

Iowa Liquor Sales

Sale activity for alcohol in the United States

Invoice/Item Number

Store Name

Zip Code

Store Number

Date

Address

County Number

City

Store Location

Category

County

State Bottle Retail

State Bottle Cost

Category Name

Vendor Name

Volume Sold G

Sale \$

Bottles Sold

Item

Item Number

Volume Sold L

description

Bottle

Vendor Number

Pack

Volume

Date.	Locations.	Type.	Quantity.	Sales.	Quantity sold.
m/d/y	Vendors Shops Addresses Cities etc.	County Item number etc.	Bottle Volume Pack	Retail Sales Cost	Volume Sold Gallons Sold Bottles Sold

Handling datatypes.

mixed datatypes

'\$' near int/float

etc.

Dropping all NaN values.

*EDA advantage.

< 8k

Address
City
Zip Code
Zip Code
Category
Category Name
Store Location

+/- 79k

County
County Number

< 10

State Bottle Cost
State Bottle Retail
Sale (Dollars)

Uncertainty in data ~ 4k

State Bottle Cost	State Bottle Retail	Bottles Sold	Sale (Dollars)	Volume Sold (Liters)
0.00	0.00	12	0.0	9.00
0.00	0.00	2	0.0	2.00
0.00	0.00	3	0.0	2.25
0.00	0.00	12	0.0	9.00
0.00	0.00	4	0.0	3.00
...
9.96	14.94	1	0.0	0.75
29.53	44.30	1	0.0	1.20

Store Number

County Number

Category Category Name

Vendor Number Vendor Name

Item Number

Invoice/Item Number

Item description

Store Location

Address

Zip Code

Date

Store Name
County

Volume Sold L

Bottles Sold
Volume Sold G

State Bottle Retail
State Bottle Cost
Sale \$

Bottle
Volume

Pack

State Bottle Retail
State Bottle Cost
Sale \$

Bottle
Volume
Pack

Price per Liter: Retail / Bottles Volume

Margin: Retail Price - Cost
(direct profit)

Total Margin: Margin * Bottles Sold

Gross Margin: Total Sales -
Cost
(Indirect profit)

Total Margin

Total profit for the number of bottles sold in that day.

