Akademia Ekonomiczno-Humanistyczna w Warszawie

SPRAWOZDANIE

INTELIGENTNA ANALIZA DANYCH

**LAB5**

**REGRESJA LINIOWA**

09.01.2022

JOANNA PRAJZENDANC

36358

MIŁOSZ SAKOWSKI

36381

**Spis treści**

[1. Cel i przebieg ćwiczenia 3](#_Toc26351)

[2. Definicje i założenia 3](#_Toc26147)

[2.1. Wyjaśnienie pojęć 3](#_Toc1342)

[3. Metryki oceny regresji liniowej 3](#_Toc22073)

[3.1. Omówienie kodu 3](#_Toc2834)

[i. Najprostszy, przykładowy zestaw danych 3](#_Toc7450)

[ii. Regresja liniowa na przykładowym zestawie danych 4](#_Toc7543)

[3.2. Zadanie #1 5](#_Toc24695)

[i. Pytanie 5](#_Toc21731)

[ii. Odpowiedź 5](#_Toc25163)

[3.3. Zadanie #2 5](#_Toc20797)

[i. Polecenie 5](#_Toc31061)

[ii. Rozwiązanie 6](#_Toc13048)

[3.4. Zadanie #3 6](#_Toc32317)

[i. Polecenie 6](#_Toc18960)

[ii. Rozwiązanie 6](#_Toc23332)

[3.5. Zadanie #4 7](#_Toc4591)

[i. Polecenie 7](#_Toc6253)

[ii. Rozwiązanie 7](#_Toc2437)

[3.6. Zadanie #5 8](#_Toc22777)

[i. Polecenie 8](#_Toc12568)

[ii. Rozwiązanie 8](#_Toc2360)

[3.7. Zadanie #6 9](#_Toc918)

[i. Polecenie 9](#_Toc24497)

[ii. Rozwiązanie 9](#_Toc12142)

[3.8. Zadanie #7 10](#_Toc762)

[i. Polecenie 10](#_Toc13116)

[ii. Rozwiązanie 10](#_Toc2099)

# Cel i przebieg ćwiczenia

Celem ćwiczenia było zapoznanie się z regresją liniową oraz sposobami jej zastosowania.

# Definicje i założenia

## Wyjaśnienie pojęć

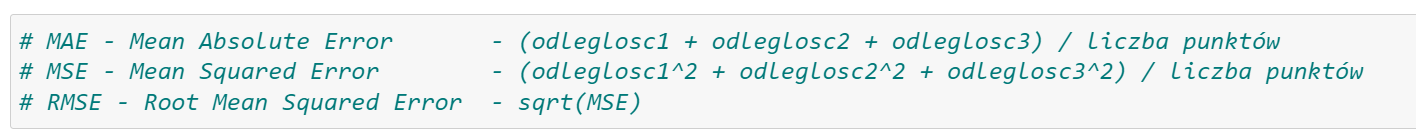
W sprawozdaniu pojawiają się następujące pojęcia:

* regresja liniowa -
* współczynnik regresji liniowej -
* metryki oceny regresji liniowej -
* metryka MAE - średnie oddalenie się przewidywanego wyniku od prawdziwej danej, które pozwala oszacować uporządkowanie zbioru, liczone według wzoru:
* metryka MSE - uśrednione kwadraty oddalenia się przewidywanego wyniku od prawdziwej danej, liczone według wzoru:
* metryka RMSE - pierwiastek kwadratowy z metryki MSE, bardziej czuła metryka pozwalająca wykryć znacząco odstające dane w zbiorze danych

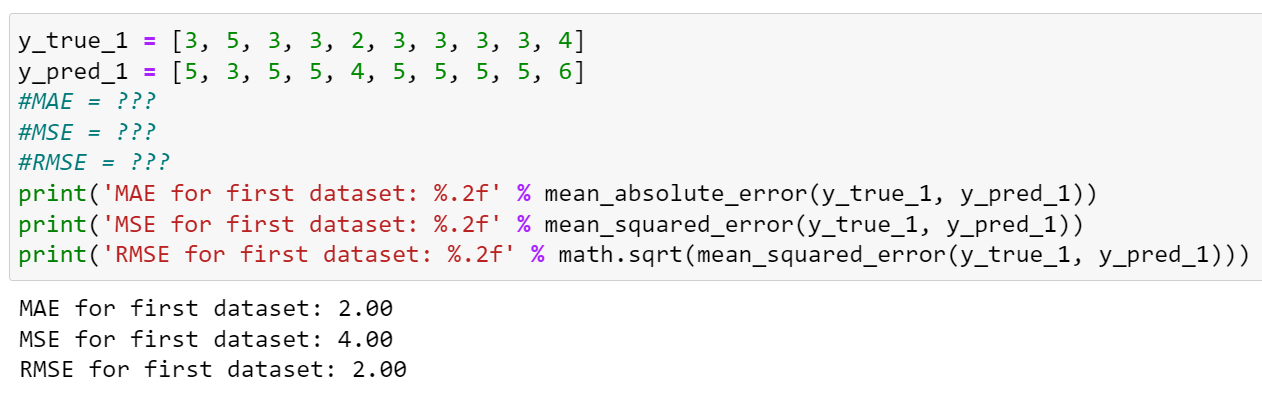
# Metryki oceny regresji liniowej

## Omówienie kodu

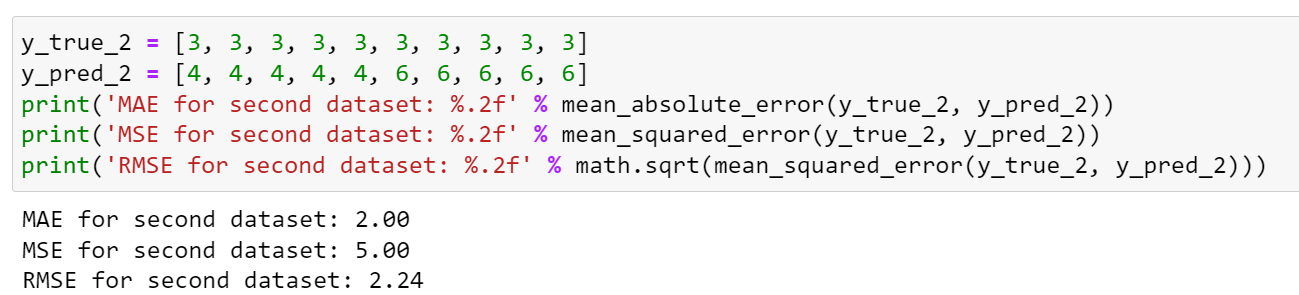
### Najprostszy, przykładowy zestaw danych



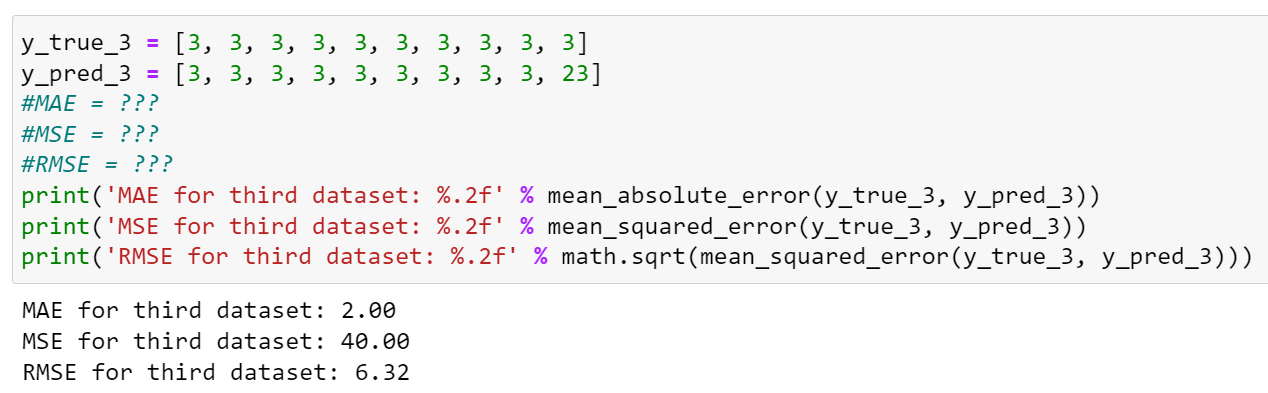
Obraz 1: Przypomnienie sposobu obliczania metryk



