

CMPE 255 Assignment 1

Farazuddin Mohammad (016176836)

October 2, 2022

1 Introduction

The dataset selected for analysis/mining is the Iowa Liquor Sales data set from the listed public datasets in the Google's BigQuery service. The dataset contains every wholesale order of liquor by all grocery stores, liquor stores, convenience stores, etc., with details about the store and location, the exact liquor brand and size, and the number of bottles ordered. You can find more details, as well as sample queries, in the GCP Marketplace here: <https://console.cloud.google.com/marketplace/details/iowa-department-of-commerce/iowa-liquor-sales>"
(bwo) Note:- Due to Limited Computational Capacity I have limited the dataset to first 10,000 instances(purchases)

My goal is to identify trends and make observations from the visualization plotted.

2 Data Set Info

Feature	DataType
invoice_and_item_number	object
date	object
store_number	object
store_name	object
address	object
city	object
zip_code	object
store_location	object
county_number	object
county	object
category	object
category_name	object
vendor_number	object
vendor_name	object
item_number	object
item_description	object
pack	int64
bottle_volume_ml	int64
state_bottle_cost	float64
state_bottle_retail	float64
bottles_sold	int64
sale_dollars	float64
volume_sold_liters	float64
volume_sold_gallons	float64
Date	datetime64[ns]
Year	int64
Month	int64
Day	int64

Above Table lists the features and the type of data.

3 Fetch Data

Below is the SQL query made to BigQuery, due to limited computational capacity I have limited the instances to only 10,000.

```
SELECT
  *
FROM
  `bigquery-public-data.iowa_liquor_sales.sales`
LIMIT 10000
```

4 Bar Plot

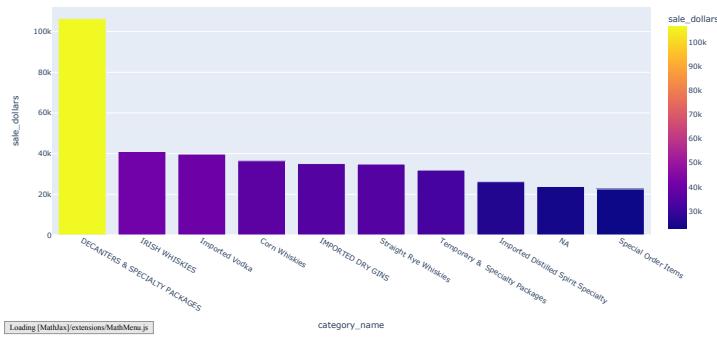


Figure 1: Top 10 Most selling Liquor Categories

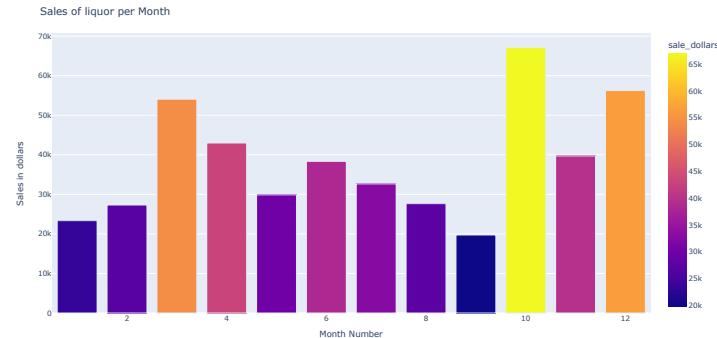


Figure 2: Cumulative Monthly Sales

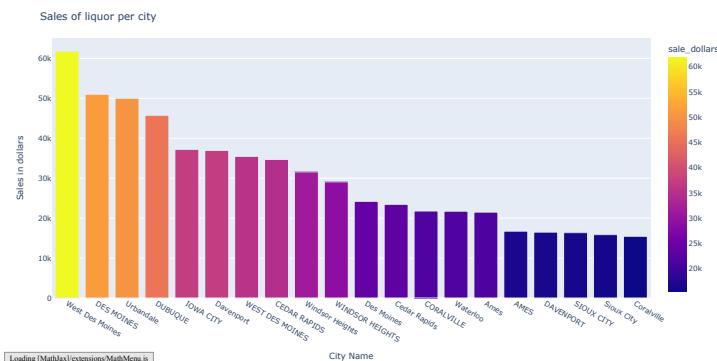


Figure 3: Sales of Top 10 Cities in Liquor Purchase

Above are the bar plots which show trends in purchases. We notice from Figure 1 that "Decanters Speciality Packages" is the most bought category of liquor, from Figure 2 we notice the October month over the years has observed highest sales in liquor and the city West Des Moines has the highest sales in liquor.

