fliptnews

January 7, 2024

```python
[1]: import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```python
[2]: import nltk
```

```python
[3]: nltk.download('popular')
```

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to /root/nltk_data…
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to /root/nltk_data…
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to /root/nltk_data…
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to /root/nltk_data…
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to /root/nltk_data…
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     /root/nltk_data…
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to /root/nltk_data…
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to /root/nltk_data…
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to /root/nltk_data…
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to /root/nltk_data…
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     /root/nltk_data…
[nltk_data]    |   Package twitter_samples is already up-to-date!
[nltk_data]    | Downloading package omw to /root/nltk_data…
[nltk_data]    |   Package omw is already up-to-date!
[nltk_data]    | Downloading package omw-1.4 to /root/nltk_data…
[nltk_data]    |   Package omw-1.4 is already up-to-date!
```

```
[nltk_data]     | Downloading package wordnet to /root/nltk_data…
[nltk_data]     |   Package wordnet is already up-to-date!
[nltk_data]     | Downloading package wordnet2021 to /root/nltk_data…
[nltk_data]     |   Package wordnet2021 is already up-to-date!
[nltk_data]     | Downloading package wordnet31 to /root/nltk_data…
[nltk_data]     |   Package wordnet31 is already up-to-date!
[nltk_data]     | Downloading package wordnet_ic to /root/nltk_data…
[nltk_data]     |   Package wordnet_ic is already up-to-date!
[nltk_data]     | Downloading package words to /root/nltk_data…
[nltk_data]     |   Package words is already up-to-date!
[nltk_data]     | Downloading package maxent_ne_chunker to
[nltk_data]     |     /root/nltk_data…
[nltk_data]     |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]     | Downloading package punkt to /root/nltk_data…
[nltk_data]     |   Package punkt is already up-to-date!
[nltk_data]     | Downloading package snowball_data to
[nltk_data]     |     /root/nltk_data…
[nltk_data]     |   Package snowball_data is already up-to-date!
[nltk_data]     | Downloading package averaged_perceptron_tagger to
[nltk_data]     |     /root/nltk_data…
[nltk_data]     |   Package averaged_perceptron_tagger is already up-
[nltk_data]     |     to-date!
[nltk_data]     |
[nltk_data]   Done downloading collection popular
```

[3]: True

[4]:
```python
df= pd.read_csv('/content/flipitnews-data.csv')
```

[5]:
```python
df.head()
```

[5]:
```
          Category                                            Article
0       Technology  tv future in the hands of viewers with home th…
1         Business  worldcom boss  left books alone  former worldc…
2           Sports  tigers wary of farrell  gamble  leicester say …
3           Sports  yeading face newcastle in fa cup premiership s…
4    Entertainment  ocean s twelve raids box office ocean s twelve…
```

[6]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2225 entries, 0 to 2224
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Category  2225 non-null   object
 1   Article   2225 non-null   object
dtypes: object(2)
```

```
memory usage: 34.9+ KB
```

[7]: `df.shape`

[7]: `(2225, 2)`

[8]:
```python
print('No. of Rows:',df.shape[0])
print('No. of Columns: ', df.shape[1])
```

```
No. of Rows: 2225
No. of Columns:  2
```

[9]: `print(df.shape[0] * df.shape[1])`

```
4450
```

[10]: `df.describe()`

[10]:
```
        Category                                           Article
count       2225                                              2225
unique         5                                              2126
top        Sports  kennedy questions trust of blair lib dem leade…
freq         511                                                 2
```

[11]: `df.dtypes`

[11]:
```
Category    object
Article     object
dtype: object
```

[12]:
```python
categories=df['Category'].value_counts()
categories
```

[12]:
```
Sports          511
Business        510
Politics        417
Technology      401
Entertainment   386
Name: Category, dtype: int64
```

[13]: `df.head(2)`

[13]:
```
      Category                                           Article
0   Technology  tv future in the hands of viewers with home th…
1     Business  worldcom boss  left books alone  former worldc…
```