Obraz 2: Obliczenie metryk dla pierwszego przykładowego zestawu danych



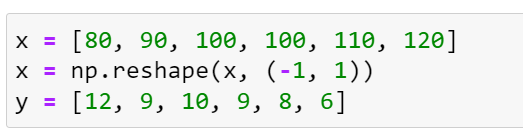
Obraz 3: Obliczenie metryk dla drugiego przykładowego zestawu danych



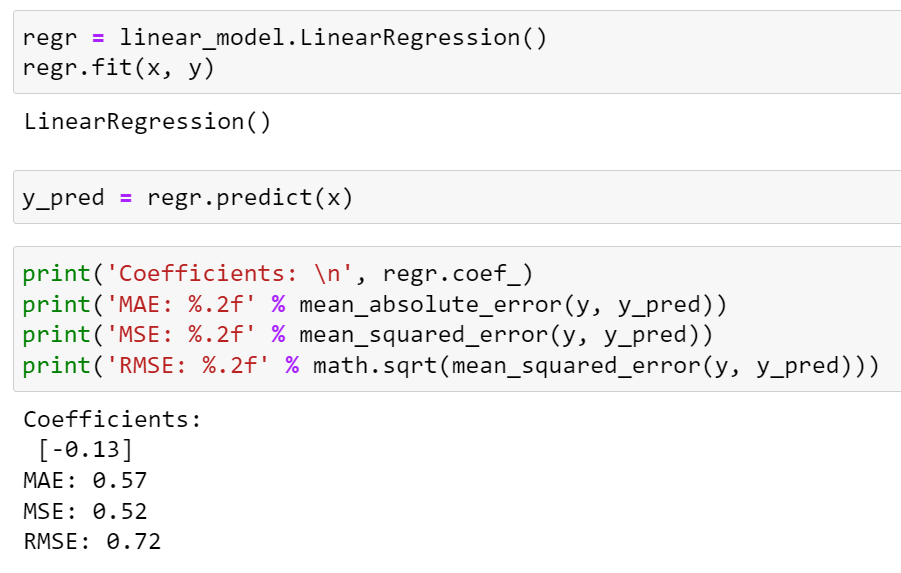
Obraz 4: Obliczenie metryk dla trzeciego przykładowego zestawu danych

Można zauważyć, że metryka RMSE jest bardzo czuła na duże odchylenia danych.

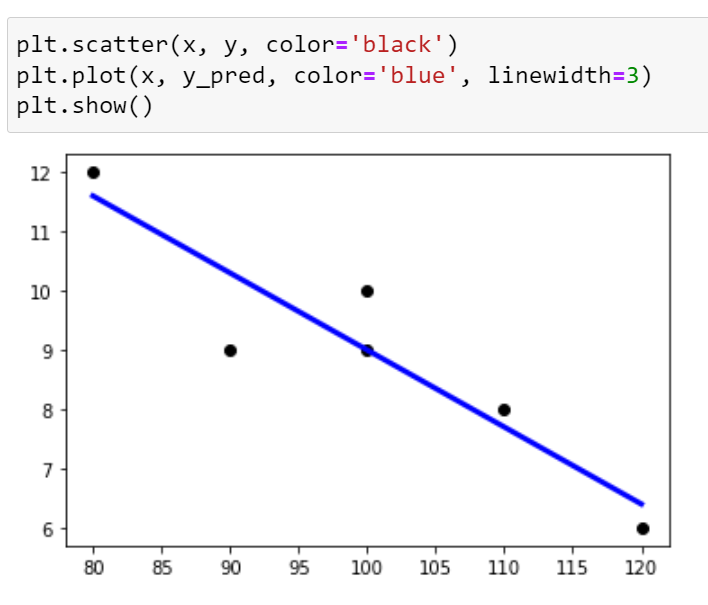
### Regresja liniowa na przykładowym zestawie danych



Obraz 5: Przygotowanie zestawu danych do regresji liniowej



Obraz 6: Regresja liniowa i obliczenie metryk



Obraz 7: Wykres dopasowanej funkcji do zestawu danych dzięki regresji liniowej

## Zadanie #1

### Pytanie

Kiedy stosować dane metryki oceny regresji?

### Odpowiedź

Kiedy analizowane dane mają charakter liniowy i można dopasować do nich wykres funkcji, w omawianym przypadku funkcji liniowej.

