Farid Lisniawan Muzakki

02-17-2021

Iowa Liquor Sales Analysis

Problem:

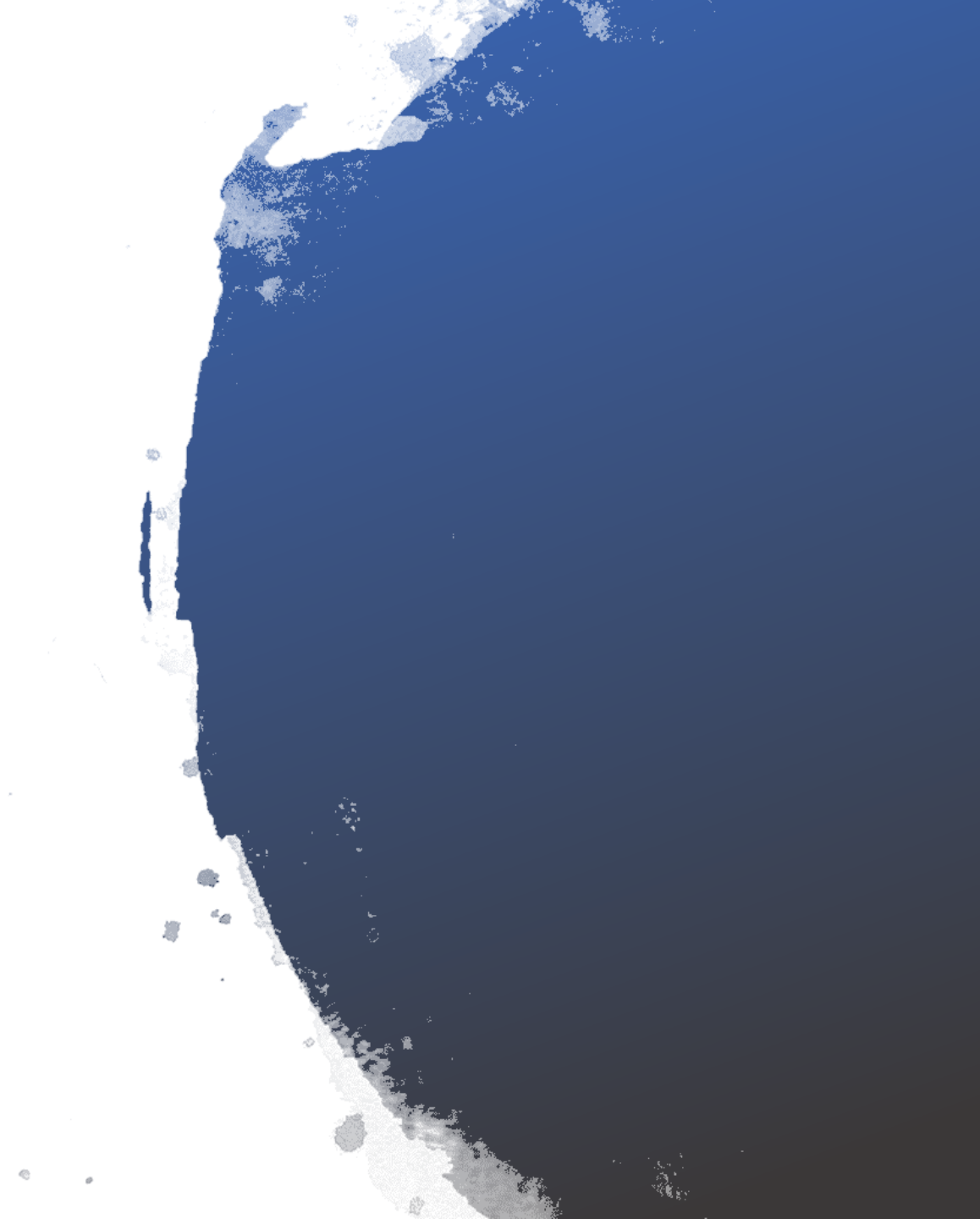
In Covid-19 Pandemic, a vendor from another state wants to develop their business. They plan to sell their products in the state of Iowa.

To realize their plan, they want to know which cities have the most potential to distribute their products and what kind of products are most frequently purchased by consumers during Covid-19 Pandemic.

Situation:

Because of the government regulation, the stores must apply Covid-19 safety protocols, such as limiting the number of visitors in store, closing the store early, etc. Many people also lost their jobs, so they don’t have money to buy liquor.

This situation makes some stores have to close their business until an indefinite time. Only some stores that have great innovations can survive this situation.



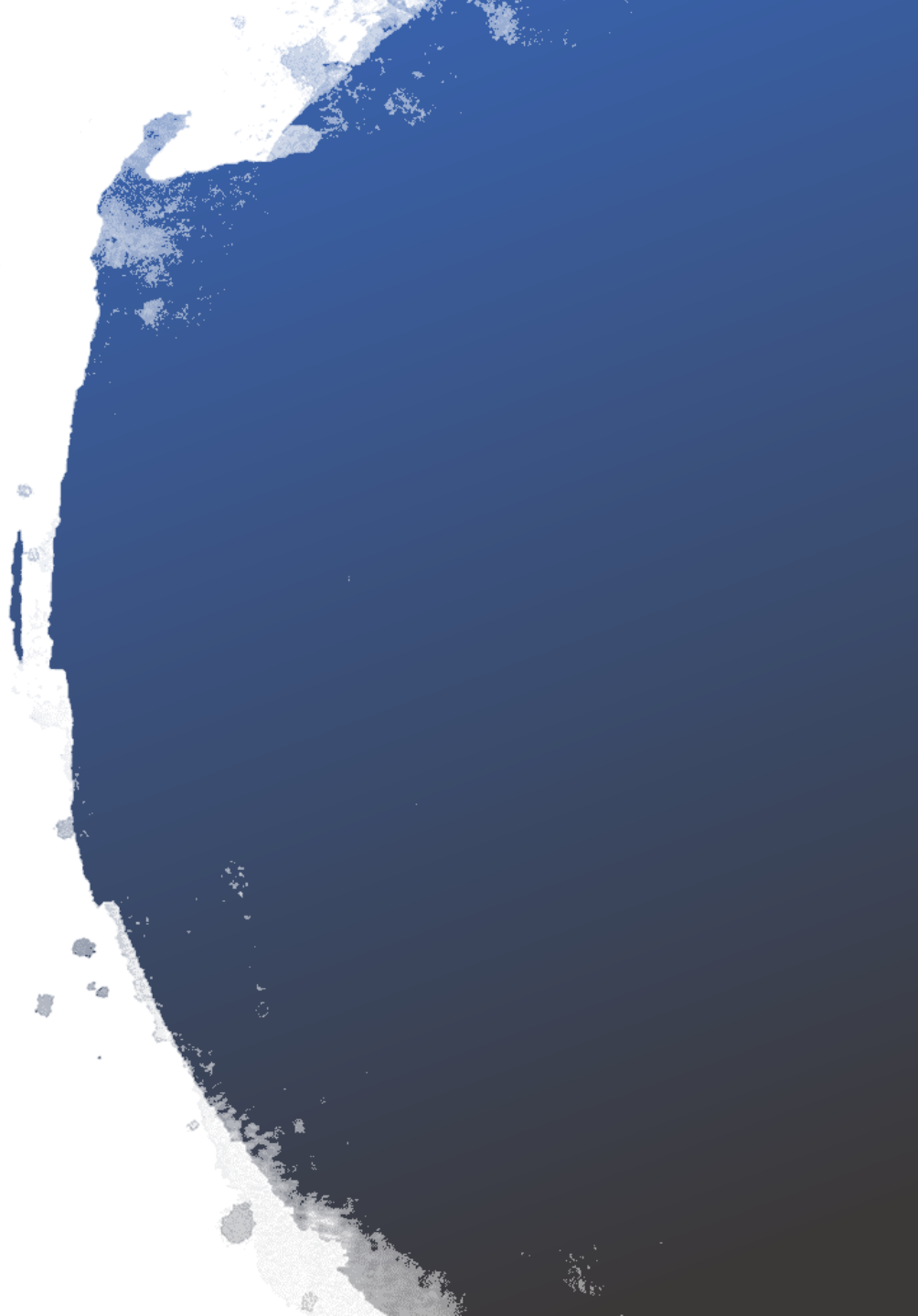
Preparing Data

To get a full insight about the effect of Covid-19 Pandemic to the liquor sales in Iowa, the data must be compared with the data before pandemic.

