

Wstęp do eksploracji danych - Praca domowa 3

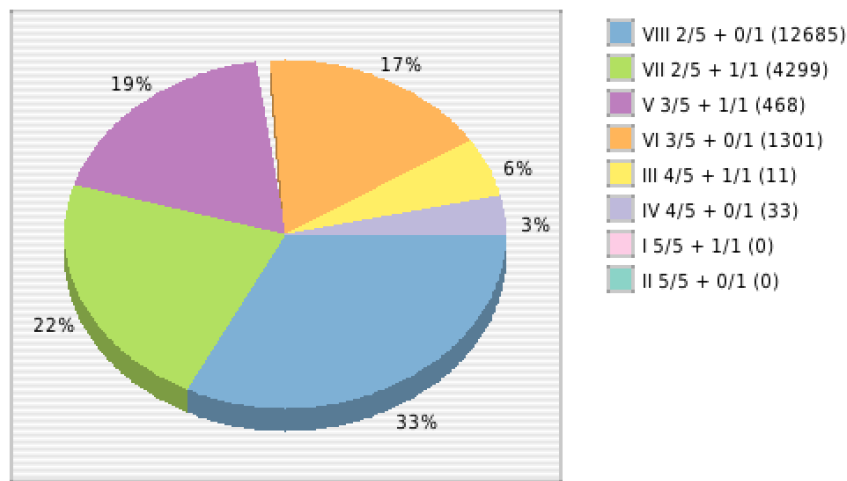
Julia Strzelczyk

Kwiecień 2024

1 Wizualizacja źródłowa



Ekstra P. podział puli [193980 zł] [07.03.2024]



Powyższa wizualizacja oraz dane pochodzą ze strony:

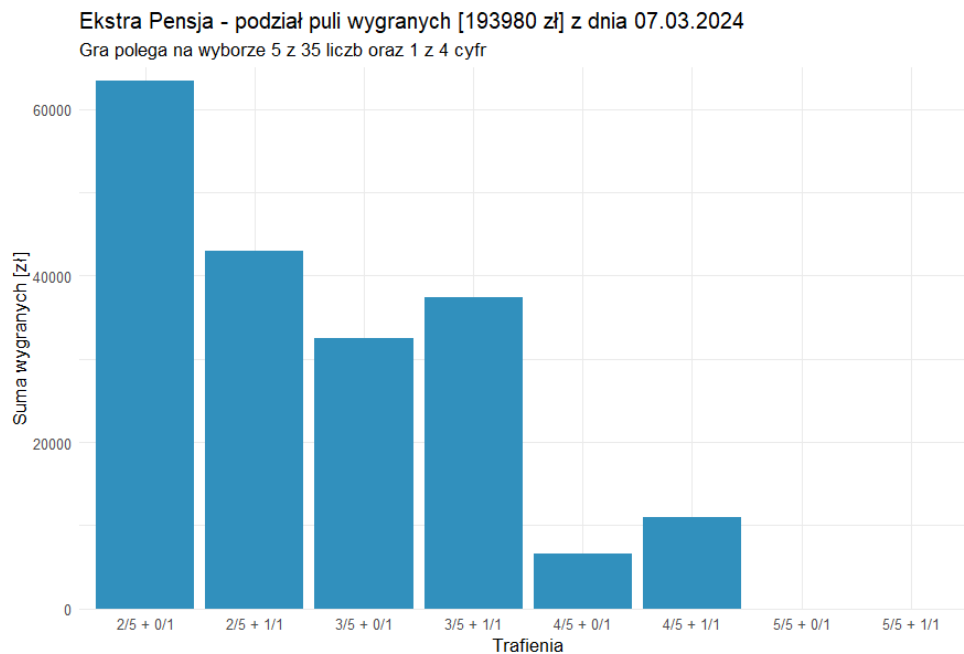
<https://www.wynikilotto.net.pl/ekstra-pensja/wyniki/07-mar-2024/>.

Wykres został opublikowany 7 marca 2024 roku.

Błędy w wizualizacji

Pierwszym problemem w tej wizualizacji jest widoczna na wykresie przerwa, która nie stanowi żadnej wartości, ponieważ przedstawione wyniki sumują się do 100%. Błędem było również zastosowanie wykresu kołowego, gdyż wiadomo, że ludzki mózg ma problem z określaniem kątów ostrych i rozwartych. Ponadto trójwymiarowość wykresu powoduje zniekształcenie wielkości i przednia część wizualizacji wygląda na większą niż w rzeczywistości. Oprócz tego trudno zrozumieć, o co chodzi w legendzie. Co więcej podane w nawiasie liczby przedstawiają liczbę osób z określoną liczbą trafień. Może to spowodować pomyłkę w prawidłowym odczytaniu wykresu, który ma obrazować sumę wygranych pieniędzy dla poszczególnych trafień.

2 Poprawiona wizualizacja



Na powyższym wykresie jesteśmy w stanie łatwo odczytać różnicę w sumie wygranych dla poszczególnych trafień. Dzięki rezygnacji z perspektywy nie zaburzamy wielkości. Ponadto dodanie podtytułu powoduje brak problemu w zrozumieniu osi x. Dodatkowo postanowiłam zrezygnować z podania informacji dotyczącej liczby osób z daną liczbą trafień, ponieważ zaburzało to czytelność treści wykresu.

Kod do wygenerowania wykresu

```
df <-  
  data.frame(  
    Trafienia = c(  
      "5/5 + 1/1",  
      "5/5 + 0/1",  
      "4/5 + 1/1",  
      "4/5 + 0/1",  
      "3/5 + 1/1",  
      "3/5 + 0/1",  
      "2/5 + 1/1",  
      "2/5 + 0/1"  
    ),  
    LiczbaGraczy = c(0, 0, 11, 33, 468, 1301, 4299, 12685),  
    Kwota_w_zl = c(5000, 25000, 1000, 200, 80, 25, 10, 5)  
  )  
library(ggplot2)  
library(dplyr)
```

```

df <- df %>%
  mutate(Wygrana_kwota = LiczbaGraczy * Kwota_w_zl)

ggplot(df, aes(x = Trafienia, y = Wygrana_kwota)) +
  geom_col(fill = "#3190bd") +
  theme_minimal() +
  labs(
    title = "Ekstra Pensja - podział puli wygranych [193980 zł] z dnia 07.03.2024",
    subtitle = "Gra polega na wyborze 5 z 35 liczb oraz 1 z 4 cyfr",
    x = "Trafienia",
    y = "Suma wygranych [zł]"
  ) +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 65000))

```