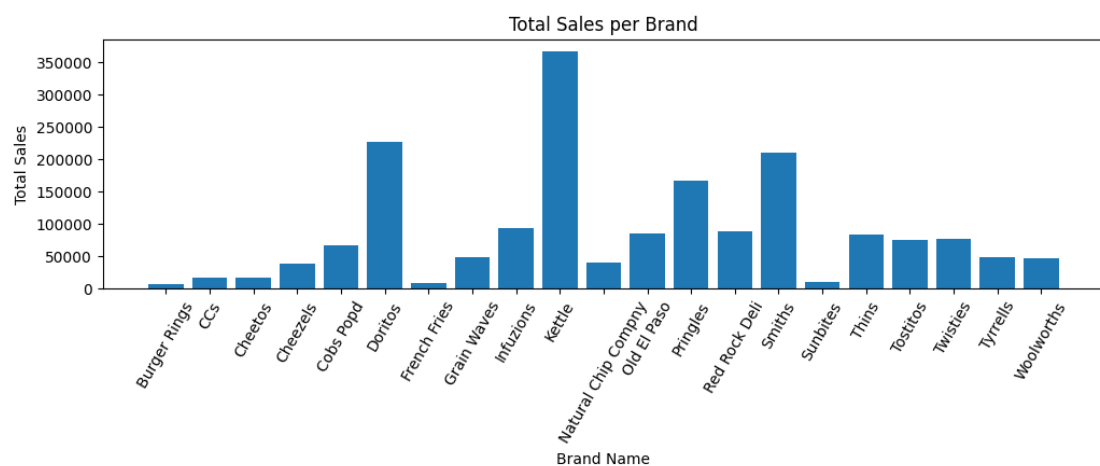
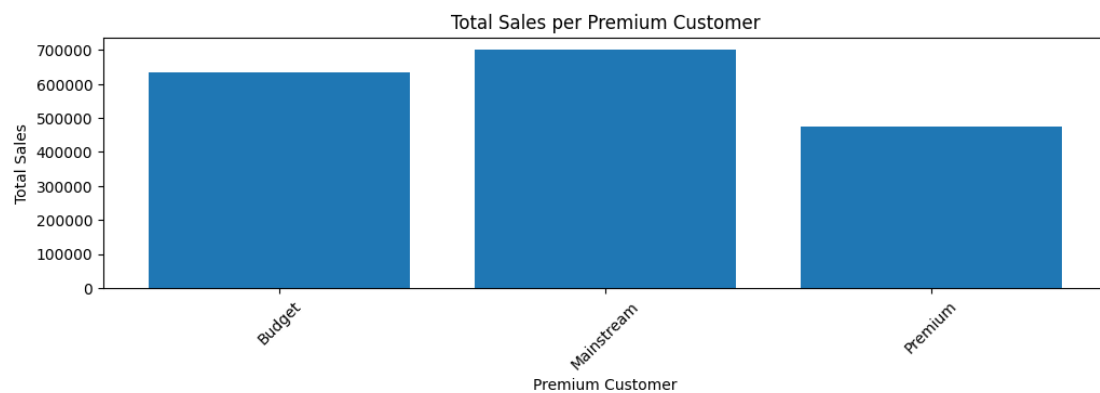
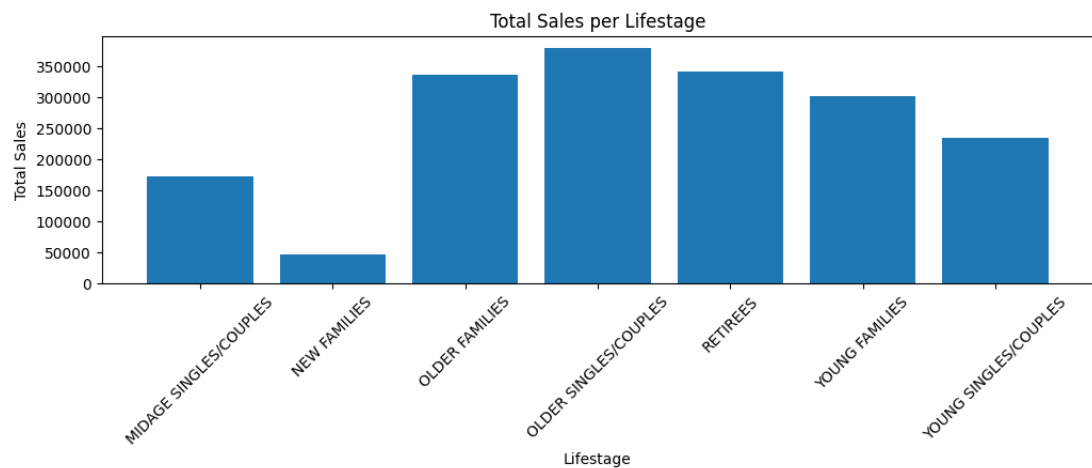
