

Abstract geometric lines in the top left corner, consisting of several thin, light brown lines forming a complex, overlapping pattern of polygons.

# PRZETWARZANIE DANYCH W JĘZYKU PYTHON

mgr inż. Mateusz Bieniek



Poniżej znajdują się zadania projektowe z R (część 2).

Proszę przepisać swoje rozwiązania do Języka Python, uzyskać identyczne wyniki i w krótkim komentarzu porównać trudność przetwarzania danych przy zastosowaniu R oraz Python.

# Zadanie 1

Plik **crypto.rds** zawiera notowania wybranych kryptowalut.

1. Wczytaj dane crypto.rds do R i zapoznaj się z nimi.
2. Wybierz z danych tylko te wiersze, które dotyczą Bitcoina.
3. Pozostaw w danych tylko kolumny Date i Close.
4. Popraw kolumnę Date w taki sposób, aby była typu Date.
5. Stwórz kolumnę Rate ( $r_t$ ) na podstawie kolumny Close ( $p_t$ ) zgodnie z następującą definicją:  $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$ .
6. Posortuj dane według kolumny Rate w porządku malejącym.

# Zadanie 2

Plik **albums.csv** zawiera następujące dane dotyczące albumów muzycznych:

- artist\_id – dentyfikator artysty,
- album\_title – tytuł albumu,
- genre – gatunek muzyczny,
- year\_of\_pub – rok publikacji,
- num\_of\_tracks – liczba piosnek na płycie,
- num\_of\_sales – liczba sprzedanych płyt,
- rolling\_stone\_critic – ocena Rolling Stone Magazine,
- mtv\_critic – ocena MTV,
- music\_maniac\_critic – ocena Music Maniac.

Przeprowadź analizę tej próbki stosując dplyr.

# Zadanie 3

Plik **suicides.rds** zawiera informacje na temat liczby popełnionych samobójstw w 101 krajach świata na przestrzeni lat 1985 – 2016 z uwzględnieniem podziału na płeć oraz grupę wiekową.

1. Wskaż pięć krajów w których na przestrzeni lat 1985 – 2016 popełniono najwięcej / najmniej samobójstw na 100k mieszkańców.
2. Dla każdego roku badania wyznacz łączną liczbę samobójstw popełnionych na 100k mieszkańców na świecie.
3. Ustal łączną liczbę samobójstw popełnionych na 100k mieszkańców na przestrzeni całej próby w podziale na płeć oraz wiek.
4. Dla każdego roku badania wskaż trzy kraje, w których odnotowano największą liczbę samobójstw.
5. Znajdź kraj w którym nastąpiła największa / najmniejsza zmiana pomiędzy liczbą samobójstw na 100k mieszkańców w najgorszym roku (najwięcej samobójstw) i najlepszym roku (najmniej samobójstw).

# Zadanie 4

Katalog **gapps** zawiera trzy pliki z informacji o aplikacjach z Google Play Store:

- **free\_apps.rds** – ocenione aplikacje darmowe,
- **paid\_apps.rds** – ocenione aplikacje płatne,
- **norat\_apps.rds** – nie ocenione aplikacje płatne i darmowe.

Połącz ten dane w jeden wspólny plik i zapisz wynik pracy w postaci pliku CSV.

# Zadanie 5

Katalog **movies** zawiera trzy pliki dotyczące filmów:

- **movies.rds** – podstawowe dane na temat filmu,
- **ratings.rds** – jednostkowe oceny jakie filmom przyznali użytkownicy,
- **tags.rds** – tagi jakie do poszczególnych filmów przypisali użytkownicy.

Wykonaj następujące zadania.

1. Wyznacz średnie oceny filmów oraz liczbę osób, które oceniły każdy film i dołącz te informacje do informacji na temat filmu.
2. Ustal czas dodania ostatniego tagu do każdego z filmów i dodaj tą informację do informacji na temat filmu.
3. Dokonaj agregacji wszystkich różnych tagów dotyczących filmu, a następnie dołącz je o informacji na temat filmu.