[14]: `df['Article'].value_counts()`

[14]: kennedy questions trust of blair lib dem leader charles kennedy has said voters now have a fundamental lack of trust of tony blair as prime minister. he said backing his party was not a wasted vote adding that with the lib dems what you see is what you get . he made his comments at the start of a day of appearances on channel five in a session on the wright stuff programme. questions from callers a studio audience and the show s presenter covered lib dem tax plans anti-terror laws and immigration. mr kennedy said during his nearly 22 years in parliament he had seen prime ministers and party leaders come and go and knew the pitfalls of british politics. 1983 was when i was first elected as an mp – so tony blair michael howard and myself were all class of 83 – and over that nearly quarter of a century the world has changed out of recognition   he said. we don t actually hear the argument any longer: lib dems good people reasonable ideas but only if we thought they could win around here – it s a wasted vote . you don t hear that because the evidence of people s senses demonstrates that it isn t a wasted vote. but he said mr blair had lost the trust of the british people. there is a fundamental lack of trust in tony blair as prime minister and in his government   he said.   what we ve got to do as a party – what i ve got to do as a leader of this party – is to convey to people that what you see is what you get. mr kennedy also used his tv appearance to defend his party s plans to increase income tax to 50% for those earning more than £100 000 saying it would apply to just 1% of the population. he said the extra revenue would allow his party to get rid of tuition and top-up fees introduce free personal care for the elderly and replace the council tax with a local income tax. mr blair has already spent a day with five and michael howard is booked for a similar session.
2
tough rules for ringtone sellers firms that flout rules on how ringtones and other mobile extras are sold could be cut off from all uk phone networks. the rules allow offenders to be cut off if they do not let consumers know exactly what they get for their money and how to turn off the services. the first month under the new rules has seen at least ten firms suspended while they clean up the way they work. the rules have been brought in to ensure that the problems plaguing net users do not spread to mobile phones. in the last couple of years ringtones wallpapers screensavers and lots of other extras for phones have become hugely popular. but fierce competition is making it difficult for firms to get their wares in front of consumers said jeremy flynn head of third party services at vodafone.   if you are not on the operator s portal you are going to have quite heavy marketing costs because it s a problem of how people discover your services   he said. to combat this many ringtone and other mobile content sellers started using a new tactic to squeeze more cash out of customers. this tactic involved signing people up for a subscription to give them for instance several ringtones per week or month instead of the single track they thought they were getting. mr flynn said that the move to using subscriptions happened over the space of a few weeks at the end of 2004. websites such as grumbletext.co.uk started getting reports from customers who were racking up large bills for phone content they did not know they had signed up for. what made us uncomfortable was that these services were not being marketed

transparently said mr flynn. people did not know they were being offered a subscription service. we saw potential for substantial consumer harm here he added. the swift adoption of subscription services led to the creation of a new code of conduct for firms that want to sell content for mobile phones. the drafting of the new rules was led by the mobile entertainment forum and the uk s phone firms. everyone is required to conform to this code of conduct said andrew bud regulatory head of the mef and executive chairman of messaging firm mblox. it s all about transparency he said. consumers have to be told what they have got themselves into and how to get out of it. the consumer has a right to be protected he said. christian harris partnership manager of mobile content firm zed said the new system was essential if consumers were to trust companies that sell ringtones and other downloads. the groundrules must be applied across the whole industry and if that s done effectively we will see the cowboys driven out he said. the new rules came in to force on 15 january and the first month under the new regime has seen many firms cautioned for not honouring them. some have been told to revamp websites so customers know what they get for their money and what they are signing up for said mr flynn. also said mr flynn vodafone has briefly cut off between eight and ten content sellers flouting the rules. we have quite draconian contracts with firms he said. we do not have to say why. we can just cut them off. under the rules consumers must be able to switch off the services by using a universal stop command sent via text message. he said the system had been designed to limit how much a consumer will pay if they inadvertently signed up for a service. the mobile is so personal that people really resent the abuse of what is effectively part of their personality said mr flynn.
2
super union merger plan touted two of britain s big trade unions could merge to form a super union of two million members. the move by amicus and the transport and general workers union (tgwu) would be a seen as a bid to carry more weight with ministers and employers. amicus has 1.2 million members and the tgwu has 800 000. any merger would have to be approved by the unions executives and their membership. it is understood meetings will be held on wednesday about the proposal. along with the gmb and unison the tgwu and amicus worked closely together in the last year to hammer out a 56-point deal with labour s leadership over equality at work holidays and pensions – the warwick agreement. both unions are remaining tight-lipped about the merger rumours but one insider pointed out to the bbc news website that nobody is denying suggestions a merger could be on the agenda when the two unions executives hold their meetings on wednesday. amicus s executive was due to meet in any case although the tgwu is holding specially scheduled talks.
2
rings of steel combat net attacks gambling is hugely popular especially with tech-savvy criminals. many extortionists are targeting net-based betting firms and threatening to cripple their websites with deluges of data unless a ransom is paid. but now deep defences are being put in place by some of the uk s biggest net firms to stop these attacks. increasing numbers of attacks and the huge amounts of data being used to try to bump a site off the web are prompting