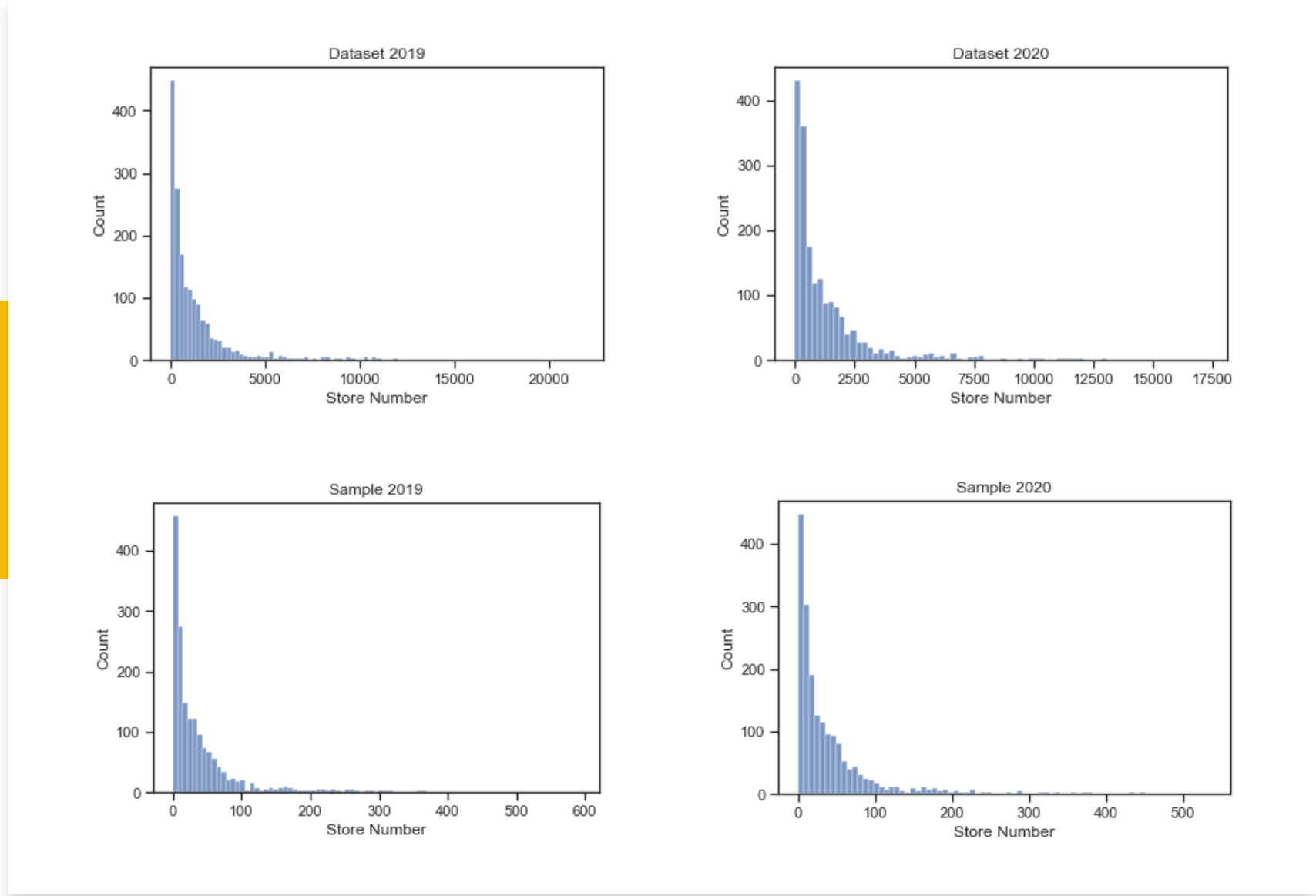
From the Iowa Liquor Sales Dataset, I extracted two new datasets, namely the liquor sales dataset during 2019 (before pandemic) and during 2020 (during pandemic).

There are over 2 million rows in each dataset. Due to the long computation time and the limitation of the hardware, I only took 3% from each dataset as a sample.

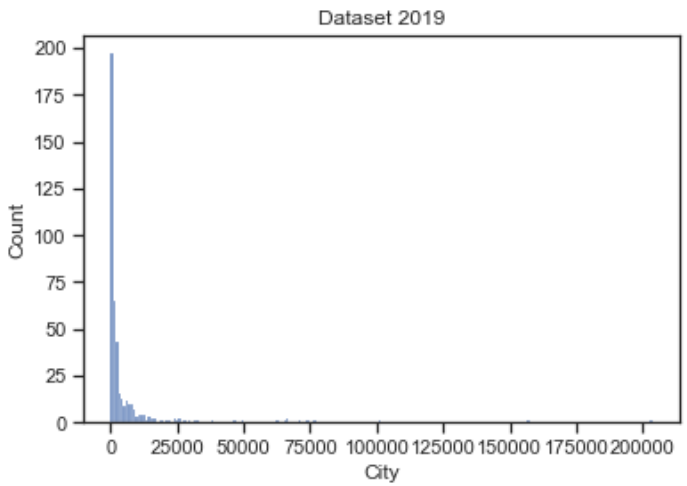
To make sure the sample has the same characteristic to the dataset, I made some histograms from five important columns in this analysis. If there are no big differences in dataset histogram and sample histogram, the sample has the same characteristic to the dataset.

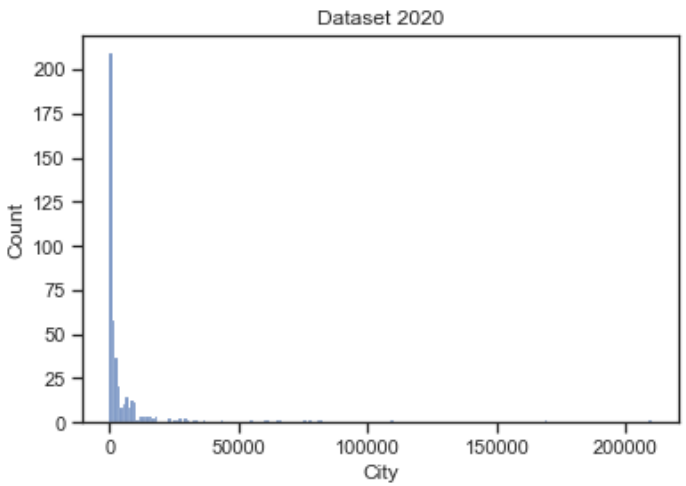


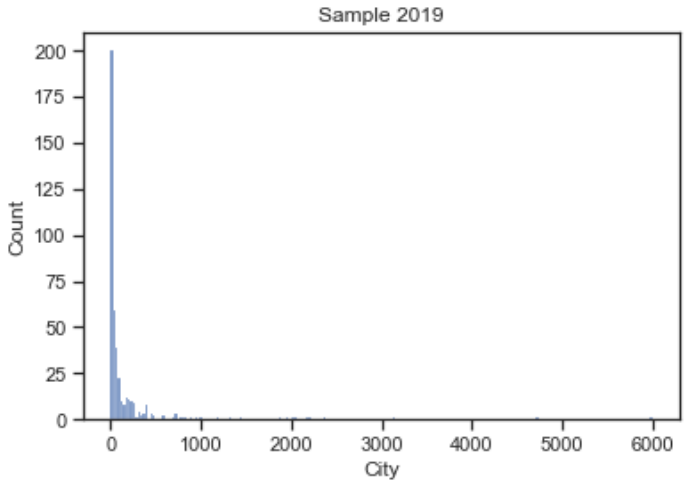
Column: ‘Store Number’

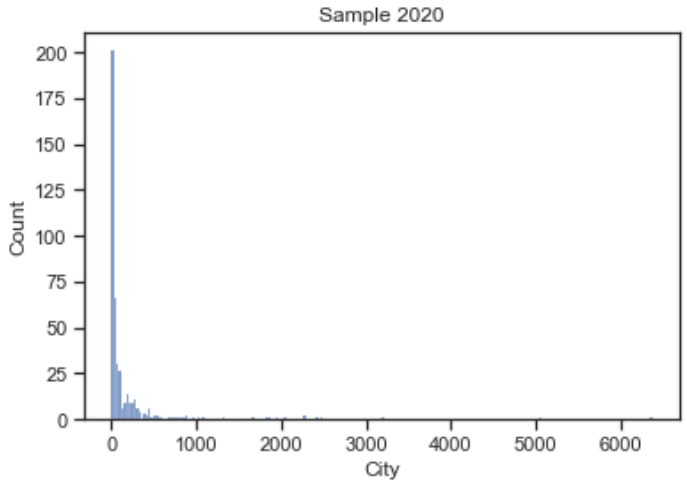


Column: ‘City’

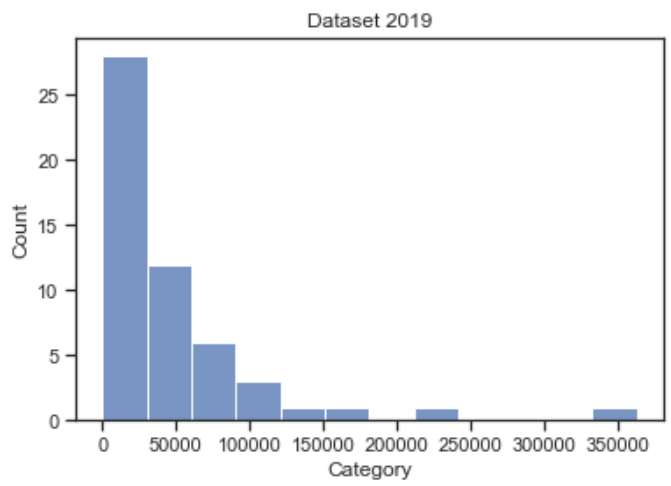

Dataset 2019
Count
City
0
25
50
75
100
125
150
175
200
0
25000
50000
75000
100000
125000
150000
175000
200000

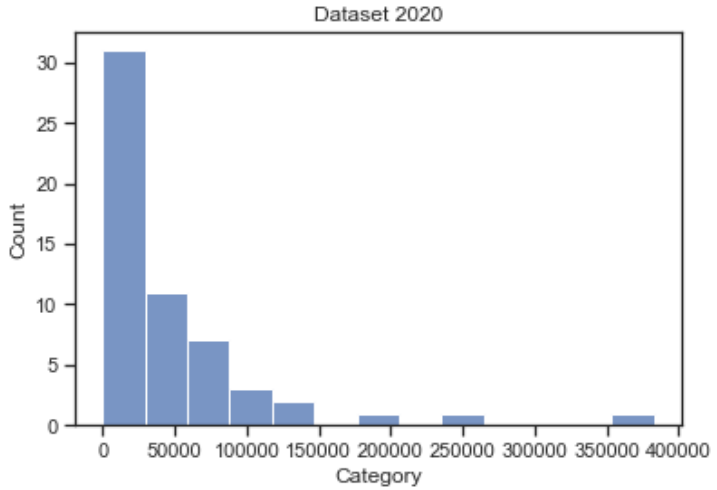

Dataset 2020
Count
City
0
25
50
75
100
125
150
175
200
0
50000
100000
150000
200000

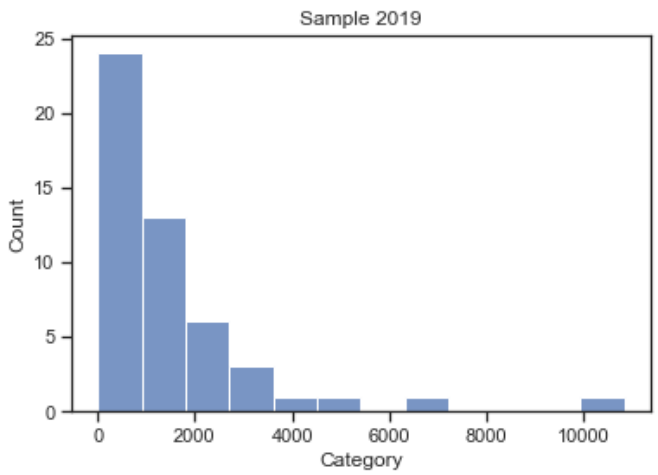

Sample 2019
Count
City
0
25
50
75
100
125
150
175
200
0
1000
2000
3000
4000
5000
6000

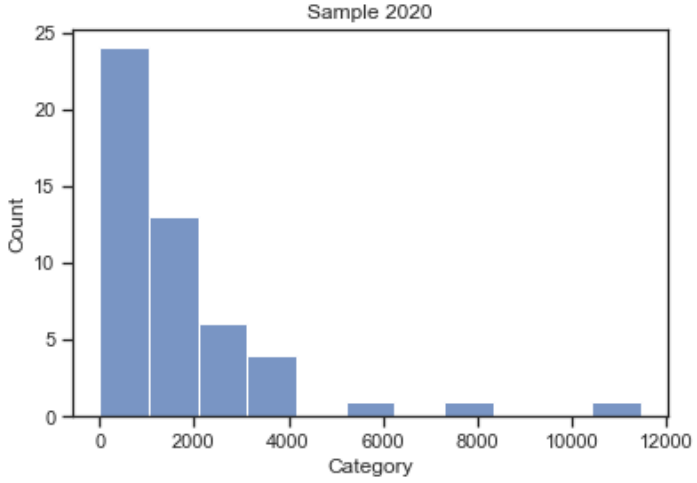

Sample 2020
Count
City
0
25
50
75
100
125
150
175
200
0
1000
2000
3000
4000
5000
6000

Column: ‘Category’









From the histograms, it can be said that there are no huge differences between sample and dataset, and they have the same characteristics.

There are some nan values in the table, so I changed it to 0.


Analyzing Data

In the public dataset, it states that the ‘state_bottle_cost’ column is the amount that the Alcoholic Beverages Division paid for each bottle of liquor ordered.

I redefined it into the amount that must be paid to the Alcoholic Beverages Division for each bottle of liquor ordered.

I made two new columns in the table. The first column is ‘State_Bottle_Income’ which is defined as the amount of income without Alcoholic Beverages Division tax for each bottle of liquor ordered.

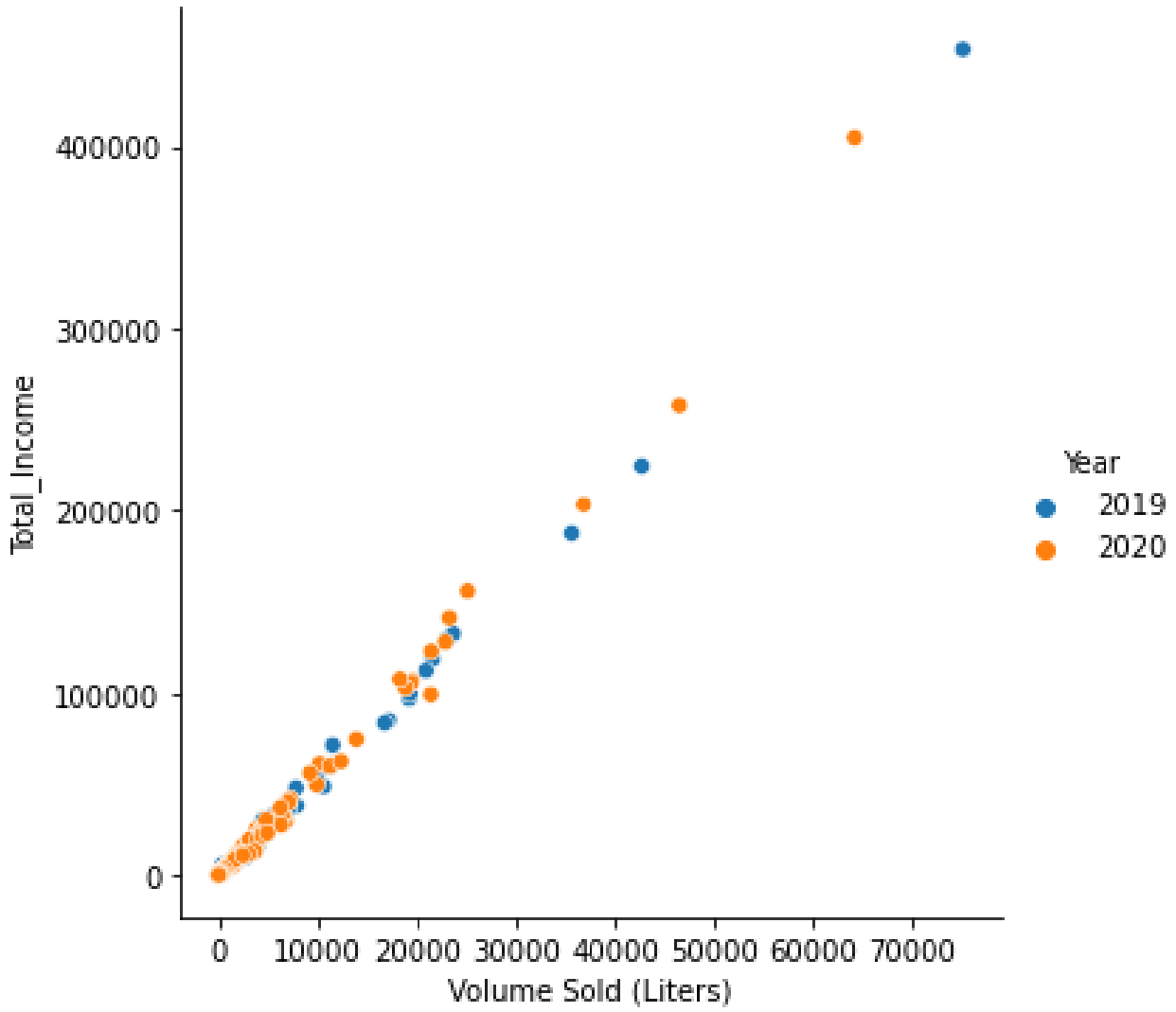
The column was calculated by subtracting the ‘State_Bottle_Retail’ column and ‘State_Bottle_Cost’ column.

The second column is ‘Total_Income’ which is defined as the total amount of income for each invoice of liquor ordered.

The column was calculated by multiplying the ‘State_Bottle_Income’ column and ‘Bottles_Sold’ column.

Then I calculated each city’s total income and total liquor volume sold (liters) in 2019 and 2020.

Total income and total volume sold are the best parameters to measure how potential a city is. I searched the linear correlation between these two columns, visually and statistically.



Correlation between income and volume sold

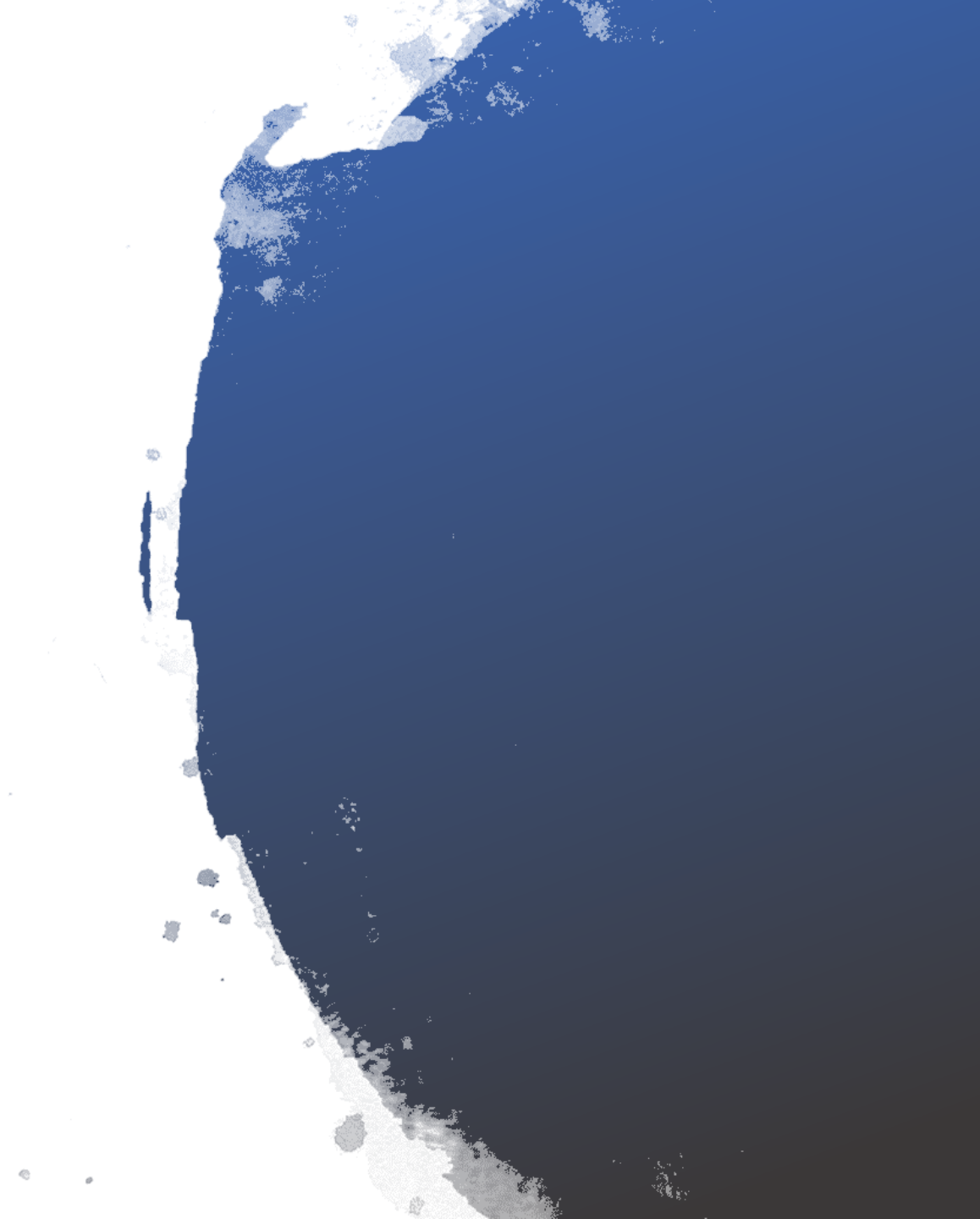
Visually, total income and total volume sold have an almost perfect linear correlation.

