COURSE OUTCOME 5

PROGRAM - 13

AIM:

Implement a simple web crawler (ensure ethical conduct)

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
import re #regular expression

html=urlopen('https://en.wikipedia.org/wiki/CNN')
bs=BeautifulSoup(html,'html.parser')
for link in bs.find('div',{'id':'bodyContent'}).find_all('a',href=re.compile('^(/wiki/)((?!:)
  if 'href' in link.attrs:
    print(link.attrs['href'])
```

```
/wiki/CNN_(disambiguation)
/wiki/CNN_Center
/wiki/Atlanta,_Georgia
/wiki/1080i
/wiki/High-definition_television
/wiki/Letterboxing_(filming)
/wiki/480i
/wiki/Standard-definition_television
/wiki/WarnerMedia
/wiki/Vice_President
/wiki/Chief_Financial_Officer
/wiki/Vice_President
/wiki/Ken_Jautz
/wiki/CNN_Airport
/wiki/CNN_Arabic
/wiki/CNN_Brazil
/wiki/CNN_Chile
/wiki/CNN_en_Espa%C3%B1ol
/wiki/CNN_Indonesia
/wiki/CNN_International
/wiki/CNN-News18
/wiki/CNN_Philippines
/wiki/CNN_Portugal
/wiki/CNN_T%C3%BCrk
/wiki/HLN_(TV_network)
/wiki/Webcast
/wiki/DirecTV
/wiki/Video_on_demand
/wiki/Dish_Network
/wiki/Bell_Satellite_TV
/wiki/Shaw_Direct
/wiki/DirecTV_Caribbean
/wiki/Tata_Play
/wiki/Internet_Protocol_television
```

```
/wiki/U-verse_TV
/wiki/Bell_Fibe_TV
/wiki/Google_Fiber
/wiki/VMedia
/wiki/Verizon_FiOS
/wiki/Streaming_media
/wiki/Hulu#Hulu_+_Live_TV_service
/wiki/Sling_TV
/wiki/YouTube_TV
/wiki/Sirius_XM_Holdings
/wiki/Pay_television
/wiki/Atlanta
/wiki/AT%26T
/wiki/WarnerMedia
/wiki/Media_proprietor
/wiki/Ted_Turner
/wiki/Reese_Schonfeld
/wiki/United_States_cable_news
/wiki/24-hour_news_cycle
/wiki/Nielsen_Corporation
/wiki/Fox_News
/wiki/MSNBC
/wiki/Breaking_news
/wiki/CNN_controversies
```

RESULT:

The program executed successfully and obtained the output.

PROGRAM - 14

AIM:

Implement a program to scrap the web page of any popular website – suggested python package is scrappy (ensure ethical conduct).

In [ ]:

```
#webpage scraper
from bs4 import BeautifulSoup
import csv
import requests
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

In [ ]:

```
webpage=requests.get("https://meesho.com/handbags-women/pl/6m6o8")
```

