# Pakiety Statystyczne

Bartłomiej Gintowt

2022-12-10

## Wstęp

W raporcie posłużymy się danymi dotyczącymi wartości odżywczych dla pozycji jedzeniowych największych sieci fastfoodowych na terenie Stanów Zjednoczonych.Dane pochodzą ze strony zawierającej różne zbiory danych dla wielu kategorii https://vincentarelbundock.github.io/Rdatasets/datasets.html. Nasze dane zawierają zmienną kategoryczną będącą nazwą restauracji z przypisanymi do nich nazwami posiłków oraz odpowiednio rozpisane wartości odżywcze każdego posiłku będące zmiennymi ciągłymi. Rozważanymi wartościami odżywczymi będą kalorie, kalorie pochodzące z samych tłuszczy, odpowiednio w gramach: liczba tłuszczy, liczba tłuszczy nasyconych, liczba tłuszczy trans, cholesterol, sód, licza węglowodanów, błonnik, cukier, białko, oraz odpowiednio w mikrogramach/10 witaminę A, w miligramach witaminę C, w miligamach/10 wapń. Każda pozycja jedzeniowa będzie przypisana do odpowiedniego indeksu. Ostatnia zmienna jest kategoryczną zmienną określającą czy dany posiłek jest sałątką czy też nie. Możliwe błędy występujące w danych są spowodowane chociażby tym, iż dana restauracja nie przeprowadziła, bądz nie upubliczniła, badań na temat niektórych wartości odżywczych czy witamin zawartych w posiłkach z menu.

W eksperymencie będziemy chcieli sobie odpowiedzieć na pytanie, czy wybrać restaurację McDonald's aby zjeść najzdrowiej jeśli chcemy zjeść obiad składający się z sałatki oraz pozycji niesałatkowej z menu. W przypadku braku sałatek w danej restauracji zamówimy pozycje niesałatkowe o równowartości kalorycznej obiadu uwzględniającym sałatkę. Słowo najzdrowiej zdefiniujemy jako duża ilość białka w stosunku do jak najmniejszej ilości tłuszczu w naszej porcji.

## Wczytanie danych

Pobrane dane ze strony w formacie csv wczytamy przy pomocy funkcji read.csv. Ustawimy nazwy kolumn, ich klasy oraz zdefiniujemy możliwe błędy.

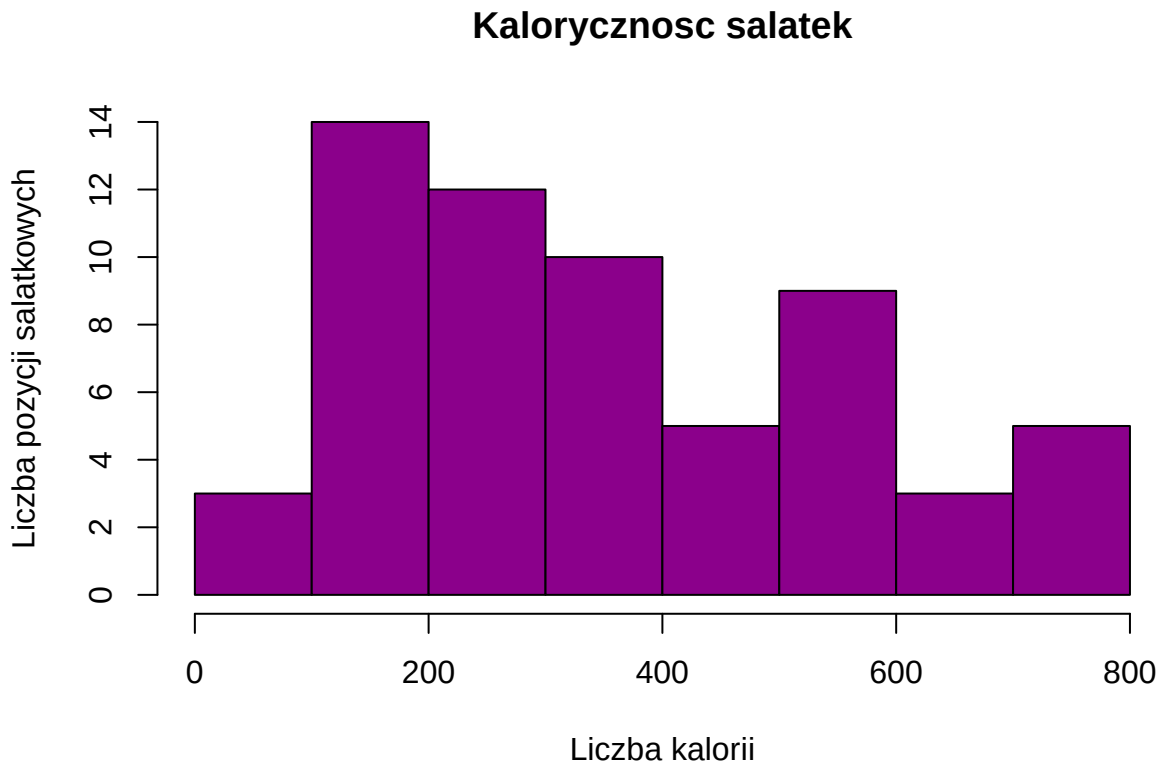
Przy pomocy pakietu "expss" nadamy każdej kolumnie odpowiednią etykietę.