Statistically, the correlation between these two columns is 0.996 which means very good.

If total income in a city is increased, then most likely total volume sold will also increase.

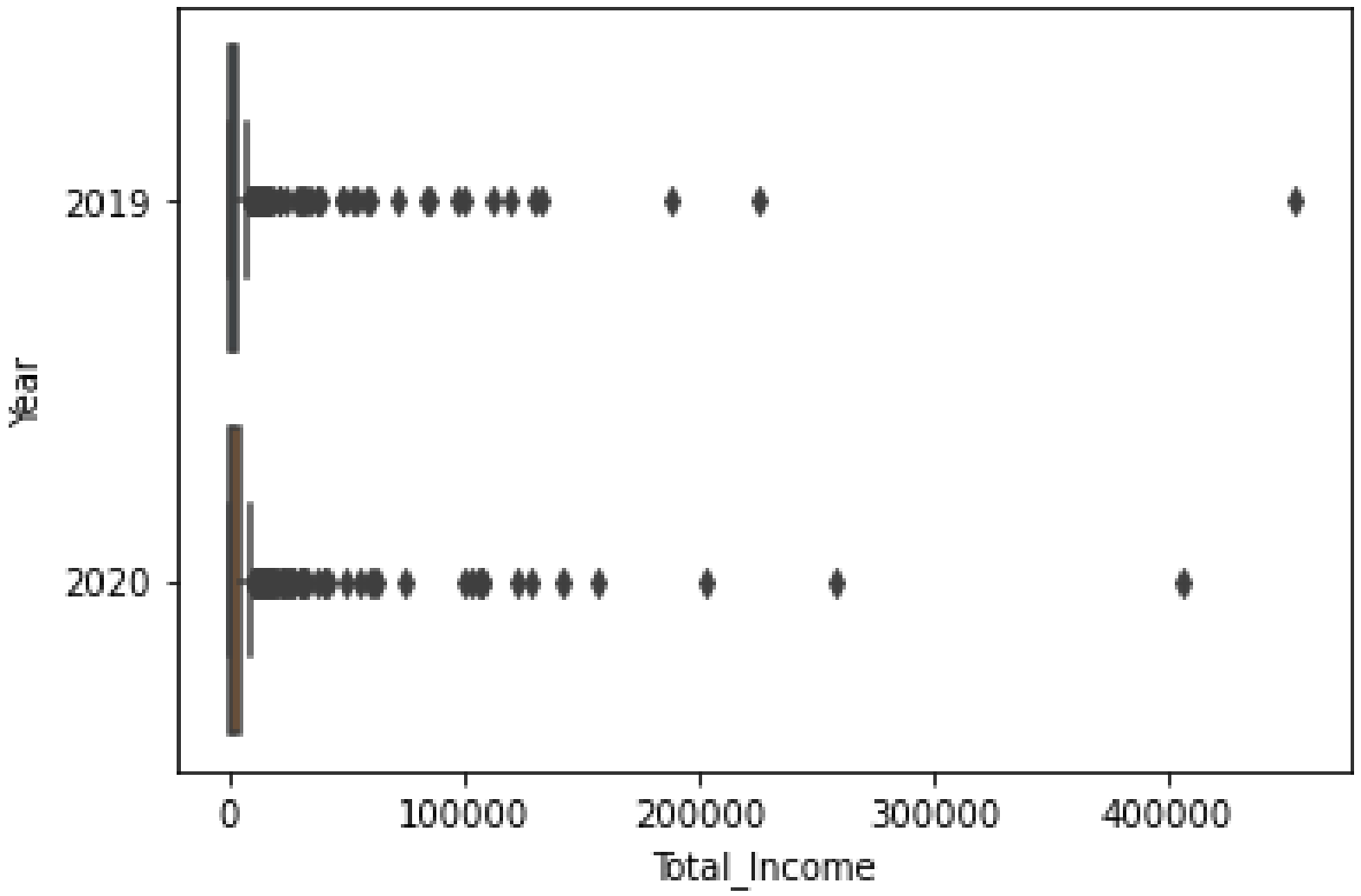
Good linear correlation between two columns means that they are dependent on each other. Therefore, one parameter is enough to measure how potential a city is.

I chose total income as a parameter to analyze a city.



I made a boxplot from the city total income table.

It seems that there are some cities that have enormous value differences to the sample statistic maximum value.



Cities Total Income Boxplots

Notes:

The vendor (the client) prefers to focus on several cities because they must build a warehouse in each city. The warehouse in a city can’t reach another city due to the long distances between cities.

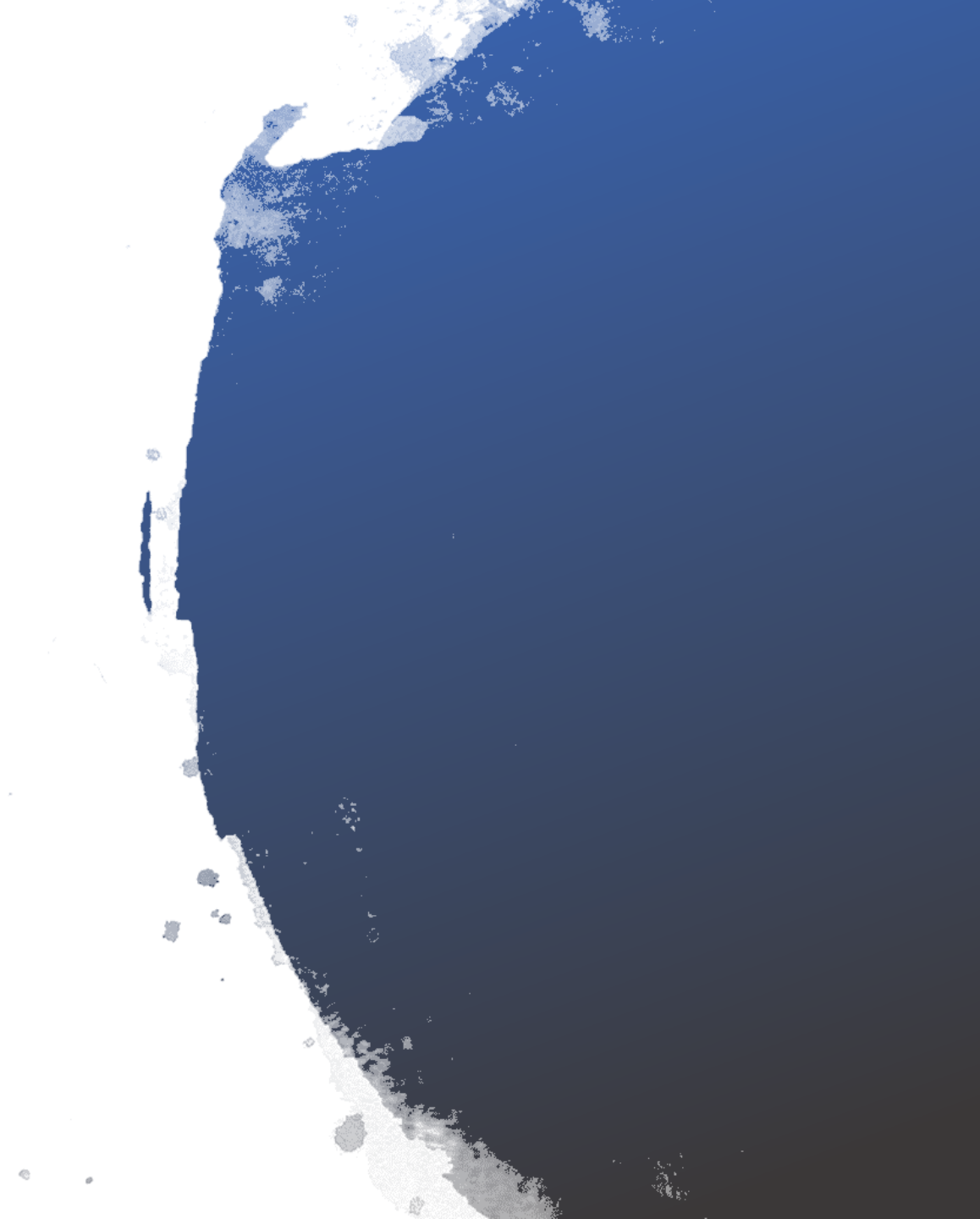
The vendor also provides all categories of liquor, and their plan is optimizing the category of liquor according to the demand in the city.

There are 71 cities which are labeled as outliers in 2019, and 74 cities in 2020. There are still too many cities to analyze, so I used other ways to find the top five most potential cities.

From those outliers, I took two types of cities. The first type is the cities which became outliers in 2019 and 2020. And the second type is the cities which just became outliers in 2020.

