Analiza

Aliaksandr Halaunia

30 04 2022

Projekt nr 2 - Bazy danych - Linie lotnicze

PODLĄCZENIE DO BAZY

```
library(odbc)
```

```
## Warning: package 'odbc' was built under R version 4.1.3
```

```
library(DBI)
con <- dbConnect(odbc(),
                 Driver = "ODBC Driver 17 for SQL Server",
                 Server = "mssql-2017.labs.wmi.amu.edu.pl",
                 Database = "dbad_flights",
                 uid = Sys.getenv("userid"),
                 pwd = Sys.getenv("pwd"),
                 port = 1443)
```

ZADANIE 1

Jakie było średnie opóźnienie przylotu?

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

Table 1: 1 records

avg_delay
15.91152

ZADANIE 2

Jakie było maksymalne opóźnienie przylotu?

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

Table 2: 1 records

max_delay
1895

ZADANIE 3

Który lot miał największe opóźnienie przylotu?

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

Table 3: 1 records

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

ZADANIE 4

Które dni tygodnia są najgorsze do podróżowania?

```
SELECT   AVG(arr_delay_new) AS arr_delay,
    CASE day_of_week
    WHEN '1' THEN 'Monday'
    WHEN '2' THEN 'Tuesday'
    WHEN '3' THEN 'Wednesday'
    WHEN '4' THEN 'Thursday'
    WHEN '5' THEN 'Friday'
    WHEN '6' THEN 'Saturday'
    WHEN '7' THEN 'Sunday'
    END AS day_of_week
  FROM Flight_delays
  GROUP BY day_of_week
  ORDER BY arr_delay DESC;
```

Table 4: 7 records

arr_delay	day_of_week
20.80747	Friday
18.04801	Monday
16.10514	Wednesday
15.64696	Thursday
15.21876	Saturday
12.88056	Tuesday
12.77606	Sunday

ZADANIE 5

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

```sql
WITH najmOp1
AS
(
    SELECT A.airline_id, airline_name, origin_city_name
    FROM Airlines A
    LEFT OUTER JOIN Flight_delays F
    ON A.airline_id = F.airline_id
    WHERE origin_city_name LIKE 'San Francisco%'
    GROUP BY origin_city_name, airline_name, A.airline_id
),
naimOp2
AS
(
    SELECT airline_id,
    AVG(arr_delay_new) AS arr_delay
    FROM Flight_delays F1
    GROUP BY airline_id
)
SELECT N.airline_name, N.origin_city_name, D.arr_delay
FROM najmOp1 N
LEFT OUTER join naimOp2 D
ON N.airline_id = D.airline_id
ORDER BY arr_delay DESC
```

Table 5: Displaying records 1 - 10

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

ZADANIE 6

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

```sql
SELECT (CAST((SELECT COUNT(*) AS number
    FROM (SELECT COUNT(airline_name) AS a
        FROM Airlines A
        LEFT OUTER JOIN Flight_delays F
        ON A.airline_id = F.airline_id
        GROUP BY airline_name
        HAVING AVG(arr_delay_new)>10) AS subq) AS float)
        /
        (SELECT COUNT(*) AS number
            FROM (SELECT COUNT(airline_name) AS a
```

```
                FROM Airlines A
                LEFT OUTER JOIN Flight_delays F
                ON A.airline_id = F.airline_id
                GROUP BY airline_name
                HAVING AVG(arr_delay_new)>0) AS subq)) AS late_propotion;
```

Table 6: 1 records

late_propotion
0.8333333

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)))
/ (STDEVP(arr_delay_new)*STDEVP(dep_delay_new))) AS 'Pearsons r' FROM Flight_delays
```

Table 7: 1 records

Pearsons r
0.97371

ZADANIE 8

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

```
WITH firs
AS
(
    SELECT A.airline_id, airline_name,
    AVG(arr_delay_new) AS arr_delay
    FROM Airlines A
    LEFT OUTER JOIN Flight_delays F
    ON A.airline_id = F.airline_id
    WHERE day_of_month between '1' and '23'
    GROUP BY airline_name, A.airline_id
),
secon
AS
(
    SELECT A.airline_id, airline_name,
    AVG(arr_delay_new) AS arr_delay
    FROM Airlines A
    LEFT OUTER JOIN Flight_delays F
    ON A.airline_id = F.airline_id
    WHERE day_of_month between '24' and '31'
    GROUP BY airline_name, A.airline_id
)
SELECT TOP 1 MAX(sub.del) AS delay_increase, sub.airline_name
FROM (SELECT (M.arr_delay - N.arr_delay) AS del, N.airline_name AS airline_name
FROM firs N
```

```
LEFT OUTER JOIN secon M
ON N.airline_id = M.airline_id
GROUP BY N.airline_name, N.arr_delay, M.arr_delay) AS sub
GROUP BY sub.airline_name
ORDER BY delay_increase DESC
```

Table 8: 1 records

delay_increase	airline_name
0.584763	Southwest Airlines Co.: WN

ZADANIE 9

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

```
SELECT DISTINCT airline_name
FROM   Flight_delays F
LEFT OUTER JOIN Airlines A
ON F.airline_id = A.airline_id
WHERE origin_city_name = 'San Francisco, CA' and dest_city_name = 'Eugene, OR'
and A.airline_id = SOME(SELECT F.airline_id FROM Flight_delays F
WHERE F.dest_city_name =  'Portland, OR'
and F.origin_city_name = 'San Francisco, CA')
```

Table 9: 2 records

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

ZADANIE 10

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

```
SELECT origin, dest, AVG(arr_delay_new) AS 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