BAD_Analytics_Project

Marcin Kostrzewski

13/05/2019

Raport został wygenerowany przy użyciu danych udostępnionych przez Departament Transportu Stanów Zjednoczonych.

Jakie było średnie opóźnienie przylotu?

```
SELECT CAST(AVG(arr_delay_new) AS NUMERIC(30,3)) AS 'Average delay (minutes)'
FROM Flight_delays
WHERE arr_delay_new IS NOT NULL;
```

Average delay (minutes)
15.912

Jakie było maksymalne opóźnienie przylotu?

```
SELECT CAST(MAX(arr_delay_new) AS NUMERIC(30,3)) AS 'Max delay (minutes)'
FROM Flight_delays
WHERE arr_delay_new IS NOT NULL;
```

Max delay (minutes)
1895

Który lot miał największe opóźnienie przylotu?

```
SELECT carrier AS 'Carrier',
       origin_city_name AS 'Origin',
       dest_city_name AS 'Destination',
       fl_date AS 'Date',
       arr_delay_new AS 'Delay (minutes)'
FROM   Flight_delays
WHERE  arr_delay_new = (SELECT MAX(arr_delay_new)
                        FROM   Flight_delays
                        WHERE  arr_delay_new IS NOT NULL);
```

Carrier	Origin	Destination	Date	Delay (minutes)
AA	Kona, HI	Los Angeles, CA	2017-07-26	1895

Które dni tygodnia są najgorsze do podróżowania?

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

Day	Average Delay (minutes)
Friday	20.80747
Monday	18.04801
Wednesday	16.10514
Thursday	15.64696
Saturday	15.21876
Tuesday	12.88056
Sunday	12.77606

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

```
SELECT   F1.carrier AS 'Carrier',
         (SELECT   AVG(F2.arr_delay_new) AS 'avg_delay'
          FROM     Flight_delays F2
          WHERE    F1.carrier = F2.carrier
          GROUP BY F2.carrier) AS 'Delay (minutes)'
FROM     Flight_delays F1
WHERE    F1.origin_city_name LIKE 'San Francisco%'
GROUP BY F1.carrier
ORDER BY "Delay (minutes)" DESC;
```

Carrier	Delay (minutes)
B6	28.841148
F9	18.980300
AA	18.375314

Carrier	Delay (minutes)
UA	16.950403
OO	16.808273
VX	13.964467
WN	13.823983
DL	12.258788
AS	7.453927
HA	4.202719

Pojawiające się w tabeli wartości ujemne oznaczają, że średnio samoloty lądowały wcześniej, niż przewidziano, czyli były przyśpieszone.

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

```
WITH reg_dlys
AS
(
SELECT   F1.carrier,
         (SELECT AVG(arr_delay_new)
          FROM   Flight_delays F2
          WHERE  F1.carrier=F2.carrier
          GROUP BY F2.carrier
          HAVING AVG(F1.arr_delay_new)>10) AS 'avg_delay'
FROM     Flight_delays F1
GROUP BY F1.carrier
),
carriers_sum
AS
(
SELECT COUNT(*)  AS 'sum'
FROM   (SELECT   COUNT(*) AS 'count'
        FROM     Flight_delays
        GROUP BY carrier) AS T
)
SELECT CAST((SELECT COUNT(*)
        FROM   reg_dlys t
        WHERE  t.avg_delay IS NOT NULL) /
                    CAST((SELECT *
                          FROM   carriers_sum) AS FLOAT) AS NUMERIC(5,3))
               AS 'Part of continuous delays';
```

Part of continuous delays
0.833

Jak opóźnienia wylotów wpływają na opóźnienia przylotów?

```
data <- DBI::dbGetQuery(con, "SELECT dep_delay_new,
       arr_delay_new
FROM   Flight_delays;")

library(knitr)
res <- cor(data, use = "complete.obs", method = "pearson")
kable(res[2:2])
```

x
0.9763465

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

```
WITH A1_23avg
AS
(
SELECT   carrier,
         AVG(arr_delay_new) AS 'avg'
FROM     Flight_delays
WHERE    day_of_month BETWEEN 1 AND 23
GROUP BY carrier
),
A24_31avg
AS
(
SELECT   carrier,
         AVG(arr_delay_new) AS 'avg'
FROM     Flight_delays
WHERE    day_of_month BETWEEN 24 AND 31
GROUP BY carrier
)
SELECT   TOP 1 T1.carrier AS 'Carrier',
               T2.avg-T1.avg AS 'Delay growth'
FROM     A1_23avg T1 INNER JOIN A24_31avg T2
                     ON T1.carrier = T2.carrier
ORDER BY T2.avg-T1.avg DESC;
```

Carrier	Delay growth
WN	0.584763

Wartości ujemne oznaczają spadek opóźnień

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

```
SELECT DISTINCT carrier
FROM   Flight_delays
WHERE  origin LIKE 'SFO'
       AND dest LIKE 'PDX'

INTERSECT

SELECT DISTINCT carrier
FROM   Flight_delays
WHERE  origin LIKE 'SFO'
       AND dest LIKE 'EUG'
```

carrier
OO
UA

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

```
SELECT   origin AS 'Origin',
         dest AS 'Destination',
         AVG(arr_delay_new) AS 'Delay'
FROM     Flight_delays
WHERE    origin IN ('MDW', 'ORD')
         AND dest IN ('SFO', 'SJC', 'OAK')
         AND crs_dep_time>1400
GROUP BY origin, dest
ORDER BY AVG(arr_delay_new) DESC;
```

Origin	Destination	Delay
ORD	SFO	22.19253
MDW	SFO	19.85714
MDW	SJC	17.20000
ORD	SJC	14.81111
MDW	OAK	12.12903