# WEB SCRAPING ON RAILYATRI.IN WEBSITE BY USING PYTHON

## PYTHON PROJECT

# Vishal .R. Yadav

Data Science and analytics with AI

MOB : 9322267388

EMAIL: yadavvishal8421@gmail.comAA

# TO GET DATA OF BEST TRAIN SCHEDULE TIMING BY SCRAPING AND ANALYSIS OF DATA

- ## AIM:
  - By using Web scraping with Python to find out the best time schedule of trains and availability of seats for our customer
  - For best facilities during traveling with train to our customer guidance

- ## OBJECTIVE:
  - Web scraping on Railyatri.in can help to gather data from website for analysis, and time schedule of train arriving and departure.
  - Also gather the knowledge of train names train numbers and which days of week it is running after scraping data from website

**WHAT IS WEB SCRAPING?**
WEB SCRAPING IS A TERM USED TO DESCRIBE THE USE OF A PROGRAM OR ALGORITHM TO EXTRACT AND PROCESS LARGE AMOUNTS OF DATA FROM THE WEB.
WHETHER YOU ARE A DATA SCIENTIST, ENGINEER, OR ANYBODY WHO ANALYZES LARGE AMOUNTS OF DATASETS, THE ABILITY TO SCRAPE DATA FROM THE WEB IS A USEFUL SKILL TO HAVE.
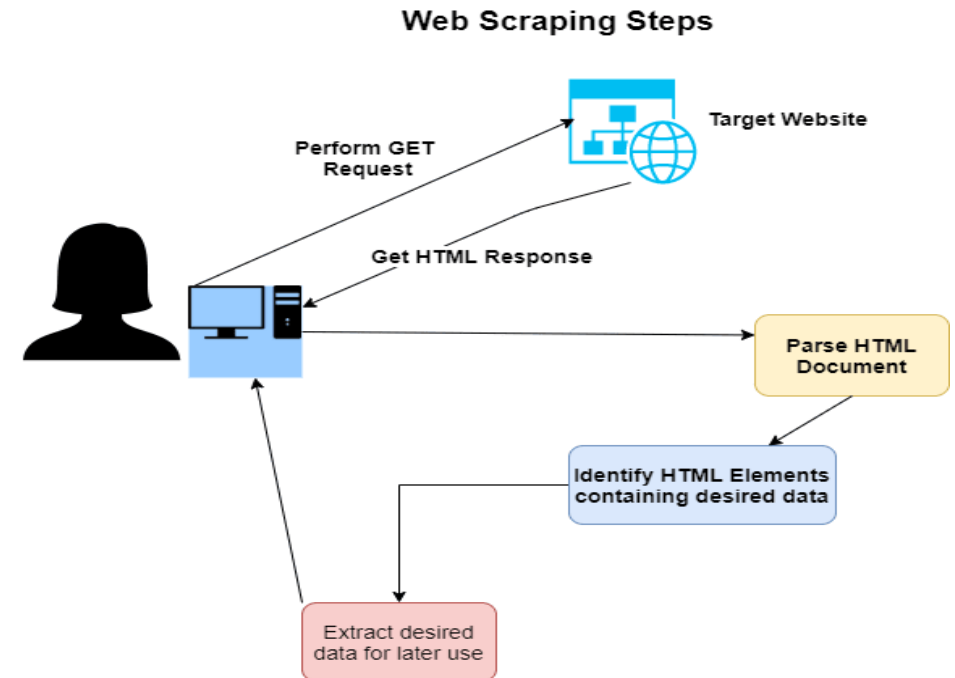


Webpages → Web Scraping → Structured Data

# STEPS REQUIRES TO SCRAP DATA

- To extract data using web scraping with python, I need to follow these basic steps:
- 1. Find the URL that you want to scrape
- 2. Inspecting the Page
- 3. Find the data you want to extract
- 4. Write the code
- 5. Run the code and extract the data
- 6. Store the data in the required format

# LIBRARIES USED FOR WEB SCRAPING

- **BeautifulSoup**: Beautiful Soup is a Python package for parsing HTML and XML documents. It creates parse trees that are helpful to extract the data easily.

- **Pandas**: Pandas is a library used for data manipulation and analysis. It is used to extract the data and store it in the desired format.

- Use **matplotlib** and **seaborn** for data visualization

# Scrappeing data from Railyatri.in website

Let us import all the necessary libraries:

```
In [1]: import pandas as pd
        import requests
        from bs4 import BeautifulSoup
```

The data that we extract is unstructured data. So we'll create empty lists to store them in a structured form,

```
In [28]: Train_Number=[]
         Train_Name=[]
         Train_Type=[]
         Run_Days=[]
         Departure_time=[]
         Arrival_Time=[]
```

