1. Plik *churn.txt* – chcemy znaleźć naturalne grupy dla promocji:

ID	numer klienta
LONGDIST	czas zamiejscowych rozmów na miesiąc
International	czas międzynarodowych rozmów na miesiąc
LOCAL	czas lokalnych rozmów na miesiąc
DROPPED	liczba przerwanych połączeń
PAY_MTHD	sposób opłacania rachunków
LocalBillType	taryfa połączeń lokalnych
LongDistanceBillType	taryfa dla połączeń zamiejscowych
AGE	wiek
SEX	płeć
STATUS	stan cywilny
CHILDREN	liczba dzieci
Est_Income	szacowany dochód
Car_Owner	właściciel samochodu
CHURNED	(3 kategorie) Current – nadal z firmą Vol – odchodzący, których firma chce utrzymać Invol – odchodzący, których firma nie chce utrzymać

2. Wstaw źródło *Plik zmiennych* i wybierz plik *churn.txt*
3. Dołącz węzeł *Tabela* i oglądnij dane (zauważ, że nie ma braków danych)
4. Dołącz węzeł *Typ* i ustaw na *Dane wejściowe* atrybuty: LONGDIST, International, LOCAL, a reszta atrybutów na *Brak*
5. Wstaw węzeł *Metoda k-średnich,* oglądnij ustawienia (defaultowo liczba klastrów = 5), zaznacz *Utwórz zmienną odległości* (powstanie pole $KMD-Metoda k-średnich – odległość między danym rekordem a środkiem klastra)
6. Do uzyskanego modelu podłącz *Tabela* i sprawdź przypisanie do klastrów poszczególnych rekordów
7. Oglądnij uzyskany model
8. Trzy klastry: 2,3,5 mają większość przypadków
9. W lewej części ekranu wybierz *Widok: Grupy*
10. Zaznacz tam największe klastry 2,3,5, a w prawym wybierz *Widok:Porównanie grup:*
    - **Klaster 5** – rzadko używają telefony, bo mają najmniejszą średnią minut dla każdego pola
    - **Klaster 3** – używa więcej long distance minut
    - **Klaster 2** – podobny co 3, ale mniej long distance i więcej local
11. Na razie trudno powiedzieć, czy to rozwiązanie jest użyteczne
12. Teraz chcemy znaleźć zależność między przynależnością do klastra a polem CHURNED
13. Wstaw węzeł wykresu *Rozkład*
    - wybierz $KMD-Metoda k-średnich w polu *Zmienna*
    - a CHURNED jako *Nakładanie* kolor
    - zaznacz *Normalizuj według koloru*
14. Wnioski:
    - wszyscy *Invol Churns* (usunięci przez firmę) są w 5 klastrze (klienci generalnie z małą aktywnością) – to że wszyscy wpadli do jednego klastra może być dobrym znakiem
    - *Vol Churns* – zazwyczaj najbardziej krytyczni są w 1 i 4, które są również najmniejsze – niesatysfakcjonujące – trzeba zastosować inną strategię

15. Ponieważ jest mało pól można użyć wykres *Rozrzutu* które pomoże zoptymalizować rozwiązania klastrowe
    - Wybierz LONGDIST na X, International na Y, a $KM-Metoda k-średnich jako *nakładanie*
16. Wnioski:
    - tylko klaster 5 ma małe wartości dla obu
    - klaster 4 zawiera klientów którzy mają wysokie wartości dla obydwu rodzajów połączeń
    - można również zobaczyć jak zwarte albo luźne są połączenia wewnątrz klastra: 2 i 5 są ściśle powiązane w porównaniu do 1 i 4
    - można też sprawdzić obiekty odległe
17. Cwiczenia:
    - Zmień liczbę klastrów i powtórz ćwiczenie – przy jakiej liczbie wynik jest najlepszy?
    - Dodaj pola **LocalBillType (**taryfa połączeń lokalnych) oraz **LongDistanceBillType** (taryfa dla połączeń zamiejscowych) – jak teraz wyglądają klastry?
    - Wykonaj podobne klastrowanie na danych *car_sales*
        - Spróbuj użyć węzła *Auto grupowanie*