**Challenge 1**

**Nama Lengkap :** Yeni Rosandi

**No. Registrasi :** 01613821009

**Kelas :** B | Data Science

**Nutrition Fact for McDonald’s Menu**

1. **Studi Kasus**

“Ray Kroc wanted to build a restaurant system that would be famous for providing food of consistently high quality and uniform methods of preparation. He wanted to serve burgers, buns, fries and beverages that tasted just the same in Alaska as they did in Alabama. To achieve this, he chose a unique path: persuading both franchisees and suppliers to buy into his vision, working not for McDonald’s but for themselves, together with McDonald’s. Many of McDonald’s most famous menu items – like the Big Mac, Filet-O-Fish, and Egg McMuffin – were created by franchisees.”

Dalam kasus ini Ray Kroc sebagai stakeholder memiliki pertanyaan sebagai berikut:

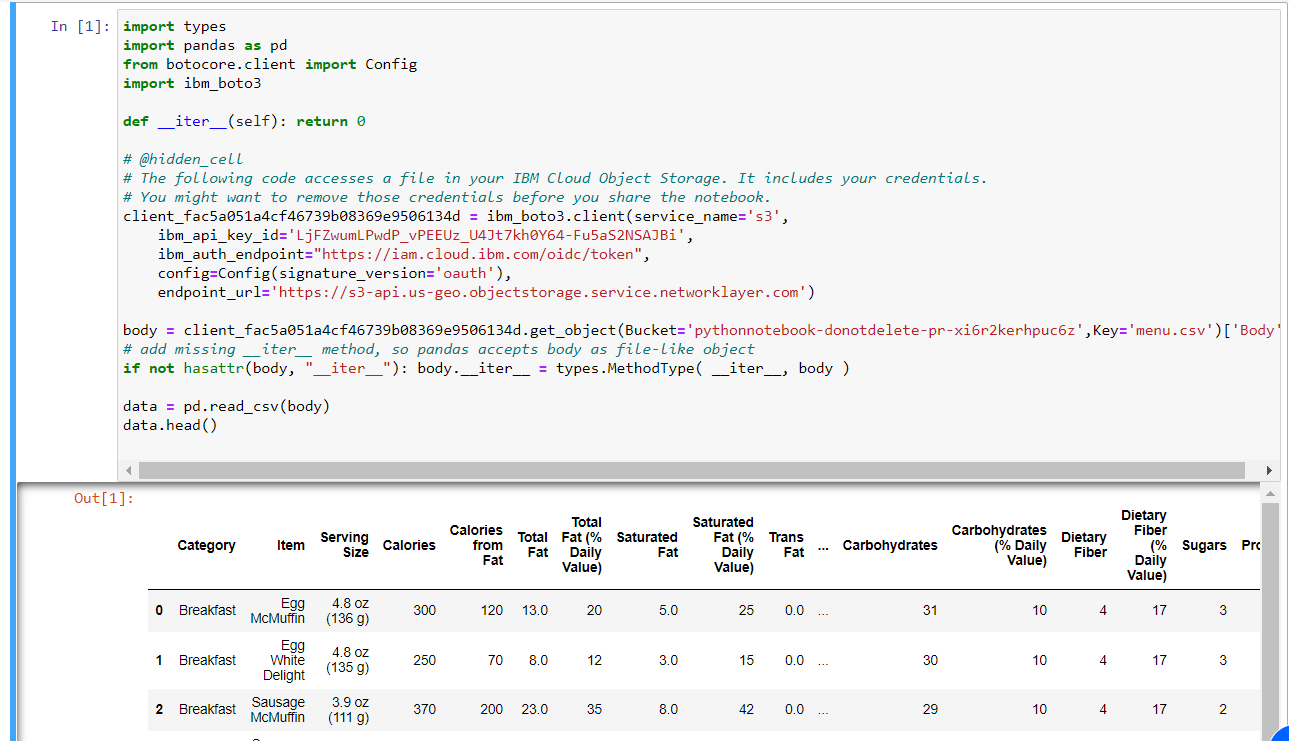
1. How many calories does the average McDonald's value meal contain?
2. How much do beverages, like soda or coffee, contribute to the overall caloric intake?
3. Does ordered grilled chicken instead of crispy increase a sandwich's nutritional value?
4. What about ordering egg whites instead of whole eggs?
5. What is the least number of items could you order from the menu to meet one day's nutritional requirements?
6. **Dataset**

Dataset didapatkan dari McDonald’s dengan judul “Nutrition Facts for McDonald's Menu” yang diperoleh melalui situs web Kaggle. Data tersebut memberikan data mengenai analisis nutrisi dari setiap items menu di McDonals’s Amerika Serikat, termasuk menu *breakfast, beef burgers, chicken and fish sandwiches, fries, salads, soda, coffee and tea, milkshakes,* dan *desserts*. Dataset ini terdiri dari 24 kolom dan 261 baris, yang berarti data tersebut memiliki item sebanyak 260. Kategori yang paling banyak sebesar 37% adalah *Coffe & Tea*, kemudian kedua adalah Breakfast dengan persentase sebesar 16%, kemudian kategori lainnya sebesar 47%.

