1. **Intro (5pts): Provide context regarding your decision maker, organization, and overall decision climate. State your research question. Explain how policy decisions will affect your organization and the broader community.**

As the CEO of a liquor sales company, making informed decisions about where and what to sell is critical for maximizing profits and staying competitive. Understanding pairwise patterns in total/per capita sales across geographies in Iowa can provide valuable insights into consumer preferences and demand in different regions. For instance, if certain zip codes or cities consistently show higher sales than others, the CEO can consider expanding their product offerings in those areas or increasing marketing efforts to attract more customers. Additionally, analyzing sales data across different liquor categories can help identify trends and inform decisions on which products to focus on or discontinue.

The decision climate for the CEO of a liquor sales company is complex and influenced by various factors such as competition, regulations, and consumer behavior. Analyzing pairwise patterns in total/per capita sales across geographies can help the CEO navigate this climate by providing evidence-based insights on market demand and opportunities. For example, if there is a correlation between average sales per zipcode and average sales in corresponding cities, the CEO can use this information to target specific areas for expansion or promotional efforts. Understanding how these patterns differ across liquor categories can also inform decisions on which products to stock in certain regions and which to prioritize for marketing campaigns.

The research question for this analysis is: What are the pairwise patterns in total/per capita sales across geographies for Iowa liquor sales, and are there correlations between average sales per zipcode and average sales in corresponding cities? Does this pattern differ across liquor categories? The results of this analysis can inform policy decisions for the liquor company as well as the broader community. For instance, if certain areas show consistently high sales, the company may consider opening new stores or increasing inventory in those regions, potentially leading to job creation and economic growth. However, policy decisions related to alcohol sales must also consider the potential impact on public health and safety, and the company should prioritize responsible marketing and sales practices.

1. **Data Summary (5pts):** Provide a short description of the nature of the provided data set and explain how these characteristics affect your analysis methodology. Summarize the data set with basic descriptive statistics as applicable.

The provided dataset contains information about sales and demographic data for cities, counties, and zip codes in the US. Specifically, there are three datasets for sales, each containing information on sales volume and sales dollars for different categories of products in cities, counties, and zip codes. There are also three demographic datasets, each containing information on population, education level, income, and race for cities, counties, and zip codes. The datasets are merged by geography to create three merged datasets: city\_data, county\_data, and zipcode\_data.

The characteristics of the dataset affect the analysis methodology in different ways. Firstly, the dataset is relatively large, with 3693 observations in the city\_data dataset alone, and this means that some analyses may require specialized computing resources to process. Secondly, the dataset is mainly composed of categorical and numerical variables, which may require different analysis techniques. Lastly, the dataset contains demographic information, which may require additional data cleaning and preparation steps to ensure that analyses are not influenced by factors such as data errors or missing data.

From the summary statistics, we can see that the mean sales dollars are much higher than the median, indicating that the distribution of sales dollars is positively skewed. The same is true for sales volume. The mean unemployment rate is 4.44, with a median of 4, indicating that the distribution of unemployment is relatively symmetrical. The median population is 1613, while the mean is 6154, indicating that the distribution of population is positively skewed. The median income is 30329, with a mean of 30606, indicating that the distribution of income is relatively symmetrical. The racial data shows that the mean and median for the different racial groups vary widely.

1. **Data Analytics (50pts):** Provide data analytics that add clarity to the research question. Thoroughly discuss insight obtained from your visualizations and analysis of aggregated, ACS and merged datasets, including trends or specific data points (Tasks 2-5). Suggest an excursion, and provide supporting analysis. Plots should be well formatted according to best practices learned in class. Discuss the advantages and challenges of performing analysis in your chosen software tool.
2. **Conclusion (10pts):** Summarize the analytical methodology and provide closure to your analytical story. Succinctly answer the research questions. Highlight the limitations of your findings and recommend future work. Do not make policy recommendations here.
3. **Policy recommendation (10pts):** Introduce a specific policy decision that your decision maker is facing. Provide a data driven recommendation for their decision. Explain probable first and second order effects of the recommendation. Explain the benefits and risks of the recommendation.