Zajęcia 4

Pomoc na dziś, materiały i dane:

Dane zajdziesz w github (<https://github.com/databricks/Spark-The-Definitive-Guide>) użyte na wykładzie

Zadanie 1

**Notatnik 1**

Wykorzystaj dane jakie chcesz lub użyj te z wykładu.

1. Użyj poniższe funkcje Nulls, fill, explode, drop, regexp\_replace, regexp\_extract, ifnull, nullIf, replace, array\_contains.

1. Użyj 3 funkcji agregujących (które według Ciebie są najciekawsze).

Zadanie 2

Stwórz 3 funkcje UDF do wybranego zestawu danych, wymyśl co mają robić w kontekście wybranych danych.

1. Dwie funkcje działające na typach liczbowych: int, double
2. Jedna funkcja na string

Zadanie 3

**Notatnik 2**

Wykorzystaj przykłady z notatnika w SQL Windowed Aggregate Functions i przepisz funkcje używając Spark API. Nie trzeba robić ostatniej kolumny tej z sales.

Zadanie 4

Użyj ostatnich danych i użyj funkcji okienkowych LEAD, LAG, ROW\_NUMBER

- Każdą z funkcji wykonaj dla ROWS i BETWEEN