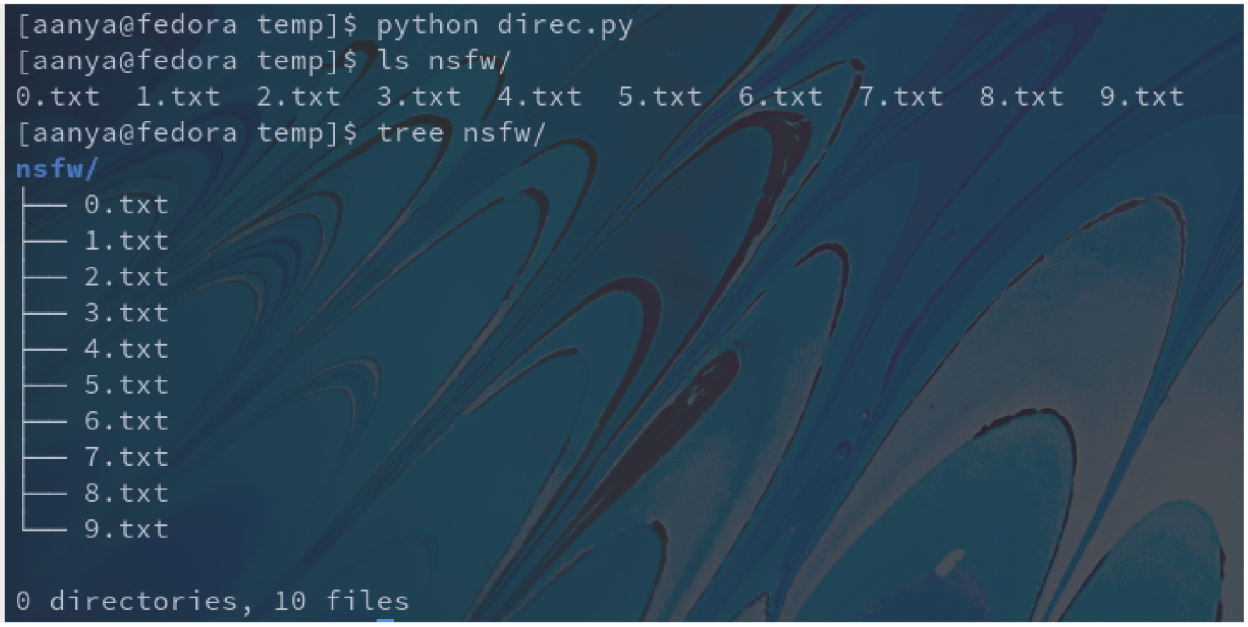
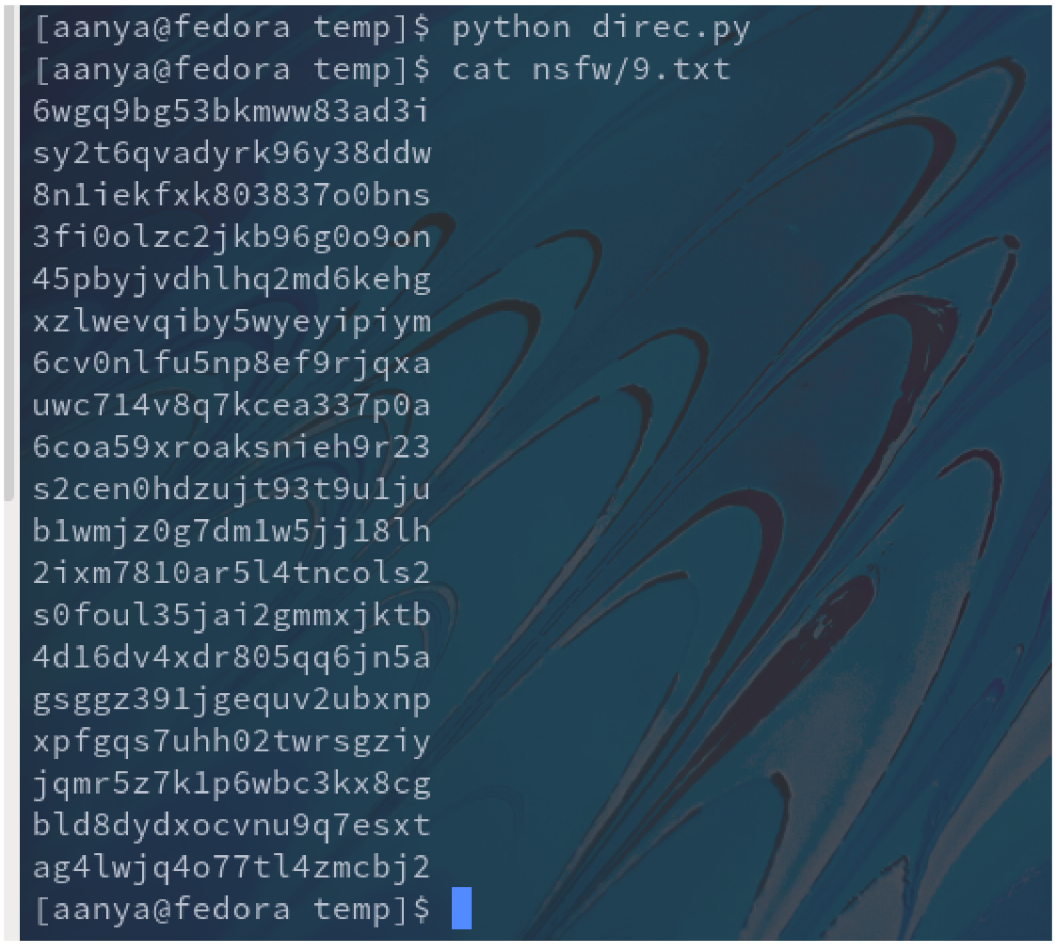
Name : Abhishek Guleri
Roll No. : 185509
Lab : Data Mining Lab
Assignment No. : 3
Branch - CSE DD