## Zadanie #2

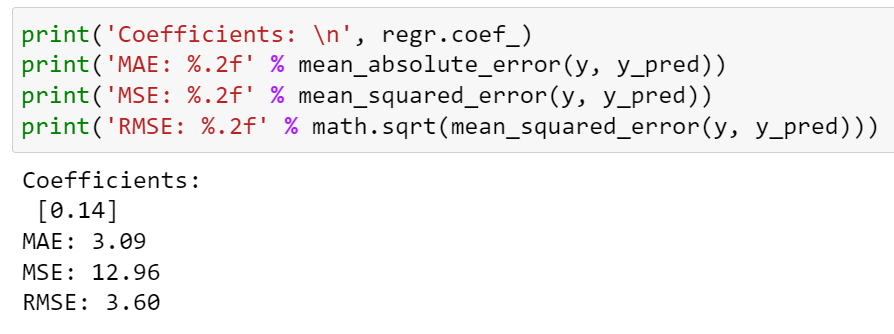
### Polecenie

Proszę zaproponować tak zbiór danych, żeby współczynnik regresji liniowej był dodatni

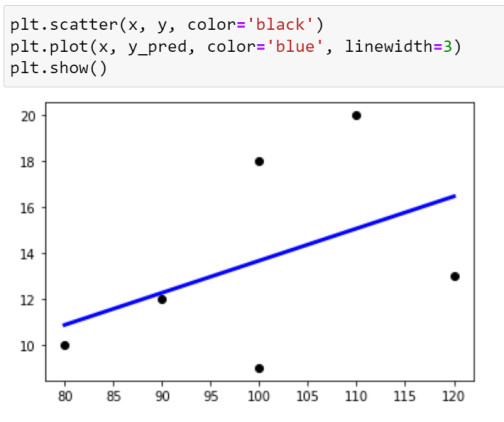
### Rozwiązanie



Obraz 8: Przykładowy zestaw danych



Obraz 9: Obliczenie metryk



Obraz 10: Wykres pokazujący dodatni współczynnik regresji liniowej

## Zadanie #3

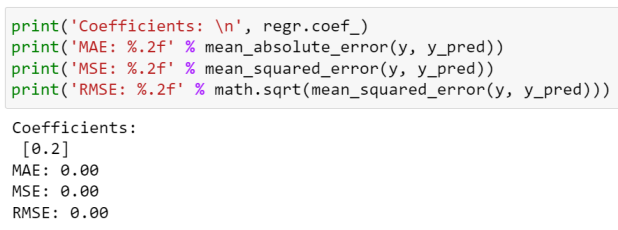
### Polecenie

Proszę zaproponować taki zbiór danych, aby błędy MAE, MSE, RMSE były równe 0

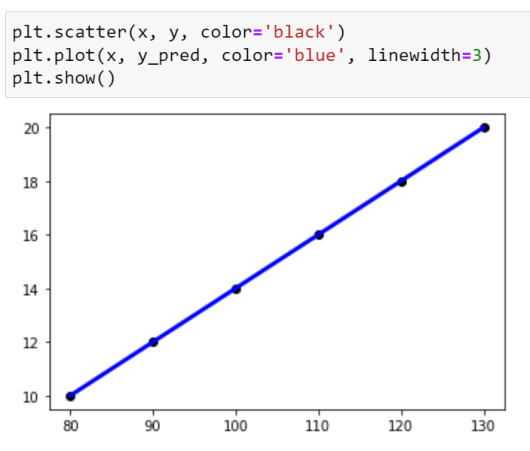
### Rozwiązanie



Obraz 11: Przykładowy zestaw danych



Obraz 12: Obliczenie metryk



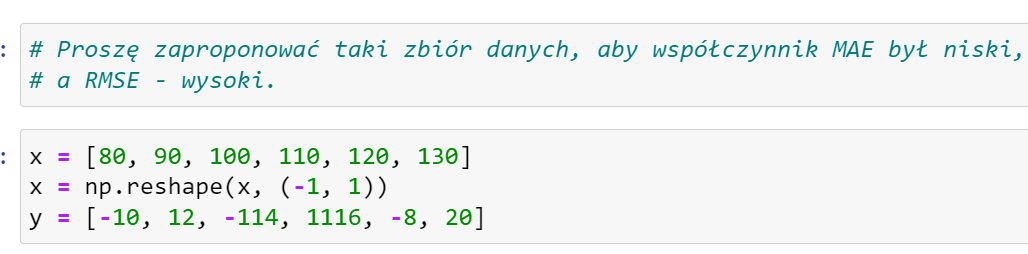
Obraz 13: Wykres pokazujący zerowe metryki