# Step 1: Find the URL that I want to scrape

For this example, I am going to scrape the **Railyatri.in** website to extract the Train Number, Train Name, Train Type, Run Days, Departure time, Arrival Time. The URL for this page is https://www.railyatri.in/time-table

```
In [2]: url='https://www.railyatri.in/time-table'
```

```
In [3]: page = requests.get(url)
        soup=BeautifulSoup(page.content,'html.parser')
        soup.prettify
```

```
Out[3]: <bound method Tag.prettify of <!DOCTYPE html>

        <html lang="en">
        <head>
        <meta charset="utf-8"/>
        <meta content="width=device-width, initial-scale=1, minimum-scale=1" name="viewport"/>
        <meta content="https://images.railyatri.in/ry_images_prod/imageCard2-1676384275.png" name="twitter:card"/>
        <meta content="Indian Railways Train Enquiry, Insights &amp; IRCTC eTicket booking - RailYatri" property="og:title">
```
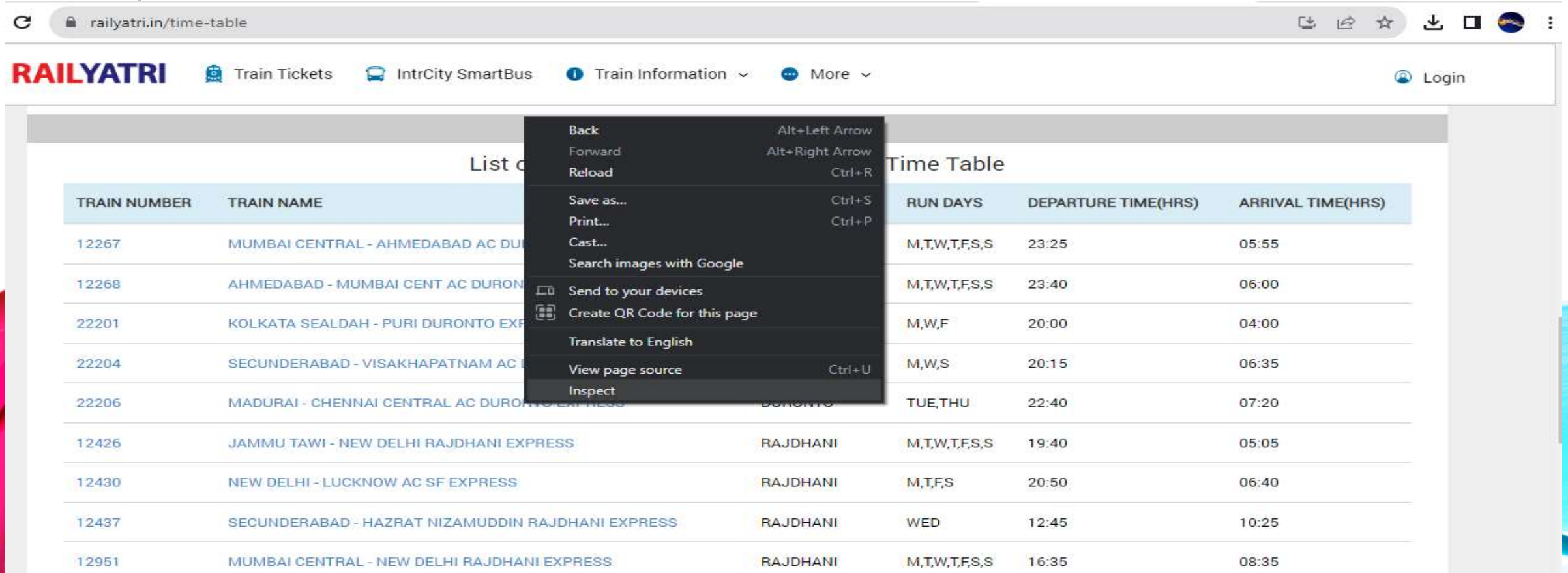
*As we got the Response 200 it means we got the excess from website to scrap data from it¶*

```
In [4]: page = requests.get(url)
        page
```

```
Out[4]: <Response [200]>
```

# STEP 2: INSPECTING THE PAGE

- The data is usually nested in tags. So, we inspect the page to see, under which tag the data we want to scrape is nested. To inspect the page, just right click on the element and click on "Inspect".

Here I got the class in which my data is present as class = "table table-condensed"

The data we want to extract is nested in <div> tags. So, I will find the div tags with those respective class-names, extract the data, and store the data in a variable. Refer to the code below:

```
In [5]: train_detail=soup.find_all(class_="table table-condensed")
print(train_detail)
print(len(train_detail))
```

```
<th>Departure time(hrs)</th>
<th>Arrival Time(hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="/time-table/12267-mumbai-central-ahmedabad-ac-duronto-exp">12267</a></td>
<td><a href="/time-table/12267-mumbai-central-ahmedabad-ac-duronto-exp">MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp</a></td>
<td>Duronto</td>
<td>M,T,W,T,F,S,S</td>
<td> 23:25</td>
<td>05:55</td>
</tr>
<tr>
<td><a href="/time-table/12268-ahmedabad-mumbai-cent-ac-duronto-exp">12268</a></td>
<td><a href="/time-table/12268-ahmedabad-mumbai-cent-ac-duronto-exp">AHMEDABAD - MUMBAI CENT AC Duronto Exp</a></td>
<td>Duronto</td>
<td>M,T,W,T,F,S,S</td>
<td> 23:40</td>
<td>06:00</td>
```

# By using libraries I got the html sourse code data now I will create a new list1 and appending my data in it

```
In [6]: list1=[]
        for i in range(0,len(train_detail)):
            list1.append(train_detail[i].get_text())
        list1[0]
```