Sumber Data: <https://www.kaggle.com/mcdonalds/nutrition-facts>

1. **Import Data**

Kode program yang digunakan adalah dalam bentuk python, dan saya menggunakan Python Notebook pada platform IBM Cloud. Pertama-tama hal yang dilakukan adalah import library yang dibutuhkan beserta data csv yang ingin digunakan, kode program dan output dapat dilihat pada gambar dibawah ini:



1. **Cek Kualitas Data**

Dibagian ini kualitas data dicek, dicari apakah ada yang null/kosong pada kolom/fitur, apakah ada angka yang tidak masuk akal dengan menggunakan kode program beserta outputnya yang dapat ditunjukan dibawah ini:

ln [2]: print(data.isnull().any())

Category False

Item False

Serving Size False

Calories False

Calories from Fat False

Total Fat False

Total Fat (% Daily Value) False

Saturated Fat False

Saturated Fat (% Daily Value) False

Trans Fat False

Cholesterol False

Cholesterol (% Daily Value) False

Sodium False

Sodium (% Daily Value) False

Carbohydrates False

Carbohydrates (% Daily Value) False

Dietary Fiber False

Dietary Fiber (% Daily Value) False

Sugars False

Protein False

Vitamin A (% Daily Value) False

Vitamin C (% Daily Value) False

Calcium (% Daily Value) False

Iron (% Daily Value) False

dtype: bool

Dapat dilihat nilai yang dikembalikan adalah *false*, sehingga dapat disimpulkan tidak ada nilai *null* pada dataset tersebut.

1. Analisis Data
2. Analisis 1

*How many calories does the average McDonald's value meal contain?*

Untuk menjawab pertanyaan tersebut, kita dapat menggunakan fungsi describe(). Describe berfungsi untuk melihat ringkasan statistik dataset tersebut. Sehingga kode program dan output yang dikeluarkan adalah:

ln [3]: print(data.describe())

Calories Calories from Fat Total Fat Total Fat (% Daily Value) \

count 260.000000 260.000000 260.000000 260.000000

mean 368.269231 127.096154 14.165385 21.815385

std 240.269886 127.875914 14.205998 21.885199

min 0.000000 0.000000 0.000000 0.000000

25% 210.000000 20.000000 2.375000 3.750000

50% 340.000000 100.000000 11.000000 17.000000

75% 500.000000 200.000000 22.250000 35.000000

max 1880.000000 1060.000000 118.000000 182.000000

Saturated Fat Saturated Fat (% Daily Value) Trans Fat Cholesterol \

count 260.000000 260.000000 260.000000 260.000000

mean 6.007692 29.965385 0.203846 54.942308

std 5.321873 26.639209 0.429133 87.269257

min 0.000000 0.000000 0.000000 0.000000

25% 1.000000 4.750000 0.000000 5.000000

50% 5.000000 24.000000 0.000000 35.000000

75% 10.000000 48.000000 0.000000 65.000000

max 20.000000 102.000000 2.500000 575.000000

Cholesterol (% Daily Value) Sodium ... Carbohydrates \

count 260.000000 260.000000 ... 260.000000

mean 18.392308 495.750000 ... 47.346154

std 29.091653 577.026323 ... 28.252232

min 0.000000 0.000000 ... 0.000000

25% 2.000000 107.500000 ... 30.000000

50% 11.000000 190.000000 ... 44.000000

75% 21.250000 865.000000 ... 60.000000

max 192.000000 3600.000000 ... 141.000000

Carbohydrates (% Daily Value) Dietary Fiber \

count 260.000000 260.000000

mean 15.780769 1.630769

std 9.419544 1.567717

min 0.000000 0.000000

25% 10.000000 0.000000

50% 15.000000 1.000000

75% 20.000000 3.000000

max 47.000000 7.000000

Dietary Fiber (% Daily Value) Sugars Protein \

count 260.000000 260.000000 260.000000

mean 6.530769 29.423077 13.338462

std 6.307057 28.679797 11.426146

min 0.000000 0.000000 0.000000

25% 0.000000 5.750000 4.000000

50% 5.000000 17.500000 12.000000

75% 10.000000 48.000000 19.000000

max 28.000000 128.000000 87.000000

Vitamin A (% Daily Value) Vitamin C (% Daily Value) \

count 260.000000 260.000000

mean 13.426923 8.534615

std 24.366381 26.345542

min 0.000000 0.000000

25% 2.000000 0.000000

50% 8.000000 0.000000

75% 15.000000 4.000000

max 170.000000 240.000000

Calcium (% Daily Value) Iron (% Daily Value)

