**CSE-3024 Web Mining**

**Lab Assignment 3**

**Alokam Nikhitha**

**19BCE2555**

Web Crawling

Question

**Experiment 4 (28.01.2022)**

1. Use BeautifulSoup or Scrapy to crawl any one of the E-commerce website of your choice and perform the same. The following information needs to be extracted from the page: (Choose any one product : e.g laptop, smartphone … etc)

a) Product Name

b) Product price

c) Product discount

d) Product image

**Problem statement:**

To Crawl any of the E-commerce website and extract the data from the page like Name, Price, Discount and image of the Product.

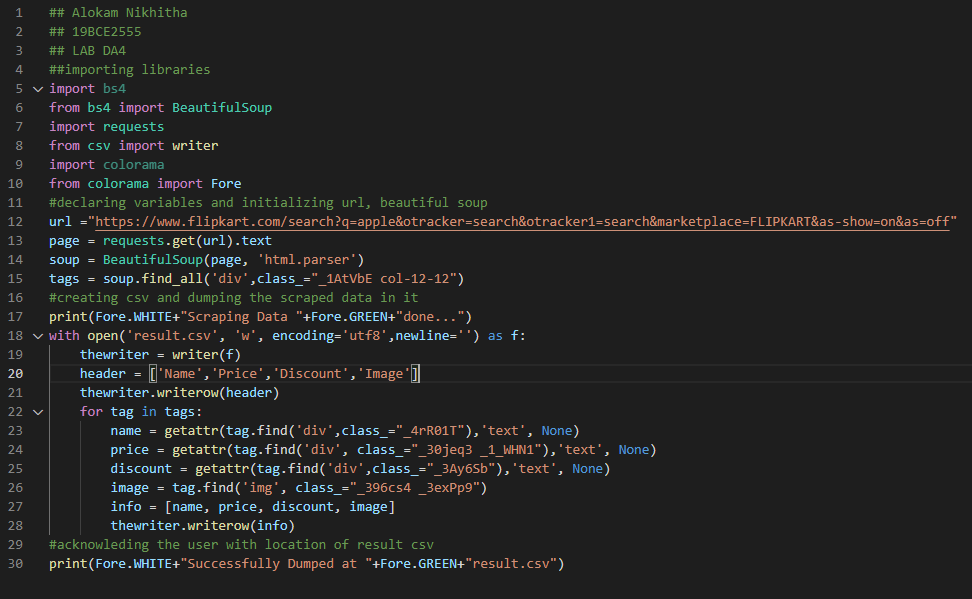
**Procedure:**

* We will Firstly import our libraries which are necessary in order to scrap the data from the website.
* Later we will declare the variables and also we will initialize the URL and also the beautifulsoup.
* Later we will create the file result.csv and dump the scrapped data into it. Here we will make a made the header with Name, Price, Discount and Image.
* Later we will collect the data of the product and add it to the CSV file.
* On running the python file the results.csv file will be created with scrapped data in it.
* In the result.csv file, the data of the product i.e, Name, Price, Discount and Image will be displayed .The image is returned in the form of a link.

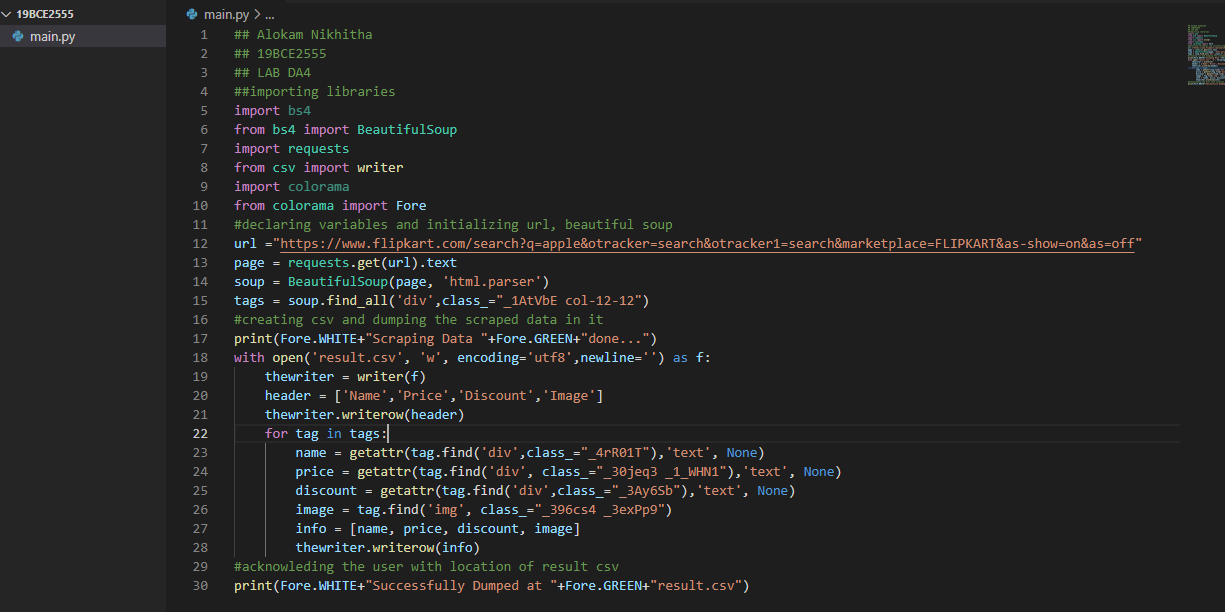
**URL of the website from which we are scrapping the data :**

"https://www.flipkart.com/search?q=apple&otracker=search&otracker1=search&marketplace=FLIPKART&as-show=on&as=off"

**Code:**

****

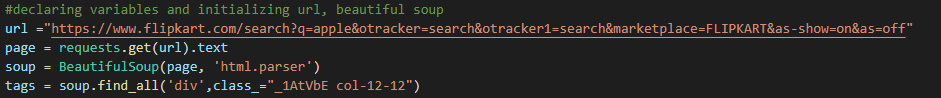
**Code Snippets and Outputs:**

****

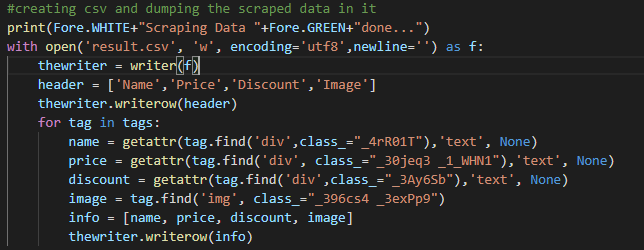
The python file is stored in the folder named ‘19BCE2555’.