Out[6]: '\n\n\nTrain Number\nTrain Name\nTrain Type\nRun Days\nDeparture time(hrs)\nArrival Time(hrs)\n\n\n\n\n12267\nMUMBAI CENTRAL - AHMEDABAD AC Duronto Exp\nDuronto\nM,T,W,T,F,S,S\n 23:25\n05:55\n\n\n12268\nAHMEDABAD - MUMBAI CENT AC Duronto Exp\nDuronto\nM,T,W,T,F,S,S\n 23:40\n06:00\n\n\n22201\nKOLKATA SEALDAH - PURI Duronto Express\nDuronto\nM,W,F\n 20:00\n04:00\n\n\n22204\nSECUNDERABAD - VISAKHAPATNAM AC Duronto Express\nDuronto\nM,W,S\n 20:15\n06:35\n\n\n22206\nMADURAI - CHENNAI CENTRAL AC Duronto Express\nDuronto\nTUE,THU\n 22:40\n07:20\n\n\n12426\nJAMMU TAWI - NEW DELHI Rajdhani Express\nRajdhani\nM,T,W,T,F,S,S\n 19:40\n05:05\n\n\n12430\nNEW DELHI - LUCKNOW AC SF Express\nRajdhani\nM,T,F,S\n 20:50\n06:40\n\n\n12437\nSECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express\nRajdhani\nWED\n 12:45\n10:25\n\n\n12951\nMUMBAI CENTRAL - NEW DELHI Rajdhani Express\nRajdhani\nM,T,W,T,F,S,S\n 16:35\n08:35\n\n\n12953\nMUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti Raj Exp\nRajdhani\nM,T,W,T,F,S,S\n 17:40\n10:55\n\n\n12019\nHOWRAH - RANCHI Shatabdi Express\nShatabdi\nM,T,W,T,F,S\n 06:05\n13:15\n\n\n12038\nLUDHIANA - NEW DELHI Shatabdi Express\nShatabdi\nT,W,T,F,S\n 16:40\n22:10\n\n\n12048\nFIROZPUR - NEW DELHI Shatabdi Express\nShatabdi\nM,S\n 04:00\n21:10\n\n\n12049\nAgra Cantt - Delhi H Nizamuddin Gatimaan Express\nShatabdi\nM,T,W,T,S,S\n 17:50\n19:30\n\n\n12050\nDelhi H Nizamuddin - Agra Cantt Gatimaan Express\nShatabdi\nM,T,W,T,S,S\n 08:10\n09:50\n\n\n11077\nPUNE - JAMMU TAWI Jhelum Express\nMAIL EXPRESS\nM,T,W,T,F,S,S\n 17:20\n10:10\n\n\n12138\nFIROZPUR CANTT - MUMBAI CST Punjab Mail SF Exp\nSUPERFAST\nM,T,W,T,F,S,S\n 21:40\n07:35\n\n\n12295\nBANGALORE CITY - DANAPUR Sanghamitra SF Exp\nSUPERFAST\nM,T,W,T,F,S,S\n 09:00\n09:20\n\n\n12307\nHOWRAH - JODHPUR SF Express\nSUPERFAST\nM,T,W,T,F,S,S\n 23:30\n08:30\n\n\n12424\nNEW DELHI - DIBRUGARH TOWN Rajdhani Express\nRAJDHANI\nM,T,W,T,F,S,S\n 13:55\n07:00\n\n\n12506\nDELHI ANAND VIHAR - GUWAHATI North East Express\nSUPERFAST\nM,T,W,T,F,S,S\n 06:45\n16:50\n\n\n12801\nPURI - NEW DELHI Purushottam SF Express\nSUPERFAST\nM,T,W,T,F,S,S\n 21:45\n04:50\n\n\n12802\nNEW DELHI - PURI Purushottam SF Express\nSUPERFAST\nM,T,W,T,F,S,S\n 22:15\n05:25\n\n\n12833\nAHMEDABAD - HOWRAH SF Express\nSUPERFAST\nM,T,W,T,F,S,S\n 00:15\n13:30\n\n\n12859\nMUMBAI CST - HOWRAH Gitanjali SF Express\nSUPERFAST\nM,T,W,T,F,S,S\n 06:00\n12:30\n\n\n12864\nYESVANTPUR - HOWRAH SF Express\nSUPERFAST\nM,T,W,T,F,S,S\n 19:35\n06:25\n\n\n'

# For Removing \n from our data I will use split function

```
In [7]: list2=[]
        for i in range(0,len(list1)):
            list2.append(list1[i].split("\n"))
        print(list2[0])
```

['', '', '', 'Train Number', 'Train Name', 'Train Type', 'Run Days', 'Departure time(hrs)', 'Arrival Time(hrs)', '', '', '', '', '12267', 'MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp', 'Duronto', 'M,T,W,T,F,S,S', ' 23:25', '05:55', '', '', '12268', 'AHME DABAD - MUMBAI CENT AC Duronto Exp', 'Duronto', 'M,T,W,T,F,S,S', ' 23:40', '06:00', '', '', '22201', 'KOLKATA SEALDAH - PURI Du ronto Express', 'Duronto', 'M,W,F', ' 20:00', '04:00', '', '', '22204', 'SECUNDERABAD - VISAKHAPATNAM AC Duronto Express', 'Dur onto', 'M,W,S', ' 20:15', '06:35', '', '', '22206', 'MADURAI - CHENNAI CENTRAL AC Duronto Express', 'Duronto', 'TUE,THU', ' 22: 40', '07:20', '', '', '12426', 'JAMMU TAWI - NEW DELHI Rajdhani Express', 'Rajdhani', 'M,T,W,T,F,S,S', ' 19:40', '05:05', '', '', '12430', 'NEW DELHI - LUCKNOW AC SF Express', 'Rajdhani', 'M,T,F,S', ' 20:50', '06:40', '', '', '12437', 'SECUNDERABAD - HA ZRAT NIZAMUDDIN Rajdhani Express', 'Rajdhani', 'WED', ' 12:45', '10:25', '', '', '12951', 'MUMBAI CENTRAL - NEW DELHI Rajdhani Express', 'Rajdhani', 'M,T,W,T,F,S,S', ' 16:35', '08:35', '', '', '12953', 'MUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti Raj E xp', 'Rajdhani', 'M,T,W,T,F,S,S', ' 17:40', '10:55', '', '', '12019', 'HOWRAH - RANCHI Shatabdi Express', 'Shatabdi', 'M,T,W,T, F,S', ' 06:05', '13:15', '', '', '12038', 'LUDHIANA - NEW DELHI Shatabdi Express', 'Shatabdi', 'T,W,T,F,S', ' 16:40', '22:10', '', '', '12048', 'FIROZPUR - NEW DELHI Shatabdi Express', 'Shatabdi', 'M,S', ' 04:00', '21:10', '', '', '12049', 'Agra Cantt - Delhi H Nizamuddin Gatimaan Express', 'Shatabdi', 'M,T,W,T,S,S', ' 17:50', '19:30', '', '', '12050', 'Delhi H Nizamuddin - Agra Cantt Gatimaan Express', 'Shatabdi', 'M,T,W,T,S,S', ' 08:10', '09:50', '', '', '11077', 'PUNE - JAMMU TAWI Jhelum Express', 'MA IL EXPRESS', 'M,T,W,T,F,S,S', ' 17:20', '10:10', '', '', '12138', 'FIROZPUR CANTT - MUMBAI CST Punjab Mail SF Exp', 'SUPERFAS T', 'M,T,W,T,F,S,S', ' 21:40', '07:35', '', '', '12295', 'BANGALORE CITY - DANAPUR Sanghamitra SF Exp', 'SUPERFAST', 'M,T,W,T, F,S,S', ' 09:00', '09:20', '', '', '12307', 'HOWRAH - JODHPUR SF Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 23:30', '08:30', '', '', '12424', 'NEW DELHI - DIBRUGARH TOWN Rajdhani Express', 'RAJDHANI', 'M,T,W,T,F,S,S', ' 13:55', '07:00', '', '', '12506', 'D ELHI ANAND VIHAR - GUWAHATI North East Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 06:45', '16:50', '', '', '12801', 'PURI - NEW DELHI Purushottam SF Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 21:45', '04:50', '', '', '12802', 'NEW DELHI - PURI Purushottam SF Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 22:15', '05:25', '', '', '12833', 'AHMEDABAD - HOWRAH SF Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 00:15', '13:30', '', '', '12859', 'MUMBAI CST - HOWRAH Gitanjali SF Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 06:00', '12:30', '', '', '12864', 'YESVANTPUR - HOWRAH SF Express', 'SUPERFAST', 'M,T,W,T,F,S,S', ' 19:35', '06:25', '', '', '']

