

Proj. Analityczny. Linie lotnicze.

Dominik Szymkowiak

1 czerwca 2019

Krótki opis danych

W bazie dbad_flights znajduje się 5 różnych tabel, jednak używana głównie była tabela “Flight_delays”. Można z niej było odczytać m.in. takie dane jak: kiedy, skąd, o której, z jakim opóźnieniem leciał samolot, o której dotarł na miejsce oraz jakiej linii lotniczej był to samolot.

Łączenie się z bazą

```
login <- readline()
```

```
haslo <- readline()
```

```
connection <- DBI::dbConnect(odbc::odbc(),  
  Driver = "SQL Server",  
  Server = "mssql-2016.labs.wmi.amu.edu.pl",  
  Database = "dbad_flights",  
  UID = login,  
  PWD = haslo,  
  Port = 1433)
```

Zadanie 1

Jakie było średnie opóźnienie przylotu?

```
SELECT AVG(arr_delay_new) 'avg_delay'  
FROM Flight_delays
```

Table 1: 1 records

| avg_delay |
|------------------|
| 15.9115212681785 |

Odpowiedź: Średnie opóźnienie każdego lotu wynosił prawie 16 minut.

Zadanie 2

Jakie było maksymalne opóźnienie przylotu?

```
SELECT MAX(arr_delay_new) 'max_delay'  
FROM Flight_delays
```

Table 2: 1 records

| max_delay |
|-----------|
| 1895 |

Odpowiedź: Maksymalne opóźnienie lotu wynosiło 1895 minut, czyli ponad 31.5h.

Zadanie 3

Który lot miał największe opóźnienie przylotu?

```
SELECT carrier, origin_city_name, dest_city_name, fl_date, arr_delay_new
FROM Flight_delays
WHERE arr_delay_new = (SELECT MAX(arr_delay_new)
                       FROM Flight_delays)
```

Table 3: 1 records

| carrier | origin_city_name | dest_city_name | fl_date | arr_delay_new |
|---------|------------------|-----------------|------------|---------------|
| AA | Kona, HI | Los Angeles, CA | 2017-07-26 | 1895 |

Odpowiedź: Największe opóźnienie przylotu miał lot z Kona do Los Angeles w dniu 26 lipca 2017r.

Zadanie 4

Które dni tygodnia są najgorsze do podróżowania?

```
SELECT weekday_name, AVG(arr_delay_new) 'avg_delay'
FROM Flight_delays
INNER JOIN Weekdays ON Flight_delays.day_of_week = Weekdays.weekday_id
GROUP BY weekday_name
ORDER BY AVG(arr_delay_new) DESC
```

Table 4: 7 records

| weekday_name | avg_delay |
|--------------|-----------|
| Friday | 20.80747 |
| Monday | 18.04801 |
| Wednesday | 16.10514 |
| Thursday | 15.64696 |
| Saturday | 15.21876 |
| Tuesday | 12.88056 |
| Sunday | 12.77606 |

Odpowiedź: Piątki i poniedziałki to najgorsze dni do podróżowania.

Zadanie 5

Które linie lotnicze latające z San Francisco (SFO) mają najmniejsze opóźnienia przylotu?

```
SELECT Airlines.airline_name, AVG(arr_delay_new) 'avg_delay'
FROM Flight_delays
INNER JOIN Airlines ON Flight_delays.airline_id = Airlines.airline_id
WHERE Airlines.airline_id IN
  (SELECT Flight_delays.airline_id FROM Flight_delays WHERE Flight_delays.origin = 'SFO')
GROUP BY Airlines.airline_name
ORDER BY AVG(arr_delay_new) DESC
```

Table 5: Displaying records 1 - 10

| airline_name | avg_delay |
|----------------------------|-----------|
| JetBlue Airways: B6 | 28.841148 |
| Frontier Airlines Inc.: F9 | 18.980300 |
| American Airlines Inc.: AA | 18.375314 |
| United Air Lines Inc.: UA | 16.950403 |
| SkyWest Airlines Inc.: OO | 16.808273 |
| Virgin America: VX | 13.964467 |
| Southwest Airlines Co.: WN | 13.823983 |
| Delta Air Lines Inc.: DL | 12.258788 |
| Alaska Airlines Inc.: AS | 7.453927 |
| Hawaiian Airlines Inc.: HA | 4.202719 |

Odpowiedź: Linie lotnicze latające z San Francisco mające najmniejsze opóźnienia przylotu to Hawaiian Airlines oraz Alaska Airlines.