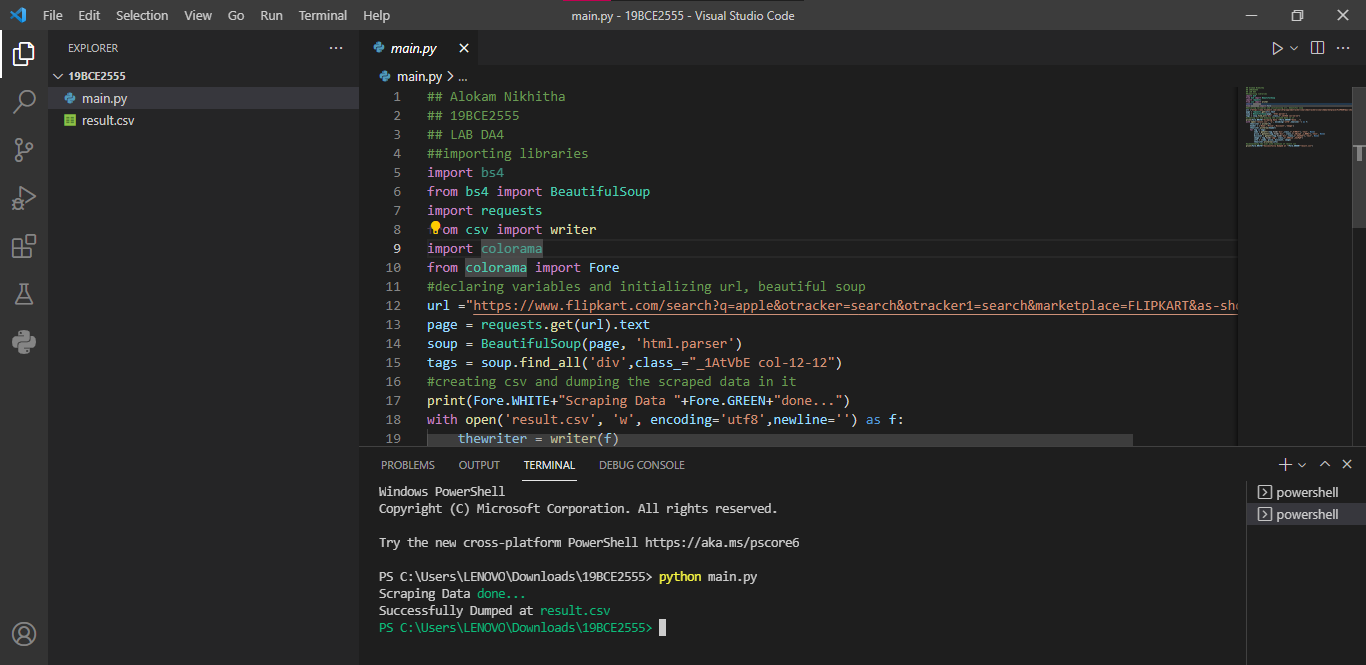
Here we are importing the necessary libraries inorder to Scrap the data.



Here we are declaring the variables and also we are Initializing the URL (here url is from ‘Flipkart’ site). And also initializing the beautifulsoup.

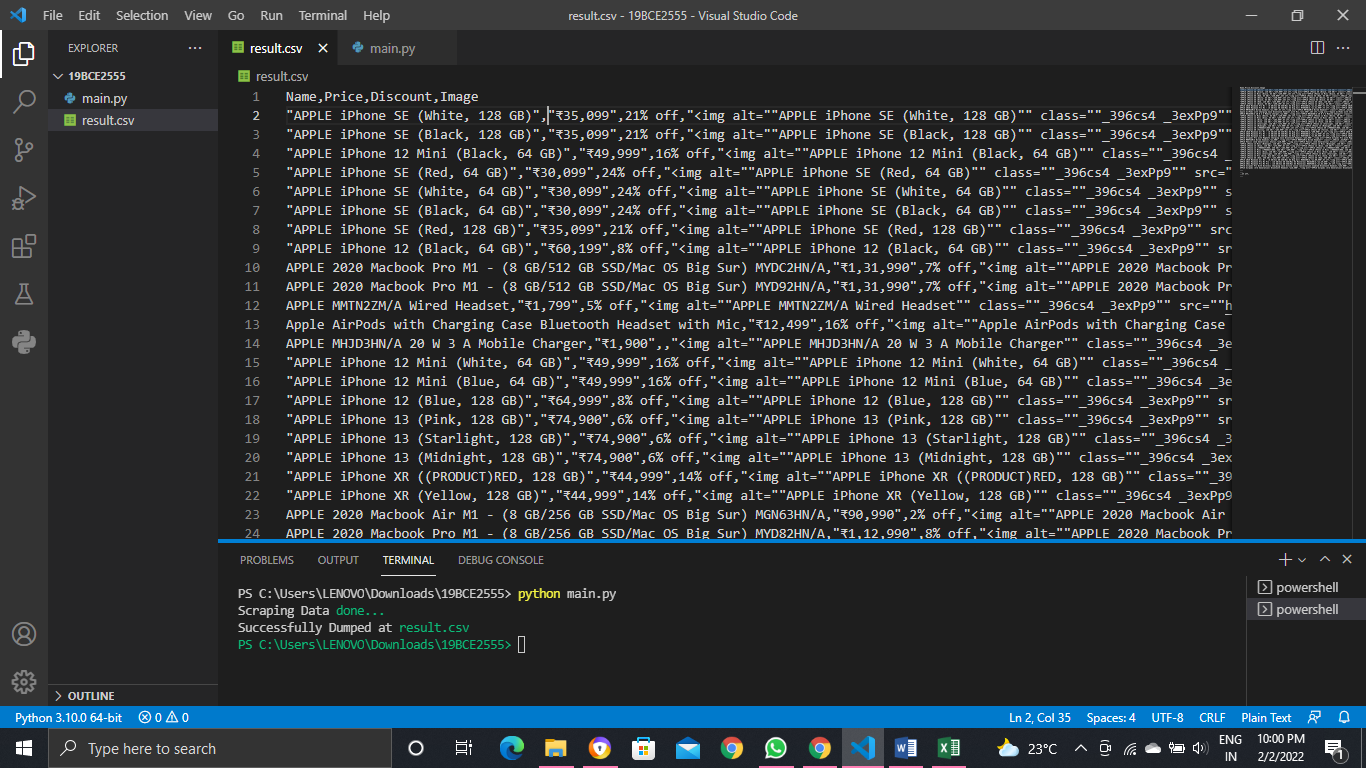


Here we are creating CSV file and dumping the scrapped data in it . Here we are opening the file result.csv and dumping the scrapped data into it. Here we made the header with Name, Price, Discount and Image. Later we are collecting the data of the product and adding it to the CSV file.