firms to adopt the measures. net firms are realising that it s not just about anti-virus and firewalls   said paul king chief security architect at cisco. there are more things that can be done in the network to protect data centres. mr king said the only way to properly combat these so-called distributed denial-of-service attacks was with intelligent net-based systems. many of the gambling sites suffering ddos attacks are in offshore data and hosting centres so any large scale data flood could knock out access to many more sites than just the one the criminals were targeting said mr king. this overspill effect was only likely to grow as attacks grow in size and scale. malcolm seagrave security expert at energis said the most common types of attacks hit sites with 10 megabytes of data over short periods of time. bigger attacks sending down 200 megabytes of traffic or more were rarely seen he said. it does feel like they are turning the dial because you see this traffic gradually growing   he said. so far there have been no attacks involving gigabytes of data said mr seagrave. however he added that it was only a matter of time before such large attacks were mounted. maria capella spokeswoman for net provider pipex said that when ddos attacks were at their height customers were getting hit every four to five days. the defences being put in place constantly monitor the streams of data flowing across networks and pluck out the traffic destined for target sites. it s about understanding what s genuine traffic and keeping attack traffic from going to the site   she said. we study the profile of their traffic and as soon as we see an anomaly in the profile that s when we start to get the backbone engineering boys to see if we are going to sustain an attack   said ms capella. this traffic can be hard to spot because ddos attacks typically use thousands of computers in many different countries each participating machine only sends a small part of the entire data flood. typically these computers have been infected by a virus or worm which reports its success and the net address of compromised machines back to the malicious hacker or hi-tech criminal that set off the virus. hijacked computers are known as zombies or bots and collections of them are called bot nets. many spammers rent out bot nets to help them anonymously send junk mail. most of the zombies are based outside the country that hosts the target site so getting the attacking pcs shut off can be difficult. often pipex and other net suppliers do get advance notice that an attack is about to happen. the serious players tend to precede an attack with some kind of ransom e-mail   said ms capella. we ask as part of the service we provide that customers notify us of anything they have in advance that would give us forewarning. once an attack is spotted dedicated net hardware takes over to remove the attack traffic and ensure that sites stay up. energis took a similar approach said mr seagrave. we have technology out there that allows us to detect attacks in minutes rather than let network engineers spend hours pulling the information together   said mr seagrave. also net firms were starting to work more closely together on the problem of ddos attacks and pool information about where they are coming from. information gathered on attacks and where they originated has led to some arrests. he said energis also did its own intelligence work to get in insight into which sites criminal gangs plan to target. we have people in places where they shouldn t be monitoring tech sites he said. sometimes though he said spotting the next victim was easy. you can

see them going alphabetically through the list with the gambling sites trying
one after another   said mr seagrave.   2
holmes starts 2005 with gb events kelly holmes will start 2005 with a series of
races in britain. holmes will make her first track appearance on home soil
since winning double olympic gold in january s norwich union international in
glasgow. she will also run in the grand prix in birmingham in february and may
defend her indoor aaa 800m title in sheffield earlier that month. i am still
competitive and still want to win   she said. i m an athlete and i can t wait
to get back on the track. she added: these events are also a great opportunity
to thank the british public for the enormous levels of support they have given
me from the moment i stepped off that plane from greece. the glasgow meeting
will see holmes compete over 1500m in a five-way match against sweden france
russia and italy.
2