5 Heatmap Plot

Below is the heatmap of the correlation matrix of the numeric features like:

- pack
- bottle volume ml
- state bottle cost
- state bottle retail
- bottles sold
- sale dollars
- volume sold liters
- volume sold gallons

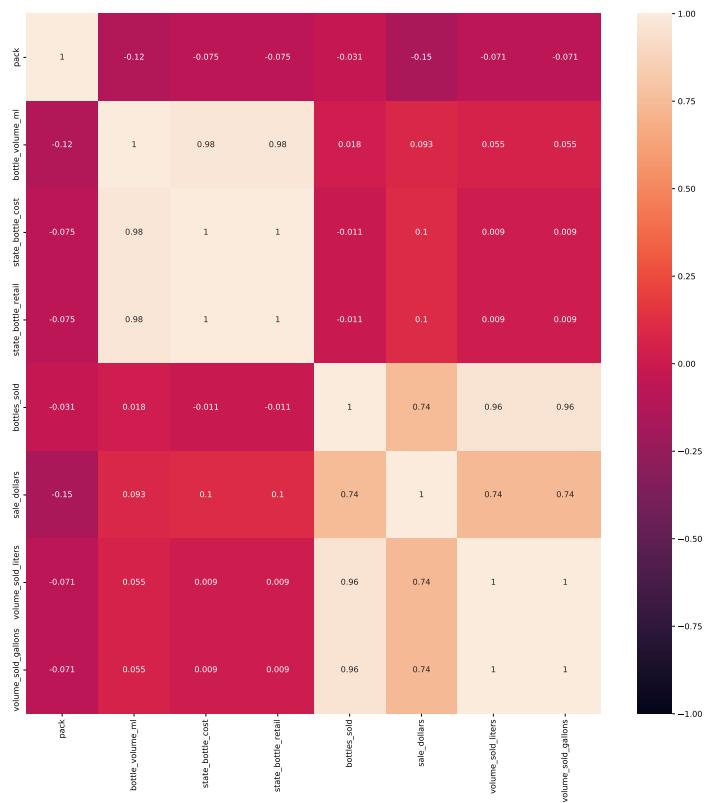


Figure 4: Heatmap plot for correlation matrix

We notice high positive correlation between sales in dollars (target variable) and volume sold in liters

6 Scatter Plot - Regression Line

Scatter Plot to analyze the positive correlation between volume sold in liters and sales in dollars observed from the above heatmap.

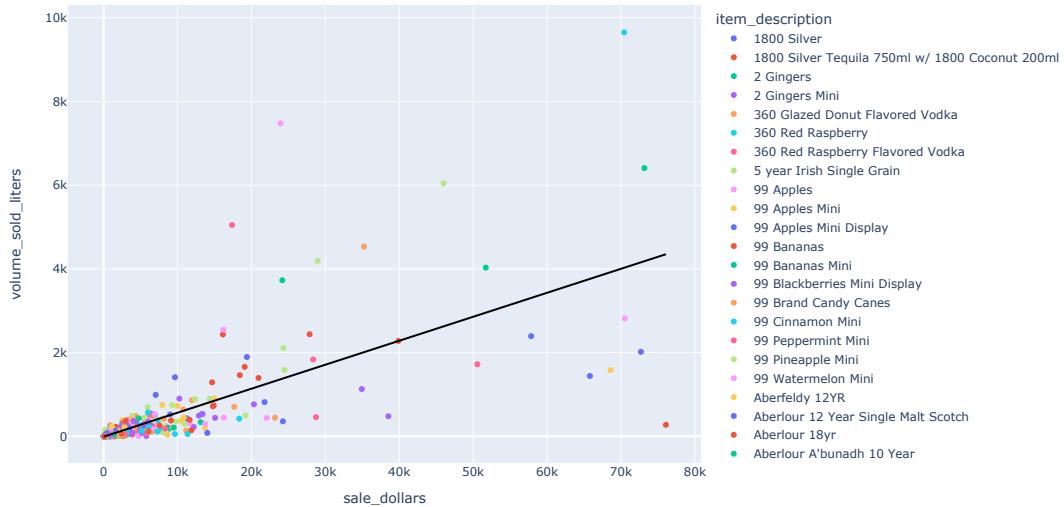


Figure 5: Scatter Plot between Volume Sold in liters and Sales in Dollars

7 Histogram Plot

Histogram plot to visualize the Distribution of number of bottles sold in a transaction. We observe that most of the transaction have total bottles less than 20 and big orders are less frequent.