For making empty list of unwanted data  I will use split function to split out by comma ( , ) separated values list.

```
In [8]: list3=[]
        for i in range(0,len(list2)):
            for n in range(0,len(list2[i])):
                list3.append(list2[i][n].split(','))
        print(list3)
```

[[''], [''], [''], ['Train Number'], ['Train Name'], ['Train Type'], ['Run Days'], ['Departure time(hrs)'], ['Arrival Time(hrs)'], [''], [''], [''], [''], ['12267'], ['MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp'], ['Duronto'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 23:25'], ['05:55'], [''], [''], ['12268'], ['AHMEDABAD - MUMBAI CENT AC Duronto Exp'], ['Duronto'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 23:40'], ['06:00'], [''], [''], ['22201'], ['KOLKATA SEALDAH - PURI Duronto Express'], ['Duronto'], ['M', 'W', 'F'], [' 20:00'], ['04:00'], [''], [''], ['22204'], ['SECUNDERABAD - VISAKHAPATNAM AC Duronto Express'], ['Duronto'], ['M', 'W', 'S'], [' 20:15'], ['06:35'], [''], [''], ['22206'], ['MADURAI - CHENNAI CENTRAL AC Duronto Express'], ['Duronto'], ['TUE', 'THU'], [' 22:40'], ['07:20'], [''], [''], ['12426'], ['JAMMU TAWI - NEW DELHI Rajdhani Express'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 19:40'], ['05:05'], [''], [''], ['12430'], ['NEW DELHI - LUCKNOW AC SF Express'], ['Rajdhani'], ['M', 'T', 'F', 'S'], [' 20:50'], ['06:40'], [''], [''], ['12437'], ['SECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express'], ['Rajdhani'], ['WED'], [' 12:45'], ['10:25'], [''], [''], ['12951'], ['MUMBAI CENTRAL - NEW DELHI Rajdhani Express'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 16:35'], ['08:35'], [''], [''], ['12953'], ['MUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti Raj Exp'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 17:40'], ['10:55'], [''], [''], ['12019'], ['HOWRAH - RANCHI Shatabdi Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'F', 'S'], [' 06:05'], ['13:15'], [''], [''], ['12038'], ['LUDHIANA - NEW DELHI Shatabdi Express'], ['Shatabdi'], ['T', 'W', 'T', 'F', 'S'], [' 16:40'], ['22:10'], [''], [''], ['12048'], ['FIROZPUR - NEW DELHI Shatabdi Express'], ['Shatabdi'], ['M', 'S'], [' 04:00'], ['21:10'], [''], [''], ['12049'], ['Agra Cantt - Delhi H Nizamuddin Gatimaan Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'S', 'S'], [' 17:50'], ['19:30'], [''], [''], ['12050'], ['Delhi H Nizamuddin - Agra Cantt Gatimaan Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'S', 'S'], [' 08:10'], ['09:50'], [''], [''], ['11077'], ['PUNE - JAMMU TAWI Jhelum Express'], ['MAIL EXPRESS'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 17:20'], ['10:10'], [''], [''], ['12138'], ['FIROZPUR CANTT - MUMBAI CST Punjab Mail SF Ex

For making empty list of unwanted data I will use split function again to split out semi colum ( ' ' ) from my data to make out its length as zero (0)

```
In [9]: for i in range(0,len(list3)):
            for n in range(0,len(list3[i])):
                if list3[i][n]=='':
                    list3[i].remove('')
        print(len(list3))
        print(list3)
```

222
[[], [], [], ['Train Number'], ['Train Name'], ['Train Type'], ['Run Days'], ['Departure time(hrs)'], ['Arrival Time(hrs)'],
[], [], [], [], ['12267'], ['MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp'], ['Duronto'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'],
[' 23:25'], ['05:55'], [], [], ['12268'], ['AHMEDABAD - MUMBAI CENT AC Duronto Exp'], ['Duronto'], ['M', 'T', 'W', 'T', 'F',
'S', 'S'], [' 23:40'], ['06:00'], [], [], ['22201'], ['KOLKATA SEALDAH - PURI Duronto Express'], ['Duronto'], ['M', 'W',
'F'], [' 20:00'], ['04:00'], [], [], ['22204'], ['SECUNDERABAD - VISAKHAPATNAM AC Duronto Express'], ['Duronto'], ['M', 'W',
'S'], [' 20:15'], ['06:35'], [], [], ['22206'], ['MADURAI - CHENNAI CENTRAL AC Duronto Express'], ['Duronto'], ['TUE', 'TH
U'], [' 22:40'], ['07:20'], [], [], ['12426'], ['JAMMU TAWI - NEW DELHI Rajdhani Express'], ['Rajdhani'], ['M', 'T', 'W',
'T', 'F', 'S', 'S'], [' 19:40'], ['05:05'], [], [], ['12430'], ['NEW DELHI - LUCKNOW AC SF Express'], ['Rajdhani'], ['M',
'T', 'F', 'S'], [' 20:50'], ['06:40'], [], [], ['12437'], ['SECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express'], ['Rajdhan
i'], ['WED'], [' 12:45'], ['10:25'], [], [], ['12951'], ['MUMBAI CENTRAL - NEW DELHI Rajdhani Express'], ['Rajdhani'], ['M',
'T', 'W', 'T', 'F', 'S', 'S'], [' 16:35'], ['08:35'], [], [], ['12953'], ['MUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti Raj
Exp'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 17:40'], ['10:55'], [], [], ['12019'], ['HOWRAH - RANCHI Shatabd
i Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'F', 'S'], [' 06:05'], ['13:15'], [], [], ['12038'], ['LUDHIANA - NEW DELHI S
hatabdi Express'], ['Shatabdi'], ['T', 'W', 'T', 'F', 'S'], [' 16:40'], ['22:10'], [], [], ['12048'], ['FIROZPUR - NEW DELHI
Shatabdi Express'], ['Shatabdi'], ['M', 'S'], [' 04:00'], ['21:10'], [], [], ['12049'], ['Agra Cantt - Delhi H Nizamuddin Gat
imaan Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'S', 'S'], [' 17:50'], ['19:30'], [], [], ['12050'], ['Delhi H Nizamuddin
- Agra Cantt Gatimaan Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'S', 'S'], [' 08:10'], ['09:50'], [], [], ['11077'], ['PU
NE - JAMMU TAWI Jhelum Express'], ['MAIL EXPRESS'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 17:20'], ['10:10'], [], [], ['121