                                        ..
gb select holmes for double bid kelly holmes has been chosen for both the 800m
and 1500m at next month s european indoor championships - despite not yet
confirming her availability. the double-olympic champion is still undecided
about whether to compete in her first european indoors in madrid. but uk
athletics chief max jones believes it is likely holmes will race. he said: it s
worth a punt isn t it she s gone back to training. if she s confident she will
pick one (distance). the programme does not permit both. if holmes does take
her place in the gb side she has already indicated that 1500m will be her
favoured distance. under championship rules she does not have to decide until
24 hours before the event begins in madrid on 4 march. jones added: she was
pleased with her run in birmingham (at last weekend s grand prix). i am
inclined to think she will (compete). it could depend on how training goes and
she will have to be close to top form. and he admitted holmes presence would
be a great boost to the rest of the norwich union gb team. she is a very good
team member and is very good to have there   he added. meanwhile there is no
surprises in the rest of the great britain team announced on tuesday. defending
60m champion jason gardener is included and will be chasing his third successive
gold medal. he will be joined in the event by mark lewis-francis - runner-up
behind him three years ago in vienna - and new boy darren chin. kelly sotherton
has been invited by championship organisers to compete against carolina kluft in
the pentathlon. sotherton finished third in the olympic heptathlon in athens
where the swede took gold. there are 16 newcomers at this level of competition
in gb s 46-strong team. the championships will be the last to be attended by
jones before he retires to be succeeded by dave collins. jones said: the
olympic games is the overwhelming target for uka and therefore it is beneficial
to start the next four-year cycle with a not-so-pressurised major games. the
european indoor championships always provide a chance for the young and
inexperienced to make a name for themselves and the current indoor season has
produced some exciting new talent. madrid is the first small step on a long
journey to beijing and the 2008 olympics but for some of the emerging athletes
this will be a very important step.   60m: d chin (belgrave harriers) j
gardener (wessex and bath) m lewis-francis (birchfield harriers). t abeyie

(woodford green with essex ladies) c lambert (belgrave harriers). d garland
(channel islands ac). j mcilroy (windsor slough eton and hounslow ac) jimmy
watkins (cardiff aac). ed jackson (aldershot farnham and district) n speaight
(belgrave harriers) j thie (cardiff aac). a baddeley (harrow ac) m farah
(windsor slough eton and hounslow ac) j mayock (barnsley ac). d greaves
(newham and essex beagles) a scott (shaftesbury barnet harriers) a turner
(notts ac). b challenger (belgrave harriers). pole vault: n buckfield (crawley
ac). n morgan (birchfield harriers) c tomlinson (newham and essex beagles). l
achike (shaftesbury barnet harriers) n douglas (oxford city ac) p idowu
(belgrave harriers). d cossins (birchfield harriers) r davenport (gloucester
ac) d garland (channel islands ac) r preddy (gloucester ac) g warburton
(cardiff aac). k endacott (city of plymouth ac) j kwakye (woodford green with
essex ladies) j maduaka (woodford green with essex ladies). k wall (basildon
ac). k holmes (ealing southall and middlesex ac). h clitheroe (preston
harriers) k holmes (ealing southall and middlesex ac). j pavey (exeter
harriers). d allahgreen (trafford ac) s claxton (belgrave harriers). j crane
(sale harriers manchester) s jones (trafford ac). j whitlock (trafford ac). j
johnson (herne hill harriers). k sotherton (birchfield harriers). d fraser
(croydon harriers) l mcconnell (shaftesbury barnet harriers) c murphy
(woodford green with essex ladies) m purkiss (southampton ac).
1
wood – ireland can win grand slam former captain keith wood believes ireland can
win only their second grand slam – and first since 1948 – in this year s rbs six
nations championship. after claiming their first triple crown for 19 years last
season wood tips his former team-mates to go one better. things have been
building up over the past few years and i think this is the year for ireland
he told bbc sport. there is a great chance to win a grand slam. a lot of things
are in our favour with england and france at home. ireland have finished
runners-up three times including last year since the old five nations became
six in 2000 and not finished outside the top three in the past five years.
despite being without flanker keith gleeson coach eddie o sullivan has not had
to contend with the sort of casualty lists that have hit england and scotland in
particular prior to the tournament. for ireland to win it we need to stay
relatively injury free and fortunately we are one of the few teams that have
done that so far   wood added. it is going to be tough and we need to take all
the luck and opportunities that come our way. ireland s last game of the
tournament is against wales in cardiff – a fixture they have not lost since
1983. but despite their traditional hospitality when the irish are visiting
wood believes wales might end their four-match losing run against england in
cardiff. so many of the major england players have either retired in the last
year or are injured that i think it will be very hard for them down in cardiff
wood added. wales have had four brilliant games in the last year or so and lost
all four so the time is right for them now to beat one of the major teams.
1
eu software patent law delayed controversial new eu rules for the patenting of
computer-based inventions have been put on hold due to a last minute
intervention from poland. poland – a large and therefore crucial eu member –