## Zadanie #4

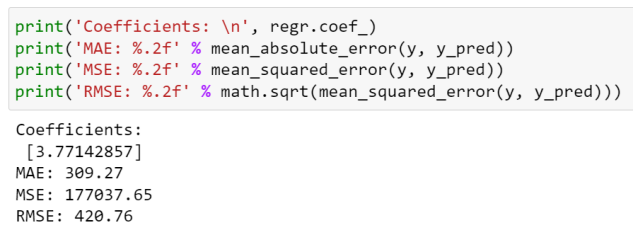
### Polecenie

Proszę zaproponować taki zbiór danych, aby współczynnik MAE był niski, a RMSE - wysoki.

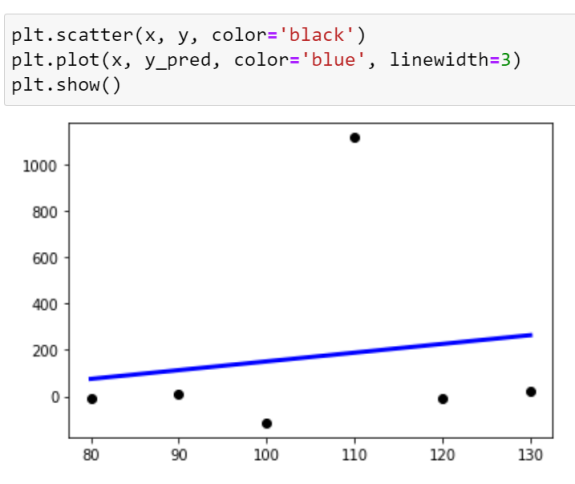
### Rozwiązanie



Obraz 14: Przykład zestaw danych



Obraz 15: Obliczenie metryk



Obraz 16: Wykres dla metryki RMSE dużo wyższej od MAE

## Zadanie #5

### Polecenie

Proszę zaproponować taki zbiór danych, aby wartość MAE była większa niż RMSE.

### Rozwiązanie

Jeśli MAE ma być większe od RMSE to tak naprawdę należy sprawdzić czy:

>

>

>

Otrzymany wzór byłby prawdą dla i mniejszych od zera; ponieważ jednak w naszych zadaniach RMSE liczymy jako pierwiastek z MSE to ostatecznie otrzymujemy tak naprawdę wartość bezwzględną z różnicy i . Różnica pomiędzy MAE i RMSE wynika jedynie z niedokładności MSE.

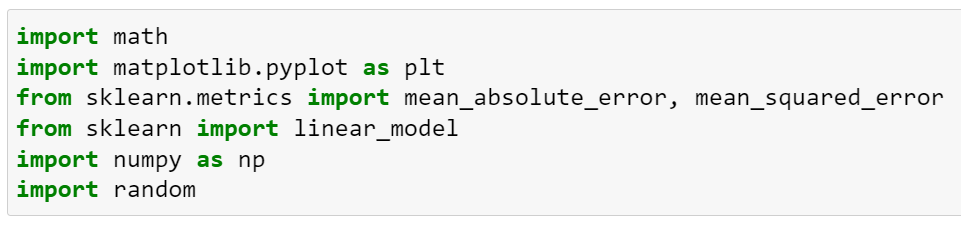
## Zadanie #6

### Polecenie

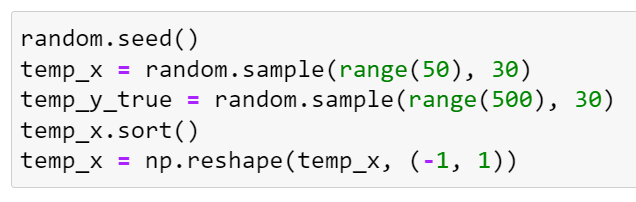
Proszę wygenerować zbiór 30 punktów w przestrzeni 2d i powtórzyć eksperyment z regresją liniową.

