Week 2 Assignment

Jonathan Ibifubara Pollyn

College of Science, Engineering and Technology, Grand Canyon University

DSC-530: Predictive Modeling

Filippo Posta

September 15, 2021

Data Preparation

The result shows that the list of the first five food items with the highest saturated fat content is,

Cheesecake 1 cake, Ice cream; vanilla; rich 16% ft1/2 gal,Yellowcake w/ chocfrstng; comml1

cake,Creme pie 1 pie,Lard 1 cup It also shows that the list of the five food items with

the highest content of saturated fat per gram is provided in the output of the

explanation part. Food items with the highest saturated fat per gram are "butter; salted

1 TBSP" and "butter; unsalted 1 tbsp". When it comes to cholesterol, the list of the five food

items with the highest cholesterol content per gram provided in the output from the results are

food items like the "eggs; raw; yolk 1 yolk". When checking for outliers, we find those food

items that are outliers at the higher end of the scale provided in the output are 63.

Data Visualization

The Contingency table is a table used as temporal storage of data to be used later in our

programs. The contingency table for the marital and response stores the information from the

marital variable and their response. It shows from the output that people who identified as being

married are more than double compared to people identified as divorced 14579 to 2743. While

only 50 people confirmed they do not know. When comparing the percentages for the marital

and response on the contingency table, it is shown that the percentages are relatively close with

no mush difference as seen without the percentage. The histogram is used to display the marital

and response, and the married and divorced are both the same. Now, when we plot the bar with

an overlay, the number of responses for married people is more. The peak of the graph is right

around 6000 on the graph.

Conclusion

The use of data preparation was demonstrated in this assignment and the importance of of adhering to that practice is key to data science projects. Data visualization helps us to visually look at our data to extract meaningful insight from the data

References List

Larose, C. D., Larose, D. T., & Larose, Chantal D., Author. (2019). *Data science using python and r.* John Wiley & Sons,inc,.