Zadanie 6

Jaka część linii lotniczych ma regularne opóźnienia, tj. jej lot ma średnio co najmniej 10 min. opóźnienia?

```
SELECT CAST(COUNT(T.airline_id) AS DECIMAL(5,1))
/
(SELECT COUNT(DISTINCT Flight_delays.airline_id) FROM Flight_delays) 'late_proportion'
FROM
  (SELECT Flight_delays.airline_id
   FROM Flight_delays
   GROUP BY Flight_delays.airline_id
   HAVING AVG(Flight_delays.arr_delay_new) > 10)
AS T
```

Table 6: 1 records

| late_proportion |
|-----------------|
| 0.833333333333 |

Odpowiedź: Około 83% lini lotniczych ma regularne opóźnienia.

Zadanie 7

Jak opóźnienia wylotów wpływają na opóźnienia przylotów?

```
SELECT (AVG(arr_delay_new*dep_delay_new)-(AVG(arr_delay_new)*AVG(dep_delay_new)))
/
(STDEV(arr_delay_new)*STDEV(dep_delay_new)) 'Pearsons r'
FROM Flight_delays
```

Table 7: 1 records

| Pearsons r |
|-------------------|
| 0.973708052333025 |

Odpowiedź: Współczynnik Pearsona między czasem opóźnienia wylotów a czasem opóźnienia przylotów wynosi ponad 0,97, a więc opóźnienia wylotów bardzo wpływają na opóźnienia przylotów.

Zadanie 8

Która linia lotnicza miała największy wzrost (różnica) średniego opóźnienia przylotów w ostatnim tygodniu miesiąca, tj. między 1-23 a 24-31 lipca?

```
SELECT TOP 1 airline_name,

((SELECT AVG(F2.arr_delay_new)
FROM Flight_delays F2
WHERE ((F2.day_of_month BETWEEN 24 AND 31) AND F2.month=7) AND F1.airline_id=F2.airline_id)
GROUP BY F2.airline_id)
-
(SELECT AVG(F3.arr_delay_new)
FROM Flight_delays F3
WHERE ((F3.day_of_month BETWEEN 1 AND 23) AND F3.month=7) AND F1.airline_id=F3.airline_id)
GROUP BY F3.airline_id)) 'delay_increase'

FROM Flight_delays F1 INNER JOIN Airlines ON F1.airline_id=Airlines.airline_id
GROUP BY airline_name, F1.airline_id
ORDER BY 'delay_increase' DESC
```

Table 8: 1 records

| airline_name | delay_increase |
|----------------------------|----------------|
| Southwest Airlines Co.: WN | 0.584763 |

Odpowiedź: Linia lotnicza mające największy wzrost średniego opóźnienia przylotów w ostatnim tygodniu miesiąca to Southwest Airlines Co.

Zadanie 9

Które linie lotnicze latają zarówno na trasie SFO › PDX (Portland), jak i SFO › EUG (Eugene)?

```
SELECT DISTINCT airline_name
FROM Flight_delays F0 INNER JOIN Airlines ON Airlines.airline_id = F0.airline_id
WHERE F0.airline_id IN
  (SELECT F1.airline_id
   FROM Flight_delays F1 WHERE (F1.origin='SFO' AND F1.dest='PDX') AND (F1.airline_id IN
    (SELECT F2.airline_id
     FROM Flight_delays F2 WHERE (F2.origin='SFO') AND (F2.dest='EUG'))))
```

Table 9: 2 records

| airline_name |
|---------------------------|
| SkyWest Airlines Inc.: OO |
| United Air Lines Inc.: UA |

Odpowiedź: Linie lotnicze latające zarówno na trasie SFO>PDX, jak i SFO>EUG to: SkyWest Airlines Inc. oraz United Air Lines Inc.

Zadanie 10

Jak najszybciej dostać się z Chicago do Stanfordu, zakładając wylot po 14:00 czasu lokalnego?

```
SELECT origin, dest, AVG(arr_delay_new) 'avg_delay'
FROM Flight_delays
WHERE (((origin='MDW' OR origin = 'ORD')
      AND (dest='SFO' OR dest='SJC' OR dest='OAK'))
      AND crs_dep_time>1400)
GROUP BY origin, dest
ORDER BY 'avg_delay' DESC
```

Table 10: 5 records

| origin | dest | avg_delay |
|--------|------|-----------|
| ORD | SFO | 22.19253 |
| MDW | SFO | 19.85714 |
| MDW | SJC | 17.20000 |
| ORD | SJC | 14.81111 |
| MDW | OAK | 12.12903 |

Odpowiedź: Najszybszą możliwą opcją dostania się do Stanfordu z Chicago jest lot z Midway do Oakland.