### Rozwiązanie

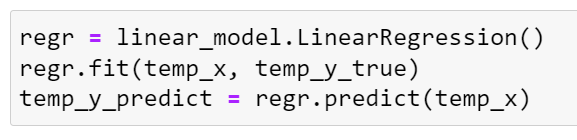
#### Generowanie zbioru



Obraz 17: Dodanie bibliotek

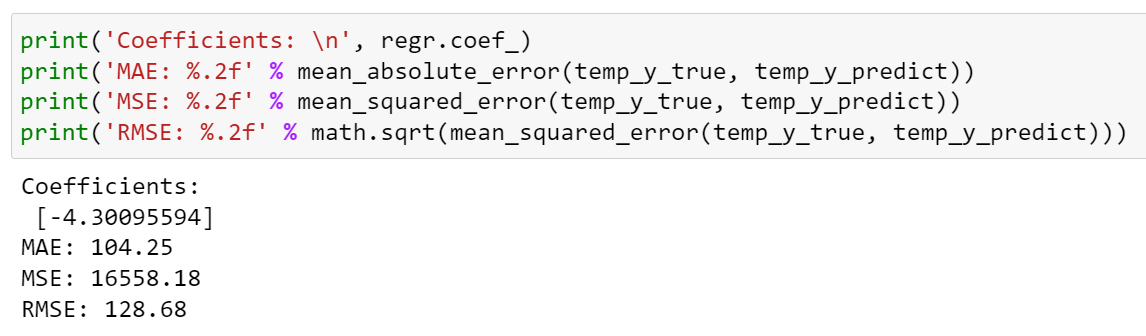


Obraz 18:Wygenerowanie zestawu danych 30 punktów

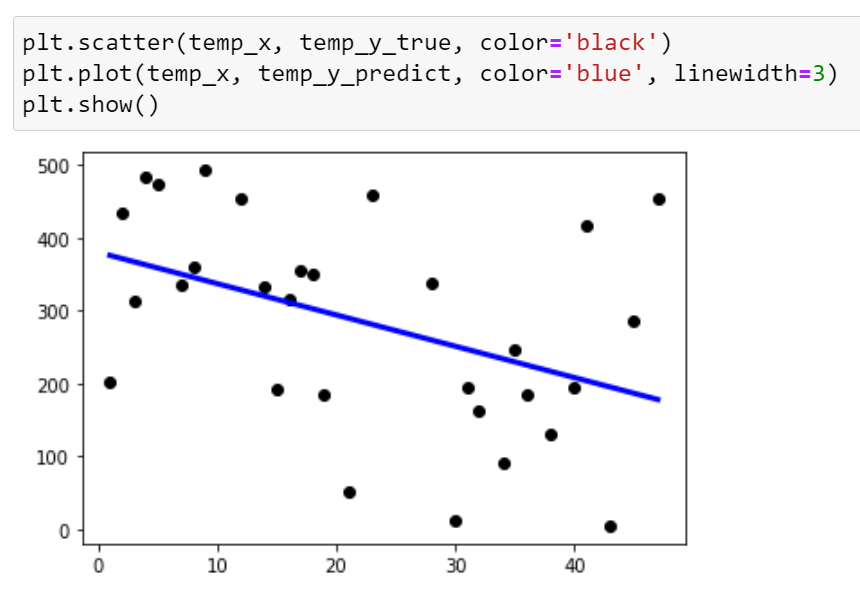


Obraz 19: Regresja liniowa

#### Wynik



Obraz 20: Obliczenie metryk



Obraz 21: Wykres regresji liniowej wygenerowanych przypadkowych punktów

## Zadanie #7

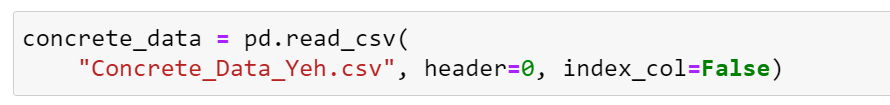
### Polecenie

Proszę powtórzyć regresję liniową w przestrzeni jednowymiarowej na innym, rzeczywistym zbiorze danych.

