

```
1 ### md
2 # Basic
3
4 > Take file name as input.
5 >
6 > Combine text of all the files.
7 >
8 > Open the file, read its content, and store it in
  one variable.
9 >
10 > Find out the number of lines.
11 >
12 > Find out the number of words.
13 >
14 > What is the total number of consumers for Branch1?
15 >
16 > What are the beverages available on Branch10,
  Branch8, and Branch1?
17 >
18 > How many times was the Icy Cappuccino ordered in
  Branch 5?
19 >
20 > What are the total number of people who ordered the
  MILD COFFEE?
21 >
22
23 ###
24 # Take file name as input
25
26 import os
27
28 from sympy.integrals.intpoly import distance_to_side
29
30 file_1 = 'Bev_BranchA.txt'
31 file_2 = 'Bev_BranchB.txt'
32 file_3 = 'Bev_BranchC.txt'
33
34 # file_1= input("Enter the name of the first file: ")
35 # file_2= input("Enter the name of the first file: ")
36 # file_3= input("Enter the name of the first file: ")
37
```

```
38 files=[file_1, file_2, file_3]
39 files
40 ###
41 # Combine text of all the files.
42
43 file_4 = 'Bev_Branch_All.txt'
44
45 with open(file_4, 'w') as out_file:
46     for file in files:
47         if os.path.exists(file):
48             with open(file, 'r') as in_file:
49                 out_file.write(in_file.read())
50 ###
51 # Open the file, read its content, and store it in
one variable.
52
53 with open(file_4, 'r') as file:
54     file_content = file.read()
55
56 print(file_content)
57
58 ###
59 # Find out the number of lines.
60
61 files.append(file_4)
62 print(files)
63
64 for file in files:
65     with open(file, 'r') as simple_file:
66         lines = simple_file.readlines()
67         print(f"There are {len(lines)} in the file {
file}.".")
68
69 ###
70 # Find out the number of words.
71 import re
72
73 for file in files:
74     if os.path.exists(file):
75         with open(file, 'r') as simple_file:
76             global words
```

```

77         words = simple_file.read().strip()
78         words = re.split(r'[, \s+]', words)
79         #print(words)
80         print(f'Number of words in {file}: {len(
words)})')
81
82 ###
83 # What is the total number of consumers for Branch1?
84
85 count = 0
86 for i in words:
87     if i == "Branch1":
88         count+=1
89 print(f"The total number of consumers for Branch1
are {count}.")
90
91 ###
92 # What are the beverages available on Branch10,
Branch8, and Branch1?
93
94 beverages = []
95
96 if os.path.exists(file_4):
97     with open (file_4,"r") as simple_file:
98         for i in simple_file.readlines():
99             if "Branch10" in i or "Branch8" in i or
"Branch1" in i:
100                 beverages.append(i.split(",")[0])
101
102                 # sort and remove repeated beverages
103                 beverages = sorted(set(beverages))
104
105 print(f"The beverages available on Branch10, Branch8
, and Branch1 are:")
106 for i in range((len(beverages))):
107     print(f"{i+1}. {beverages[i]}")
108
109 ###
110 # How many times was the Icy Cappuccino ordered in
Branch 5?
111

