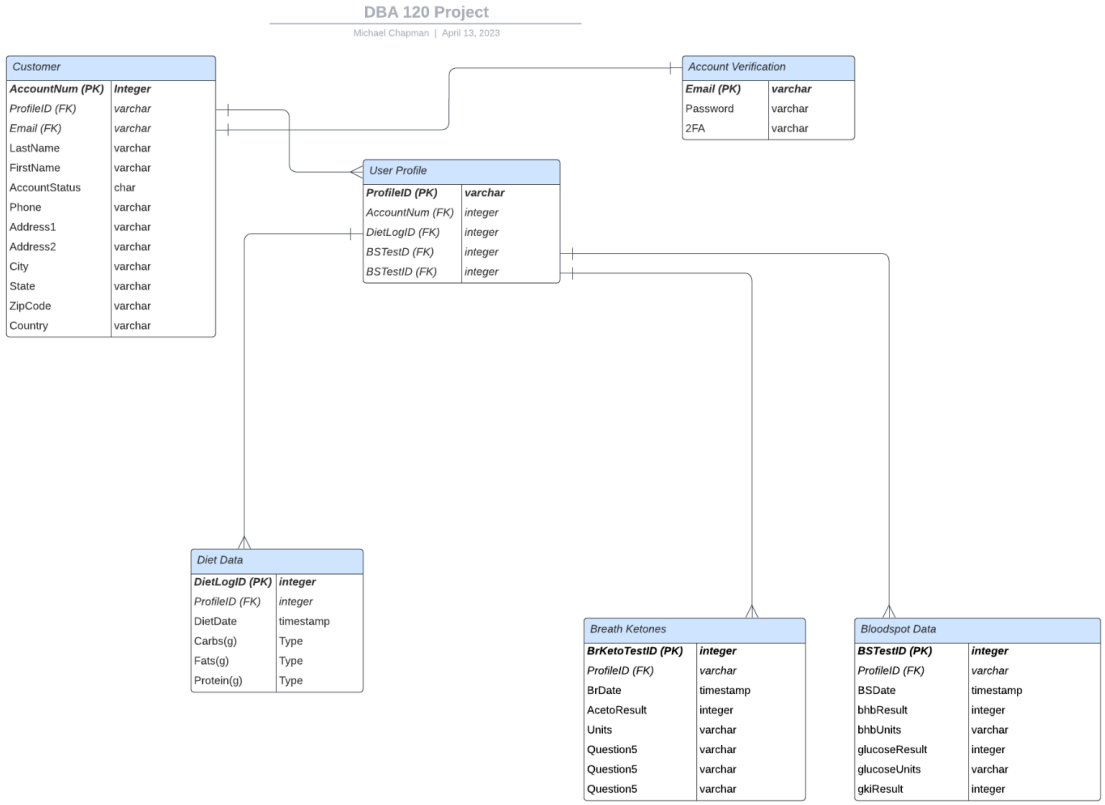
Michael Chapman
DBA-120 Project

## Project Overview

For my project, I wanted to create a database to be able to store and run analytics on health data. I have recently done an experiment where I went on a ketogenic (low carbohydrate) diet and tracked multiple indices of health data. I tracked biometric data such as weight along with blood glucose, blood and breath ketones, and diet macronutrient data. I wanted to build something that would work as a repository for this information.

## ERD

As posted in GitHub, below is an image of the database model for this project. This was used to create the database.

