Carolina Salas

Helen Zamudio

Uri Luna

Professor Woo

CIS 4560-01

5/1/2023

Lab Tutorial

\*SCP Part navigation command, remember you must be on your computer, your DESKTOP to do this command. Open a notepad to copy & paste code off tutorial for no data inconsistency and two GitBash terminals.

-----------------------------BEGINNING OF TUTORIAL------------------------------------

Carolina Salas@DESKTOP-GDBR7EJ MINGW64 ~ /\*File is located on desktop \*/

$ cd Desktop/

Carolina Salas@DESKTOP-GDBR7EJ MINGW64 ~/Desktop

$ ls

'3060 System Analysis and Design'/

'4100 Hardware & Software Architecture'/

'4560 Into to Big Data'/

'4580 E-Business Appl Dvl w Oracle'/

'4840 Communication Systems'/

'Business communication'/

Iowa\_Liquor\_Sales.csv… // located here

\*TRANSFER CURRENT FILE TO LOCAL DIRECTORY TO THE REMOTE DIRECTORY

Carolina Salas@DESKTOP-GDBR7EJ MINGW64 ~/Desktop

$ scp Iowa\_Liquor\_Sales.csv csalas25@129.153.114.72:/home/csalas25

csalas25@129.153.114.72's password:

Iowa\_Liquor\_Sales.csv 100% 11 767MBs 00:00

\*NEXT STEPS

Carolina Salas@DESKTOP-GDBR7EJ MINGW64 ~/Desktop

$ ssh csalas25@129.153.114.72

csalas25@129.153.114.72's password:\*\*\*\*\*

Last login: Mon May 1 21:47:53 2023 from 104-14-48-33.lightspeed.irvnca.sbcglobal.net

-bash-4.2$ ls \*CHECKING IF FILE TRANSFERRED SUCCESFFULY

dictionary.tsv fileToTransferViaSSH.txt \_\_MACOSX ratings\_2012.txt senti\_out.csv

driver\_data Iowa\_Liquor\_Sales.csv minion\_tweets.tar.gz ratings\_2013.txt time\_zone\_map.tsv

driver\_data.zip csv products.tsv ratings\_2013.txt.1 Tweets

\*MAKING DIRECTORY TO PUT FILE INTO, CHECKING AND DON’T COPY -BASH-4.2$

-bash-4.2$ hdfs dfs -mkdir temp

-bash-4.2$ hdfs dfs -mkdir temp/Iowa\_Liquor\_Sales

-bash-4.2$ hdfs dfs -put Iowa\_Liquor\_Sales.csv temp/Iowa\_Liquor\_Sales

-bash-4.2$ hdfs dfs -ls temp

-bash-4.2$ hdfs dfs -ls temp/Iowa\_Liquor\_Sales

\*Provides a way to connect to a Hive server and execute HiveQL statements

-bash-4.2$ beeline

INFO : Concurrency mode is disabled, not creating a lock manager

No rows affected (0.074 seconds)

0: jdbc:hive2://bigdaiun0.sub03291929060.trai> CREATE database IF NOT EXISTS csalas25;

0: jdbc:hive2://bigdaiun0.sub03291929060.trai> USE csalas25;

\*CREATE TABLE

0: jdbc:hive2://bigdaiun0.sub03291929060.trai> CREATE TABLE IF NOT EXISTS Iowa\_Liquor\_Sales (

invoice\_number INT,

item\_number INT,

transaction\_date DATE,

store\_number INT,

store\_name STRING,

address STRING,

city STRING,

zip\_code INT,

store\_location STRING,

county\_number INT,

county STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

STORED AS TEXTFILE LOCATION '/user/csalas25/temp/Iowa\_Liquor\_Sales/'TBLPROPERTIES ('skip.header.line.count'='1');

INFO : Completed executing command(queryId=hive\_20230501221926\_c8fc2255-664e-4393-bd15-ad80c5bff1e9); Time taken: 0.367 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

No rows affected (0.401 seconds)

DESCRIBE Iowa\_Liquor\_Sales;

\*LOADING DATA INTO TABLE

0: jdbc:hive2://bigdaiun0.sub03291929060.trai> LOAD DATA INPATH '/user/csalas25/temp/Iowa\_Liquor\_Sales/Iowa\_Liquor\_Sales.csv' INTO TABLE Iowa\_Liquor\_Sales;

