

Eksploracja danych - etap 1

Krzysztof Nasuta 193328, Filip Dawidowski 193433, Aleks Iwicki 193354

1. Ogólny opis zbioru

Departament Transportu Stanów Zjednoczonych (DOT), za pośrednictwem Biura Statystyki Transportu, monitoruje punktualność krajowych lotów realizowanych przez dużych przewoźników lotniczych. Zbiorcze informacje na temat liczby lotów punktualnych, opóźnionych, odwołanych oraz przekierowanych są publikowane w comiesięcznym raporcie „Air Travel Consumer Report” oraz w tym zestawie danych dotyczącym opóźnień i odwołań lotów z 2015 roku.

Pojedynczy wiersz zbioru danych reprezentuje pojedynczy lot. Zbiór danych zawiera informacje o czasie odlotu i przylotu, czasie opóźnienia, przyczynie opóźnienia, a także inne szczegóły dotyczące lotu.

2. Określenie celu eksploracji i kryteriów sukcesu

Celem eksploracji jest predykcja, czy lot będzie opóźniony, czy nie. Uznajemy, że lot jest opóźniony, jeśli atrybut `ARRIVAL_DELAY` jest większy niż 15. Dodatkowo chcemy zrozumieć, które czynniki mają największy wpływ na opóźnienia lotów.

Kryterium sukcesu jest osiągnięcie dokładności klasyfikacji opóźnienia na poziomie 85%

3. Charakterystyka zbioru

- Pochodzenie: [Kaggle](#)
- Liczba przykładów: 5819080
- Format: CSV (3 pliki: `flights.csv` - właściwy zbiór, `airports.csv` - informacje o lotniskach, `airlines.csv` - informacje o liniach lotniczych)
- Ilość zbiorów danych: 1

4. Opis atrybutów

Nazwa	Typ	Opis
YEAR	Numeryczny	Rok lotu
MONTH	Numeryczny	Miesiąc lotu
DAY	Numeryczny	Dzień miesiąca lotu
DAY_OF_WEEK	Numeryczny	Dzień tygodnia lotu
AIRLINE	Tekst	Indentyfikator linii lotniczej
FLIGHT_NUMBER	Numeryczny	Numer lotu
TAIL_NUMBER	Tekst	Numer rejestracyjny samolotu
ORIGIN_AIRPORT	Tekst	Kod IATA lotniska wylotu
DESTINATION_AIRPORT	Tekst	Kod IATA lotniska przylotu
SCHEDULED_DEPARTURE	Numeryczny	Planowany czas odlotu (w formacie HHMM)
DEPARTURE_TIME	Numeryczny	Czas odlotu (w formacie HHMM)
DEPARTURE_DELAY	Numeryczny	Całkowite opóźnienie odlotu (w minutach)
TAXI_OUT	Numeryczny	Ilość minut spędzonych na kołowaniu przed odlotem

Nazwa	Typ	Opis
WHEELS_OFF	Numeryczny	Czas startu (w formacie HHMM)
SCHEDULED_TIME	Numeryczny	Planowany czas lotu (w minutach)
ELAPSED_TIME	Numeryczny	Całkowity czas lotu (w minutach) = AIR_TIME+TAXI_IN+TAXI_OUT
AIR_TIME	Numeryczny	Czas lotu (w minutach)
DISTANCE	Numeryczny	Dystans lotu (w milach)
WHEELS_ON	Numeryczny	Czas lądowania (w formacie HHMM)
TAXI_IN	Numeryczny	Ilość minut spędzonych na kołowaniu po lądowaniu
SCHEDULED_ARRIVAL	Numeryczny	Planowany czas przylotu (w formacie HHMM)
ARRIVAL_TIME	Numeryczny	Czas przylotu (w formacie HHMM) = WHEELS_ON+TAXI_IN
ARRIVAL_DELAY	Numeryczny	Całkowite opóźnienie przylotu (w minutach) = ARRIVAL_TIME-SCHEDULED_ARRIVAL
DIVERTED	Prawda/Fałsz	Czy lot był przekierowany? (1-tak, 0-nie)
CANCELLED	Prawda/Fałsz	Czy lot był odwołany? (1-tak, 0-nie)
CANCELLATION_REASON	Tekst	Przyczyna odwołania lotu (A - przewoźnik, B - pogoda, C - National Air System, D - Bezpieczeństwo), tylko dla odwołanych lotów
AIR_SYSTEM_DELAY	Numeryczny	Opóźnienie spowodowane przez system lotniczy (w minutach)
SECURITY_DELAY	Numeryczny	Opóźnienie spowodowane przez kontrole bezpieczeństwa (w minutach)
AIRLINE_DELAY	Numeryczny	Opóźnienie spowodowane przez przewoźnika (w minutach)
LATE_AIRCRAFT_DELAY	Numeryczny	Opóźnienie spowodowane przez samolot (w minutach)
WEATHER_DELAY	Numeryczny	Opóźnienie spowodowane przez pogodę (w minutach)

