In [ ]:

```
webpage
```

Out[ ]:

```
<Response [200]>
```

In [ ]:

```
soup=BeautifulSoup(webpage.content,"html.parser")
```

In [ ]:

```
print(soup.prettify()) # used to nest tags
```

```
<!DOCTYPE html>
<html lang="en">
 <head>
  <link as="font" href="/fonts/mier-fonts.css?display=optional" rel="preload" type="font/
woff2"/>
  <link href="/fonts/mier-fonts.css?display=optional" rel="stylesheet"/>
  <title>
   Handbags for Women - Buy Ladies Handbag Designs Online at Best Price | Meesho
  </title>
  <meta content="Shop for trendy Handbags for Women Check out online our range of Handbag
s for ladies and girls in so many different shapes and designs at best prices online at m
eesho" name="description"/>
  <meta content="meesho,reseller,products,reselling" property="keywords"/>
  <meta content="Handbags for Women - Buy Ladies Handbag Designs Online at Best Price | M
eesho" property="og:title"/>
  <meta content="Shop for trendy Handbags for Women Check out online our range of Handbag
s for ladies and girls in so many different shapes and designs at best prices online at m
eesho" property="og:description"/>
  <meta content="website" property="og:type"/>
  <meta content="Meesho" property="twitter:creator"/>
  <meta content="Handbags for Women - Buy Ladies Handbag Designs Online at Best Price | M
eesho" property="twitter:title"/>
  <meta content="Shop for trendy Handbags for Women Check out online our range of Handbag
s for ladies and girls in so many different shapes and designs at best prices online at m
eesho" property="twitter:description"/>
  <meta content="hN0QXvUaNeMlD8XXGkR9z5vDDiSrPQJ8ioMDa7eAMEU" name="google-site-verificat
ion"/>
  <meta content="summary" property="twitter:card"/>
  <link href="https://meesho.com/handbags-women/pl/6m6o8" rel="canonical"/>
```

```
  <link href="https://meesho.com/handbags-women/pl/6m6o8" hreflang="en" rel="alternate"/>
  <link href="https://meesho.com/hi/handbags-women/pl/f6mso" hreflang="hi" rel="alternate
"/>
  <script type="text/javascript">
   (function(e, s, t) {
               var t = t || {};
                   t.id = e;
                   var versaTagObj = {
                       $: [],
                       onready: function(e) {
                           this.$.push(e)
                       },
                   };
               var n = s.getElementsByTagName("script")[0],
                   r = s.createElement("script");
                   r.options = t;
                   r.async = !0;
                   r.src = "https://secure-ds.serving-sys.com/SemiCachedScripts/ebOneTag.
js?id=" + e;
                   r.options = t;
                   n.parentNode.insertBefore(r, n)
           })("1073747206", document, {
               dataLayer: "dataLayer"
           })
  </script>
  <link as="image" href="https://images.meesho.com/images/products/38188106/np9p0_512.jpg
" rel="preload"/>
  <meta charset="utf-8"/>
  <meta content="IE=edge" http-equiv="X-UA-Compatible"/>
  <meta content="width=device-width,initial-scale=1,maximum-scale=1,user-scalable=no" nam
e="viewport"/>
  <script defer="" src="https://www.googletagmanager.com/gtag/js?id=UA-113318580-2">
  </script>
  <script defer="">
   window.dataLayer = window.dataLayer || [];
               function gtag(){dataLayer.push(arguments);}
               gtag('js', new Date());

               gtag('config', 'UA-113318580-2', {
                 page_path: window.location.pathname,
               });
  </script>
  <script defer="">
   (function(w,d,s,l,i){w[l]=w[l]||[];w[l].push({'gtm.start':
          new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagName(s)[0],
          j=d.createElement(s),dl=l!='dataLayer'?'&l='+l:'';j.defer=true;j.src=
          'https://www.googletagmanager.com/gtm.js?id='+i+dl;f.parentNode.insertBefore(j,
f);
          })(window,document,'script','dataLayer','GTM-WCJQFLD');
  </script>
  <noscript>
   <img alt="facebook" height="1" src="https://www.facebook.com/tr?id=771814629829015&amp
;ev=PageView&amp;noscript=1" width="1"/>
  </noscript>
  <meta content="23" name="next-head-count"/>
  <noscript data-n-css="">
  </noscript>
  <link as="script" href="/_next/static/chunks/29.e844e135ae48111a6f54.js" rel="preload"/
>
  <link as="script" href="/_next/static/chunks/webpack-b2687f991651e7a65c91.js" rel="prel
oad"/>
  <link as="script" href="/_next/static/chunks/framework.57a22ac5870571c2eff5.js" rel="pr
eload"/>
  <link as="script" href="/_next/static/chunks/commons.7991f999a7315064944e.js" rel="prel
oad"/>
  <link as="script" href="/_next/static/chunks/main-42280db73908a476c888.js" rel="preload
"/>
  <link as="script" href="/_next/static/chunks/e835fc4de3c3612e6724f6ce6fcbab3c00ee3a2c.2
c2977314fcf57c1f948.js" rel="preload"/>
  <link as="script" href="/_next/static/chunks/6dc84bcc6813d527638a4811ea5ac2bf6e0387b4.8
6889062eec0e5d87462.js" rel="preload"/>
  <link as="script" href="/_next/static/chunks/4ac4af99e14a75ae402f6da6b5657d2d43fae348.0
```