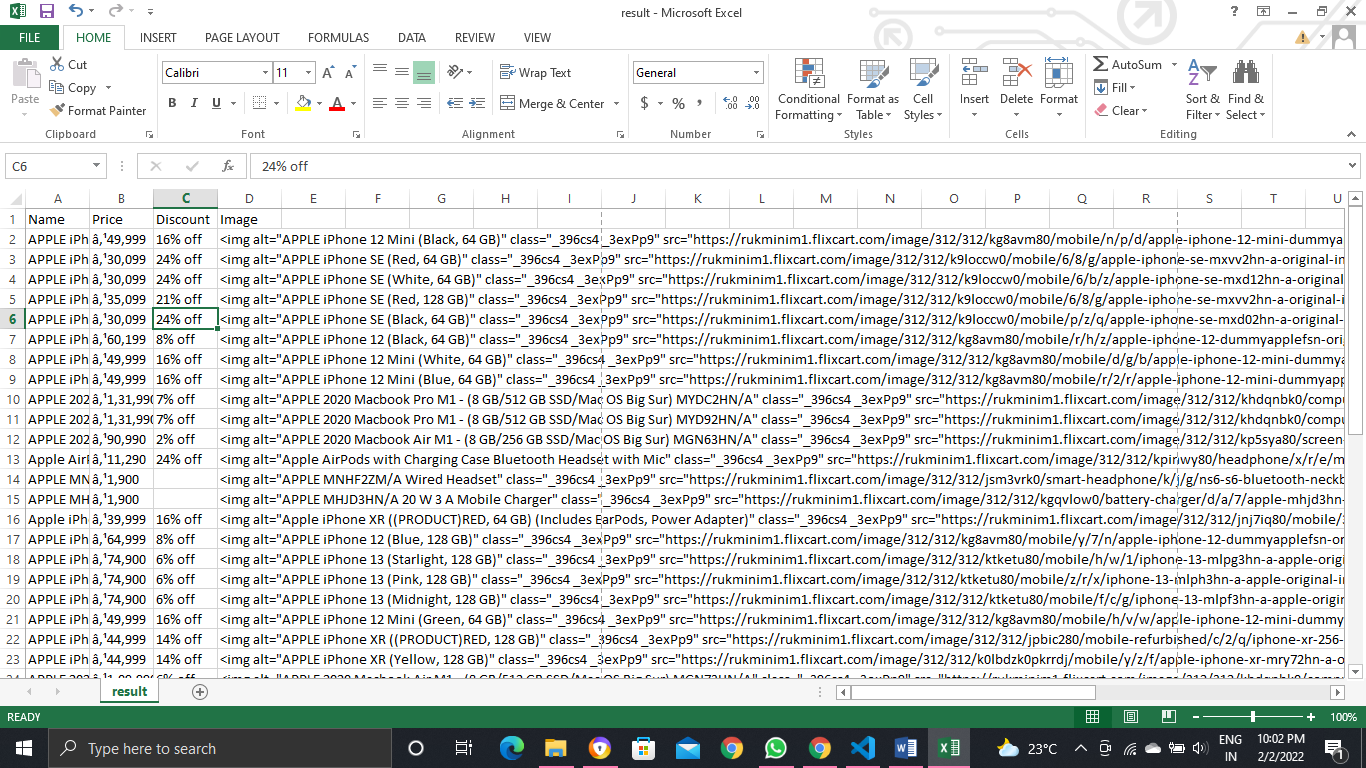


Here we are running the code using the command

‘python main.py’ in the terminal. Here the “result.csv” file got created in the folder after running the code.