# Analiza danych

## Kaloryczność

### Pozycje sałatkowe

Przeanalizujemy wykres słupkowy zawierający kaloryczność wszystkich pozycji sałatkowych z naszej bazy danych.

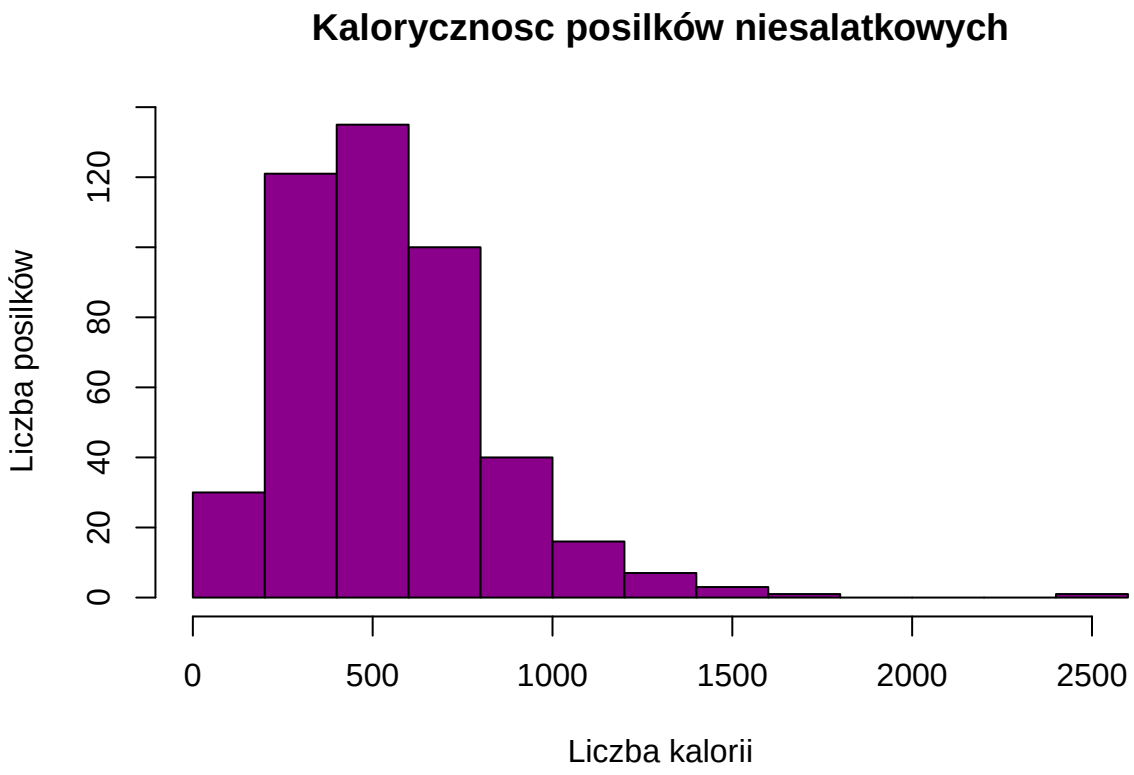



Rozpatrując powyższy histogram zauważalna jest tendencja malejącej liczby pozycji dla rosnącej liczby kalorii. Wyznaczymy średnią wartość kalorii dla sałatki w rozważanym histogramie.

Wykorzystując funkcję mean otrzymujemy średnią wartość kaloryczną pozycji sałatkowwych równą 352.459 kalorie.

**Pozycje niesałatkowe**

Teraz sprawdzimy histogram dla pozycji niesałatkowych.




Rozpatrując powyższy histogram zauważalnie najwięcej posiłków niesałatkowych jest o kaloryczności w przedziale $(250, 750)$. Ponownie wyznaczymy średnią dla rozważanego histogramu.

Otrzymujemy średnią wartość kaloryczną posiłku niesałatkowego równą 556.5198 kalorie.

Zdecydowalnie zauważalna jest przewaga kaloryczności dań niesałatkowych. Zawierają one średnio o 204.0608 więcej kalorii.

