

Analiza bazy Linie Lotnicze

Dawid Piskadło

30.05.2019

Wstępna uwaga:

W zadaniach 1,2,4,5,10 dane są zaokrąglone w dół do pełnych minut.

Zadanie 1:

Zapytanie:

```
SELECT (FLOOR(AVG(arr_delay_new))) AS średnie_opóźnienie
FROM dbad_flights.dbo.Flight_delays
```

Wynik:

Średnie opóźnienie przylotu było równe 15 minut.

Zadanie 2:

Zapytanie:

```
SELECT FLOOR(MAX(arr_delay_new)) AS średnie_opóźnienie
FROM dbad_flights.dbo.Flight_delays
```

Wynik:

Największe opóźnienie przylotu było równe 1895 minut.

Zadanie 3:

Zapytanie:

```
SELECT TOP 1 carrier,
           fl_num,
           origin_city_name,
           dest_city_name,
           fl_date,
           arr_delay_new

FROM dbad_flights.dbo.Flight_delays AS F1
     INNER JOIN dbad_flights.dbo.Airlines AS F2
       ON F1.airline_id = F2.airline_id
ORDER BY F1.arr_delay_new DESC
```

Wynik:

Największe opóźnienie przylotu miał następujący lot:

| carrier | fl_num | origin_city_name | dest_city_name | fl_date | arr_delay_new |
|---------|--------|------------------|-----------------|------------|---------------|
| AA | 58 | Kona, HI | Los Angeles, CA | 2017-07-26 | 1895 |

Zadanie 4:

Zapytanie:

```
SELECT  weekday_name,  
        FLOOR(AVG(arr_delay_new)) AS avg_delay  
FROM    dbad_flights.dbo.Flight_delays AS F1  
        INNER JOIN dbad_flights.dbo.Weekdays AS W1  
          ON F1.day_of_week = W1.weekday_id  
GROUP BY weekday_name,weekday_id  
ORDER BY AVG(arr_delay_new) DESC
```

Wynik:

Tabela najgorszych dni do podróżowania pod względem opóźnień przylotu prezentuje się następująco:

| weekday_name | avg_delay |
|--------------|-----------|
| Friday | 20 |
| Monday | 18 |
| Wednesday | 16 |
| Thursday | 15 |
| Saturday | 15 |
| Tuesday | 12 |
| Sunday | 12 |

Zadanie 5:

Zapytanie:

```
SELECT  airline_name,  
        FLOOR(AVG(F1.arr_delay_new)) AS średnie_opóźnienie  
FROM    dbad_flights.dbo.Airlines AS A1  
        INNER JOIN dbad_flights.dbo.Flight_delays AS F1  
          ON F1.airline_id = A1.airline_id  
WHERE   F1.airline_id IN (SELECT F2.airline_id  
                          FROM dbad_flights.dbo.Flight_delays AS F2  
                          WHERE F2.origin = 'SFO')  
GROUP BY airline_name  
ORDER BY FLOOR(AVG(arr_delay_new)) DESC
```

Wynik:

Linie latające z San Francisco posiadające najmniejsze opóźnienia to odpowiednio JetBlue Airways, American Airlines i Frontier Airlines. Tabela tych linii prezentuje się następująco:

| airline_name | średnie_opóźnienie |
|----------------------------|--------------------|
| JetBlue Airways: B6 | 28 |
| Frontier Airlines Inc.: F9 | 18 |
| American Airlines Inc.: AA | 18 |
| United Air Lines Inc.: UA | 16 |
| SkyWest Airlines Inc.: OO | 16 |
| Virgin America: VX | 13 |
| Southwest Airlines Co.: WN | 13 |
| Delta Air Lines Inc.: DL | 12 |

| airline_name | średnie_opóźnienie |
|----------------------------|--------------------|
| Alaska Airlines Inc.: AS | 7 |
| Hawaiian Airlines Inc.: HA | 4 |

Zadanie 6:

Zapytanie:

```
SELECT ROUND((CONVERT(float,COUNT(*),2)/(SELECT COUNT(DISTINCT airline_id)
FROM dbad_flights.dbo.Flight_delays)),3)
FROM (SELECT A1.airline_name
FROM dbad_flights.dbo.Airlines AS A1
INNER JOIN dbad_flights.dbo.Flight_delays AS F1
ON F1.airline_id = A1.airline_id
GROUP BY airline_name
HAVING AVG(arr_delay_new)>10) AS T1
```

Wynik:

Stosunek liczby linii lotniczych posiadających regularne opóźnienia do liczby wszystkich linii lotniczych znajdujących się w tabeli Flight_delays wynosi 0.833.