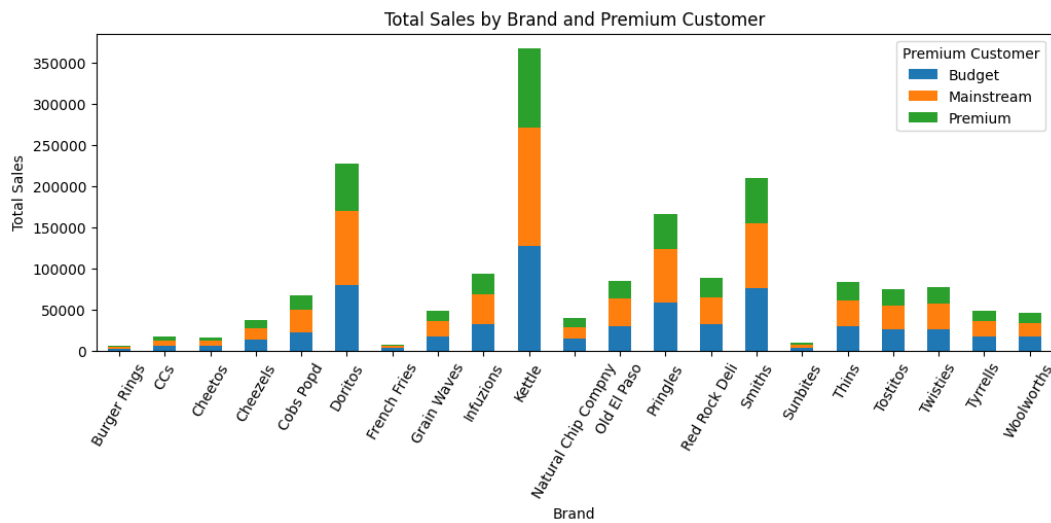
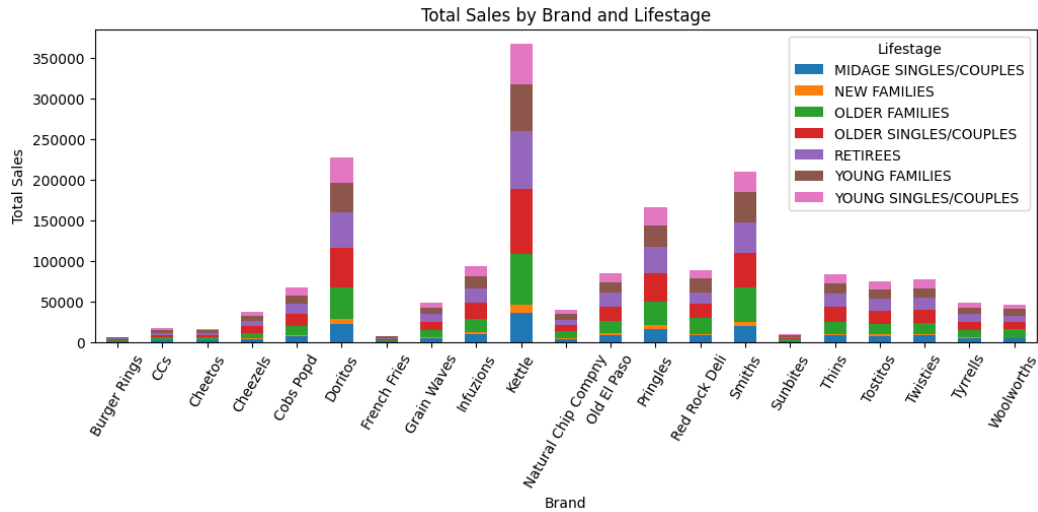


