Zadanie 1

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

Zadanie 2

1. Wyznacz 10 najbardziej uczęszczanych tras lotniczych (Origin Dest)
2. Wyznacz 10 tras, w których najczęściej występują opóźnienia (ArrDelayMinutes > 15)
3. Wskaż lotniska na których (Dest) są największe średnie opóźnienia (ArrDelayMinutes) (wyznacz 5 lotnisk)
4. Wyznacz godziny(ArrTime) oraz dni tygodnia (DayOfWeek), w których występują największe średnie opóźnienia (ArrDelayMinutes)
5. Wyznacz najczęściej odwiedzane miasta (DestCityName)(bez nazw stanów)

Zadanie 3 (cz. 1)

Zaprojektuj proces Spark, który wyznacza wartości zgodnie z definicją:

KOLUMNA	DEFINICJA
FLIGHT_DATE	Data lotu (dzień) (FlightDate)
MAX_DELAY	Maksymalne opóźnienie w danym okresie (ArrDelayMinutes)
AVG_DELAY	Średnie opóźnienie (ArrDelayMinutes)
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