TASK:-

Your task is to develop a Python command line tool that takes the architecture (amd64, arm64, mips etc.) as an argument and downloads the compressed Contents file associated with it from a Debian mirror. The program should parse the file and output the statistics of the top 10 packages that have the most files associated with them.
An example output could be:

./package_statistics.py amd64

1. <package name 1> <number of files>
2. <package name 2> <number of files>
.....
10. <package name 10> <number of files>

Time Taken:-

2 ~ 2.5 hours (including building report)

CODE:-

```
#!/usr/bin/python3
import os
import gzip
from heapq import nlargest
import io
import sys


NUMBER_OF_RESULTS = 10

Package_Dict = {}

'''This program simply fetch file and downloads it to parse on later.
Parsing the results will provide us with the required/expected output'''

class DebianPacks():
    '''class'''
    def __init__(self):
        '''takes the argument passed from command line'''
        archi = sys.argv[1]
```

```
        # data to be saved in this fileformat
        file_name = "Contents-{arch}.gz".format(arch=archi)
        url = self.get_url(file_name)
        print("Downloading file: %s" % file_name)
        print("Downloading from: %s" % url)
        self.download_file(url)
        print("Parsing file: %s " % file_name)
        self.read_file(file_name)
    def download_file(self, file):
        '''using wget to fetch file'''
        wget_command = "wget {} --no-check-certificate".format(file)
        os.system(wget_command)
    def read_file(self, file):
        '''reading and parsing data'''
        gz = gzip.open(file, 'rb')
        f = io.BufferedReader(gz)
        for line in f:
            line = line.decode("utf-8")
            line = line.rstrip()
            file_name, space, package_name = line.rpartition(' ')
            package_name = package_name.split(',')
            for package in package_name:
                # get package name
                package = package.rpartition('/')[2]
                # Uniquessness in keys
                if package not in Package_Dict.keys():
                    Package_Dict[package] = []
                Package_Dict[package].append(file_name)
        gz.close()
        i = 0
        line = '-' * 100
        print(line)
        print('{:<20s}{:<30s}{:>20s}'.format("#", "Package", "Count"))
        print(line)
        #Heap queue is used for processing
        #heapq.nlargest provides results sorted order from largest to smallest
        for package in nlargest(NUMBER_OF_RESULTS,
                                Package_Dict, key=lambda e:
len(Package_Dict[e])):
            print(
                "{:<20s}{: <30} {: >20}"
                .format("{}.".format(i+1), package, len(Package_Dict[package]))
            )
            i += 1
    def get_url(self, filename):
```

```
        '''mirror'''
        url = "http://ftp.uk.debian.org/debian/dists/stable/main/"
        file_url = "{deb_mirror}{filename}".format(
            filename=filename,
            deb_mirror=url,
        )
        return file_url
if __name__ == "__main__":
    DebianPacks()
```

Command to run :-

```
./package_statistics.py arm464
```

Output:

```
root@master:# ./package_statistics.py arm64

Downloading file: Contents-arm64.gz

Downloading from: http://ftp.uk.debian.org/debian/dists/stable/main/Contents-
arm64.gz

--2022-07-22 06:49:09--
http://ftp.uk.debian.org/debian/dists/stable/main/Contents-arm64.gz

Resolving ftp.uk.debian.org (ftp.uk.debian.org)... 78.129.164.123,
2001:1b40:5600:ff80:f8ee::1

Connecting to ftp.uk.debian.org (ftp.uk.debian.org)|78.129.164.123|:80...
connected.

HTTP request sent, awaiting response... 200 OK

Length: 9782525 (9.3M) [application/octet-stream]

Saving to: ‘Contents-arm64.gz.6’
```

```
Contents-arm64.gz.6
100%[=============================================================================
=========>]   9.33M  2.23MB/s    in 4.9s



2022-07-22 06:49:14 (1.89 MB/s) - ‘Contents-arm64.gz.6’ saved [9782525/9782525]



Parsing file: Contents-arm64.gz

------------------------------------------------------------------------------
------------------

#                   Package                                     Count

------------------------------------------------------------------------------
------------------

1.                  piglit                                       51784

2.                  esys-particle                                18015

3.                  libboost1.74-dev                             14333

4.                  acl2-books                                   12668

5.                  golang-1.15-src                               9015

6.                  liboce-modeling-dev                           7457

7.                  zoneminder                                    7002

8.                  paraview-dev                                  6178

9.                  locales-all                                   5956

10.                 linux-headers-5.10.0-16-arm64                 5858
```

