

Big Data

# Organizacyjnie

Prowadzący:

dr Mariusz Rafało

[mrafalo@sggw.edu.pl](mailto:mrafalo@sggw.edu.pl)

<http://mariuszrafalo.pl> (hasło: BIG)

# Zadanie 1

1. Utwórz zbiór danych, zawierający ID (`customerID`) klienta oraz miesięczny abonament (`MonthlyCharges`)
2. Wyznacz licznosci klientów względem rodzaju podpisanej umowy (`Contract`). Zaprezentuj na wykresie.
3. Ilu klientów posiada wiele linii telefonicznych (`MultipleLines`)?

# Zadanie 2

1. Jaka jest minimalna i maksymalna długość stażu (`tenure`)
2. Ile miesięcznie średnio płacą mężczyźni a ile kobiety (`gender`)
3. Czy klienci, którzy podpisali umowę na rok (`Contract`), płacą więcej od pozostałych (średnio)
4. Utwórz zmienną, będącą ilorazem `TotalCharges` i `tenure`

# Zadanie 3

1. Utwórz zmienną `churn2`, która przyjmuje tylko wartości 0 (dla `Churn = No`) i 1 (dla `Churn = Yes`)
2. Wykonaj kroki:
  - a) Podziel klientów na przedziały związane z lojalnością (`tenure`). Klientów o stażu krótszym niż 6 miesięcy zignoruj:
    - i. 6-12 m-cy
    - ii. 12-36 m-cy
    - iii. >36 m-cy
  - b) Zapisz wyniki w bazie danych Hive, w tabeli `tenureResults`
  - c) Przeanalizuj zjawisko Churn w wyznaczonych grupach

# Zadanie 4

1. Wyznacz 5 klientów, o najwyższych płatnościach całkowitych (`TotalCharges`)
2. Wyznacz średnie opłaty abonamentowe (`monthlyCharges`) dla klientów którzy odeszli i dla klientów, którzy nie odeszli (`Churn`)
3. Wyznacz średnie oraz maksymalne wartości opłat dla kontraktów rocznych i miesięcznych (`monthlyCharges`)
4. Utwórz kolumnę `genderShort`, która przyjmuje wartości M i F

Dziękuję za uwagę