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**

Stwórz nowy notatnik w którym wykonasz poniższe zadania. Wykorzystaj dane jakie chcesz lub użyj tych 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 2 funkcje UDF do wybranego zestawu danych, wymyśl co mają robić w kontekście wybranych danych. Jedna funkcja z dekoratorem @pandas\_udf druga standardowa.

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

## Zadanie 3

**Notatnik 2**

Wykorzystaj przykłady z notatnika w SQL Windowed Aggregate Functions (cmd 11) i przepisz funkcje używając Spark API.

Zadanie 1

Do tego notatnika dopisz użycie funkcji okienkowych LEAD, LAG, FIRST\_VALUE, LAST\_VALUE, ROW\_NUMBER