DIGITAL ASSIGNMENT-3

Practice Questions

December 18, 2024

SAI VARUN AITHA
saivarun.aitha18103@gmail.com

```
In [ ]:  !pip install spacy
```

```
In [ ]:  !python -m spacy download en_core_web_sm
```

```
In [3]:  import spacy
         import re
```

```
In [4]:  nlp = spacy.load("en_core_web_sm")
```

```
In [5]:  text = '''
         Look for data to help you address the question. Governments are good
         sources because data from public research is often freely available. Good
         places to start include http://www.data.gov/, and http://www.science.
         gov/, and in the United Kingdom, http://data.gov.uk/.
         Two of my favorite data sets are the General Social Survey at http://www3.norc.o
         and the European Social Survey at http://www.europeansocialsurvey.org/.
         '''
```

```
In [6]:  url_pattern = r"http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[!*\\(\\),]|(?:%[0-9a-fA-
```

```
In [7]:  doc = nlp(text)
         urls = [match.group() for match in re.finditer(url_pattern, text)]
         print("Extracted URLs:", urls)
```

Extracted URLs: ['http://www.data.gov/,', 'http://www.science.', 'http://data.go
v.uk/.', 'http://www3.norc.org/gss+website/,', 'http://www.europeansocialsurvey.o
rg/.']

```
In [ ]:
```