

# Zadanie 1

1. Dla każdego lotniska, wyznacz średnią wartość opóźnienia w godzinach, do dwóch miejsc po przecinku
2. Wyznacz dodatkową kolumnę z trasą lotu, bez względu na kierunek lotu
3. Podziel opóźnienie na przedziały: brak opóźnienia, małe, średnie, duże
4. Dodaj kolumnę z nazwą dnia tygodnia
5. Wyznacz jednocześnie liczbę lotów i maksymalne opóźnienie w lotach do Los Angeles (względem lotniska wylotu)

# Zadanie 2

1. Wyznacz 10 najbardziej uczęszczanych tras lotniczych
2. Wyznacz 10 tras, w których najczęściej występują opóźnienia
3. Wskaż lotniska na których są największe średnie opóźnienia (wyznacz 5 lotnisk)
4. Wyznacz godziny oraz dni tygodnia, w których występują największe średnie opóźnienia
5. Wyznacz najczęściej odwiedzane miasta (bez nazw stanów)

# Zadanie 3 (cz. 1)

Zaprojektuj proces Spark, który wyznacza wartości zgodnie z definicją:

KOLUMNA	DEFINICJA
FLIGHT_DATE	Data analizy (dzień)
MAX_DELAY	Maksymalne opóźnienie w danym okresie
AVG_DELAY	Średnie opóźnienie
NUMBER_OF_DELAYS_1	Liczba opóźnień dłuższych niż 1h
NUMBER_OF_DELAYS_2	Liczba opóźnień dłuższych niż 2h
NUMBER_OF_FLIGHTS	Liczba lotów w danym okresie

# Zadanie 3 (cz. 2)

Utwórz w Hive tabelę FLIGHT\_DETAILS o następujących kolumnach:

KOLUMNA	TYP
ID	UID
FLIGHT_DATE	STRING
MAX_DELAY	DOUBLE
AVG_DELAY	DOUBLE
NUMBER_OF_DELAYS_1	INT
NUMBER_OF_DELAYS_2	INT
NUMBER_OF_FLIGHTS	INT
CALC_DAY	CURRENT DATE