



Project Objective: Identify actionable business insights by formulating key questions to guide our analysis of datasets with enough size > 200MB.



Analysis Tools: We used Hadoop and Spark for data processing, while leveraging Tableau for dynamic visualization of our findings to illustrate key insights clearly.



Expected Outcomes: Aim to uncover valuable business insights and provide informed recommendations that will drive business growth and enhance operational efficiency through detailed data analysis.

Dataset Overview

- **Dataset Composition:** Our "*Liquor_sales*" dataset, sourced from [Kaggle](#), encompasses roughly *3 million* records across 24 columns, documenting detailed spirits purchases by Iowa Class "E" liquor licenses from January 2021 to January 2022.
- **Dataset Details:** Each record in the dataset provides comprehensive insights, capturing the date of purchase, product specifics like type, brand, and alcohol content, as well as financial data including cost and quantity purchased.
- **Analytical Applications:** The extensive data enables deep analysis of regional and temporal sales trends, aiding retailers in stock optimization and marketing, while providing policymakers data to gauge the economic dynamics of liquor sales in Iowa.

Dataset in Excel

category	category_name	vendor_name	item_number	item_description	pack	bottle_v e_r
1081600	Whiskey Liqueur	SAZERAC COMPANY INC	64870	Fireball Cinnamon Whiskey	48	
1081200	Cream Liqueurs	McCormick Distilling Co.	65200	Tequila Rose Liqueur	12	
1031100	American Vodkas	DIAGEO AMERICAS	38008	Smirnoff 80prf PET	6	
1031100	American Vodkas	SAZERAC NORTH AMERICA	36648	Caliber Vodka	12	
1012200	Scotch Whiskies	DIAGEO AMERICAS	4626	Buchanan Deluxe 12YR	12	
1032100	Imported Vodkas	CONSTELLATION BRANDS INC	34821	Svedka 80prf	6	
1032100	Imported Vodkas	PERNOD RICARD USA	34006	Absolut Swedish Vodka 80prf	12	
1031100	American Vodkas	LUXCO INC	36308	Hawkeye Vodka	6	
1012200	Scotch Whiskies	DIAGEO AMERICAS	5318	Johnnie Walker Double Black	6	
1091300	Neutral Grain Spirits Fla	OLE SMOKY DISTILLERY LLC	86739	Ole Smoky Apple Pie Moonshine	8	
1062500	Flavored Rum	BACARDI USA INC	43051	Bacardi Dragon Berry	12	
1081200	Cream Liqueurs	DIAGEO AMERICAS	68036	Baileys Original Irish Cream	12	
1031100	American Vodkas	FIFTH GENERATION INC	38178	Titos Handmade Vodka	6	
1081400	American Schnapps	Phillips Beverage	84617	Phillips Root Beer Schnapps	12	
1011400	Tennessee Whiskies	DIAGEO AMERICAS	26656	George Dickel #12	12	
1051100	American Brandies	LUXCO INC	55506	Paramount Cherry Brandy	12	
1081400	American Schnapps	SAZERAC NORTH AMERICA	84172	99 Bananas Mini	10	
1081400	American Schnapps	Jim Beam Brands	82867	Dekuyper Watermelon Pucker	12	
1012100	Canadian Whiskies	DIAGEO AMERICAS	10807	Crown Royal Regal Apple	12	
1011200	Straight Bourbon Whiski	Heaven Hill Brands	17956	Evan Williams Black	12	
1051100	American Brandies	E & J Gallo Winery	52595	E & J VS PET	12	
1081400	American Schnapps	Jim Beam Brands	82847	Dekuyper Luscious Peachtree	12	
1081400	American Schnapps	SAZERAC COMPANY INC	82957	Firewater Cinnamon Schnapps	12	
1082200	Imported Schnapps	SAZERAC COMPANY INC	69611	Dr McGillicuddys Apple Pie	12	
1031200	American Flavored Vodka	E & J Gallo Winery	39492	New Amsterdam Pink Whitney	12	
1011200	Straight Bourbon Whiski	Jim Beam Brands	19068	Jim Beam	6	
1022200	100% Agave Tequila	PROXIMO	87402	Jose Cuervo Especial Silver	48	
1021200	American Flavored Vodka	Heaven Hill Brands	10227	Deep Eddy Ruby Red Grapefruit	12	

Dataset Imported in Pyspark

```
>>> liquor_sales_df.printSchema()
root
|-- invoice_and_item_number: string (nullable = true)
|-- date: string (nullable = true)
|-- store_number: integer (nullable = true)
|-- store_name: string (nullable = true)
|-- category: double (nullable = true)
|-- category_name: string (nullable = true)
|-- vendor_name: string (nullable = true)
|-- item_number: integer (nullable = true)
|-- item_description: string (nullable = true)
|-- pack: integer (nullable = true)
|-- bottle_volume_ml: integer (nullable = true)
|-- state_bottle_cost: double (nullable = true)
|-- state_bottle_retail: double (nullable = true)
|-- bottles_sold: integer (nullable = true)
|-- sale_dollars: double (nullable = true)
|-- volume_sold_gallons: double (nullable = true)
|-- city: string (nullable = true)

>>>
```

1.Stores and Categories Performance: How do store order volumes and liquor product sales volumes correlate with total dollar sales at both the store and product levels?

2.Pricing Strategy: How does pricing affect the volume and profitability of liquor sales, and what are the optimal price points for various categories to enhance revenue?

3.Vendor and Category Profitability: Which vendors and product categories contribute most to the store's profitability, and how can this information improve vendor negotiations and inventory decisions?



Question 1: Stores and Categories Performance

Liquor Stores

```
... )
>>> top_stores.show(20, False)
```

store_name	formatted_total_sales_dollars	formatted_total_volume_sold	formatted_dollar_to_volume_ratio
Hy-Vee #3 / BDI / Des Moines	13,266,416.86	187,461.240	70.7689
Central City 2	12,901,621.86	181,086.020	71.2458
Hy-Vee Wine and Spirits / Iowa City	5,821,589.25	97,310.230	59.8250
Costco Wholesale #788 / WDM	5,000,957.79	89,863.030	55.6509
Benz Distributing	4,592,046.20	63,074.290	72.8038
Wilkie Liquors	4,278,719.74	71,816.990	59.5781
Sam's Club 8162 / Cedar Rapids	3,887,888.06	66,024.810	58.8853
I-80 Liquor / Council Bluffs	3,680,217.43	48,315.680	76.1703
Sam's Club 6344 / Windsor Heights	3,546,451.17	60,694.570	58.4311
Lot-A-Spirits	3,402,977.63	48,028.730	70.8530
Hy-Vee Food Store / Urbandale	3,298,320.24	39,712.200	83.0556
Sam's Club 6979 / Ankeny	3,233,622.07	56,069.310	57.6719
Another Round / DeWitt	3,076,358.85	43,279.730	71.0808
Hy-Vee Food Store / Coralville	2,885,988.44	41,627.080	69.3296
Hy-Vee / Wauke	2,831,663.98	37,254.900	76.0078
Central City Liquor, Inc.	2,799,272.61	26,774.170	104.5512
Sam's Club 6514 / Waterloo	2,733,032.37	52,845.640	51.7173
Costco Wholesale #1111 / Coralville	2,711,095.10	50,398.880	53.7928
Happy's Wine & Spirits	2,707,078.75	37,635.400	71.9291
Hy-Vee Wine and Spirits / WDM	2,631,598.90	34,964.360	75.2652

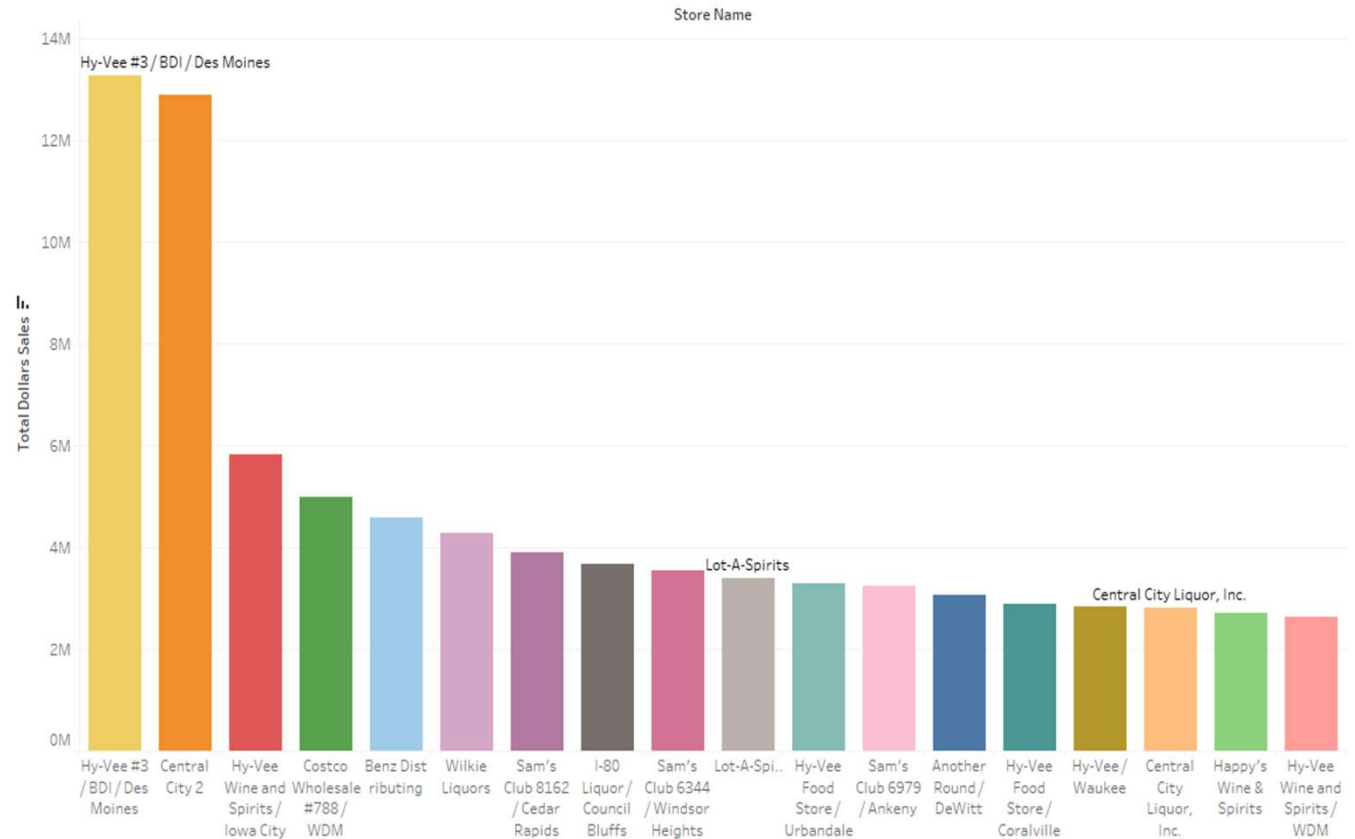
Liquor Categories

```
... )
>>> top_categories.show(20, False)
```

category_name	formatted_total_sales_dollars	formatted_total_volume_sold	formatted_dollar_to_volume_ratio
American Vodkas	65,582,684.00	1,603,247.150	40.9062
Canadian Whiskies	50,322,883.27	846,501.210	59.4481
Straight Bourbon Whiskies	36,948,852.92	377,571.990	97.8591
Whiskey Liqueur	26,664,116.55	327,699.950	81.3675
100% Agave Tequila	25,359,203.78	171,832.120	147.5813
Spiced Rum	24,549,501.08	415,171.880	59.1309
Tennessee Whiskies	17,219,118.57	146,473.460	117.5579
Imported Brandies	16,646,780.08	90,520.270	183.9011
Imported Vodkas	15,239,358.28	224,434.090	67.9013
Blended Whiskies	13,477,222.83	238,734.640	56.4527
Mixto Tequila	11,258,758.49	178,015.010	63.2461
American Flavored Vodka	11,213,426.84	199,830.150	56.1148
Imported Cordials & Liqueurs	10,903,561.22	93,719.370	116.3427
Irish Whiskies	9,880,194.07	84,860.790	116.4283
Cream Liqueurs	9,545,889.06	106,311.010	89.7921
Flavored Rum	9,514,014.85	161,823.640	58.7925
Cocktails /RTD	9,079,479.92	302,664.870	29.9985
Temporary & Specialty Packages	8,631,775.17	75,375.250	114.5174
Imported Schnapps	7,113,738.03	98,804.280	71.9983
White Rum	7,097,933.79	170,746.380	41.5700

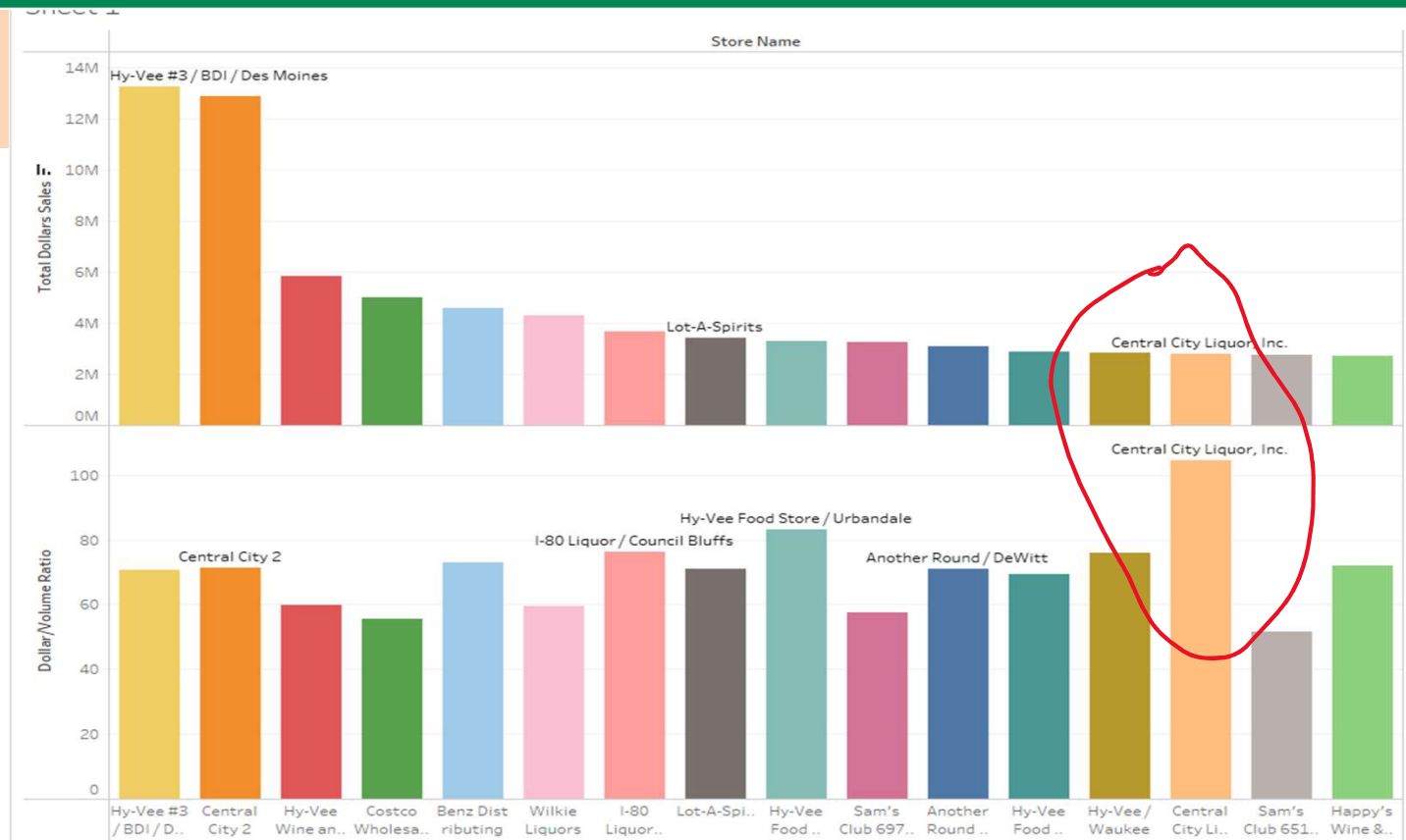
Top 20 Liquor Stores based of Total Revenues

- ❖ Based on the revenue, two major stores are leading:
- ❖ **Hy-Vee #3 DB / Des Moines**
- ❖ **Central City 2**



Top 20 Liquor Stores based of Total Revenues

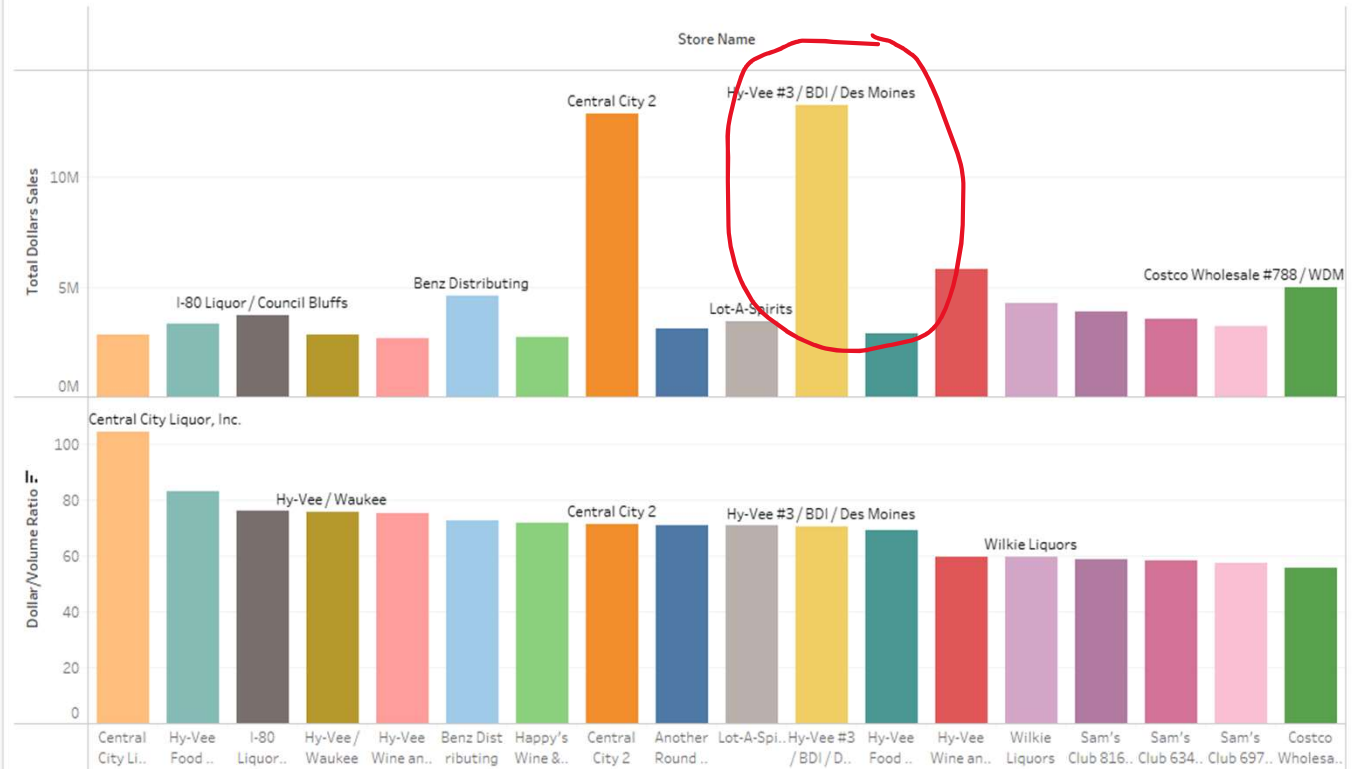
- ❑ The dollar-to-volume ratio provides a refined measure of store performance, revealing differences in effectiveness beyond mere sales and volume metrics.
- ❑ **Central City Liquor, Inc.** boasts a higher dollar-to-volume ratio of 104.55 on \$2.8M in revenue.
- ❑ Despite its strong sales, **Costco Wholesale #788 / WDM** has a lower dollar-to-volume ratio of \$55.6, indicating a strategy that prioritizes volume over higher profit margins per unit