```

```
112 count = 0
113
114 if os.path.exists(file_4):
115     with open(file_4, 'r') as simple_file:
116         for i in simple_file.readlines():
117             if "ICY_cappuccino" in i and "Branch5"
in i:
118                 count+=1
119
120 print(f"The Icy Cappuccino was ordered {count} times
in Branch 5.")
```

```
121 ###
122 # What are the total number of people who ordered
the MILD COFFEE?
123
124 count = 0
125 if os.path.exists(file_4):
126     with open(file_4, "r") as simple_file:
127         for i in simple_file.readlines():
128             if "Mild_Coffee" in i:
129                 count+=1
130
131 print(f"The total number of people who ordered the
MILD COFFEE are {count}")
```

```
132 ### md
133 # Advance Task
134
135 > Find out the top 10 most frequent words.
136 >
137 > Create a visualization of word frequencies.
138 >
139 > What is the most consumed beverage per branch?
140 >
141 > What is the most consumed beverage overall?
142 >
143 > How many times was ☐Special Lite☐ ordered overall?
144 >
145 > GUI to upload file.
146 >
147 > Find the most important word.
148
```

```
149 ###
150 # Find out the top 10 most frequent words.
151
152 from nltk.tokenize import word_tokenize
153 from collections import Counter
154
155 if os.path.exists(file_4):
156     with open(file_4, 'r') as simple_file:
157         t_words = simple_file.read().strip()
158         t_words = word_tokenize(t_words)
159         t_words = [word for word in t_words if word
            != ","]
160
161         # Count the frequency of each word
162         word_freq = Counter(t_words)
163         top_10 = word_freq.most_common(10)
164
165         print("Top 10 most frequent words:")
166         for word, count in top_10:
167             print(f"{word}: {count}")
168
169
170


---


171 ###
172 # Create a visualization of word frequencies.
173 import matplotlib.pyplot as plt
174 import numpy as np
175
176 if os.path.exists(file_4):
177     with open(file_4, 'r') as simple_file:
178         t_words = simple_file.read().strip()
179         t_words = word_tokenize(t_words)
180         t_words = [word for word in t_words if word
            != ","]
181
182         # Count the frequency of each word
183         word_freq = Counter(t_words)
184         top_10 = word_freq.most_common(10)
185
186         # Plot the top 10 words
187         plt.figure(figsize=(10,5))
```

```

188         plt.bar([word for word, count in top_10], [
            count for word, count in top_10], color='skyblue',)
189
190         plt.xticks(rotation=90, fontsize=12)
191
192         x = np.arange(len(top_10))
193         y = [count for word, count in top_10]
194         # Line plot for the top 10 words
195         plt.plot(x, y, color='red')
196
197
198         plt.xlabel("Words")
199         plt.ylabel("Frequency")
200         plt.title("Top 10 most frequent words")
201         plt.show()
202
203 ###
204 # What is the most consumed beverage per branch?
205
206 branch = []
207 beverage = []
208
209 Branch1 = []
210 Branch2 = []
211 Branch3 = []
212 Branch4 = []
213 Branch5 = []
214 Branch6 = []
215 Branch7 = []
216 Branch8 = []
217 Branch9 = []
218
219 # Check if the file exists
220 if os.path.exists(file_4):
221     with open(file_4, "r") as simple_file:
222         for words in simple_file.readlines():
223             words = words.strip() # Remove leading/
trailing whitespaces
224             beverage_name, branch_name = words.split
225             ("","")