Tools used

1. Autopep8

```
autopep8 package_statistics.py --in-place
```

2. Pylint3

```
root@master:# pylint3 package_statistics.py
No config file found, using default configuration
************* Module package_statistics
C: 63, 0: Trailing whitespace (trailing-whitespace)
C:  1, 0: Missing module docstring (missing-docstring)
C: 11, 0: Constant name "Package_Dict" doesn't conform to UPPER_CASE naming style (invalid-name)
R: 29, 4: Method could be a function (no-self-use)
C: 35, 8: Variable name "gz" doesn't conform to snake_case naming style (invalid-name)
C: 36, 8: Variable name "f" doesn't conform to snake_case naming style (invalid-name)
W: 40,23: Unused variable 'space' (unused-variable)
R: 33, 4: Method could be a function (no-self-use)
R: 64, 4: Method could be a function (no-self-use)

-------------------------------------------------------------------
Your code has been rated at 8.16/10 (previous run: 8.16/10, +0.00)
```

3. cProfiler – important output is colored in Yellow color below

```
#python3 -m cProfile package_statistics.py amd64 > cProfiler.txt

Downloading file: Contents-amd64.gz

Downloading from: http://ftp.uk.debian.org/debian/dists/stable/main/Contents-amd64.gz

Parsing file: Contents-amd64.gz
```

```
-----------------------------------------------------------------------------------------------
#               Package                         Count
-----------------------------------------------------------------------------------------------
1.              piglit                          51784
2.              esys-particle                   18015
3.              libboost1.74-dev                14333
4.              acl2-books                      12668
5.              golang-1.15-src                 9015
6.              liboce-modeling-dev             7457
7.              zoneminder                      7002
8.              paraview-dev                    6178
9.              linux-headers-5.10.0-16-amd64   6150
10.             linux-headers-5.10.0-13-amd64   6149
       12035628 function calls (12019925 primitive calls) in 11.972 seconds


  Ordered by: standard name


  ncalls  tottime  percall  cumtime  percall filename:lineno(function)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:103(release)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:143(__init__)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:147(__enter__)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:151(__exit__)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:157(_get_module_lock)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:176(cb)
     4/1    0.000    0.000    0.001    0.001 <frozen importlib._bootstrap>:211(_call_with_frames_removed)
      35    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:222(_verbose_message)
       1    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:232(_requires_builtin_wrapper)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:307(__init__)
       3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:311(__enter__)
```

```
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:318(__exit__)
    12    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:321(<genexpr>)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:35(_new_module)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:369(__init__)
     4    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:403(cached)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:416(parent)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:424(has_location)
     1    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:433(spec_from_loader)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:504(_init_module_attrs)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:564(module_from_spec)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:58(__init__)
   3/1    0.000    0.000    0.001    0.001 <frozen importlib._bootstrap>:651(_load_unlocked)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:707(find_spec)
     1    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:728(create_module)
     1    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:736(exec_module)
     1    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:753(is_package)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:78(acquire)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:780(find_spec)
     7    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:843(__enter__)
     7    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:847(__exit__)
     3    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:870(_find_spec)
   3/1    0.000    0.000    0.001    0.001 <frozen importlib._bootstrap>:936(_find_and_load_unlocked)
   3/1    0.000    0.000    0.001    0.001 <frozen importlib._bootstrap>:966(_find_and_load)
