

Analiza danych - Projekt v.0.1

Kamil Kukielka, Michał Zakielarz, Klaudia Kopeć

2024-04-29

Zaczytanie danych i ich przygotowanie

W tym projekcie będziemy korzystać z kilku zestawów danych, tak aby wykazać, że pomimo danych z różnych dziedzin ich predykcja jest w jakimś sposób możliwa

Zestaw 1

Nasz pierwszy zestaw dotyczy różnych wskaźników pewnego przedsiębiorstwa, które zmieniają się w czasie. Dane te były aktualizowane co miesiąc i obejmują zakres od 01.01.2015 do 01.02.2020 Teraz przedstawimy fragment naszych danych aby wiedzieć z czym mamy doczynienia. ## Zbiór 1 Zawiera on przychodów naszego przedsiębiorstwa.

##	Period	Revenue
## 1	01.01.2015	16010072.1195
## 2	01.02.2015	15807587.4498
## 3	01.03.2015	22047146.0236
## 4	01.04.2015	18814583.2943
## 5	01.05.2015	14021479.6117
## 6	01.06.2015	16783928.5221
## 7	01.07.2015	19161892.1949
## 8	01.08.2015	15204984.2967
## 9	01.09.2015	20603939.9751
## 10	01.10.2015	20992874.7801
## 11	01.11.2015	14993369.6576
## 12	01.12.2015	27791807.6398

Zbiór 2

Zawiera ilość sprzedarzy w naszej firmie

##	Period	Sales_quantity
## 1	01.01.2015	12729
## 2	01.02.2015	11636
## 3	01.03.2015	15922
## 4	01.04.2015	15227
## 5	01.05.2015	8620
## 6	01.06.2015	13160
## 7	01.07.2015	17254
## 8	01.08.2015	8642

## 9	01.09.2015	16144
## 10	01.10.2015	18135
## 11	01.11.2015	10841
## 12	01.12.2015	22113

Zbiór 3

Zawiera średni koszt produkcji

##	Period	Average_cost
## 1	01.01.2015	1257.76354148
## 2	01.02.2015	1358.50699981
## 3	01.03.2015	1384.69702447
## 4	01.04.2015	1235.60670482
## 5	01.05.2015	1626.62176470
## 6	01.06.2015	1275.37450776
## 7	01.07.2015	1110.57680508
## 8	01.08.2015	1759.42887025
## 9	01.09.2015	1276.25990926
## 10	01.10.2015	1157.58890434
## 11	01.11.2015	1383.02459714
## 12	01.12.2015	1256.80855786

Zbiór 4

Zawiera informację o średniej liczbie pracowników w regionie (rocznie)

##	Period	Average_annual_payroll_of_region
## 1	01.01.2015	30024676
## 2	01.02.2015	30024676
## 3	01.03.2015	30024676
## 4	01.04.2015	30024676
## 5	01.05.2015	30024676
## 6	01.06.2015	30024676
## 7	01.07.2015	30024676
## 8	01.08.2015	30024676
## 9	01.09.2015	30024676
## 10	01.10.2015	30024676
## 11	01.11.2015	30024676
## 12	01.12.2015	30024676

Zbiór 2

Obejmuje średnią dzienną temperaturę w Mumbaiu. Nasz zbiór zawiera więcej danych takich jak wilgoć, prędkość czy kierunek wiatru, jednak my skupimy się tylok na temperaturze

##	Data	Temperatura
## 1	01-01-2016	28.4
## 2	02-01-2016	26.8
## 3	03-01-2016	25.5
## 4	04-01-2016	26.4

```
## 5 05-01-2016      27.1
## 6 06-01-2016      26.9
## 7 07-01-2016      26.1
## 8 08-01-2016      26.6
## 9 09-01-2016      26.3
## 10 10-01-2016     26.0
## 11 11-01-2016     26.1
## 12 12-01-2016     25.1
```

Zbiór 3

Zawiera on kwartalne dane o długu publicznym USA (podany w milionach USD)

```
##      Data      Dług
## 1 1966-01-01 320999
## 2 1966-04-01 316097
## 3 1966-07-01 324748
## 4 1966-10-01 329319
## 5 1967-01-01 330947
## 6 1967-04-01 322893
## 7 1967-07-01 335896
## 8 1967-10-01 344663
## 9 1968-01-01 349473
## 10 1968-04-01 345369
## 11 1968-07-01 354743
## 12 1968-10-01 358029
```

Zamiana na szereg czasowy

Teraz kiedy mamy już nasze dane musimy je zamienić na szeregi czasowe

```
z1ts1 <- ts(z1df1$Revenue,start=c(2015,1),frequency = 12)
z1ts2 <- ts(z1df2$Sales_quantity,start=c(2015,1),frequency = 12)
z1ts3 <- ts(z1df3$Average_cost,start=c(2015,1),frequency = 12)
z1ts4 <- ts(z1df4$Average_annual_payroll_of_regiion,start=c(2015,1),frequency = 12)
z2ts1 <- ts(z2df1$Temperatura, start = c(2016,1,1), frequency = 365)
z3ts1 <- ts(z3df1$Dług, start = c(1966,1), frequency = 4)
```