

# Case Study: Analyzing Iowa Liquor Sales Data

Isit Pokharel

01/08/2024

## 1. What impact did Covid have on the overall liquor market in Iowa?

At the onset of the pandemic, the consumption of alcohol in Iowa saw a notable increase, with mean spending on liquor rising from \$28.33 million per month before the World Health Organization (WHO) officially declared the pandemic on March 11, 2020, to \$35.34 million per month afterward

Refer to Figure 1.1 for more data.

### a. What trends evolved over the next 3-18 months?

A clear seasonal pattern is present in monthly liquor sales in Iowa, with January exhibiting the lowest consumption and October through December seeing peak sales. Beyond this, there was a pre-existing upward trend in alcohol consumption between 2018 and end of 2019. However, the pandemic's onset marked a pivotal turning point. Figure 1.2 displays the monthly alcohol sales time series alongside Non-Seasonal trend component. Notably, it reveals a sharp uptick in alcohol sales trend coinciding with the pandemic's onset, which has persisted till end of 2021. This demonstrates a substantial post-pandemic surge in alcohol consumption even after considering the seasonality of alcohol sales time series.

### b. Was there a notable shift in the types of products purchased in terms of pack size?

In 2019, as illustrated in Figure 1.3, the most significant growth in alcohol categories, as measured by the number of bottles in a case or pack size, was observed in liquors packaged in 48-packs and 10-packs, with annual percentage growth rates of 26.8% and 25.8% respectively. Notably, within the 48-pack category, alcohol products primarily consisted of bottles sizes of 100 ml or 200 ml, indicating a substantial increase in the demand for smaller-sized alcohol products during that period. In the case of 10-packs, a closer examination revealed that a majority of the growth was attributed to purchases of 50 ml bottles. In 2018, 50 ml bottles represented only a minor portion, accounting for just 1.4% of purchases within the 10-pack category. However, by 2019, the purchase of 50 ml bottles surged significantly, constituting a substantial 43.9% of total purchases within the same category. This striking shift underscores a noteworthy change in consumer preferences, with a pronounced inclination towards smaller bottle sizes within specific pack categories from 2018 to 2019.

However, In 2020, after the onset of pandemic, the most significant growth was observed in alcohol products packaged in 6-pack as illustrated in Figure 1.3. Nearly 90% of purchases within the 6-pack category consisted of bottles of sizes 750 ml and 1750 ml in 2019 and 2020<sup>1</sup>. The 26% growth in alcohol products packaged in 6-pack category from 2019 to 2020 suggests a considerable increase in the consumption of 750 ml and 1750 ml bottle sizes. This shift in purchasing behavior may be attributed to the impact of the COVID-19 pandemic, as larger pack bottle sizes became more favorable due to lockdown measures, reducing the need for frequent visits to liquor stores.

In 2021, there was a noteworthy surge (7.81% increase) in sales within the alcohol products packaged in 12-pack, these alcohol products predominantly consisted of bottles of sizes 750 ml and 1750 ml. This growth pattern in 12-pack liquor bottles mirrors the trend seen during the 2019-2020 period when a significant increase was observed in the 6-pack category. This pattern underscores a consistent consumer preference for larger-sized alcohol bottles during pandemic.

## 2. Which are the fastest growing types of liquor (e.g., vodka, tequila, rum, etc.)? How has market share changed over time?

To identify the fastest-growing liquor types and track changes in market share over time, I categorized the unique liquor categories into the following broader groups: *Whiskey, Vodka, Liqueur, Rum, Tequila, Cocktail, Liqueur, Scotch, Specialty, Grain Spirits, Unspecified, Distilleries* <sup>2</sup>.

**Market Share Rankings:** Examining monthly market share trends since the pandemic's onset, we find a consistent top five: *Whiskey, Vodka, Liqueur, Rum, and Tequila*, respectively. While there was a brief moment when Rum temporarily surpassed Liqueur to claim the third position, the overall ranking has remained steady throughout the pandemic. Refer to Figure 1.4 for a visual representation.

**Sales Growth Analysis:** Tequila stands out as a remarkable performer in terms of sales growth. Since 2018, Tequila has shown consistent year-over-year growth, with each growth rate surpassing the previous one. Specifically, Tequila recorded impressive growth rates of 9.5%, 22.8%, and 29.7% from 2018-2019, 2019-2020, and 2020-2021, respectively.

Conversely, Brandy exhibited growth rates of 7.1% and 28.9% between 2018-2019 and 2019-2020. However, its sales experienced a decline of 4.1% between 2020 and 2021.

In conclusion, Tequila emerges as the clear frontrunner among the fastest-growing liquor type. Refer to Figure 1.5 for a detailed visual representation.

### a. Write a function that takes a list of liquor types as an input and visualizes the market share over time for each of those.

Please refer to the Jupyter notebook file provided along with this report.

<sup>1</sup>Details on the breakdown of pack sizes by bottle volumes are available in the Jupyter notebook.

<sup>2</sup>Details on the exact mappings of Alcohol Category given in dataset to broader category defined here are available in Jupyter Notebook.

b. **What is driving the growth in tequila sales? Increases in average price or increases in volume sold?**

Tequila sales growth can be attributed to a combination of factors, with a significant role played by the increase in volume sold from 2018 to 2020, and price playing a comparable role from 2020 to 2021.

Year-over-Year, Tequila has consistently shown robust sales volume growth, with growth rates of 7.2%, 13.7%, and 18.7% from 2018 - 2019, 2019 - 2020, and 2020 - 2021, respectively and average price of Tequila has also experienced gradual growth, with rates of 2.1%, 2.3%, and 10.3% during respective periods. While sales revenue growth from 2018 - 2020 was primarily driven by volume increases, the period from 2020 - 2021 witnessed a notable contribution from price increases. Refer to Figure 1.6 for a greater detail.

3. **Grouping individual store brands together (e.g., all of Walmart, Liquor Barn, Hy-Vee, etc.), who are the top 10 retailers by year?**

Retailer Rank	2018	2019	2020	2021
1	HY-VEE	HY-VEE	HY-VEE	HY-VEE
2	LICKETY LIQUOR	LICKETY LIQUOR	LICKETY LIQUOR	LICKETY LIQUOR
3	FAREWAY	FAREWAY	FAREWAY	FAREWAY
4	WALMART	WALMART	WALMART	SAM'S CLUB
5	SAM'S CLUB	SAM'S CLUB	SAM'S CLUB	CASEY'S
6	CASEY'S	CASEY'S	CASEY'S	WALMART
7	KUM & GO	KUM & GO	COSTCO	COSTCO
8	COSTCO	COSTCO	KUM & GO	KUM & GO
9	WALGREENS	WALGREENS	TARGET	TARGET
10	TARGET	TARGET	WALGREENS	I-80 LIQUOR

3

4. **In late 2019 Heaven Hill Brands bought a portfolio of liquor brands from Constellation Brands. What percentage of Heaven Hill's growth in 2020 can be attributed to the acquisition?**

Heaven Hill Brands' purchase of a portfolio of liquor brands consisted solely of *Black Velvet* Canadian Whiskey. To calculate the sales of Heaven Hills Brands attributed to this acquisition, I isolated the sales data for Canadian Whiskey by filtering the rows that contained 'Black Velvet' in the *Item Description* column. Additionally, to determine the total sales attributed to Heaven Hill Brands, I extracted all the rows from the dataset that corresponded to the brands listed on the official Heaven Hill Distillery *website*. By comparing the sales figures before and after the acquisition, we can accurately calculate the percentage of Heaven Hill's growth in 2020 that can be attributed to this strategic acquisition.

Date	Black Velvet Total Sales (Mil. \$)	Heaven Hills Total Sales (Mil. \$)
2018	16.114	1.869
2019	16.222	2.136
2020	17.259	2.976
2021	16.308	3.332

I used the following formula to calculate the growth in 2020 that can be attributed to the acquisition.

$$\text{Percentage Growth from Acquisition} = \left( \frac{\text{Revenue from Acquisition in 2020}}{\text{Total Revenue Growth in 2020}} \right) \times 100$$

Based on the data provided in table, and the formula described above, 85.29% percentage of Heaven Hill's growth in 2020 can be attributed to the acquisition.

5. **What data integrity issues did you discover? How could you (or how did you) solve/account for these?**

a. **Comment on any data issues you discovered and what assumptions you used to deal with them.**

**Duplicates:** There were no duplicates in dataset which i verified using *Invoice/Item Number* column.

**Missing Values:** There were some columns with missing values but none of the columns necessary for analysis done in this case study had more than 0.7% of the total values missing and were dropped on case by case basis instead of blanket removing all missing rows at one. A decision was made to drop them because these were not numeric values, if numeric values were missing there are ways to impute them instead of dropping them.

<sup>3</sup>Details on the how these figures were calculated are available in Jupyter Notebook.

**Column Types:** Columns were also adjusted. Especially, columns that were of type numeric but where no arithmetic operations could be done were converted to str format, for example packs, zip codes, Store number etc. Date was changed to datetime object.

**Numeric Data Accuracy:** To ensure the accuracy of numeric columns, I used the following formulas to check data:

- Sale (Dollar) was validated using the formula  $\text{Bottles Sold} \times \text{State Bottle Retail}$ .
- Volume Sold (Liters) was checked with the formula:  $(\text{Bottle Volume (ml)} \times \text{Bottles Sold}) / 1,000$ .
- Volume Sold (Gallons) was checked using the formula:  $(\text{Volume Sold (Liters)}) / 3.7854117$ .

I removed rows where Sale (Dollars) didn't match the expected sales amount, even after considering a 20% discount. No discrepancies were found in Volume Sold (Liters) and Volume Sold (Gallons). These validation steps ensured numeric column accuracy for confident analysis.

b. **What's a simple solution to solving data quality issues?**

When dealing with time series data, especially, using formulas to verify sub-aggregates against aggregates or employing outlier tests to identify sudden spikes or drops can quickly reveal data quality issues. Additional time series tests may involve detecting repeating values within a series where such repetitions are unexpected. For example, in a time series representing population data, discovering that the population in 2018 matches that of 2019 would prompt further investigation. Some time series, like policy rates or fed funds rates, are particularly sensitive to data changes. Implementing specialized tests to scrutinize value changes in these series is another effective way to assess data quality. Lastly, performing null value checks is a fundamental step in testing missing values in time series.

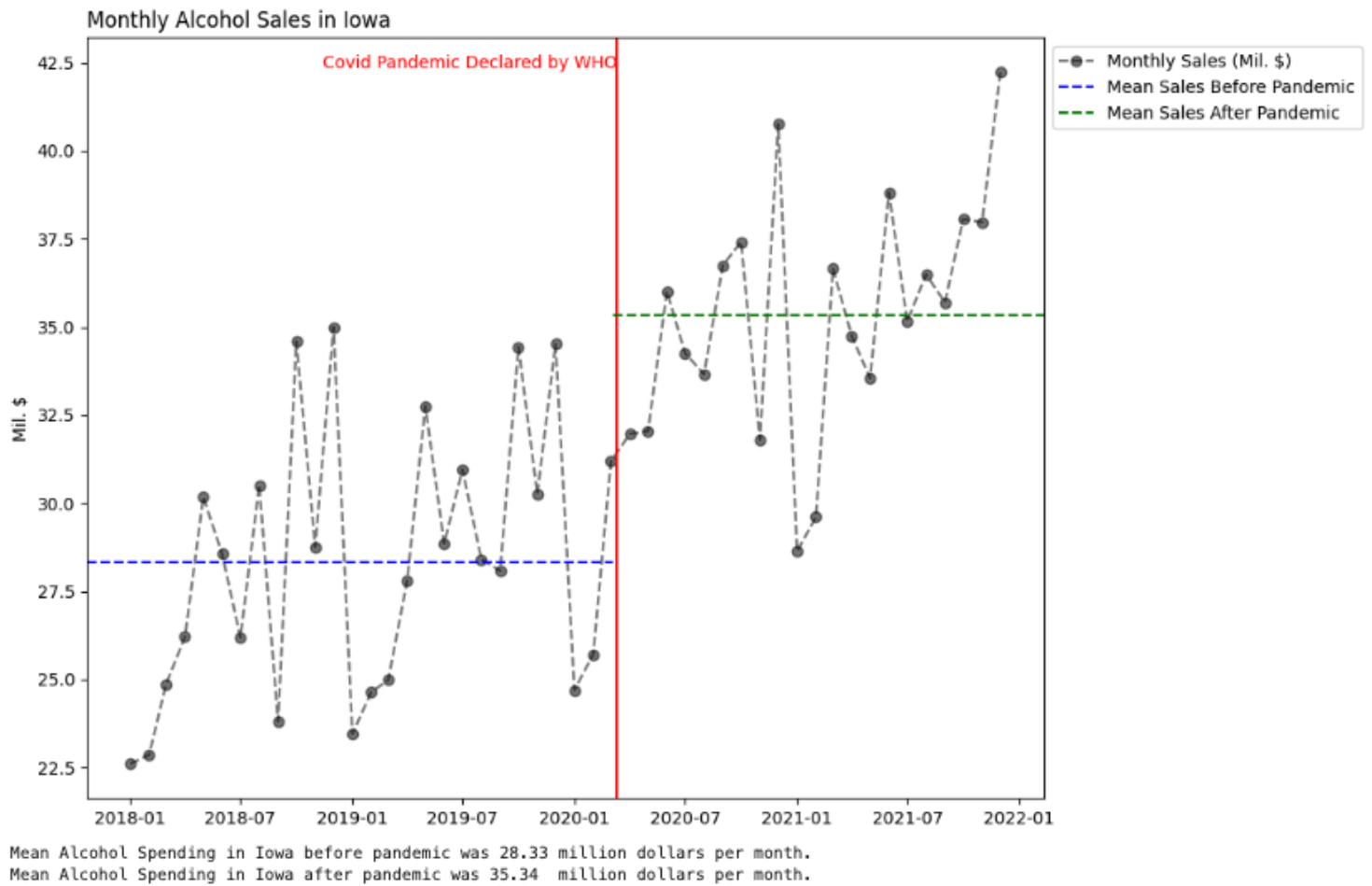
c. **What would be a more scalable solution to deal with quality issues?**

Modern data technology stacks, such as DBT, come with integrated testing features when constructing data transformation layers. Utilizing DBT tests to apply some of the straightforward solutions mentioned earlier can effectively identify erroneous data before it gets incorporated into production tables or is presented to clients.

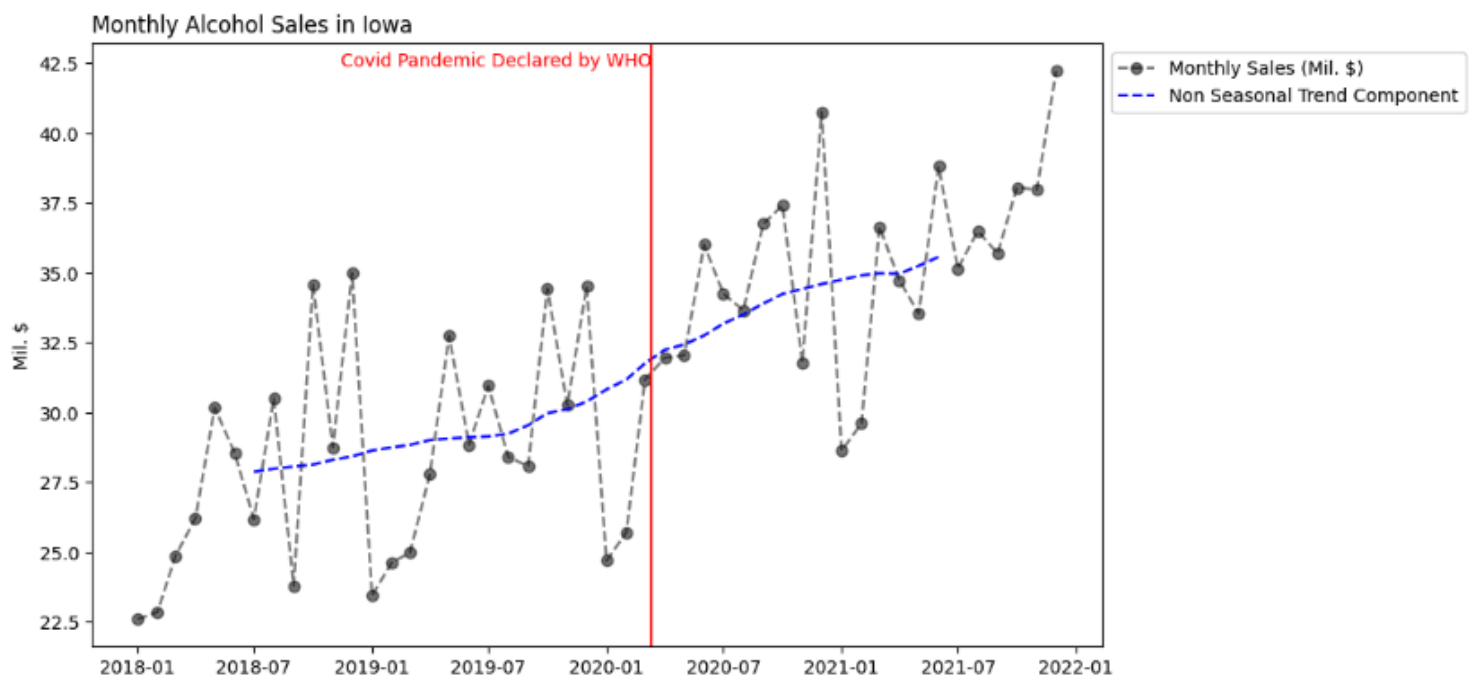
Within DBT, you have the capability to designate specific time series as "not null" or assess if categorical values belong to predefined categories. Establishing range limits for particular series and defining connections between multiple time series using predefined formulas can offer a scalable approach for handling data quality concerns.

# 1 Appendix A: Additional Charts

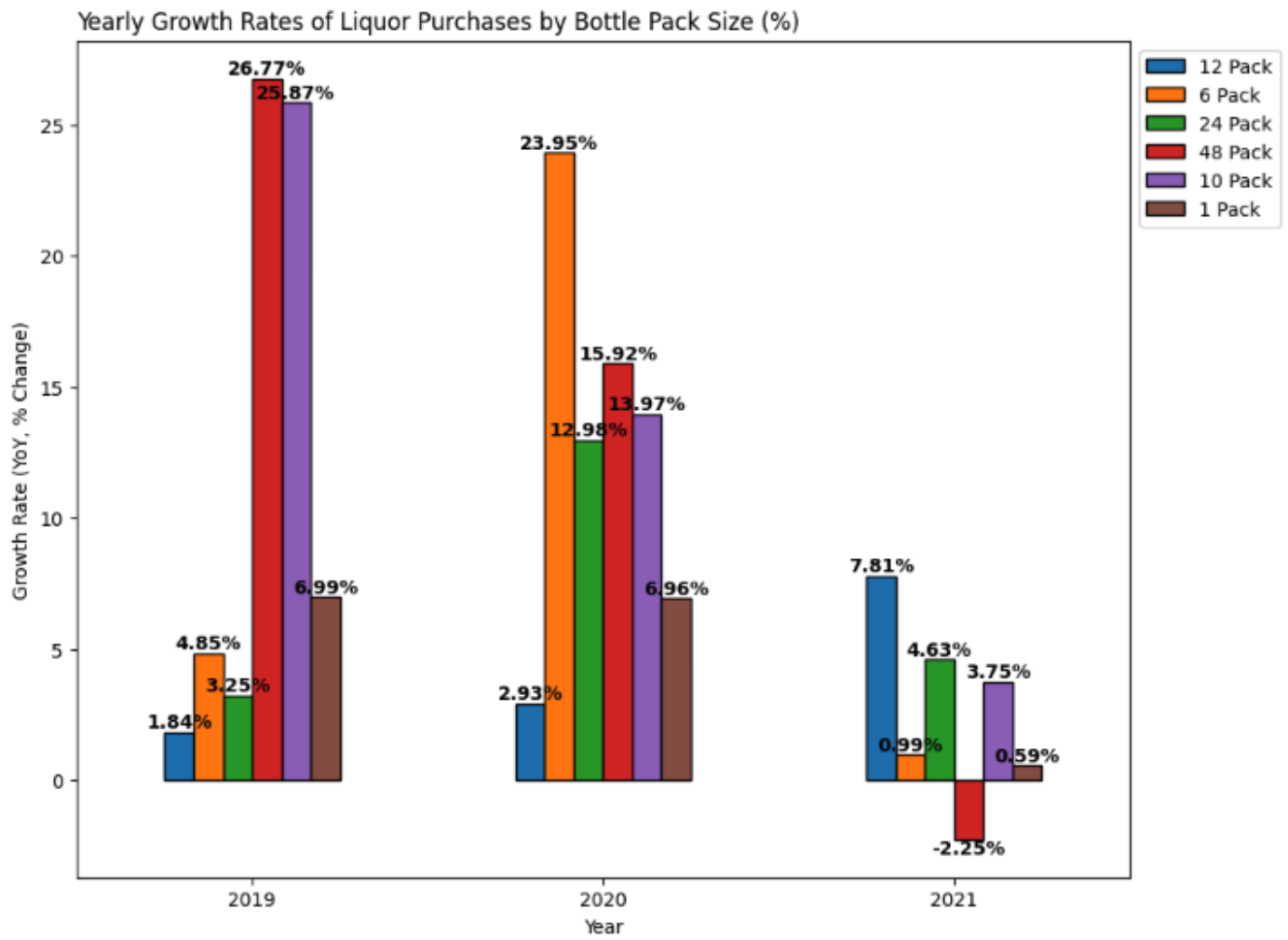
## 1.1 Monthly Alcohol Sales: Before and After Pandemic



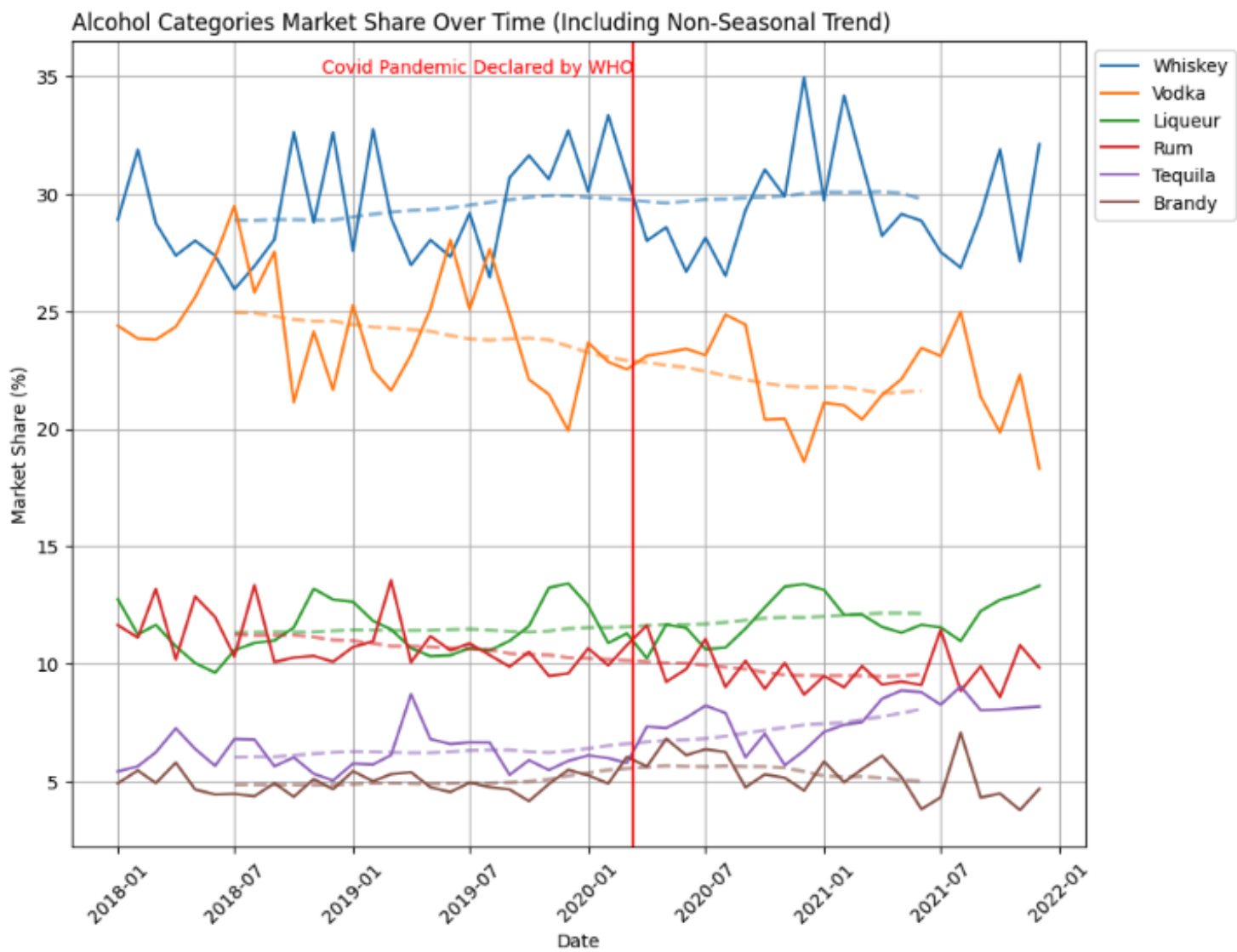
## 1.2 Non-Seasonal Trend in Alcohol Sales in Iowa



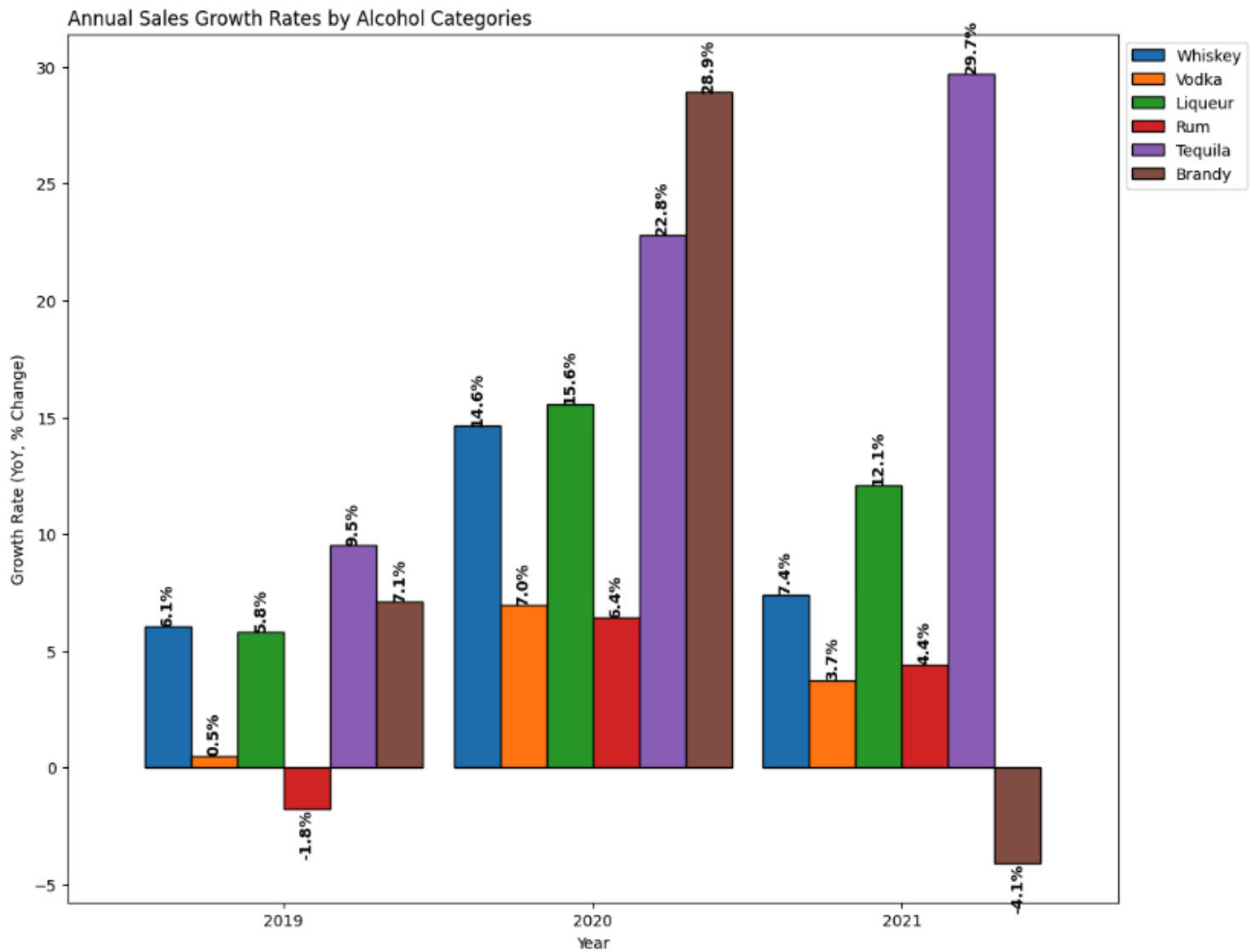
### 1.3 Growth Rate of Liquor Purchased by Pack Size



1.4 Market Share by Alcohol Categories



## 1.5 Year Over Year Sales Growth by Alcohol Categories



## 1.6 Tequilla's Average Retail Price and Average Volume Sold Growth, YoY

	Date	Total Sales (Mil. \$)	Volume Sold (Thous. Liters)	Average Price (\$)	Sales Growth(%)	Price Growth(%)	Volume Growth(%)
0	2018	19.929	866.306	21.790	NaN	NaN	NaN
1	2019	21.784	928.461	22.245	9.310	2.092	7.175
2	2020	26.682	1055.468	22.757	22.482	2.298	13.679
3	2021	34.538	1253.498	25.109	29.443	10.340	18.762