KF4h"],"gssp":true,"customServer":true}
</script>
<script nomodule="" src="/_next/static/chunks/polyfills-104f08ba039f3cb24a4b.js">
</script>
<script async="" src="/_next/static/chunks/29.e844e135ae48111a6f54.js">
</script>
<script async="" src="/_next/static/chunks/webpack-b2687f991651e7a65c91.js">
</script>
<script async="" src="/_next/static/chunks/framework.57a22ac5870571c2eff5.js">
</script>
<script async="" src="/_next/static/chunks/commons.7991f999a7315064944e.js">
</script>
<script async="" src="/_next/static/chunks/main-42280db73908a476c888.js">
</script>
<script async="" src="/_next/static/chunks/e835fc4de3c3612e6724f6ce6fcbab3c00ee3a2c.2c2
977314fcf57c1f948.js">
</script>
<script async="" src="/_next/static/chunks/6dc84bcc6813d527638a4811ea5ac2bf6e0387b4.868
89062eec0e5d87462.js">
</script>
<script async="" src="/_next/static/chunks/4ac4af99e14a75ae402f6da6b5657d2d43fae348.077
89f8b8e8d5ff9ad60.js">
</script>
<script async="" src="/_next/static/chunks/f9cba575b2453906b7c34e64790f3360ede18f05.ca2
2077dc27bebcf7f90.js">
</script>
<script async="" src="/_next/static/chunks/64264fc7fef057efed475660fcbb653751b382c9.27b
8b3ccc42065377b03.js">
</script>
<script async="" src="/_next/static/chunks/pages/_app-766894b99df542c6c06b.js">
</script>
<script async="" src="/_next/static/chunks/279dde0a44deca2be40a3292cc11e949397c3321.d88
d09ff3fa4d89901d6.js">
</script>
<script async="" src="/_next/static/chunks/6752a23c9e2875d8afbd31a353b7f78231ca3881.ef3
367054aea0e5ba848.js">
</script>
<script async="" src="/_next/static/chunks/1a4ac8e632d753151428ffa1a0e7d9e8789a915a.b62
b191e7789e6d711cc.js">
</script>
<script async="" src="/_next/static/chunks/6e60e7ca507f193393b6f559aa62cfd471cb8be9.ebc
d2691250bcb8cda4a.js">
</script>
<script async="" src="/_next/static/chunks/e22ca4685f8364bfc5982f1b8e44b6407a032cf1.80b
334724c11715fb6e6.js">
</script>
<script async="" src="/_next/static/chunks/a45ddadc1c2718bbaca1306bddd98c08cc8a65e3.4a9
4d1d33c995945ea14.js">
</script>
<script async="" src="/_next/static/chunks/ac0d8dfae10d3f45eb3e72f4ed51cfb77a69602b.ab9
9c819b1a5a9839b81.js">
</script>
<script async="" src="/_next/static/chunks/a02e2299ce456baad815636cc8b3579bbe00d7e3.f11
f53c1f9d5d113bb46.js">
</script>
<script async="" src="/_next/static/chunks/pages/%5Bcategory%5D/pl/%5Bpage_id%5D-26ef4f
4bbeaf63b4b559.js">
</script>
<script async="" src="/_next/static/TJQ9SDug4-bArKXxso1H7/_buildManifest.js">
</script>
<script async="" src="/_next/static/TJQ9SDug4-bArKXxso1H7/_ssgManifest.js">
</script>
</body>
</html>

In [ ]:

```
names=soup.find_all('p',class_='Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__Produ
ctTitle_Desktop-sc-j0e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQheP
S')
```

In [ ]:

```
names
```

Out[ ]:

```
[<p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elegant Fashionable Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Gorgeous Attractive Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Classic Fancy Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Classic Versatile Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Gorgeous Versatile Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Trendy Fancy Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elegant Versatile Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Ravishing Stylish Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Ravishing Alluring Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Classic Versatile Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elegant Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elite Fashionable Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Voguish Fancy Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Classic Stylish Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elegant Attractive Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elite Versatile Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Classic Versatile Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Elite Fancy Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Trendy Fashionable Women Handbags</p>,
 <p class="Text__StyledText-sc-oo0kvp-0 bWSOET NewProductCard__ProductTitle_Desktop-sc-j0
e7tu-4 cQhePS NewProductCard__ProductTitle_Desktop-sc-j0e7tu-4 cQhePS" color="greyT2" fon
t-size="16px" font-weight="book">Gorgeous Classy Women Handbags</p>]
```

In [ ]:

```
Bag_names=[]
for i in range(0,len(names)):
  Bag_names.append(names[i].get_text())
```

In [ ]:

```
Bag_names
```

Out[ ]:

```
['Elegant Fashionable Women Handbags',
 'Gorgeous Attractive Women Handbags',
 'Classic Fancy Women Handbags',
 'Classic Versatile Women Handbags',
 'Gorgeous Versatile Women Handbags',
 'Trendy Fancy Women Handbags',
 'Elegant Versatile Women Handbags',
 'Ravishing Stylish Women Handbags',
 'Ravishing Alluring Women Handbags',
 'Classic Versatile Women Handbags',
 'Elegant Women Handbags',
 'Elite Fashionable Women Handbags',
 'Voguish Fancy Women Handbags',
 'Classic Stylish Women Handbags',
 'Elegant Attractive Women Handbags',
 'Elite Versatile Women Handbags',
 'Classic Versatile Women Handbags',
 'Elite Fancy Women Handbags',
 'Trendy Fashionable Women Handbags',
 'Gorgeous Classy Women Handbags']
```