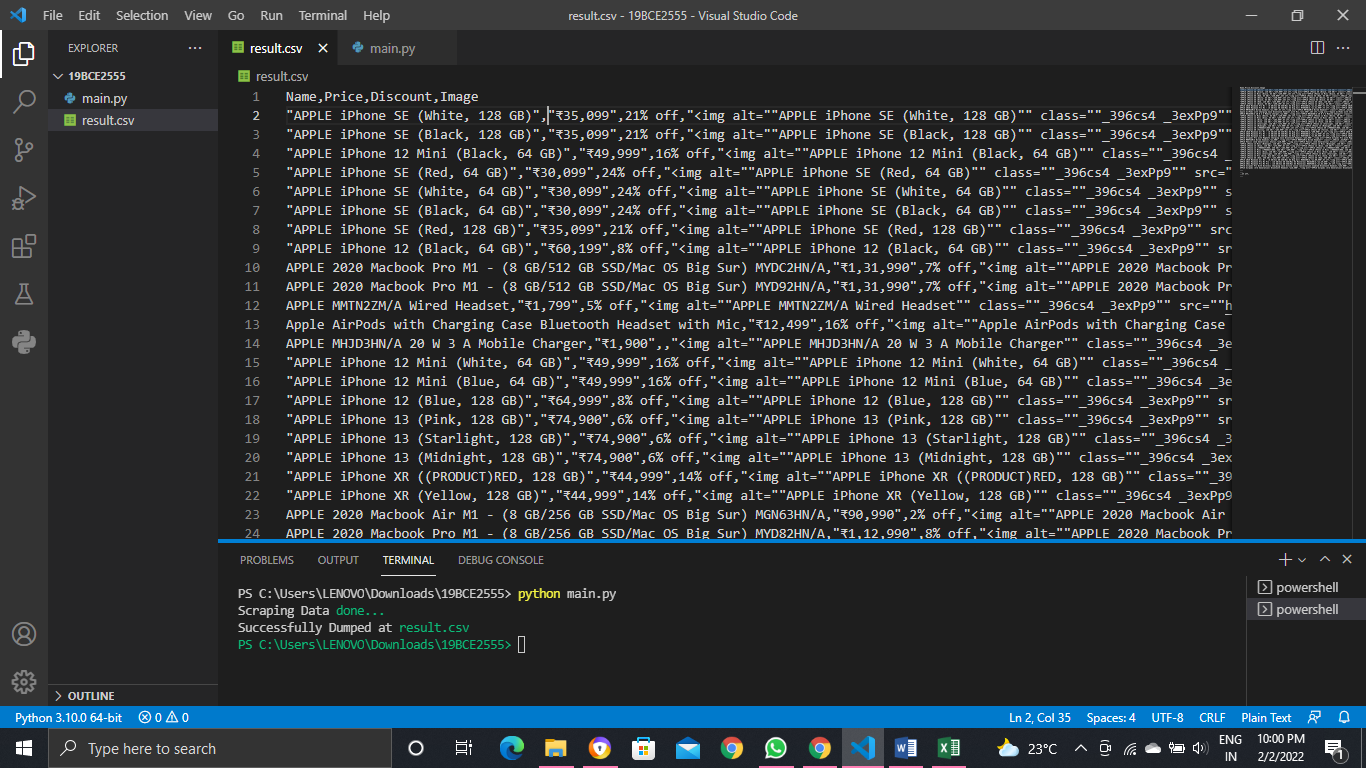


In the result.csv we can see the data that is scrapped from the Flipkart website.



Here is the Excel sheet in which we collected the scrapped data. The name of the product, price and Discount are collected. The image is returned in the form of a link.

**Results and Output**



Here is the list of data that is being scrapped from the website and dumped into result.csv file and which is formed in the folder of the code after running the code using beautifulsoup.



This is the view of Excel file in which the data is being stored.

Web Crawling Using Scrapy

Question

**Experiment 4 b**

1. Use BeautifulSoup or Scrapy to crawl any one of the E-commerce website of your choice and perform the same. The following information needs to be extracted from the page: (Choose any one product : e.g laptop, smartphone … etc)

a) Product Name

b) Product price

c) Product discount

d) Product image

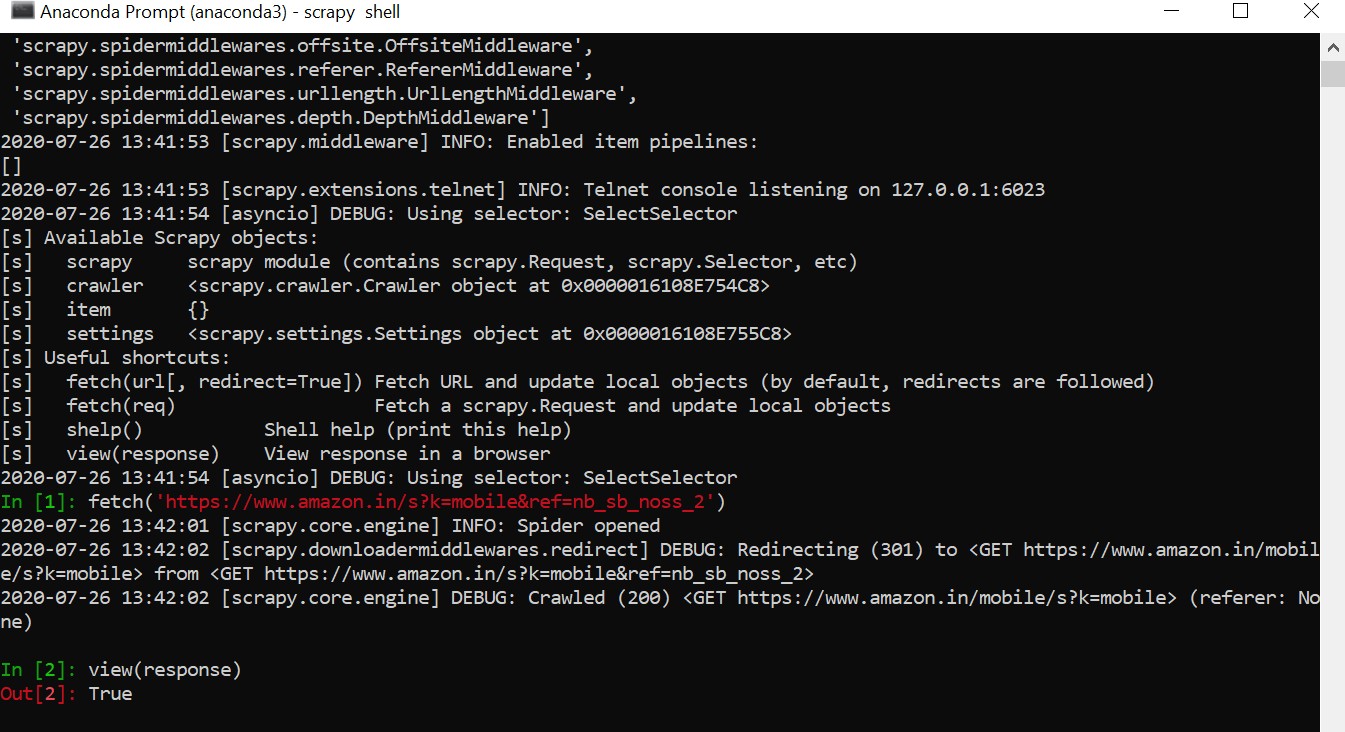
**Problem statement:**

To Crawl any of the E-commerce website and extract the data from the page like Name, Price, Discount and image of the Product using only Scrapy.