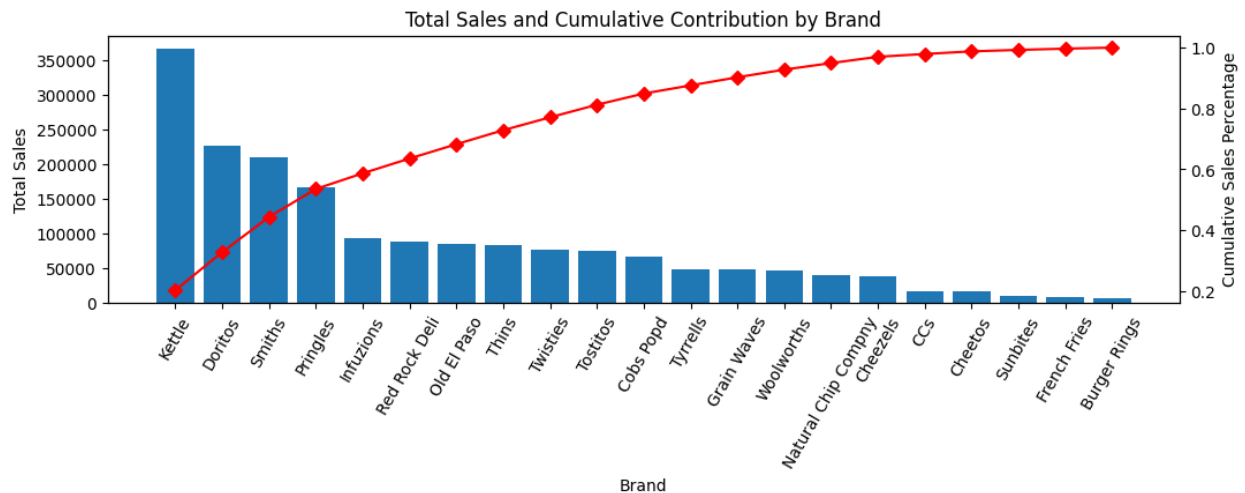
First, I analyzed key metrics in the dataset to find total sales, drivers of sales, and the source of highest sales. I broke total sales into three categories: by lifestage, by premium customer, and by brand. I summarized the results in bar charts:



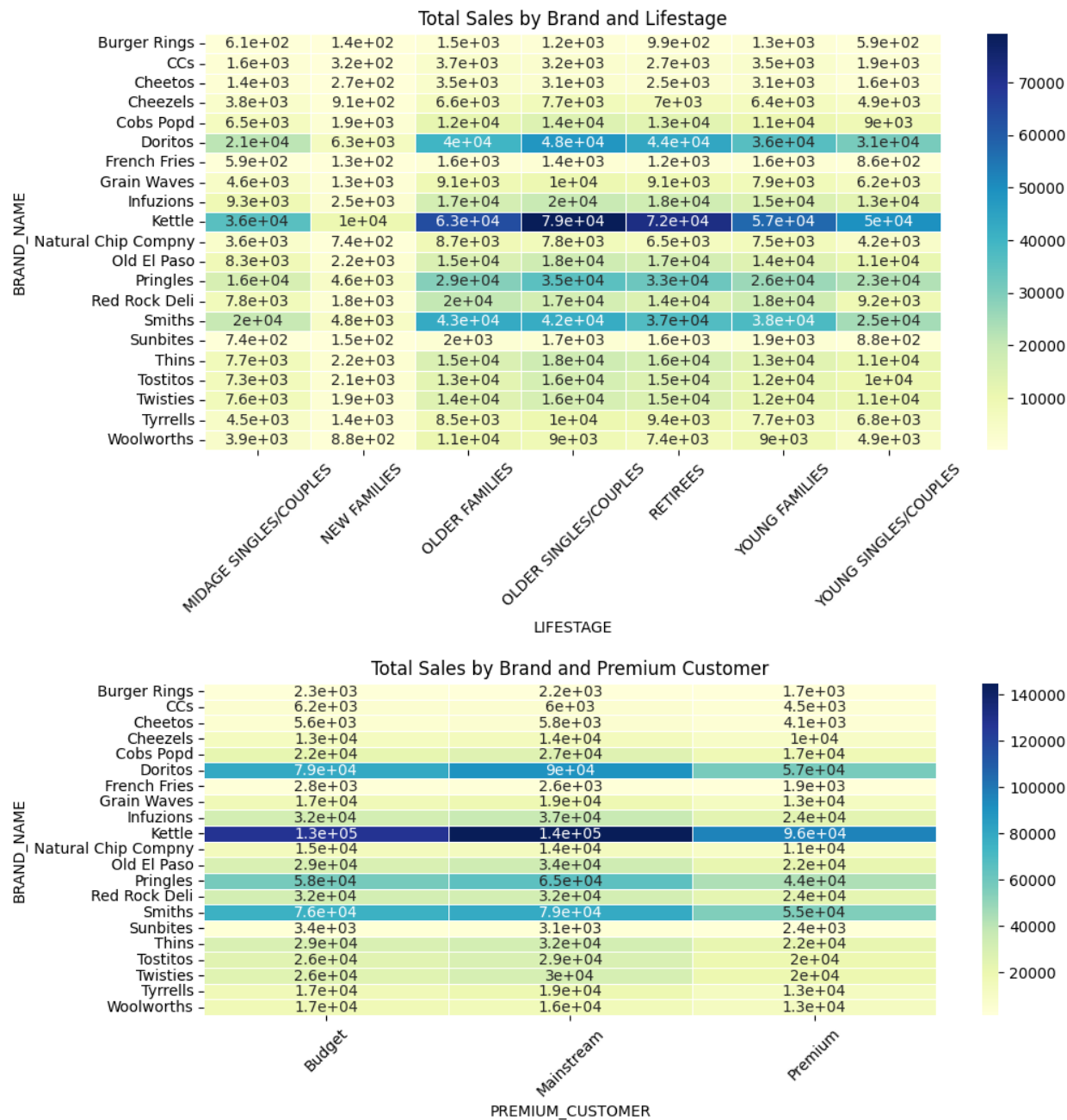
Some keys findings are that new families have much lower total sales compared to the other lifestyle groups, and older people tend to have higher total sales. For the premium customer category, we can see that mainstream customers have the highest total sales, and premium has the lowest. For the brands Kettle, Doritos, and Smiths have the highest total sales. Burger Rings, French Fries, and Sunbites have the lowest total sales. I also created a breakdown of total sales per brand for these 2 customer segments.



