Course Project

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

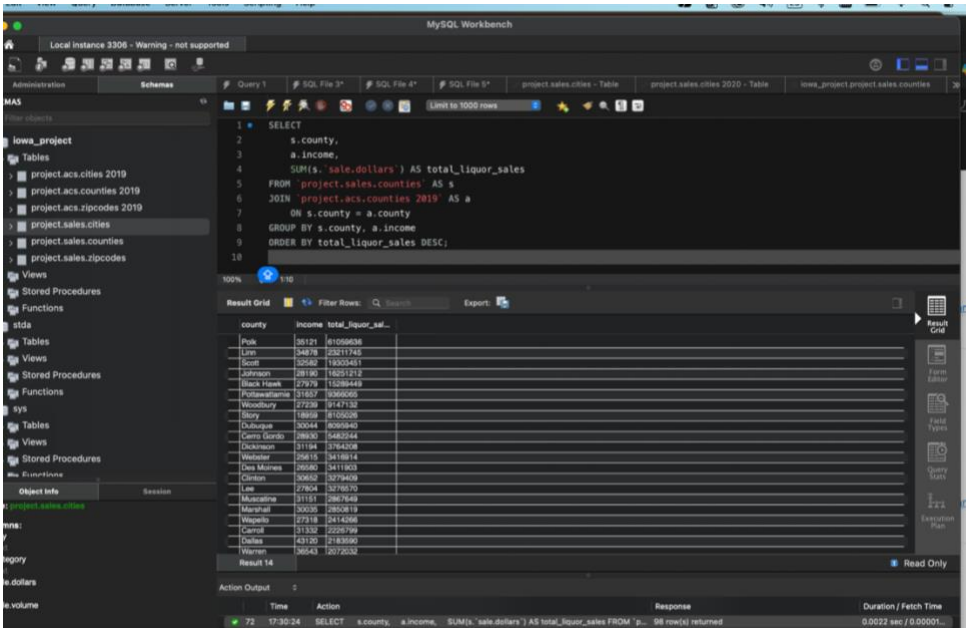
1. INTRODUCTION

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;
```



City-Level Interpretation :

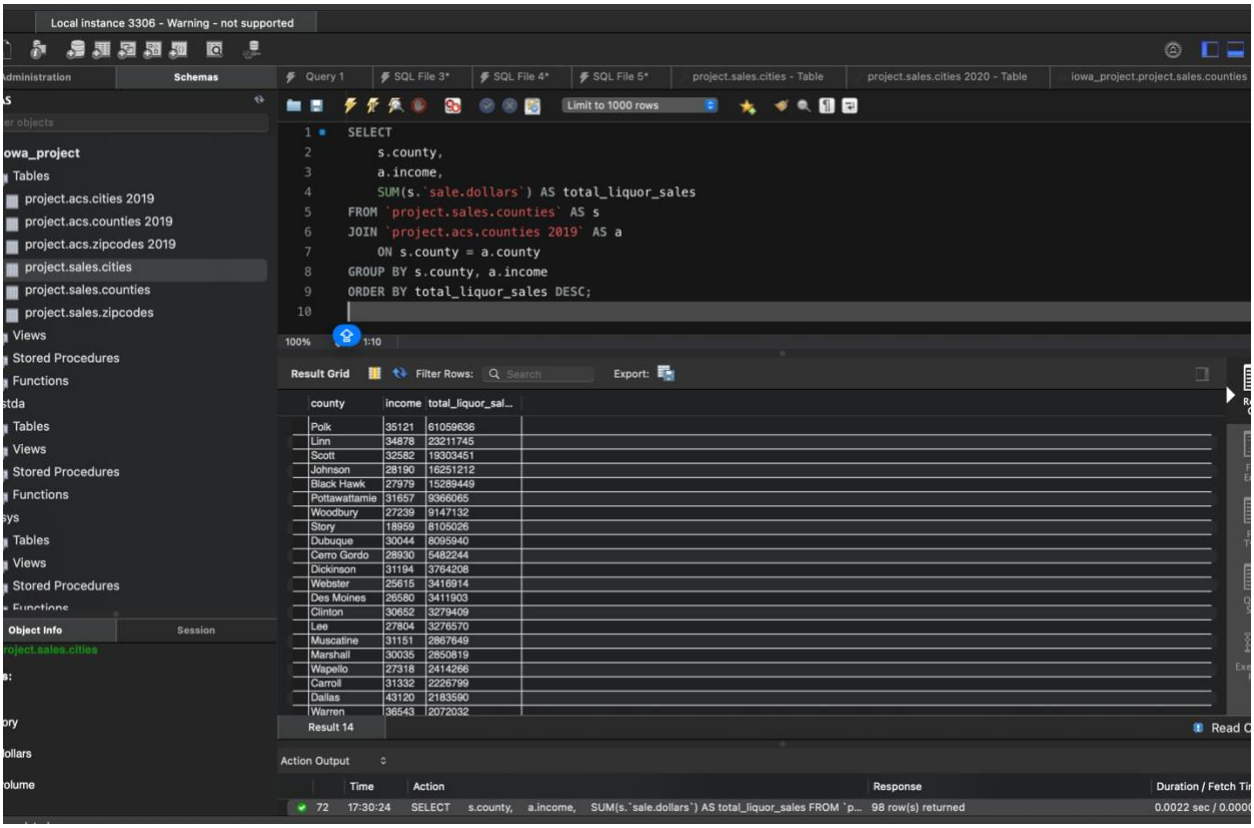
At the city level, we joined 2020 liquor sales data with 2019 ACS income data using city as the matching key. After calculating the total liquor sales per city, we found that cities with higher income generally tend to have higher total liquor sales. However, some lower-income cities also showed high 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;



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) :

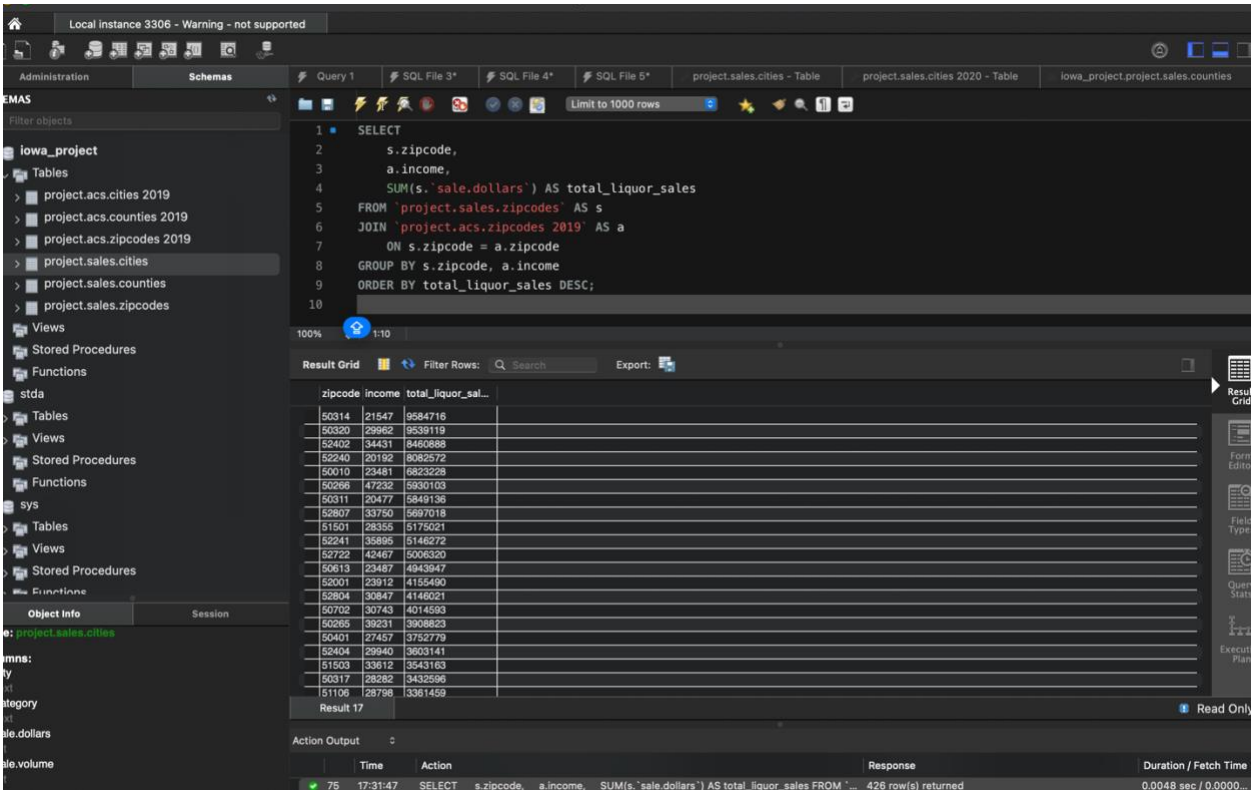
SELECT

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;
```



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.