In [ ]:

```
Price=soup.find_all('h5',class_='Text__StyledText-sc-oo0kvp-0 hiHdyy')
Price
```

Out[ ]:

```
[<h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->192</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->342</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->207</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->0<!-- --> <span class="Text__StyledText-sc-oo0kvp-0 fMjoAc" color="
greyT2" font-size="12px" font-weight="demi">onwards</span></h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->234</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->464</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->239</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->227</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->282</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->214</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->171</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->233</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->137</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->274</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->200</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->234</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->206</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->232</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->210</h5>,
```

```
eight= bold >₹<!-- -->210</h5>,
 <h5 class="Text__StyledText-sc-oo0kvp-0 hiHdyy" color="greyBase" font-size="24px" font-w
eight="bold">₹<!-- -->192</h5>]
```

In [ ]:

```
price_list=[]
for i in range(0,len(Price)):
  price_list.append(Price[i].get_text())
price_list
```

Out[ ]:

```
['₹192',
 '₹342',
 '₹207',
 '₹0 onwards',
 '₹234',
 '₹464',
 '₹239',
 '₹227',
 '₹282',
 '₹214',
 '₹171',
 '₹233',
 '₹137',
 '₹274',
 '₹200',
 '₹234',
 '₹206',
 '₹232',
 '₹210',
 '₹192']
```

In [ ]:

```
import pandas as pd
df=pd.DataFrame()
```

In [ ]:

```
df['Names']=Bag_names
df['Price']=price_list
```

In [ ]:

```
df
```

Out[ ]:

	Names	Price
0	Elegant Fashionable Women Handbags	₹192
1	Gorgeous Attractive Women Handbags	₹342
2	Classic Fancy Women Handbags	₹207
3	Classic Versatile Women Handbags	₹0 onwards
4	Gorgeous Versatile Women Handbags	₹234
5	Trendy Fancy Women Handbags	₹464
6	Elegant Versatile Women Handbags	₹239
7	Ravishing Stylish Women Handbags	₹227
8	Ravishing Alluring Women Handbags	₹282
9	Classic Versatile Women Handbags	₹214
10	Elegant Women Handbags	₹171
11	Elite Fashionable Women Handbags	₹233

	Names	Price
12	Voguish Fancy Women Handbags	₹137
13	Classic Stylish Women Handbags	₹274
14	Elegant Attractive Women Handbags	₹200
15	Elite Versatile Women Handbags	₹234
16	Classic Versatile Women Handbags	₹206
17	Elite Fancy Women Handbags	₹232
18	Trendy Fashionable Women Handbags	₹210
19	Gorgeous Classy Women Handbags	₹192

In [ ]:

```
df.to_csv('Bags.csv',index=False)
```

In [ ]:

```
pd.read_csv("Bags.csv")
```

Out[ ]:

	Names	Price
0	Elegant Fashionable Women Handbags	₹192
1	Gorgeous Attractive Women Handbags	₹342
2	Classic Fancy Women Handbags	₹207
3	Classic Versatile Women Handbags	₹0 onwards
4	Gorgeous Versatile Women Handbags	₹234
5	Trendy Fancy Women Handbags	₹464
6	Elegant Versatile Women Handbags	₹239
7	Ravishing Stylish Women Handbags	₹227
8	Ravishing Alluring Women Handbags	₹282
9	Classic Versatile Women Handbags	₹214
10	Elegant Women Handbags	₹171
11	Elite Fashionable Women Handbags	₹233
12	Voguish Fancy Women Handbags	₹137
13	Classic Stylish Women Handbags	₹274
14	Elegant Attractive Women Handbags	₹200
15	Elite Versatile Women Handbags	₹234
16	Classic Versatile Women Handbags	₹206
17	Elite Fancy Women Handbags	₹232
18	Trendy Fashionable Women Handbags	₹210
19	Gorgeous Classy Women Handbags	₹192

RESULT:

The program executed successfully and obtained the output.