count 260.000000 260.000000

mean 20.973077 7.734615

std 17.019953 8.723263

min 0.000000 0.000000

25% 6.000000 0.000000

50% 20.000000 4.000000

75% 30.000000 15.000000

max 70.000000 40.000000

Vitamin A (% Daily Value) Vitamin C (% Daily Value) \

count 260.000000 260.000000

mean 13.426923 8.534615

std 24.366381 26.345542

min 0.000000 0.000000

25% 2.000000 0.000000

50% 8.000000 0.000000

75% 15.000000 4.000000

max 170.000000 240.000000

Calcium (% Daily Value) Iron (% Daily Value)

count 260.000000 260.000000

mean 20.973077 7.734615

std 17.019953 8.723263

min 0.000000 0.000000

25% 6.000000 0.000000

50% 20.000000 4.000000

75% 30.000000 15.000000

max 70.000000 40.000000

Berdasarkan data yang didapatkan, terdapat data yang harus diperhatikan untuk menjawab pertanyaan tersebut, yaitu pada bagian:

Calories Calories from Fat Total Fat Total Fat (% Daily Value) \

count 260.000000 260.000000 260.000000 260.000000

mean 368.269231 127.096154 14.165385 21.815385

std 240.269886 127.875914 14.205998 21.885199

min 0.000000 0.000000 0.000000 0.000000

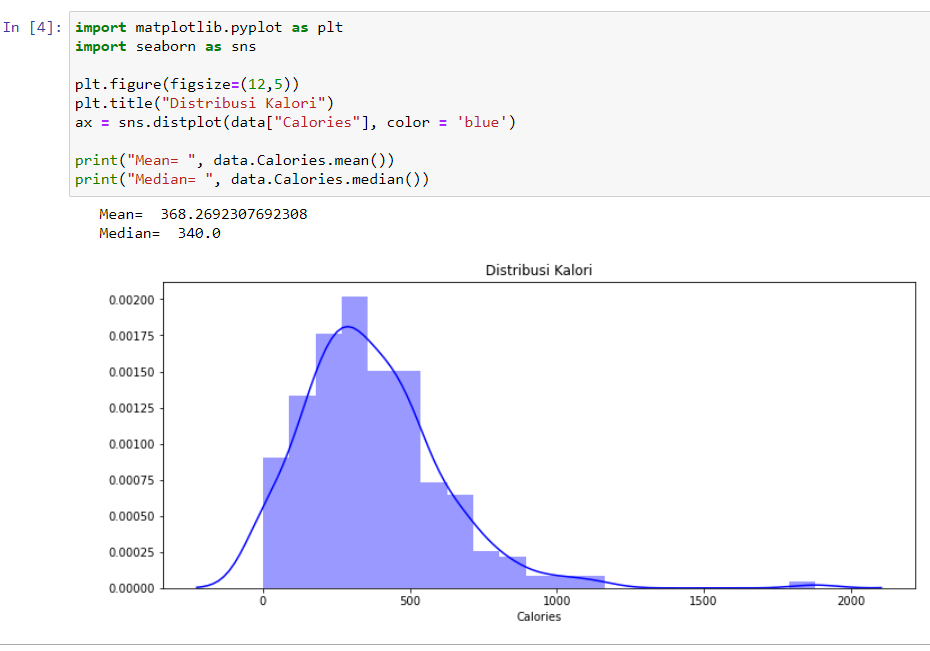
25% 210.000000 20.000000 2.375000 3.750000

50% 340.000000 100.000000 11.000000 17.000000

75% 500.000000 200.000000 22.250000 35.000000

max 1880.000000 1060.000000 118.000000 182.000000

Sehingga berapa banyak **rata-rata kalori** yang terkandung dalam menu McDonald adalah sebanyak **368.269231 kalori**, dan kalori dari lemak sebanyak 127.096154 dari lemak total sebesar 14.165385. Distribusi kalori dapat digambarkan pada grafik berikut ini:

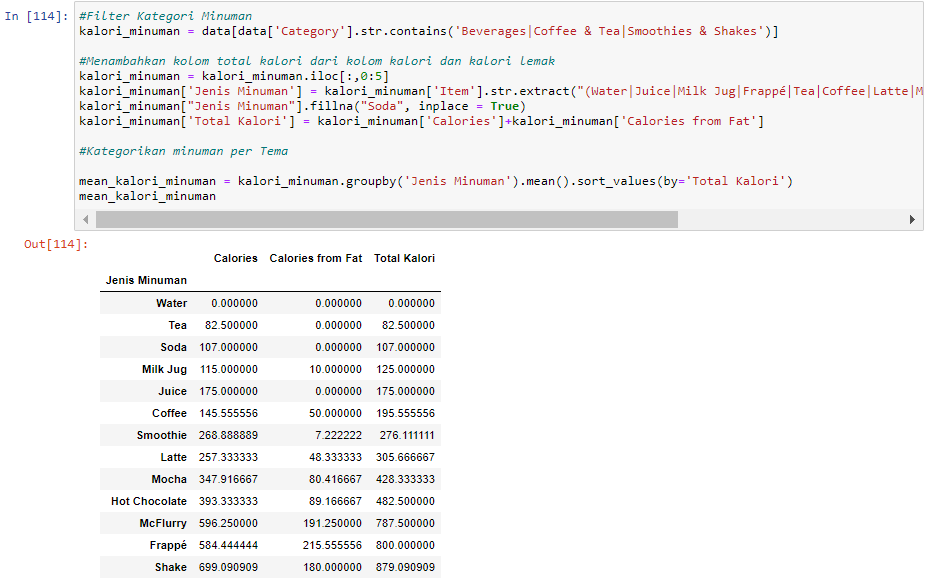


Grafik diatas diplot menggunakan library Matplotlib. Library Seaborn ini sendiri adalah library yang dibangun diatas Matplotlib sebagai tipe plot tambahan, agar grafik terlihat lebih cantik.

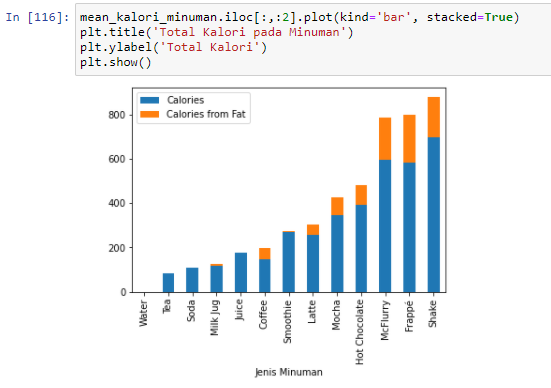
1. Analisis 2

*How much do beverages, like soda or coffee, contribute to the overall caloric intake?*

Untuk menjawab pertanyaan tersebut, kita terlebih dahulu buat data frame baru khusus untuk menampilkan data minuman, dari 13 jenis minuman yang diekstrak dari nama yang sama. Dengan menampilkan hanya kolom Category, Calories, Calories from Fat, dan menambahkan satu kolom dengan nama “Total Kalori” yang berisi nilai sum dari Calories dan Calories from Fat. Lalu dataframe tersebut di grup menjadi 3 kategori dan dicari nilai rata-ratanya untuk mendapatkan nilai keseluruhan. Kode program beserta outputnya dapat dilihat pada gambar berikut ini:



Output diatas adalah tabel rata-rata dalam bentuk tabel. Selanjutnya membuat visualisasi data agar data dapat lebih mudah dipahami. Visualisasi yang digunakan adalah barplot dengan kedua nilai dijadikan tumpang tindih yang dapat dilihat dibawah ini:

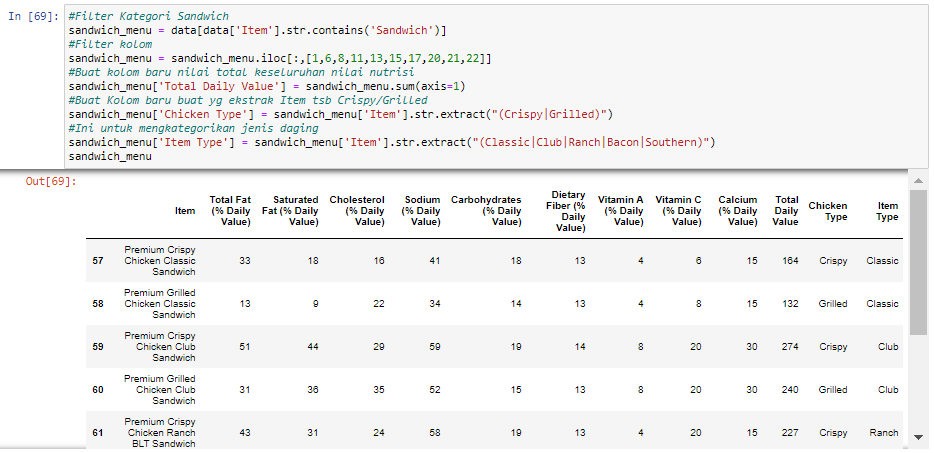


Visualisasi data selesai. Sehingga kita dapat melihat bahwa setiap minuman memiliki kontribusi kalori sebesar:

1. Air adalah minuman yang tidak berkalori
2. Teh menduduki peringkat kedua dan yang pertama untuk minuman berkalori rendah dengan total kalori sebesar 82
3. Dilanjutkan dengan soda yang menduduki peringkat ke 3 dengan kalori sebesar 107
4. Peringkat ke-4 Milk Jug dengan total kalori sebesar 115
5. Peringkat ke-5 Jus dengan total kalori sebesar 175
6. Peringkat ke-6 Kopi dengan total kalori sebesar 195.556
7. Peringkat ke-7 Smoothie dengan total kalori sebesar 276.11
8. Peringkat ke-8 Latte dengan total kalori sebesar 305.667
9. Peringkat ke-9 Mocha dengan total kalori sebesar 428.333
10. Peringkat ke-10 Hot Chocolate dengan total kalori sebesar 482.500
11. Peringkat ke-11 McFlurry dengan total kalori sebesar 787.500
12. Peringkat ke-12 Frappe dengan total kalori sebesar 800.000
13. Dan total kalori paling tinggi pertama pada peringkat ke-13 adalah Shake dengan total kalori sebesar 979.091
14. Analisis 3

Does ordered grilled chicken instead of crispy increase a sandwich's nutritional value?

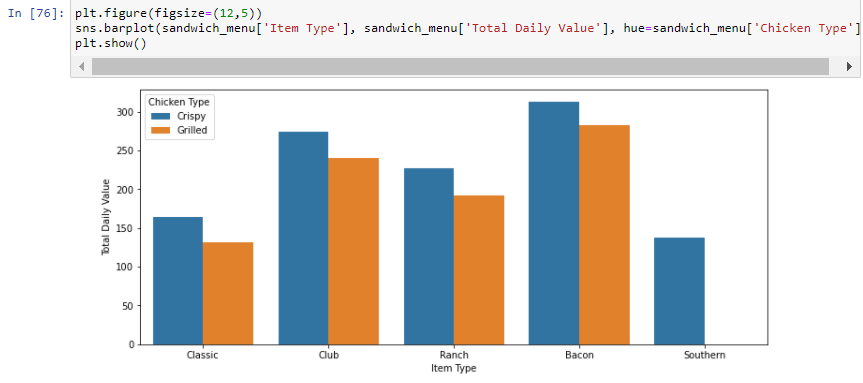
Untuk menjawab pertanyaan tersebut, terlebih dahulu buat data frame baru khusus untuk Sandwich, karena pada pertanyaan ini hanya membutuhkan data yang dikategorikan Sandwich (lebih spesifik). Kolom yang digunakan juga diseleksi sehingga kolom hanya terdiri dari nama item, tipe ayam, tipe daging, dan nutrisi yang terkandung di dalam items seperti lemak, kolesterol, sodium, karbohidrat, serat, vitamin A, vitamin C, kalsium, zat besi, dan total nilai harian. Sehingga dataframe yang dibuat adalah sebagai berikut:

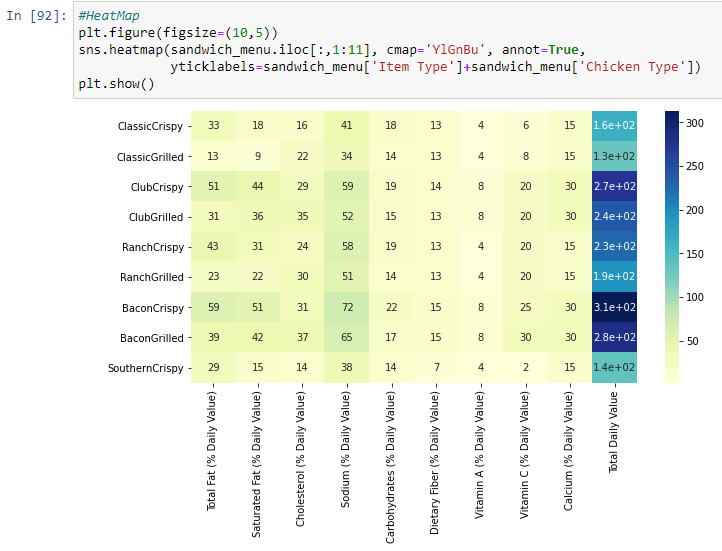


Setelah dataframe khusus sandwich dibuat maka dapat dimulai untuk plotting grafik untuk menganalisa atau mengetahui nilai total nutrisi harian per tipe ayam (Chicken Type).



Data diatas terlihat Ayam Crispy lebih banyak nilai total nutrisi harian dibandingkan Ayam Bakar. Untuk melihat lebih spesifik lagi dapat dihat melalui grafik bar lagi namun lebih spesifik berdasarkan tipe item. Hasil yang didapatkan adalah:

Berdasarkan gambar diatas pada ke-5 tipe item terlihat jenis ayam Crispy lebih banyak nutrisinya dibandingkan ayam bakar. Kemudian mari kita pecah lebih rinci lagi dengan menampilkan kanduangan tiap nutrisi yang ada melalui heatmap plot. Hasil yang didapatkan adalah:



Pada plot HeatMap ini dapat dilihat alasan mengapa Ayam Crispy memiliki total nilai harian yang lebih tinggi. Hal tersebut dikarenakan terlihat perbedaan yang sangat signifikan pada nilai Total Fat dimana antara Crispy dan Grilled memiliki selisih yang paling besar dibandingkan nutrisi lainnya yaitu sebesar 20%.

Sehingga pertanyaan *“Does ordered grilled chicken instead of chicken crispy increase a sandwich's nutritional value?”* adalah **Tidak**. Dikarenakan ayam bakar memiliki beberapa nilai nutrisi lebih rendah daripada ayam krispy.

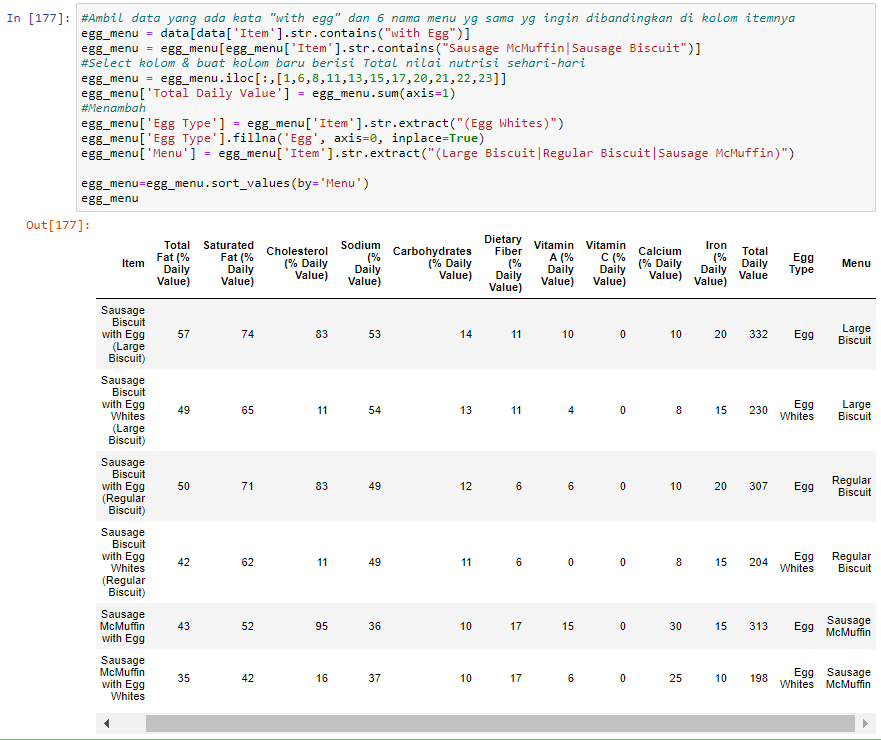
1. Analisis 4

What about ordering egg whites instead of whole eggs?

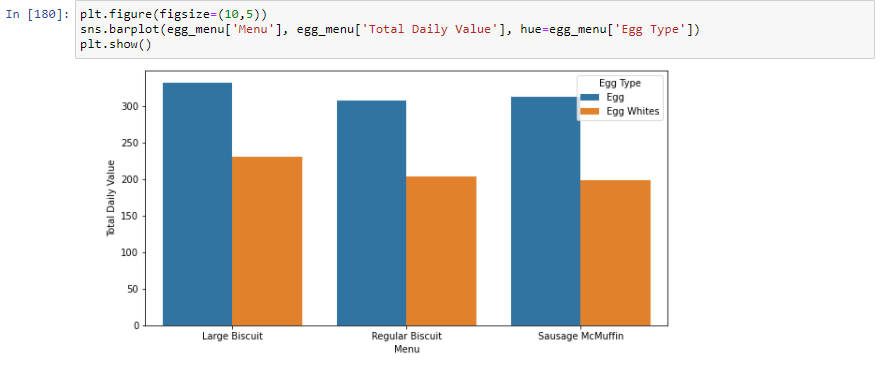
Untuk menjawab pertanyaan tersebut, terlebih dahulu buat data frame baru khusus untuk menu yang memiliki kalimat “with egg white” atau “with egg”, pada kasus ini kita akan membandingkan nutrisi pada menu yang sama dengan tambahan telur yang berbeda (egg white atau egg). Pertama-tama mari kita lihat pada data csv tersebut ada berapa data yang memiliki kalimat yang diharapkan menggunakan fitur find:

|  |  |
| --- | --- |
| **No.** | **Menu with Egg/Egg White** |
| 1. | Bacon, Egg & Cheese Bagel with Egg Whites |
| 2. | Bacon, Egg & Cheese Biscuit with Egg Whites (Large Biscuit) |
| 3. | Bacon, Egg & Cheese Biscuit with Egg Whites (Regular Biscuit) |
| 4. | Bacon, Egg & Cheese McGriddles with Egg Whites |
| 5. | Big Breakfast with Egg Whites (Large Biscuit) |
| 6. | Big Breakfast with Egg Whites (Regular Biscuit) |
| 7. | Sausage Biscuit with Egg (Large Biscuit) |
| 8. | Sausage Biscuit with Egg (Regular Biscuit) |
| 9. | Sausage Biscuit with Egg Whites (Large Biscuit) |
| 10. | Sausage Biscuit with Egg Whites (Regular Biscuit) |
| 11. | Sausage McMuffin with Egg |
| 12. | Sausage McMuffin with Egg Whites |
| 13. | Sausage, Egg & Cheese McGriddles with Egg Whites |

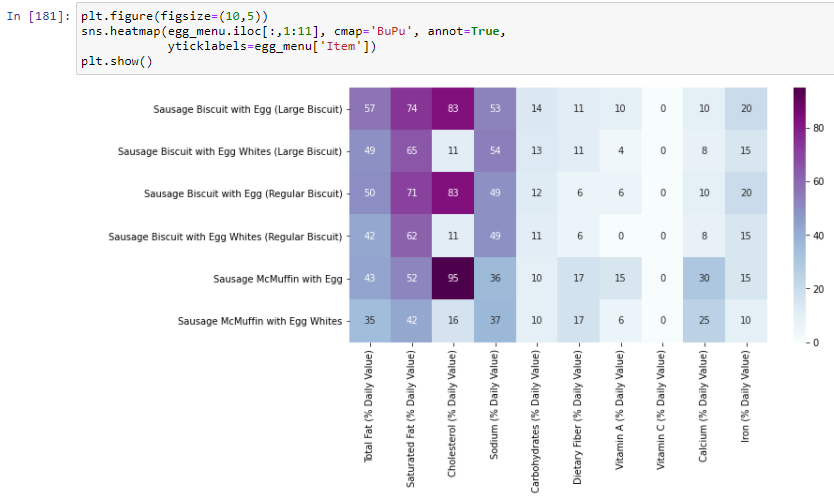
Data diatas menampilkan 13 data namun hanya 6 data yang memiliki menu yang dapat dibandingkan. Sehingga data frame yang dibuat adalah sebagai berikut:



Setelah dataframe dibuat maka tahap selanjutnya adalah visualisasi data. Kita akan membandingkan data menu dengan tipe egg dan egg white menggunakan barplot:



Seperti grafik diatas dapat dilihat bahwa total nutrisi harian lebih banyak pada makanan yang diberikan telur utuh, bukan hanya telur putih. Untuk melihat alasan lebih rinci, kita dapat lihat pada HeatMap plot sebagai berikut:



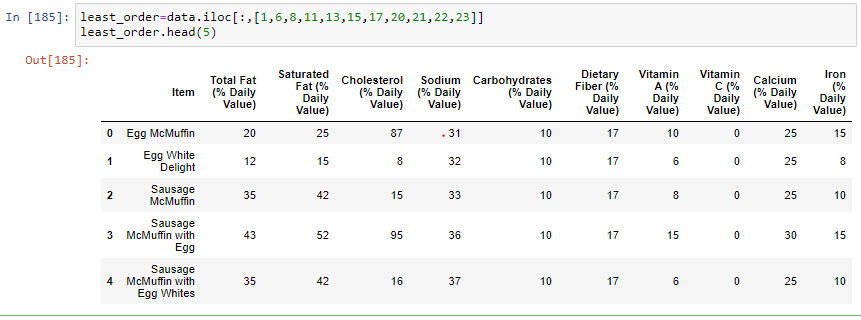
Pada plot HeatMap ini dapat dilihat alasan mengapa Makanan dengan telur utuh memiliki total nilai nutrisi harian yang lebih tinggi. Hal tersebut dikarenakan terlihat perbedaan yang sangat signifikan pada nilai kolesterol dimana antara Egg dan Egg White memiliki selisih yang paling besar dibandingkan nutrisi lainnya yaitu sebesar 80%. Analisis ini membuktikan bahwa telur kuning yang ada pada telur mengandung kolesterol yang banyak.

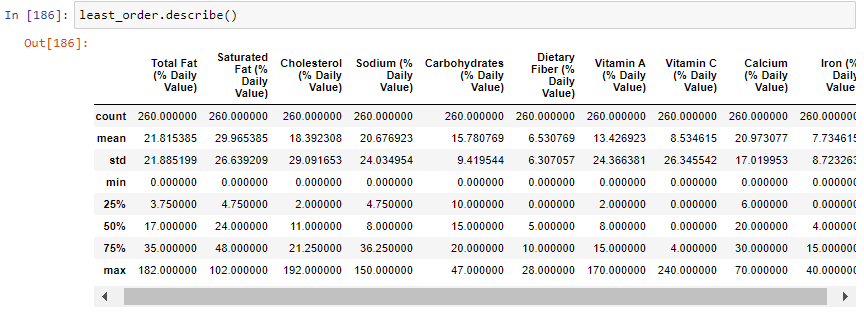
Sehingga pertanyaan *“What about ordering egg whites instead of whole eggs?”* adalah **Tidak**. Dikarenakan telur putih memiliki nilai nutrisi yang lebih rendah dibandingkan telur yang utuh.

1. Analisis 5

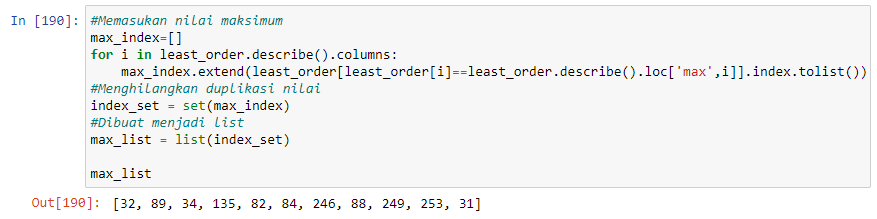
What is the least number of items could you order from the menu to meet one day's nutritional requirements?

Untuk menjawab pertanyaan tersebut, terlebih dahulu buat data frame baru dengan kolom yang digunakan masih sama dengan yang sebelumnya dikarenakan pada analisis ini bertujuan untuk mencari kombinasi item yang memungkinkan untuk memenuhi kebutuhan nutrisi dalam sehari. Dataframe dan ringkasan statistik dapat dilihat sebagai berikut:

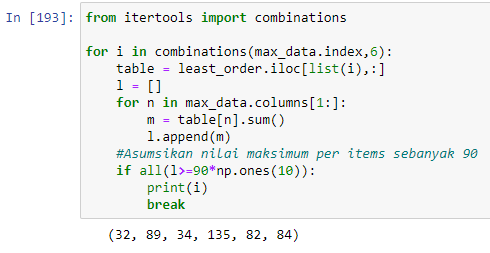


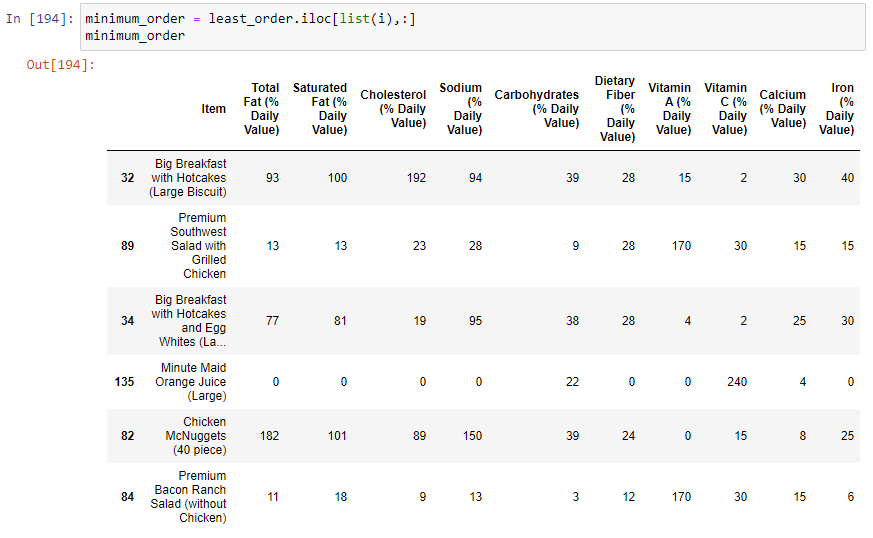


Kemudian menggunakan nilai max pada describe sebagai acuan dalam memilih items apa yang dapat memenuhi nutrisi keseharian kita. Hasil yang didapatkan sebagai berikut:



Data tersebut masih belum sampai disitu, masih ada tahap filter Kembali mencari item mana saja yang cocok untuk memenuhi nutrisi harian. Dengan diasumsikan untuk memenuhi nutrisi harian membutuhkan minimum 90 %Daily Value pada setiap kolom. Sehingga hasil yang didapat sebagai berikut:





Berdasarkan hasil analisis kombinasi makanan yang memungkinkan dapat memenuhi nutrisi harian ada 6 kombinasi, diantaranya adalah:

1. Big Breakfast with Hotcakes (Large Biscuit)
2. Premium Southwest Salad with Grilled Chicken
3. Big Breakfast with Hotcakes and Egg White (Large Biscuit)
4. Minute Maid Orange Juice (Large)
5. Chicken McNuggets (40 piece)
6. Premium Bacon Ranch Salad (Without Chicken)

Ke-6 kombinasi makanan diatas adalah saran berdasarkan nilai nutrisi yang terkandung, namun belum tentu makanan yang ada disebutkan diatas adalah kombinasi makanan yang sehat. Kita harus bisa cermat dan berhati-hati dengan yang disebut “overnutrition” dari sebagaian nutrisi seperti sodium, lemak, kolesterol, dan nutrisi lainnya yang masing-masing memiliki batasan. Berikut adalah manfaat yang dapat dirasakan jika mengkonsumsi ke-6 kombinasi tersebut:

