

# analiza.Rmd

Tomasz Adamczyk

17 04 2020

## Zadanie 1

Jakie było średnie opóźnienie przylotu?

```
zadanie1 <- sqlQuery(con, "SELECT AVG(arr_delay_new) AS avg_delay FROM Flight_delays;")
```

avg_delay
15.91152

## Zadanie 2

Jakie było maksymalne opóźnienie przylotu?

```
zadanie2 <- sqlQuery(con, "SELECT MAX(arr_delay_new) AS max_delay FROM Flight_delays;")
```

max_delay
1895

## Zadanie 3

Który lot miał największe opóźnienie przylotu? [przewoźnik, miasto wylotu, miasto przylotu, data lotu, opóźnienie]

```
zadanie3 <- sqlQuery(con, "SELECT TOP 1 carrier, origin_city_name, dest_city_name, fl_date, arr_delay_new FROM Flight_delays ORDER BY arr_delay_new DESC;")
```

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

## Zadanie 4

Które dni tygodnia są najgorsze do podróżowania? [tabela zawierająca dla każdego dnia tygodnia średni czas opóźnień]

```
zadanie4 <- sqlQuery(con, "SELECT W.weekday_name, AVG(F.arr_delay_new) AS avg_delay FROM
Weekdays W INNER JOIN Flight_delays F ON W.weekday_id = F.day_of_week
GROUP BY W.weekday_name ORDER BY avg_delay DESC")
```

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

## Zadanie 5

Które linie lotnicze latające z San Francisco (SFO) mają najmniejsze opóźnienia przylotu? [tabela zawierająca nazwę przewoźnika oraz średnie opóźnienie z jego wszystkich lotów]

```
zadanie5 <- sqlQuery(con, "SELECT A.airline_name, AVG(F1.arr_delay_new) AS avg_delay FROM
Airlines A INNER JOIN Flight_delays F ON A.airline_id = F.airline_id
INNER JOIN Flight_delays F1 ON F.airline_id = F1.airline_id WHERE
F.origin = 'SFO' GROUP BY A.airline_name ORDER BY avg_delay DESC")
```

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

## Zadanie 6

Jaka część linii lotniczych ma regularne opóźnienia, tj. jej lot ma średnio co najmniej 10 min. opóźnienia? [tylko linie lotnicze występujące w tabeli Flight\_delays]

```
zadanie6 <- sqlQuery(con, "SELECT (SELECT TOP 1 COUNT(*) OVER() FROM Flight_delays
GROUP BY airline_id HAVING AVG(arr_delay_new) > 10) / CAST((SELECT
TOP 1 COUNT(*) OVER() FROM Flight_delays GROUP BY airline_id)
AS REAL) AS late_proportion")
```

late_proportion
0.8333333

## Zadanie 7

Jak opóźnienia wylotów wpływają na opóźnienia przylotów? [współczynnik korelacji Pearsona między czasem opóźnienia wylotów a czasem opóźnienia przylotów]

```
zadanie7 <- sqlQuery(con, "SELECT (AVG(dep_delay_new * arr_delay_new)
- (AVG(dep_delay_new) * AVG(arr_delay_new))) / (STDEVP(dep_delay_new)
* STDEVP(arr_delay_new)) AS 'Pearsons r' FROM Flight_delays")
```

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? [nazwa przewoźnika oraz wzrost]

```
zadanie8 <- sqlQuery(con, "SELECT TOP 1 S.airline_name, SUM(S.average_delay) AS
delay_increase FROM (SELECT A.airline_name, AVG(F.arr_delay_new) AS
average_delay FROM Flight_delays F INNER JOIN Airlines A ON
A.airline_id = F.airline_id WHERE (F.fl_date BETWEEN '2017-07-24' AND
'2017-07-31') GROUP BY A.airline_name UNION SELECT A.airline_name,
-AVG(F.arr_delay_new) AS average_delay FROM Flight_delays F
INNER JOIN Airlines A ON A.airline_id = F.airline_id WHERE (F.fl_date
BETWEEN '2017-07-01' AND '2017-07-23') GROUP BY A.airline_name) AS S
GROUP BY S.airline_name ORDER BY delay_increase DESC")
```

airline_name	delay_increase
Southwest Airlines Co.: WN	0.584763

## Zadanie 9

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

```
zadanie9 <- sqlQuery(con, "SELECT DISTINCT A.airline_name FROM Airlines A INNER JOIN
Flight_delays F ON A.airline_id = F.airline_id INNER JOIN
Flight_delays F1 ON F.airline_id = F1.airline_id WHERE (F.origin
= 'SFO' AND F.dest = 'PDX') AND (F1.origin = 'SFO' AND F1.dest
= 'EUG') ORDER BY A.airline_name ASC")
```

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? [tabela zawierająca jako miejsce wylotu Midway (MDW) lub O'Hare (ORD), jako miejsce przylotu San Francisco (SFO), San Jose (SJC) lub Oakland (OAK) oraz średni czas opóźnienia przylotu dla wylotów po 14:00 czasu lokalnego (atrybut crs\_dep\_time); wyniki pogrupowane po miejscu wylotu i przylotu, posortowane malejąco]

```
zadanie10 <- sqlQuery(con, "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 dest, origin ORDER BY avg_delay DESC")
```

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