PROGRAM - 15

AIM :

Problems may be designed for the following topics so that students can get hands on experience in using python for natural language processing: • Part of Speech tagging • N-gram and smoothening • Chunking

DATASET :

Reviews.csv

```
import nltk
from nltk import word_tokenize
nltk.download('punkt')
nltk.download('average_perceptron_tagger')
nltk.download('tagsets')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Error loading average_perceptron_tagger: Package
[nltk_data]     'average_perceptron_tagger' not found in index
[nltk_data] Downloading package tagsets to /root/nltk_data...
[nltk_data]   Package tagsets is already up-to-date!
True
```

```
nltk.help.upenn_tagset()
```

```
$: dollar
    $ -$ --$ A$ C$ HK$ M$ NZ$ S$ U.S.$ US$
'': closing quotation mark
    ' ''
(: opening parenthesis
    ( [ {
): closing parenthesis
    ) ] }
,: comma
    ,
--: dash
    --
.: sentence terminator
    . ! ?
:: colon or ellipsis
    : ; ...
CC: conjunction, coordinating
    & 'n and both but either et for less minus neither nor or plus so
    therefore times v. versus vs. whether yet
```

```
    CD: numeral, cardinal
        mid-1890 nine-thirty forty-two one-tenth ten million 0.5 one forty-
        seven 1987 twenty '79 zero two 78-degrees eighty-four IX '60s .025
        fifteen 271,124 dozen quintillion DM2,000 ...
    DT: determiner
        all an another any both del each either every half la many much nary
        neither no some such that the them these this those
    EX: existential there
        there
    FW: foreign word
        gemeinschaft hund ich jeux habeas Haementeria Herr K'ang-si vous
        lutihaw alai je jour objets salutaris fille quibusdam pas trop Monte
        terram fiche oui corporis ...
    IN: preposition or conjunction, subordinating
        astride among uppon whether out inside pro despite on by throughout
        below within for towards near behind atop around if like until below
        next into if beside ...
    JJ: adjective or numeral, ordinal
        third ill-mannered pre-war regrettable oiled calamitous first separable
        ectoplasmic battery-powered participatory fourth still-to-be-named
        multilingual multi-disciplinary ...
    JJR: adjective, comparative
        bleaker braver breezier briefer brighter brisker broader bumper busier
        calmer cheaper choosier cleaner clearer closer colder commoner costlier
        cozier creamier crunchier cuter ...
    JJS: adjective, superlative
        calmest cheapest choicest classiest cleanest clearest closest commonest
        corniest costliest crassest creepiest crudest cutest darkest deadliest
        dearest deepest densest dinkiest ...
    LS: list item marker
        A A. B B. C C. D E F First G H I J K One SP-44001 SP-44002 SP-44005
        SP-44007 Second Third Three Two * a b c d first five four one six three
        two
    MD: modal auxiliary
        can cannot could couldn't dare may might must need ought shall should
        shouldn't will would
    NN: noun, common, singular or mass
        common-carrier cabbage knuckle-duster Casino afghan shed thermostat
        investment slide humour falloff slick wind hyena override subhumanity
```

```
import nltk
nltk.download('averaged_perceptron_tagger')
sentence=word_tokenize("Third wave of corona virus is here")
nltk.pos_tag(sentence)
```

```
    [nltk_data] Downloading package averaged_perceptron_tagger to
    [nltk_data]     /root/nltk_data...
    [nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
    [('Third', 'JJ'),
     ('wave', 'NN'),
     ('of', 'IN'),
     ('corona', 'NN'),
     ('virus', 'NN'),
     ('is', 'VBZ'),
     ('here', 'RB')]
```

```
#n grams in sentence analysis
import numpy as np
import pandas as pd
from sklearn import model_selection,naive_bayes,svm


df=pd.read_csv('/content/sample_data/Reviews.csv',engine='python',error_bad_lines=False)
```

```
/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:2882: FutureWarr


  exec(code_obj, self.user_global_ns, self.user_ns)
```

```
df.head()
```

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessD
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144826 entries, 0 to 144825
Data columns (total 10 columns):
 #   Column                  Non-Null Count   Dtype 
---  ------                  --------------   ----- 
 0   Id                      144826 non-null  int64 
 1   ProductId               144826 non-null  object
 2   UserId                  144826 non-null  object
 3   ProfileName             144820 non-null  object
 4   HelpfulnessNumerator    144826 non-null  int64 
 5   HelpfulnessDenominator  144826 non-null  int64 
 6   Score                   144826 non-null  int64 
 7   Time                    144826 non-null  int64 
 8   Summary                 144821 non-null  object
 9   Text                    144826 non-null  object
dtypes: int64(5), object(5)
memory usage: 11.0+ MB
```