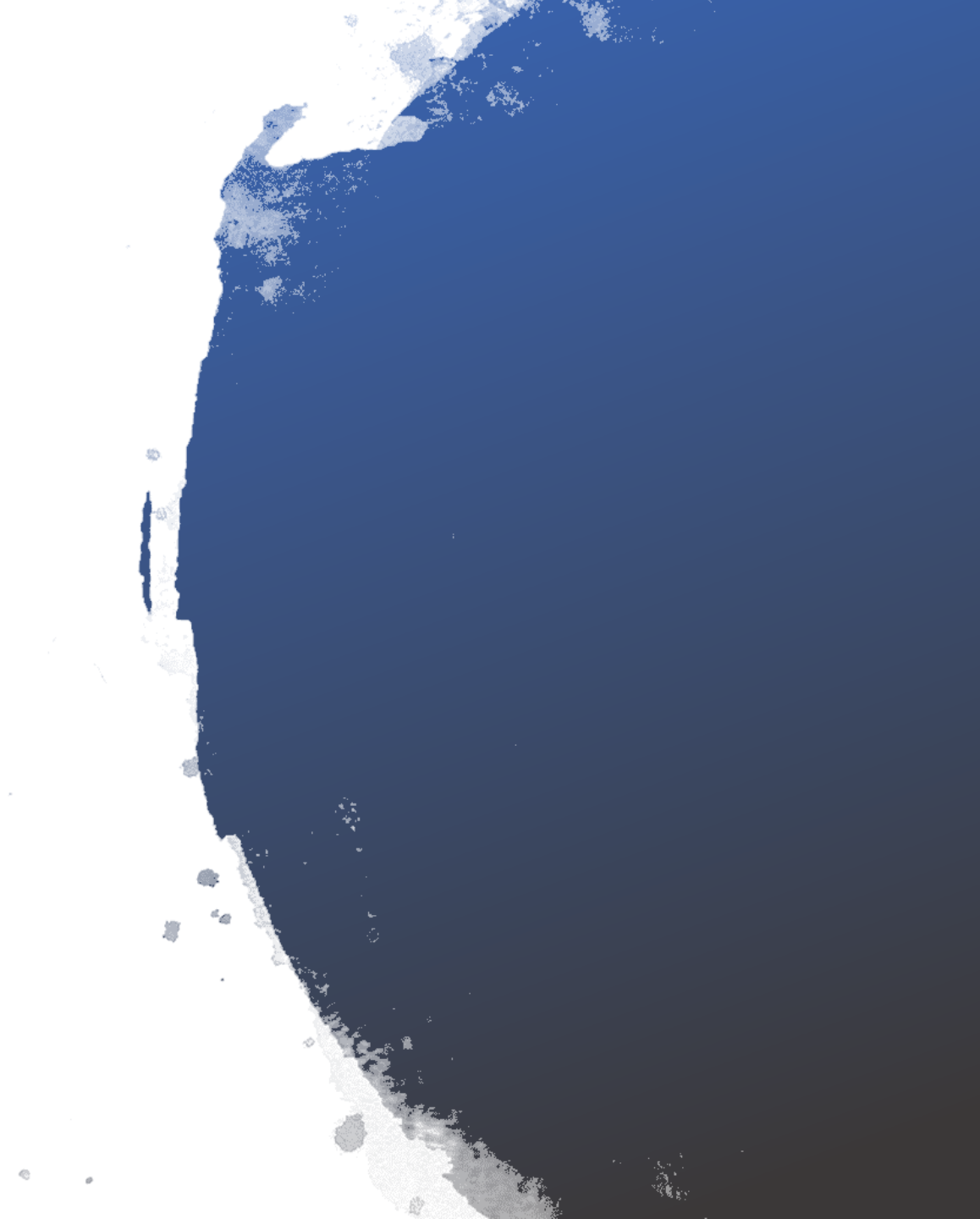
I calculated the growth percentage for the cities that I took. And I chose five cities with the best growth percentage from 2019 and 2020.

A positive and huge growth percentage means that most of the stores in the cities were able to face Covid-19 Pandemic and had great innovation to reach their consumers.



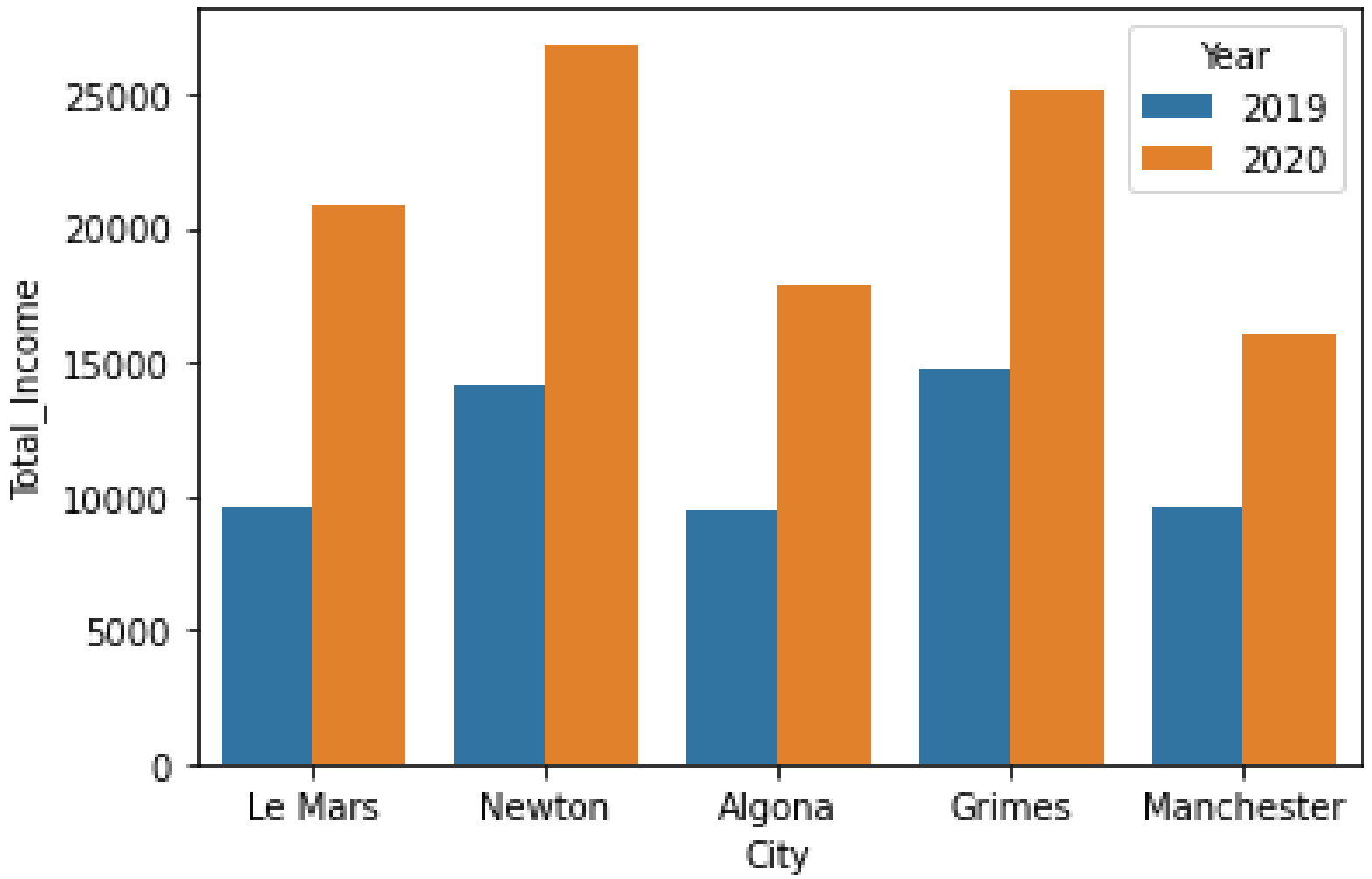
This situation is an advantage for the vendor because their products are still in demand by consumers in this city, even the demand is increasing.

After some analysis, I got the top five most potential cities.



Rank	City	Income Growth Percentage
1	Le Mars	118.24%
2	Newton	91.05%
3	Algona	89.31%
4	Grimes	69.73%
5	Manchester	66.58%