has requested more time to consider the issue especially as it relates to the patenting of software. critics say the law would favour large companies over small innovative ones. they say it could have massive ramifications for developments such as open source software. polish ministers want to see the phrasing of the text of the directive on the patentability of computer-implemented inventions changed so that it excludes the patenting of software. the planned law has ignited angry debate about whether the eu should allow the patenting of computer programs and internet business methods as currently happens in the us. so for instance us-based amazon.com holds a patent on one-click shopping . critics claim the law which the eu says is needed to harmonise with the us is little more than a law on ideas. patent rights it is claimed protect inventors but in this case it is exclusively the interests of the big companies that are protected   eva lichtenberg a austrian green member of the european parliament said in a statement. smaller innovative firms cannot afford the patenting and legal costs that the directive would inflict on them she added. the european parliament has already clashed with the european union on the issue and there have been accusations that the process to decide the issue has been undemocratic. the directive has been subject to several previous delays.

1

mansfield 0-1 leyton orient an second-half goal from andy scott condemned mansfield to a ninth successive game without a win. early in the second half wayne carlisle s cross was met by scott and he blasted the ball home from just outside the penalty area. as orient chased a second mansfield had to clear alan white s header off the line and kevin pilkington saved well from michael simpson. by the end mansfield fans were chanting for the head of chairman keith haslam. pilkington talbot buxton dimech artell corden murray curtis neil warne barker. subs not used: white mcintosh wood lloyd herron. harrison lockwood donny barnard white mackie scott saah simpson carlisle lee barnard ibehre. subs not used: morris wardley newey zakuani mcmahon. scott 51. 3 803 s mathieson (cheshire).

1

souness delight at euro progress boss graeme souness felt newcastle were never really in danger of going out of the uefa cup against heerenveen. an early own goal followed by an alan shearer strike earned them a 2-1 win and a place in the uefa cup last 16. obviously with winning in the first leg it gave us a great advantage he said after the 4-2 aggregate victory. we got our goals early and in the minds of some players the job was done but then they got a goal and perhaps made us a bit nervous. shearer s goal moved him within 12 of jackie milburn s club scoring record of 200 for the magpies. but souness said he did not think beating the record would have any bearing on his decision to retire at the end of the season. i think if he got it this year he would want to stay next year anyway he added. he struck the ball very well - he always has done - and i think it was the power and pace that beat the goalkeeper. souness also paid tribute to laurent robert who was at the heart of much of united s attacking play. in the first half he did really well and did everything you want from a wide player. more of the same in future please   he said.

```
     1
Name: Article, Length: 2126, dtype: int64
```