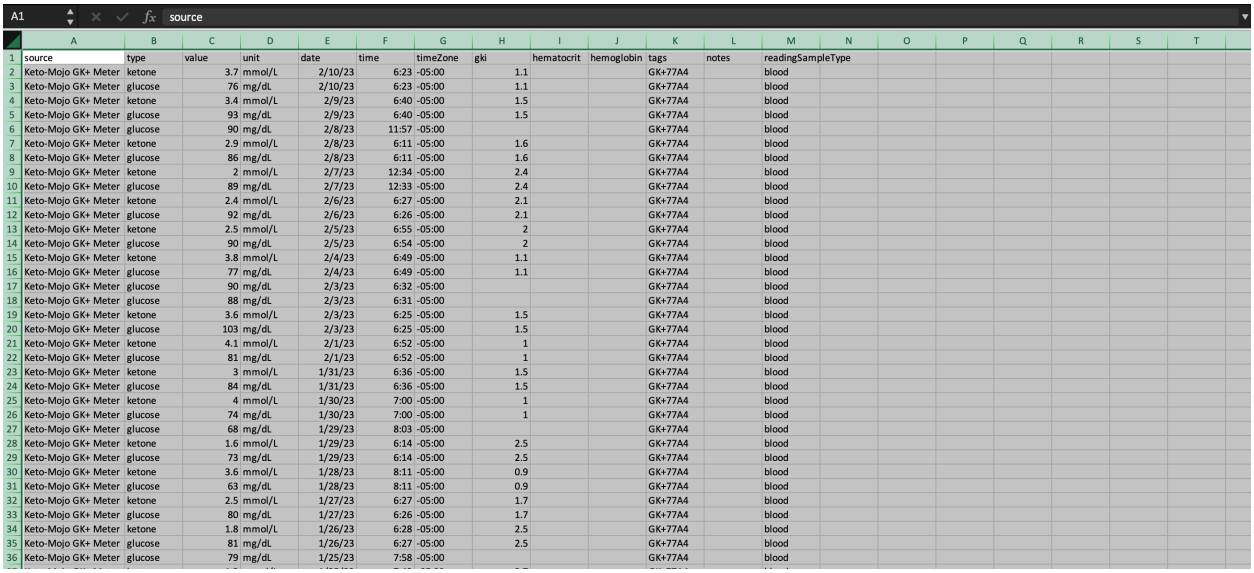



## Part 2 – ETL Operations

I first had to extract data from multiple sources to input the data into the database. I extracted data from three software applications (KetoMojo, Biosense, and Cronometer). These files were in the form of .csv and required transformation to load properly into the mySQL database.