Income Growth Percentage in Top Five Cities



Income Growth in Top Five Cities

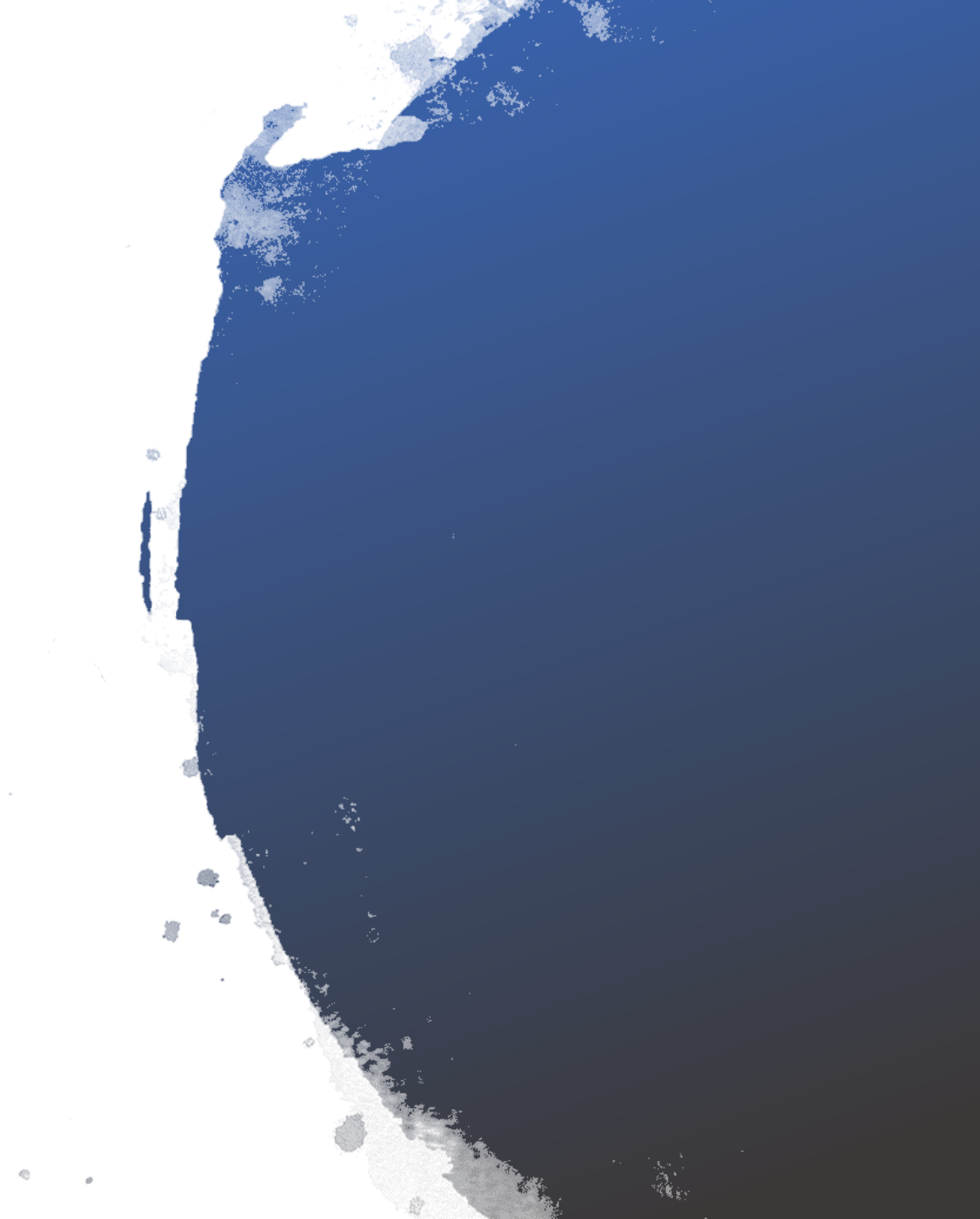
Notes:

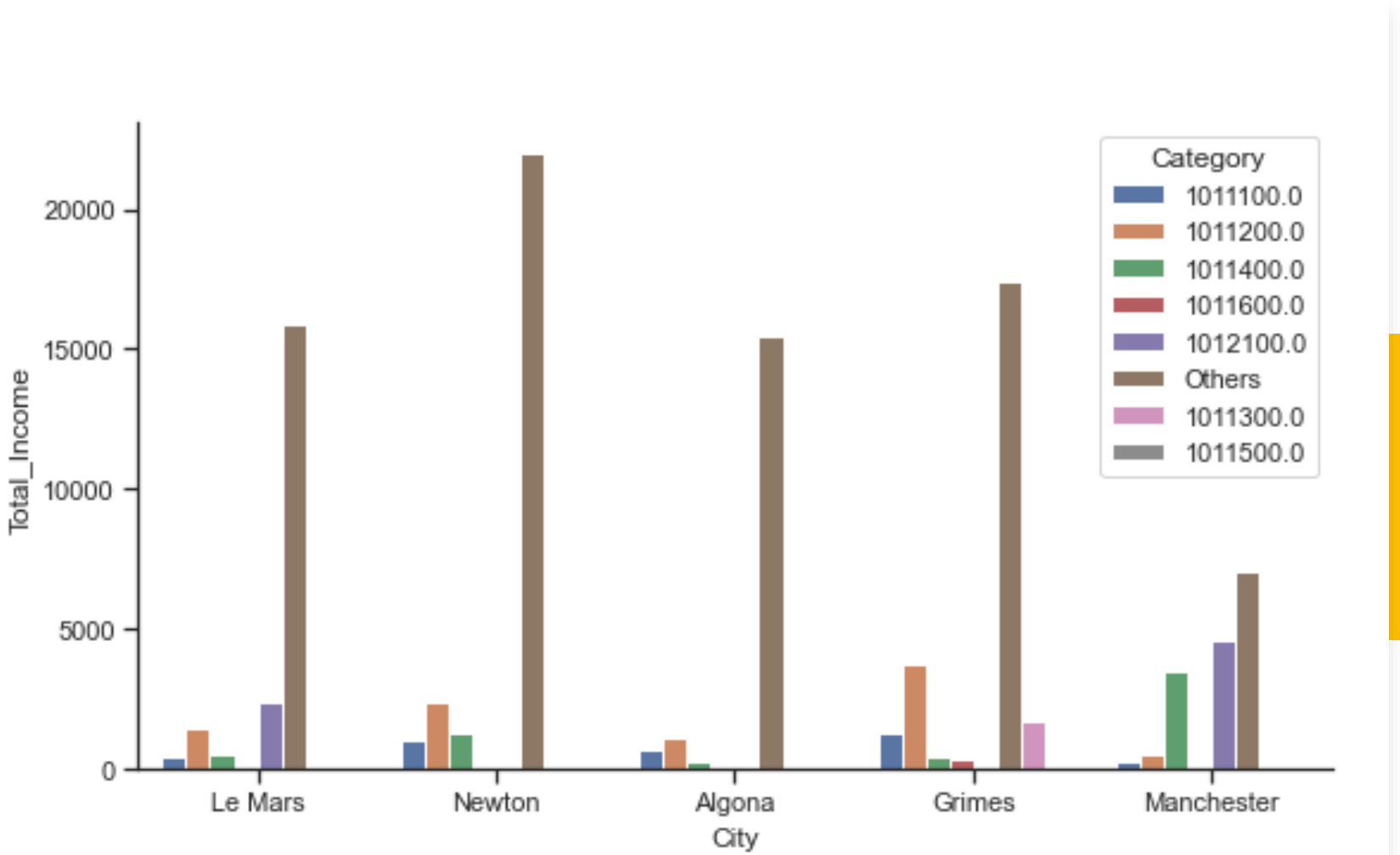
It is possible if there are any cities that are not outliers and have a huge income growth percentage. But their total incomes are too small, the vendor wants to get income as high as possible.

Outliers fulfill this requirement because they have higher income compared to other cities in general.

From those five cities, I extracted five liquor categories that are the most purchased by consumers for each city.

By knowing the most purchased liquor categories in each city, the vendor can maintain their product stocks in the warehouse. The vendor also can focus on marketing for several categories in each city.





The Most Purchased Liquor Categories

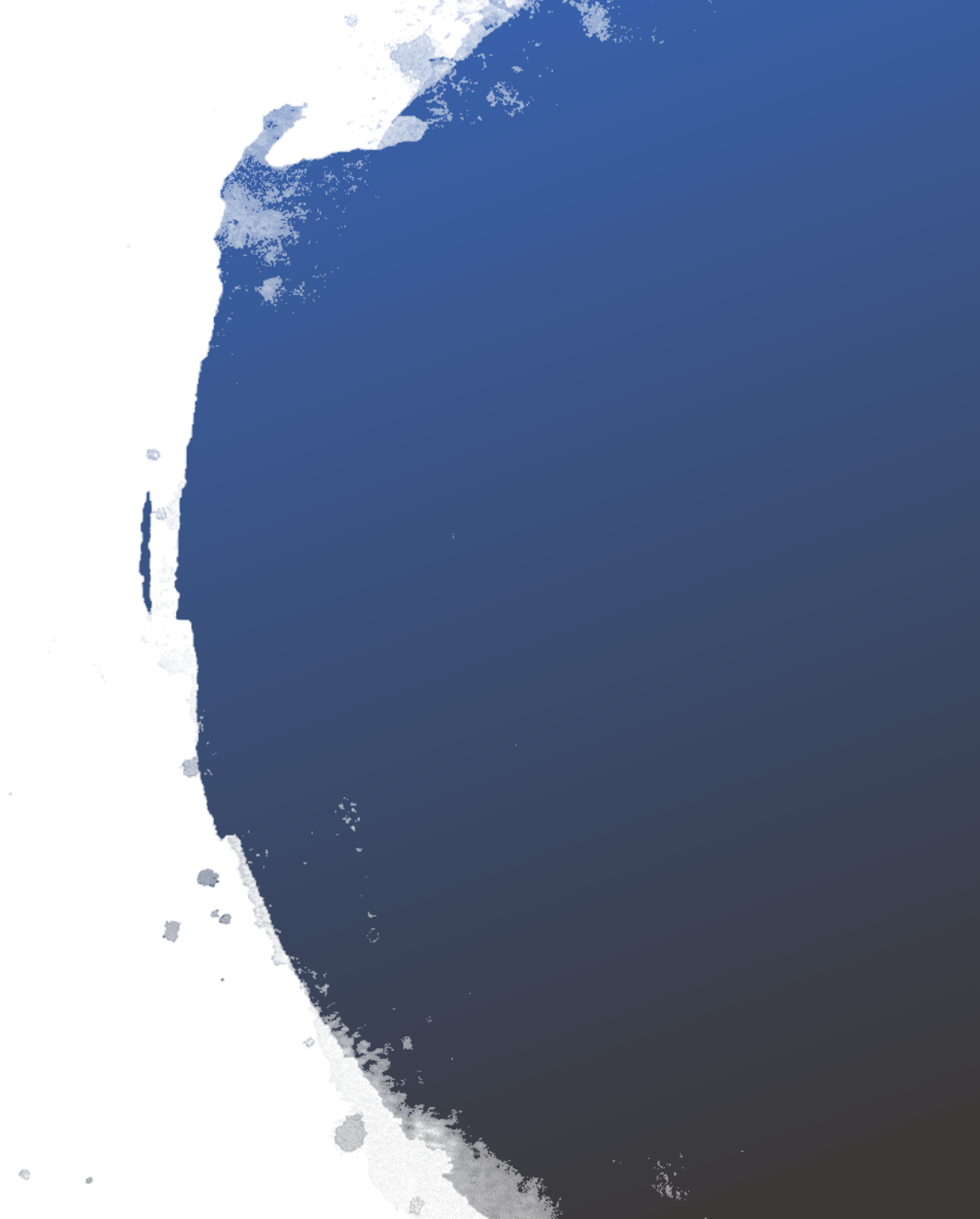
In all cities, there are many liquor categories with small purchased quantities, and the sum of those categories is higher than the most five purchased liquor categories.

