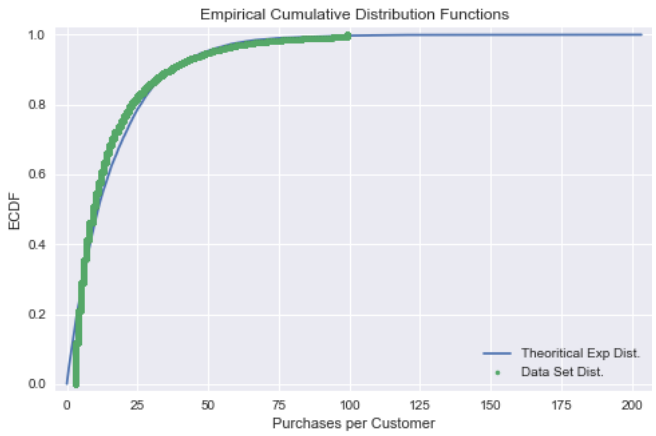
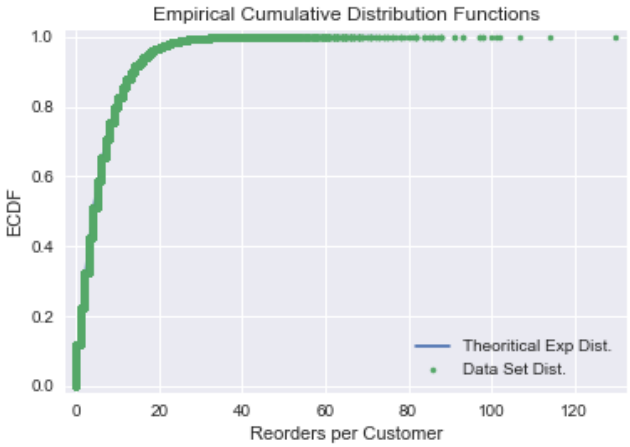
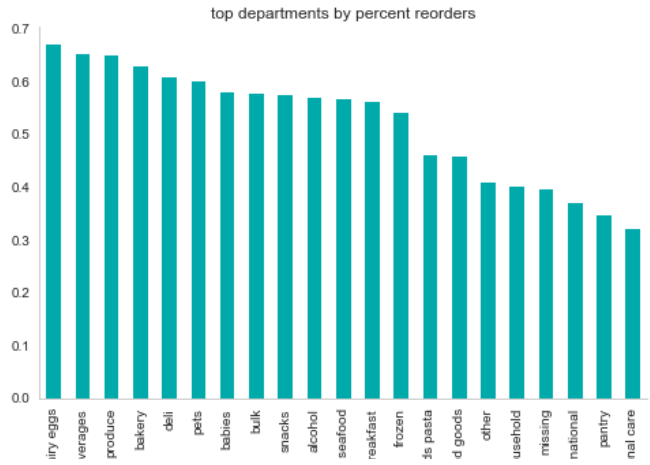
Beginning with some basic statistics, we start using the describe attribute to get the feature's mean , standard deviation, median, and quartiles. We are interested in reorders. So lets look a purchases. Counting up and plotting purchases and reorders we see a classic exponential decay suggesting an exponential distribution might be a good approximation of the purchases and reorder distributions. Plotting these overlaid on one another we see that an exponential distribution does fit the data well.

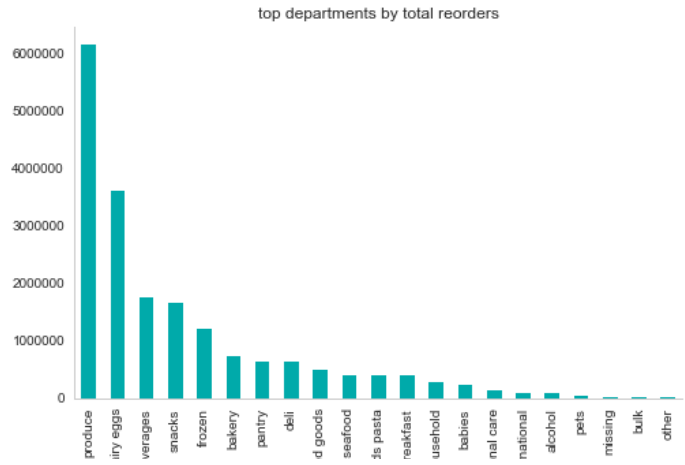




Using the properties of the exponential distribution, some probability questions can be easily answered. We see that 99% of customers purchased at least 74 items at each order, 95% purchased 48 items and 66% purchased at least 17 items. This gives us a sense of what a typical customer looks like and could help businesses make appropriate decisions in relation to the majority of their typical customers. The also helps in identifying outliers in the data, what are large purchases and what are large reorder amounts. Not surprising we see that the number of purchases has a large correlation to the amount of reorders. The more you buy things the more likely you are to buy things again, not surprising but reaffirming. We will later as we develop more features test correlation again.