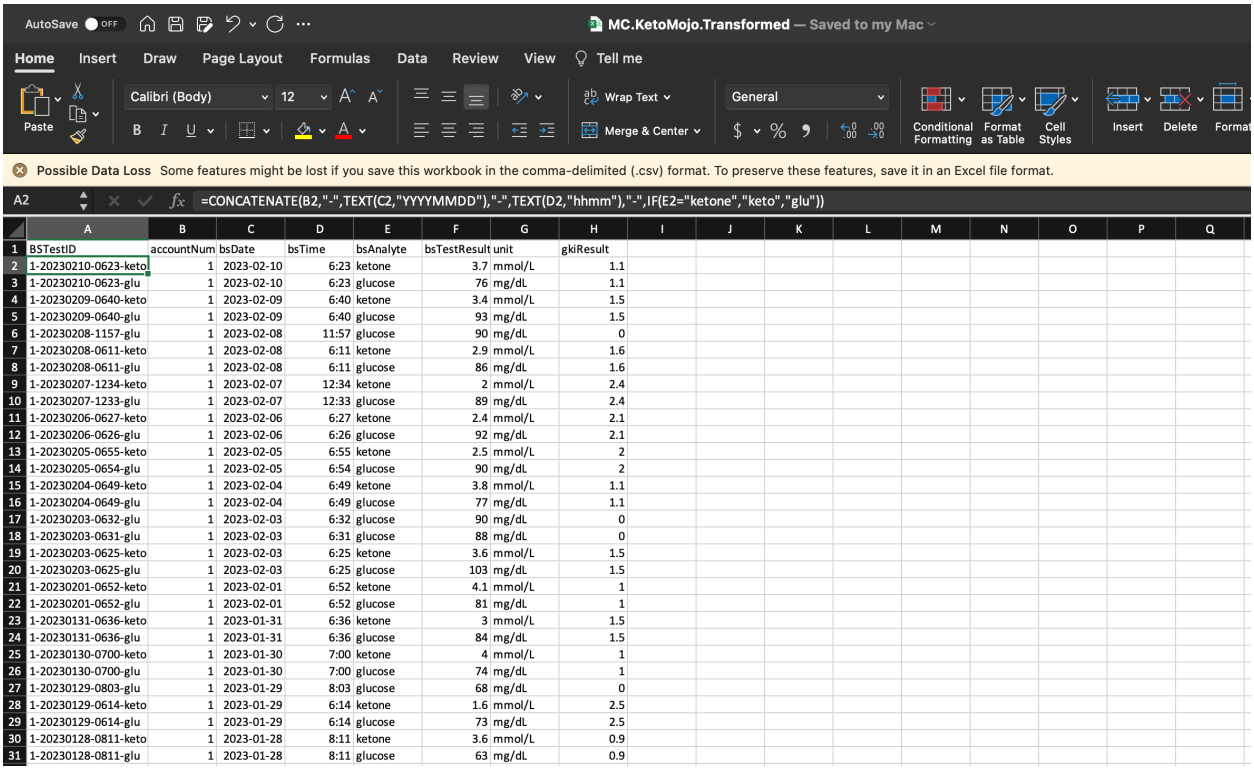
Michael Chapman
DBA-120 Project

## **Ketomojo Transformation**



| source | type | value | unit | date | time | timeZone | gki | hematocrit | hemoglobin | tags | notes | readingSampleType |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Keto-Mojo GK+ Meter | ketone | 3.7 | mmol/L | 2/10/23 | 6:23 | -05:00 | 1.1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 76 | mg/dL | 2/10/23 | 6:23 | -05:00 | 1.1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 3.4 | mmol/L | 2/9/23 | 6:40 | -05:00 | 1.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 93 | mg/dL | 2/9/23 | 6:40 | -05:00 | 1.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 90 | mg/dL | 2/8/23 | 11:57 | -05:00 | | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 2.9 | mmol/L | 2/8/23 | 6:11 | -05:00 | 1.6 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 86 | mg/dL | 2/8/23 | 6:11 | -05:00 | 1.6 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 2 | mmol/L | 2/7/23 | 12:34 | -05:00 | 2.4 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 89 | mg/dL | 2/7/23 | 12:33 | -05:00 | 2.4 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 2.4 | mmol/L | 2/6/23 | 6:27 | -05:00 | 2.1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 92 | mg/dL | 2/6/23 | 6:26 | -05:00 | 2.1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 2.5 | mmol/L | 2/5/23 | 6:55 | -05:00 | 2 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 90 | mg/dL | 2/5/23 | 6:54 | -05:00 | 2 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 3.8 | mmol/L | 2/4/23 | 6:49 | -05:00 | 1.1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 77 | mg/dL | 2/4/23 | 6:49 | -05:00 | 1.1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 90 | mg/dL | 2/3/23 | 6:32 | -05:00 | | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 88 | mg/dL | 2/3/23 | 6:31 | -05:00 | | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 3.6 | mmol/L | 2/3/23 | 6:25 | -05:00 | 1.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 103 | mg/dL | 2/3/23 | 6:25 | -05:00 | 1.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 4.1 | mmol/L | 2/1/23 | 6:52 | -05:00 | 1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 81 | mg/dL | 2/1/23 | 6:52 | -05:00 | 1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 3 | mmol/L | 1/31/23 | 6:36 | -05:00 | 1.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 84 | mg/dL | 1/31/23 | 6:36 | -05:00 | 1.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 4 | mmol/L | 1/30/23 | 7:00 | -05:00 | 1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 74 | mg/dL | 1/30/23 | 7:00 | -05:00 | 1 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 68 | mg/dL | 1/29/23 | 8:03 | -05:00 | | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 1.6 | mmol/L | 1/29/23 | 6:14 | -05:00 | 2.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 73 | mg/dL | 1/29/23 | 6:14 | -05:00 | 2.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 3.6 | mmol/L | 1/28/23 | 8:11 | -05:00 | 0.9 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 63 | mg/dL | 1/28/23 | 8:11 | -05:00 | 0.9 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 2.5 | mmol/L | 1/27/23 | 6:27 | -05:00 | 1.7 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 80 | mg/dL | 1/27/23 | 6:26 | -05:00 | 1.7 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | ketone | 1.8 | mmol/L | 1/26/23 | 6:28 | -05:00 | 2.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 81 | mg/dL | 1/26/23 | 6:27 | -05:00 | 2.5 | | | GK+77A4 | | blood |
| Keto-Mojo GK+ Meter | glucose | 79 | mg/dL | 1/25/23 | 7:58 | -05:00 | | | | GK+77A4 | | blood |

