**#Anas Amir**

**#FA21-BSE-012**

**#Question 1**

import requests

from bs4 import BeautifulSoup

import pandas as pd

# Send a GET request to the URL

url = "https://www.imdb.com/chart/top/"

response = requests.get(url)

# Parse the HTML content

soup = BeautifulSoup(response.content, "html.parser")

# Find the list of top 250 movies

movies = soup.find\_all("td", class\_="titleColumn")

# Extract data for each movie

data = []

for movie in movies:

title = movie.find("a").text

year = movie.find("span", class\_="secondaryInfo").text.strip("()")

rating = movie.find\_next\_sibling("td", class\_="ratingColumn").find("strong").text

# Extracting duration

movie\_page\_url = "https://www.imdb.com" + movie.find("a")["href"]

movie\_page\_response = requests.get(movie\_page\_url)

movie\_page\_soup = BeautifulSoup(movie\_page\_response.content, "html.parser")

duration = movie\_page\_soup.find("time").text.strip()

data.append([title, year, duration, rating])

# Convert the data into a DataFrame

df = pd.DataFrame(data, columns=["Title", "Year", "Duration", "IMDb Rating"])

# Save the DataFrame to a CSV file

df.to\_csv("imdb\_top\_250\_movies.csv", index=False)

**#Anas Amir**

**#FA21-BSE-012**

**#Question 2**

import requests

from bs4 import BeautifulSoup

import pandas as pd

# Send a GET request to the URL

url = "https://space-facts.com/mars/"

response = requests.get(url)

# Parse the HTML content

soup = BeautifulSoup(response.content, "html.parser")

# Find the table containing Mars planet profile data

table = soup.find("table", {"id": "tablepress-p-mars-no-2"})

# Extract data from the table

data = []

for row in table.find\_all("tr"):

cols = row.find\_all("td")

if len(cols) == 2: # Ensure each row has exactly two columns

data.append([cols[0].text.strip(), cols[1].text.strip()])

# Convert the data into a DataFrame

df = pd.DataFrame(data, columns=["Parameter", "Value"])

# Save the DataFrame to an Excel file

df.to\_excel("mars\_planet\_profile.xlsx", index=False)