Name: - L Prathyusha

In [2]:

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
```

HTML Code of the Website that i created

a) scrape all the Information like, heading, image files, content, etc.,

In [3]:

```
myFile=open('C:/Users/Prathyu Lachireddy/Desktop/Assignment/19111344_L Prathyusha.html','r'
soup=BeautifulSoup(myFile,"html5lib")
print(soup.prettify())
```

```
<html lang="en">
 <head>
  <title>
   Name : - L Prathyusha (19111344), WSDA
  </title>
  <meta charset="utf-8"/>
  <meta content="width=device-width, initial-scale=1" name="viewport"/>
  <style>
   body {
  font-family: Arial, Helvetica, sans-serif;
}
  </style>
 </head>
 <body bgcolor="E6E6FA">
  <h1>
   Name : - L Prathyusha (19111344), WSDA
  </h1>
  <h1>
   Heading, paragraph, image (gif), background colour, Table, List
```

Title and Head of the data

In [4]:

```
page_title=soup.title.text
page_title
```

Out[4]:

```
'Name : - L Prathyusha (19111344), WSDA '
```

In [5]:

```
page_title=soup.title
page_title
```

Out[5]:

```
<title>Name : - L Prathyusha (19111344), WSDA </title>
```

In [10]:

```
# Extract body of page
page_body = soup.body
page_body
```

Out[10]:

```
<body bgcolor="E6E6FA">


<h1>Name : - L Prathyusha (19111344), WSDA </h1>
<h1>Heading, paragraph, image (gif), background colour, Table, List</h1>

<img alt="Web Scraping" height="317" src="C:\Users\Prathyu Lachireddy\Desk
top\Assignment\i.jpeg" width="480"/>
<img alt="Web Scraping" height="317" src="C:\Users\Prathyu Lachireddy\Desk
top\Assignment\p.gif" width="480"/>
<img alt="Web Scraping" height="317" src="C:\Users\Prathyu Lachireddy\Desk
top\Assignment\n.jpeg" width="480"/>

<style>
.city {
  background-color: DarkSlateGray;
  color: white;
  border: 2px solid black;
```

In [11]:

```
# Extract head of page
page_head = soup.head
page_head
```

Out[11]:

```
<head>
<title>Name : - L Prathyusha (19111344), WSDA </title>
<meta charset="utf-8"/>
<meta content="width=device-width, initial-scale=1" name="viewport"/>
<style>
body {
  font-family: Arial, Helvetica, sans-serif;
}
</style>
</head>
```

Style

In [13]:

```
# To find all style types
style=soup.find_all("style")
style
```

Out[13]:

```
[<style>
 body {
   font-family: Arial, Helvetica, sans-serif;
 }
 </style>,
 <style>
 .city {
   background-color: DarkSlateGray;
   color: white;
   border: 2px solid black;
   margin: 20px;
   padding: 20px;
 }
 </style>,
 <style>
 table {
   font-family: arial, sans-serif;
   border-collapse: collapse;
   width: 100%;
 }

 td, th {
   border: 1px solid #dddddd;
   text-align: left;
   padding: 8px;
 }

 tr:nth-child(even) {
   background-color: #dddddd;
 }
 </style>]
```

Image

In [ ]:

```
import requests
from bs4 import BeautifulSoup
# Make a request
page = requests.get("C:/Users/Prathyu Lachireddy/Desktop/Assignment/19111344_L Prathyusha.h
soup = BeautifulSoup(page.content, 'html.parser')

# Create top_items as empty list
image_data = []

# Extract and store in top_items according to instructions on the left
images = soup.select('img')
for image in images:
    src = image.get('src')
    alt = image.get('alt')
    image_data.append({"src": src, "alt": alt})

print(image_data)
```

Paragraph on biography and importance of analytics

In [24]:

```
soup.find_all('div')
```

Out[24]:

```
[<div class="city">
   <h2>BIOGRAPHY</h2>
   <p>I am L Prathyusha. i am pursuing my undergraduate from Christ Universt
iy lavasa. i really do enjoy dancing, doing art works.
 and i am doing BBA (Business Analytics) and i would like to talk about the
importance of analytics,
 Analytics allow you to quantify the effects of making a change to your mark
eting strategy,
  and that's invaluable to the process of improving and optimizing online ma
rketing campaigns.
  The biggest benefit of utilizing proper analytics is being able to identif
y strengths and weaknesses.
 For example, letâ€™s say you run a blog for your car detailing business. Yo
uâ€™re just starting out, and arenâ€™t sure what kinds of posts will bring y
ou the most traffic, or provide the most value to your readers.
 If youâ€™re using analytics, youâ€™ll be able to measure which blog posts a
ttract the most traffic, which get the least traffic, which have a high boun
ce rate, a low bounce rate, and so on. It will be easy to tell which blog po
sts are performing better or worse than others.
 Key idea is to collect data about the organization and use them to improve
operations. Raw form of data is not of any use.If you are trying to bring an
y significant improvement in your business, then analytics is your best bet
to bring about an informed transformation.
 </p>
 </div>]
```

b) identify the tables available in that and convert into CSV file

Table to dataframe to CSV

In [34]:

```
path = 'C:/Users/Prathyu Lachireddy/Desktop/Assignment/19111344_L Prathyusha.html'

# empty list
data = []

list_header = []
soup = BeautifulSoup(open(path),'html.parser')
header = soup.find_all("table")[0].find("tr")

for items in header:
    try:
        list_header.append(items.get_text())
    except:
        continue

# for getting the data
HTML_data = soup.find_all("table")[0].find_all("tr")[1:]

for element in HTML_data:
    sub_data = []
    for sub_element in element:
        try:
            sub_data.append(sub_element.get_text())
        except:
            continue
    data.append(sub_data)

# Storing the data into Pandas
# DataFrame
dataFrame = pd.DataFrame(data = data, columns = list_header)
dataFrame
```

Out[34]:

	S no	Company	Country	Sector
0	1	Google	USA	Infomation Technology
1	2	Infosys	India	Infomation technology
2	3	WISE	Italy	Medical
3	4	Tesla	USA	Automobile
4	5	ECARX	China	Automotive
5	6	Prada	Italy	Fashion
6	7	Cisco systems inc	USA	Computers and electronics
7	8	Goldman sachs	USA	Financial services
8	9	PriceWater Coopers	USA	Financial services
9	10	Sun pharma	India	pharmaceutical

In [35]:

```
dataFrame.to_csv('19111344_assignment.csv')
```