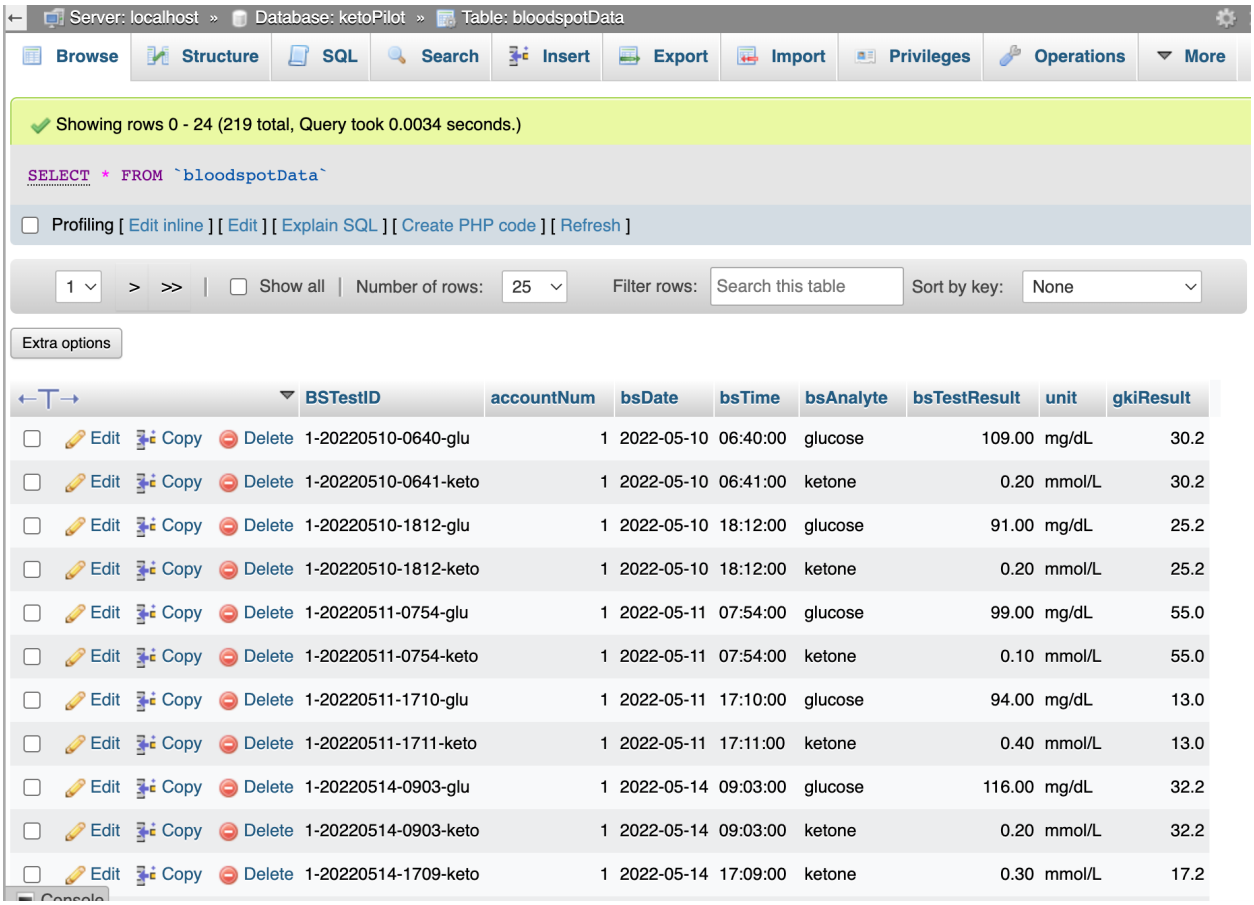
The above CSV had unnecessary columns that were deleted. Also, the primary key needed to be created which was done using the formula listed in a concatenate function (listed in the image below). This also required the transformation of some of the data, such as data, into a different formation.




