# Assignment 2

Data Management and Business Intelligence

POLITIS KOSTAS
VAIDOMARKAKIS PANAGIOTIS

# Contents

# Dataset Collection

For the purpose of the current assignment, first we searched for our dataset in Kaggle. We selected the dataset “Iowa Liquor Sales”. According to the dataset description, the Iowa Department of Commerce requires that every store that sells alcohol in bottled form for off-the-premises consumption must hold a class "E" liquor license (an arrangement typical of most of the state alcohol regulatory bodies). All alcoholic sales made by stores registered thusly with the Iowa Department of Commerce are logged in the Commerce department system, which is in turn published as open data by the State of Iowa.

This data is a representative sample of sale activity for alcohol in the United States. It shows the sales of Iowa liquor vendors to stores all over the United States and it can be used to answer many questions like: how much alcohol is sold and consumed in the United States? What kind? Who is the biggest vendor? Etc.

The available columns of the database are the following:

- Invoice_and_item_number: Concatenated invoice and line number associated with the liquor order. This provides a unique identifier for the individual liquor products included in the store order.
- PPDATE: Date of order.
- Store_number: Unique number assigned to the store who ordered the liquor.
- Store_name: Name of store who ordered the liquor.
- Address: Address of store who ordered the liquor.
- City: City where the store who ordered the liquor is located.
- Zip_code: Zip code where the store who ordered the liquor is located.
- County_number: Zip code where the store who ordered the liquor is located.
- Store Location: Geocoded location of store who ordered the liquor.
- County: County where the store who ordered the liquor is located.
- Category: Category code associated with the liquor ordered.
- Category_name: Category of the liquor ordered.
- Vendor_number: The vendor number of the company for the brand of liquor ordered.
- Vendor_name: The vendor’s name of the company for the brand of liquor ordered.
- Item_number: Item number for the individual liquor product ordered.
- Item_description: Description of the individual liquor product ordered.
- Pack: The number of bottles in a case for the liquor ordered.
- Bottle_volume_ml: Volume of each liquor bottle ordered in milliliters.
- State_bottle_cost: The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered.
- State_bottle_retail: The amount the store paid for each bottle of liquor ordered.
- Bottles_sold: The number of bottles of liquor ordered by the store.
- Sale_dollars: Total cost of liquor order (number of bottles multiplied by the state bottle retail).

- Volume_sold_liters: Total volume of liquor ordered in liters. (i.e. (Bottle Volume (ml) x Bottles Sold)/1,000)
- Volume_sold_gallons: Total volume of liquor ordered in gallons. (i.e. (Bottle Volume (ml) x Bottles Sold)/3785.411784)

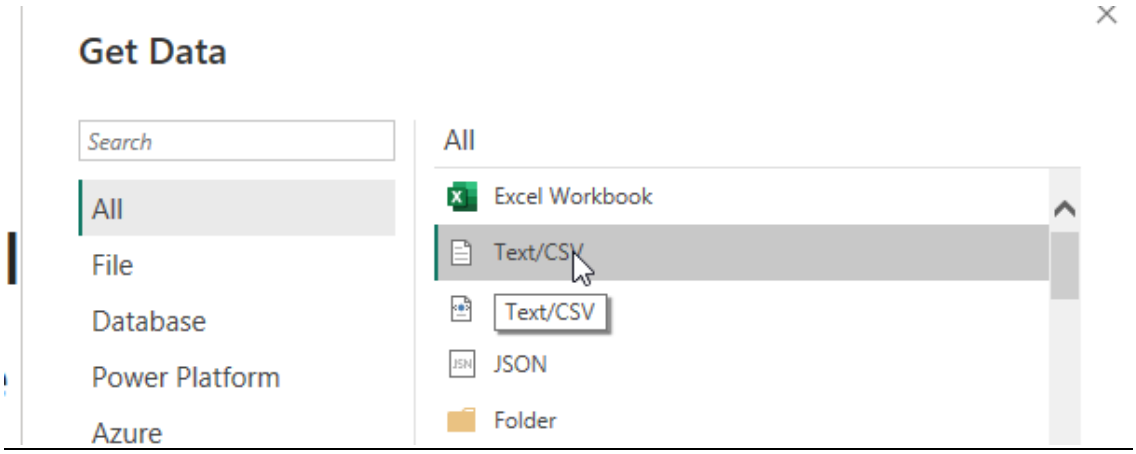
From a business point of view, we will try to present a view regarding the overall business activity of the Iowa vendors for the sake of the Iowa vendors association and take a closer look on the overall profitability of the Iowa liquor vendors as a whole. More specifically, our business analysis will try to answer the following questions:

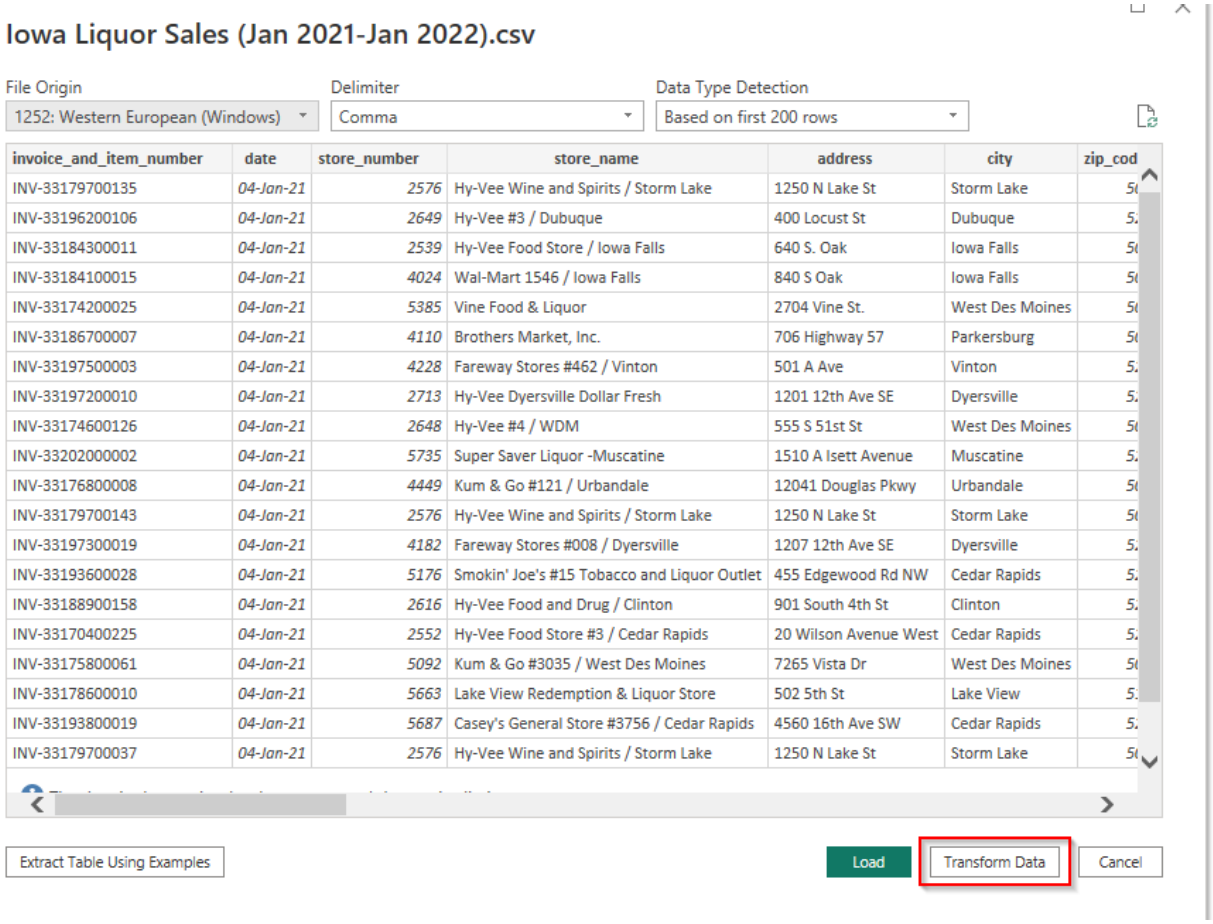
1. Which is the most sold category and individual liquor in the United States?
2. Which vendor is the most profitable? Which is each vendor's most profitable liquor category in terms of sales?
3. Which is the most profitable county and city for Iowa vendor association in terms of revenue and bottles sold?
4. Does Iowa vendor association reach its monthly target in terms of Revenue and total sales?
5. What is the evolution of Iowa's profitability in the last one year in total? What is its profitability for each category and each individual liquor?
6. How would a change in price potentially affect the total profit?

## Data Cleaning

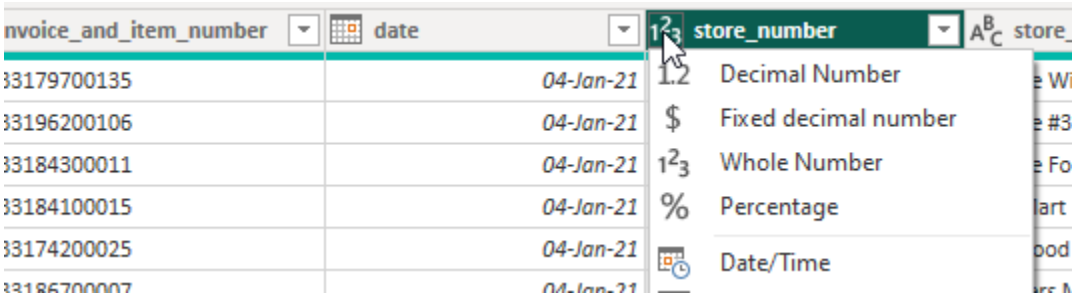
After downloading the dataset in csv file, we proceeded by checking the dataset for inconsistencies and by transforming it to a final form in order to insert it in SQL server. Due to its big size (2.5 M rows) we used the data transformation feature of Power BI.

After inserting the csv ins Power BI, we selected the transform data option:





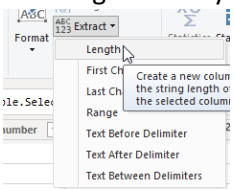

First, we checked the data type of each row and we made sure that it is in the correct format. For example, the column Date must be date/time, store number must be whole number, sale_dollars must be decimal number etc.



Secondly, we checked if there is any kind of inconsistency between the values of each row and make some logic tests to check their validity. These tests serve the dual purpose of both fixing inconsistency between the values of each row and of acquiring a better understanding of our data.

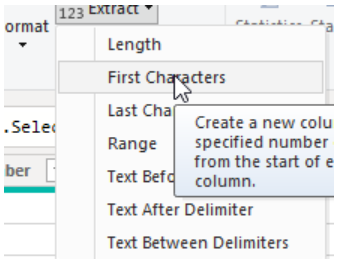
For example:

- Zip Code: We created a new column based to its length, to check whether all the zip codes had 5 digits as they should.

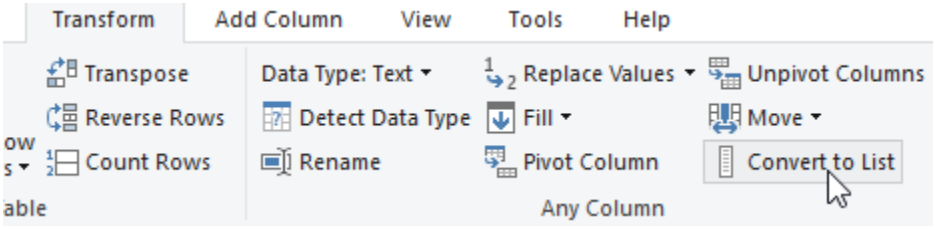



-

- Address: First we extracted the first character of each row.

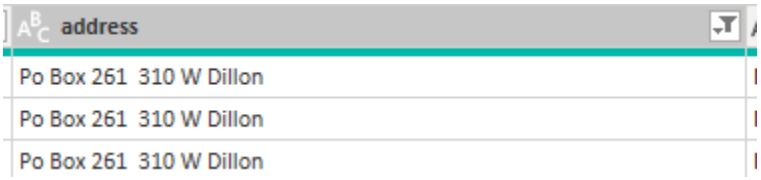



Then we turned the new column into a list and used the remove duplicates function to keep only the unique values.




| | List |
|---|---|
| 1 | 9 |
| 2 | 2 |
| 3 | 7 |
| 4 | 1 |
| 5 | 5 |
| 6 | 4 |
| 7 | 3 |
| 8 | 8 |
| 9 | 6 |
| 10 | H |
| 11 | O |
| 12 | P |

We noticed that there are some values that begin with a letter. However, after checking them out, they seem to be legit values that make sense thus we did not proceed to any transformation.

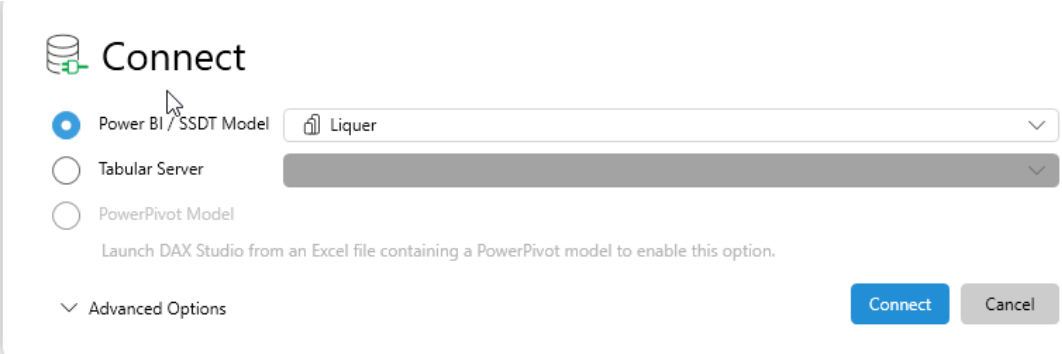



- Invoice_and_item_number: After extracting the three first characters we checked whether very value starts with the chars “INV”.

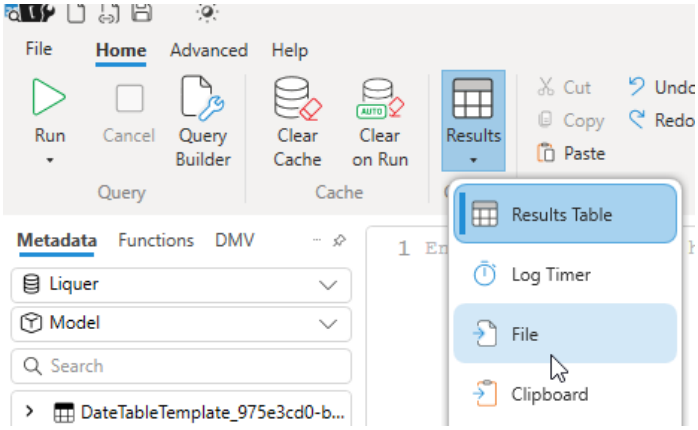
The above are some indicative tests that we did in order to check data validity. As a final test, we searched for empty values in each row. Even though many rows contained empty values, at this stage we leaved them as they are.

After finalizing our data, we proceeded by exporting the final table in a new CSV file in order to import it to our database. For that purpose, we used DAX studio.

On the connection table we selected on the PowerBI/SSDT Model our Power BI file Liquor.

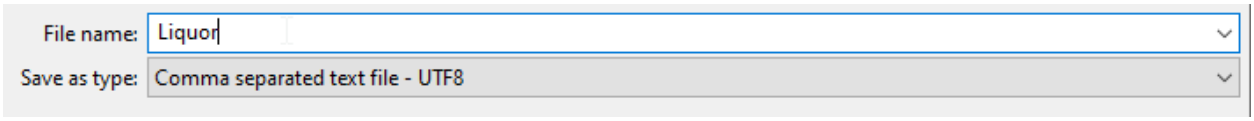



After the connection was established, we selected the exporting option of the query to be on file format.




By running the query below and we selected to save the results on a CSV format.

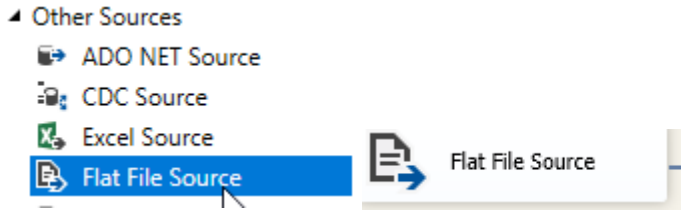
```
// Generated DAX Query
EVALUATE
('Iowa Liquor Sales (Jan 2021-Jan 2022)' )
```



# Liquor database import

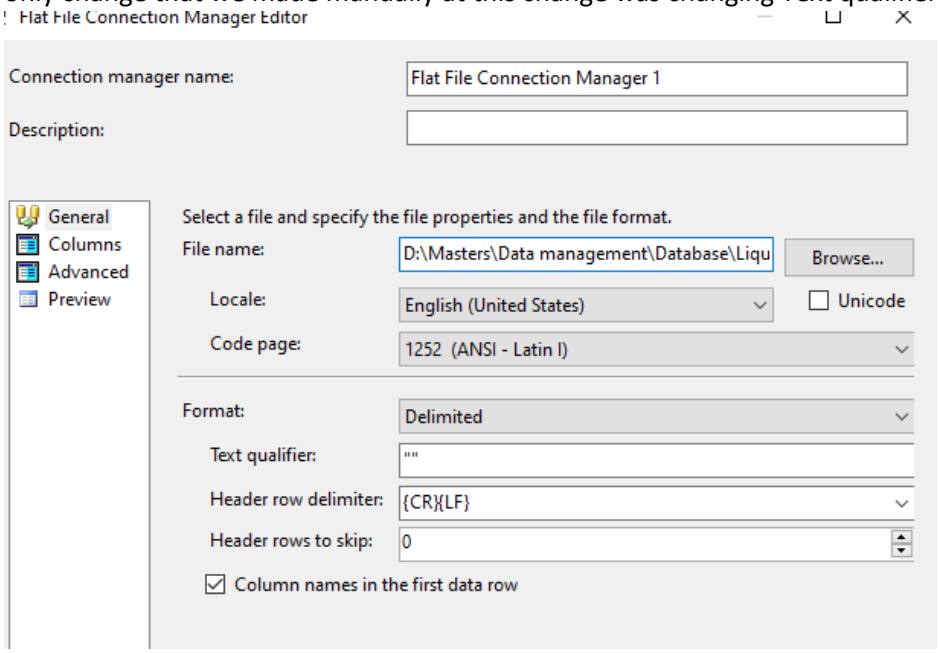
In order to insert the liquor database into MS SQL we used SSIS.  First, we inserted a Data Flow Task.


Within the Data Flow Task, we added a Flat File Source which was used to upload the csv file into SSIS environment.

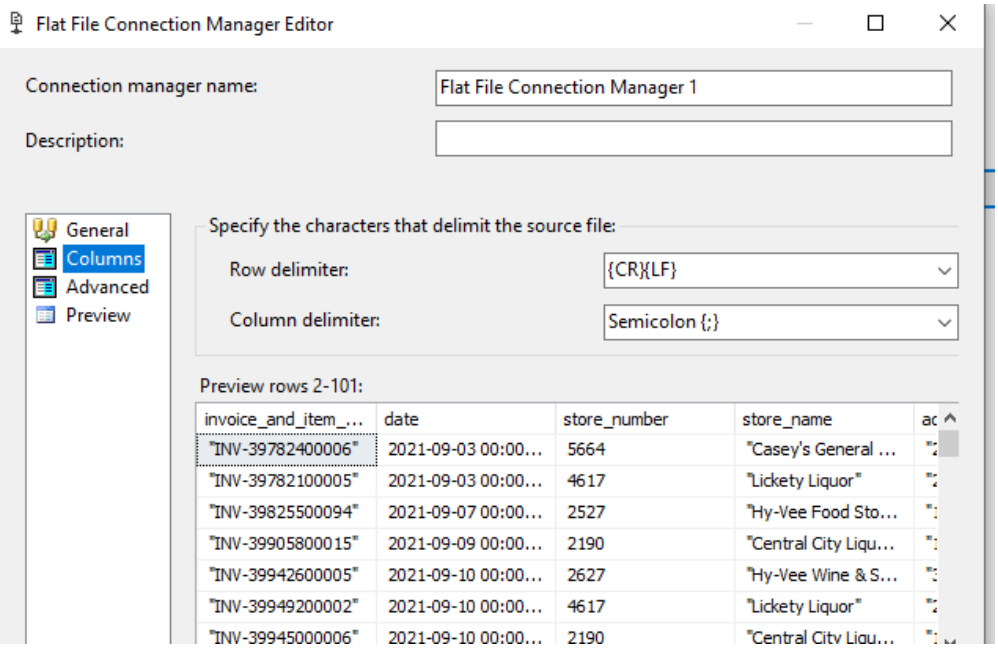



On the Flat File Source, we selected for a new Flat File Conmnection.

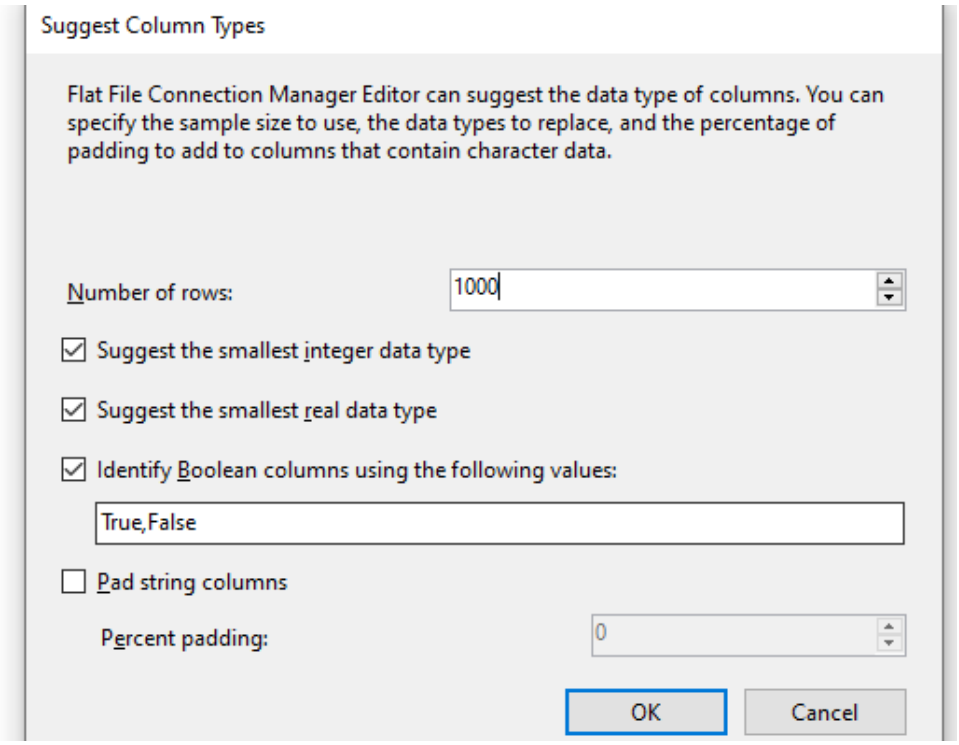

We then proceeded by choosing the csv file and adjusting the parameters wherever was needed. The only change that we made manually at this change was changing Text qualifier to “”.



On the tab Columns, we set the Column delimiter as Semicolon and we got a first preview regarding the values of the rows.




On the tab Advanced, first we clicked on the Suggest Types option. After setting the number of rows to 1000 to have a better sample, we clicked on ok.



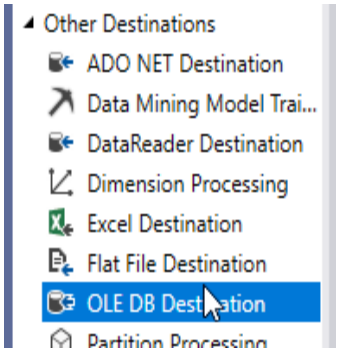
String columns and date columns were adjusted correctly as DT_STR and DT_DATE respectively.

However, most of the numerical values were updated as “single-byte signed integer [DT_I]”. Due to the small data type size, this created many problems during the import process on SQL server. Thus, for all int values (county_number, vendor_number etc.) we changed the type to “eight-byte signed integer [DT_I4]” and for all float values (salle_dollars, bottles_sold etc.) we changed the type to float [DT_R4].

Finally, on the preview tab we checked some of the data to make sure that each column had been identified properly. For example, that no data from one column had overflowed to another.

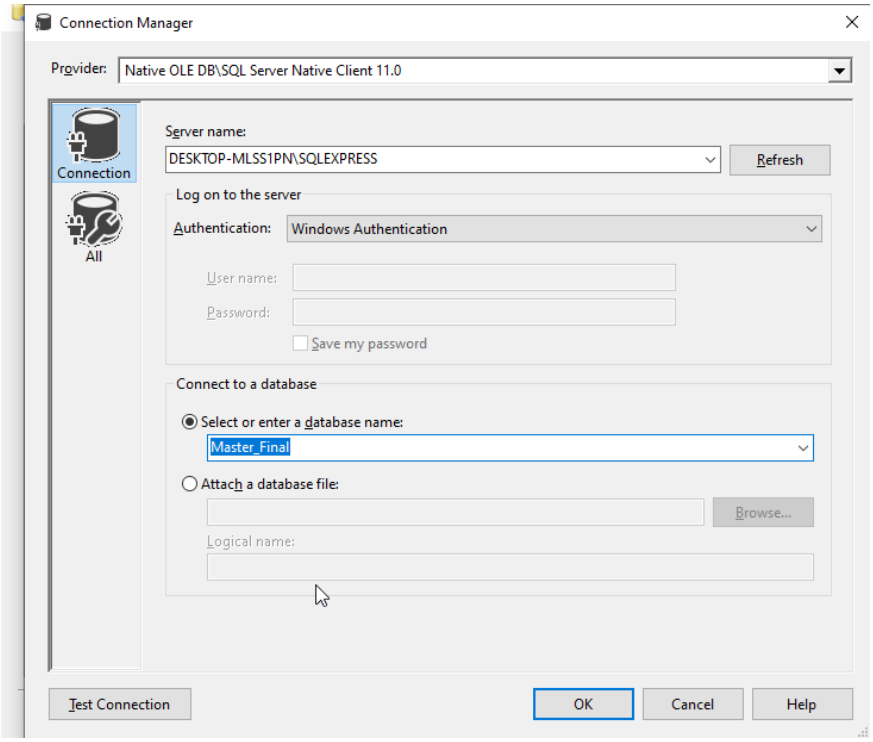
Having established that the CSV has been uploaded and that all of its rows and values have been defined correctly, we proceeded by creating a connection between the uploaded CSV file and the SQL server.

First, we inserted the OLE DB Destination to the Data Flow and made a connection between the Flat File Source and OLE DB Destination.

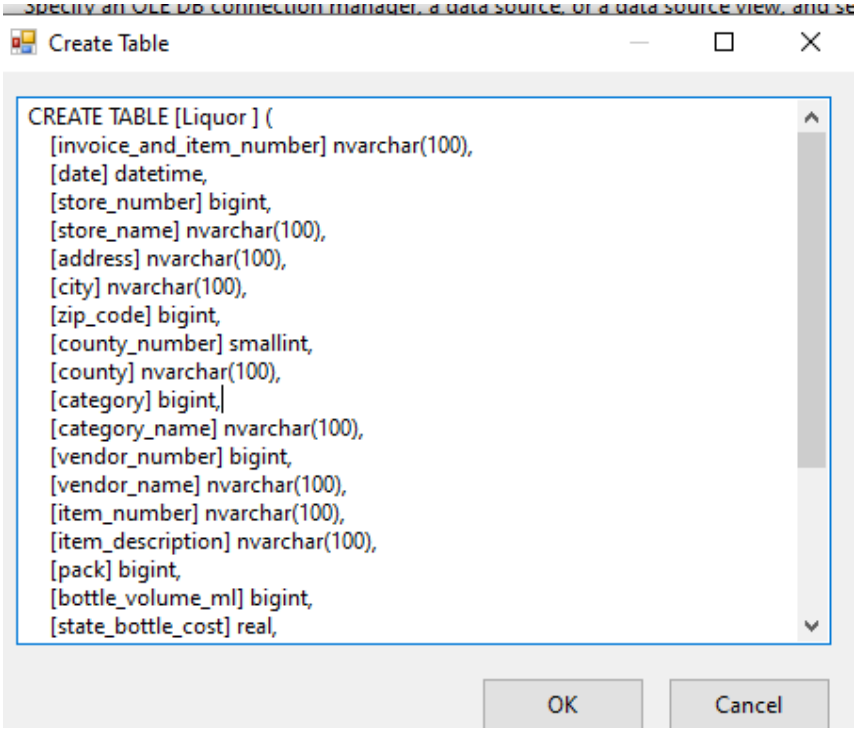




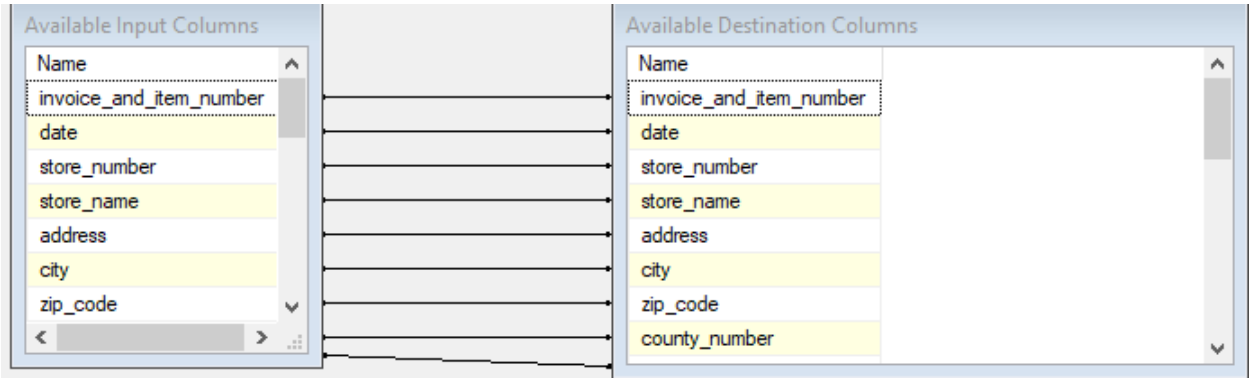
On the Connection Manger tab, we selected “New” in order to establish a new SQL connection. After selecting the server, the authentication method and the database where the table will be created, we selected ok.

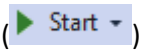



After the connection on the database was established, we selected to automatically create a new table where the CSV file would be inserted. After making some adjustments like changing the table name, set all nvarchar types to nvarchar (250) we selected ok.



On the Mappings tab we checked whether all the columns from the CSV have been mapped correctly with the table columns.



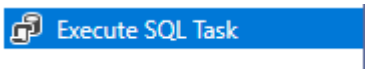

Having made sure that everything is ok on both the Flat File Source and OLEB DB Destination we clicked on start ( 

) to begin the importing procedure.

After the importing process ended successfully, we made a select statement on our newly created table to ensure that everything was imported successfully.

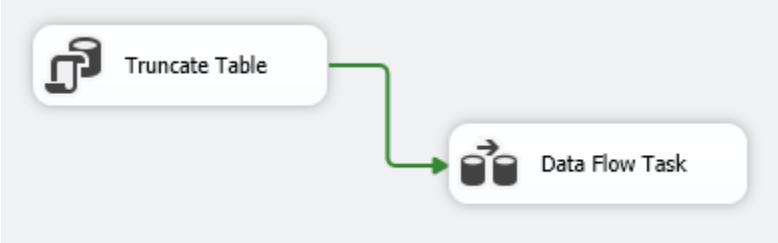
| | invoice_and_item_number | PPDATE | store_number | store_name | address | city | zip_code | county_number | county | category | category_name |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | INV-33337000096 | 2021-01-11 | 2552 | Hy-Vee Food Store #3 / Cedar Rapids | 20 Wilson Avenue West | Cedar Rapids | 52404 | 57 | LINN | 1081400 | American Schnapps |
| 2 | INV-33348400010 | 2021-01-11 | 5955 | Gowrie Grocery | 1119 Market St | Gowrie | 50543 | 94 | WEBSTER | 1081400 | American Schnapps |
| 3 | INV-33347000046 | 2021-01-11 | 5441 | Liquor on the Corner | 300 East Main | Cherokee | 51012 | 18 | CHEROKEE | 1081400 | American Schnapps |
| 4 | INV-33401800090 | 2021-01-12 | 5423 | Stammer Liquor Corp | 615 2nd Ave | Sheldon | 51201 | 71 | OBRIEN | 1081400 | American Schnapps |
| 5 | INV-33369500048 | 2021-01-12 | 3943 | Good and Quick Co | 519 Lincolnway | Nevada | 50201 | 85 | STORY | 1081400 | American Schnapps |
| 6 | INV-33422200014 | 2021-01-13 | 3682 | Jumbos | 110 E Main St | Wesley | 50483 | 55 | KOSSUTH | 1081400 | American Schnapps |
| 7 | INV-33516100009 | 2021-01-15 | 4472 | Quick Shop Foods / Centerville | 1023 South Main | Centerville | 52544 | 4 | APPANOOSE | 1081400 | American Schnapps |
| 8 | INV-42144500018 | 2021-11-18 | 2582 | Hy-Vee #2 Food Store / Mason City | 551 S Illinois Ave | Mason City | 50401 | 17 | CERRO GORD | 1081400 | American Schnapps |

On the SSIS Control Flow, we also added an Execute SQL Task to take place before the Data Flow Task.




There, after establishing connection we added a query to truncate the liquor table each time we run the procedure in order not to insert duplicate values by mistake.

| | |
|---|---|
| Connection | **DESKTOP-MLSS1PN\SQLEXPRESS.Master_Final** |
| SQLSourceType | **Direct input** |
| SQLStatement | **TRUNCATE TABLE Liquor** |




Having successfully imported uploaded the dataset on MS SQL, we proceed by replacing the empty values of the rows that detected in the Data Cleaning step with NULL.

# Creation of dimension table and fact table

In order to create the appropriate dimension tables first we checked which columns are categorical and thus will be used as dimensions and which ones are measures and will be part of the fact table. The following columns were used to create the dimension tables:

- Store_Name
- Address
- City
- County
- Zip_code
- Category
- Item_description
- Vendor_name
- Date

While the following columns were included on the fact table:

- Invvoice_and_item_number
- Pack
- Bottle_volume_ml
- State_bottle_cost
- State_bottle_retail
- Bottles_sold
- Sale_dollars
- Volume_sold_liters

The column “Invvoice_and_item_number” is unique for each line so it was used as a primary key.

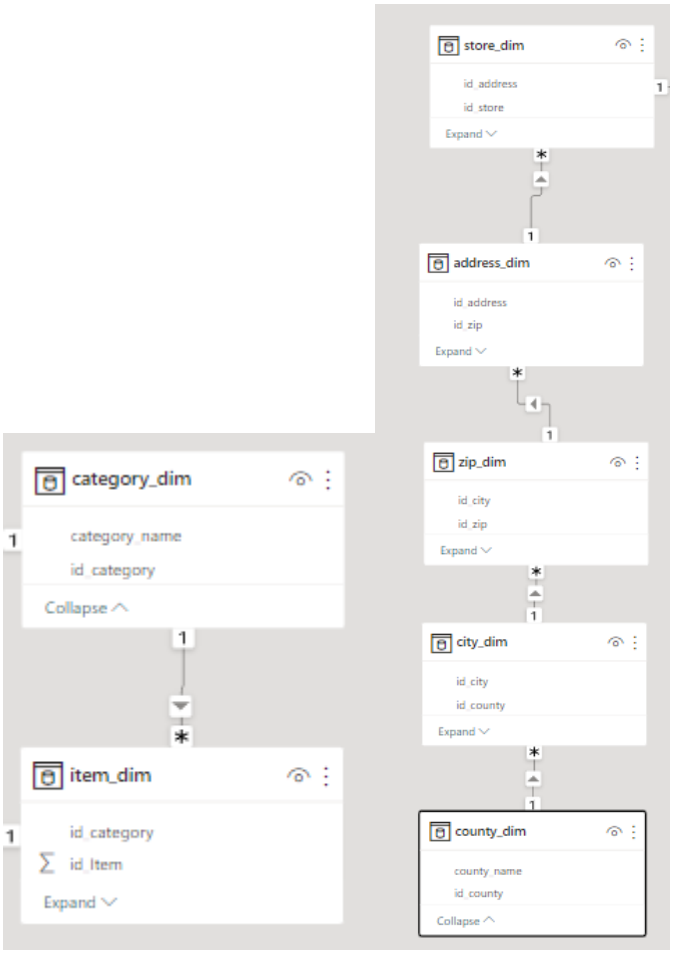
Many of the columns that were used for the dimension tables already had some corresponsive unique numerical columns. However, many of these values seemed to not follow a strict numerical order. For example, column “item_number” did not begin its numbering from 1.

Moreover, there seemed to be values that do not correspond to a unique number. For example, the value "Fireball Cinnamon Whiskey Party Cooler" from the “item” column had a “item_number” of both 101046 and 101047.

For the two above mentioned reasons and in order to simplify the unique id assignment, we created a new mapping for each label when creating the dimension table.

There was also a level of cardinality between the above-mentioned columns, which lead us to create several levels of relationship with a parent-child relationship between the tables.

The parent/child relationship between the dimension tables looks like this:




Before proceeding with the creating of the dimension tables, we made a check whether each child column had a unique relationship between the parent columns. For example, to check whether for all the available rows of each city only belongs to one county etc.

We found out some values that do not abide by the one-to-many relationship. However, that was a very small percentage compared the overall data, and we could proceed with the creation of the parent-child relationship despite them. These values were replaced with NULL as we did not really know to which parent id they corresponded to.

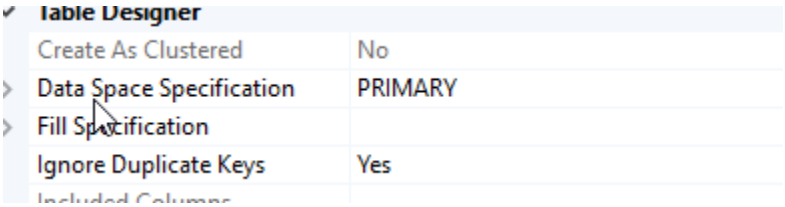
We then proceeded by creating the dimension tables. For each table we defined an ID column as int which was set as primary key and it was the unique number identifier for each label. We also defined a second column as nchar(255) as the label. We did not allow null values in our dimension tables for the ID and Label columns.

| Column Name | Data Type | Allow Nulls |
|---|---|---|
| id_category | int | ☐ |
| category_name | nchar(255) | ☐ |

For the primary key, we set it as identity with increment of one.

| | |
|---|---|
| Has Non-SQL Server Subscriber | No |
| Identity Specification | Yes |
| (Is Identity) | Yes |
| Identity Increment | 1 |
| Identity Seed | 1 |

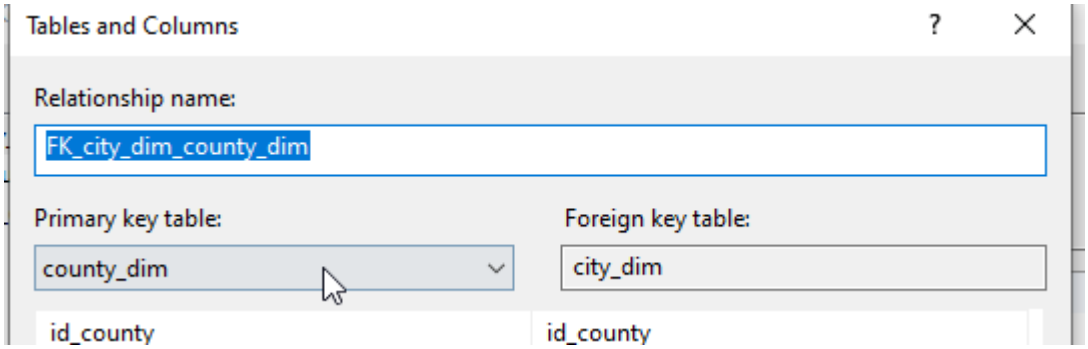
For the label column we created an index and set it to ignore any duplicate keys, in order to avoid duplicate value inserts.




For child tables, we also added a third int type column which is the foreign key that is connected with the parent's table primary key.

| Column Name | Data Type | Allow Nulls |
|---|---|---|
| id_city | int | ☐ |
| city_name | nchar(255) | ☐ |
| id_county | int | ☑ |
| | | ☐ |

From the create relations option, we set it as a foreign key and linked it with the corresponsive primary key.




For the date dimension we also added as a separate int columns the year, month and day of each date.

| Column Name | Data Type | Allow Nulls |
|---|---|---|
| DATE | date | ☐ |
| YEAR | int | ☐ |
| MONTH | int | ☐ |
| DAY | int | ☐ |
| | | ☐ |

After creating the dimension tables, we then wrote the queries to insert the values into each table from the original liquor table.

```
insert into city_dim(city_name,id_county)
select distinct L.city, C.id_county  from Liquor as L  left  join   county_dim as C on  L.county=C.county_name
```

Having created the dimension tables as well, we proceeded by creating the fact table. On the fact table apart from the measures and invoice column as primary key, we added the child dimension tables as well

We then set the appropriate foreign keys between the fact table and the dimension tables.

| | Column Name | Data Type | Allow Nulls |
|---|---|---|---|
| 🔑 | Invoice | nchar(255) | ☐ |
| | Date | date | ☑ |
| | Store | int | ☑ |
| | Ventor | int | ☑ |
| | Item | int | ☑ |
| | Pack | int | ☑ |
| | Bottle_volume_ml | float | ☑ |
| | State_bottle_cost | float | ☑ |
| | State_bottle_retail | float | ☑ |
| | Bottles_sold | float | ☑ |
| | Sales_dollars | float | ☑ |
| | Volume_sold_liters | float | ☑ |
| | Volume_sold_gallons | float | ☑ |

Having created the fact table, we insert the values into the table with the below query.

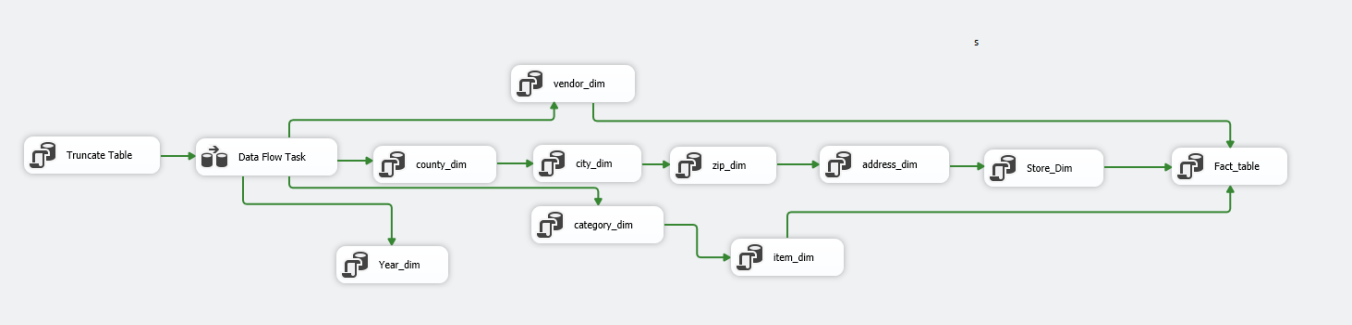
```
insert into Liquer_fact select
invoice_and_item_number,date_dim.date,store_dim.id_store,vendor_dim.vendor_id,id_Item,pack,bottle_volume_ml,state_bottle_cost,state_bottle_retail,
bottles_sold,sale_dollars,volume_sold_liters,volume_sold_gallons
from Liquer_v2

inner join store_dim  on store_dim.store_name=Liquer_v2.store_name
inner join item_dim  on Liquer_v2.item_description=item_dim.item_name
inner join date_dim  on date_dim.date=Liquer_v2.date
inner join vendor_dim  on Liquer_v2.vendor_name=vendor_dim.vendor_name
```

We also added three additional columns which would serve as measures for the final cube. These columns are the following:

- Total_Revenue= Bottles_sold*State_bottle_retail
- Total_Cost= Bottles_sold*State_bottle_cost
- Total_Profit= Total_Revenue- Total_Cost

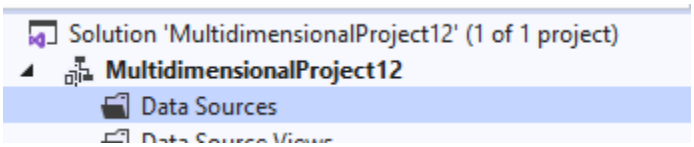
After updating the fact table as well, we recreated the data inserting process of each table on SSIS with Execute SQL Task. The flow of the importing process is depicted below:

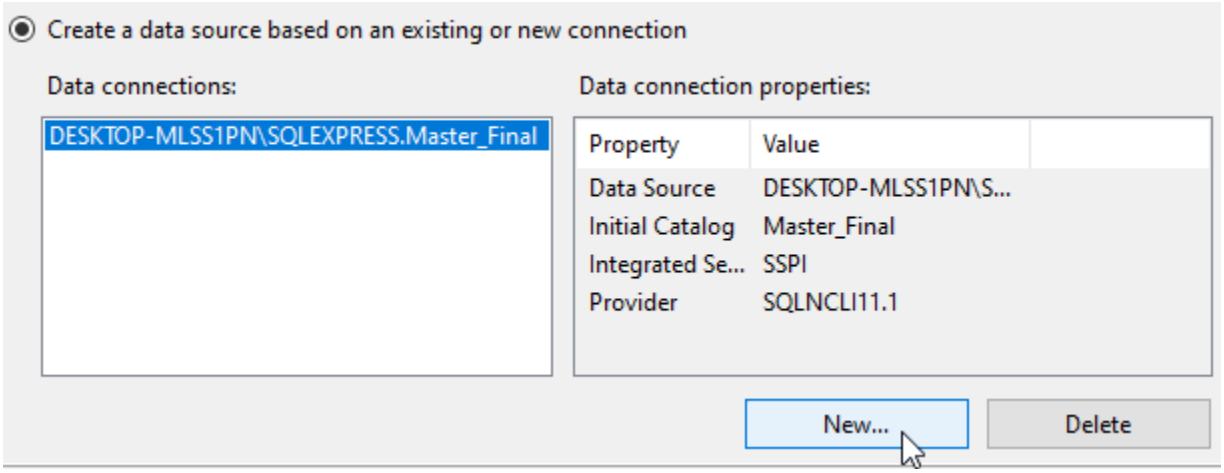


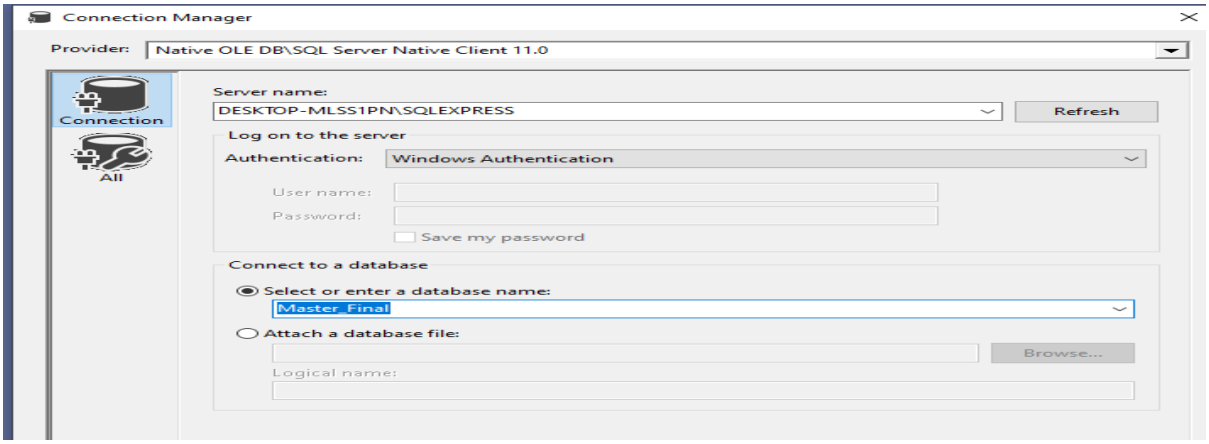

# Cube Creation

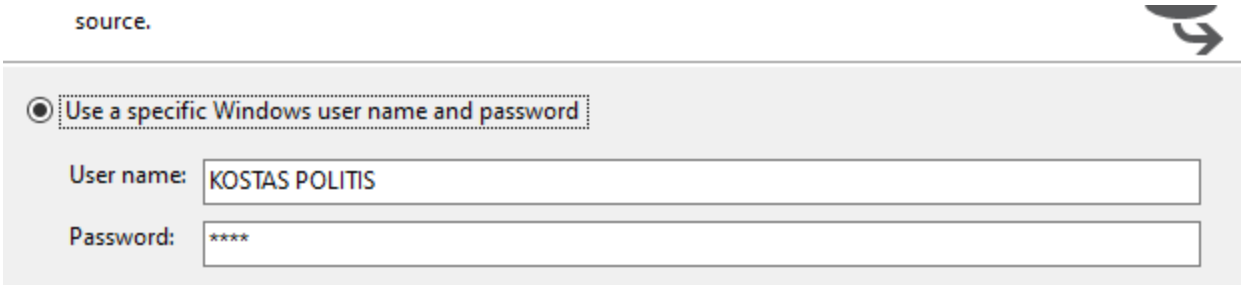
Having created both the fact table and the dimension tableσ we then proceeded by creating the cube using SSAS.

First, we set a data source.

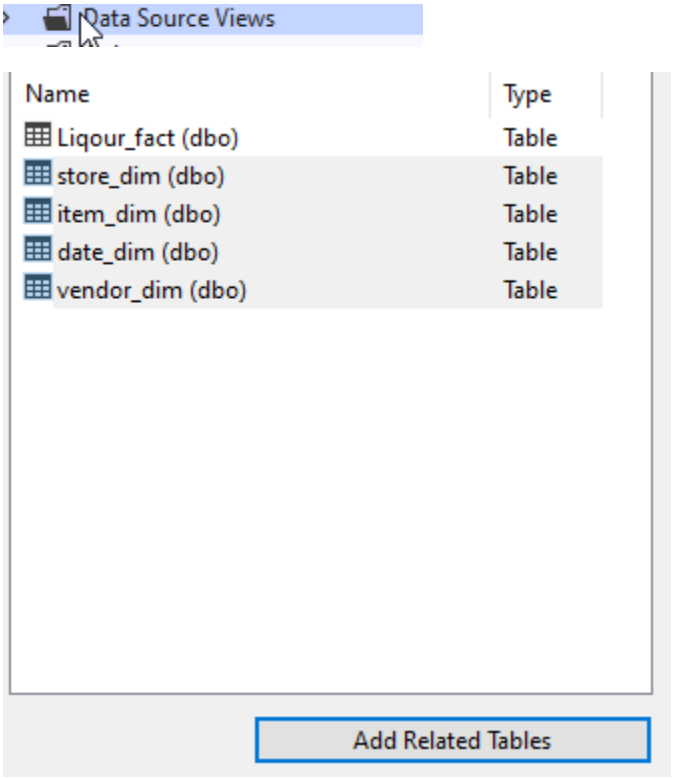


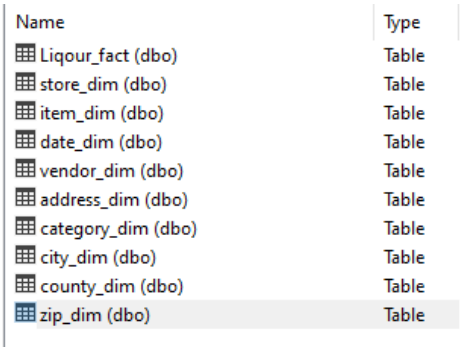



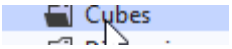



After establishing the data source, we then set the data source views.

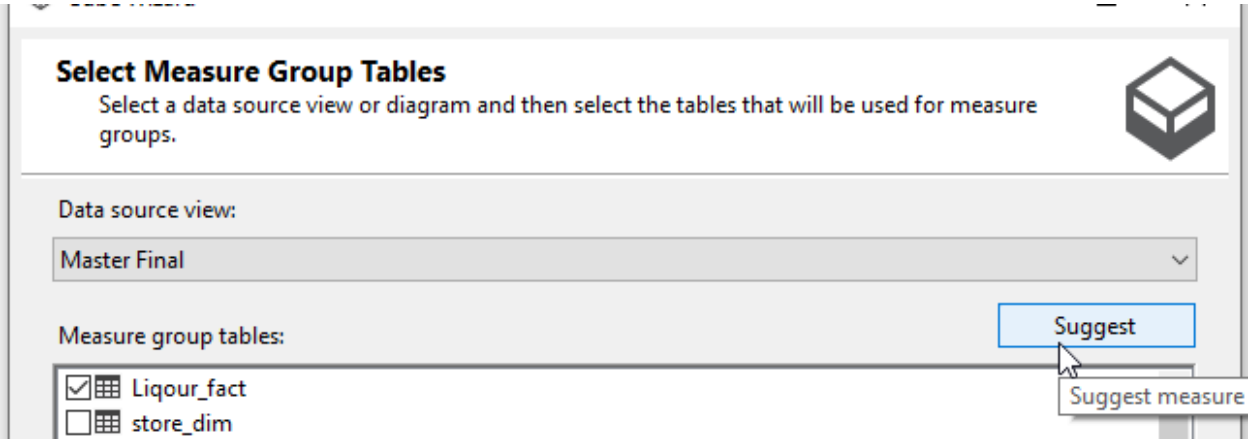



First, we added the fact table and clicked on “Add Related Tables”. However, it only added the dimension tables that are directly related to the fact table while missing their parent tables such as county, category etc. Thus, we added the rest of the parent dimension tables manually.

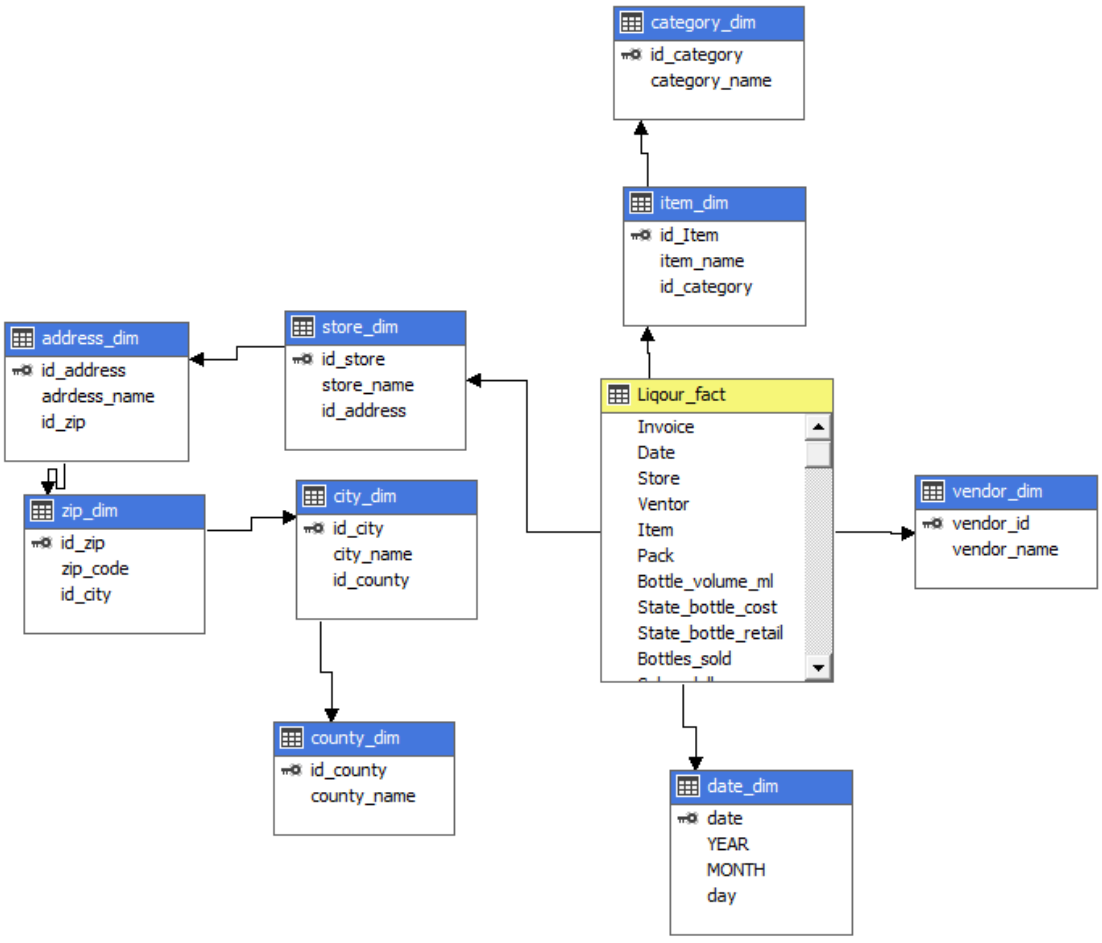


The next step was to create the cube. 

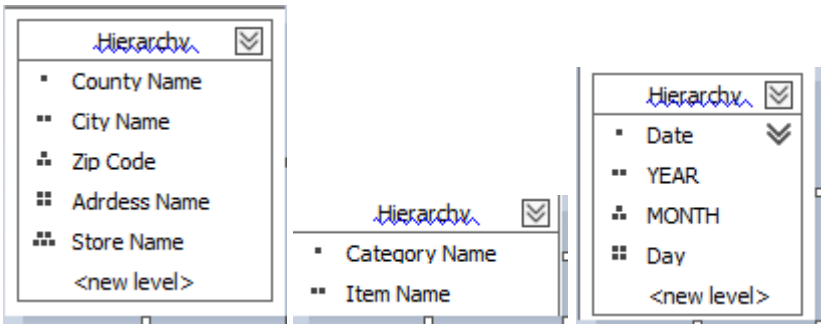

By using the uploaded tables, it automatically recognized and selected our fact table.



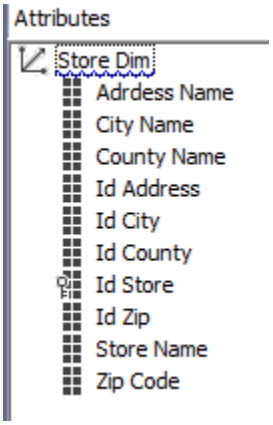

By following the steps until the end, the cube was created as below:



Before processing the cube, we made some final adjustments. For the dimensions store_dim, date_dim and item_dim we set the properly the hierarchy between the parent and child dimensions.

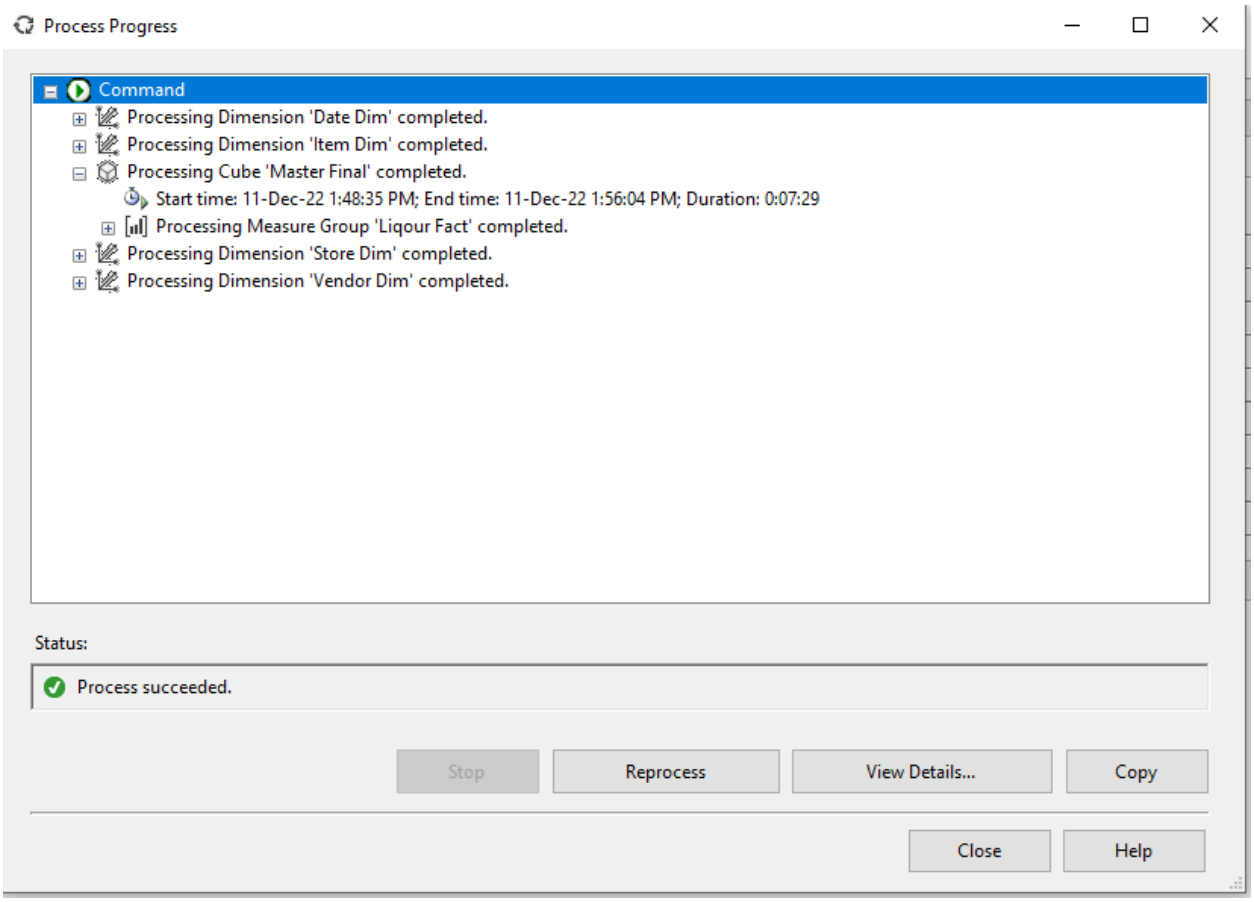



Moreover, for all the dimensions, we added the label names on the Attributes menu.



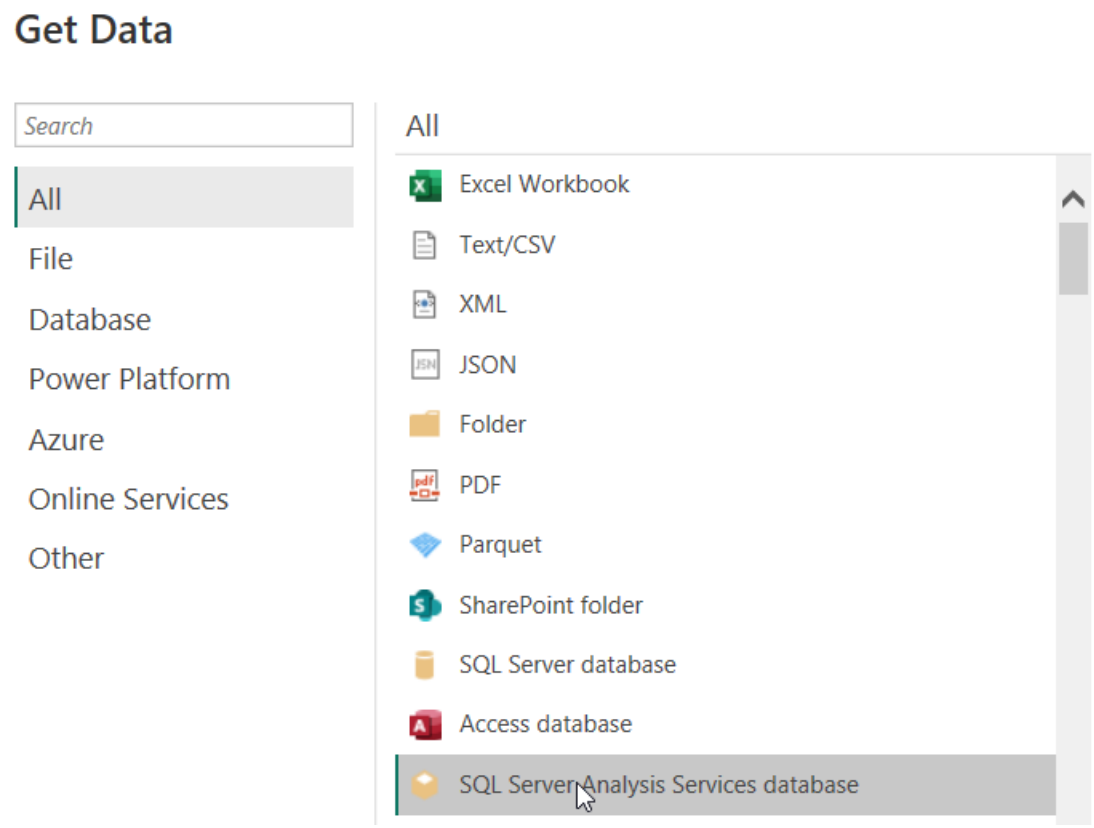

After making the final adjustments to our cube, we selected process in order to continue with the calculations.

After the calculations had been completed, we proceeded with uploading the cube on Power BI and visualizing the data.

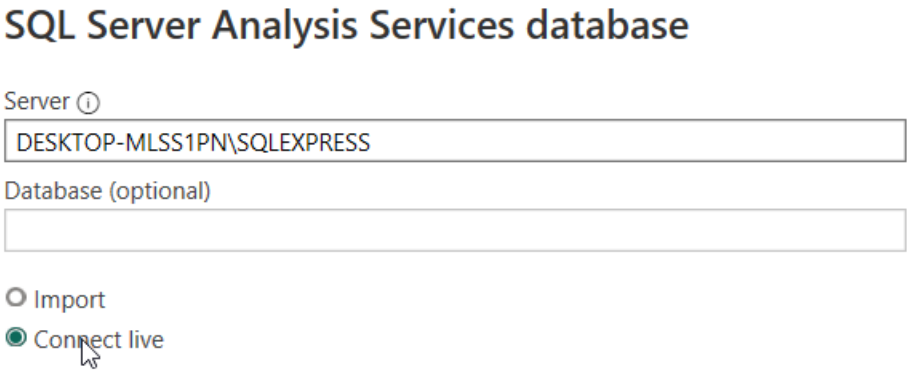

Process Progress
Command
Processing Dimension 'Date Dim' completed.
Processing Dimension 'Item Dim' completed.
Processing Cube 'Master Final' completed.
Start time: 11-Dec-22 1:48:35 PM; End time: 11-Dec-22 1:56:04 PM; Duration: 0:07:29
Processing Measure Group 'Liqour Fact' completed.
Processing Dimension 'Store Dim' completed.
Processing Dimension 'Vendor Dim' completed.
Status:
Process succeeded.
Stop
Reprocess
View Details...
Copy
Close
Help

# Power BI

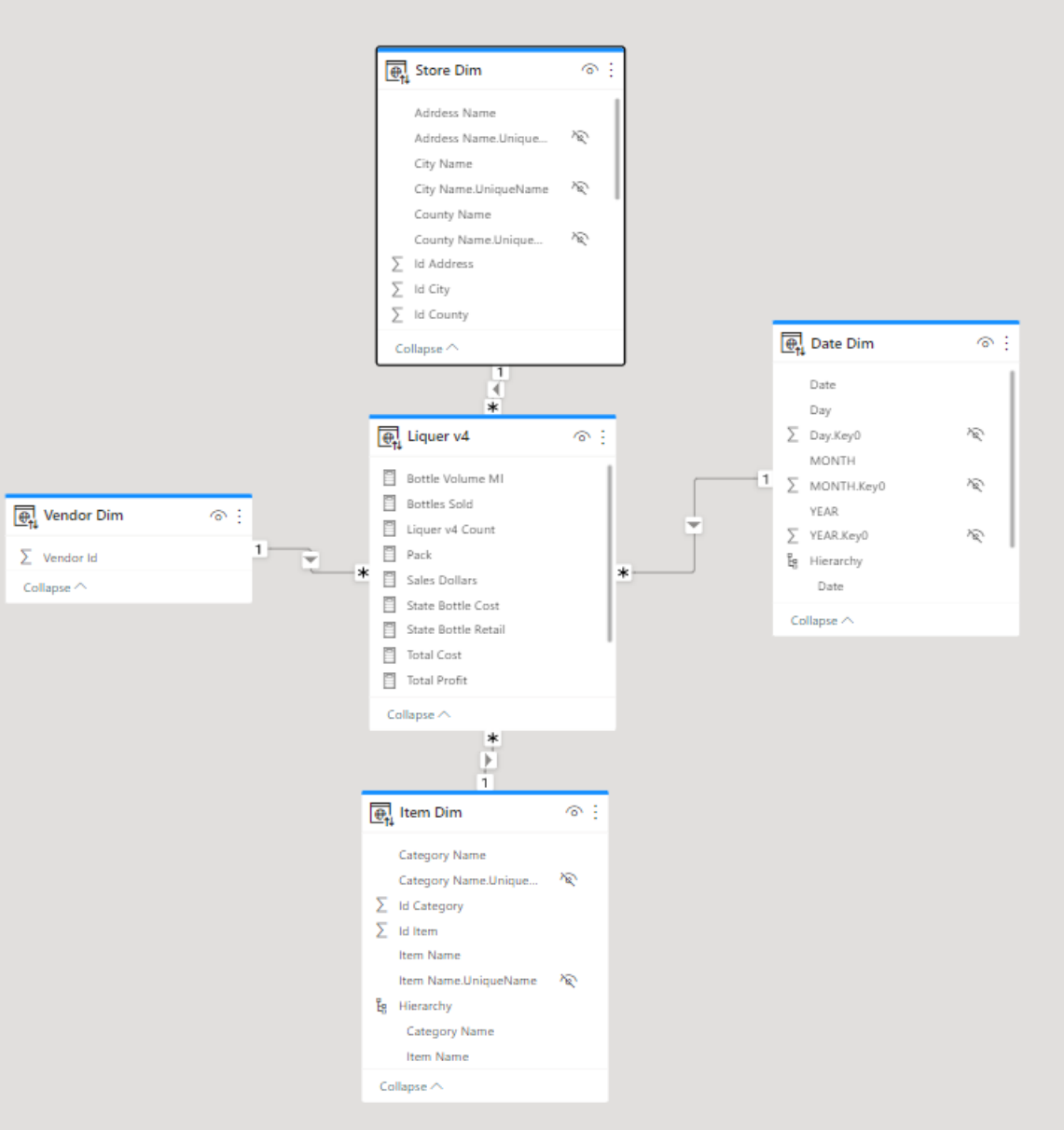
We selected the get data option and the 'SQL Server Analysis Services Database".




We selected the Connect live option and clicked on ok.

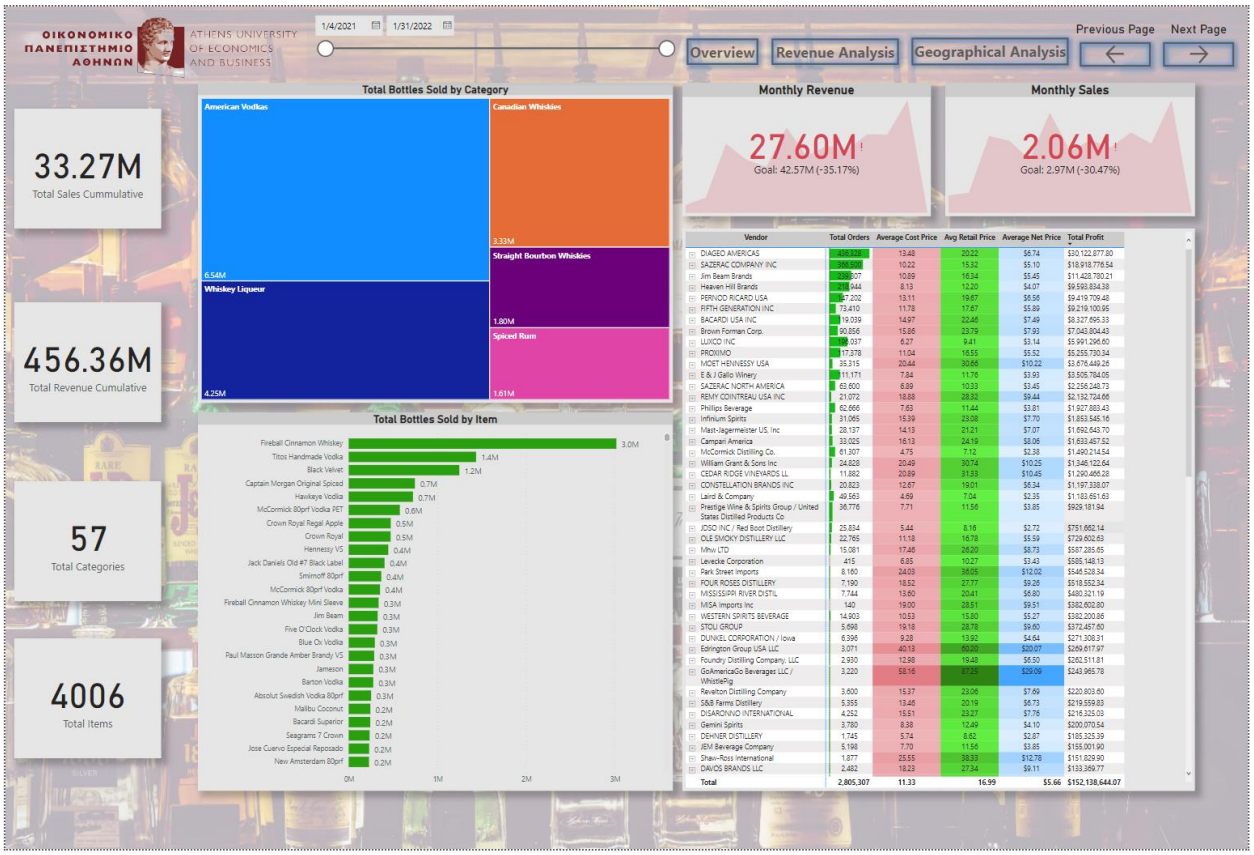



When the cube was uploaded, the relationship schema on Power Bi looks like below:

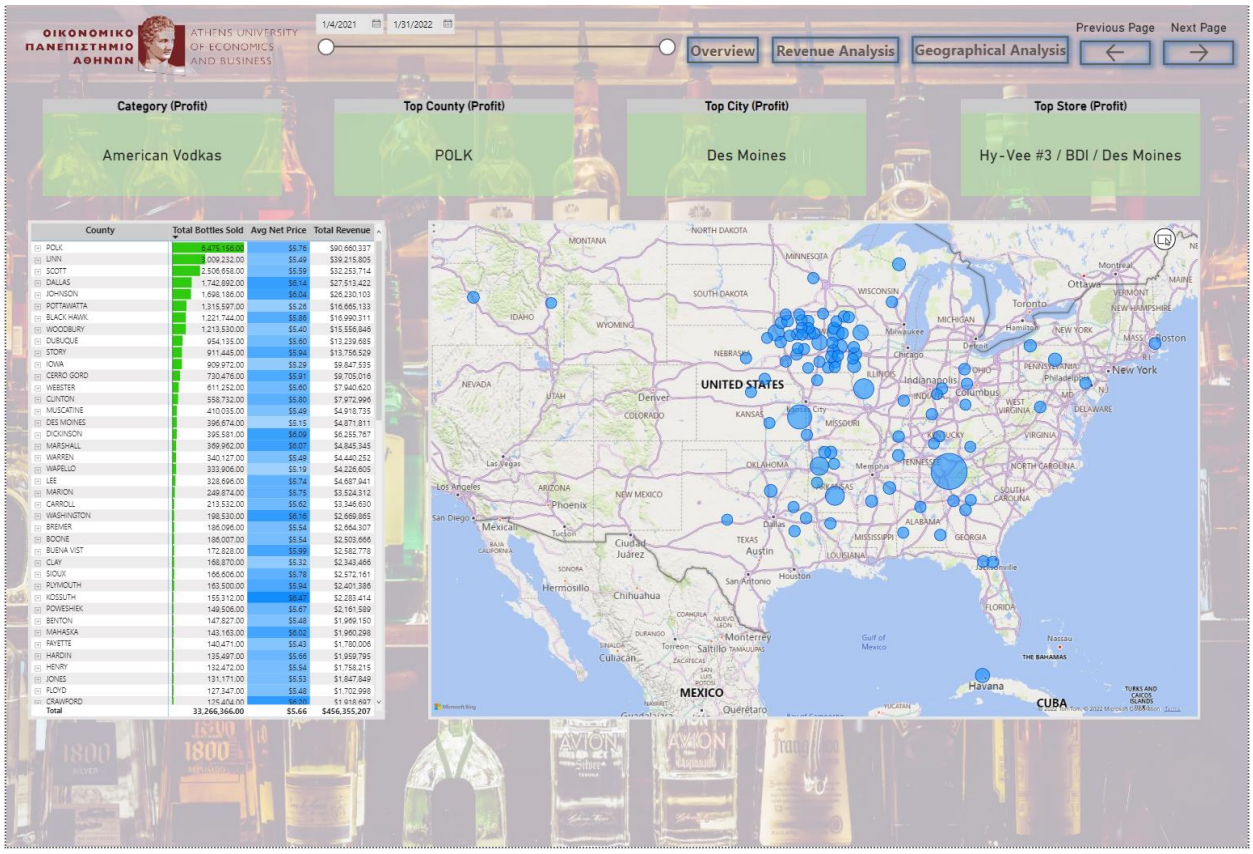



We also used the import option in order to be able to create additional measures within the Power Bi.

By adding the appropriate visualizations, we created two reporting pages that looks like the screenshots below:




ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
1/4/2021
1/31/2022
Overview
Revenue Analysis
Geographical Analysis
Previous Page
Next Page
Date
Item
Category
Total Orders and Target Orders
2,805K
Total Revenue and Target Revenue
456.36M
Weekly Profit vs Adjusted Profit
Total Profit
Adjusted Profit
Price Adjustment %
$16.99
Avg Retail Price
$18.69
Adjusted Price
Weekly Profit
Year
Total Profit vs Total Cost
Total Revenue
Sum of Total_Profit
Total Cost
100% Agave Tequila
68,545,125 Total Volume
$8,460,981.67 Total Profit
876,172.00 Total Bottles Sold
$13,798,761.995 Adjusted Profit
Aged Dark Rum
7,474,900 Total Volume
$384,815.25 Total Profit
55,645.00 Total Bottles Sold
$627,156.2291 Adjusted Profit
American Brandies
52,223,600 Total Volume
$1,877,427.37 Total Profit
845,632.00 Total Bottles Sold
$3,058,422.4946 Adjusted Profit
American Cordials & Liqueur
37,139,375 Total Volume
$1,515,771.80 Total Profit
523,867.00 Total Bottles Sold
$2,470,236.826 Adjusted Profit
American Cordials & Liqueurs
257,250 Total Volume
$11,236.48 Total Profit
2,704.00 Total Bottles Sold
$18,315.4624 Adjusted Profit
American Distilled Spirit Specialty
8,293,925 Total Volume
$524,462.82 Total Profit


ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
1/4/2021
1/31/2022
Overview
Revenue Analysis
Geographical Analysis
Previous Page
Next Page
Category (Profit)
American Vodkas
Top County (Profit)
POLK
Top City (Profit)
Des Moines
Top Store (Profit)
Hy-Vee #3 / BDI / Des Moines
County
Total Bottles Sold
Avg Net Price
Total Revenue
POLK 8,475,156.00 $5.76 $90,660,337
LINN 3,009,232.00 $5.49 $39,215,805
SCOTT 2,506,658.00 $5.59 $32,253,714
DALLAS 1,742,892.00 $6.14 $27,513,422
JOHNSON 1,698,186.00 $6.04 $26,230,103
POTTAWATTA 1,316,597.00 $5.26 $16,665,133
BLACK HAWK 1,221,744.00 $5.66 $16,990,311
WOODBURY 1,213,530.00 $5.40 $15,556,848
DUBUQUE 954,135.00 $5.60 $13,239,685
STORY 911,445.00 $5.94 $13,756,529
IOWA 909,972.00 $5.29 $9,847,535
CERRO GORD 730,476.00 $5.91 $9,705,016
WEBSTER 611,252.00 $5.80 $7,940,620
CLINTON 558,732.00 $5.80 $7,972,996
MUSCATINE 410,035.00 $5.49 $4,918,735
DES MOINES 396,674.00 $5.15 $4,871,811
DICKINSON 395,581.00 $6.09 $6,255,767
MARSHALL 369,962.00 $6.07 $4,845,345
WARREN 340,127.00 $5.49 $4,440,252
WAPELLO 333,906.00 $5.19 $4,226,605
LEE 328,696.00 $5.74 $4,687,941
MARION 249,874.00 $5.75 $3,524,312
CARROLL 213,532.00 $5.62 $3,346,630
WASHINGTON 198,530.00 $5.76 $2,669,865
BREMER 186,096.00 $5.94 $2,664,307
BOONE 186,007.00 $5.54 $2,503,666
BUENA VIST 172,828.00 $5.99 $2,582,778
CLAY 168,870.00 $5.32 $2,343,466
SIOUX 166,606.00 $5.78 $2,572,161
PLYMOUTH 163,500.00 $5.94 $2,401,386
KOSSUTH 155,312.00 $5.47 $2,283,414
POWESHIEK 149,506.00 $5.67 $2,161,589
BENTON 147,827.00 $5.48 $1,969,150
MAHASKA 143,163.00 $6.02 $1,960,298
FAYETTE 140,471.00 $5.43 $1,780,006
HARDIN 136,497.00 $5.66 $1,959,795
HENRY 132,472.00 $5.94 $1,758,215
JONES 131,171.00 $5.53 $1,847,849
FLOYD 127,347.00 $5.48 $1,702,998
CRAWFORD 126,404.00 $5.66 $1,918,697
Total 33,266,366.00 $5.66 $456,355,207
UNITED STATES
MEXICO
CUBA

Apart from the measures that were already calculated form the cubes, we added some additional ones using the DAX language of the power bi. Some indicative measures are the following:

Prev Month Orders/ Prev Month Revenue

In order to make some month-to-month comparisons we created the above measures using the following DAX code.

```
Prev Month Orders = CALCULATE([Total Orders], DATEADD(date_dim[date],-1,MONTH))
```

```
Prev Month Revenue = CALCULATE([Total Revenue], DATEADD(date_dim[date],-1,MONTH))
```

Target Orders/ Target Revenue

And to create the targeted orders and revenue, based on the above calculation we used the code below.

```
Target Orders = [Prev Month Orders]*1.1
```

```
Target Revenue = [Prev Month Revenue]*1.1
```

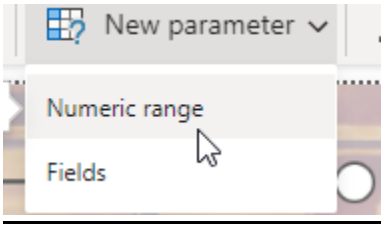
We also wanted to show some scenarios regarding the profitability, by changing the average retail price. For this purpose, first we recalculated the revenue as below.

```
Total Revenue = SUMX(Liquer_v3,Liquer_v3[Bottles_sold]*Liquer_v3[State_bottle_retail])
```

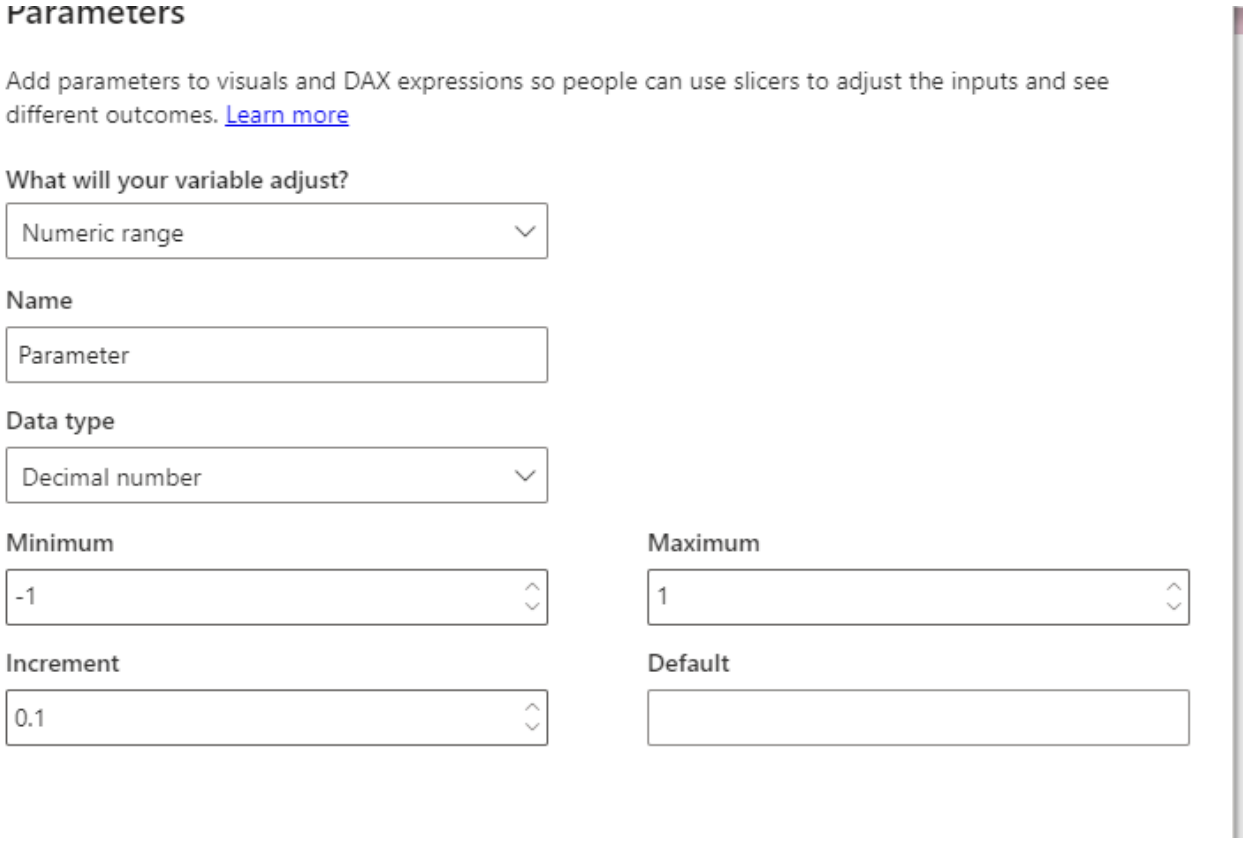
And based on this revenue we recalculated the Total profit etc as below.

```
Profit = [Total Revenue]-[Total Cost]
```

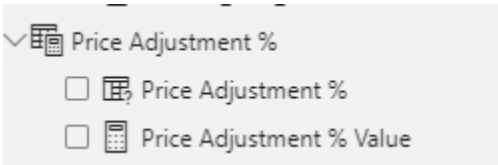
For the price adjustment value, we used the option New Parameter/Numeric Range.




And set the parameters as below.




Thus we created a price adjustment measure which we combined with the rest measures to create the appropriate calculations.




This parameter was used to create an adjusted revenue, which then used to calculate adjusted profit etc.

```
Adjusted Revenue = SUMX(Liquer_v3,Liquer_v3[Bottles_sold]*(1+'Price Adjustment %'[Price Adjustment % Value])*[Adjusted Price])
```

We also created an adjusted price measure.

```
Adjusted Price = [Avg Retail Price]*(1+'Price Adjustment %'[Price Adjustment % Value])
```

We then created a slicer with the price adjustment measure as well as the required graphs as they will be presented below.

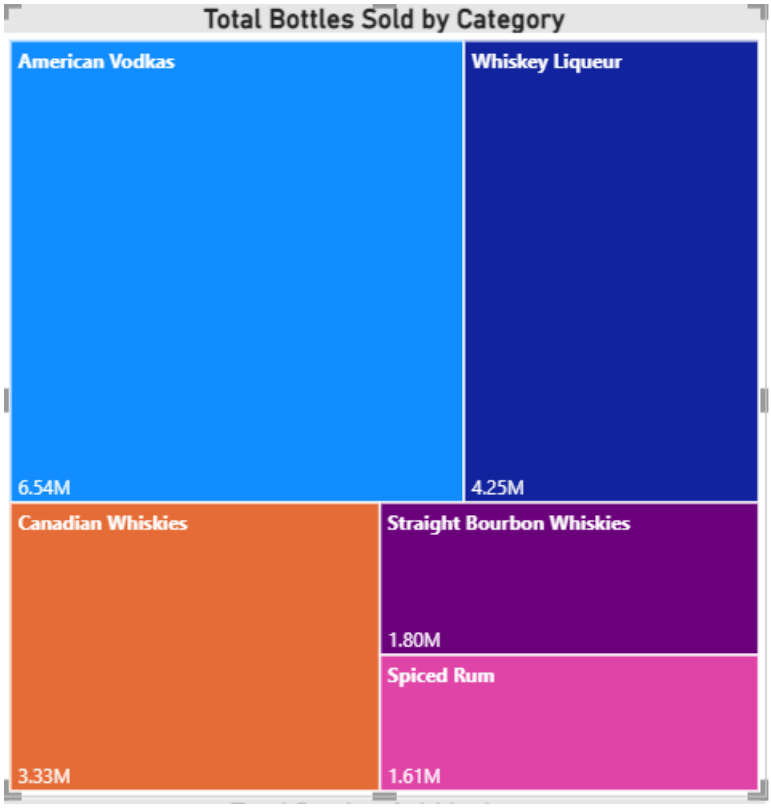
## Which is the most sold category and individual liquor in the United States?

From the below graph, we can see the top 5 liquor categories in terms total bottles sold on the American stores on the last one year. The three dominant categories are:

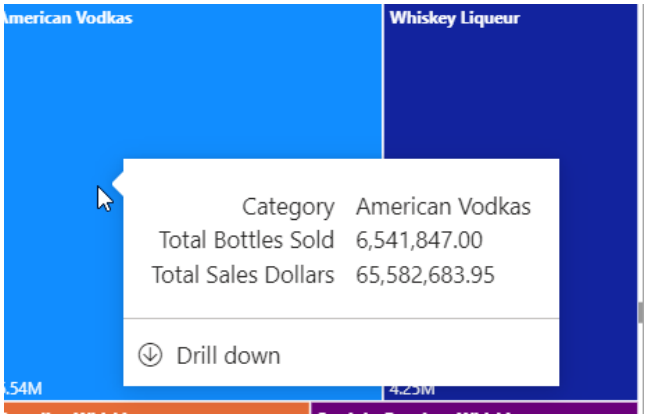
- American Vodkas with 6.54 million bottles sold
- Whiskey Liquor with 4.25 million bottles sold
- Canadian Whiskies with3.33 million bottles sold

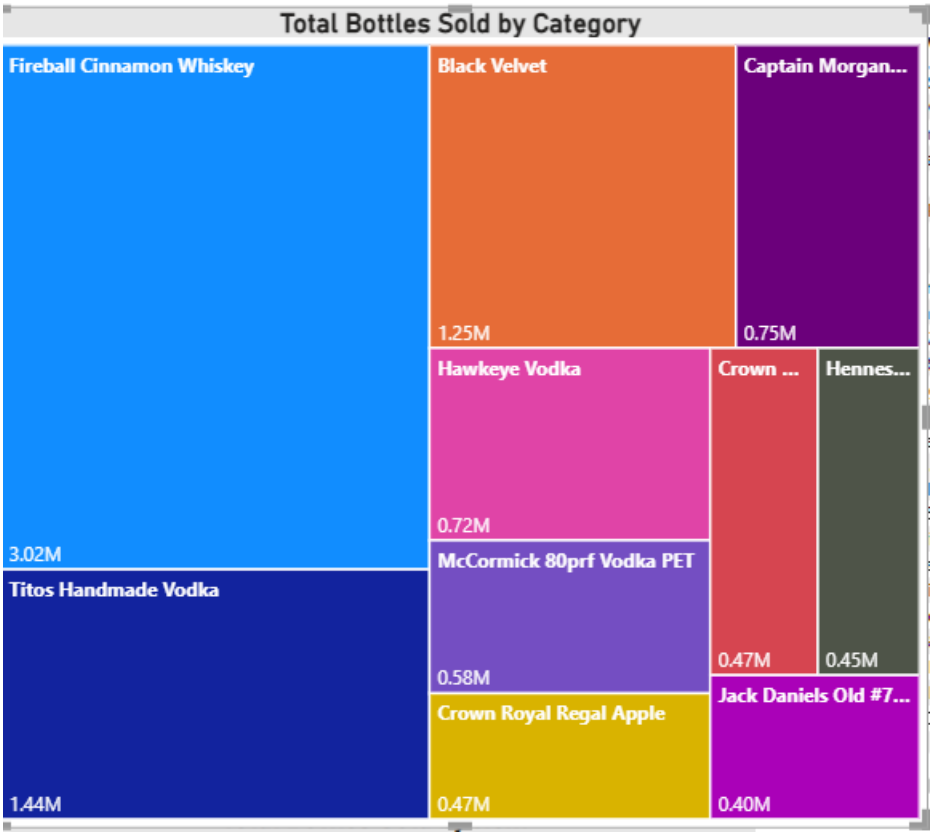
On the fourth and fifth place are Straight Bourbon Whiskies and Spiced Rum which are relatively close regarding their total bottles sold.

We see that even though American Vodkas category has the highest demand, whiskies in general tend to be preferred by American stores, as the three whiskey categories combined make a total of 9.35 million on total bottles sold.



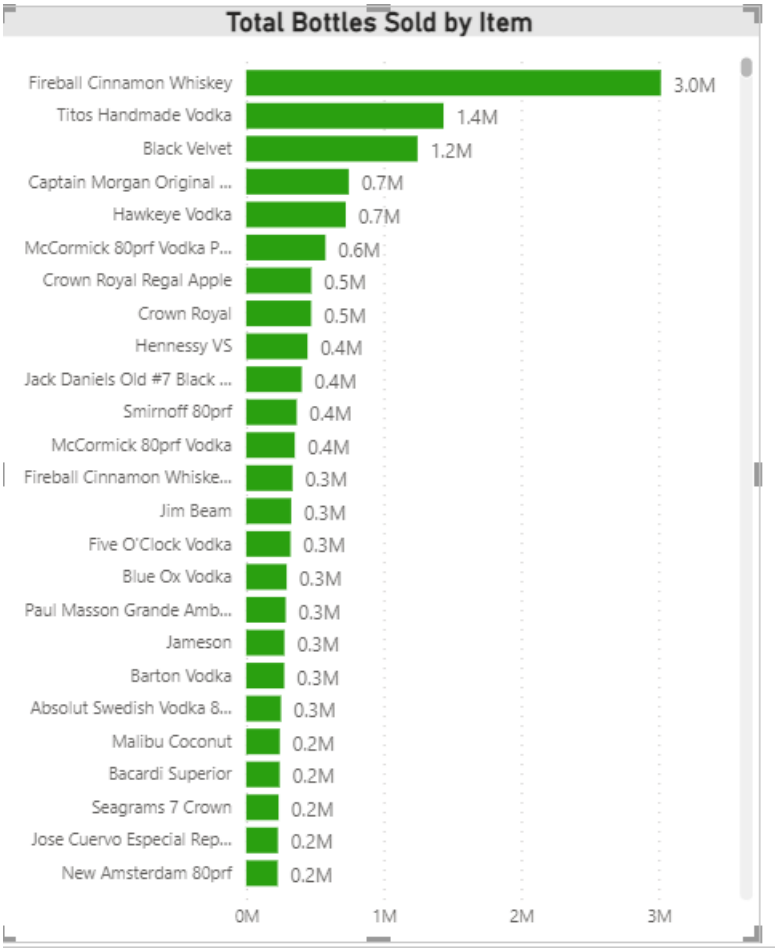

By hovering over each category, we can also see its total sales in dollars as additional information.




What is even more interested, is that on item level the top 2 liquors are Fireball Cinnamon Whiskey and Black Velvet which are both whiskeys. Thus, even though American Vodkas in general is most sold liquor category in USA, the most wanted liquors are Whiskeys.

This can also be confirmed on the below graph where the total bottles sold for each item is depicted.




The Iowa vendor association could follow two strategies based on the above.

- Because vodkas and whiskeys are the dominant categories, a sudden shift on demand for either of the two could lead to potential problems for the vendors. Thus, they could try to differentiate their portfolio by trying to promote other liquor categories as well.
- On the other hand, because whiskeys are preferred, they could focus on whiskey categories even more and claim a bigger market share on this category.

## Which vendor is the most profitable? Which is each vendor's most profitable liquor category in terms of sales?

From the graph below we can see that the two top vendors by total orders are DIAGEO AMERICAS and SAZERAC COMPANY INC.

The difference between the vendors is also shown in Total Profits. On the same graph we can also see the average cost for each vendor as well as the average retail price which they sell their liquors on stores.

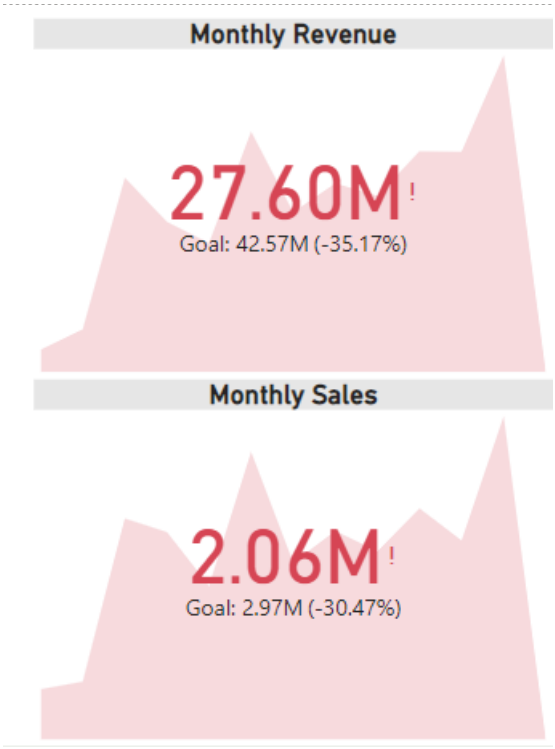
The top 5 vendors seem to dominate the market and create an oligopoly. Moreover, even though the average net price that they secure is not the highest among the competitors, due their big number of orders they manage to create the most profit.

| Vendor | Total Orders | Average Cost Price | Avg Retail Price | Average Net Price | Total Profit |
|---|---|---|---|---|---|
| DIAGEO AMERICAS | 436,328 | 13.48 | 20.22 | $6.74 | $30,122,877.80 |
| SAZERAC COMPANY INC | 366,500 | 10.22 | 15.32 | $5.10 | $18,918,776.54 |
| Jim Beam Brands | 239,307 | 10.89 | 16.34 | $5.45 | $11,428,780.21 |
| Heaven Hill Brands | 218,944 | 8.13 | 12.20 | $4.07 | $9,593,834.38 |
| LUXCO INC | 196,037 | 6.27 | 9.41 | $3.14 | $5,991,296.60 |
| PERNOD RICARD USA | 147,202 | 13.11 | 19.67 | $6.56 | $9,419,709.48 |
| BACARDI USA INC | 119,039 | 14.97 | 22.46 | $7.49 | $8,327,695.33 |
| PROXIMO | 117,378 | 11.04 | 16.55 | $5.52 | $5,255,730.34 |
| E & J Gallo Winery | 111,171 | 7.84 | 11.76 | $3.93 | $3,505,784.05 |
| Brown Forman Corp. | 90,856 | 15.86 | 23.79 | $7.93 | $7,043,804.43 |
| FIFTH GENERATION INC | 73,410 | 11.78 | 17.67 | $5.89 | $9,219,100.95 |
| SAZERAC NORTH AMERICA | 63,600 | 6.89 | 10.33 | $3.45 | $2,256,248.73 |
| Phillips Beverage | 62,666 | 7.63 | 11.44 | $3.81 | $1,927,883.43 |
| McCormick Distilling Co. | 61,307 | 4.75 | 7.12 | $2.38 | $1,490,214.54 |
| Laird & Company | 49,563 | 4.69 | 7.04 | $2.35 | $1,183,651.63 |
| Prestige Wine & Spirits Group / United States Distilled Products Co | 36,776 | 7.71 | 11.56 | $3.85 | $929,181.94 |
| MOET HENNESSY USA | 35,315 | 20.44 | 30.66 | $10.22 | $3,676,449.26 |
| Campari America | 33,025 | 16.13 | 24.19 | $8.06 | $1,633,457.52 |
| Infinium Spirits | 31,065 | 15.39 | 23.08 | $7.70 | $1,853,545.16 |
| Mast-Jagermeister US, Inc | 28,137 | 14.13 | 21.21 | $7.07 | $1,692,643.70 |
| JDSO INC / Red Boot Distillery | 25,834 | 5.44 | 8.16 | $2.72 | $751,662.14 |
| William Grant & Sons Inc | 24,828 | 20.49 | 30.74 | $10.25 | $1,346,122.64 |
| **Total** | **2,805,307** | **11.33** | **16.99** | **$5.66** | **$152,138,644.07** |

On a drilldown level, we can see the best seller for the top vendor is Canadian Whiskies. The same goes for all the top 5 vendors. Thus, vendors who focus more on whiskey selling seems to have more profitability.

| Vendor | Total Orders | Average Cost Price | Avg Retail Price | Average Net Price | Total Profit |
|---|---|---|---|---|---|
| **DIAGEO AMERICAS** | **436,328** | **13.48** | **20.22** | **$6.74** | **$30,122,877.80** |
| Canadian Whiskies | 103,106 | 15.32 | 22.98 | $7.66 | $8,887,877.97 |
| Spiced Rum | 72,082 | 9.91 | 14.87 | $4.96 | $6,441,968.09 |
| American Vodkas | 50,709 | 8.32 | 12.49 | $4.17 | $2,303,712.36 |
| Blended Whiskies | 32,348 | 8.22 | 12.33 | $4.11 | $1,065,102.26 |
| American Flavored Vodka | 31,706 | 8.51 | 12.77 | $4.26 | $1,050,124.51 |
| Scotch Whiskies | 17,350 | 26.25 | 39.38 | $13.13 | $1,116,760.54 |
| Temporary & Specialty Packages | 16,873 | 16.00 | 24.01 | $8.01 | $1,225,523.65 |
| Cream Liqueurs | 16,663 | 19.30 | 28.96 | $9.65 | $1,449,311.46 |

We have also added a comparison between the current revenue and the previous month revenue of the vendors in order to monitor the profitability each month. We can see that there is a huge decrease both in revenue and in total sales on the latest month. Thus, actions must be made in order to improve the total profit of the next periods.

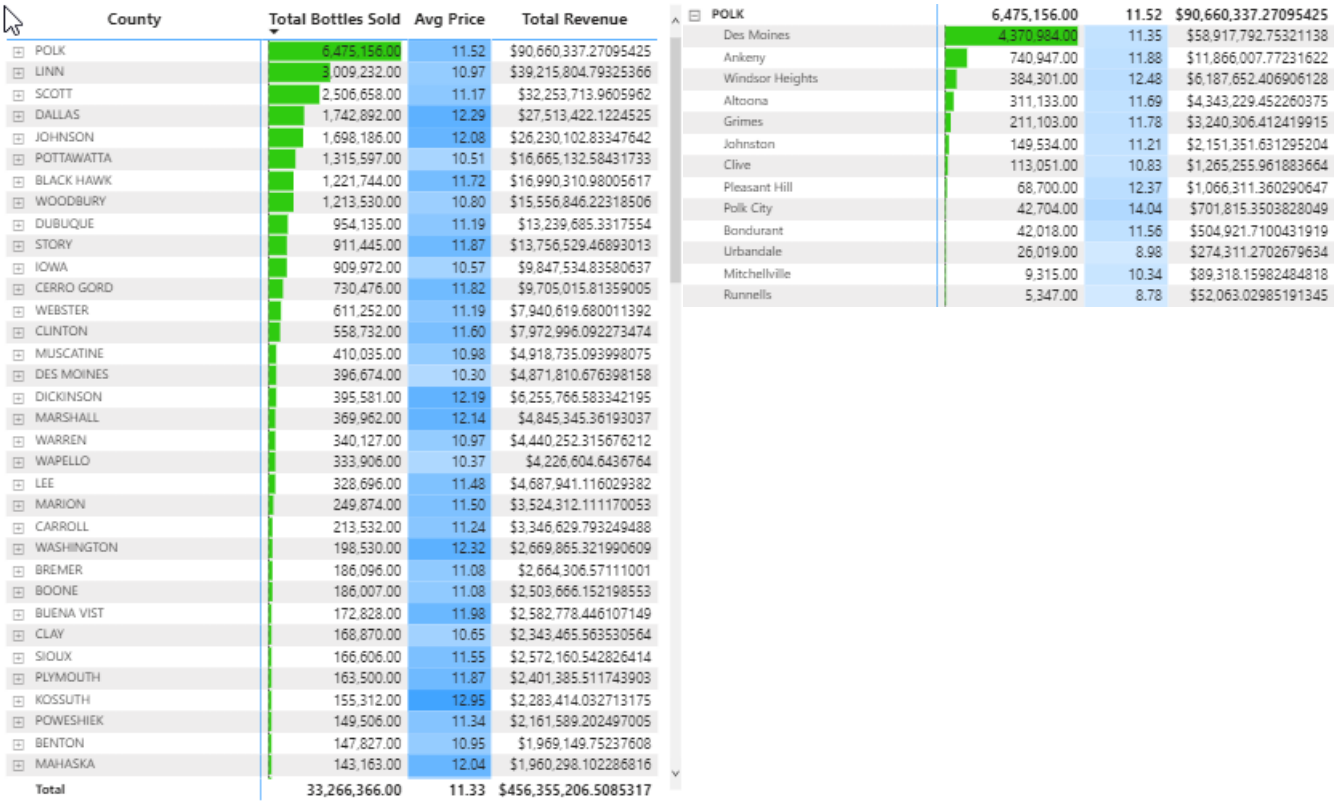


## Which is the most profitable county and city for Iowa vendor association in terms of revenue and bottles sold??

From the graph below we get an overall picture regarding the total bottles that are sold to each county as well as the total revenue that is produced from that county.

The top 3 counties make up the 1/3 of the overall bottles of liquor sold throughout the USA. Thus, the Iowa vendors has a very huge exposure to the liquor demand of these three counties. A sudden change on the preferences of the consumers of these counties, could affect the profitability of Iowa vendors to a greater degree.

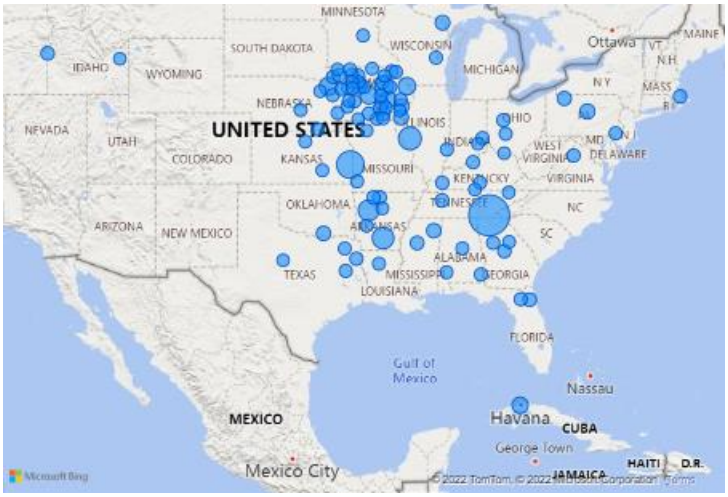
On a second level, we can see the total import detail for each city of each county. The most profitable city in matters of both revenue and total bottles sold is De Moines of the county Polk.



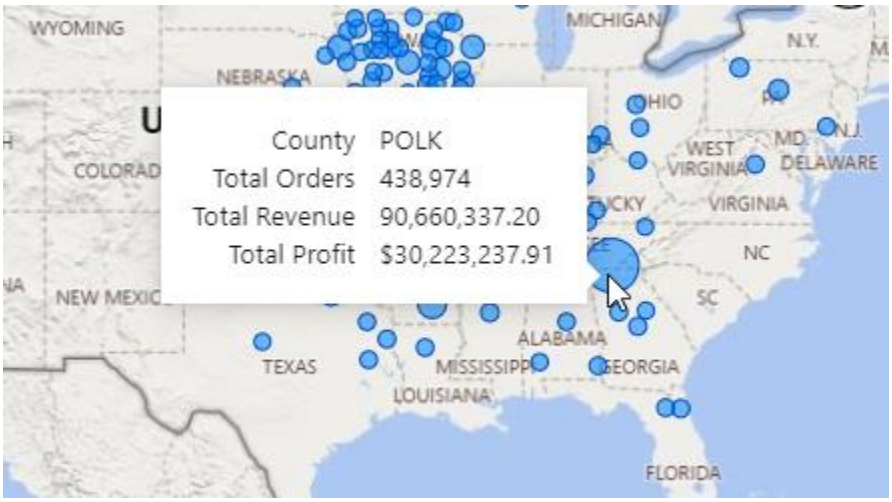
| County | Total Bottles Sold | Avg Price | Total Revenue |
|---|---|---|---|
| POLK | 6,475,156.00 | 11.52 | $90,660,337.27095425 |
| LINN | 3,009,232.00 | 10.97 | $39,215,804.79325366 |
| SCOTT | 2,506,658.00 | 11.17 | $32,253,713.9605962 |
| DALLAS | 1,742,892.00 | 12.29 | $27,513,422.1224525 |
| JOHNSON | 1,698,186.00 | 12.08 | $26,230,102.83347642 |
| POTTAWATTA | 1,315,597.00 | 10.51 | $16,665,132.58431733 |
| BLACK HAWK | 1,221,744.00 | 11.72 | $16,990,310.98005617 |
| WOODBURY | 1,213,530.00 | 10.80 | $15,556,846.22318506 |
| DUBUQUE | 954,135.00 | 11.19 | $13,239,685.3317554 |
| STORY | 911,445.00 | 11.87 | $13,756,529.46893013 |
| IOWA | 909,972.00 | 10.57 | $9,847,534.83580637 |
| CERRO GORD | 730,476.00 | 11.82 | $9,705,015.81359005 |
| WEBSTER | 611,252.00 | 11.19 | $7,940,619.680011392 |
| CLINTON | 558,732.00 | 11.60 | $7,972,996.092273474 |
| MUSCATINE | 410,035.00 | 10.98 | $4,918,735.093998075 |
| DES MOINES | 396,674.00 | 10.30 | $4,871,810.676398158 |
| DICKINSON | 395,581.00 | 12.19 | $6,255,766.583342195 |
| MARSHALL | 369,962.00 | 12.14 | $4,845,345.36193037 |
| WARREN | 340,127.00 | 10.97 | $4,440,252.315676212 |
| WAPELLO | 333,906.00 | 10.37 | $4,226,604.6436764 |
| LEE | 328,696.00 | 11.48 | $4,687,941.116029382 |
| MARION | 249,874.00 | 11.50 | $3,524,312.111170053 |
| CARROLL | 213,532.00 | 11.24 | $3,346,629.793249488 |
| WASHINGTON | 198,530.00 | 12.32 | $2,669,865.321990609 |
| BREMER | 186,096.00 | 11.08 | $2,664,306.57111001 |
| BOONE | 186,007.00 | 11.08 | $2,503,666.152198553 |
| BUENA VIST | 172,828.00 | 11.98 | $2,582,778.446107149 |
| CLAY | 168,870.00 | 10.65 | $2,343,465.563530564 |
| SIOUX | 166,606.00 | 11.55 | $2,572,160.542826414 |
| PLYMOUTH | 163,500.00 | 11.87 | $2,401,385.511743903 |
| KOSSUTH | 155,312.00 | 12.95 | $2,283,414.032713175 |
| POWESHIEK | 149,506.00 | 11.34 | $2,161,589.202497005 |
| BENTON | 147,827.00 | 10.95 | $1,969,149.75237608 |
| MAHASKA | 143,163.00 | 12.04 | $1,960,298.102286816 |
| **Total** | **33,266,366.00** | **11.33** | **$456,355,206.5085317** |

| | | | |
|---|---|---|---|
| POLK | 6,475,156.00 | 11.52 | $90,660,337.27095425 |
| Des Moines | 4,370,984.00 | 11.35 | $58,917,792.75321138 |
| Ankeny | 740,947.00 | 11.88 | $11,866,007.77231622 |
| Windsor Heights | 384,301.00 | 12.48 | $6,187,652.406906128 |
| Altoona | 311,133.00 | 11.69 | $4,343,229.452260375 |
| Grimes | 211,103.00 | 11.78 | $3,240,306.412419915 |
| Johnston | 149,534.00 | 11.21 | $2,151,351.631295204 |
| Clive | 113,051.00 | 10.83 | $1,265,255.961883664 |
| Pleasant Hill | 68,700.00 | 12.37 | $1,066,311.360290647 |
| Polk City | 42,704.00 | 14.04 | $701,815.3503828049 |
| Bondurant | 42,018.00 | 11.56 | $504,921.7100431919 |
| Urbandale | 26,019.00 | 8.98 | $274,311.2702679634 |
| Mitchellville | 9,315.00 | 10.34 | $89,318.15982484818 |
| Runnells | 5,347.00 | 8.78 | $52,063.02985191345 |

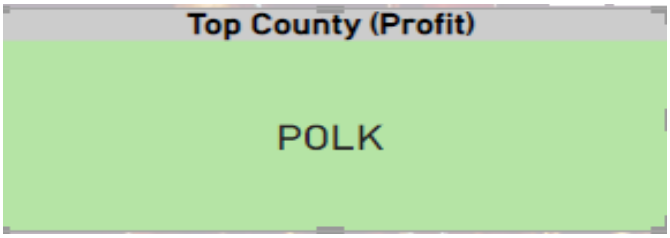
For a better and more interactive view of the total exports in USA, we have added a map where the size of each bubble represents the total orders of each county on the first level, and of each city on a second level.

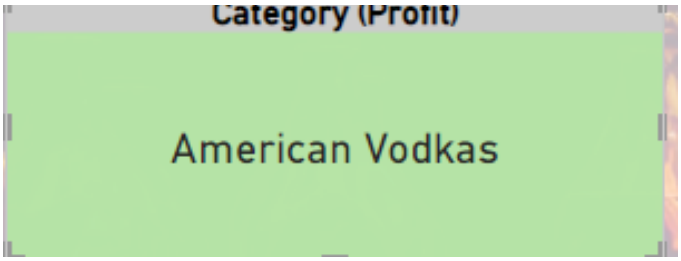



By hovering over each bubble, we can we see as additional information the total revenue and thet total profit for each county or city.




Finally, for a better control the Iowa vendors association can see at any time the top importing city in terms of profit as well as the top importing store, top importing county and the most profitable liquor category.

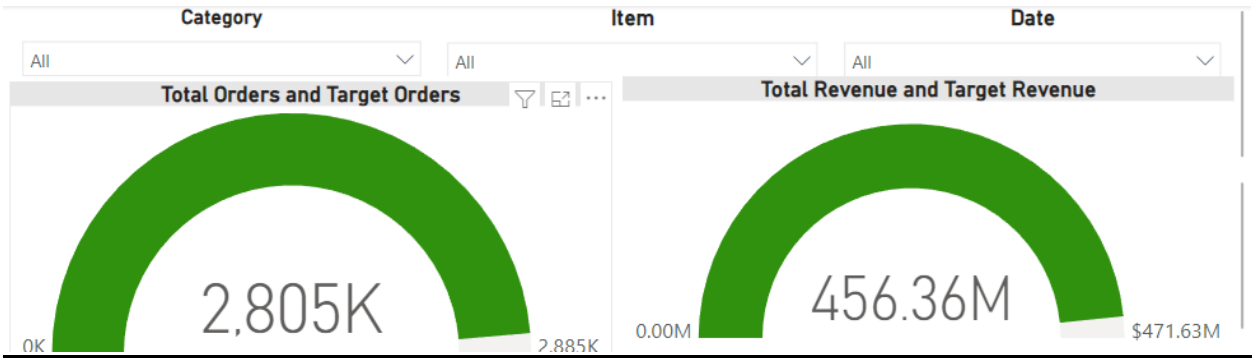




## Does Iowa reach its monthly target in terms of Revenue and total sales?

In order to monitor the monthly productiveness, we have added two gauge charts. We have selected as a target for each month to be a 10% increase in total orders and target revenue over the previous month.

The target is set at first level an overall level of all liquors and category. However, we can filter the graphs and check the target for each liquor and category.

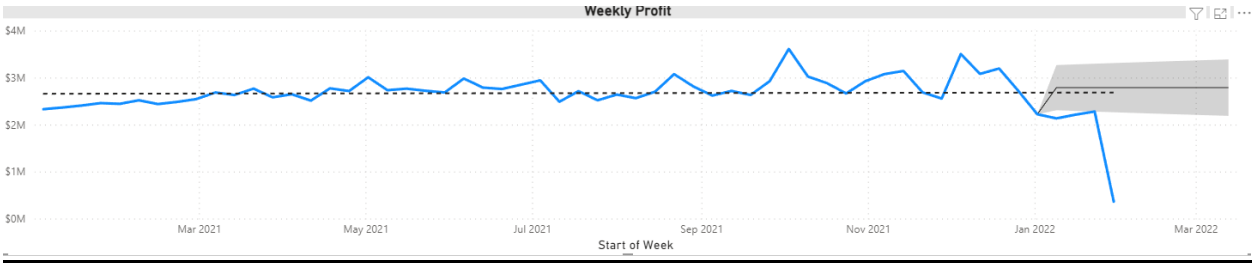
It seems that on the last month Iowa vendors were close to the monthly target, even though they did not manage to reach the target by a small margin.



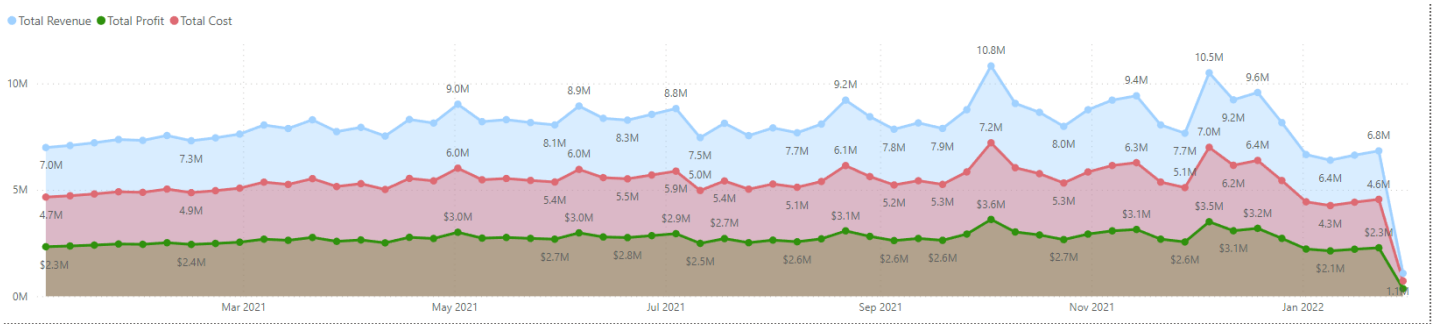
## What is the evolution of Iowa's profitability in the last one year in total? What is its profitability for each category and each individual liquor?

From the below graph we see the weekly profit of Iowa vendors on the last one year. We see that there is generally a steady trend in profits on around 2.8M per week. Only the last week there was an extreme decline that reduced the profit below 1M which was due to being very close to New Year's Eve.

Thus, due to this extreme occasion, for the future forecast we decided to exclude the 4 last periods, and the diagram seems to forecast a profit close to the average value for the next 10 weeks.

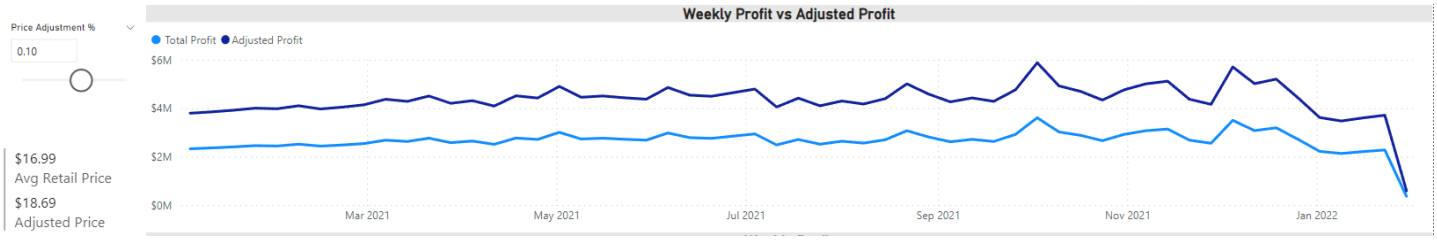


We have also a graph to monitor the weekly profit in relation to revenue and total cost. Each increase in revenue is followed by a similar increase in cost leading to a steady weekly profit. Thus, there seems to be a steady net price and the profit is driven mainly from the overall demand.




## How would a change in price potentially affect the total profit?

The below graph shows the weekly profit in comparison to adjusted profit in case of a percentage change in the retail price and if every other factor, like the overall demand, stays the same. Thus, we can make various kind of scenarios of how a change in price would affect the profit in general as well as the profit for a particular liquor or category.




To have an even better overview of the effect that a price change would have, we have added a multi-row card graph where we can see the adjusted and total profit for each liquor given their total volume sold and their total bottles sold.