

```
1  import re
2  import pandas as pd
3  from collections import Counter
4
5
6  def taskA(data):
7
8      #collect all rows with the given passage id.
9      res = []
10     count = 0
11     for row in data["passageID"]:
12         if row == 8004107:
13             res.append(count)
14             count += 1
15
16     #Print wanted information.
17     print("taskA:")
18     for i in range(len(res)):
19         current = data["passage"][res[i]]
20         print("passage length: " + str(len(current.split())))
21         print("character count: " + str(len(current)))
22     print("\n")
23
24 def taskB(data):
25     count = 0
26     for row in data["queryID"]:
27         if row == 494835:
28             res = count
29             break
30         count += 1
31     passage = data["passage"][count]
32     passageID = data["passageID"][count]
33
34     print("taskB:")
35     print("passage: " + passage)
36     print("passageID: " + str(passageID))
37
38
39 def taskC(data):
40     counter = Counter(list(data["passageID"]))
41     tupellist = []
42     for item in counter.items():
43         tupellist.append((item[0],item[1]))
44
45     tupellist.sort(key=lambda a: a[1])
46     maxTupel = tupellist[-1]
47
48
49     counts = []
50     count = 0
51     for row in data["passageID"]:
52         if row == tupellist[-1][0]:
53             counts.append(count)
54             count += 1
55
56     print("taskC:")
57     print("rowIDs of the passages: " + str(counts))
58     queries = []
59     for i in range(len(counts)):
```

```
60     querys.append(data["queryID"][counts[i]])
61     print("queryIDs of the passages: " + str(querys))
62
63
64 def main() -> None:
65     FILEPATH = "msmarco-passagetest2019-top1000.tsv"
66     passagetestCSV = pd.read_csv(FILEPATH, sep='\t', names = ["queryID", "passageID",
67 "query", "passage"])
68     collectionCSV = pd.read_csv(FILEPATH, sep='\t')
69
70     #a solution
71     taskA(passagetestCSV)
72     # print(taskA(collectionCSV))
73
74     #b solution
75     taskB(passagetestCSV)
76     # print(taskB(collectionCSV))
77
78     #c solution
79     taskC(passagetestCSV)
80     # taskC(collectionCSV)
81
82 main()
```