***Final Assignment – Workearly***

*Βήμα 1ο:*

*Κατέβασα το αρχείο από το GitHub* ([finance\_liquor\_sales.sql](https://github.com/Workearly/Final-Assignment/blob/main/finance_liquor_sales.sql" \o "finance_liquor_sales.sql)) *και στη συνέχεια με το ακόλουθο query πήρα τα δεδομένα για τις χρονιές 2016-2019:*

Graphical user interface, text, application, email

Description automatically generated

*Βήμα 2ο:*

*Έκανα export τα δεδομένα σε αρχείο csv:*

*A screenshot of a computer

Description automatically generated*

*Βήμα 3ο:*

*Αποθήκευσα το αρχείο στο Pycharm και μετέτρεψα τα δεδομένα σε DataFrame, μέσω της βιβλιοθήκης Pandas.*

*Text

Description automatically generated*

*Βήμα 4ο:*

*Υπολόγισα το συνολικό αριθμό των μπουκαλιών που πουλήθηκαν και στη συνέχεια έκανα εισαγωγή στο DataFrame νέα στήλη που αντιστοιχεί στο ποσοστό πώλησης κάθε μπουκαλιού:*

*Text

Description automatically generated*

*Βήμα 5ο:*

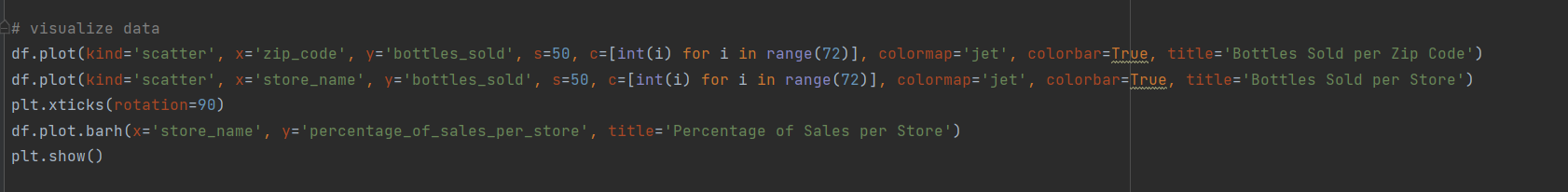
*Έκανα GroupBy το DataFrame με βάση το zip\_code, store\_name, item\_description με χρήση aggregation sum(), για να έχουμε στο τέλος το συνολικό ποσοστό πώλησης ανά κατάστημα και τον αριθμό μπουκαλιών που πουλήθηκαν:*

*Text

Description automatically generated*

*Βήμα 6ο:*

*Με χρήση του command DataFrame.plot της Pandas έκανα απεικόνιση των δεδομένων για την εύρεση του δημοφιλέστερου προϊόντος:*

**

*Και προέκυψαν τα ακόλουθα διαγράμματα:*

*Chart, scatter chart

Description automatically generated*

*Chart, scatter chart

Description automatically generated*

*A picture containing chart

Description automatically generated*

*Δυσκολίες κατά το assignment:*

* *Άνοιγμα του αρχείου sql κατά το κατέβασμα από το GitHub, καθώς ότα το ανέβαζα στο mySQL Workbench, μου έβγαζε κείμενο σε HTML.*
* *Ποια ήταν τα σωστά δεδομένα που έπρεπε να γίνουν GroupBy.*
* *Πως να αντλήσω τα δεδομένα που έκανα groupby(zip\_code κλπ.) για να τα χρησιμοποιήσω μετά στα plot. Λύθηκε με το as\_index=False, το οποίο αντί για index τα κάνει στήλες.*
* *Παραμέτρους στο plot, καθώς και πως να φτιάξω παλέτα χρωμάτων, ώστε κάθε παρατήρηση να εμφανίζεται με διαφορετικό χρώμα.*