1 Detecting Null Values & Missing Values

```
[15]: df.isnull()/len(df)*100
```

```
[15]:       Category  Article
      0          0.0      0.0
      1          0.0      0.0
      2          0.0      0.0
      3          0.0      0.0
      4          0.0      0.0
      …          …        …
      2220       0.0      0.0
      2221       0.0      0.0
      2222       0.0      0.0
      2223       0.0      0.0
      2224       0.0      0.0

      [2225 rows x 2 columns]
```

```
[16]: # Missing Values

sns.heatmap(df.isnull())
plt.show()
```

```
[17]: df.duplicated().sum()
```

```
[17]: 99
```

```
[18]: df.drop_duplicates(keep=False, inplace=True)
```
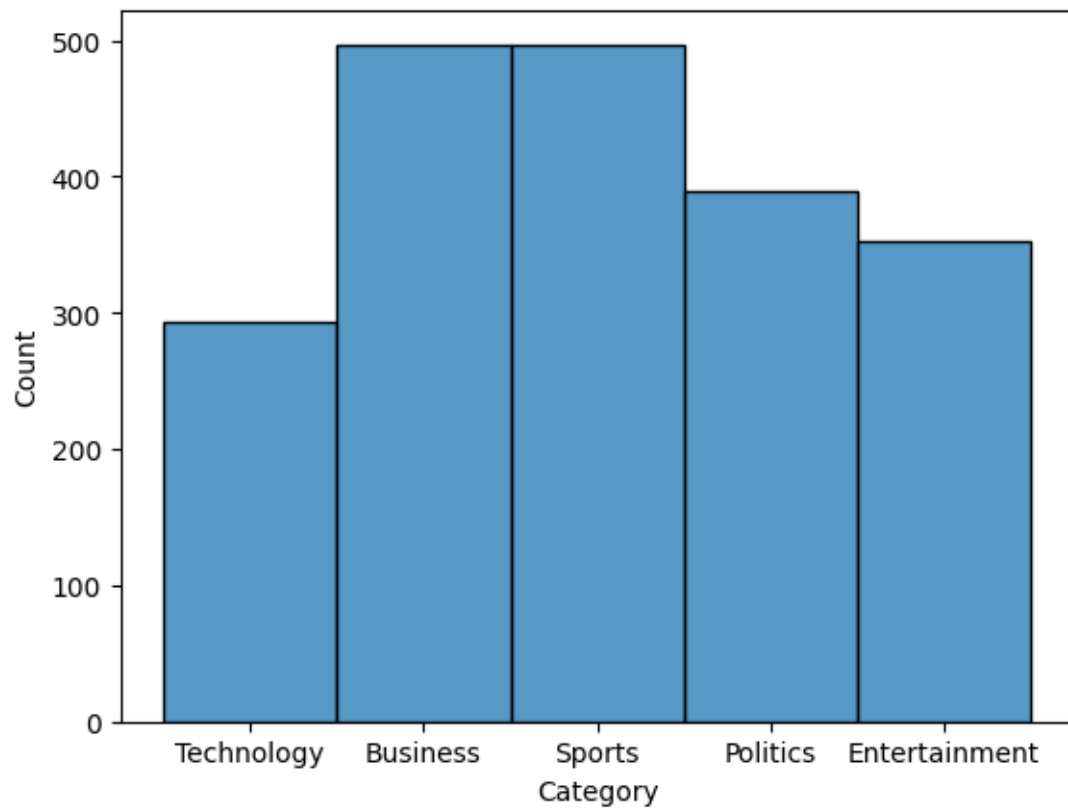
```
[19]: df.fillna(0, inplace=True)
```