| BSTestID | accountNum | bsDate | bsTime | bsAnalyte | bsTestResult | unit | gkiResult |
|---|---|---|---|---|---|---|---|
| 1-20230210-0623-keto | 1 | 2023-02-10 | 6:23 | ketone | 3.7 | mmol/L | 1.1 |
| 1-20230210-0623-glu | 1 | 2023-02-10 | 6:23 | glucose | 76 | mg/dL | 1.1 |
| 1-20230209-0640-keto | 1 | 2023-02-09 | 6:40 | ketone | 3.4 | mmol/L | 1.5 |
| 1-20230209-0640-glu | 1 | 2023-02-09 | 6:40 | glucose | 93 | mg/dL | 1.5 |
| 1-20230208-1157-glu | 1 | 2023-02-08 | 11:57 | glucose | 90 | mg/dL | 0 |
| 1-20230208-0611-keto | 1 | 2023-02-08 | 6:11 | ketone | 2.9 | mmol/L | 1.6 |
| 1-20230208-0611-glu | 1 | 2023-02-08 | 6:11 | glucose | 86 | mg/dL | 1.6 |
| 1-20230207-1234-keto | 1 | 2023-02-07 | 12:34 | ketone | 2 | mmol/L | 2.4 |
| 1-20230207-1233-glu | 1 | 2023-02-07 | 12:33 | glucose | 89 | mg/dL | 2.4 |
| 1-20230206-0627-keto | 1 | 2023-02-06 | 6:27 | ketone | 2.4 | mmol/L | 2.1 |
| 1-20230206-0626-glu | 1 | 2023-02-06 | 6:26 | glucose | 92 | mg/dL | 2.1 |
| 1-20230205-0655-keto | 1 | 2023-02-05 | 6:55 | ketone | 2.5 | mmol/L | 2 |
| 1-20230205-0654-glu | 1 | 2023-02-05 | 6:54 | glucose | 90 | mg/dL | 2 |
| 1-20230204-0649-keto | 1 | 2023-02-04 | 6:49 | ketone | 3.8 | mmol/L | 1.1 |
| 1-20230204-0649-glu | 1 | 2023-02-04 | 6:49 | glucose | 77 | mg/dL | 1.1 |
| 1-20230203-0632-glu | 1 | 2023-02-03 | 6:32 | glucose | 90 | mg/dL | 0 |
| 1-20230203-0631-glu | 1 | 2023-02-03 | 6:31 | glucose | 88 | mg/dL | 0 |
| 1-20230203-0625-keto | 1 | 2023-02-03 | 6:25 | ketone | 3.6 | mmol/L | 1.5 |
| 1-20230203-0625-glu | 1 | 2023-02-03 | 6:25 | glucose | 103 | mg/dL | 1.5 |
| 1-20230201-0652-keto | 1 | 2023-02-01 | 6:52 | ketone | 4.1 | mmol/L | 1 |
| 1-20230201-0652-glu | 1 | 2023-02-01 | 6:52 | glucose | 81 | mg/dL | 1 |
| 1-20230131-0636-keto | 1 | 2023-01-31 | 6:36 | ketone | 3 | mmol/L | 1.5 |
| 1-20230131-0636-glu | 1 | 2023-01-31 | 6:36 | glucose | 84 | mg/dL | 1.5 |
| 1-20230130-0700-keto | 1 | 2023-01-30 | 7:00 | ketone | 4 | mmol/L | 1 |
| 1-20230130-0700-glu | 1 | 2023-01-30 | 7:00 | glucose | 74 | mg/dL | 1 |
| 1-20230129-0803-glu | 1 | 2023-01-29 | 8:03 | glucose | 68 | mg/dL | 0 |
| 1-20230129-0614-keto | 1 | 2023-01-29 | 6:14 | ketone | 1.6 | mmol/L | 2.5 |
| 1-20230129-0614-glu | 1 | 2023-01-29 | 6:14 | glucose | 73 | mg/dL | 2.5 |
| 1-20230128-0811-keto | 1 | 2023-01-28 | 8:11 | ketone | 3.6 | mmol/L | 0.9 |
| 1-20230128-0811-glu | 1 | 2023-01-28 | 8:11 | glucose | 63 | mg/dL | 0.9 |

This was then imported into mySQL databased:

Michael Chapman
DBA-120 Project

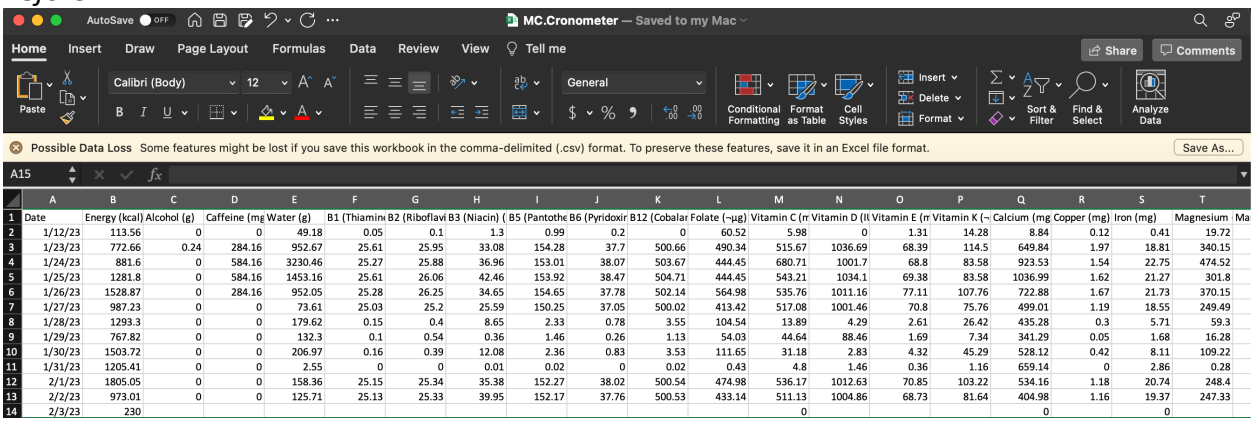



| | BSTestID | accountNum | bsDate | bsTime | bsAnalyte | bsTestResult | unit | gkiResult |
|---|---|---|---|---|---|---|---|---|
| Edit Copy Delete | 1-20220510-0640-glu | 1 | 2022-05-10 | 06:40:00 | glucose | 109.00 | mg/dL | 30.2 |
| Edit Copy Delete | 1-20220510-0641-keto | 1 | 2022-05-10 | 06:41:00 | ketone | 0.20 | mmol/L | 30.2 |
| Edit Copy Delete | 1-20220510-1812-glu | 1 | 2022-05-10 | 18:12:00 | glucose | 91.00 | mg/dL | 25.2 |
| Edit Copy Delete | 1-20220510-1812-keto | 1 | 2022-05-10 | 18:12:00 | ketone | 0.20 | mmol/L | 25.2 |
| Edit Copy Delete | 1-20220511-0754-glu | 1 | 2022-05-11 | 07:54:00 | glucose | 99.00 | mg/dL | 55.0 |
| Edit Copy Delete | 1-20220511-0754-keto | 1 | 2022-05-11 | 07:54:00 | ketone | 0.10 | mmol/L | 55.0 |
| Edit Copy Delete | 1-20220511-1710-glu | 1 | 2022-05-11 | 17:10:00 | glucose | 94.00 | mg/dL | 13.0 |
| Edit Copy Delete | 1-20220511-1711-keto | 1 | 2022-05-11 | 17:11:00 | ketone | 0.40 | mmol/L | 13.0 |
| Edit Copy Delete | 1-20220514-0903-glu | 1 | 2022-05-14 | 09:03:00 | glucose | 116.00 | mg/dL | 32.2 |
| Edit Copy Delete | 1-20220514-0903-keto | 1 | 2022-05-14 | 09:03:00 | ketone | 0.20 | mmol/L | 32.2 |
| Edit Copy Delete | 1-20220514-1709-keto | 1 | 2022-05-14 | 17:09:00 | ketone | 0.30 | mmol/L | 17.2 |

## Cronometer/Biosense Transformation

Similar extraction and transformations were executed for both the other application data sources. With all three of these transformations, some edits had to be made to the database setup to allow for appropriate primary key data types. Usually, this was changing the primary key from an INT type to a VARCHAR type. Otherwise, the primary key values were not as intuitive as they could be.

*Before*




| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Date | Energy (kcal) | Alcohol (g) | Caffeine (mg | Water (g) | B1 (Thiamin | B2 (Riboflavi | B3 (Niacin) ( | B5 (Pantothe | B6 (Pyridoxir | B12 (Cobalar | Folate (¬µg) | Vitamin C (r | Vitamin D (IU | Vitamin E (r | Vitamin K (¬ | Calcium (mg | Copper (mg) | Iron (mg) | Magnesium | Ma |
| 2 | 1/12/23 | 113.56 | 0 | 0 | 49.18 | 0.05 | 0.1 | 1.3 | 0.99 | 0.2 | 0 | 60.52 | 5.98 | 0 | 1.31 | 14.28 | 8.84 | 0.12 | 0.41 | 19.72 | |
| 3 | 1/23/23 | 772.66 | 0.24 | 284.16 | 952.67 | 25.61 | 25.95 | 33.08 | 154.28 | 37.7 | 500.66 | 490.34 | 515.67 | 1036.69 | 68.39 | 114.5 | 649.84 | 1.97 | 18.81 | 340.15 | |
| 4 | 1/24/23 | 881.6 | 0 | 584.16 | 3230.46 | 25.27 | 25.88 | 36.96 | 153.01 | 38.07 | 503.67 | 444.45 | 680.71 | 1001.7 | 68.8 | 83.58 | 923.53 | 1.54 | 22.75 | 474.52 | |
| 5 | 1/25/23 | 1281.8 | 0 | 584.16 | 1453.16 | 25.61 | 26.06 | 42.46 | 153.92 | 38.47 | 504.71 | 444.45 | 543.21 | 1034.1 | 69.38 | 83.58 | 1036.99 | 1.62 | 21.27 | 301.8 | |
| 6 | 1/26/23 | 1528.87 | 0 | 284.16 | 952.05 | 25.28 | 26.25 | 34.65 | 154.65 | 37.78 | 502.14 | 564.98 | 535.76 | 1011.16 | 77.11 | 107.76 | 722.88 | 1.67 | 21.73 | 370.15 | |
| 7 | 1/27/23 | 987.23 | 0 | 0 | 73.61 | 25.03 | 25.2 | 25.59 | 150.25 | 37.05 | 500.02 | 413.42 | 517.08 | 1001.46 | 70.8 | 75.76 | 499.01 | 1.19 | 18.55 | 249.49 | |
| 8 | 1/28/23 | 1293.3 | 0 | 0 | 179.62 | 0.15 | 0.4 | 8.65 | 2.33 | 0.78 | 3.55 | 104.54 | 13.89 | 4.29 | 2.61 | 26.42 | 435.28 | 0.3 | 5.71 | 59.3 | |
| 9 | 1/29/23 | 767.82 | 0 | 0 | 132.3 | 0.1 | 0.54 | 0.36 | 1.46 | 0.26 | 1.13 | 54.03 | 44.64 | 88.46 | 1.69 | 7.34 | 341.29 | 0.05 | 1.68 | 16.28 | |
| 10 | 1/30/23 | 1503.72 | 0 | 0 | 206.97 | 0.16 | 0.39 | 12.08 | 2.36 | 0.83 | 3.53 | 111.65 | 31.18 | 2.83 | 4.32 | 45.29 | 528.12 | 0.42 | 8.11 | 109.22 | |
| 11 | 1/31/23 | 1205.41 | 0 | 0 | 2.55 | 0 | 0 | 0.01 | 0.02 | 0 | 0.02 | 0.43 | 4.8 | 1.46 | 0.36 | 1.16 | 659.14 | 0 | 2.86 | 0.28 | |
| 12 | 2/1/23 | 1805.05 | 0 | 0 | 158.36 | 25.15 | 25.34 | 35.38 | 152.27 | 38.02 | 500.54 | 474.98 | 536.17 | 1012.63 | 70.85 | 103.22 | 534.16 | 1.18 | 20.74 | 248.4 | |
| 13 | 2/2/23 | 973.01 | 0 | 0 | 125.71 | 25.13 | 25.33 | 39.95 | 152.17 | 37.76 | 500.53 | 433.14 | 511.13 | 1004.86 | 68.73 | 81.64 | 404.98 | 1.16 | 19.37 | 247.33 | |
| 14 | 2/3/23 | 230 | | | | | | | | | | | 0 | | | | 0 | | 0 | | |

Michael Chapman
DBA-120 Project

*After*

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | dietLogID | accountNum | dietDate | energyCalori | carbGrams | fatGrams | proteinGrams | | | | | | | |
| 2 | dietLog1-202 | 1 | 1/12/23 | 113.56 | 1.22 | 10.48 | 1.33 | | | | | | | |
| 3 | dietLog1-202 | 1 | 1/23/23 | 772.66 | 20.65 | 52.03 | 50.42 | | | | | | | |
| 4 | dietLog1-202 | 1 | 1/24/23 | 881.6 | 20.28 | 54.24 | 71.23 | | | | | | | |
| 5 | dietLog1-202 | 1 | 1/25/23 | 1281.8 | 19.89 | 87.23 | 104.19 | | | | | | | |
| 6 | dietLog1-202 | 1 | 1/26/23 | 1528.87 | 20.08 | 129.07 | 60.4 | | | | | | | |
| 7 | dietLog1-202 | 1 | 1/27/23 | 987.23 | 14.16 | 82.04 | 44.57 | | | | | | | |
| 8 | dietLog1-202 | 1 | 1/28/23 | 1293.3 | 15.92 | 100.2 | 73.61 | | | | | | | |
| 9 | dietLog1-202 | 1 | 1/29/23 | 767.82 | 7.38 | 68.35 | 26.34 | | | | | | | |
| 10 | dietLog1-202 | 1 | 1/30/23 | 1503.72 | 20.52 | 108.58 | 100.16 | | | | | | | |
| 11 | dietLog1-202 | 1 | 1/31/23 | 1205.41 | 13 | 100.58 | 60.21 | | | | | | | |
| 12 | dietLog1-202 | 1 | 2/1/23 | 1805.05 | 15.23 | 151.02 | 85.07 | | | | | | | |
| 13 | dietLog1-202 | 1 | 2/2/23 | 973.01 | 9.69 | 66.16 | 76.59 | | | | | | | |
| 14 | dietLog1-202 | 1 | 2/3/23 | 230 | 1 | 18 | 16 | | | | | | | |

These two data sets were loaded into the database after these transformations.

The last three tables were input manually with the following code:

```
63  INSERT INTO customers VALUES
64      (DEFAULT,'michaeljchapman@students.abtech.edu','Chapman','Michael','Asheville','NC',28803,'USA'),
65      (DEFAULT,'madeup@students.abtech.edu','Donald','Ashley','Asheville','NC',28803,'USA');
66
67  INSERT INTO accountSecurity VALUES
68      ('michaeljchapman@students.abtech.edu',(SELECT accountNum FROM customers WHERE email='michaeljchapman@students.abtech.edu'),'indiana',12345),
69      ('madeup@students.abtech.edu',(SELECT accountNum FROM customers WHERE email='madeup@students.abtech.edu'),'southcarolina',67890);
70
71  INSERT INTO userProfiles VALUES
72  SELECT dietLogID,accountNum
73  FROM dietData
74  WHERE accountNum = 1;
75
76  INSERT INTO userProfiles VALUES
77  SELECT BSTestID, accountNum
78  FROM bloodspotData
79  WHERE accountNum = 1;
80
81  INSERT INTO userProfiles VALUES
82  SELECT BRKetoTestID, accountNum
83  FROM breathKetones
84  WHERE accountNum = 1;
```

The rest of the database setup code is listed in the GitHub site.