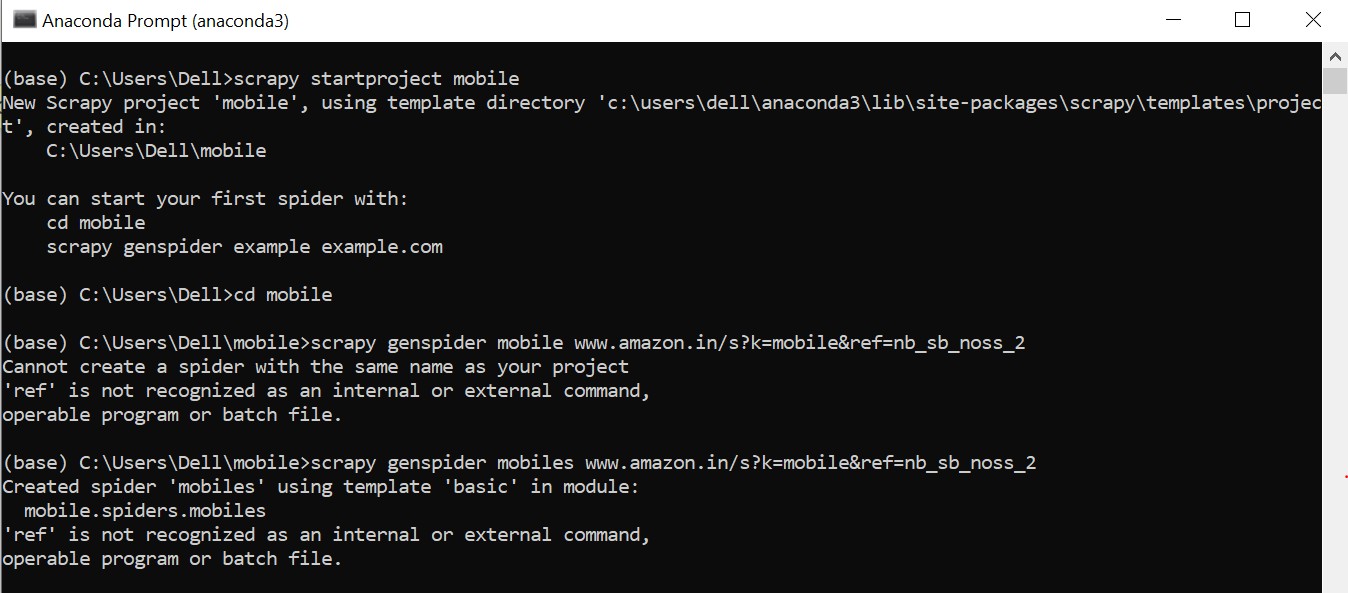
**Procedure:**

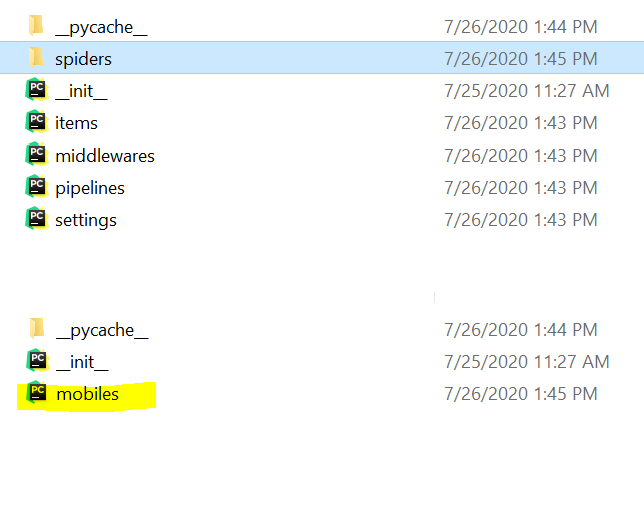
* + Firstly we install scrapy package with “pip install scrapy” in anaconda prompt
  + Later, we can start Shell by “scrapy shell”
  + Then Crawler run in the shell by use of the fetch and using view(response) to view fetched data.
  + An object should be created for the scrapper by “scrapy startproject mobile”
  + Create folder named “mobile”and move to that particular folder using command“cd mobile”
  + Create a python(.py) file inside the “spider” folder by using the command “scrapy genspider ..url..”
  + Here I scrapped data of amazons mobile as by product so the same url is pasted here.
  + Then python code is written in the file.
  + We can view the output in the terminal on typing “scrapy crawl ..name.. “ on teriminal
  + Finally it is exported as csv file using command “scrapy crawl mob -o data.csv”.

**Installing Scarpy in Anaconda .**



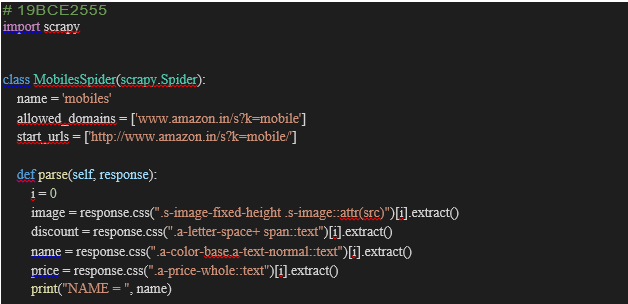
#creating scrapy project as name mobiles:

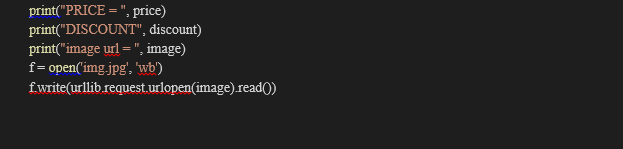




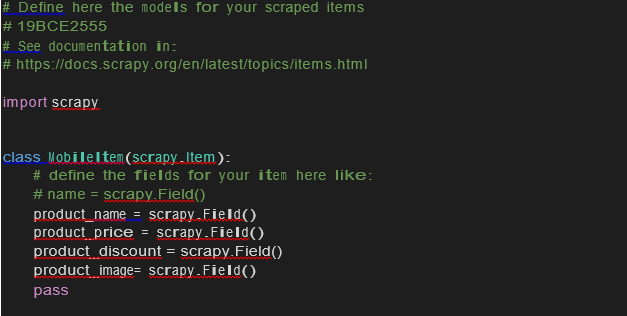
**Code:**

#mobiles.py:

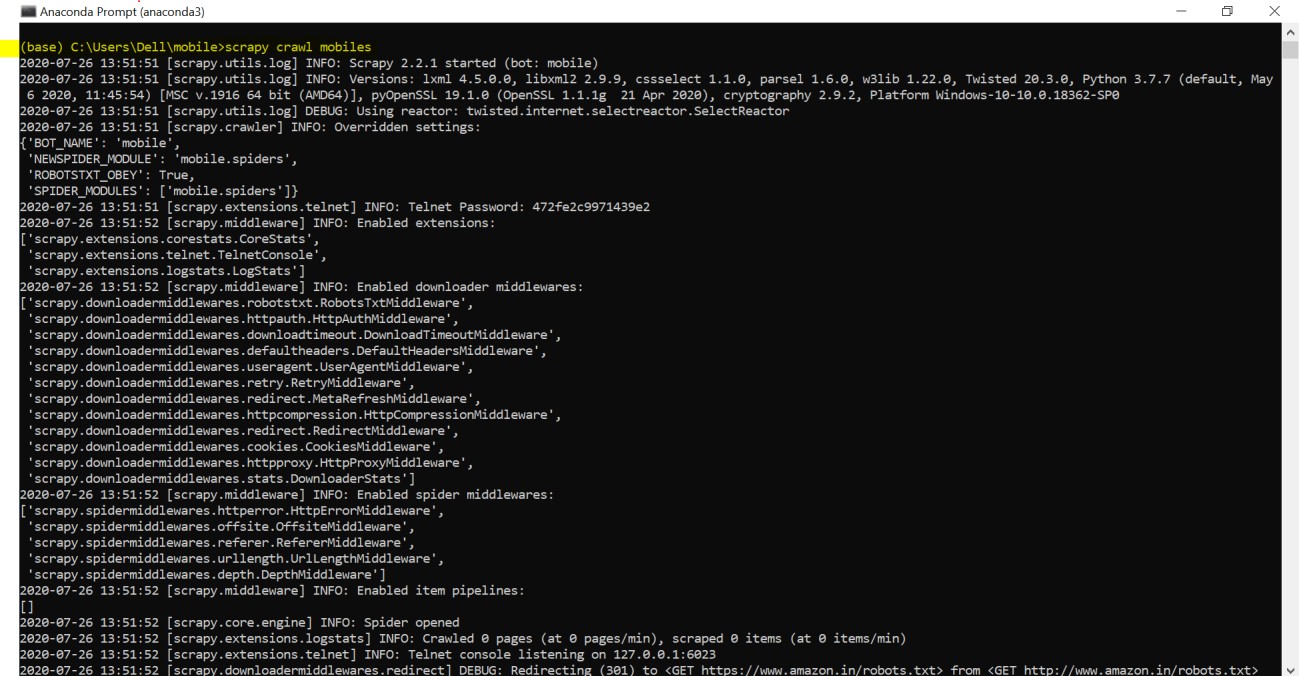


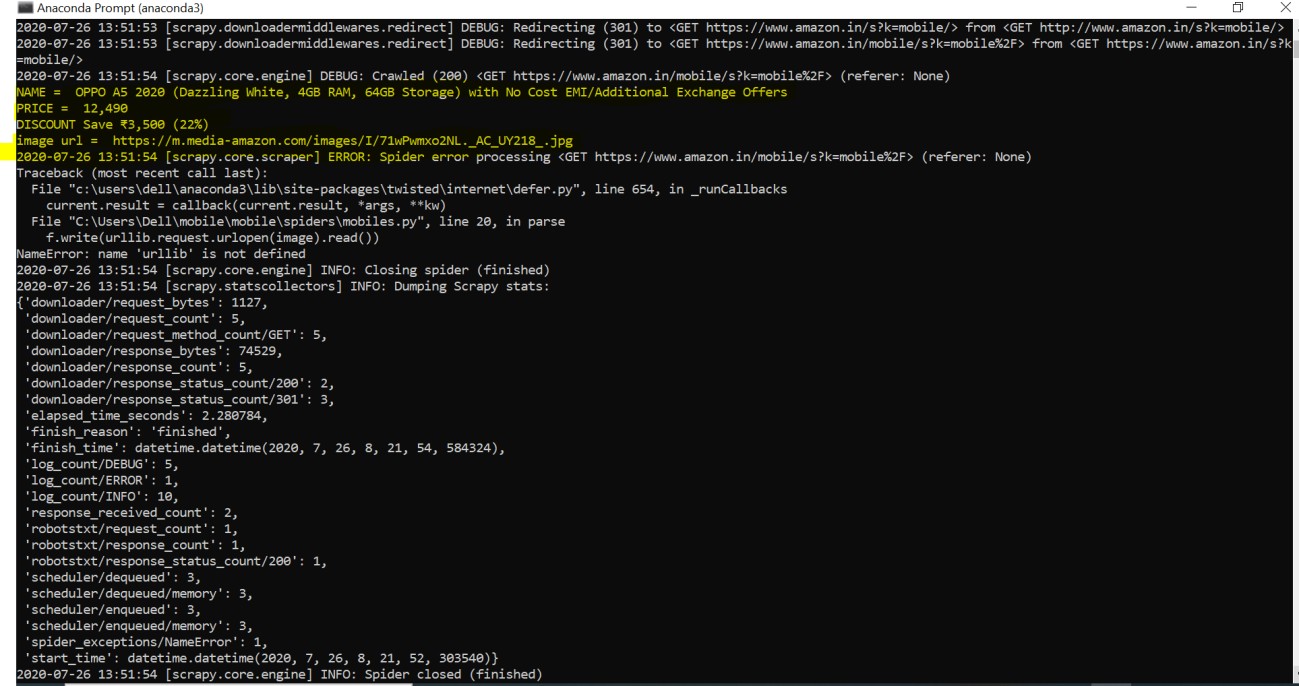


Items.py



**Output**



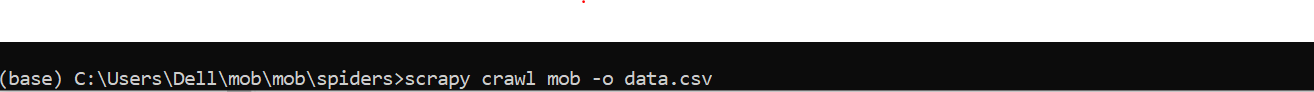
The basic information of the product is highlighted below 

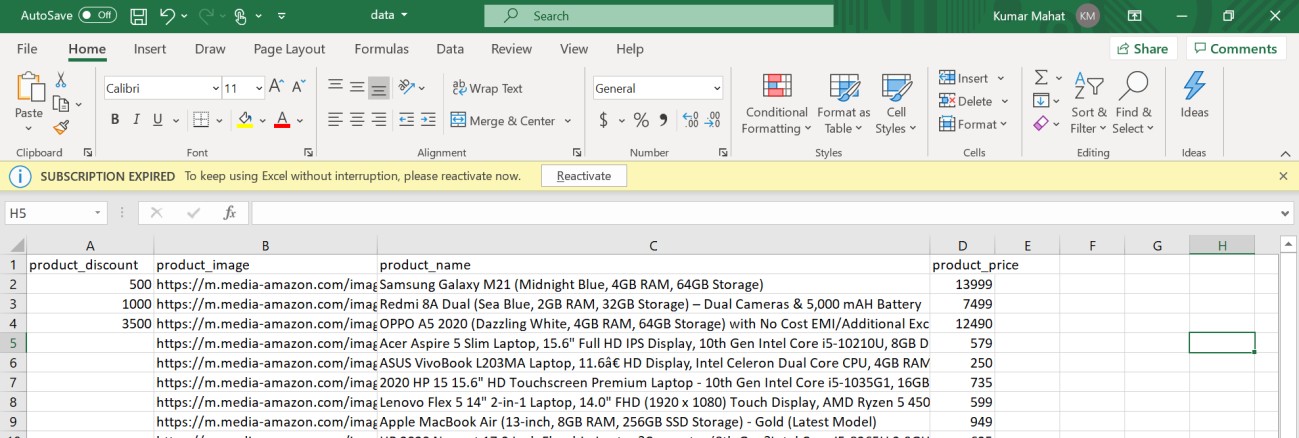
By clicking on the link extracted from the webpage we get the following image: Link: https://m.media-amazon.com/images/I/71wPwmxo2NL.\_AC\_UY218\_.jpg