Liquor Stores ranked based on the dollar/volume ratio

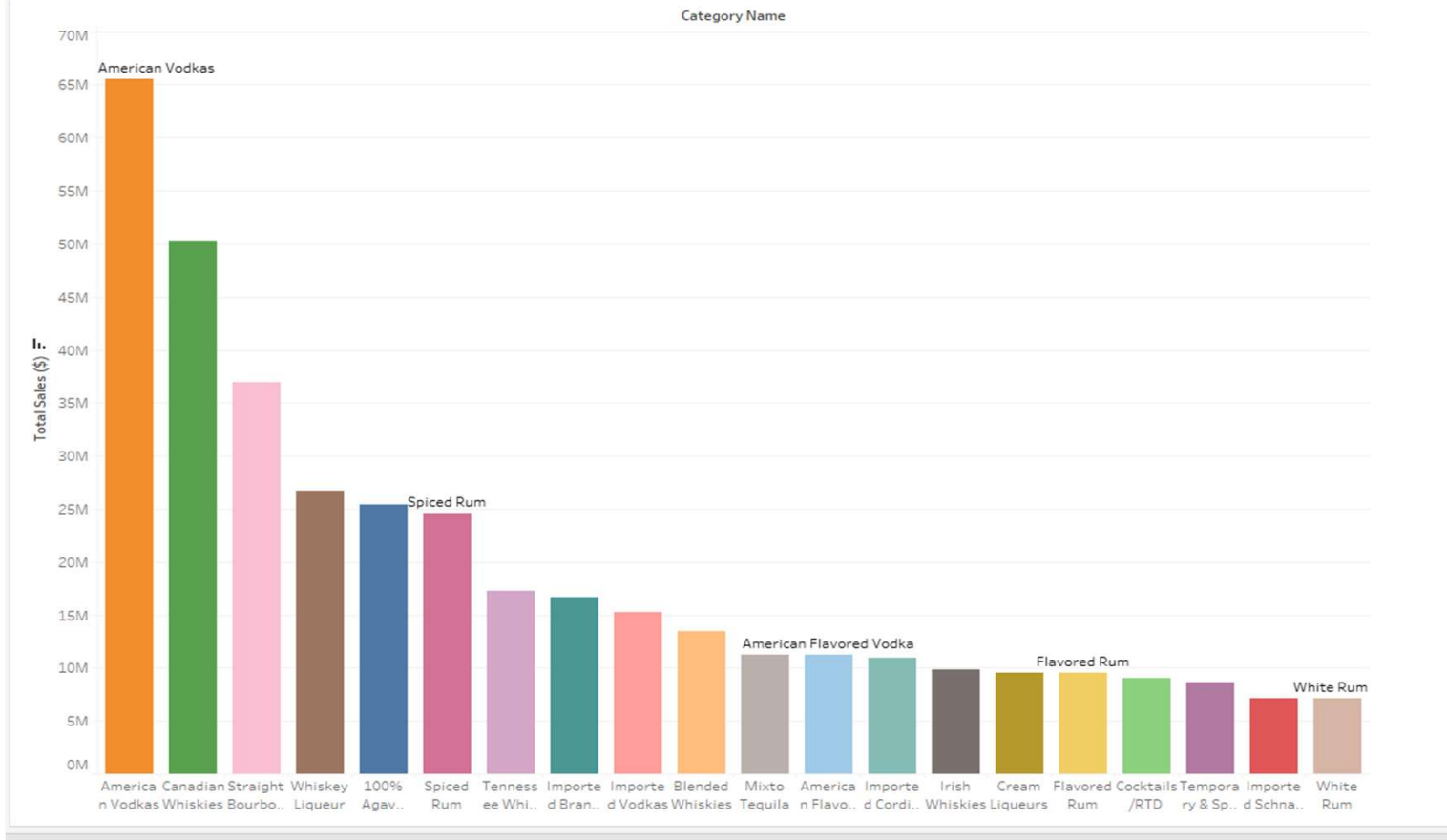
- ❑ The two top are ranked 11th and 8th based on dollar to volume ratio
- ❑ The two store with more revenue potential are **Central City Liquor** and **H-Vee Food Store Urbandale**

Sheet 1



Top 20 Liquor Categories ranked by revenues

- Rank shows that **American Vodkas**, lead with the most revenue 65M
- Followed by **Canadian Whiskies** and **Straight Bourbon Whiskies**



Top 20 Liquor Categories ranked by revenues

- ❑ Even though with the highest revenue of \$65M, **American Vodkas** yet exhibit a lower profitability ratio at **\$41**, ranked 7th
- ❑ **Imported Brandies**, despite lower total revenues of **\$16M**, achieve the highest dollar-to-volume profitability at **\$184** per gallon
- ❑ 100% Agave Tequila demonstrates a balanced approach with a dollar-to-volume ratio of **\$147.38**, indicating a strong market position.



Conclusion and Recommendations

- ❑ **Optimize Merchandising Strategies:** Capitalize on the success of high-performing stores like **Hy-Vee #3 / BDI** by analyzing and adopting their effective practices in product selection. Additionally, explore merchandising strategies that enhance the **dollar-to-volume ratio**, to increase their profitability without compromising on sales volume.
- ❑ **Enhance Inventory Allocation:** Given the varying dollar-to-volume ratios, stores could optimize inventory levels based on sales performance and profitability. For instance, **Hy-Vee Food Store / Urbandale** and **Central City Liquor, Inc.**, which show relatively high dollar-to-volume ratios, suggest that customers there may prefer premium products. The business could increase the stock of higher-end liquors in these stores.
- ❑ **Expand Premium Selections:** With **Imported Brandies** having the highest dollar-to-volume ratio, there's an indication that consumers are willing to pay more for premium products. The business should consider expanding its selection of premium brands and specialty products in other high-margin categories, such as **Whiskey Liqueur** and **100% Agave Tequila**, to capitalize on this trend

Question 2

- **"How does pricing affect the volume and profitability of liquor sales, and how do adjustments in pricing influence the sales quantities across different categories?"**



Sample Data Overview

```
>>> df_categories = spark.read.csv("file:///home/kiran/Downloads/selected_liquor_columns.csv", header=True, inferSchema=True)
>>> df_categories
DataFrame[category: double, state_bottle_retail: double, bottles_sold: int, sale_dollars: double, item_description: string, category_name: string]
>>> df_categories.show()
```

category	state_bottle_retail	bottles_sold	sale_dollars	item_description	category_name
1011400.0	11.25	3	33.75	Jack Daniels Old ...	Tennessee Whiskies
1081600.0	1.35	48	64.8	Fireball Cinnamon...	Whiskey Liqueur
1051100.0	19.5	1	19.5	E & J VS	American Brandies
1041100.0	10.38	6	62.28	Caliber Gin	American Dry Gins
1011200.0	13.49	3	40.47	Makers Mark Replica	Straight Bourbon ...
1012100.0	15.68	6	94.08	Black Velvet Apple	Canadian Whiskies
1012100.0	14.25	6	85.5	Windsor Canadian PET	Canadian Whiskies
1012100.0	21.75	1	21.75	Canadian Club Whisky	Canadian Whiskies
1012100.0	11.03	1	11.03	Crown Royal Mini	Canadian Whiskies
1012100.0	7.85	12	94.2	Black Velvet Toas...	Canadian Whiskies
1081600.0	8.0	24	192.0	Fireball Cinnamon...	Whiskey Liqueur
1081600.0	8.0	24	192.0	Fireball Cinnamon...	Whiskey Liqueur
1042100.0	10.49	3	31.47	Tanqueray Gin	Imported Dry Gins
1022200.0	21.0	2	42.0	Cazadores Reposado	100% Agave Tequila
1011400.0	13.59	3	40.77	Jack Daniels Tenn...	Tennessee Whiskies
1011100.0	15.75	1	15.75	Red Stag Black Ch...	Blended Whiskies
1011100.0	3.14	24	75.36	Five Star	Blended Whiskies
1081400.0	7.11	1	7.11	Arrow Mcdales But...	American Schnapps
1012100.0	2.34	48	112.32	Black Velvet	Canadian Whiskies
1012100.0	49.49	12	593.88	Crown Royal Regal...	Canadian Whiskies

```
only showing top 20 rows
>>>
```

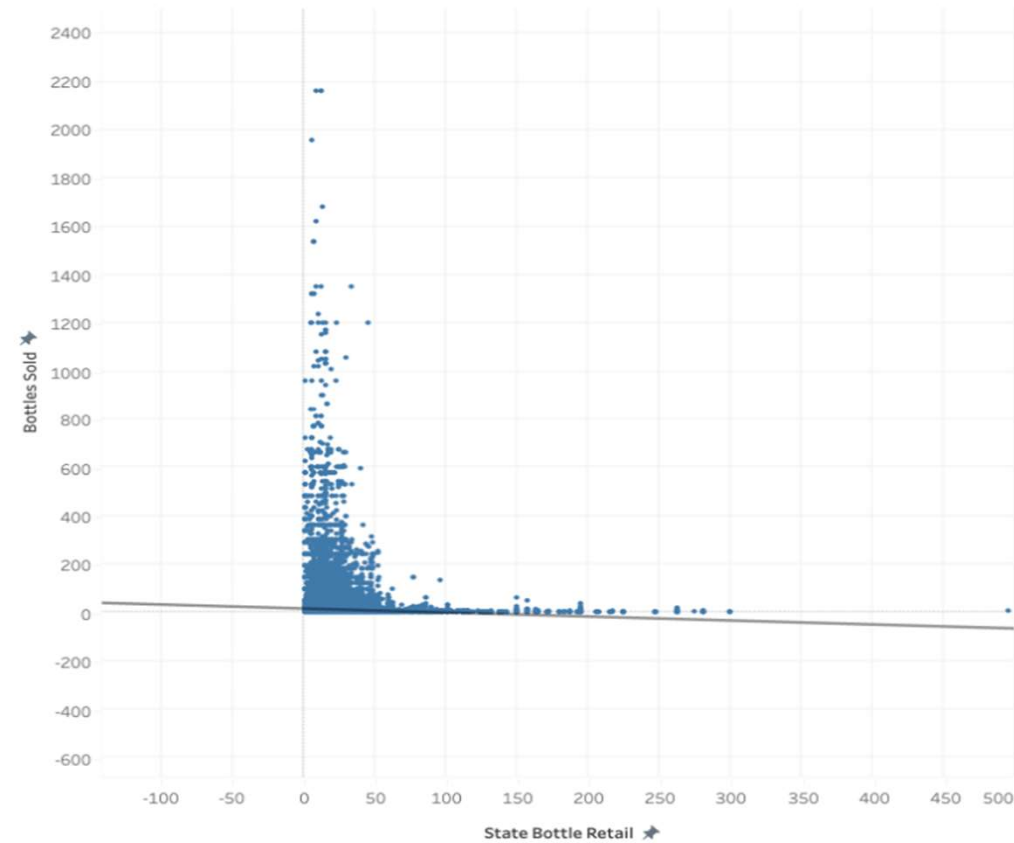
Correlation Analysis

we are examining the strength and direction of the relationship between the prices of products and their sales volume and profitability.

```
>>> correlations.show()
+-----+-----+
| price_volume_corr| price_profit_corr|
+-----+-----+
| -0.06570430284499904| 0.09645475239875996|
+-----+-----+
```

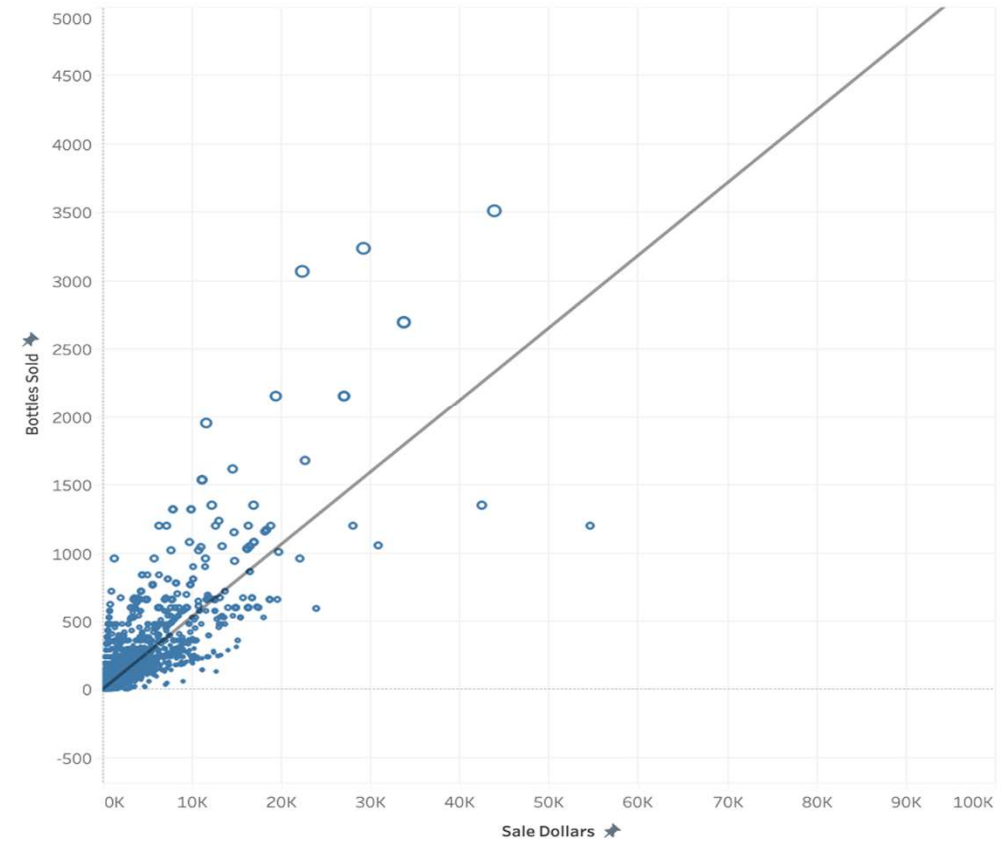

State Bottle Retail vs Bottle sold

Sheet 2



Sale Dollars vs Bottle Sold

Sheet 1



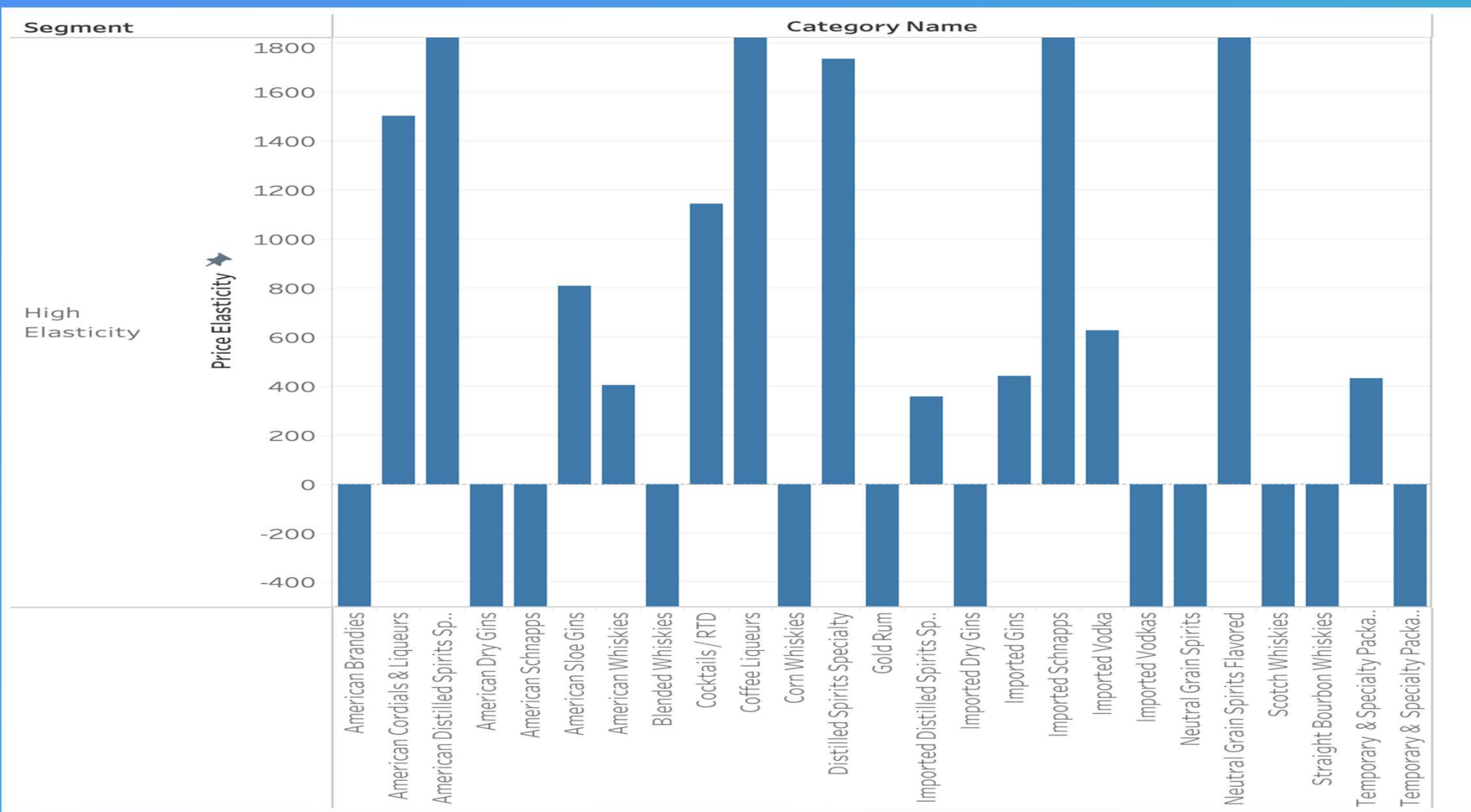
- **Price Elasticity Calculation:** Calculate the price elasticity of demand for each category by analyzing the percentage change in sales volume relative to the percentage change in pricing.
- **State Bottle Retail (\$/P):** The price you see on the tag for each bottle of liquor.
- **Total Bottles Sold:** How many bottles have been sold.
- **Total Sales:** The total money earned from selling bottles.
- **Prev Bottles Sold:** How many bottles were sold before this.
- ΔQ : How much the number of sold bottles has gone up or down since last time.
- ΔP : How much the price has gone up or down since last time.
- $\Delta Q/Q$: Tells you by what percentage the number of sold bottles has changed.
- $\Delta P/P$: Tells you by what percentage the price has changed.
- **Price Elasticity:** This number shows if changing the price, a little makes a big difference in sales.
- **Segment:** This tells us whether the liquor type is very responsive ("High Elasticity") or not much responsive ("Low Elasticity") to price changes.

```
>>> sampled_df_no_duplicates.show(5)
+-----+-----+-----+-----+-----+-----+-----+-----+
|state_bottle_retail|total_bottles_sold|total_sales|prev_bottles_sold|ΔQ|ΔQ/Q|prev_price|ΔP|
|ΔP/P|Price_Elasticity|category|category_name|
+-----+-----+-----+-----+-----+-----+-----+-----+
|7.01|48790|342017.89999999994|20336|28454|58.31932773109244|6.99|0.019999999999999574|0
.2853067047075546|204.40924369748333|1701100.0|Temporary & Speci...|
|46.02|41313|1901224.2599999884|935|40378|97.73678987243724|45.96|0.060000000000002274|0.
13037809647979634|749.6411783215652|1011400.0|Tennessee Whiskies|
|47.01|89|4183.89|40|49|55.0561797752809|47.0|0.009999999999999801|0.0
21272069772384623|2588.19101123647|1901200.0|Special Order Items|
|10.13|7532|76299.16000000009|4441|3091|41.03823685608072|10.11|0.02000000000000135|0.
19743336623890767|207.85866967603485|1062400.0|Spiced Rum|
|13.53|19553|261770.87|30827|-11274|-57.65867130363627|13.52|0.009999999999999787|0
.0739098300073894|-780.1218227382153|1011300.0|Single Barrel Bou...|
+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 5 rows

>>> █
```

```
>>> top_rows = segmented_df.orderBy("segment").show(5)
+-----+-----+-----+-----+-----+-----+-----+-----+
|state_bottle_retail|total_bottles_sold|total_sales|prev_bottles_sold|ΔQ|ΔQ/Q|prev_price|ΔP|
|ΔP/P|Price_Elasticity|segment|
+-----+-----+-----+-----+-----+-----+-----+-----+
|2.7|505920|1365983.9999999672|7244|498676|98.5681530676787|2.69|0.01000000000000231|0.3
703703703703789|266.13401328272636|High Elasticity|
|5.01|263142|1318341.4199999475|18569|244573|92.94335377856822|5.0|0.009999999999999787|0.19
960079840318934|465.6462024306368|High Elasticity|
|3.2|31442|100614.3999999999|485|30957|98.45747725971631|3.17|0.03000000000000025|0.9
375000000000078|105.02130907702987|High Elasticity|
|4.83|114074|550977.4199999977|2385|111689|97.9092518891246|4.82|0.009999999999999787|0.20
703933747411568|472.9016866244819|High Elasticity|
|4.97|116068|576857.9599999968|30920|85148|73.36044387772685|4.95|0.019999999999999574|0.4
024144869215206|182.3007030361551|High Elasticity|
+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 5 rows

>>> █
```



- **High Elasticity:** Categories with bars above the horizontal axis have positive elasticity, meaning sales increase when the price drops and decrease when the price rises. Taller bars indicate greater sensitivity. For these categories, even a small change in price could lead to a large change in the quantity sold.
- **Low or Negative Elasticity:** Categories with bars below the axis have low or negative elasticity. Here, changes in price have a less significant impact on the quantity sold, or sales might even go up when the price increases.

Some of the High Elasticity Categories

- 1.American Cordials & Liqueurs
- 2.American Distilled Spirits Specialty
- 3.Cocktails / RTD
- 4.Coffee Liqueurs
- 5.Distilled Spirits Specialty

Question 3

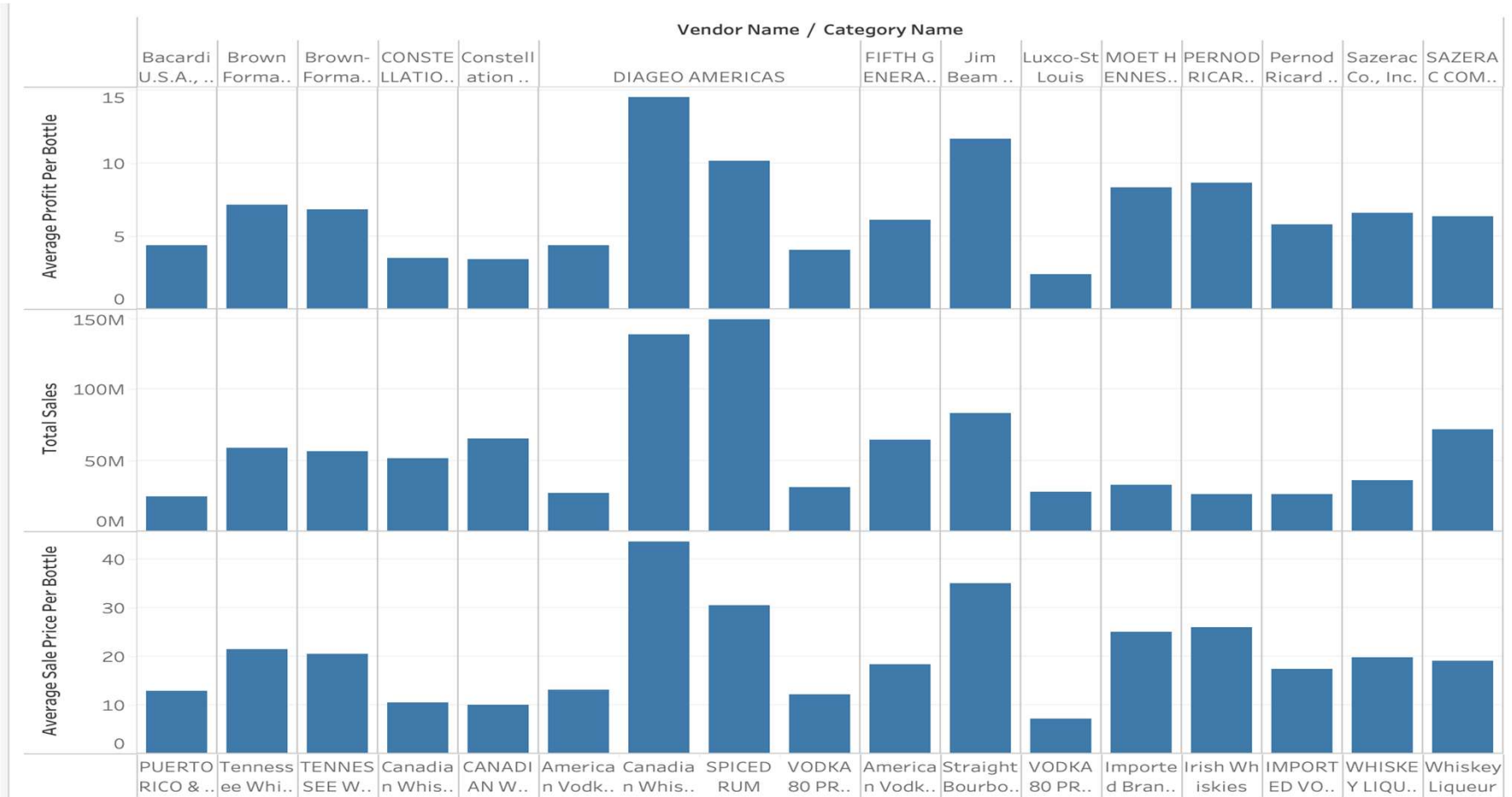
- Which vendors contribute most to the store's profitability, and how can this information improve vendor negotiations and inventory decisions?



Query Result

Vendor Name	Category Name	Total Profit	Total Sales	Average Profit Per Bottle	Average Sale Price Per Bottle
DIAGEO AMERICAS	Canadian Whiskies	2.6804918950000834E7	8.02488490299996E7	7.143848396334013	21.424394272627705
Diageo Americas	SPICED RUM	2.5952129039998062E7	7.775624414999816E7	5.1133499153228	15.31912275558262
SAZERAC COMPANY INC	Whiskey Liqueur	2.38787284500009E7	7.174731151000008E7	6.373573813244071	19.117334218285077
DIAGEO AMERICAS	Spiced Rum	2.3628504130001575E7	7.096484845000362E7	5.069567079424951	15.20599687699381
Constellation Win...	CANADIAN WHISKIES	2.2026625379998032E7	6.469269810000175E7	3.3845670638528884	10.042223192482565
FIFTH GENERATION INC	American Vodkas	2.1627639500000138E7	6.442162636000207E7	6.150680331593812	18.449947073815423
Brown Forman Corp.	Tennessee Whiskies	1.958391525000071E7	5.874403819000196E7	7.183007817679969	21.543973012614366
Diageo Americas	CANADIAN WHISKIES	1.9366998640000068E7	5.8066014049998485E7	7.404055269845713	22.202664198667147
Brown-Forman Corp...	TENNESSEE WHISKIES	1.873814374000024E7	5.620814805000143E7	6.819289520872533	20.452059417471514
CONSTELLATION BRA...	Canadian Whiskies	1.7244171029998474E7	5.1554946730003774E7	3.4788128994573055	10.427228472269473
Jim Beam Brands	Straight Bourbon ...	1.5640158669999797E7	4.697402999999916E7	5.825280035162035	17.472580338466685
Sazerac Co., Inc.	WHISKEY LIQUEUR	1.2055908609999834E7	3.615643032999957E7	6.610578343158289	19.828460102107012
Jim Beam Brands	STRAIGHT BOURBON ...	1.203136812999982E7	3.59614809899989E7	5.906379084629999	17.66844211543308
MOET HENNESSY USA	Imported Brandies	1.0963034169999948E7	3.2588282969998047E7	8.349207574082298	25.040803564273254
Diageo Americas	VODKA 80 PROOF	1.0264779020000178E7	3.0681460090000045E7	4.074191538863648	12.18520342723558
Luxco-St Louis	VODKA 80 PROOF	9330587.69999944	2.79671475200023E7	2.4213715272187644	7.257971501464157
DIAGEO AMERICAS	American Vodkas	8882327.290000418	2.6641164470002297E7	4.408409892135771	13.21718030171438
Pernod Ricard USA...	IMPORTED VODKA	8786084.850000104	2.628494331999779E7	5.842784648613119	17.48440154825601
PERNOD RICARD USA	Irish Whiskies	8774770.189999944	2.603133397999879E7	8.700731483325647	26.097863307443273
Bacardi U.S.A., Inc.	PUERTO RICO & VIR...	8308912.209999895	2.4624935310000185E7	4.3719666005355	12.980464924783734

Vendor/Category vs Sales/Profits



Contributions to Profit

1. As **DIAGEO AMERICAS**, stands out prominently, particularly in the categories of Canadian Whiskies and Spiced Rum.
 - Marketing efforts, such as joint promotional campaigns or sponsored events with DIAGEO AMERICAS could drive sales and enhance brand visibility for both parties.
 - Exclusivity in stocking, or first access to new product lines, and prioritize shelf space and positioning for Canadian Whiskies and Spiced Rum, and the new line products could give the company a competitive edge in the market.
2. **Sazerac Company**'s presence in the Whiskey Liqueur category have a balanced approach between volume and profitability whereas it's average sale price suggests a room for pricing strategy.
 - Competitive pricing strategies might increase the volume of sales for Sazerac products, which could offset the lower margin per unit and lead to higher overall profits capturing a larger market share.