No rows affected (1.598 seconds)

---------------------------------------STEP 2 ----------------------------------------

\*RUNNING QUERIES AGAINST TABLE

1. Find the total number of invoices for each store

0: jdbc:hive2://bigdaiun0.sub03291929060.trai> beeline

SELECT store\_number, COUNT(\*) as total\_invoices

FROM Iowa\_liquor\_Sales

GROUP BY store\_number;

INFO : Completed executing command(queryId=hive\_20230501222407\_5c6bfb82-94fa-4c71-9868-8b5382579a59); Time taken: 84.459 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

+---------------+-----------------+

| store\_number | total\_invoices |

+---------------+-----------------+

| 1011400 | 4587 |

| 1031000 | 4497 |

| 1031080 | 100864 |

| 1031090 | 351 |

| 1031100 | 25356 |

| 1031200 | 49690 |

| 1032080 | 32050 |

| 1032100 | 4966 |

| 1041200 | 720 |

| 1051150 | 303 |

| 1062300 | 23408 |

| 1081100 | 1027 |

| 1081210 | 134 |

| 1081250 | 200 |

| 1081350 | 158 |

| 1081370 | 1416 |

| 1082100 | 86 |

| 1082200 | 2491 |

| 1091200 | 344 |

| 1091300 | 158 |

| 1701200 | 201 |

| 1011300 | 19452 |

| 1012210 | 7546 |

| 1012400 | 1661 |

| 1041300 | 70 |

| 1042100 | 11092 |

| 1051120 | 1084 |

| 1052010 | 10549 |

| 1062050 | 998 |

| 1062200 | 35411 |

| 1070000 | 872 |

| 1081015 | 942 |

| 1081020 | 99 |

| 1081030 | 5880 |

| 1081200 | 22176 |

| 1081240 | 804 |

| 1081300 | 13526 |

| 1081305 | 2655 |

| 1081600 | 31358 |

| 1081700 | 5119 |

| 1082000 | 3656 |

| 1011700 | 194 |

| 1011800 | 40 |

| 1022100 | 36843 |

| 1022200 | 4017 |

| 1032000 | 1118 |

| 1081230 | 632 |

| 1081340 | 1075 |

| 1081380 | 2498 |

| 1081400 | 9210 |

| 1081900 | 9315 |

| 1082015 | 4 |

| 1092000 | 167 |

| 1901100 | 13 |

| 1011200 | 54420 |

| 1011250 | 738 |

| 1012100 | 86706 |

| 1012300 | 8397 |

| 1022300 | 39 |

| 1031110 | 103 |

| 1032200 | 23663 |

| 1051100 | 8203 |

| 1052100 | 3092 |

| 1062310 | 41859 |

| 1071100 | 28408 |

| 1081000 | 777 |

| 1081220 | 623 |

| 1081312 | 2704 |

| 1081335 | 963 |

| 1501100 | 3 |

| 1901200 | 181 |

| 1011100 | 41471 |

| 1041100 | 24134 |

| 1041150 | 1180 |

| 1051010 | 18198 |

| 1051110 | 4435 |

| 1081010 | 4436 |

| 1081315 | 2879 |

| 1081355 | 955 |

| 1081365 | 755 |

| NULL | 36830259 |

| 1011500 | 4758 |

| 1011600 | 879 |

| 1012200 | 18382 |

| 1051140 | 1115 |

| 1062100 | 2615 |

| 1062400 | 7797 |

| 1062500 | 4537 |

| 1081317 | 935 |

| 1081330 | 5237 |

| 1081390 | 13245 |

| 1081500 | 845 |

| 1082900 | 19284 |

| 1091000 | 48 |

| 1091100 | 223 |

| 1091400 | 61 |

| 1092100 | 20 |

| 1101100 | 2003 |

| 1700000 | 1 |

| 1701100 | 3832 |

+---------------+-----------------+

100 rows selected (84.914 seconds)

1. Calculate the total sales for each store:

SELECT store\_number, SUM(item\_number) as total\_sales

FROM Iowa\_Liquor\_Sales

GROUP BY store\_number;

INFO : Completed executing command(queryId=hive\_20230501222954\_df41332b-d3b7-4f83-84df-740ab07d800e); Time taken: 152.39 seconds

INFO : OK

INFO : Concurrency mode is disabled, not creating a lock manager

+---------------+--------------+

| store\_number | total\_sales |

+---------------+--------------+

| NULL | 41544 |

| 1011100 | 2448196 |

| 1011200 | 3201500 |

| 1011250 | 46083 |

| 1011300 | 1108416 |

| 1011400 | 271764 |

| 1011500 | 282582 |

| 1011600 | 52293 |

| 1011700 | 12306 |

| 1011800 | 1942 |

| 1012100 | 4819815 |

| 1012200 | 1038190 |

| 1012210 | 443555 |

| 1012300 | 478424 |

| 1012400 | 98116 |

| 1022100 | 2142471 |

| 1022200 | 240817 |

| 1022300 | 2457 |

| 1031000 | 230937 |

| 1031080 | 5794966 |

| 1031090 | 21381 |

| 1031100 | 1524990 |

| 1031110 | 6495 |

| 1031200 | 2855009 |

| 1032000 | 56813 |

| 1032080 | 1823229 |

| 1032100 | 293992 |

| 1032200 | 1372208 |

| 1041100 | 1385987 |

| 1041150 | 66319 |

| 1041200 | 39260 |

| 1041300 | 3434 |

| 1042100 | 632279 |

| 1051010 | 979926 |

| 1051100 | 452808 |

| 1051110 | 241661 |

| 1051120 | 62055 |

| 1051140 | 56480 |

| 1051150 | 15859 |

| 1052010 | 594810 |

| 1052100 | 196783 |

| 1062050 | 57152 |

| 1062100 | 155193 |

| 1062200 | 2062780 |

| 1062300 | 1310454 |

| 1062310 | 2410909 |

| 1062400 | 439954 |

| 1062500 | 253270 |

| 1070000 | 45521 |

| 1071100 | 1651762 |

| 1081000 | 36690 |

| 1081010 | 259944 |

| 1081015 | 55490 |

| 1081020 | 6591 |

| 1081030 | 332184 |

| 1081100 | 55550 |

| 1081200 | 1234542 |

| 1081210 | 6427 |

| 1081220 | 31511 |

| 1081230 | 30938 |

| 1081240 | 40539 |

| 1081250 | 11252 |

| 1081300 | 743398 |

| 1081305 | 159224 |

| 1081312 | 146405 |

| 1081315 | 160312 |

| 1081317 | 53826 |

| 1081330 | 299855 |

| 1081335 | 49351 |

| 1081340 | 61972 |

| 1081350 | 8582 |

| 1081355 | 49860 |

| 1081365 | 42174 |

| 1081370 | 83481 |

| 1081380 | 152923 |

| 1081390 | 703222 |

| 1081400 | 506943 |

| 1081500 | 48640 |

| 1081600 | 1815613 |

| 1081700 | 292843 |

| 1081900 | 530557 |

| 1082000 | 219078 |

| 1082015 | 69 |

| 1082100 | 5289 |

| 1082200 | 127312 |

| 1082900 | 1111015 |

| 1091000 | 2486 |

| 1091100 | 14311 |

| 1091200 | 18497 |

| 1091300 | 8187 |

| 1091400 | 3475 |

| 1092000 | 7792 |

| 1092100 | 942 |

| 1101100 | 116808 |

| 1501100 | 91 |

| 1700000 | 48 |

| 1701100 | 209766 |

| 1701200 | 9929 |

| 1901100 | 584 |

| 1901200 | 11126 |

+---------------+--------------+

100 rows selected (152.771 seconds)

3.Find the top 10 stores with the highest total sales

SELECT store\_number, SUM(item\_number) as total\_sales

FROM lowa\_liquor\_Sales

GROUP BY store\_number

ORDER BY total\_sales DESC

LIMIT 10;

+---------------+--------------+

| store\_number | total\_sales |

+---------------+--------------+

| 1031080 | 11589932 |

| 1012100 | 9639630 |

| 1011200 | 6403000 |

| 1031200 | 5710018 |

| 1011100 | 4896392 |

| 1062310 | 4821818 |

| 1022100 | 4284942 |

| 1062200 | 4125560 |

| 1032080 | 3646458 |

| 1081600 | 3631226 |

+---------------+--------------+

4. find county where its called Tama

SELECT store\_number, store\_name, address, city, county

FROM lowa\_liquor\_Sales

WHERE county = 'Tama';