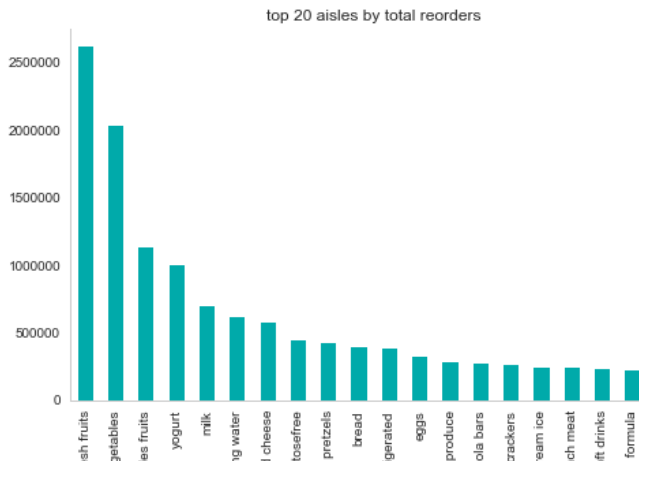
More interesting things from the data exploration, we see the top reorder products and top purchases. Not surprising again there is a heavy correlation, the more things are reordered the more likely those items will be the most purchased items. A retailer will definitely want to be on both of these lists. Looking at these lists we see Bananas at the top. Customer who like bananas seem to reorder bananas and purchase a lot of bananas. We are interested in predicting reorders so lets focus on that topic. We see that 60% of items in a basket are typically reorders. We see from the graphs below that dairy eggs, beverages, and produce lead the department in both volume of orders and portion of reorders.

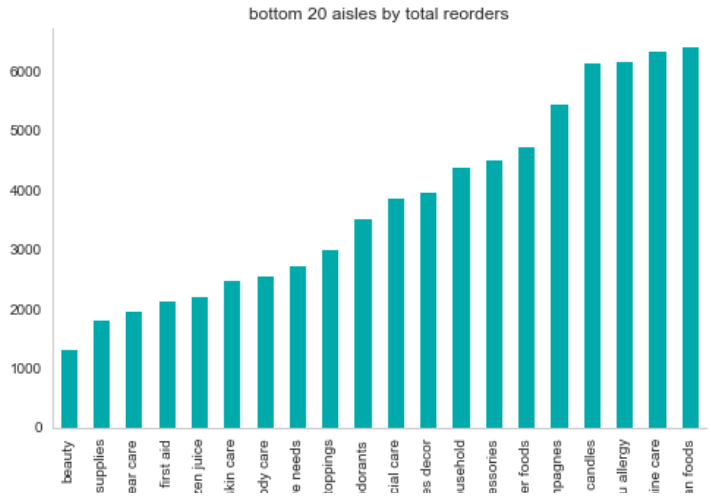


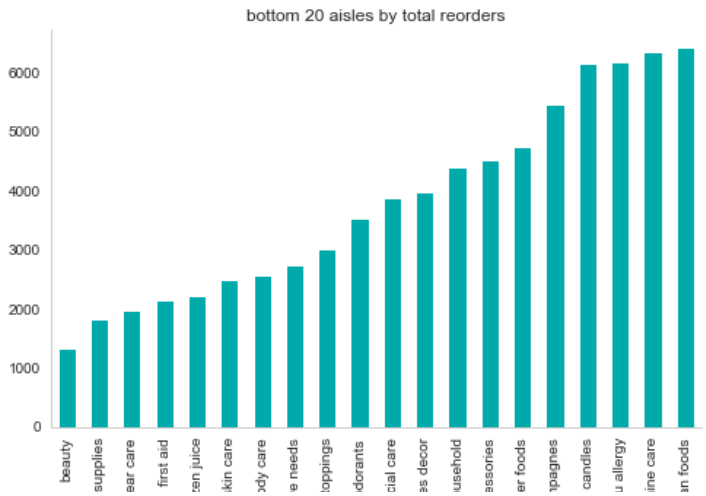


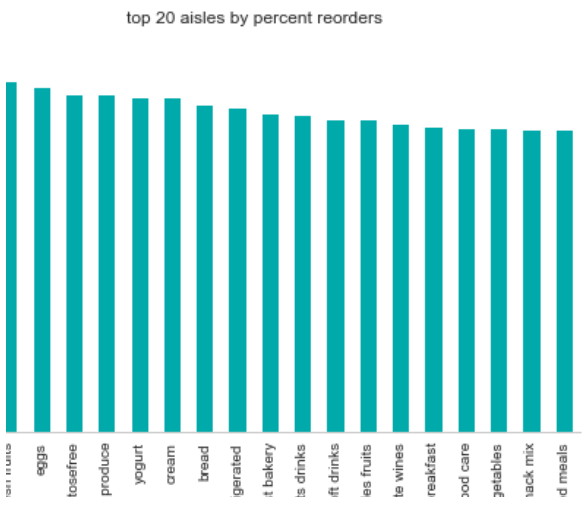
The aisles with the most reorders by portion is:

milk	0.781428
water seltzer sparkling water	0.729593
fresh fruits	0.718104
eggs	0.705366
soy lactose free	0.692551









The bottom aisles for reorders are

spices seasonings	0.152391
baking supplies decor	0.167229
first aid	0.194812
kitchen supplies	0.195377
beauty	0.212062

It is interesting to note that the percent reorders by department is different from the total reorders by department. The largest take away we see is that perishable items seem to drive the reorder purchase habits.