# For Removing empty list I use REMOVE function

```
In [29]:  for i in range(len(list3)-1,-1,-1):
              if len(list3[i])==0:
                  list3.remove(list3[i])
          print(list3)
```

[['12267'], ['MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp'], ['Duronto'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 23:25'], ['0
5:55'], ['12268'], ['AHMEDABAD - MUMBAI CENT AC Duronto Exp'], ['Duronto'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 23:40'],
['06:00'], ['22201'], ['KOLKATA SEALDAH - PURI Duronto Express'], ['Duronto'], ['M', 'W', 'F'], [' 20:00'], ['04:00'], ['2220
4'], ['SECUNDERABAD - VISAKHAPATNAM AC Duronto Express'], ['Duronto'], ['M', 'W', 'S'], [' 20:15'], ['06:35'], ['22206'], ['M
ADURAI - CHENNAI CENTRAL AC Duronto Express'], ['Duronto'], ['TUE', 'THU'], [' 22:40'], ['07:20'], ['12426'], ['JAMMU TAWI -
NEW DELHI Rajdhani Express'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 19:40'], ['05:05'], ['12430'], ['NEW DELH
I - LUCKNOW AC SF Express'], ['Rajdhani'], ['M', 'T', 'F', 'S'], [' 20:50'], ['06:40'], ['12437'], ['SECUNDERABAD - HAZRAT NI
ZAMUDDIN Rajdhani Express'], ['Rajdhani'], ['WED'], [' 12:45'], ['10:25'], ['12951'], ['MUMBAI CENTRAL - NEW DELHI Rajdhani E
xpress'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 16:35'], ['08:35'], ['12953'], ['MUMBAI CENT - HAZRAT NIZAMUD
DIN August Kranti Raj Exp'], ['Rajdhani'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 17:40'], ['10:55'], ['12019'], ['HOWRAH -
RANCHI Shatabdi Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'F', 'S'], [' 06:05'], ['13:15'], ['12038'], ['LUDHIANA - NEW D
ELHI Shatabdi Express'], ['Shatabdi'], ['T', 'W', 'T', 'F', 'S'], [' 16:40'], ['22:10'], ['12048'], ['FIROZPUR - NEW DELHI Sh
atabdi Express'], ['Shatabdi'], ['M', 'S'], [' 04:00'], ['21:10'], ['12049'], ['Agra Cantt - Delhi H Nizamuddin Gatimaan Expr
ess'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'S', 'S'], [' 17:50'], ['19:30'], ['12050'], ['Delhi H Nizamuddin - Agra Cantt Gati
maan Express'], ['Shatabdi'], ['M', 'T', 'W', 'T', 'S', 'S'], [' 08:10'], ['09:50'], ['11077'], ['PUNE - JAMMU TAWI Jhelum Ex
press'], ['MAIL EXPRESS'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 17:20'], ['10:10'], ['12138'], ['FIROZPUR CANTT - MUMBAI C
ST Punjab Mail SF Exp'], ['SUPERFAST'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 21:40'], ['07:35'], ['12295'], ['BANGALORE CI
TY - DANAPUR Sanghamitra SF Exp'], ['SUPERFAST'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 09:00'], ['09:20'], ['12307'], ['HO
WRAH - JODHPUR SF Express'], ['SUPERFAST'], ['M', 'T', 'W', 'T', 'F', 'S', 'S'], [' 23:30'], ['08:30'], ['12424'], ['NEW DELH

# Removing unwanted data from our list by indexing

*Now removing the Detail list all values in our main list before passing it in a loop*

```
In [13]: print(list3[0:6])
         list3[0:6]=[]        # for removing values from index 0 to 6
         list3
         len(list3)

         [['Train Number'], ['Train Name'], ['Train Type'], ['Run Days'], ['Departure time(hrs)'], ['Arrival Time(hrs)']]

Out[13]: 156
```

# Creating empty list for storing our data after passing out it

*Empty list for appending our different data*

```
In [14]: Train_Number=[]
         Train_Name=[]
         Train_Type=[]
         Run_Days=[]
         Departure_time=[]
         Arrival_Time=[]
```

# AFTER CREATING A LIST I GOT MY HOLE DATA IN A SINGLE LIST

AS I SEE IN MY LIST MY DATA WAS ASSEMBLE LINE BY LINE AS TRAIN NUMBER, NAME, TYPE, RUN DAYS, DEPARTURE TIME, ARRIVAL TIME AND THEN ITS REPEATS AGAIN BY SEQUENCE WISE

**So to over ciome with this I will use fore-loop and hopping by index of my data to append it in my lists**

```
In [18]:   #append values in lisst one by one by using hopping or jumper in main list

           for i in range(0,len(list3),6):
               Train_Number.append(list3[i])

           for i in range(1,len(list3),6):
               Train_Name.append(list3[i])

           for i in range(2,len(list3),6):
               Train_Type.append(list3[i])

           for i in range(3,len(list3),6):
               Run_Days.append(list3[i])

           for i in range(4,len(list3),6):
               Departure_time.append(list3[i])

           for i in range(5,len(list3),6):
               Arrival_Time.append(list3[i])
```