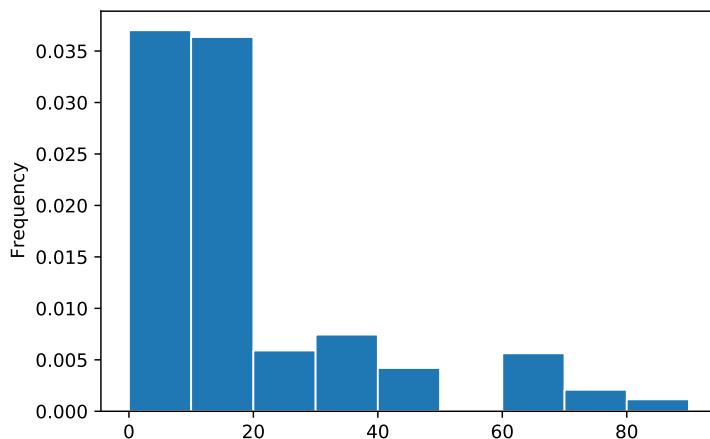


Figure 6: Dash board page

8 Line Plot

Line plot to observe the Monthly Purchase trends of Top 10 Cities with highest purchases. October month is where we observe highest sales.



Figure 7: Line plot between Sales in Dollars and Months over the years

References

- [bwo] Iowa Liquor Sales Dataset. <https://console.cloud.google.com/marketplace/details/iowa-department-of-commerce/iowa-liquor-sales>.

9 Dashboard Screen Shots



Figure 8: Dash board page

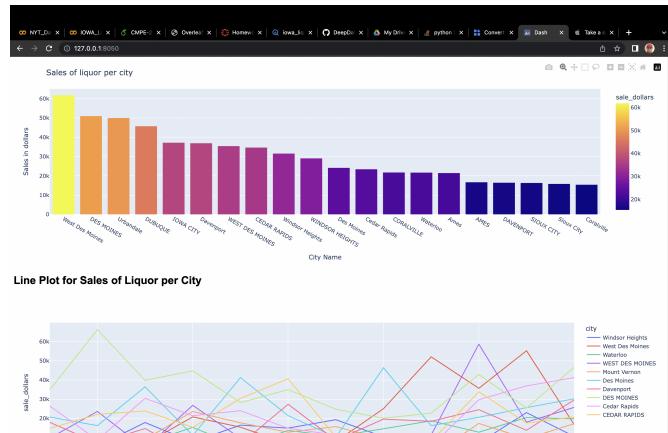


Figure 9: Dash board page

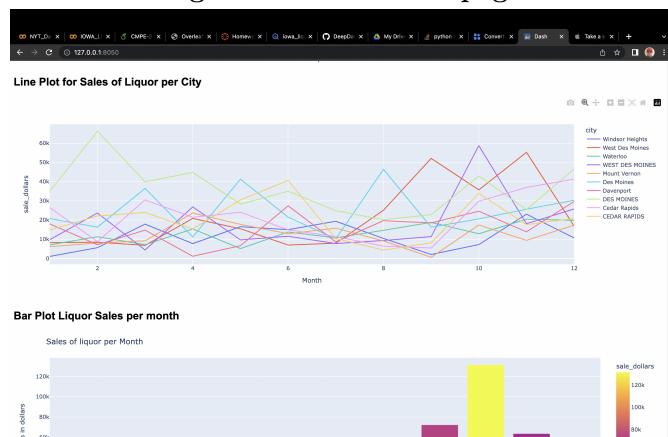


Figure 10: Dash board page



Figure 11: Dash board page

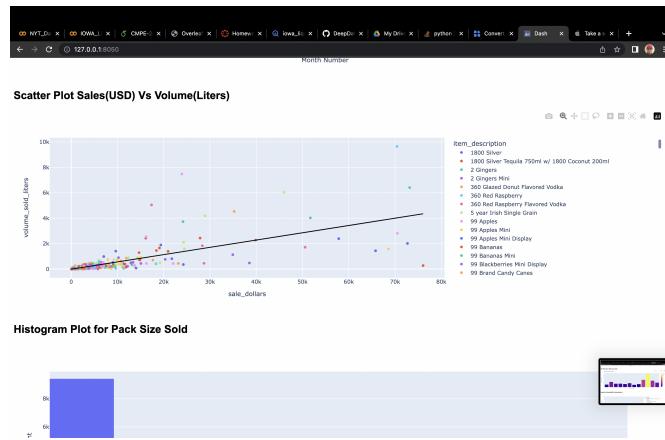


Figure 12: Dash board page



Figure 13: Dash board page



Figure 14: Dash board page