

# **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;
```

The screenshot shows the MySQL Workbench interface with a query editor and results grid. The query joins 2020 liquor sales data with 2019 ACS income data by city. The results show 14 rows of data, ordered by total liquor sales in descending order. The execution time was 0.0022 sec / 0.00001.

```

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;

```

county	income	total_liquor_sales
Polk	38121	6105936
Linn	34878	23211745
Story	32222	20311723
Arapahoe	28190	19251212
Black Hawk	27979	15289449
Pottawattamie	31657	13655966
Harrison	27723	13525223
Story	18691	6105926
Dubuque	30044	6069040
Cerro Gordo	28930	14620344
Clinton	31171	14511217
Weld	25815	14161014
Des Moines	26580	3411903
Clinton	30652	279409
Washington	25220	2781000
Muscatine	31151	2867649
Marshall	30035	2850519
Wapello	27318	2414596
Cass	17171	2381000
Dallas	43120	2185990
Warren	36543	2072032

Action Output: 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;

Time: 0.0022 sec / 0.00001 Duration / Fetch Time:

### 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 the MySQL Workbench interface with the following details:

- Query Editor:** Contains the SQL code:

```

1 SELECT
2     s.county,
3     a.income,
4     SUM(s.`sale.dollars`) AS total_liquor_sales
5 FROM `project.acs.cities 2019` 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
    
```
- Result Grid:** Displays the results of the query as a table:

county	income	total_liquor_sales
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	28930	5482244
Dickinson	31194	3764208
Webster	25615	3416914
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
- Timeline:** Shows the execution time and duration of the query.

### 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;

```

The screenshot shows the MySQL Workbench interface with the following details:

- Query Editor:** Contains the SQL code provided above.
- Result Grid:** Displays the output of the query, which includes columns for zipcode, income, and total\_liquor\_sales. The data shows various zip codes across different income levels contributing to the total sales.
- Object Info:** Shows the schema and structure of the tables involved in the query.
- Session:** Provides information about the current session, including the number of rows returned (426).

zipcode	income	total_liquor_sales
50314	21547	9584716
50320	29962	9539119
52402	34431	8460888
52240	20192	8082572
50010	23481	6823228
50269	41738	5930103
50311	20377	5849136
52807	28750	5671618
51501	28355	5175021
52241	38405	5146272
52722	49467	5096920
50613	29487	4943947
52001	29912	4155490
52804	30847	4146021
50702	30743	4014593
50265	38231	3908823
50401	27457	3752779
52404	29940	3603141
51503	38612	3543163
50317	28282	3432596
51106	28798	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.