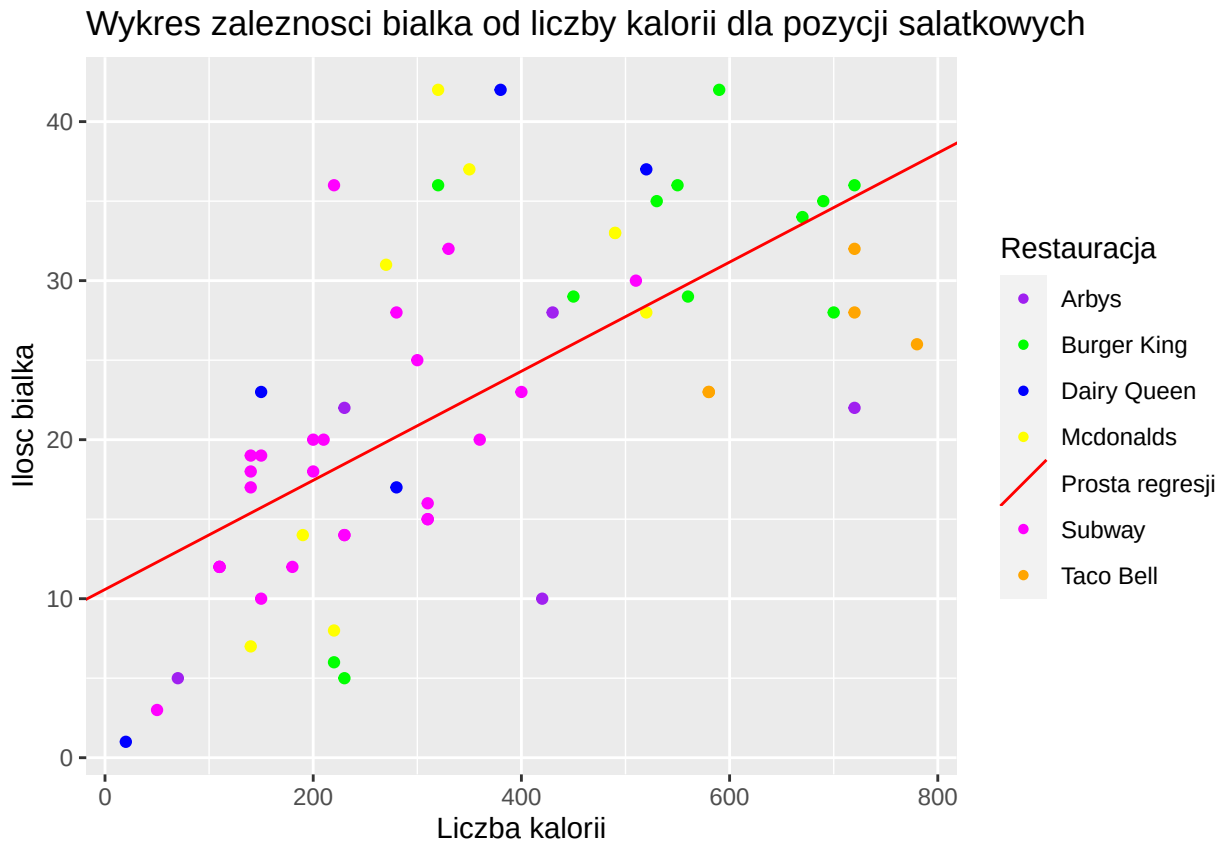
## Zawartość białka w zależności od kalorii

Będziemy chcieli na wykresach rozproszenia opisać zależność białka od ilości spożywanych kalorii. Do każdego wykresu rozproszenia będziemy chcieli dopasować prostą regresji metodą najmniejszych kwadratów dla wszystkich restauracji oraz znaleźć restaurację, która posiadać będzie najwiekszy współczynnik ilości makroskładnika do ilości kalorii.

### Pozycje sałatkowe

Wykres gramów białka w zależności od ilości spożywanych kalorii w sałatce dla wszystkich restauracji zwierających pozycje sałatkowe oraz dopasowana do nich krzywa wyznaczona metodą najmniejszych kwadratów.

Wzór na prostą wyznaczoną metodą najmniejszych kwadratów: $Y = a \cdot X + b$.




Wyznaczone współczynniki prostej: $a = 0.03430611, b = 10.58063$. W średniej porcji sałatki liczącej 352.459 kalorii otrzymujemy średnio 22.67213 gramów białka.

Wyznaczymy prostą regresji dla każdej restauracji z osobna i sprawdzimy, w której liczba białka dla średniej liczby kalorii jest największa.

Dla Arbys liczba białka wynosi 16.95203 gramów.

Dla Burger King liczba białka wynosi 21.33441 gramów.

Dla Dairy Queen liczba białka wynosi 29.99949 gramów.

