]: `Counter(all_emojis_list).most_common(10)[2][0]`

Out[54]: `'❤️'`

In [55]: `emojis = [Counter(all_emojis_list).most_common(10)[i][0] for i in range(10)]`

In []:

In []:

In [56]: `Counter(all_emojis_list).most_common(10)[0][1]`

Out[56]: `36987`

In [57]: `Counter(all_emojis_list).most_common(10)[1][1]`

Out[57]: `33453`


```
In [58]: Counter(all_emojis_list).most_common(10)[2][1]
```

```
Out[58]: 31119
```

```
In [59]: freqs = [Counter(all_emojis_list).most_common(10)[i][1] for i in range(10)]
```

```
In [60]: freqs
```

```
Out[60]: [36987, 33453, 31119, 8694, 8398, 5719, 5545, 5476, 5359, 5147]
```

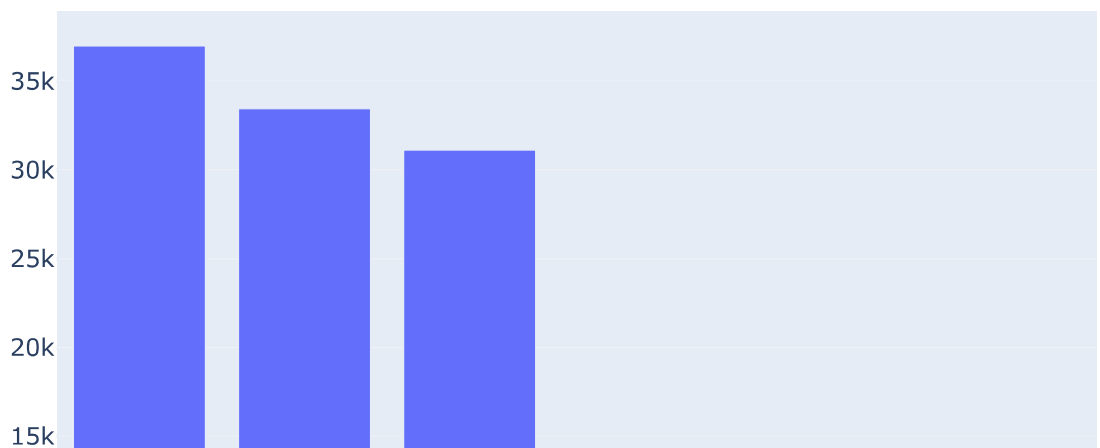
```
In [ ]:
```

```
In [ ]:
```

```
In [61]: import plotly.graph_objs as go  
from plotly.offline import iplot
```

```
In [62]: trace = go.Bar(x=emojis , y=freqs)
```

```
In [63]: iplot([trace])
```



In [64]: *## Conclusions : Majority of the customers are happy as most of them are using*

In []: