

Data Management and Visual Analytics Final Report

Group 1: Ben Ahnen, Michael Gerot, Ethan Hahn, Blake McClung, Dylan Sambrano

Professor Weigo Fan

December 4th, 2020

Analyzing Iowa's Liquor Retail Environment

Business Problem and Key Questions

With the onset of the COVID-19 pandemic, the liquor retail environment in the U.S. has changed drastically. As people feel less comfortable visiting crowded bars and restaurants to consume liquor, they have largely turned to stores and online deliveries. According to a study by Nielsen, there has been a 38.6% increase in retail liquor sales compared to last year. The same study showed that there was a 441% increase in online liquor purchases, which was the fastest growing vertical in all economic sectors. Along with these trends, some states like Iowa have even legalized to-go alcohol sales in response to bars and dine-in restaurants having restrictions on the number of customers allowed in the establishment. The rapidly changing business environment for liquor sales presents an opportunity for entrepreneurs to serve customers in new ways. However, it is crucial for any incoming businesses to understand the current market dynamics and consumer preferences before entering a market.

For the final project, our team sought to analyze liquor retail sales data to obtain actionable business insights. To narrow the scope of our analysis, we focused specifically on the state of Iowa. We investigated the following key questions to understand the dynamics of Iowa's liquor retail market and consumer preferences:

- 1. What is the size and make-up of Iowa's current liquor retail market?*
- 2. Who are the largest liquor vendors in the state by sales?*
- 3. Which retailers have the strongest market position?*
- 4. Which liquor products are the most popular?*
- 5. Which regions purchase the most liquor?*

Dataset Description

We utilized data collected by the Iowa Alcoholic Beverages Division (data.iowa.gov). This dataset includes vendor invoice data for all transactions of class E Liquor Licenses in the State of Iowa from January 1, 2019 to September 30th, 2020. Class E Liquor Licenses do not include bars or restaurants, only stores that sell alcoholic beverages. The dataset included the location of the sale, the stores and vendors involved in the transaction, product descriptions and categories, the sale quantity, containers per package, and volume of containers.

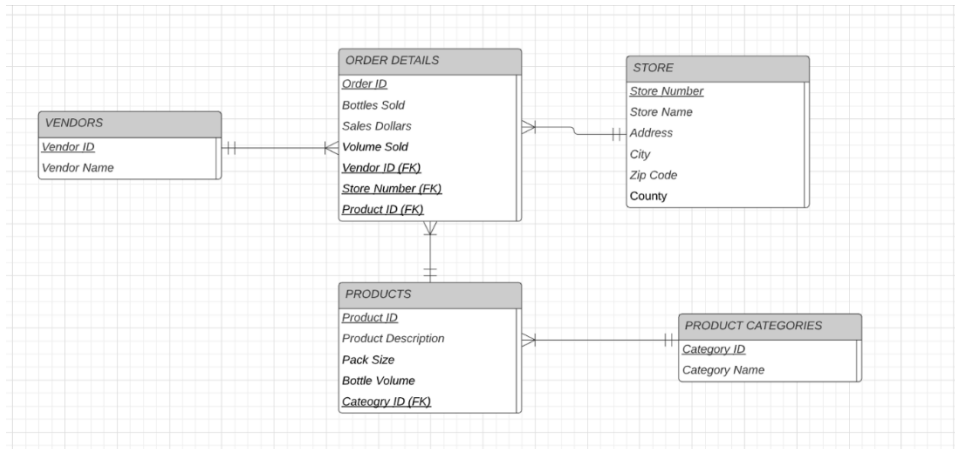
Field Name	Data Type	Description
Vendor ID	NUMBER	ID of Vendor
Vendor Name	VARCHAR2(255)	Name of Vendor
Store Number	NUMBER	Unique Store Number
Store Name	VARCHAR2(255)	Name of Store
Address	VARCHAR2(255)	Address of Store
City	VARCHAR2(50)	City of Store
Zip Code	NUMBER	Zip Code of Store
County	VARCHAR2(50)	County of Store
Product ID	NUMBER(25,0)	Unique ID of Product Sold
Product Description	VARCHAR2(250)	Description of Product
Pack	NUMBER(10,0)	Pack Size of Order
Bottle Volume	NUMBER(10,0)	Bottle Volume of Order
Category ID	VARCHAR2(50)	Unique Category Product
Category Name	VARCHAR2(255)	Name of Category
Order ID	NUMBER(25,0)	ID of the Order
Bottles Sold	NUMBER(25,0)	Number of Bottles Sold
Sales Dollars	NUMBER(25,0)	Amount of sales (\$)
Volume Sold	NUMBER(25,0)	Amount of Product Sold

Database Design and Methodology

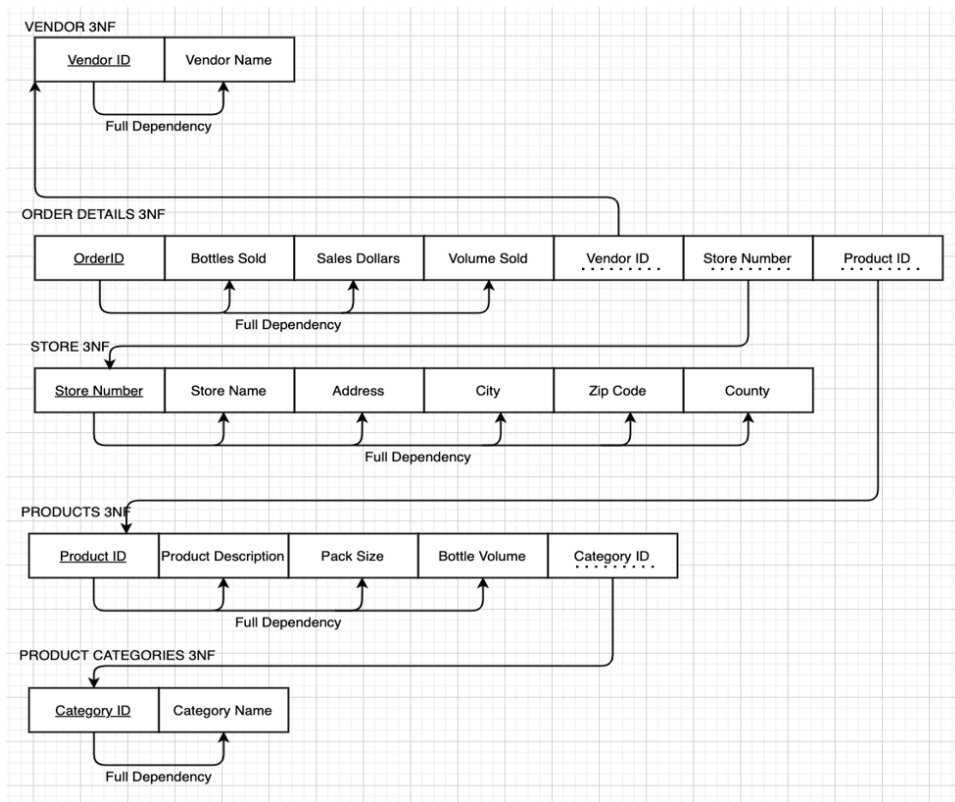
We started our database design by identifying 5 entities in the dataset: Vendors, Stores, Order Details, Products, and Product Categories. Each vendor is identified by a unique Vendor ID. Each store is identified by a unique Store Number. All Order Details are identified by a unique Order ID. Each Product is identified by a unique Product ID. Lastly, each Product Category is identified by a unique Category ID.

We determined the following business relationships from the dataset. Each Order Detail (invoice) has exactly one Vendor, Store and Product included. Additionally, each individual Vendor, Store, and Product must be included in at least one, or many different Order Detail instances. Each Product belongs to exactly one Product Category, and each Product Category must be associated with at least one, or many different Products.

The Entity Relationship Diagram with all five entities can be seen below:



Once our Entity Relationship Diagram was created and verified, we created our relational schema to show the overall logical structure of our database. After our initial ERD steps, our data was already in 3NF. We know this because we have no multivalued attributes, no partial dependencies, and no transitive dependencies. We have shown this by mapping out the dependencies, and as you can see, our data only contains full dependencies. The relational schema (with dependencies added) is shown below:



After normalizing our database structure and creating a logical model of our data, we began the process of creating a physical database in Oracle. We created separate CSV files in R for each of our entities.

After creating separate files for our 5 distinct entities, we began cleaning the data. We removed duplicates for all of our entities that contained redundant records. Additionally, we also identified a significant number of sales transaction records that would refer to vendors or stores that did not exist in their respective entities. For all these phantom records, we deleted the record to ensure referential integrity between our tables.

With our data cleaned, we exported the data in Oracle and created the necessary primary key and foreign key relationships between the tables based on the relevant business rules we established through our ERD and Relational Schema design. We then moved onto our analysis, where we queried the data in Oracle, and exported the 3NF Oracle database to Tableau to create visualizations.

Data Analysis in SQL

For our data analysis in SQL, we developed queries to address each of our key questions.

1. What is the size and make-up of Iowa's current liquor retail market?

To understand the overall liquor environment in Iowa, we queried the dataset for general summary statistics. Shown below we have three queries that sum the total sales dollars, number of bottles sold, and fluid volume (in gallons) of all transactions in the observed period. In a similar query, we summed the number of unique vendors, stores, products, and product categories in the dataset. The results illustrated the size of the Iowa liquor retail market, the vendor environment, competition in the market, and diversity of product offerings.

```
Select 'Total Sales' as Statistic, Sum(Sales_Dollars) as Memo
From Order_Details
UNION
Select 'Total Bottles' as Statistic, Sum(Bottles_Sold) as Memo
From Order_Details
UNION
Select 'Total Volume' as Statistic, Sum(Volume_Sold) as Memo
From Order_Details
```

	STATISTIC	MEMO
1	Total Bottles	6666937
2	Total Sales	86415304.75
3	Total Volume	5735350.44

2. Who are the largest liquor vendors in the state by sales?

To identify the largest vendors in the state, we created a query grouped by vendor name to count the number of transactions, and sum the sales dollar and bottles sold, then ordered it descending. To execute this query, we assigned aliases to the Vendor and Order Details tables, then joined them by the vendor ID. The results illustrated the key vendors in the market by both volume and total sales value. As a potential retail business owner, these vendors may be important suppliers to understand before entering the market.

```
Select Vendor_Name, Count(Distinct Order_ID) as Transactions,
```

Sum(Sales_Dollars), Sum(Bottles_Sold)

From Vendor v, Order_Details o

Where v.Vendor_ID = o.Vendor_ID

Group By Vendor_Name

Order by 2 DESC

	VENDOR_NAME	TRANSACTIONS	SUM(SALES_DOLLARS)	SUM(BOTTLES_SOLD)
1	DIAGEO AMERICAS	99355	18222305.14	987470
2	Jim Beam Brands	56048	6689675	467639
3	LUXCO INC	49188	4425061	555370
4	SAZERAC COMPANY INC	45407	7128217.75	855954
5	CONSTELLATION BRANDS INC	37573	5400068.53	545023
6	PERNOD RICARD USA	35734	6291062.43	356339
7	SAZERAC NORTH AMERICA	28870	2856771.35	373058
8	Heaven Hill Brands	26389	2327786.21	208693
9	BACARDI USA INC	24291	3331870.18	193839
10	PROXIMO	22553	2979472.78	199796

3. Which retailers have the strongest market position?

To evaluate the largest competitors in the retail market, we created a query grouped by store name to count the number of transactions, and sum the sales dollar and bottles sold, then ordered it descending. To execute this query, we assigned aliases to the Store and Order Details tables, then joined them by the store number attribute. The results illustrated the top end competitors in the market and revealed that many top stores fall under the same corporate parent company, such as Hy-Vee.

Select Store_Name, Count(Distinct Order_ID) as Transactions,

Sum(Sales_Dollars), Sum(Bottles_Sold)

From Store s, Order_Details o

Where s.Store_Number = o.Store_Number

Group By Store_Name

Order by 2 DESC

	STORE_NAME	TRANSACTIONS	SUM(SALES_DOLLARS)	SUM(BOTTLES_SOLD)
1	Hy-Vee #3 / BDI / Des Moines	5215	2746632.71	158900
2	Central City 2	4557	2986311.2	168587
3	Central City Liquor, Inc.	4121	507052.81	34777
4	Hy-Vee Wine and Spirits / Iowa City	3932	1434187.54	97659
5	Hy-Vee Wine and Spirits / Bettendorf	3857	517852.74	35831
6	Hy-Vee #4 / WDM	3388	562709.68	34512
7	Hy-Vee Food Store / Coralville	3345	688682.86	42645
8	Hy-Vee Food Store #1 / Ames	3265	413381.11	28276
9	Benz Distributing	3116	893588.35	49969
10	Hy-Vee Food Store / Cedar Falls	3008	268773.61	17117

To drill down further into our store data, we executed a similar query of sales volume and dollar values, but we added a condition of store names that included “Hy-Vee”. Our team performed this query for other large retailers such as Target and Wal-Mart as well. These results helped us to identify the key branches of major competitors in the Iowa market.

Select Store_Name, Count(Distinct Order_ID) as Transactions,

Sum(Sales_Dollars), Sum(Bottles_Sold)

From Store s, Order_Details o

Where s.Store_Number = o.Store_Number

AND Store_Name LIKE '%Hy-Vee%'

Group By Store_Name

Order by 2 DESC

	STORE_NAME	TRANSACTIONS	SUM(SALES_DOLLARS)	SUM(BOTTLES_SOLD)
1	Hy-Vee #3 / BDI / Des Moines	5215	2746632.71	158900
2	Hy-Vee Wine and Spirits / Iowa City	3932	1434187.54	97659
3	Hy-Vee Wine and Spirits / Bettendorf	3857	517852.74	35831
4	Hy-Vee #4 / WDM	3388	562709.68	34512
5	Hy-Vee Food Store / Coralville	3345	688682.86	42645
6	Hy-Vee Food Store #1 / Ames	3265	413381.11	28276
7	Hy-Vee Food Store / Cedar Falls	3008	268773.61	17117
8	Hy-Vee Wine & Spirits #2 / Davenport	2987	699754.59	42688
9	Hy-Vee Food Store / Muscatine	2844	256456.35	16783
10	Hy-Vee #1044 / Burlington	2841	393591.08	26122

4. Which liquor products are the most popular?

Next, we moved onto understanding consumer preferences for liquor in Iowa. We created a query grouped by product description to count the number of transactions, and sum the sales dollar and bottles sold, then ordered it descending. To execute this query, we assigned aliases to the Product and Order Details tables, then joined them by the product ID attribute. The results showed our

team which products were the most popular from a volume and total sales perspective. An incoming retailer would certainly want to carry these top products, and would need to identify key vendors from which these products could be purchased.

```
Select Product_Description, Count(Distinct Order_ID)
as Transactions, Sum(Sales_Dollars), Sum(Bottles_Sold)
From Order_Details o, Product p
Where p.Product_ID = o.Product_ID
Group by Product_Description
Order By 3 DESC
```

PRODUCT_DESCRIPTION	TRANSACTIONS	SUM(SALES_DOLLARS)	SUM(BOTTLES_SOLD)
1 Titos Handmade Vodka	11825	4429026.84	230591
2 Black Velvet	17304	3338496.88	322708
3 Crown Royal	8326	2806258.23	111034
4 Jack Daniels Old #7 Black Label	7602	2644939.44	103632
5 Captain Morgan Spiced Rum	7603	2537329.68	148090
6 Hennessy VS	7035	1819947.47	101762
7 Jameson	4199	1664348.68	61673
8 Crown Royal Regal Apple	6563	1635439.07	75814
9 Hawkeye Vodka	15036	1548721.21	236705
10 Fireball Cinnamon Whiskey	6490	1485404.79	109458

To further analyze consumer preferences, our team created an additional query to calculate total sales for product categories with names that included popular styles of alcohol, such as tequila. To execute this query, we had to join the Order Details and Product tables by product ID, and the Product and Product Categories tables by category ID. We made similar queries for other popular types of alcohol, and the results illustrated the key varieties of the most popular items.

```
Select Category_Name, sum(Sales_Dollars)
From Order_Details o, Product p, Product_Categories c
Where p.Product_ID = o.Product_ID
AND p.Category_ID = c.Category_ID
AND Category_Name like '%Tequila%'
Group By Category_Name
```

CATEGORY_NAME	SUM(SALES_DOLLARS)
1 100% Agave Tequila	2844599.04
2 Mixto Tequila	2043188.71

5. Which regions purchase the most liquor?

Lastly, we sought to identify the regions of Iowa that have the greatest demand for liquor. We created a query grouped by city to count the number of transactions, and sum the sales dollar and bottles sold, then ordered it descending. To execute this query, we assigned aliases to the Store and Order Details tables, then joined them by the store number attribute. The results illustrated the cities in Iowa that purchase the most liquor by volume and total sales. One could interpret these cities as opportunities to tap into large existing market, or as saturated markets to avoid. Further analysis is needed to know for sure.

```
Select City, Count(Distinct Order_ID) as Transactions,
Sum(Sales_Dollars), Sum(Bottles_Sold)
From Store s, Order_Details o
Where s.Store_Number = o.Store_Number
Group By City
Order by 2 DESC
```

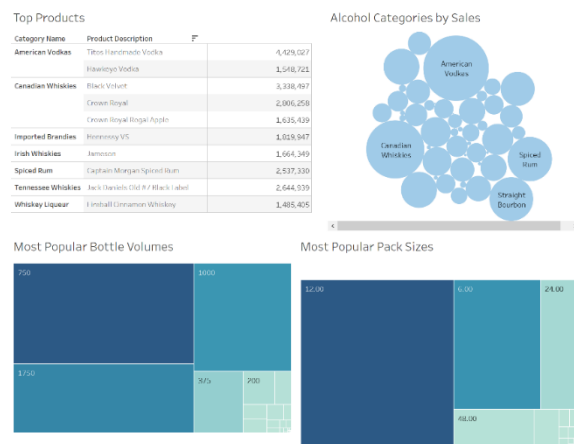
CITY	TRANSACTIONS	SUM(SALES_DOLLARS)	SUM(BOTTLES_SOLD)
1 Des Moines	52507	11084121.46	876327
2 Cedar Rapids	34497	4668046.68	396559
3 Davenport	29064	4982648.38	428860
4 Council Bluffs	20074	3184273.31	258012
5 Sioux City	19367	3074332.7	239022
6 Waterloo	18741	2885868.48	255747
7 West Des Moines	18486	3358565.07	218402
8 Ames	15428	2218574.42	151568
9 Iowa City	13966	2397592.04	178215
10 Dubuque	13960	2020829.54	152071

Findings and Visualizations

Visualization #1 – Products

After querying the product data, we created visualizations to depict the trends for products. We identified a few major findings from our product visualization. First, we identified that the alcohol product market in Iowa is dominated by significant large cap alcohol brand names, such as Tito's, Captain Morgan, and Fireball. However, while there are large brands that dominate a large portion of sales, there are still quite a few alcohol categories that are significant in sales dollars but do not have a major brand name, such as straight bourbon.

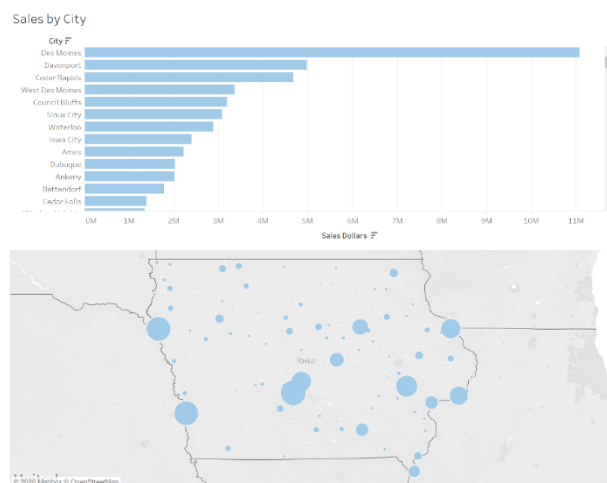
Additionally, we also identified unsurprising trends for both bottle volumes and pack sizes. Bottle volume was led by 750 mL products, which is the most common liquor volume, and pack size was dominated significantly by 12 packs.



Visualization #2 – Geography

One of the major questions that we had going into our project was how alcohol sales was decomposed by city/county within the state of Iowa. As you can see by our visualization below, we, unsurprisingly, found that alcohol sales have the most density in the most highly populated and urban areas of the state, such as the Des Moines metropolitan area, Cedar Rapids, and Council Bluffs.

However, one finding that surprised us was the density of alcohol sales within the major college towns in Iowa. Iowa City and Ames are positioned as 8th and 9th highest, respectively, in liquor sales within the state. However, we rationalized this conclusion was attributable to the fact that our data was limited to all Class E liquor licenses, which would not include the bars.

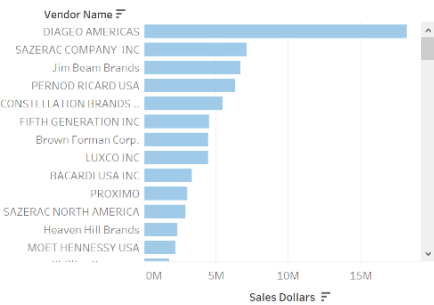


Visualization #3 – Vendors

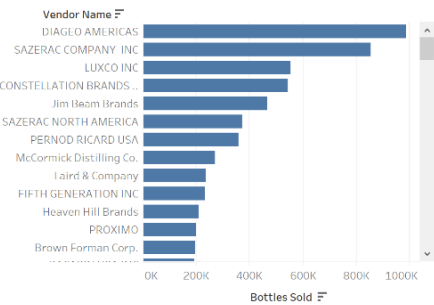
The vendor category was the category that we were most uncertain about prior to querying the data due to our unfamiliarity with the alcohol supply chain landscape within the state of Iowa. However, we did make some interesting conclusions about vendors from our visualizations. First, we noticed that Diageo, an U.K. liquor supplier, is the largest distributor of liquor in the state of Iowa in both terms of sales dollars and volumes. However, after Diageo, there is significant fragmentation in the liquor market from a vendor perspective. Additionally, we noticed that some vendors compete with liquor in terms of price and other in terms of volumes, which identifies the relevance of competitive advantage in Iowa’s liquor markets.

Lastly, we were also curious about how vendors related to a store’s and product’s success. We found that while some vendors have larger market share in terms of dollars and bottles sold, there is very little correlation between the number of stores serviced nor products offered a vendor’s market share. Diageo, which has 3x as many sales dollars in liquor as its next closest competitor, distributes to a similar number of stores and distributes a similar number of products to its competitors. We concluded that the vendor landscape in Iowa for liquor is very complex and interesting.

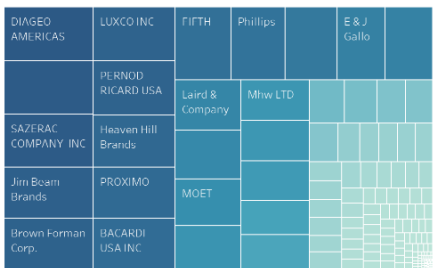
Sales Dollars by Vendor



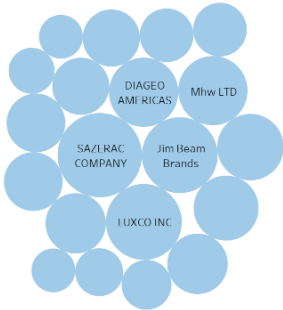
Bottles Sold by Vendor



Stores Serviced by Vendors



Products Supplied by Vendors

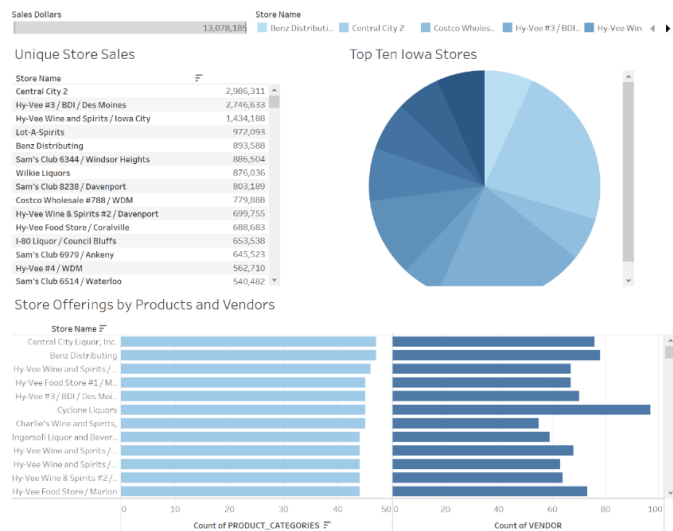


Visualization #4 – Stores

After identifying trends for vendors, we were curious about what the store landscape looked like for liquor in the state of Iowa. Our first finding was similar to the vendor visualization in that there are a select group of stores that dominate alcohol sales within the state of Iowa.

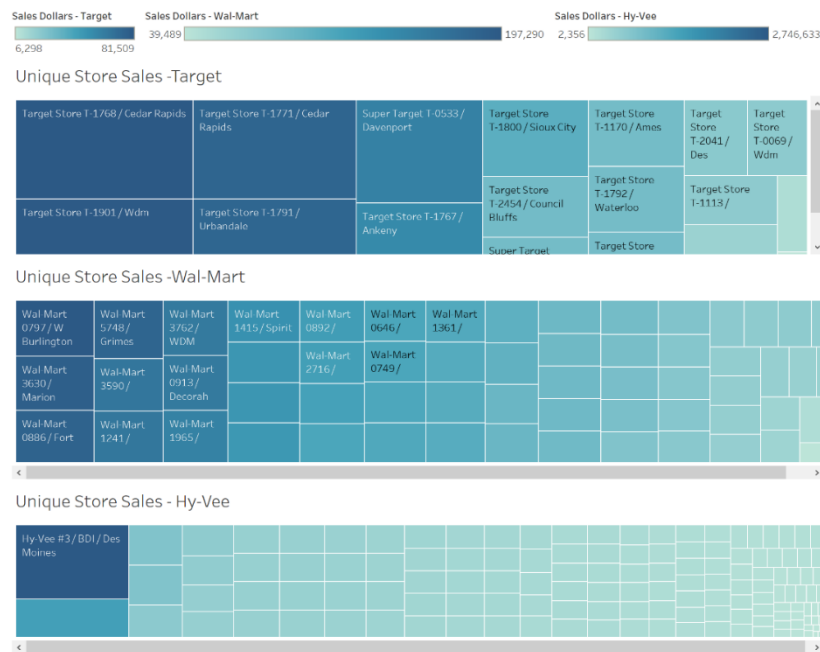
However, after the top 3 stores, which make up about ~40% of sales, there is significant fragmentation within the market.

Additionally, we were also analyzed how a store's success was related to the number of products it offers and vendors it utilizes. We concluded that there is a direct relationship between the number of products that a store offers and its total sales, so it is clear that customers appreciate and prioritize product variety. However, on the contrary, we found that there is no relationship/association between the number of vendors that a store uses. A store that uses 50 vendors for its supply chain can be as successful as a store that uses 100 vendors.



Visualization #5 – Corporations

One major finding that we had when analyzing stores was that several of the major stores were all organized by a few major corporations, such as Hy-Vee, Wal-Mart, and Target. We also identified other gas stations such as Kum-n-Go and Casey's; however, their total sales dwarfed in comparison to the major wholesale retailers. In analyzing these three major corporations shown below, one major finding that we made was these corporations have significant sales in the major locations that we talked about in our geography visualization, such as Des Moines and Cedar Rapids. However, outside of those major locations, the total sales for these corporations is not as significant as one would expect.



Key Takeaways

Based on the data our group gathered, cleaned, and analyzed we were able to come up with some key findings and takeaways. First, we found that organizations selling liquor in Iowa compete based on either pure volume or price. This is important because if you are planning on entering this market its crucial to identify your competitive strategy whether it be low pricing or high volume. Additionally, we found that there was a strong relationship between sales and the number of products offered. Therefore, by having more variation in your products you would expect higher sales than offering less variation in your products. However, we were unable to identify a strong relationship between sales and vendors. This means that using one vender over another doesn't lead to substantially more sales.

Currently urban areas like Des Moines tend to lead in sales for Iowa, and we found there were opportunities to enter a less concentrated market in more rural areas. We recommend that someone looking to enter this market avoid cities with heavy corporate influence. Thus, avoiding areas with Targets, Walmart's, and Hy-Vee's because it is very hard to compete for liquor sales with these large corporate stores. In conclusion, we recommend avoiding large cities as they are heavily concentrated markets with corporate influence and there are more opportunities in rural areas.

Challenges Faced and Future Considerations

Our analysis was primarily focused on understanding the dynamics and preferences of the liquor market in Iowa. Given more time, we could expand on our analysis to include time-series trends in total purchases and product preferences to get a timely view of the market. Additionally, with

further time series analysis, we could analyze the specific effects of the COVID-19 pandemic on liquor purchases for the state of Iowa and individual cities. Unfortunately, the large volume of the data, challenges of working with date attributes, and time constraints of the project, limited the scope of our investigation. In the future, we could dig deeper into this data to tell a variety of new stories and draw further actionable insights for entrepreneurs and other interested parties.