BK and MD Nutritional Menu Analysis

Shaquiel Pashtunyar

DSC0530 Summer 2022

The analysis I performed was to look at the menu of McDonalds and Burger King to identify if there was any correlation between the nutritional facts on their menu. Knowing that these are two popular fast-food chains that have high caloric menu items, I hypothesized that there would be a correlation with the rise in calories to that of the fat content. It was also possible that there may be a correlation between other variables in the data set, but this is one that I had a strong feeling about. After looking into the data, I saw that other caloric items, like sugar or protein didn’t appear to have the strongest trend with calories. The reason for this was that items on the menu like soda would heavily affect these numbers, or at least that was a theory of mine.

I did find a good correlation between fat content and calories when I started to make scatter plots, but I ran into an issue. The issue in my EDA was that there were clearly some outliers in the data set for fat content, otherwise it looked like a very linear trend. I made a second dataset dropping any menu item that had sausage or bacon, the root cause of the high fat content items. By doing this my correlation analysis went from 0.24 to 0.84. The difference led to a very linear looking chart with correlation numbers that proved my initial hypothesis. We can take our study further by looking at looking at linear and regression models that helped drive these ideas home.

Looking back at this study, looking at the entire menu of each of the fast-food restaurants might not have been the best idea. I can speculate that the drinks or desserts on the menu will have high sugar content and low fats or proteins. We saw the effect of having breakfast food on the menu and how that made our clear correlation non-relevant. This speculation may be true, and the driving factor of many high calorie zero fat items. It would be a good idea to filter the data set next time and try to make correlations based on the categories of items, rather than the whole menu. If I were to redo this study, I would break the data into 4-5 categories, and be able to dig much deeper into the analysis of each. This would lead me to a more diverse range of variables and correlations that I can graph out and study.

One challenge I faced was the hypothesis testing. I tried to run a permutation correlation, but it came back with the value of zero. I can see from Pearson and Spearman correlations that this is not accurate so I likely built the model wrong but could not solve how to fix it. In the end my hypothesis was correct, and I learned a lot about the correlation, the effect of outliers, and the data set in general by using an EDA approach to taking a problem, gathering data, analyzing it, creating a model and drawing some conclusions.