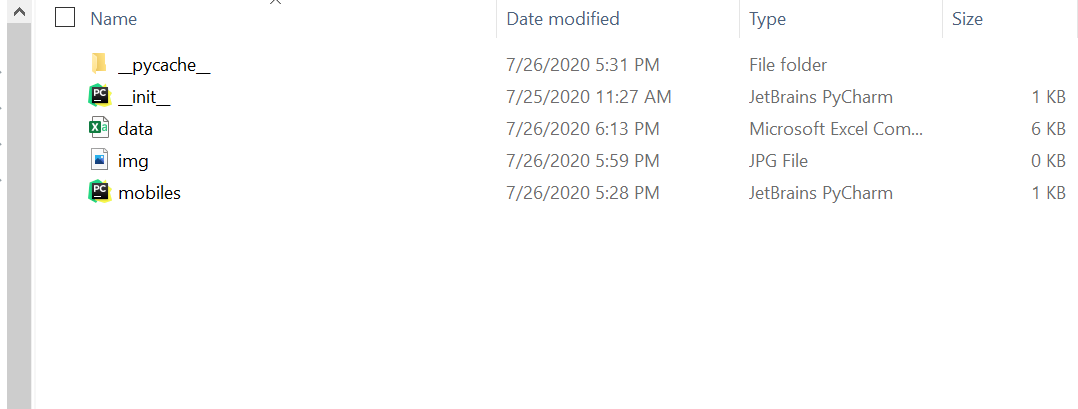


**Results and Output**

#Exporting scrapped data as csv



We can see that the data is scrapped and it is dumped in excel sheet



Encoding

Question

**Experiment 5**

Write a python program to perform the following encoding and decoding for the EVEN numbers between 1-20

1. Unary
2. Elias Gamma
3. Elias Delta
4. Golomb (b=10)

**Problem statement:**

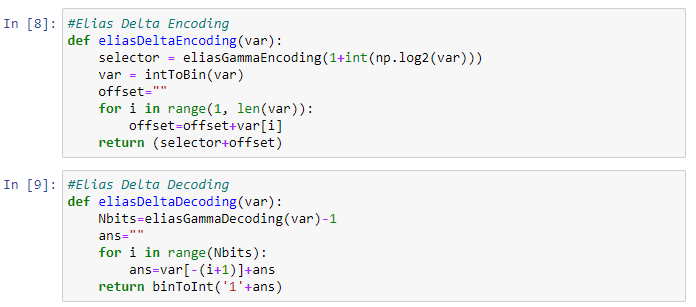
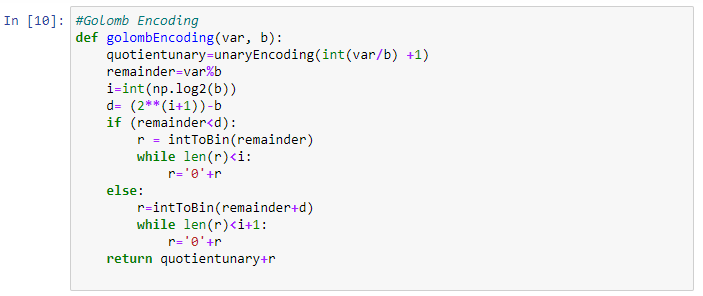
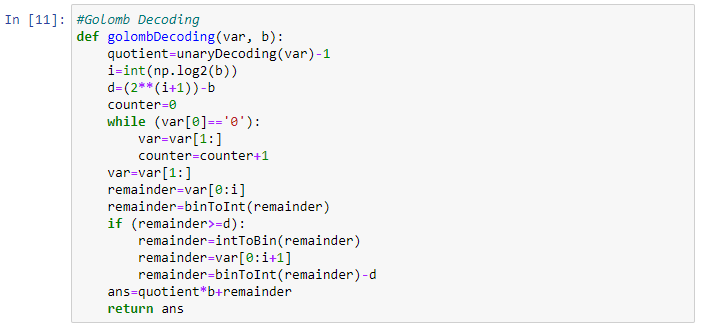
To perform the following encoding and decoding for the EVEN numbers between 1-20

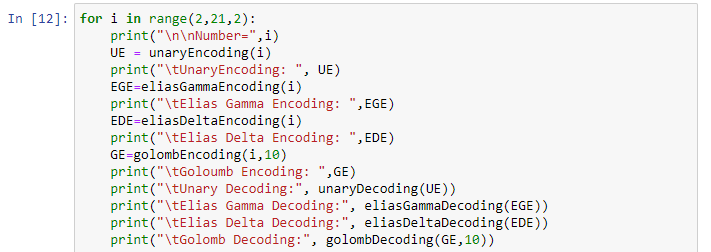
**Procedure:**

* Firstly, we will import the necessary numpy library to use mathematical functions like logarithm in our code.
* Next, We will create 2 functions, one to convert integer to binary and the other for converting binary to integer.
* Next, We will write respective functions for each and every method given.
* The functions that are corresponding to Unary Encoding, Unary Decoding, Elias Gamma Encoding, Elias Gamma Decoding, Elias Delta Encoding, Elias Delta Decoding, Golomb Encoding and Golomb Decoding.
* In main program, we will run a loop from numbers 2 to 21 with a jump of 2 to in order to get even numbers in the range from 1-20.
* We will finally perform the above functions to each of the iterators in the above loop