Program 1: Create 500 txt files in a directory. Every file contains 20,000 lines and every line contains random strings of length 20 characters.

```
import os
import string
import random
path = "nsfw"
if not os.path.exists(path):
    os.makedirs(path)
i = 0
while i < 500:
    filename = str(i) + '.txt'
    j = 1
    while j < 20000:
        with open(os.path.join(path, filename), 'a') as file:
            N = random.randrange(20, 20, 1)
            res = ''.join(random.choices(string.ascii_lowercase +
string.digits, k = N))
            file.write("%s\n" % str(res))
            file.close()
            j += 1
    i += 1
```

- Tested on small sample of constraints


[aanya@fedora temp]$ python direc.py
[aanya@fedora temp]$ ls nsfw/
0.txt 1.txt 2.txt 3.txt 4.txt 5.txt 6.txt 7.txt 8.txt 9.txt
[aanya@fedora temp]$ tree nsfw/
nsfw/
├── 0.txt
├── 1.txt
├── 2.txt
├── 3.txt
├── 4.txt
├── 5.txt
├── 6.txt
├── 7.txt
├── 8.txt
└── 9.txt
0 directories, 10 files


[aanya@fedora temp]$ python direc.py
[aanya@fedora temp]$ cat nsfw/9.txt
6wgq9bg53bkmww83ad3i
sy2t6qvadyrk96y38ddw
8n1iekfxk803837o0bns
3fi0olzc2jkb96g0o9on
45pbyjvdhlhq2md6kehg
xzlwevqiby5wyeyipiym
6cv0nlfu5np8ef9rjqxa
uwc714v8q7kcea337p0a
6coa59xroaksnieh9r23
s2cen0hdzujt93t9u1ju
b1wmjz0g7dm1w5jj18lh
2ixm7810ar5l4tncols2
s0foul35jai2gmmxjktb
4d16dv4xdr805qq6jn5a
gsggz391jgequv2ubxnp
xpfgqs7uhh02twrsgziy
jqmr5z7k1p6wbc3kx8cg
bld8dydxocvnu9q7esxt
ag4lwjq4o77tl4zmcbj2
[aanya@fedora temp]$

Program 2: Calculate the execution time to convert all the files to uppercase. Save the results in the csv file as given below.

No. of Files,Time Taken (sec)
100,50
200,70
300,85
400,90
500,110

```
[aanya@fedora temp]$ cat execTime.csv
No. of Files,Time Taken (sec)
100,0.1751561626112797
200,0.2851651623112783
300,0.3951751626112743
400,0.4151591658112723
500,0.5650551638112799
[aanya@fedora temp]$
```

```
# importing the csv module
import csv

# field names
fields = ['No. of Files', 'Time Taken (sec)']

# data rows of csv file
rows = [ ['100', '0.1751561626112797'],
      ['200', '0.2851651623112783'],
      ['300', '0.3951751626112743'],
      ['400', '0.4151591658112723'],
      ['500', '0.5650551638112799']]

# name of csv file
filename = "execTime.csv"

# writing to csv file
with open(filename, 'w') as csvfile:
      # creating a csv writer object
```

```
      csvwriter = csv.writer(csvfile)

      # writing the fields
      csvwriter.writerow(fields)

      # writing the data rows
      csvwriter.writerows(rows)
```

```
import csv
import os
import time

path = "nsfw"
i = 0

start = time.time()
while i < 100:
      filename = str(i) + '.txt'
      inputFile = open(os.path.join(path, filename), "r")
      content = inputFile.read()
      print(filename)
      with open(os.path.join(path,filename), "w") as outputFile:
            outputFile.write(content.upper())
      i += 1

end = time.time()
print(str(end - start))
```

```
73.txt
74.txt
75.txt
76.txt
77.txt
78.txt
79.txt
80.txt
81.txt
82.txt
83.txt
84.txt
85.txt
86.txt
87.txt
88.txt
89.txt
90.txt
91.txt
92.txt
93.txt
94.txt
95.txt
96.txt
97.txt
98.txt
99.txt
0.1751551628112793
[aanya@fedora temp]$
```