     1    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap>:997(_handle_fromlist)
     8    0.000    0.000    0.000    0.000 <frozen
importlib._bootstrap_external>:1080(_path_importer_cache)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:1117(_get_spec)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:1149(find_spec)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:1228(_get_spec)
```

```
     6    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:1233(find_spec)
     4    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:263(cache_from_source)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:361(_get_cached)
     2    0.000    0.000    0.000    0.000 <frozen
importlib._bootstrap_external>:393(_check_name_wrapper)
     6    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:41(_relax_case)
     2    0.000    0.000    0.000    0.000 <frozen
importlib._bootstrap_external>:430(_validate_bytecode_header)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:485(_compile_bytecode)
     4    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:52(_r_long)
     2    0.000    0.000    0.000    0.000 <frozen
importlib._bootstrap_external>:524(spec_from_file_location)
    32    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:57(_path_join)
    32    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:59(<listcomp>)
     4    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:63(_path_split)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:669(create_module)
   2/1    0.000    0.000    0.001    0.001 <frozen importlib._bootstrap_external>:672(exec_module)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:743(get_code)
    10    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:75(_path_stat)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:800(__init__)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:825(get_filename)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:830(get_data)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:840(path_stats)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:85(_path_is_mode_type)
     2    0.000    0.000    0.000    0.000 <frozen importlib._bootstrap_external>:94(_path_isfile)
     1    0.000    0.000    0.000    0.000 _compression.py:1(<module>)
 15688    0.008    0.000    0.010    0.000 _compression.py:12(_check_not_closed)
     1    0.000    0.000    0.000    0.000 _compression.py:120(seek)
     1    0.000    0.000    0.000    0.000 _compression.py:150(tell)
     1    0.000    0.000    0.000    0.000 _compression.py:33(DecompressReader)
```

```
     1    0.000    0.000    0.000    0.000 _compression.py:36(readable)
     1    0.000    0.000    0.000    0.000 _compression.py:39(__init__)
     1    0.000    0.000    0.000    0.000 _compression.py:59(close)
     1    0.000    0.000    0.000    0.000 _compression.py:63(seekable)
 15688    0.080    0.000    0.915    0.000 _compression.py:66(readinto)
     1    0.000    0.000    0.000    0.000 _compression.py:9(BaseStream)
    12    0.000    0.000    0.000    0.000 _weakrefset.py:36(__init__)
     4    0.000    0.000    0.000    0.000 abc.py:132(__new__)
     4    0.000    0.000    0.000    0.000 abc.py:135(<setcomp>)
     1    0.000    0.000    0.000    0.000 gzip.py:107(seekable)
     1    0.000    0.000    0.000    0.000 gzip.py:110(GzipFile)
     1    0.000    0.000    0.000    0.000 gzip.py:123(__init__)
     1    0.000    0.000    0.000    0.000 gzip.py:20(open)
 15687    0.024    0.000    1.007    0.000 gzip.py:271(read)
1465559    0.194    0.000    0.194    0.000 gzip.py:298(closed)
     1    0.000    0.000    0.000    0.000 gzip.py:302(close)
     1    0.000    0.000    0.000    0.000 gzip.py:343(readable)
     1    0.000    0.000    0.000    0.000 gzip.py:352(seek)
     1    0.000    0.000    0.000    0.000 gzip.py:377(_GzipReader)
     1    0.000    0.000    0.000    0.000 gzip.py:378(__init__)
     3    0.000    0.000    0.000    0.000 gzip.py:385(_init_read)
     2    0.000    0.000    0.000    0.000 gzip.py:389(_read_exact)
     1    0.000    0.000    0.001    0.001 gzip.py:4(<module>)
     3    0.000    0.000    0.000    0.000 gzip.py:405(_read_gzip_header)
 15688    0.069    0.000    0.803    0.000 gzip.py:438(read)
 15686    0.022    0.000    0.178    0.000 gzip.py:489(_add_read_data)
     1    0.000    0.000    0.000    0.000 gzip.py:493(_read_eof)
     1    0.000    0.000    0.000    0.000 gzip.py:69(_PaddedFile)
     1    0.000    0.000    0.000    0.000 gzip.py:74(__init__)
```

```
15692    0.072    0.000    0.108    0.000 gzip.py:80(read)
15686    0.019    0.000    0.022    0.000 gzip.py:93(prepend)
    1    0.009    0.009    0.021    0.021 heapq.py:524(nlargest)
    1    0.000    0.000    0.000    0.000 heapq.py:569(<listcomp>)
    1    0.000    0.000    0.000    0.000 heapq.py:583(<listcomp>)
