

Final Assignment Report

Georgios Anastasiadis

1 Dataset

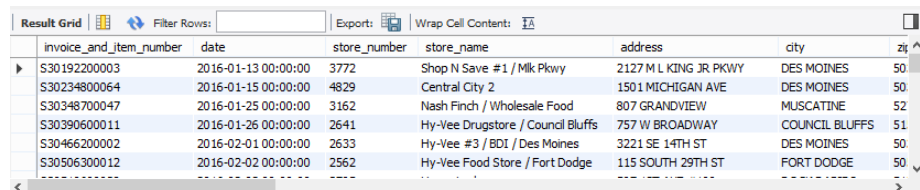
Αρχικά κατέβασα το Dataset `finance_liquor_sales.sql` το οποίο περιείχε τα απαραίτητα δεδομένα για την ολοκλήρωση του project. Έπειτα φόρτωσα το αρχείο στο MySQL Workbench ώστε να μπορέσω να το επεξεργαστώ.

2 CSV file export

Στο συγκεκριμένο βήμα έπρεπε να εξάγουμε της όλες τις στήλες για τις οποίες ισχύει ότι το **έτος ανήκει στο διάστημα [2016,2019]**. Για την περάτωση του παραπάνω βήματος χρησιμοποιήθηκε το παρακάτω query.

```
7  /*Query to get all the columns of the table between the years 2016-2019*/
8  •  SELECT *
9     FROM finance_liquor_sales
10     WHERE date >= '2016-01-01' and date <='2019-12-31'
11     ORDER BY date ASC;
12  |
```

Η εκτέλεση του συγκεκριμένου query έδωσε το παρακάτω αποτέλεσμα



The screenshot shows the MySQL Workbench interface with the 'Result Grid' tab selected. It displays the results of the SQL query, showing columns: invoice_and_item_number, date, store_number, store_name, address, city, and zip. The results are sorted by date in ascending order.

| invoice_and_item_number | date | store_number | store_name | address | city | zip |
|-------------------------|---------------------|--------------|-----------------------------------|-----------------------|----------------|-----|
| S30192200003 | 2016-01-13 00:00:00 | 3772 | Shop N Save #1 / Milk Pkwy | 2127 M L KING JR PKWY | DES MOINES | 50 |
| S30234800064 | 2016-01-15 00:00:00 | 4829 | Central City 2 | 1501 MICHIGAN AVE | DES MOINES | 50 |
| S30348700047 | 2016-01-25 00:00:00 | 3162 | Nash Finch / Wholesale Food | 807 GRANDVIEW | MUSCATINE | 52 |
| S30390600011 | 2016-01-26 00:00:00 | 2641 | Hy-Vee Drugstore / Council Bluffs | 757 W BROADWAY | COUNCIL BLUFFS | 51 |
| S30466200002 | 2016-02-01 00:00:00 | 2633 | Hy-Vee #3 / BDI / Des Moines | 3221 SE 14TH ST | DES MOINES | 50 |
| S30506300012 | 2016-02-02 00:00:00 | 2562 | Hy-Vee Food Store / Fort Dodge | 115 SOUTH 29TH ST | FORT DODGE | 50 |

το οποίο έκανα export σε CSV αρχείο κλικάροντας το **Export**.

3 Python queries

Αρχικά έκανα **import** την απαραίτητη βιβλιοθήκη, ώστε δημιουργήσω ένα dataframe από το `finance_liquor_sales(2016-2019).csv`.

```
1 import pandas as pd
2
3 df = pd.read_csv("finance_liquor_sales(2016-2019).csv")
```

Το πρώτο ερώτημα αφορούσε την εξαγωγή των πιο δημοφιλών αντικείμενων βάσει πωλήσεων για κάθε ένα `zip_code`. Παρακάτω φαίνεται το κομμάτι κώδικα που αντιστοιχεί στην υλοποίηση του παραπάνω ερωτήματος.

```
5 #Item sold based on zip_code
6 item_per_zipcode = df.groupby(['zip_code', 'item_description']).agg({'bottles_sold': 'max'}).reset_index()
7
8 #We find the item with max sales based on zip_code
9 popular_item = item_per_zipcode.sort_values('bottles_sold', ascending=False).drop_duplicates('zip_code')
10 popular_item = popular_item.sort_values('zip_code')
11
12 print(popular_item, '\n')
```

Η εκτέλεση του παραπάνω κώδικα έδωσε το παρακάτω αποτέλεσμα

```
C:\Users\George\Desktop\Final_Assignment\venv\Scripts\python.exe C:/Users/George/Desktop/Final_Assign
zip_code      item_description  bottles_sold
0  50010.0      Member's Mark Spiced Rum      288
1  50022.0      Paramount Triple Sec          4
2  50111.0      Saints N Sinners Apple Pie     108
3  50131.0      Platinum 7x Vodka             48
4  50158.0      Hennessy VS                   24
7  50263.0      Jagermeister w/2 Shot Glasses   84
8  50265.0      Kahlua Coffee                  72
```

Run Python Packages TODO Python Console Problems Terminal Services

Το δεύτερο ερώτημα ζητούσε να βρεθεί το ποσοστό των μπουκαλιών που πωλήθηκαν από κάθε κατάστημα. Παρακάτω φαίνεται το κομμάτι κώδικα που αντιστοιχεί στην υλοποίηση του παραπάνω ερωτήματος.

```
14 #Total amount of bottles sold
15 sum_of_bottles_sold = df['bottles_sold'].sum()
16 #sold bottles for every store
17 sb_for_each_store = df.groupby(['store_number', 'store_name']).agg({'bottles_sold': 'sum'})
18 #The result as a percentage
19 sb_for_each_store['bottles_sold'] = ((sb_for_each_store['bottles_sold']/sum_of_bottles_sold)*100).round(3)
20 #Rename the column 'bottles_sold'
21 sb_for_each_store.rename(columns={'bottles_sold': 'Percentage_of_bottles_sold'}, inplace=True)
22
23 print(sb_for_each_store.to_string())
```

Η εκτέλεση του παραπάνω κώδικα έδινε το παρακάτω αποτέλεσμα

| store_number | store_name | Percentage_of_bottles_sold |
|--------------|---------------------------------------|----------------------------|
| 2178 | Double D Liquor Store | 0.018 |
| 2465 | Sid's Beverage Shop | 0.089 |
| 2512 | Hy-Vee Wine and Spirits / Iowa City | 1.384 |
| 2515 | Hy-Vee Food Store #1 / Mason City | 1.171 |
| 2522 | Hy-Vee Wine and Spirits / Spirit Lake | 0.852 |
| 2538 | Hy-Vee Food Store #3 / Waterloo | 0.319 |
| 2544 | Hy-Vee Food Store / Marshalltown | 0.106 |

4 Data presentation

Στο τελευταίο στάδιο πραγματοποιήθηκε η οπτικοποίηση των δεδομένων με την χρήση του **Tableau**. Στο παρακάτω link: <https://public.tableau.com/app/profile/george.anastasiadis> υπάρχει το dashboard.