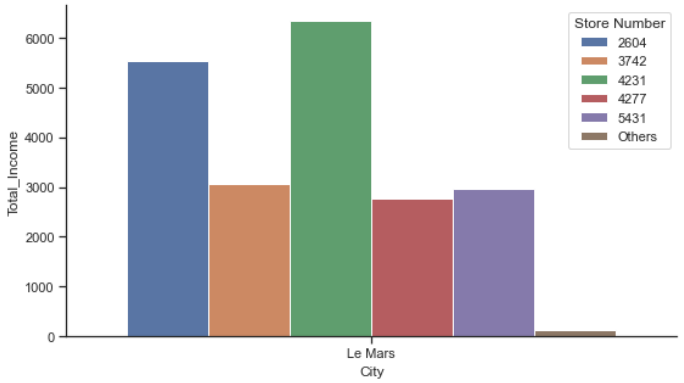
In Le Mars and Manchester city, liquor category 1012100 is the most purchased liquor. Meanwhile, in Newton, Algona, and Grimes city, liquor category 1011200 is the most purchased liquor.

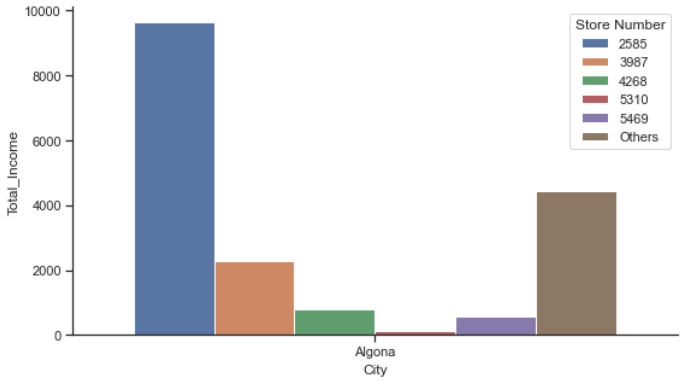
In each city, maybe there are some stores which have a big number of liquor transactions and are leading the liquor sales.

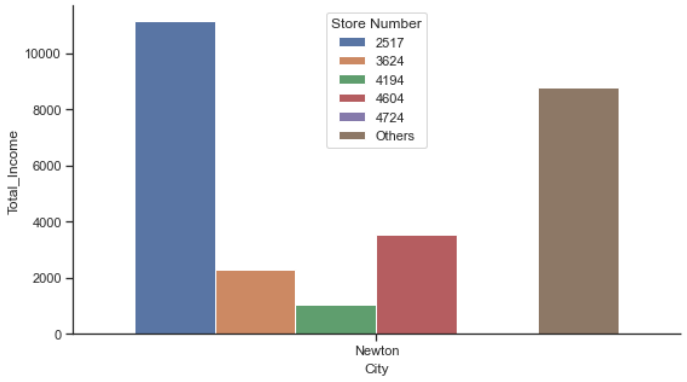
By knowing those stores, the vendor can make a strategy and approach them to sell the vendor products. It will help the vendor brand to be popular faster.

I extracted the most five popular stores in Le Mars, Newton, and Algona city.









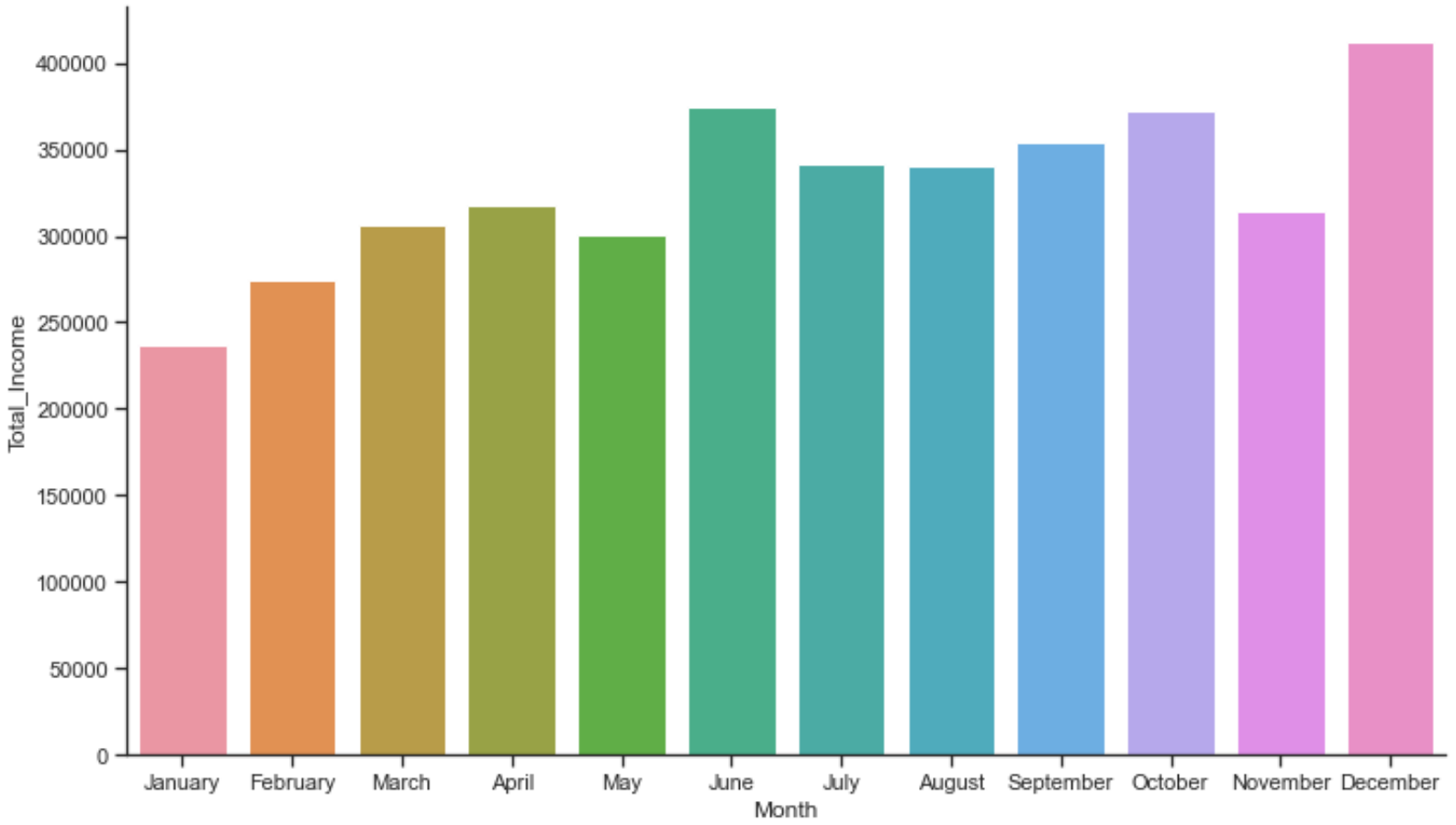
The most popular stores in Le Mars, Algona, and Newton city

In Algona and Newton city, there is a big store which dominates the liquor sales. Meanwhile, in Le Mars city, there are two stores which lead the liquor sales.

Additional Insight

There is some insight from the dataset which can be useful for the vendor.

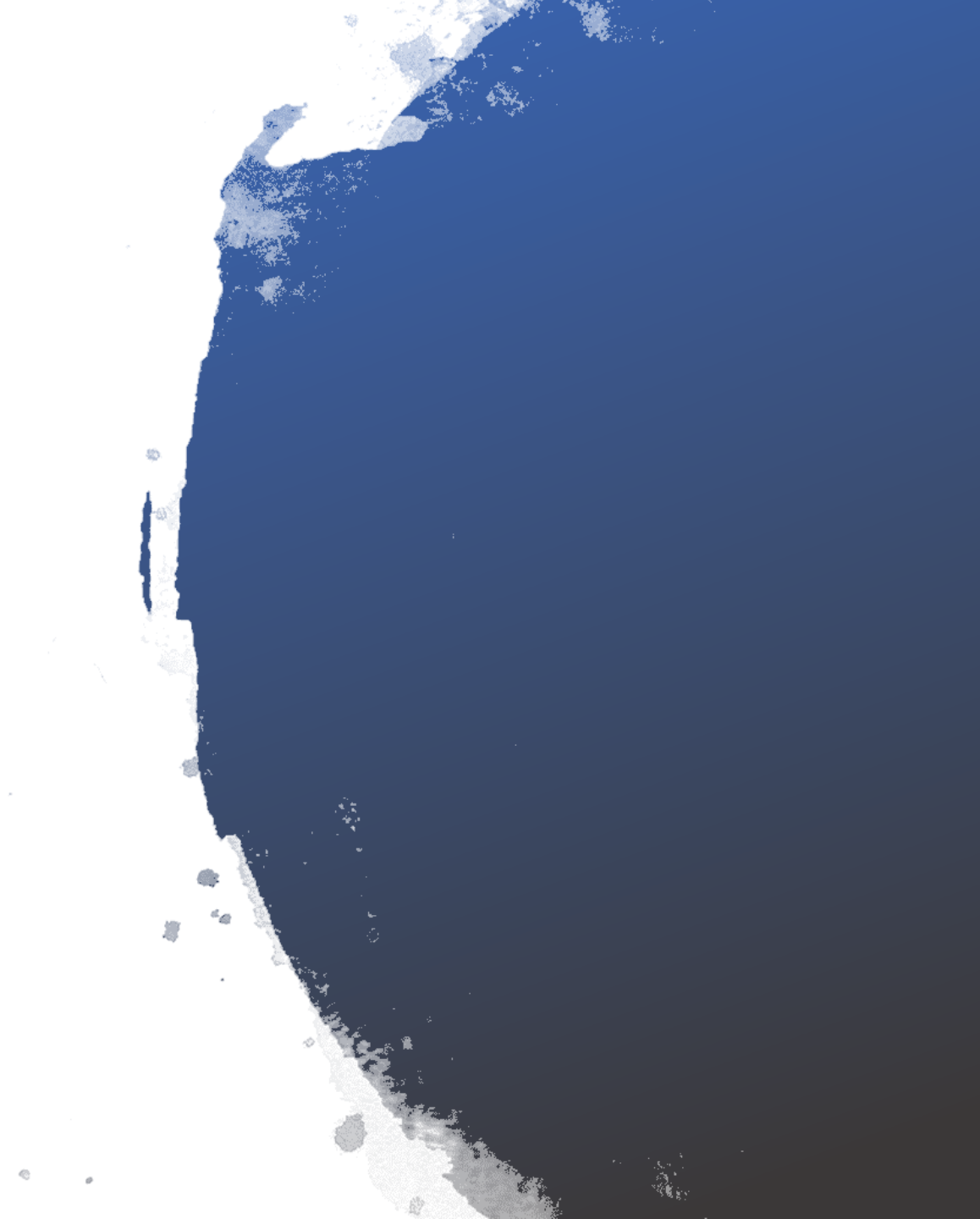
The first insight is about the monthly income from liquor sales in the state of Iowa in 2020.

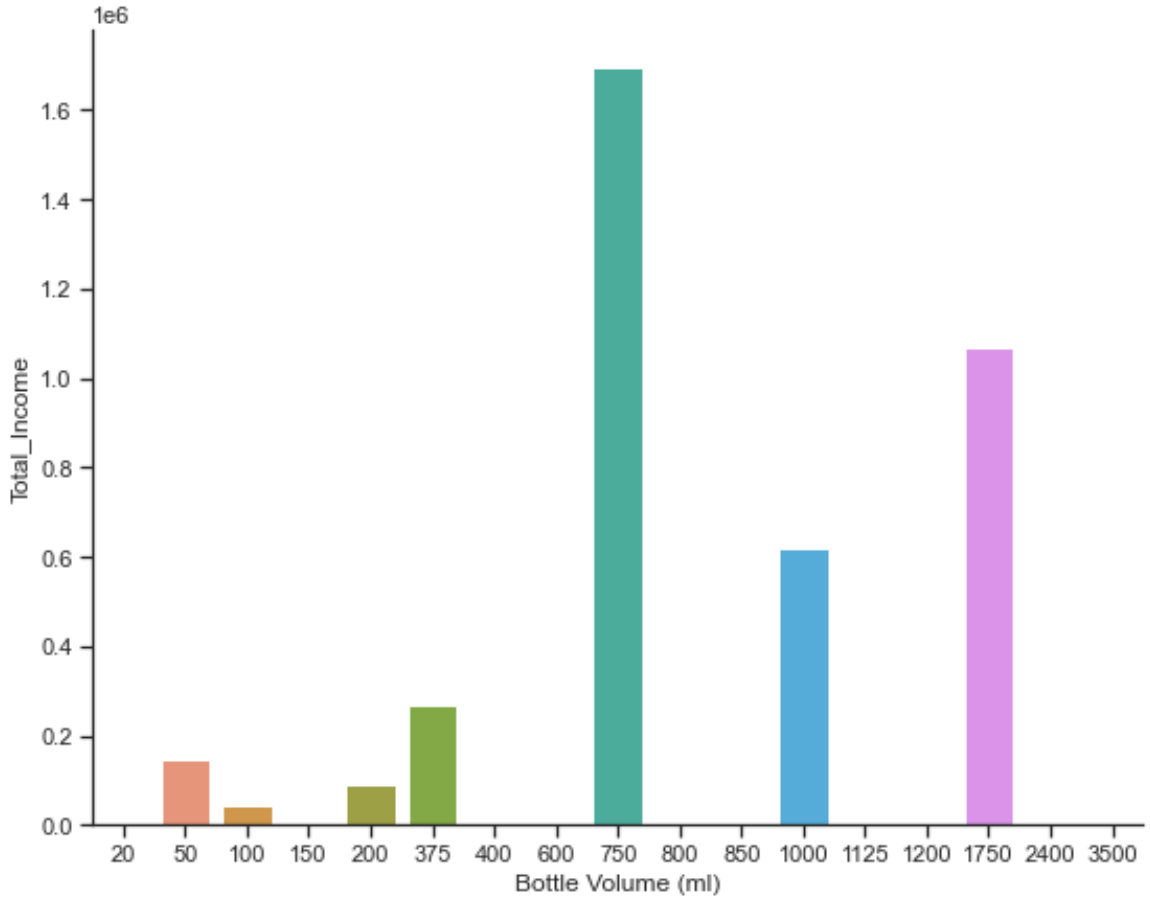


Monthly Income in The State of Iowa

From the monthly income graphic, it seems that there is a pattern which is repeated every six months. The monthly income increases from the first month until the fourth month, significant decrease in the fifth month, and significant increase in the sixth month.

The second insight is about the volume of liquor bottles that people in the state of Iowa like the most. The bottle size or volume is one of the things that affect consumers' decisions. If the bottle size is right for them, they will most likely will buy that product.





The Most Popular Bottle Volume in The State of Iowa

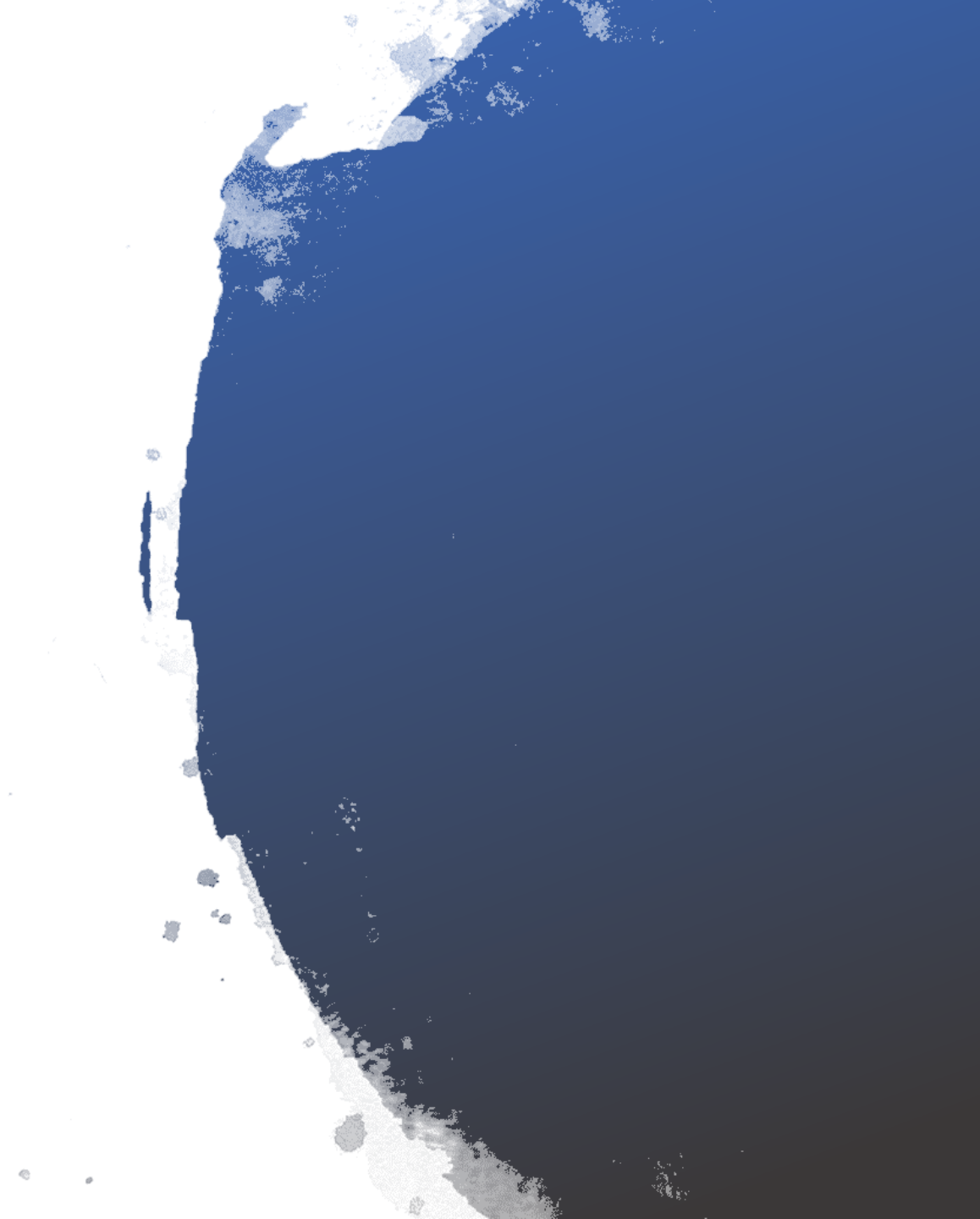
The graphic shows that the people in the state of Iowa most liked liquor bottles with 750 ml volume. The second most liked liquor bottle is the bottle with 1750 ml volume.

Conclusion

There is an almost perfect linear correlation between total income from liquor sales and total liquor volume sold.

The five most potential cities, with the biggest income growth percentage during Covid-19 Pandemic, are Le Mars, Newton, Algona, Grimes, Manchester.

In Le Mars and Manchester city, liquor category 1012100 is the most purchased liquor. Meanwhile, in Newton, Algona, and Grimes city, liquor category 1011200 is the most purchased liquor.



In Algona and Newton city, there is a big store which dominates the liquor sales. Meanwhile, in Le Mars city, there are two stores which lead the liquor sales.

It seems that there is a pattern which is repeated every six months in 2020 monthly income from liquor sales in the state of Iowa.

The people in the state of Iowa most like liquor bottles with 750 ml volume.


Thank You!