For the brands with the top total sales, young people and young families, along with older people are contributing the most towards the sales. Mainstream and budget customers are also contributing more than premium customers. To further understand which brands are contributing most towards sales I created a Pareto chart.

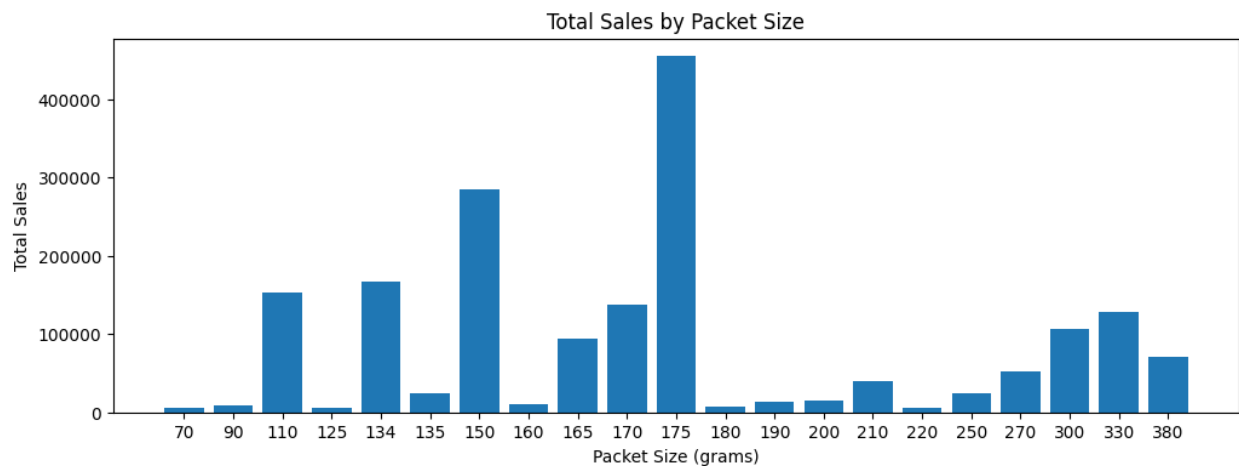


From this we can see that Kettle, Doritos, Smiths, and Pringles make up over 50% of the total sales. Lastly, I created pivot tables to see a more detailed breakdown of total sales for lifestage and premium for each brand. The numbers represent the total sales, and the colors show the intensity of sales, with darker colors meaning more sales, and therefore groups that should be targeted.

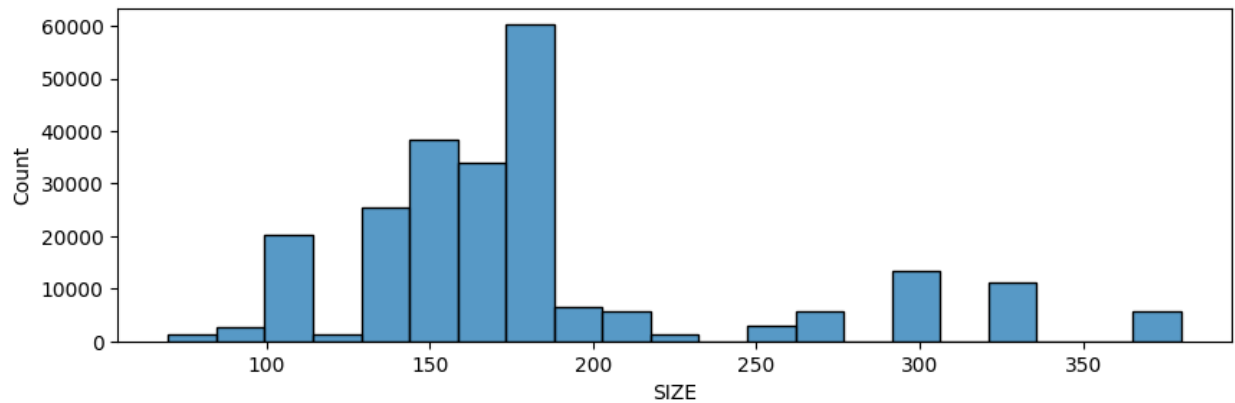


From all these results I recommend the manager focuses on selling Kettle, Doritos, Smiths, and Pringles brand chips. These brands are the drivers of sales and will help make the most money for the store. To do this, advertisements should be targeted towards older customers (older families, older singles/couples, and retirees) along with young families and young singles/couples, since these customer groups purchase the most from those top brands. Furthermore, budget and mainstream customer groups should be prioritized, since the premium group has lower total sales

for the top brands. Lastly, I investigated the packet size. I first looked at total sales per packet size. 175g has the largest total sales per packet size, with 150g coming in second.



I also investigated the distribution of packet sizes, to see if the size was evenly distributed.



Packet size is not evenly distributed. Most of the packets are between 100 to 200g.

I also calculated the correlation between packet size and total sales, which was 0.44. This means that there is a positive association between total sales. The manager should prioritize packets with sizes 100g- 200g since these have the highest sales, along with the previously mentioned brands and customer segments to target.

For more detailed analysis and further visualizations of the customer segments and packet size please see the python notebook for question 2. The notebook also includes all of my relevant work for data cleaning, processing (removing outliers), and high-level data summaries, including how I extracted brand name and packet size.