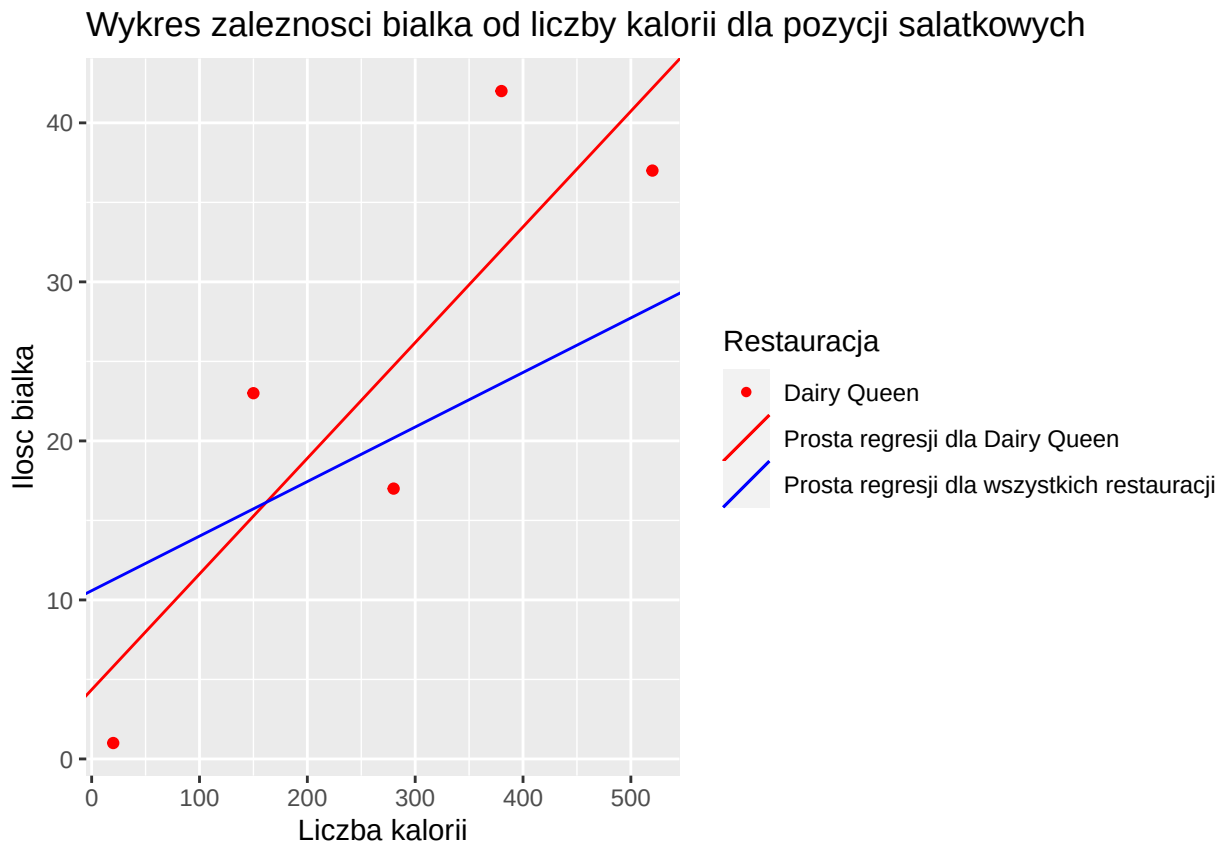
Dla Mcdonalds liczba białka wynosi 27.11618 gramów.

Dla Subway liczba białka wynosi 23.66432 gramów.

Dla Taco Bell liczba białka wynosi 17.3346 gramów.

Otrzymujemy, iż najlepszy stosunek białka w średniej procji sałatkowej posiadają sałatki restauracji Dairy Queen, który wynosi 29.99949 gramów białka na 352.459 spożytych kalorii.