# Appending the values in a list by using append function

```
In [15]: for i in range(0,len(list3),6):
             Train_Number.append(list3[i])

         print(Train_Number)
```

[['12267'], ['12268'], ['22201'], ['22204'], ['22206'], ['12426'], ['12430'], ['12437'], ['12951'], ['12953'], ['12019'], ['120
38'], ['12048'], ['12049'], ['12050'], ['11077'], ['12138'], ['12295'], ['12307'], ['12424'], ['12506'], ['12801'], ['12802'],
['12833'], ['12859'], ['12864']]

# After appending the values I got a list inside a list so for removing this extra list I will use fore-loop

```
In [16]: for i in range(0,len(Train_Number)):
             for n in range(0,len(Train_Number[i])):
                 Train_Number[i]=Train_Number[i][n]
         print(Train_Number)
```

['12267', '12268', '22201', '22204', '22206', '12426', '12430', '12437', '12951', '12953', '12019', '12038', '12048', '12049',
'12050', '11077', '12138', '12295', '12307', '12424', '12506', '12801', '12802', '12833', '12859', '12864']

# We will do above operation with all other remaining lists also
## FOR TRAIN NAME

```
In [19]: for i in range(0,len(list3),6):
             Train_Name.append(list3[i])

         print(Train_Name)
```

```
[['MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp'], ['AHMEDABAD - MUMBAI CENT AC Duronto Exp'], ['KOLKATA SEALDAH - PURI Duronto Ex
press'], ['SECUNDERABAD - VISAKHAPATNAM AC Duronto Express'], ['MADURAI - CHENNAI CENTRAL AC Duronto Express'], ['JAMMU TAWI -
NEW DELHI Rajdhani Express'], ['NEW DELHI - LUCKNOW AC SF Express'], ['SECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express'], ['M
UMBAI CENTRAL - NEW DELHI Rajdhani Express'], ['MUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti Raj Exp'], ['HOWRAH - RANCHI Shat
abdi Express'], ['LUDHIANA - NEW DELHI Shatabdi Express'], ['FIROZPUR - NEW DELHI Shatabdi Express'], ['Agra Cantt - Delhi H Ni
zamuddin Gatimaan Express'], ['Delhi H Nizamuddin - Agra Cantt Gatimaan Express'], ['PUNE - JAMMU TAWI Jhelum Express'], ['FIRO
ZPUR CANTT - MUMBAI CST Punjab Mail SF Exp'], ['BANGALORE CITY - DANAPUR Sanghamitra SF Exp'], ['HOWRAH - JODHPUR SF Express'],
['NEW DELHI - DIBRUGARH TOWN Rajdhani Express'], ['DELHI ANAND VIHAR - GUWAHATI North East Express'], ['PURI - NEW DELHI Purush
ottam SF Express'], ['NEW DELHI - PURI Purushottam SF Express'], ['AHMEDABAD - HOWRAH SF Express'], ['MUMBAI CST - HOWRAH Gitan
jali SF Express'], ['YESVANTPUR - HOWRAH SF Express'], ['12267'], ['12268'], ['22201'], ['22204'], ['22206'], ['12426'], ['1243
0'], ['12437'], ['12951'], ['12953'], ['12019'], ['12038'], ['12048'], ['12049'], ['12050'], ['11077'], ['12138'], ['12295'],
['12307'], ['12424'], ['12506'], ['12801'], ['12802'], ['12833'], ['12859'], ['12864']]
```

```
In [20]: for i in range(0,len(Train_Name)):
             for n in range(0,len(Train_Name[i])):
                 Train_Name[i]=Train_Name[i][n]
         print(Train_Name)
```

```
['MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp', 'AHMEDABAD - MUMBAI CENT AC Duronto Exp', 'KOLKATA SEALDAH - PURI Duronto Expres
s', 'SECUNDERABAD - VISAKHAPATNAM AC Duronto Express', 'MADURAI - CHENNAI CENTRAL AC Duronto Express', 'JAMMU TAWI - NEW DELHI
Rajdhani Express', 'NEW DELHI - LUCKNOW AC SF Express', 'SECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express', 'MUMBAI CENTRAL -
NEW DELHI Rajdhani Express', 'MUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti Raj Exp', 'HOWRAH - RANCHI Shatabdi Express', 'LUDH
IANA - NEW DELHI Shatabdi Express', 'FIROZPUR - NEW DELHI Shatabdi Express', 'Agra Cantt - Delhi H Nizamuddin Gatimaan Expres
s', 'Delhi H Nizamuddin - Agra Cantt Gatimaan Express', 'PUNE - JAMMU TAWI Jhelum Express', 'FIROZPUR CANTT - MUMBAI CST Punjab
Mail SF Exp', 'BANGALORE CITY - DANAPUR Sanghamitra SF Exp', 'HOWRAH - JODHPUR SF Express', 'NEW DELHI - DIBRUGARH TOWN Rajdhan
i Express', 'DELHI ANAND VIHAR - GUWAHATI North East Express', 'PURI - NEW DELHI Purushottam SF Express', 'NEW DELHI - PURI Pur
ushottam SF Express', 'AHMEDABAD - HOWRAH SF Express', 'MUMBAI CST - HOWRAH Gitanjali SF Express', 'YESVANTPUR - HOWRAH SF Expr
ess', '12267', '12268', '22201', '22204', '22206', '12426', '12430', '12437', '12951', '12953', '12019', '12038', '12048', '120
49', '12050', '11077', '12138', '12295', '12307', '12424', '12506', '12801', '12802', '12833', '12859', '12864']
```

# FOR TRAIN TYPE LIST:

```
In [21]:  for i in range(0,len(list3),6):
              Train_Type.append(list3[i])

          print(Train_Type)
```