Bottle Volume

measured in L

Isolation Forest

0.7273926783998858

Mean Squared err = 188.1441431291468
Mean Abs err = 3.257868817685284

0.6656026828432925

Mean Squared err = 58.243627334666066
Mean Abs err = 2.957945603221284

Elliptic Envelope

0.7273926783998858

Mean Squared err = 188.1441431291468
Mean Abs err = 3.257868817685284

0.7288592577885218

Mean Squared err = 46.29865185726856
Mean Abs err = 2.9200742242853424

0.9700257854142832

Mean Squared err = 43.471374046404016
Mean Abs err = 0.6231902923905559

0.9598659297006594

Mean Squared err = 8.580622464052235
Mean Abs err = 0.5570057362497742

LR

RFR

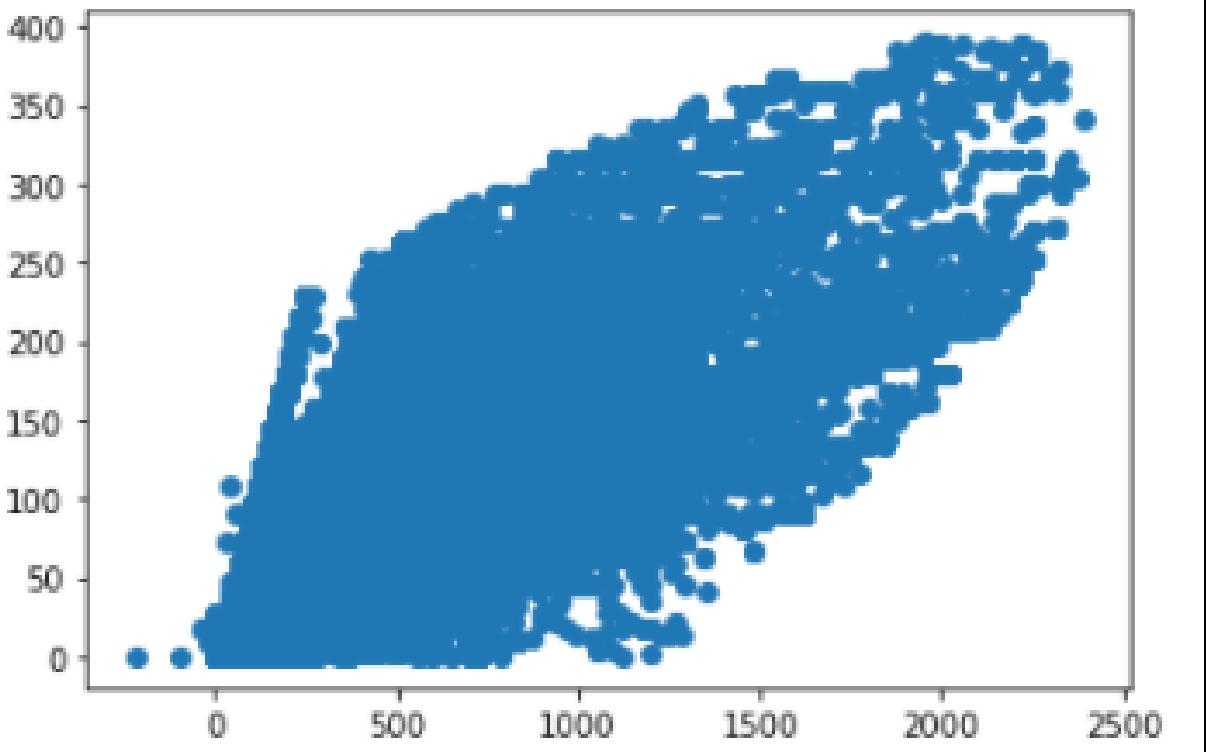
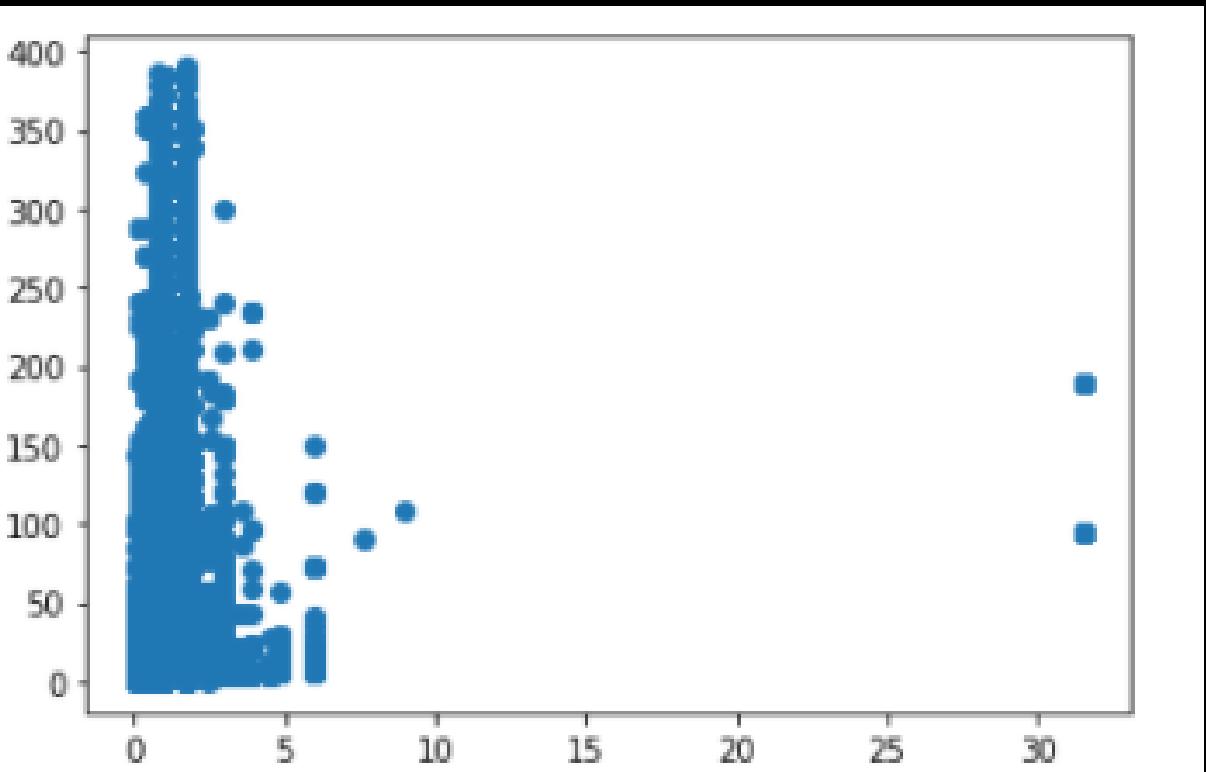
0.9700257854142832

Mean Squared err = 43.471374046404016
Mean Abs err = 0.6231902923905559

0.9656020312539905

Mean Squared err = 6.985564978014035
Mean Abs err = 0.5475214428941693

*contamination index not optimally choosed
removed ~ 12k



StandartScaler

Score: 0.9593178537225047

Mean Squared err = 6.886368807101966

Mean Abs err = 0.5466900647813306

Q/A