

Notes On Data

Derek R Neilson

September 24, 2024

Abstract

This document contains notes on the data. The notes are intended to demonstrate how I filter and manipulate the data, and are purely for my instructor to review.

1 Introduction

In data analysis, the ability to effectively filter and manipulate data is crucial for extracting meaningful insights. This document outlines the methodologies and tools I employ to pre-process and analyze the dataset. The primary focus is on cleaning the data, handling missing values, and transforming data to suit the analytical objectives. These notes serve as a comprehensive guide for understanding my data processing workflow.

2 Data Collection

The data was collected from <https://fdc.nal.usda.gov/>. The dataset is 2.9GB and is labeled **Branded** and is in JSON format. I chose this dataset because it is large and one can assume that it has the most rows because it is so large.

To download the data, I used the following commands:

```
wget https://fdc.nal.usda.gov/fdc-datasets/  
    ↪ FoodData_Central_branded_food_csv_2024-04-18.zip  
unzip FoodData_Central_branded_food_csv_2024-04-18.zip
```