[['Duronto'], ['Duronto'], ['Duronto'], ['Duronto'], ['Duronto'], ['Rajdhani'], ['Rajdhani'], ['Rajdhani'], ['Rajdhani'], ['Raj
dhani'], ['Shatabdi'], ['Shatabdi'], ['Shatabdi'], ['Shatabdi'], ['Shatabdi'], ['MAIL EXPRESS'], ['SUPERFAST'], ['SUPERFAST'],
['SUPERFAST'], ['RAJDHANI'], ['SUPERFAST'], ['SUPERFAST'], ['SUPERFAST'], ['SUPERFAST'], ['SUPERFAST'], ['SUPERFAST'], ['1226
7'], ['12268'], ['22201'], ['22204'], ['22206'], ['12426'], ['12430'], ['12437'], ['12951'], ['12953'], ['12019'], ['12038'],
['12048'], ['12049'], ['12050'], ['11077'], ['12138'], ['12295'], ['12307'], ['12424'], ['12506'], ['12801'], ['12802'], ['1283
3'], ['12859'], ['12864']]

```
In [22]:  for i in range(0,len(Train_Type)):
              for n in range(0,len(Train_Type[i])):
                  Train_Type[i]=Train_Type[i][n]
          print(Train_Type)
```

['Duronto', 'Duronto', 'Duronto', 'Duronto', 'Duronto', 'Rajdhani', 'Rajdhani', 'Rajdhani', 'Rajdhani', 'Rajdhani', 'Shatabdi',
'Shatabdi', 'Shatabdi', 'Shatabdi', 'Shatabdi', 'MAIL EXPRESS', 'SUPERFAST', 'SUPERFAST', 'SUPERFAST', 'RAJDHANI', 'SUPERFAST',
'SUPERFAST', 'SUPERFAST', 'SUPERFAST', 'SUPERFAST', 'SUPERFAST', '12267', '12268', '22201', '22204', '22206', '12426', '12430',
'12437', '12951', '12953', '12019', '12038', '12048', '12049', '12050', '11077', '12138', '12295', '12307', '12424', '12506',
'12801', '12802', '12833', '12859', '12864']

# AFTER CREATING LIST OF DATA WE FILL IT IN A DATAFRAME BY USING PANDAS

```
In [24]: import pandas as pd
         Train_Schedule=pd.DataFrame({
             'Train Number':Train_Number,
             'Train Name':Train_Name,
             'Train Type':Train_Type,
             'Run Days':Run_Days,
             'Departure time':Departure_time,
             'Arrival Time':Arrival_Time},index=range(101, len(Train_Number)+101))
         Train_Schedule
```

Out[24]:

| | Train Number | Train Name | Train Type | Run Days | Departure time | Arrival Time |
|---|---|---|---|---|---|---|
| 101 | 12267 | MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp | Duronto | [M, T, W, T, F, S, S] | 23:25 | 05:55 |
| 102 | 12268 | AHMEDABAD - MUMBAI CENT AC Duronto Exp | Duronto | [M, T, W, T, F, S, S] | 23:40 | 06:00 |
| 103 | 22201 | KOLKATA SEALDAH - PURI Duronto Express | Duronto | [M, W, F] | 20:00 | 04:00 |
| 104 | 22204 | SECUNDERABAD - VISAKHAPATNAM AC Duronto Express | Duronto | [M, W, S] | 20:15 | 06:35 |
| 105 | 22206 | MADURAI - CHENNAI CENTRAL AC Duronto Express | Duronto | [TUE, THU] | 22:40 | 07:20 |
| 106 | 12426 | JAMMU TAWI - NEW DELHI Rajdhani Express | Rajdhani | [M, T, W, T, F, S, S] | 19:40 | 05:05 |
| 107 | 12430 | NEW DELHI - LUCKNOW AC SF Express | Rajdhani | [M, T, F, S] | 20:50 | 06:40 |
| 108 | 12437 | SECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express | Rajdhani | [WED] | 12:45 | 10:25 |
| 109 | 12951 | MUMBAI CENTRAL - NEW DELHI Rajdhani Express | Rajdhani | [M, T, W, T, F, S, S] | 16:35 | 08:35 |
| 110 | 12953 | MUMBAI CENT - HAZRAT NIZAMUDDIN August Kranti ... | Rajdhani | [M, T, W, T, F, S, S] | 17:40 | 10:55 |
| 111 | 12019 | HOWRAH - RANCHI Shatabdi Express | Shatabdi | [M, T, W, T, F, S] | 06:05 | 13:15 |
| 112 | 12038 | LUDHIANA - NEW DELHI Shatabdi Express | Shatabdi | [T, W, T, F, S] | 16:40 | 22:10 |
| 113 | 12048 | FIROZPUR - NEW DELHI Shatabdi Express | Shatabdi | [M, S] | 04:00 | 21:10 |

## FOR CHECKING OUR CURRENT WORKING DIRECTORY WE USE FOLLOWING CODES

```
In [25]: import os
         os.getcwd()

Out[25]: 'C:\\Users\\admin\\Downloads'
```

## FOR SAVING OUR CURRENT FILE INTO CSV FILE

```
In [26]: Train_Schedule.to_csv('Vishal Railyatri.in website Web Scrapping ',index=True)
         # index must be true for getting same index in our csv file
```

