

```
In [1]: import numpy as np
import pandas as pd
import nltk    ## for text
import re      ## regular expression
from nltk.corpus import stopwords
from nltk.tokenize import sent_tokenize
from nltk.stem.porter import PorterStemmer  ## it converts the word into its base form
```

```
In [2]: df = pd.read_csv('tennis.csv')
```

```
In [3]: df['article_text'][0]
```

```
Out[3]: "Maria Sharapova has basically no friends as tennis players on the WTA Tour. The Russian player has no problems in openl
y speaking about it and in a recent interview she said: 'I don't really hide any feelings too much. I think everyone kno
ws this is my job here. When I'm on the courts or when I'm on the court playing, I'm a competitor and I want to beat eve
ry single person whether they're in the locker room or across the net. So I'm not the one to strike up a conversation abo
ut the weather and know that in the next few minutes I have to go and try to win a tennis match. I'm a pretty competitiv
e girl. I say my hellos, but I'm not sending any players flowers as well. Uhm, I'm not really friendly or close to many
players. I have not a lot of friends away from the courts.' When she said she is not really close to a lot of players, i
s that something strategic that she is doing? Is it different on the men's tour than the women's tour? 'No, not at all.
I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because you're
categorized, you're a tennis player, so you're going to get along with tennis players. I think every person has differen
t interests. I have friends that have completely different jobs and interests, and I've met them in very different parts
of my life. I think everyone just thinks because we're tennis players we should be the greatest of friends. But ultimate
ly tennis is just a very small part of what we do. There are so many other things that we're interested in, that we d
o.'"
```

```
In [7]: from nltk.tokenize import sent_tokenize
sentences = []
for s in df['article_text']:
    sentences.append(sent_tokenize(s))

sentences = sentences[0]
```

```
In [8]: sentences
```

```
Out[8]: ['Maria Sharapova has basically no friends as tennis players on the WTA Tour.',
"The Russian player has no problems in openly speaking about it and in a recent interview she said: 'I don't really hid
e any feelings too much.",
'I think everyone knows this is my job here.',
"When I'm on the courts or when I'm on the court playing, I'm a competitor and I want to beat every single person wheth
er they're in the locker room or across the net. So I'm not the one to strike up a conversation about the weather and kno
w that in the next few minutes I have to go and try to win a tennis match.",
"I'm a pretty competitive girl.",
"I say my hellos, but I'm not sending any players flowers as well.",
"Uhm, I'm not really friendly or close to many players.",
"I have not a lot of friends away from the courts.",
'When she said she is not really close to a lot of players, is that something strategic that she is doing?',
"Is it different on the men's tour than the women's tour?",
"'No, not at all.",
"I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because yo
u're categorized, you're a tennis player, so you're going to get along with tennis players.",
'I think every person has different interests.',
'I have friends that have completely different jobs and interests, and I've met them in very different parts of my lif
e.',
"I think everyone just thinks because we're tennis players we should be the greatest of friends.",
'But ultimately tennis is just a very small part of what we do.',
"There are so many other things that we're interested in, that we do.'"]
```

```
In [10]: w_d = {}

clean_sentences = pd.Series(sentences).str.replace("[^a-zA-Z]", " ")
clean_sentences = [s.lower() for s in clean_sentences]
stop_words = stopwords.words('english')
def remove_stopwords(sen):
    sen_new = " ".join([i for i in sen if i not in stop_words])
    return sen_new
clean_sentences = [remove_stopwords(r.split()) for r in clean_sentences]
```

```
In [11]: sentence_vectors = []
for i in clean_sentences:
    if len(i) != 0:
        v = sum([w_d.get(w, np.zeros((100,))) for w in i.split()]/(len(i.split())+0.001))
    else:
        v = np.zeros((100,))
    sentence_vectors.append(v)
```

```
In [12]: sim_mat = np.zeros([len(sentences), len(sentences)])
from sklearn.metrics.pairwise import cosine_similarity
for i in range(len(sentences)):
    for j in range(len(sentences)):
        if i != j:
            sim_mat[i][j] = cosine_similarity(sentence_vectors[i].reshape(1,100), sentence_vectors[j].reshape(1,100))[0,0]
```

```
In [13]: import networkx as nx

nx_graph = nx.from_numpy_array(sim_mat)
scores = nx.pagerank(nx_graph)
```

```
In [14]: ranked_sentences = sorted(((scores[i],s) for i,s in enumerate(sentences)), reverse=True)
for i in range(5):
    print("ARTICLE:")
    print(df['article_text'][i])
    print('\n')
    print("SUMMARY:")
    print(ranked_sentences[i][1])
    print('\n')
```

ARTICLE:

Maria Sharapova has basically no friends as tennis players on the WTA Tour. The Russian player has no problems in openly speaking about it and in a recent interview she said: 'I don't really hide any feelings too much. I think everyone knows this is my job here. When I'm on the courts or when I'm on the court playing, I'm a competitor and I want to beat every single person whether they're in the locker room or across the net. So I'm not the one to strike up a conversation about the weather and know that in the next few minutes I have to go and try to win a tennis match. I'm a pretty competitive girl. I say my hellos, but I'm not sending any players flowers as well. Uhm, I'm not really friendly or close to many players. I have not a lot of friends away from the courts.' When she said she is not really close to a lot of players, is that something strategic that she is doing? Is it different on the men's tour than the women's tour? No, not at all. I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because you're categorized, you're a tennis player, so you're going to get along with tennis players. I think every person has different interests. I have friends that have completely different jobs and interests, and I've met them in very different parts of my life. I think everyone just thinks because we're tennis players we should be the greatest of friends. But ultimately tennis is just a very small part of what we do. There are so many other things that we're interested in, that we do.'

SUMMARY:

When she said she is not really close to a lot of players, is that something strategic that she is doing?

In [ ]: