

Course Project

Analysis of Iowa Liquor Sales and Income Using SQL (City, County, and Zip Code Level)

1. INTRODUCTION

The goal of this project is to analyze the relationship between liquor sales and income in the state of Iowa using SQL. Six datasets were used in this analysis, including liquor sales and ACS income data at the city, county, and zip code levels. The liquor sales data represents activity from 2020, while the income and demographic data comes from the 2019 American Community Survey (ACS). By joining these tables using SQL, we were able to explore how income levels relate to liquor purchasing behavior at different geographic scales.

2. CITY-LEVEL ANALYSIS (2020 Sales + 2019 Income)

SQL Query Used (City Level JOIN)

```
SELECT
    s.city,
    a.income,
    SUM(s.`sale.dollars`) AS total_liquor_sales
FROM `project.sales.cities` AS s
JOIN `project.acs.cities 2019` AS a
    ON s.city = a.city
GROUP BY s.city, a.income
ORDER BY total_liquor_sales DESC;
```

MySQL Workbench

Local Instance 3306 - Warning - not supported

Administration Schemas Query 1 SQL File 3* SQL File 4* SQL File 5* project.sales.counties - Table project.sales.counties 2019 - Table Iowa_project.project.sales.counties

MAI

1 SELECT
2 s.county,
3 a.income,
4 SUM(s.sale.dollars) AS total_liquor_sales
5 FROM `project.sales.counties` AS s
6 JOIN `project.acs.counties 2019` AS a
7 ON s.county = a.county
8 GROUP BY s.county, a.income
9 ORDER BY total_liquor_sales DESC;
10

100% 1/10

Result Grid Filter Rows: Search Export

county	income	total_liquor_sales
Polk	28121	8100906
Linn	24876	23211745
Scott	32640	18005481
Jefferson	28130	16221212
Black Hawk	27979	11086448
Polk	31457	9360605
Woodbury	27238	8747128
Murray	18859	8105026
Outagamie	30044	8009493
Grant	28632	5485544
Dickinson	31194	3764308
Webster	29415	3418814
Lee	26630	3411803
Christian	30652	3279409
Lee	27804	3276070
Muskegon	31151	2867649
Marshall	30035	2800819
Winnebago	27719	2414266
Carroll	31152	2228759
Oshkosh	43120	2183000
Winneconne	36543	2072032

Results 14

Action Output

Time	Action	Response	Duration / Fetch Time
72 17:30:24	SELECT s.county, a.income, SUM(s.sale.dollars) AS total_liquor_sales FROM p... 88 row(s) returned		0.0022 sec / 0.00001...

City-Level Interpretation :

At the city level, we joined 2020 liquor sales data with 2019 ACS income data using city as the matching key. After calculating the total liquor sales per city, we found that cities with higher income generally tend to have higher total liquor sales. However, some lower-income cities also showed high sales due to factors such as population size, tourism, and concentration of liquor stores. This suggests that income plays an important role in liquor consumption, but it is not the only determining factor.

3. COUNTY-LEVEL ANALYSIS (2020 Sales + 2019 Income) :

SQL Query Used (County Level JOIN) :

SELECT

s.county,

a.income,

SUM(s.`sale.dollars`) AS total_liquor_sales

FROM `project.sales.counties` AS s

JOIN `project.acs.counties 2019` AS a

ON s.county = a.county

GROUP BY s.county, a.income

ORDER BY total_liquor_sales DESC;

The screenshot shows a SQL IDE interface with a query editor and a results grid. The query is as follows:

```

1 SELECT
2     s.county,
3     a.income,
4     SUM(s.`sale.dollars`) AS total_liquor_sales
5 FROM `project.sales.counties` AS s
6 JOIN `project.acs.counties 2019` AS a
7 ON s.county = a.county
8 GROUP BY s.county, a.income
9 ORDER BY total_liquor_sales DESC;
10

```

The results grid displays the following data:

county	income	total_liquor_sal...
Polk	35121	61059636
Linn	34878	23211745
Scott	32582	19303451
Johnson	28190	16251212
Black Hawk	27979	15289449
Pottawattamie	31657	9366065
Woodbury	27239	9147132
Story	18959	8105026
Dubuque	30044	8095940
Cerro Gordo	26930	5462244
Dickinson	31194	3764208
Webster	25515	3418914
Des Moines	26580	3411903
Clinton	30652	3279409
Lee	27804	3276570
Muscatine	31151	2867649
Marshall	30035	2850819
Wapello	27318	2414266
Carroll	31332	2226799
Dallas	43120	2183590
Warren	36543	2072032

The bottom of the screenshot shows the Action Output tab with the following information:

Time	Action	Response	Duration / Fetch Time
72 17:30:24	SELECT s.county, a.income, SUM(s.`sale.dollars`) AS total_liquor_sales FROM `p...	98 row(s) returned	0.0022 sec / 0.0000

County-Level Interpretation :

At the county level, we joined the liquor sales and income data using county name as the matching key. The results show that counties with higher income generally generate more total liquor sales. However, several mid-income counties also display strong liquor sales, likely due to higher population and economic activity. This indicates that both income and population influence total liquor revenue at the county level.

4. ZIP CODE-LEVEL ANALYSIS (2020 Sales + 2019 Income)

SQL Query Used (Zip-Level JOIN) :

SELECT

s.zipcode,

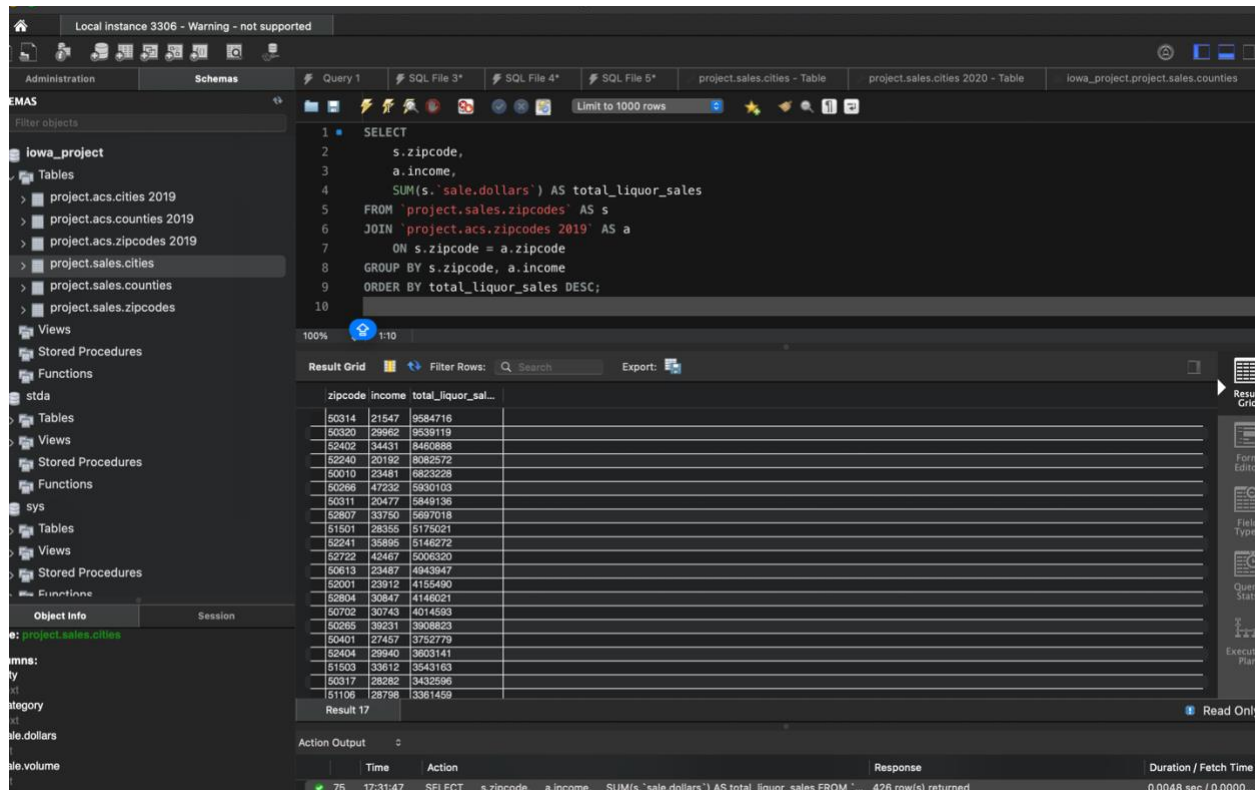
a.income,

SUM(s.`sale.dollars`) AS total_liquor_sales

```

FROM `project.sales.zipcodes` AS s
JOIN `project.acs.zipcodes 2019` AS a
    ON s.zipcode = a.zipcode
GROUP BY s.zipcode, a.income
ORDER BY total_liquor_sales DESC;

```



zipcode	income	total_liquor_sales
50314	21547	9584716
50320	29962	9539119
52402	34431	8460888
52240	20192	8082572
50010	23481	6823228
50256	47232	5930163
50311	20477	5849136
52807	33750	5897018
51501	28356	5175021
52241	35866	5146272
52722	42467	5006320
50613	23487	4943947
53001	23912	4115590
52804	30847	4146021
50702	30743	4014593
50265	39231	3908823
50401	27457	3752779
52404	29940	3603141
51503	33612	3543163
50317	28282	3432596
51136	728796	3361459

Zip Code–Level Interpretation :

At the zip code level, I combined 2020 liquor sales with 2019 income data using zip code as the matching key. The results show that liquor spending varies significantly even within the same city. Some high-income zip codes demonstrate very high liquor sales, while some lower-income areas also report strong sales. This suggests that local retail availability and consumer behavior strongly influence liquor purchasing patterns.

5. COMPARATIVE ANALYSIS (ALL THREE LEVELS) :

By analyzing liquor sales and income at the city, county, and zip code levels, we observed that income is positively related to liquor sales across all three geographic scales. The city-level analysis revealed broad economic patterns, the county-level analysis showed regional trends, and the zip code-level analysis captured highly localized spending behavior. Across all levels, population size, tourism, and density of liquor stores also played a major role in total sales.

6. FINAL CONCLUSION :

This project demonstrated how SQL can be used to integrate multiple large datasets and analyze real-world economic relationships. By joining liquor sales data from 2020 with ACS income data from 2019 at three geographic levels, we were able to explore how income relates to liquor purchasing behavior in Iowa. The results suggest that higher income areas generally show higher liquor sales, but other factors such as population size and retail concentration also strongly influence total sales. Overall, this project highlights the power of SQL for conducting meaningful economic and business data analysis.