

FINAL ASSIGNMENT REPORT

Τα βήματα που ακολούθησα για την δημιουργία του προγράμματος έχουν ως εξής:

- Για αρχή έκανα import τα packages που χρειάστηκα κατά την διάρκεια του κώδικα.
- Στη συνέχεια χρησιμοποίησα την εντολή `read_csv()` του pandas για να διαβάσω το αρχείο που περιλαμβάνει τα δεδομένα από το 2016 ως το 2019.
- Ταξινόμησα τα δεδομένα με φθίνουσα σειρά του zip code (ίσως να ήταν και περιττό).
- Έκανα μια ταξινόμηση τα δεδομένα κατά zip και είδος προϊόντος αθροίζοντας παράλληλα και τα μπουκάλια που πουλήθηκαν κάθε φορά.
- Ακολούθησε μια επιπλέον ταξινόμηση ανα κατάσταση με άθροισμα πάλι τον αριθμό των μπουκαλιών που πουλήθηκαν στο καθένα.
- Δημιούργησα μια λίστα με στοιχεία τα ονόματα των καταστημάτων.
- Στη πρώτη «for», αθροίζω όλα τα μπουκάλια που πουλήθηκαν από όλα τα καταστήματα και στη συνέχεια στη δεύτερη διαιρώ τις πωλήσεις κάθε καταστήματος με το προηγούμενο άθροισμα και το πολλαπλασιάζω με το 100, προκειμένου να βρώ το ποσοστό των πωλήσεων που αντιστοιχεί στο κάθε κατάστημα. Το ποσοστό αυτό εισάγεται κάθε φορά στη λίστα «per».
- Δημιουργώ ένα data frame με την τελευταία λίστα με indexes τα στοιχεία του store.
- Από το `csv3` παίρνω μόνο τους δείκτες του zip code σε μια λίστα.
- Χρησιμοποιώ δυο εντολές `random` για να δημιουργήσω 2 λίστες με τυχαία χρώματα, τα οποία θα χρειαστώ για την επεξεργασία των διαγραμμάτων.
- Κάνω print τα data frames με τα ποσοστά και τα 10 πρώτα στοιχεία από την `csv3`.
- Δημιουργώ δυο subplots. Το πρώτο είναι ένα διάγραμμα bar σε οριζόντια στοίχιση, το οποίο περιλαμβάνει τα τα ποσοστά των πωλήσεων ανα κατάστημα. Στο δεύτερο διάγραμμα παριστάνεται σε scatter μορφή το πλήθος των μπουκαλιών που πουλήθηκαν ανα zip.

Δυσκολεύτηκα να καταλάβω τον τρόπο που δουλεύουν τα data frames με αποτέλεσμα να μη μπορώ να τα εισάγω στην αρχή μέσα στα plots. Εν τέλει βρήκα έναν τρόπο, κάνοντας τα δεδομένα, στήλες.

Παρακάτω φαίνονται σε screenshots, ο κώδικας, τα αποτελέσματα του καθώς και τα αντίστοιχα διαγράμματα.

```

1 import pandas as pd
2 import matplotlib.pyplot as mp
3 import random
4
5 csv = pd.read_csv('2016-2019.csv')
6
7 csv2 = csv.sort_values(by='zip_code', ascending=False)
8
9 csv3 = csv2.groupby(['zip_code', 'item_description'])['bottles_sold'].sum()
10 bs = csv2.groupby('store_name')['bottles_sold'].sum()
11
12 store = bs.index
13
14 sam = 0
15
16 per = []
17
18 for i in range(len(bs)):
19     sam += bs[i]
20
21 for j in range(len(bs)):
22     per.append((bs[j]/sam)*100)
23
24 perdf = pd.DataFrame(per, index=store, columns=['percentage (%)'])
25
26 zip = csv3.index.get_level_values('zip_code')
27
28 no_of_colors = len(zip)
29 colors1 = ['#'+''.join([random.choice('0123456789ABCDEF') for i in range(6)])
30             for j in range(no_of_colors)]
31
32 no_of_colors = len(per)
33 colors2 = ['#'+''.join([random.choice('0123456789ABCDEF') for i in range(6)])
34             for j in range(no_of_colors)]
35
36 print(perdf)
37 print(" ")
38 print(csv3.head(10))
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```

/Users/antonissapounas/PycharmProjects/final_assignment/venv/bin/python /Users/antonissapounas/PycharmProjects/final_assignment/final_assignment.py
percentage (%)
store_name
Bootleggin' Barzini's Fin      0.017746
Burlington Shell              0.425989
CVS Pharmacy #8526 / Cedar Rapids  0.124224
Cedar Ridge Vineyards         1.597161
Central City 2                 5.323869
...
Speedy Gas N Shop             0.124224
Tequila's Liquor Store        0.851819
The Ox & Wren Spirits and Gifts  0.035492
Tobacco Hut #14 / Council Bluffs  0.070985
Wilkie Liquors                27.684117

[61 rows x 1 columns]

zip_code  item_description      bottles_sold
58010.0  Member's Mark Spiced Rum      288
58022.0  Paramount Triple Sec          4
58111.0  Saints N Sinners Apple Pie    188
58131.0  Platinum 7x Vodka            48
58158.0  Hennessy VS                   24
        Kahlua Coffee                  3
        Kamora Coffee Liqueur          6
58263.0  Jagermeister w/2 Shot Glasses  84
58265.0  Kahlua Coffee                  72
58266.0  Avion Silver w/Powell & Mahoney Margarita Mix  98
Name: bottles_sold, dtype: int64

```