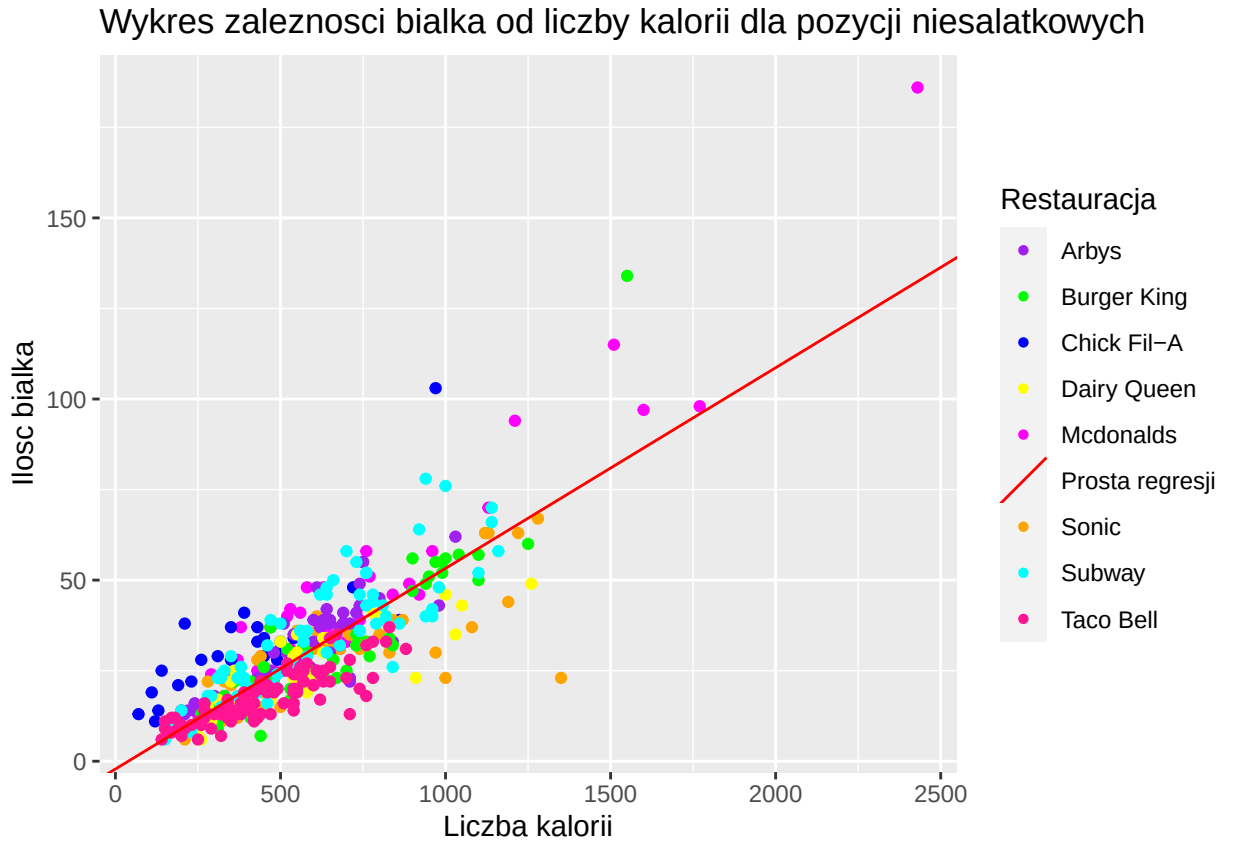
Wykres rozproszenia dla sałatek w restauracji Dairy Queen oraz porównanie prostej regresji dla wszystkich restauracji i dopasowanej dla Dairy Queen.




Zauważyć możemy, iż proste się przecinają na wysokości około 160 kalorii i wówczas prosta regresji dla restauracji Dairy Queen posiada wyższy współczynnik ilości białka. Dla rozpatrywanej przez nas wartości średnich kalorii w sałatce będzie to o aż 7.32736 gramów białka więcej. Otrzymane wnioski są jednak bardzo mało precyzyjne a model jest niezwykle uproszczony zważywszy na ilość danych z restauracji Dairy Queen, których jest zaledwie 5.

**Pozycje niesałatkowe**

Wykres gramów białka w zależności od ilości spożywanych kalorii w pozycji niesałatwkoej dla wszystkich restauracji oraz dopasowana do nich krzywa wyznaczona metodą najmniejszych kwadratów.




Wyznaczone współczynniki prostej: $a = 0.05539463, b = -2.153257$ W średniej porcji dla pozycji niesałatkowej liczącej 554.8899 kalorie otrzymujemy średnio 28.67495 gramów białka.

Wyznaczymy prostą regresji dla każdej restauracji z osobna i sprawdzimy, w której liczba białka dla średniej liczby kalorii jest największa.

Dla Arbys liczba białka wynosi 30.84705 gramów.

Dla Burger King liczba białka wynosi 26.36652 gramów.

Dla Chick Fil-A liczba białka wynosi 42.20813 gramów.

Dla Dairy Queen liczba białka wynosi 24.97774 gramów.

Dla Mcdonalds liczba białka wynosi 33.00549 gramów.

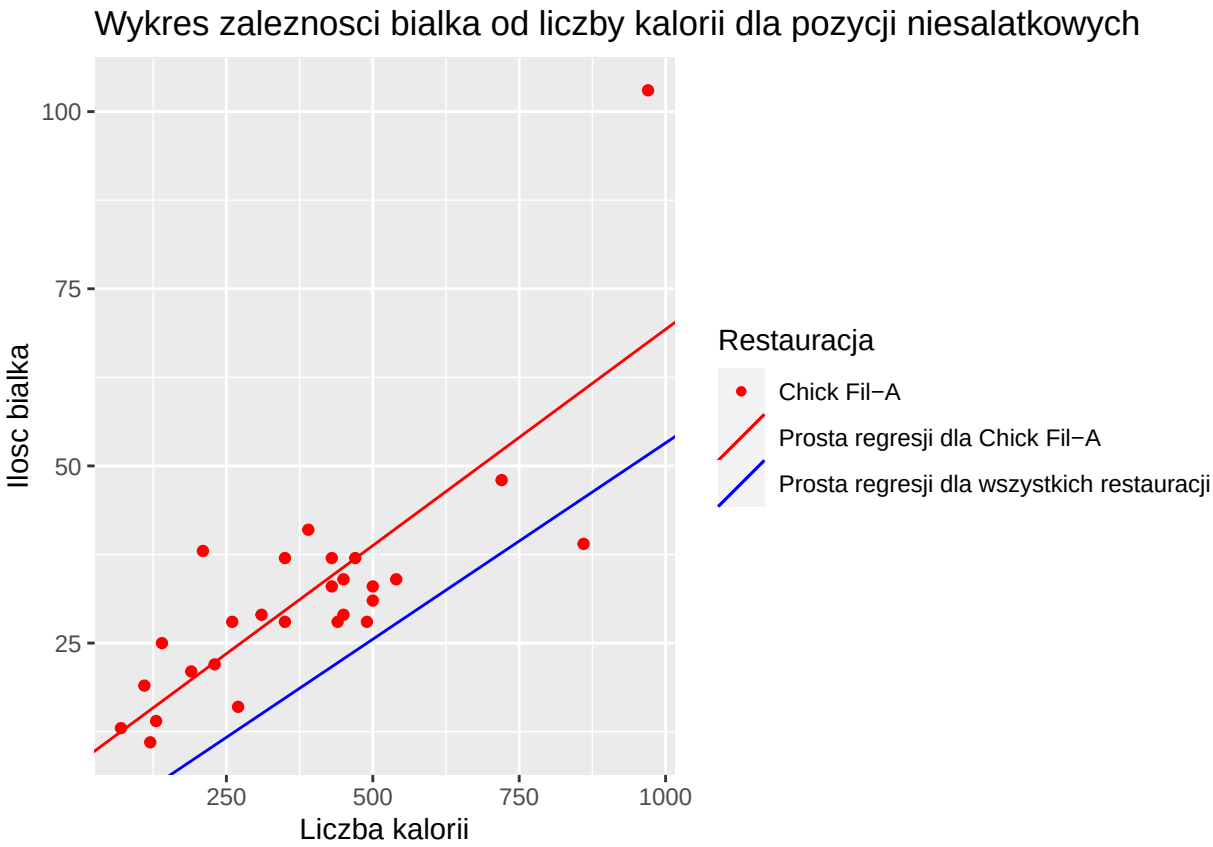
Dla Sonic liczba białka wynosi 25.6869 gramów.

Dla Subway liczba białka wynosi 32.16767 gramów.

Dla Taco Bell liczba białka wynosi 21.09848 gramów.

Otrzymujemy, iż najlepszy stosunek białka w średniej procji niesałatkowej posiadają pozycje restauracji Chick Fil-A, który wynosi 42.20813 gramów białka na 554.8899 spożytych kalorii.

Wykres rozproszenia dla pozycji niesałatkowych w restauracji Chick Fil-A oraz porównanie prostej regresji dla wszystkich restauracji i dopasowanej dla Chick Fil-A.

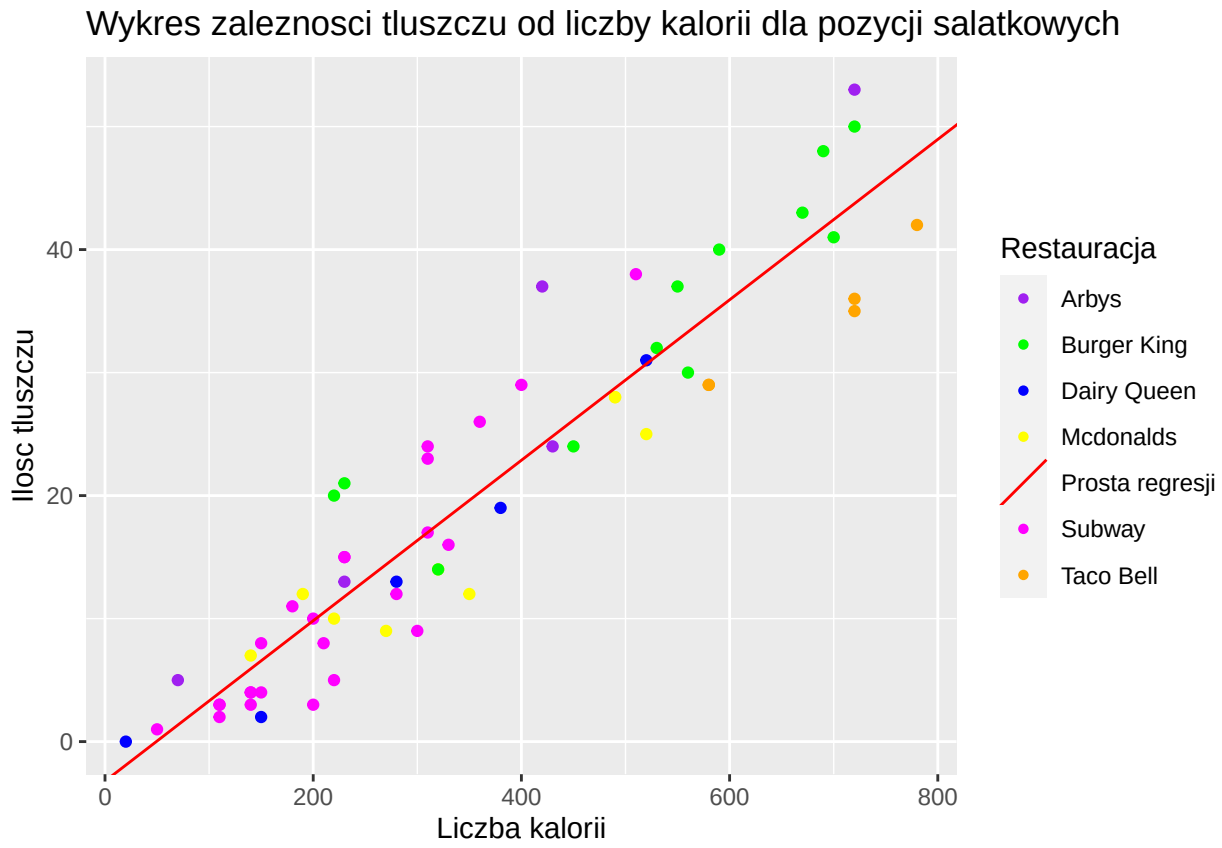



Na powyższym wykresie zauważyć możemy, iż liczba białka w zależności od kalorii dla restauracji Chick Fil-A jest zdecydowanie wieksza niż dla wszystkich restauracji. Dla średniej porcji pozycji niesałatkowej wynoszącej 554.8899 kalorie będzie to aż o 13.53318 gramóW białka więcej. Tym razem posiadamy jednak znacząco większą ilość danych dla rozpatrywanej restauracji przez co zwiększona precyzja modelu umożliwa nam wyciągniecie bardziej trafnych wniosków.

## Ilość tłuszczu w zależności od kalorii

Naszym zadaniem będzie wyznaczenie ile gramów tłuszczu znajdować się bedzie w średniej porcji sałatki i pozycji niesałatkowej. Podobnie jak w przypadku białka wykorzystamy model regresji wyznaczony przy pomocy metody najmniejszych kwadratów.

**Sałatki**




Wyznaczone współczynniki prostej: $a = 0.065214, b = -3.214771$. W średniej porcji sałatki liczącej 352.459 kalorii otrzymujemy średnio 19.77049 gramów tłuszczu.

Wyznaczymy teraz ilość gramów tłuszczu w średniej porcji dla każdej restauracji z osobna.

Dla Arbys liczba tłuszczu wynosi 24.77239 gramów.

Dla Burger King liczba tłuszczu wynosi 23.25195 gramów.

Dla Dairy Queen liczba tłuszczu wynosi 18.29239 gramów.

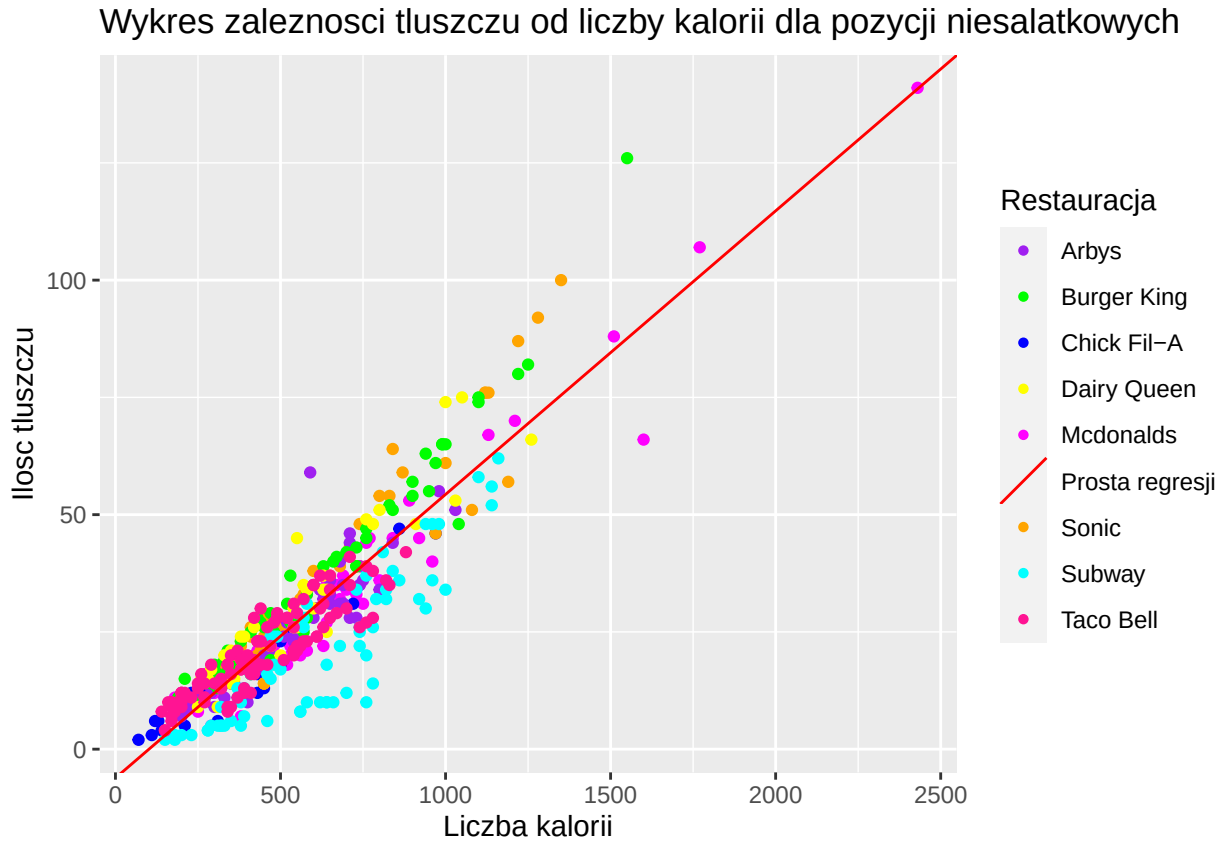
Dla Mcdonalds liczba tłuszczu wynosi 17.23484 gramów.

Dla Subway liczba tłuszczu wynosi 22.29565 gramów.

Dla Taco Bell liczba tłuszczu wynosi 15.40493 gramów.

Zauważamy, iż najwięcej tłuszczu w pozycjach sałatkowych posiada restauracja Arbys, która w średniej porcji 352.459 kalorii zawiera 24.77239 gramów białka.

**Pozycje niesałatkowe**




Wyznaczone współczynniki prostej: $a = 0.06045329, b = -6.136846$. W średniej porcji sałatki liczącej 554.8899 kalorii otrzymujemy średnio 27.50661 gramów tłuszczu.

Dla Arbys liczba tłuszczu wynosi 27.48732 gramów.

Dla Burger King liczba tłuszczu wynosi 32.44017 gramów.

Dla Chick Fil-A liczba tłuszczu wynosi 24.66056 gramów.

Dla Dairy Queen liczba tłuszczu wynosi 31.15785 gramów.

Dla Mcdonalds liczba tłuszczu wynosi 26.40137 gramów.

Dla Sonic liczba tłuszczu wynosi 31.23681 gramów.

Dla Subway liczba tłuszczu wynosi 18.61405 gramów.

Dla Taco Bell liczba tłuszczu wynosi 25.87321 gramów.

Zatem najwiekszą ilość gramów w średniej porcji posiadają pozycje restauracji Sonic.

## Co najlepiej zjeść

Będziemy teraz chcieli znaleźć restaurację, w której współczynnik białka do ilości tłuszczu w średniej porcji będzie największy. Naszym celem jest zjedzenie obiadu składającego się sałatki oraz pozycji niesałatkowej. Uwzględniając, iż nie wszystkie restauracje posiadają sałatki w swojej ofercie, wybierzemy w takich równowartość kaloryczną średniej sałatki i niesałatki ale w postaci posiłków niesałatkowych.

Wyznaczamy współczynnik ilości białka do ilości tłuszczu dla średnich wartości, korzystając z poprzednich wiliczeń ilości makroskładnika na ilość kalorii.

Dla Arbys współczynnik wynosi 0.914645.

Dla Burger King współczynnik wynosi 0.8565113.

Dla Chick Fil-A współczynnik wynosi 1.834459.

Dla Dairy Queen współczynnik wynosi 1.111769.

Dla Mcdonalds współczynnik wynosi 1.377793.

Dla Sonic współczynnik wynosi 0.9068336.

Dla Subway współczynnik wynosi 1.364762.

Dla Taco Bell współczynnik wynosi 0.9310759.

Otrzymujemy, że najwyższy współczynnik białka do ilości tłuszczu w naszym obiedzie posiada restauracja Chick Fil-A wynoszący 1.834459.

# Wnioski

Przeprowadzając analizę dla danych z amerykańskich restauracji otrzymaliśmy, że średnia wartość kaloryczna sałatek wynosi 352.459, natomiast pozycji niesałatkowych 556.5198. Po przeanalizowaniu wykresów zależności białka od liczby kalorii oraz tłuszczu od liczby kalorii, postanowiliśmy wyznaczyć parametry prostej regesji liniowej za pomocą metody najmniejszych kwadratów. Dzięki wyliczonym parametrom obliczyliśmy liczbę białka i tłuszczu spożytego w średniej porcji dla dań sałatkowych i niesałatkowych. Następnie wyznaczyliśmy współczynnik liczby białka od liczby tłuszczu dla średnich porcji i wybraliśmy restaurację o najwyższej wartości współczynnika. Powołując się na powyższą analizę danych możemy wywnioskować, iż postawione przez nas pytanie badawcze jest fałszywe. Mcdonald‘s nie jest restauracją, do której chcemy się udać aby zjeść najzdrwoszy obiad, w naszym tłumaczeniu jak największą ilość białka w stosunku do jak najmniejszej liczby tłuszczu w porcji. Restauracją tą jest Chick Fil-A, której współczynnik ilości białka do tłuszczu wynosi 1.834459.