```
df.shape
```

```
(144826, 10)
```

```
df.isnull().sum()
```

```
Id                         0
ProductId                  0
UserId                     0
ProfileName                6
HelpfulnessNumerator       0
HelpfulnessDenominator     0
Score                      0
Time                       0
Summary                    5
Text                       0
dtype: int64
```

```
df.dropna(inplace=True)
```

```
df.isnull().sum()
```

```
Id                         0
ProductId                  0
UserId                     0
ProfileName                0
HelpfulnessNumerator       0
HelpfulnessDenominator     0
Score                      0
Time                       0
Summary                    0
Text                       0
dtype: int64
```

```
df['Score'].value_counts()
```

```
5    91090
4    21197
1    13315
3    11407
2     7806
Name: Score, dtype: int64
```

```
df['postive ratings']=np.where(df['Score']>=3,1,0)
```

```
df.head()
```

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessD
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	

```
df['postive ratings'].value_counts()
```

```
1    123694
0     21121
Name: postive ratings, dtype: int64
```

```
#highly imbalanced
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(df["Text"],df["postive ratings"],random_state=
```

```
print(x_train)
```

```
7417      This is our second time using your black winte...
92282     I think there has been enough written about th...
135398    With all the coffees around it is refreshing t...
128933    I bought a 12 poack of these because the price...
49792     I really like this thinkThin bar but was surpr...
                              ...
122116    This product, Xylosweet, has no aftertaste and...
55371     I will admit that I was worried about buying t...
123022    Frist off i order it took 4days to come and it...
59363     Love these tea biscuits. The price was great f...
103912    This product is great for the price. My MAJOR ...
Name: Text, Length: 108611, dtype: object
```

```
print(y_train)
```

```
7417      1
92282     1
135398    1
128933    1
49792     1
          ..
122116    1
```

```
55371     1
123022    1
59363     1
103912    1
Name: postive ratings, Length: 108611, dtype: int64
```

```
df.head()
```

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessD
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	

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

```
vector=CountVectorizer(min_df=5,ngram_range=(1,2)).fit(x_train)
```

```
vector.get_feature_names()
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning
  warnings.warn(msg, category=FutureWarning)
['00',
 '00 also',
 '00 am',
 '00 and',
 '00 at',
 '00 bag',
 '00 bottle',
 '00 box',
 '00 br',
 '00 but',
 '00 can',
 '00 cheaper',
 '00 dollars',
 '00 each',
 '00 for',
 '00 if',
 '00 in',
 '00 including',
 '00 is',
```

```
 '00 it',
 '00 later',
 '00 less',
 '00 more',
 '00 not',
 '00 of',
 '00 off',
 '00 on',
 '00 or',
 '00 per',
 '00 plus',
 '00 pm',
 '00 pound',
 '00 shipping',
 '00 since',
 '00 so',
 '00 that',
 '00 the',
 '00 they',
 '00 this',
 '00 to',
 '00 was',
 '00 which',
 '00 with',
 '00 worth',
 '00 you',
 '000',
 '000 000',
 '000 calorie',
 '000 calories',
 '000 feet',
 '000 mg',
 '000 miles',
 '000 years',
 '00pm',
 '01',
```

```
len(vector.get_feature_names())
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: F
  warnings.warn(msg, category=FutureWarning)
196616
```

```
x_train_vectorized=vector.transform(x_train)
```

```
from sklearn.naive_bayes import MultinomialNB
model=MultinomialNB()
model.fit(x_train_vectorized,y_train)
```

```
MultinomialNB()
```

```
pred=model.predict(vector.transform(x_test))


print(pred)
```

```
     [1 1 1 ... 1 1 1]
```

```
from sklearn.metrics import classification_report,confusion_matrix


print(confusion_matrix(y_test,pred))
print(classification_report(y_test,pred))
```

```
     [[ 4207  1094]
      [ 1890 29013]]
                   precision    recall  f1-score   support

                0       0.69      0.79      0.74      5301
                1       0.96      0.94      0.95     30903

         accuracy                           0.92     36204
        macro avg       0.83      0.87      0.84     36204
     weighted avg       0.92      0.92      0.92     36204
```

```
#auc score
from sklearn.metrics import roc_auc_score
print("AUC score is:",roc_auc_score(y_test,pred))
```

```
     AUC score is: 0.8662323668958428
```

RESULT:

The program executed successfully and obtained the output.