# Zajęcia 7

Zadanie 1

Partycjonowanie.

Przerób notatnik z wykładów (Partycjonowanie) i dodatkowy lab (Partycjonowanie Lab), przetestuj, jak zachowuje się Spark, kiedy zmieniasz ilość partycji. Dla porównania możesz użyć lokalnych zasobów.

dbfs:/mnt/training/wikipedia/pageviews/pageviews\_by\_second.tsv

[databricks-spark-training/dataset at master · TomLous/databricks-spark-training · GitHub](https://github.com/TomLous/databricks-spark-training/tree/master/dataset)

Zadanie 2

Przejdź przez notatniki (wybrany język) Complex Data Types i Heigher Order Functions i przerób przykłady żeby zapoznać się z dostępnymi funkcjami.

Zadanie 3

Stwórz projekt w IntelliJ lub PyCharm, wykorzystaj podany przykład jest trochę popsuty (lekki challenge).

<https://github.com/cegladanych/spark-maven-demo.git>

Dodaj 3 metody jedną wczytującą plik, drugą robiącą jakąś transformacje i trzecią zapisującą dane.

Sprawdź czy jesteś w stanie stworzyć paczkę jar lub wheel/egg.