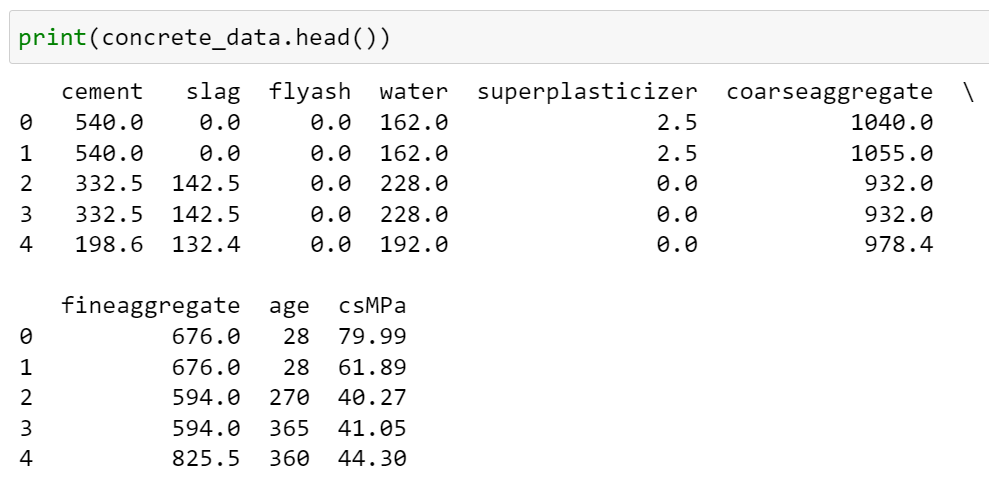
### Rozwiązanie

#### Zbiór danych

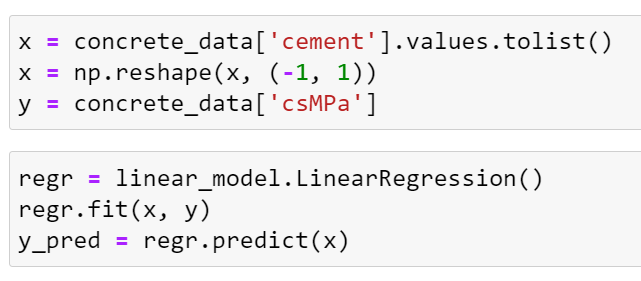
Zbiór danych pobrany został ze strony: <https://www.kaggle.com/maajdl/yeh-concret-data>, dla pokazania przykładu regresji liniowej bez wstępnej analizy, kolumna Y została wybrana na podstawie opracowania <https://www.kaggle.com/emrearslan123/concrete-compressive-strength-prediction>.



Obraz 22: Import danych z pliku



Obraz 23: Podgląd danych z pliku

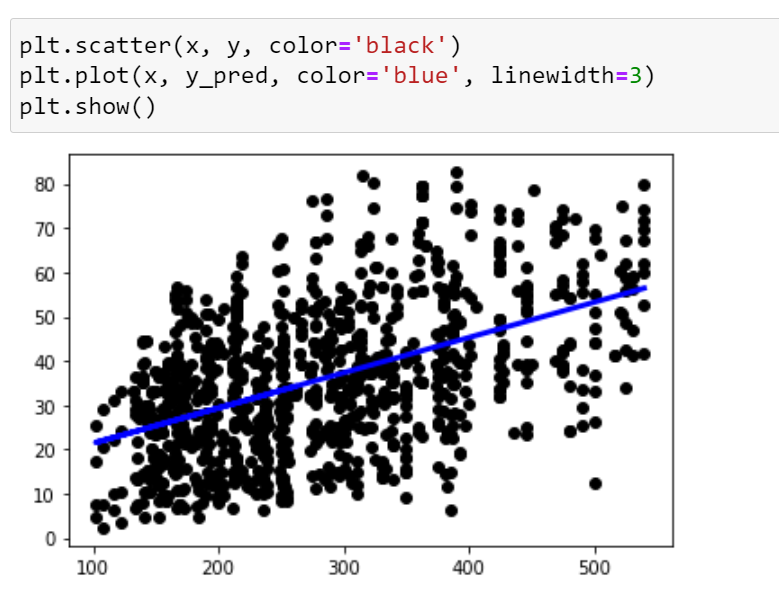


Obraz 24: Regresja liniowa

#### Wynik

#### 

Obraz 25: Obliczenie metryk



Obraz 26: Wykres regresji liniowej wybranych danych z pliku