

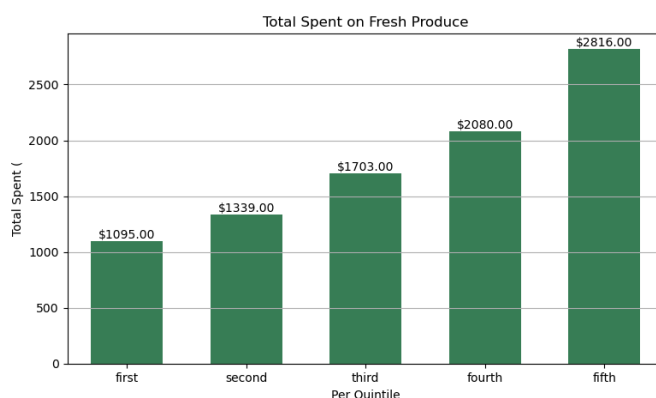
Taylor Korec
Final Project Summary

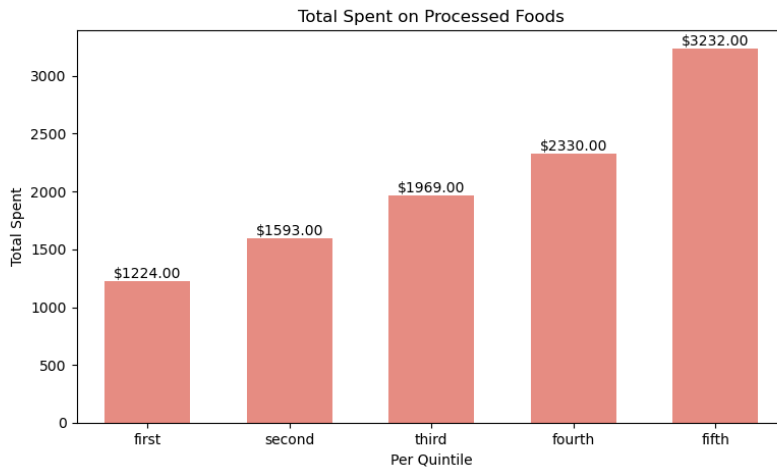
For the final project I chose to perform data cleaning, analysis and visualization on the consumer expenditure survey for 2023 from the Bureau of Labor Statistics. This survey created five groups by income before taxes into quintiles each the 20% percentile. Then recorded the mean spent in many different categories from home, travel, food, education, etc.. My questions about this dataset were: Out of the five quintiles of income, which quintile spends the most on fresh produce, Which quintile spent the most on processed foods, What were the top food categories that consumers spent the most on for each quintile of income, and Is there a significant difference in the spending on each food category between each quintiles?

I used the pandas package to import the survey dataset and place it into a pandas DataFrame. Then I explored the DataFrame and began the cleaning process, like dropping rows, renaming, columns, changing data types using the decimal package, dropping rows with NaN values, and renaming the categories in the rows. Then after answering the questions I created visualizations using the matplotlib.pyplot and seaborn packages, in forms of bar plots, stacked barplot and a scatter line plot.

For this analysis my clients would be grocery store chains. This analysis gives insights on consumer behavior between the different quintiles of income that can help with decision making like where they will place certain stores, what products will they advertise the most, and who will they advertise these products to.

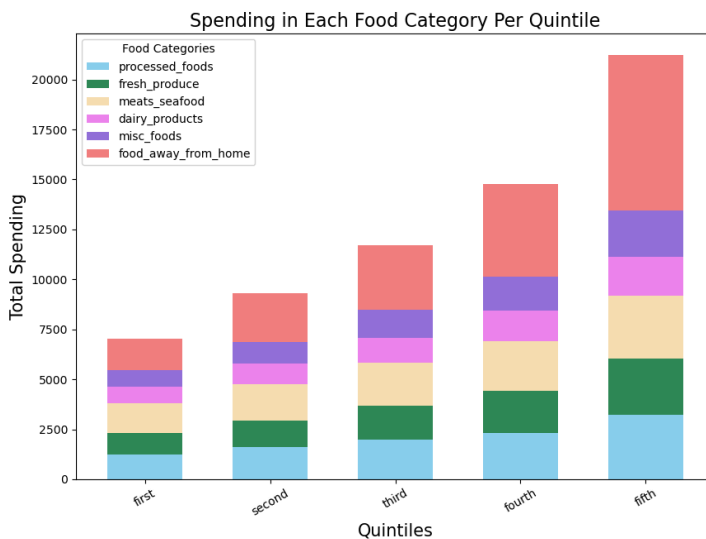
From question 1: Out of the five quintiles of income, which quintile spends the most on fresh produce and question 2: Which quintile spent the most on processed foods, I got similar results. I found that quintile 5 so the highest income spent the most in both categories, and down from 4 to 1.





For question 3: What were the top food categories that consumers spent the most on for each quintile of income?

I had to total the individual categories into more board categories then find the highest for each group. It was found that each group spent the most on food away from home meaning food from restaurants and food made out of town. When I took away the food away from home category I found that quintiles 1-4 spent the most on meats and seafood where quintile 5 spent the most on processed foods.



For question 4: Is there a significant difference in the spending on each food category between each quintile? I created a line plot and found that each category of food had a very similar purchase pattern besides the food away from home category.