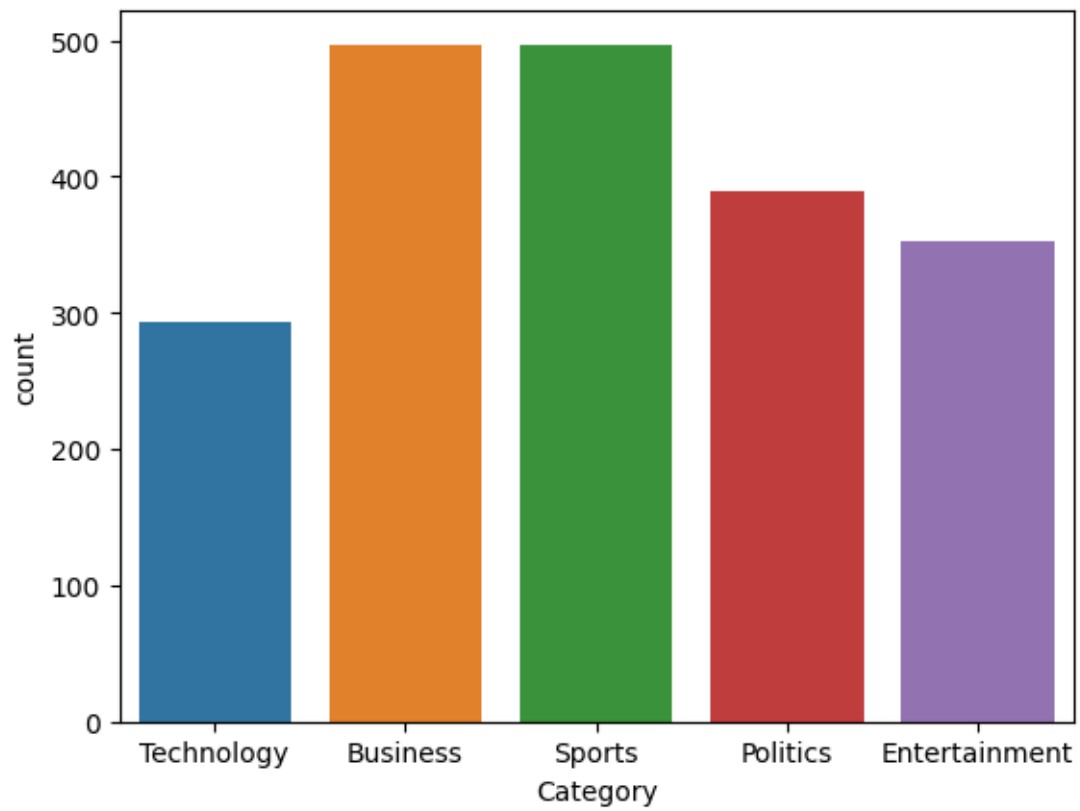## 2 Missing Value Treatment

```
[20]: df['Category'] = df['Category'].fillna(df['Category'].mode()[0])
      df['Article'] = df['Article'].fillna(df['Article'].mode())
```
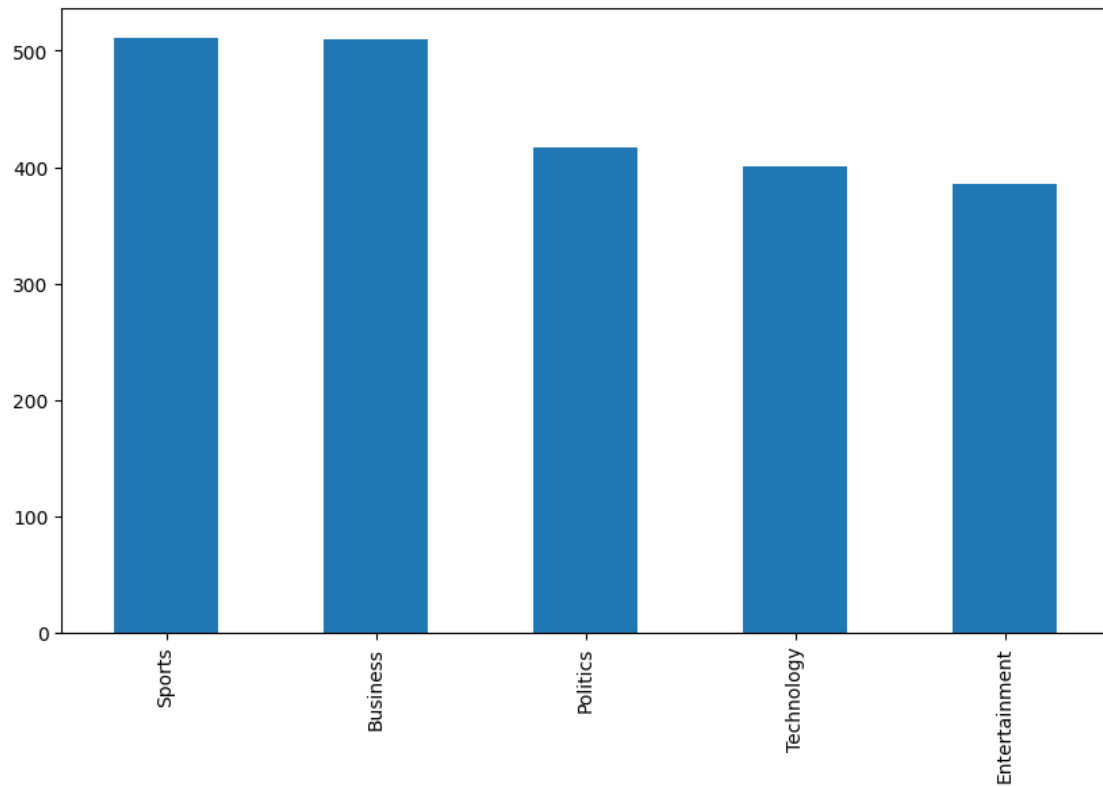
3 Univariate Analysis

```
[21]: sns.histplot(df['Category'])
      plt.show()
```