Zadanie 7:

Zapytania:

```
SELECT arr_delay_new FROM dbad_flights.dbo.Flight_delayss
```

```
SELECT dep_delay_new FROM dbad_flights.dbo.Flight_delay
```

R:

```
avg_arr_delay_conv <- as.vector(unlist(avg_arr_delay))
```

```
avg_dep_delay_conv <- as.vector(unlist(avg_dep_delay))
```

```
avg_dep_arr_temp_result <- cor.test(avg_dep_delay_conv,avg_arr_delay_conv, method = "pearson")
```

```
avg_dep_arr_correlation <- round(avg_dep_arr_temp_result$estimate,3)
```

Wynik:

Współczynnik korelacji Pearsona między opóźnieniami przylotów a opóźnieniami wylotów wynosi 0.976.

Zadanie 8:

Zapytanie:

```
SELECT TOP 1 A1.airline_name,
(AVG(F1.arr_delay_new)-(SELECT AVG(F2.arr_delay_new)
FROM dbad_flights.dbo.Flight_delays AS F2
JOIN dbad_flights.dbo.Airlines AS A2
ON F2.airline_id = A2.airline_id
WHERE A2.airline_name = A1.airline_name AND F2.month = 7 AND F2.
GROUP BY A2.airline_name)) AS 'delay_increase'
FROM dbad_flights.dbo.Airlines AS A1
```

```

INNER JOIN dbad_flights.dbo.Flight_delays AS F1
    ON A1.airline_id = F1.airline_id
WHERE F1.month = 7 AND F1.day_of_month > 23
GROUP BY A1.airline_name
ORDER BY test DESC

```

Wynik:

Linia posiadającą najwyższy przyrost średniego opóźnienia między 01-23 lipca a 24-31 lipca była następująca linia lotnicza:

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

Zadanie 9:

Zapytanie:

```

SELECT airline_name
FROM dbad_flights.dbo.Airlines AS A1
    INNER JOIN dbad_flights.dbo.Flight_delays AS F1
        ON F1.airline_id = A1.airline_id
    INNER JOIN dbad_flights.dbo.Flight_delays AS F2
        ON F1.airline_id = F2.airline_id
WHERE ((F1.origin = 'SFO' AND F1.dest = 'PDX') OR (F1.origin = 'SFO' AND F1.dest = 'EUG')) AND
(((F2.origin = 'SFO' AND F2.dest = 'PDX') OR (F2.origin = 'SFO' AND F2.dest = 'EUG')) AND F1.dest != F2.dest)
GROUP BY airline_name

```

Wynik:

Na trasach zarówno SFO->PDX jak i SFO->EUG latają poniższe linie lotnicze:

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

Zadanie 10:

Zapytanie:

```

SELECT origin,
    dest,
    FLOOR(AVG(arr_delay_new)) AS avg_delay
FROM dbad_flights.dbo.Flight_delays
WHERE (origin = 'MDW' OR origin = 'ORD') AND (dest = 'SFO' OR dest = 'SJC' OR dest = 'OAK') AND crs_dep = 1
GROUP BY origin,dest
ORDER BY AVG(arr_delay_new) DESC

```

Wynik:

Według danych statystycznych najszybszym sposobem na podróż z Chicago do Stanfordu przy założeniu wylotu po godzinie 14:00 będzie lot z Midway do Oakland. Tabela średnich opóźnień dla poszczególnych tras lotu prezentuje się następująco:

| origin | dest | avg_delay |
|--------|------|-----------|
| ORD | SFO | 22 |
| MDW | SFO | 19 |
| MDW | SJC | 17 |
| ORD | SJC | 14 |
| MDW | OAK | 12 |