INFO : Concurrency mode is disabled, not creating a lock manager

+---------------+-------------+----------+-------+---------+

| store\_number | store\_name | address | city | county |

+---------------+-------------+----------+-------+---------+

+---------------+-------------+----------+-------+---------+

No rows selected (0.274 seconds)

5. Find the store with the most invoice number

SELECT store\_number, store\_name

FROM Iowa\_Liquor\_Sales

GROUP BY store\_number, store\_name

ORDER BY COUNT(invoice\_number) DESC

LIMIT 1;

+---------------+--------------------+

| store\_number | store\_name |

+---------------+--------------------+

| NULL | 703 W PLEASANT ST |

+---------------+--------------------+

1 row selected (210.477 seconds)

-----------------------------LAST STEP OF TUTORIAL------------------------------------

\*Last step is to create a visual diagram with Tableau.

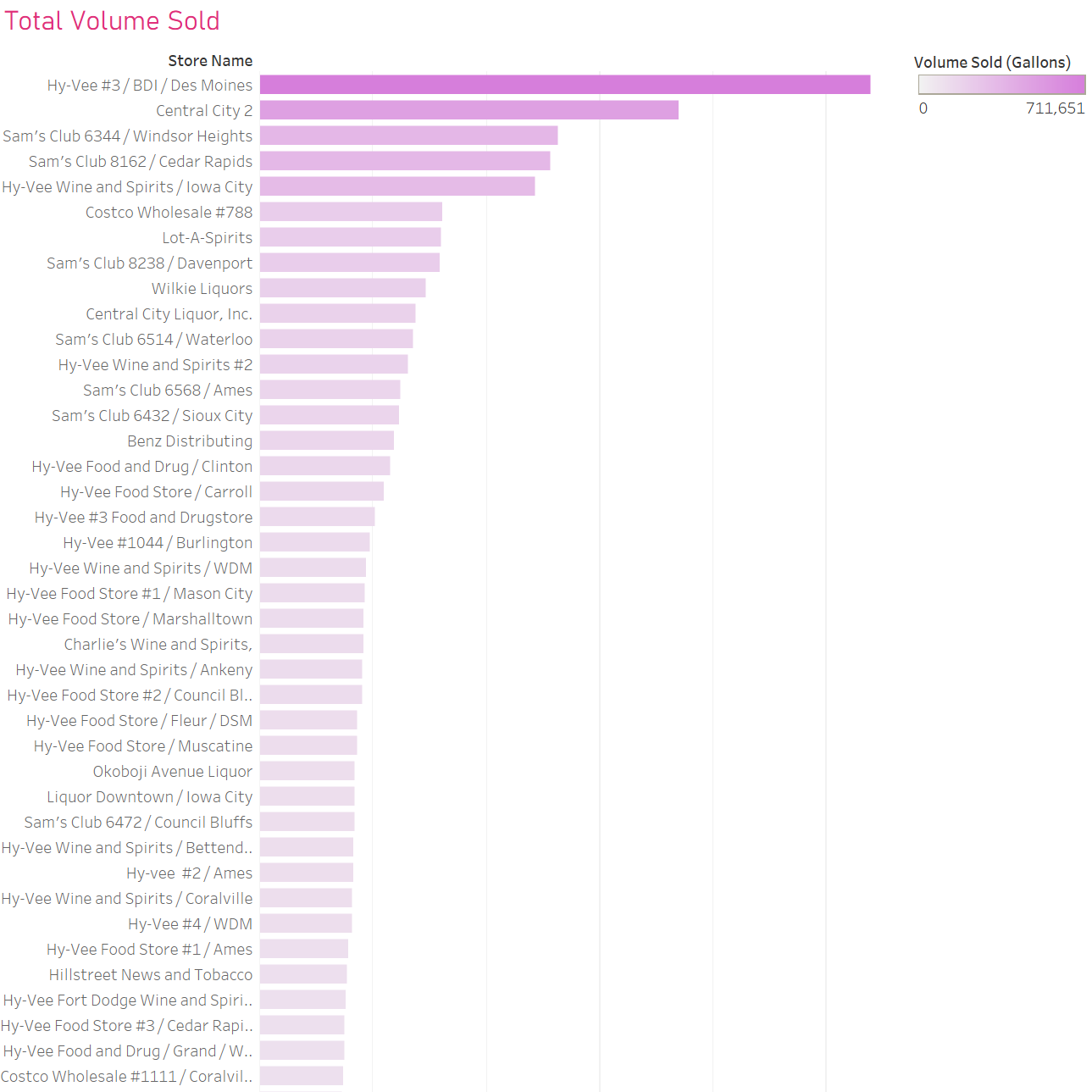
Step one: download Tableau Desktop App, if you don’t have it.

[Start your free trial of Tableau](https://www.tableau.com/products/trial)

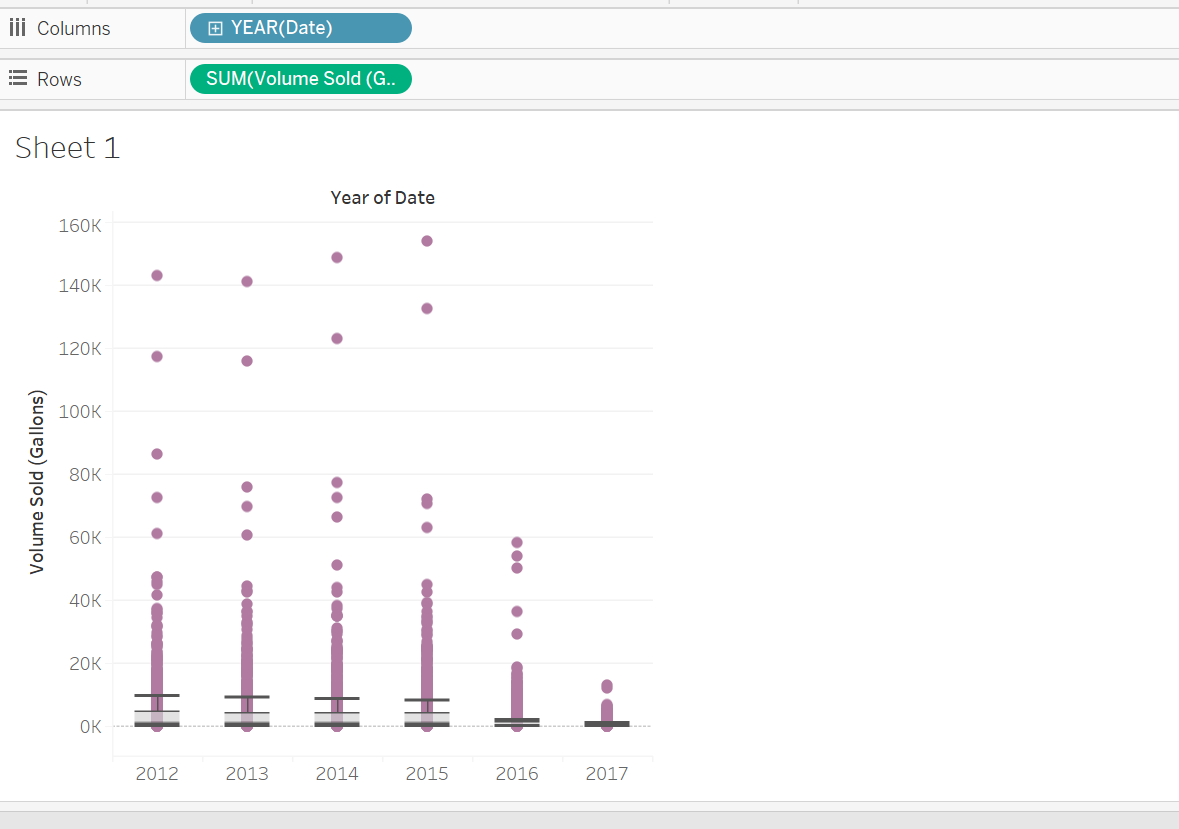
Step two: Upload Iowa\_liquor\_Sales.csv to Tableau…might take a minute to upload.

Step three: Choose Volume Sold (Gallons) as column and Store name which will show the store with the most sold in gallons.

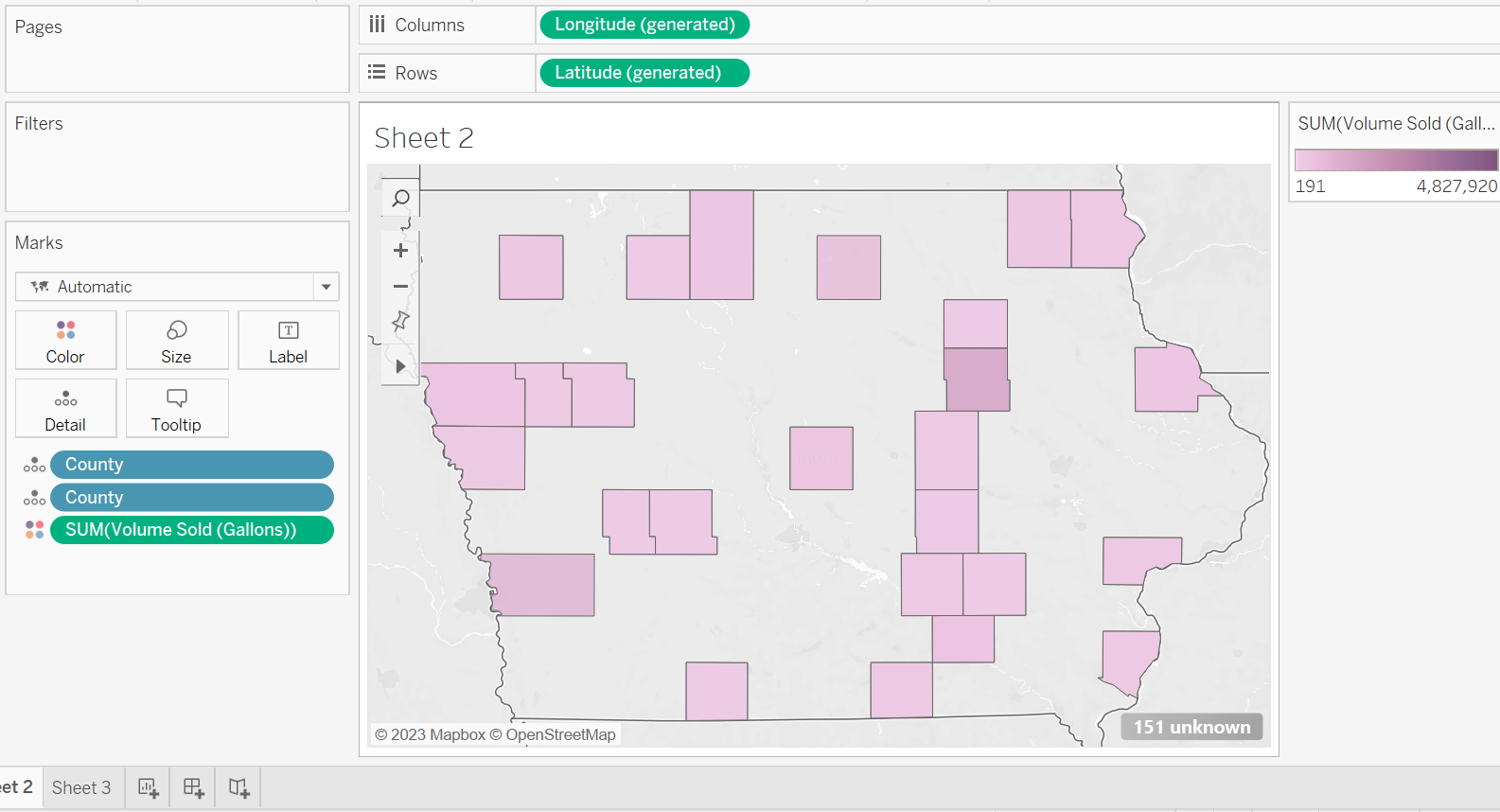
Create a Visual Representation 1 where columns are store name and rows are volume sold in gallon and then select horizontal bars.



Create a Visual Representation 2 with year(date) as columns and rows as SUM([Volume Sold (Gallons)]) then select Line map in show me.



Create a Visual Representation 3 with Country, Sum(Volume Sold(Gallons)) , and in show me button pick GEO which will then naturally the row and column will turn into Latitude and longitude (generated)



Create a Visual Representation 4 Store Location as row, and year (Date), and Quarter(Date) (This one will naturally appear)