## FOR READING OUR SAVED FILE

```
In [28]: pd.read_csv('Vishal Railyatri.in website Web Scrapping')
```
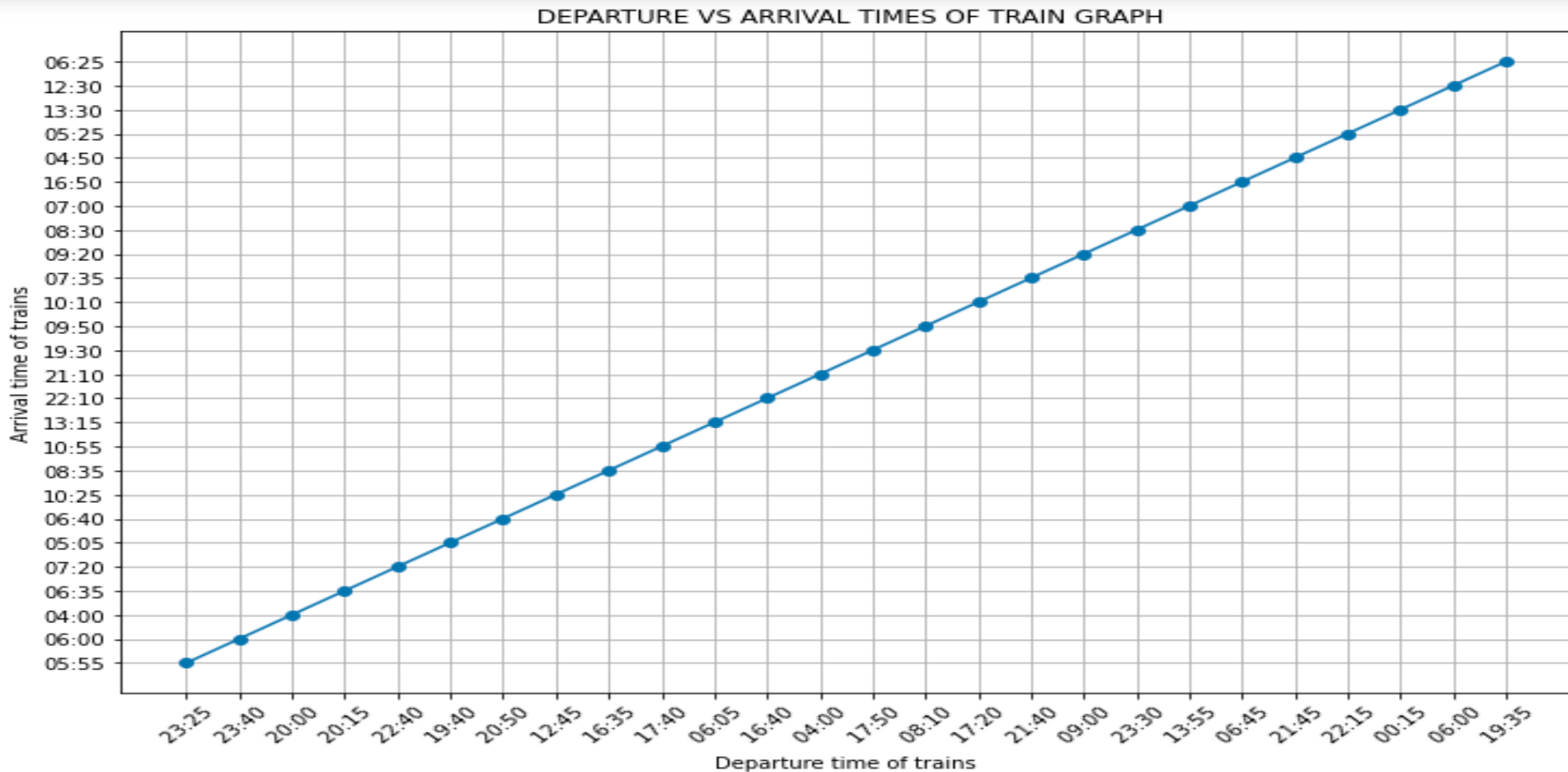
Out[28]:

| | Unnamed: 0 | Train Number | Train Name | Train Type | Run Days | Departure time | Arrival Time |
|---|---|---|---|---|---|---|---|
| 0 | 101 | 12267 | MUMBAI CENTRAL - AHMEDABAD AC Duronto Exp | Duronto | ['M', 'T', 'W', 'T', 'F', 'S', 'S'] | 23:25 | 05:55 |
| 1 | 102 | 12268 | AHMEDABAD - MUMBAI CENT AC Duronto Exp | Duronto | ['M', 'T', 'W', 'T', 'F', 'S', 'S'] | 23:40 | 06:00 |
| 2 | 103 | 22201 | KOLKATA SEALDAH - PURI Duronto Express | Duronto | ['M', 'W', 'F'] | 20:00 | 04:00 |
| 3 | 104 | 22204 | SECUNDERABAD - VISAKHAPATNAM AC Duronto Express | Duronto | ['M', 'W', 'S'] | 20:15 | 06:35 |
| 4 | 105 | 22206 | MADURAI - CHENNAI CENTRAL AC Duronto Express | Duronto | ['TUE', 'THU'] | 22:40 | 07:20 |
| 5 | 106 | 12426 | JAMMU TAWI - NEW DELHI Rajdhani Express | Rajdhani | ['M', 'T', 'W', 'T', 'F', 'S', 'S'] | 19:40 | 05:05 |
| 6 | 107 | 12430 | NEW DELHI - LUCKNOW AC SF Express | Rajdhani | ['M', 'T', 'F', 'S'] | 20:50 | 06:40 |
| 7 | 108 | 12437 | SECUNDERABAD - HAZRAT NIZAMUDDIN Rajdhani Express | Rajdhani | ['WED'] | 12:45 | 10:25 |

# FOR VISUALIZATION OF EXTRACTED DATA WE USE FOLLOWING LIBRARIES

```python
In [29]: import matplotlib.pyplot as plt
```

```python
In [42]: plt.figure(figsize=(11,7))
         plt.xlabel('Departure time of trains')
         plt.ylabel('Arrival time of trains')
         plt.title(' DEPARTURE VS ARRIVAL TIMES OF TRAIN GRAPH')
         plt.xticks(rotation=45)
         plt.plot(Departure_time, Arrival_Time, marker='o')
         plt.grid(True)
         plt.show()
```

# WE GOT THE GRAPH OF TIMES OF TRAIN DEPARTURE VS ARRIVAL



DEPARTURE VS ARRIVAL TIMES OF TRAIN GRAPH

# PROJECT OUTCOME

- Here my aim is, by using web scraping with python to find out the best arrival and departure timings of indian railways from railyatri.in website

- As for the graph there is continuous line in a graph so the conclusion by this was that we can departure in a train at any day of weeks as their was an train available from their sourse everyday

- Main conclusion , was that indian railways giving much facilities for their passengers by providing availability of train each day of weeks

THANK YOU