[22]:
```
sns.countplot(x="Category",data=df)
plt.show()
```

```
[23]: categories.plot(kind='bar',figsize=(10,6))
      plt.show()
```

4   Text Data Preprocessing

```
[24]: import re
      from nltk.stem import WordNetLemmatizer
      from nltk.corpus import stopwords
```

```
[25]: article=df['Article']
```

```
[31]: y= df['Category']
```

```
[ ]: nltk.word_tokenize(article)
```

```
[29]: wordnet= WordNetLemmatizer()
```

```
[30]: corpus=[]
```

5 Using Lemmatzier

```python
for i in range(len(article)):
    review= re.sub('[^a-ZA-Z]',article[i])
    review= review.lower()
    review= review.split()
    review= [wordnet.lemmatize(word) for word in review if word not in
    ↪set(stopwords.words('english'))]
    review=' '.join(review)

    corpus.append(review)
```

6 Implementing Bag Of Words(BOW)

```python
from sklearn.feature_extraction.text import CountVectorizer

cv= CountVectorizer()
X=cv.fit_transform(corpus).toarray()
```

```python
from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test= train_test_split(X,y,test_size=0.
  ↪20,random_state=0)
```

7 Using Naive Bayes

```python
from sklearn.naive_bayes import MultinomialNB
```

```python
naive= MultinomialNB()

naive.fit(X_train,y_train)
```

```python
y_pred= naive.predict(X_test)
```

8 Model Evaluation Metrics

```python
from sklearn.metrics import accuracy_score
```

```python
accuracy= accuracy_score(y_test,y_pred)

print(accuracy)
```

```python
from sklearn.metrics import precision_score
```

```python
precise= precision_score(y_test,y_pred)
print(precise)
```

9 Using HyperParameter & RandomForest Algorithm

```python
from sklearn.model_selection import GridSearchCV
```

```python
clf = GridSearchCv(estimator, param_grid, cv, scoring)
```

```python
_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20,
 ↪random_state = 42)
```

10 Building Random Forest Classifier

```python
from sklearn.ensemble import RandomForestClassifier
```

```python
rfc = RandomForestClassifier()
```

```python
forest_params = [{'max_depth': list(range(10, 15)), 'max_features':
 ↪list(range(0,14))}]

clf = GridSearchCV(rfc, forest_params, cv = 10, scoring='accuracy')

clf.fit(X_train, y_train)
```

```python
from sklearn.metrics import accuracy_score
```

```python
accuracy= accuracy_score(y_test,y_pred)

print(accuracy)
```

```python
from sklearn.metrics import precision_score

precise= precision_score(y_test,y_pred)
print(precise)
```