McDonalds vs BK DSC520 Final Project 2

Shaquiel Pashtunyar

2022-08-08

Final Project Step 2

- I downloaded 3 data sets, all from Kaggle in order to get the raw data for this project
- The main data set I plan to use, has nutritional values for both BK and McDonalds

How to import and clean my data

I have 3 CSVs that I wish to import, all of them are saved to my local desktop The nutrional_values CSV was formatted incorrectly where the csv delimter was a semicolon (;) and not a comma (,). I went ahead and fixed this within before trying to upload the files

To clean the data, I dropped all the columns that did not match, I then renamed the matching columns so they had the same name. Also adding the source for each dataset helps me find the variable to compare against The last step was to merge the various CSVs into one datatable that contained the nutritonal values from all 3 data sources

What does the final data set look like?

The final dataset is a combined data table of values from all the CSVs. It has all the columns renamed to be easily readable and extra columns/data has been removed. Look into the RMD markdown file to see all edits the dataset, but here is the head of the final table

```
##         Chain                                Item               Type
## 1 Burger King                    Whopper Sandwich Whopper Sandwiches
## 2 Burger King         Whopper Sandwich with Cheese Whopper Sandwiches
## 3 Burger King      Bacon & Cheese Whopper Sandwich Whopper Sandwiches
## 4 Burger King               Double Whopper Sandwich Whopper Sandwiches
## 5 Burger King Double Whopper Sandwich with Cheese Whopper Sandwiches
## 6 Burger King              Triple Whopper Sandwich Whopper Sandwiches
##   Serving.Size.(g) Calories Total.Fat.(g) Saturated.Fat.(g) Trans.Fat.(g)
## 1              270      660            40                12             1
## 2              292      740            46                16             2
## 3              303      790            51                17             2
## 4              354      900            58                20             3
## 5              377      980            64                24             3
## 6              438     1130            75                28             4
##   Chol.(mg) Sodium.(mg) Total.Carb (g) Total.Sugar.(g) Protein.(g)
## 1         5          90            980                2          11
## 2       115        1340             50               11          32
## 3       125        1560             50               11          35
## 4       175        1050             49               11          48
```

```
## 5        195         1410              50              11            52
## 6        255         1120              49              11            67
```

Questions to answer

- Questions for future steps.
- What information is not self-evident?
- What are different ways you could look at this data?
- How do you plan to slice and dice the data?
- How could you summarize your data to answer key questions?
- What types of plots and tables will help you to illustrate the findings to your questions?
- Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.
- Questions for future steps.

Answers

- I need to be able to select specific items from the data frame and index them in order to make my comparisons, which will not be easy. I have to find a whooper, big mac, and an indian equivalent to compare

- What was not self evident was the chain data in the India data set and this needed to be added in so we could have 1 data frame

- You can start to look at this by making histograms or charts of various key items on all 3 menus to see if the caloric differences are present, like comparing all the bugers at each joint, then the fries

- The data will likely be sliced by category. Like the above statement, french fries is an easy one. Comparing medium fries at all 3 can tell us a lot about the sodium and calorie levels

- The data will likely be summarized by creating plots, graphs, and comparisons between various key items in order to visualize the differences and compare them

- I plan to make histograms for the values within one chain, and bar charts for the different caloric values

- I will consider some ML techniques as I work on week 10 to see if any can help me with my analysis

- For future steps I think the most important thing I can do is go further with my data manipulation. Split the data into smaller data frames with only the important information so I can create graphs and charts that easily compare the values between items. This is going to take a lot of time doing the backend processing work, but if successful, the cleanliness of the data will shine and I will be able to

make my calculations, coorelations, and charts very easily to answer all of my questions