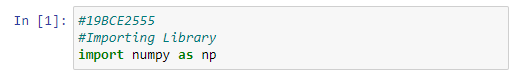
**Code:**

****

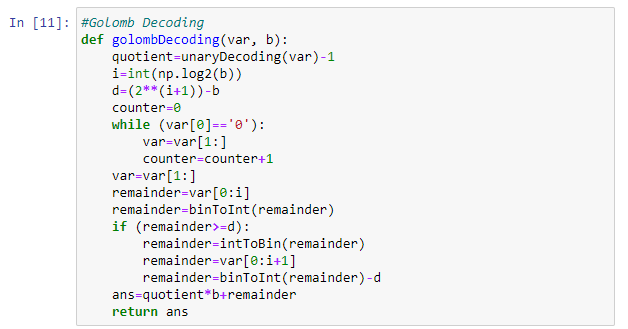
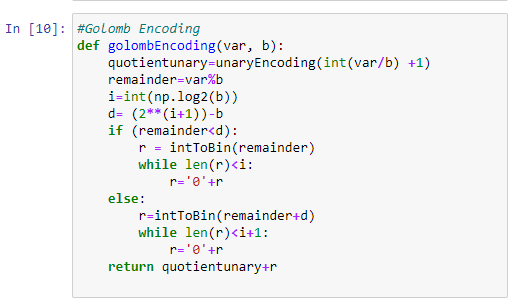
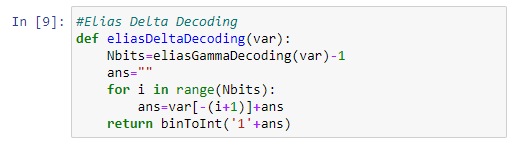
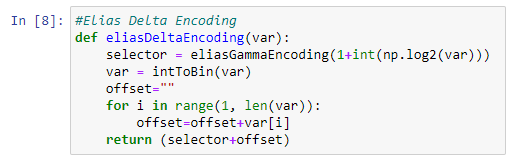
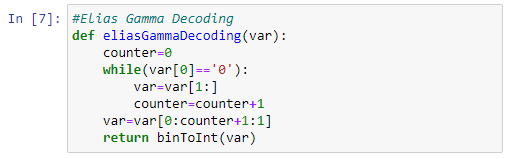
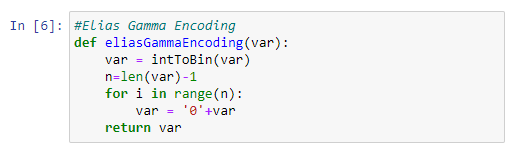
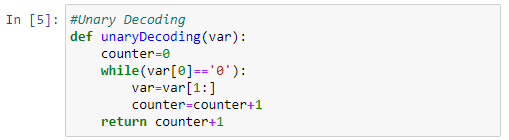
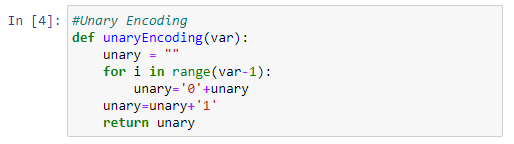
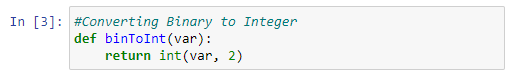
****   



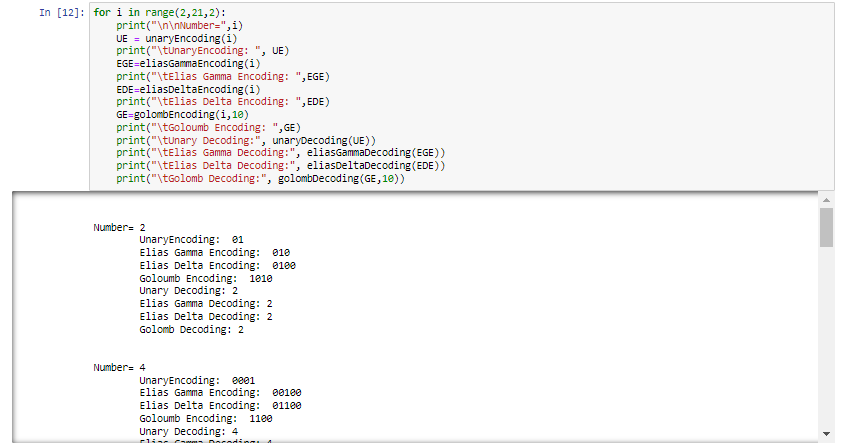
**Code Snippets and Outputs:**



Here we are importing the libraries that are required.

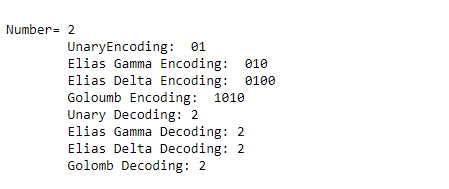


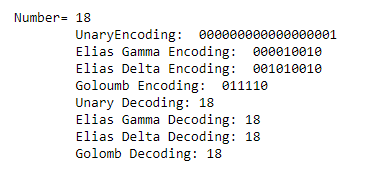
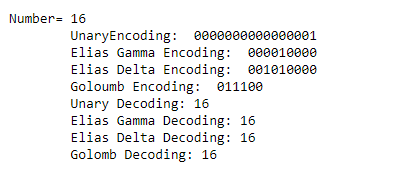
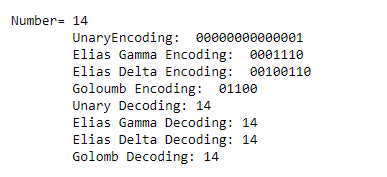
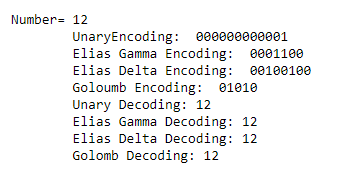
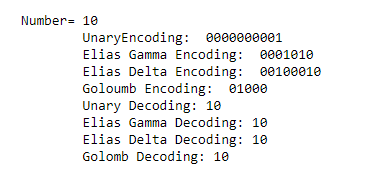
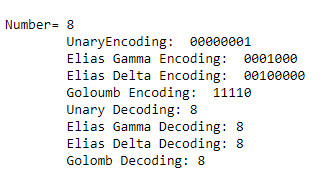
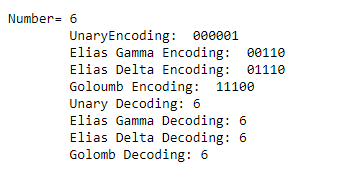
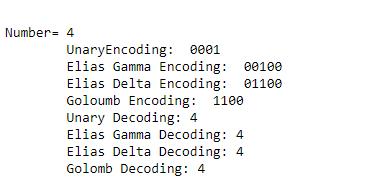
Here, we had defined all the ten functions that are described in procedure.

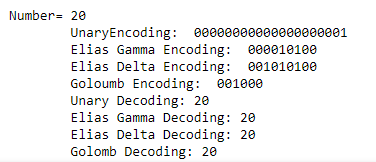


Here we are running a loop in order to iterate the even numbers in range 1-20 and then use the above functions to get our results.

**Results and Output**

****

****

****