Project 5:Sentiment Analysis for Marketing

PHASE-3: DEVELOPMENT PART 1

Loading of dataset:

	Index	Product	Helpful_count	Total _coun	URL	Review_ country	Reviewed_a t	Review_text	Review_ratin g	Product_c ompany	Profile_ name
0	0	Apple iPhone XR (64GB) - Black	5,087 people found this helpful	24	https://www.amazon.in/Apple-iPhone-XR-64GB-Bla...	India	12-12-2018	NOTE:	3.0 out of 5 stars	Apple	Sameer Patil
1	1	Apple iPhone XR (64GB) - Black	2,822 people found this helpful	6	https://www.amazon.in/Apple-iPhone-XR-64GB-Bla...	India	17-11-2018	Very bad experience	1.0 out of 5 stars	Apple	Amazon Customer
2	2	Apple iPhone XR (64GB) - Black	1,798 people found this helpful	0	https://www.amazon.in/Apple-iPhone-XR-64GB-Bla...	India	27-01-2019	Amazing camera	5.0 out of 5 stars	Apple	A
3	3	Apple iPhone XR (64GB) - Black	1,366 people found this helpful	14	https://www.amazon.in/Apple-iPhone-XR-64GB-Bla...	India	02-05-2019	The product is not so good.	1.0 out of 5 stars	Apple	Shubham
4	4	Apple iPhone XR (64GB) - Black	536 people found this helpful	5	https://www.amazon.in/Apple-iPhone-XR-64GB-Bla...	India	24-05-2019	Excellent	5.0 out of 5 stars	Apple	Lokesh

Preprocessing:

df=pd.read.csv('/kaggle/input/apple-iphone-11-reviews-from-amazon-com/apple_iphone_11_reviews.csv')

df.head()

df.isna().sum()

index	0
product	0
helpful_count	0
total_comments	0
url	0
review_country	0
reviewed_at	0
review_text	3
review_rating	0
product_company	0
profile_name	0
review_title	2

dtype: int64

df.dropna(inplace=true)
df.shape
df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 5007 entries, 0 to 5009
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   index           5007 non-null   int64
 1   product         5007 non-null   object
 2   helpful_count   5007 non-null   object
 3   total_comments  5007 non-null   int64
 4   url             5007 non-null   object
 5   review_country  5007 non-null   object
 6   reviewed_at     5007 non-null   object
 7   review_text     5007 non-null   object
 8   review_rating   5007 non-null   object
 9   product_company 5007 non-null   object
 10  profile_name    5007 non-null   object
 11  review_title    5007 non-null   object
dtypes: int64(2), object(10)
memory usage: 508.5+ KB
```

```
df['reviewed_at']=pd.to_datetime(df['reviewed_at'])
print('All index column values are unique?: ',len(df.index)==df.index.nunique())
df.drop(['index'], axis=1, inplace=True)
```

All index column values are unique?: True

```
def find_likes(x):
    likes=x.split()[0]
    if likes=='One': return 1
    elif ',' in likes: return int(likes.replace(',', ''))
    else: return int(likes)
df['likes']=df['helpful_count'].apply(find_likes)
def find_rating(x):
    rating=x.split()[0]
    return float(rating)

df['rating']=df.review_rating.apply(find_rating)
df.rating.value_counts()
```

5.0   3730
4.0   718
1.0   319
3.0   153
2.0   87
Name: rating, dtype: int64

```
df.drop(['helpful_count',  'review_rating'], axis=1, inplace=True)
df.drop(['url'],  axis=1, inplace=True)
numeric=df.select_dtypes('number').columns
categoric=df.select_dtypes('object').columns
print('Numeric Columns: ', numeric)
print('Categoric Columns: ', categoric)
```

```
Numeric Columns:
  Index(['total_comments', 'likes', 'rating'], dtype='object')
Categoric Columns:
  Index(['product', 'review_country', 'review_text', 'product_company',
       'profile_name', 'review_title'],      dtype='object')
```