    1    0.000    0.000    0.000    0.000 package_statistics.py:16(DebianPacks)
    1    0.000    0.000   11.971   11.971 package_statistics.py:18(__init__)
    1    0.000    0.000   11.972   11.972 package_statistics.py:2(<module>)
    1    0.000    0.000    5.557    5.557 package_statistics.py:29(download_file)
    1    3.345    3.345    6.414    6.414 package_statistics.py:33(read_file)
29939    0.010    0.000    0.012    0.000 package_statistics.py:58(<lambda>)
    1    0.000    0.000    0.000    0.000 package_statistics.py:64(get_url)
    4    0.000    0.000    0.000    0.000 {built-in method __new__ of type object at 0x9d22c0}
    1    0.000    0.000    0.000    0.000 {built-in method _heapq.heapify}
   47    0.000    0.000    0.000    0.000 {built-in method _heapq.heapreplace}
    2    0.000    0.000    0.000    0.000 {built-in method _imp._fix_co_filename}
   13    0.000    0.000    0.000    0.000 {built-in method _imp.acquire_lock}
    1    0.000    0.000    0.000    0.000 {built-in method _imp.create_builtin}
    1    0.000    0.000    0.000    0.000 {built-in method _imp.exec_builtin}
    3    0.000    0.000    0.000    0.000 {built-in method _imp.is_builtin}
    2    0.000    0.000    0.000    0.000 {built-in method _imp.is_frozen}
   13    0.000    0.000    0.000    0.000 {built-in method _imp.release_lock}
    2    0.000    0.000    0.000    0.000 {built-in method _struct.unpack}
    6    0.000    0.000    0.000    0.000 {built-in method _thread.allocate_lock}
    6    0.000    0.000    0.000    0.000 {built-in method _thread.get_ident}
    6    0.000    0.000    0.000    0.000 {built-in method builtins.__build_class__}
    3    0.000    0.000    0.000    0.000 {built-in method builtins.any}
  3/1    0.000    0.000   11.972   11.972 {built-in method builtins.exec}
   75    0.000    0.000    0.000    0.000 {built-in method builtins.getattr}
```

```
    15    0.000    0.000    0.000    0.000 {built-in method builtins.hasattr}
    11    0.000    0.000    0.000    0.000 {built-in method builtins.isinstance}
     1    0.000    0.000    0.000    0.000 {built-in method builtins.iter}
124083    0.016    0.000    0.016    0.000 {built-in method builtins.len}
    16    0.000    0.000    0.000    0.000 {built-in method builtins.print}
     4    0.000    0.000    0.000    0.000 {built-in method from_bytes}
     1    0.000    0.000    0.000    0.000 {built-in method io.open}
     2    0.000    0.000    0.000    0.000 {built-in method marshal.loads}
     7    0.000    0.000    0.000    0.000 {built-in method posix.fspath}
     4    0.000    0.000    0.000    0.000 {built-in method posix.getcwd}
    10    0.000    0.000    0.000    0.000 {built-in method posix.stat}
     1    5.557    5.557    5.557    5.557 {built-in method posix.system}
 15689    0.155    0.000    0.155    0.000 {built-in method zlib.crc32}
     2    0.000    0.000    0.000    0.000 {built-in method zlib.decompressobj}
     1    0.000    0.000    0.000    0.000 {function DecompressReader.close at 0x7f4da8c89e18}
1475850    0.185    0.000    0.185    0.000 {method 'append' of 'list' objects}
 15688    0.024    0.000    0.024    0.000 {method 'cast' of 'memoryview' objects}
     2    0.000    0.000    0.000    0.000 {method 'close' of '_io.BufferedReader' objects}
1449867    0.355    0.000    0.355    0.000 {method 'decode' of 'bytes' objects}
 15686    0.425    0.000    0.425    0.000 {method 'decompress' of 'zlib.Decompress' objects}
     1    0.000    0.000    0.000    0.000 {method 'disable' of '_lsprof.Profiler' objects}
     2    0.000    0.000    0.000    0.000 {method 'endswith' of 'str' objects}
    24    0.000    0.000    0.000    0.000 {method 'format' of 'str' objects}
     6    0.000    0.000    0.000    0.000 {method 'get' of 'dict' objects}
     4    0.000    0.000    0.000    0.000 {method 'items' of 'dict' objects}
    36    0.000    0.000    0.000    0.000 {method 'join' of 'str' objects}
1475850    0.170    0.000    0.170    0.000 {method 'keys' of 'dict' objects}
31378/15687    0.094    0.000    0.973    0.000 {method 'read' of '_io.BufferedReader' objects}
     2    0.000    0.000    0.000    0.000 {method 'read' of '_io.FileIO' objects}
```

```
     1    0.000    0.000    0.000    0.000 {method 'replace' of 'str' objects}
2925737    0.555    0.000    0.555    0.000 {method 'rpartition' of 'str' objects}
1449935    0.228    0.000    0.228    0.000 {method 'rstrip' of 'str' objects}
     1    0.000    0.000    0.000    0.000 {method 'seek' of '_io.BufferedReader' objects}
     1    0.000    0.000    0.000    0.000 {method 'sort' of 'list' objects}
1449867    0.355    0.000    0.355    0.000 {method 'split' of 'str' objects}
     1    0.000    0.000    0.000    0.000 {method 'startswith' of 'str' objects}
```

4. Graphviz

```
pycallgraph graphviz -- ./package_statistics.py
```