```

```
226         branch.append(branch_name)
227         beverage.append(beverage_name)
228
229     # Distribute beverages into the respective
branches
230     for i in range(len(branch)):
231         if branch[i] == "Branch1":
232             Branch1.append(beverage[i])
233         elif branch[i] == "Branch2":
234             Branch2.append(beverage[i])
235         elif branch[i] == "Branch3":
236             Branch3.append(beverage[i])
237         elif branch[i] == "Branch4":
238             Branch4.append(beverage[i])
239         elif branch[i] == "Branch5":
240             Branch5.append(beverage[i])
241         elif branch[i] == "Branch6":
242             Branch6.append(beverage[i])
243         elif branch[i] == "Branch7":
244             Branch7.append(beverage[i])
245         elif branch[i] == "Branch8":
246             Branch8.append(beverage[i])
247         elif branch[i] == "Branch9":
248             Branch9.append(beverage[i])
249
250     # Print frequency counts for each branch to
debug
251     # print("Branch1 frequencies:", Counter(Branch1
))
252     # print("Branch2 frequencies:", Counter(Branch2
))
253     # print("Branch3 frequencies:", Counter(Branch3
))
254     # print("Branch4 frequencies:", Counter(Branch4
))
255     # print("Branch5 frequencies:", Counter(Branch5
))
256     # print("Branch6 frequencies:", Counter(Branch6
))
257     # print("Branch7 frequencies:", Counter(Branch7
))
```

```

258     # print("Branch8 frequencies:", Counter(Branch8
    ))
259     # print("Branch9 frequencies:", Counter(Branch9
    ))
260
261     # Function to get the most common beverage
262     def most_common_beverage(branch_list):
263         if branch_list:
264             return Counter(branch_list).most_common(
1)[0][0]
265         return None
266
267     # Print the most common beverage for each branch
268     print(f"The most common beverage in Branch1 is {
most_common_beverage(Branch1)}")
269     print(f"The most common beverage in Branch2 is {
most_common_beverage(Branch2)}")
270     print(f"The most common beverage in Branch3 is {
most_common_beverage(Branch3)}")
271     print(f"The most common beverage in Branch4 is {
most_common_beverage(Branch4)}")
272     print(f"The most common beverage in Branch5 is {
most_common_beverage(Branch5)}")
273     print(f"The most common beverage in Branch6 is {
most_common_beverage(Branch6)}")
274     print(f"The most common beverage in Branch7 is {
most_common_beverage(Branch7)}")
275     print(f"The most common beverage in Branch8 is {
most_common_beverage(Branch8)}")
276     print(f"The most common beverage in Branch9 is {
most_common_beverage(Branch9)}")
277
278     ###
279     #What is the most consumed beverage overall?
280
281     from collections import Counter
282     import os
283
284     if os.path.exists(file_4):
285         with open (file_4,"r") as simple_file:
286             beverages = simple_file.read().strip()

```



```

287         beverages = word_tokenize(beverages)
288         beverages = [word for word in beverages if
word != ","]
289         beverages = [i for i in beverages if "Branch
" not in i]
290
291         # Count the frequency of each beverage
292         beverage_freq = Counter(beverages)
293         most_common_beverage = beverage_freq.
most_common(1)[0][0]
294         print(f"The most common beverage overall is
{most_common_beverage}")


---


295 ###
296 # How many times was Special Lite ordered overall?
297
298 count = 0
299 if os.path.exists(file_4):
300     with open(file_4,"r") as simple_file:
301         beverages = simple_file.read().strip()
302         beverages = word_tokenize(beverages)
303         beverages = [word for word in beverages if
word != ","]
304
305         for i in range(len(beverages)-1):
306             if beverages[i] == "Special_Lite":
307                 count+=1
308         print(f"The Special_Lite was ordered {count}
times overall.")


---


309 ###
310 #GUI to upload file.
311
312 import tkinter as tk
313 from tkinter import filedialog
314
315 def upload_file():
316     file_path = filedialog.askopenfilename()
317     print(f"File uploaded: {file_path}")
318
319 root = tk.Tk()
320 root.title("File Upload")
321 root.geometry("400x200")

```

```

322
323 upload_button = tk.Button(root, text="Upload File",
    command=upload_file)
324 upload_button.pack()
325
326 root.mainloop()
327 ###
328 # Find the most important word.
329
330 from nltk.corpus import stopwords
331 from nltk.tokenize import word_tokenize
332 from collections import Counter
333 import os
334
335 # Initialize set of stopwords
336 stop_words = set(stopwords.words("english"))
337
338 # Initialize dictionary to store words
339 word_freq = {}
340
341 # Check if the file exists
342 file_4 = 'Bev_Branch_All.txt' # Replace with the
    actual file
343 if os.path.exists(file_4):
344     with open(file_4, "r") as simple_file:
345         text = simple_file.read().strip()
346         words = word_tokenize(text)
347         words = [word for word in words if word
    != ","]
348
349
350     # Count the frequency of each word
351     word_freq = Counter(words)
352
353     # Find the most important word
354
355     most_important_word = word_freq.most_common(
1)[0][0]
356     print(f"The most important word is \"{
most_important_word}\".")

```