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?

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

Average delay (minutes)
8.295

Jakie było maksymalne opóźnienie przylotu?

```sql
SELECT CAST(MAX(arr_delay)/60 AS NUMERIC(30,3)) AS 'Max delay (hours)'
FROM Flight_delays
WHERE arr_delay IS NOT NULL;
```

Max delay (hours)
31.583

Który lot miał największe opóźnienie przylotu?

```sql
SELECT carrier AS 'Carrier',
       origin_city_name AS 'Origin',
       dest_city_name AS 'Destination',
       fl_date AS 'Date',
       arr_delay AS 'Delay (minutes)'
FROM   Flight_delays
WHERE  arr_delay = (SELECT MAX(arr_delay)
                    FROM   Flight_delays
                    WHERE  arr_delay 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?

```sql
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) AS 'Average Delay (minutes)'
FROM     Flight_delays
GROUP BY day_of_week
ORDER BY AVG(arr_delay) DESC;
```

Day	Average Delay (minutes)
Friday	14.452013
Monday	10.537501
Thursday	8.479856
Wednesday	8.456190
Saturday	7.544555
Tuesday	4.631525
Sunday	4.211660

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

```sql
SELECT   F1.carrier AS 'Carrier',
         (SELECT   AVG(F2.arr_delay) 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)" ASC;
```

Carrier	Delay (minutes)
AS	-1.1098174
HA	-0.2698805
DL	2.7830355

Carrier	Delay (minutes)
VX	6.0663173
UA	6.7303330
WN	8.7777055
OO	9.0998019
F9	10.1060463
AA	11.1863163
B6	21.4594984

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)
          FROM   Flight_delays F2
          WHERE  F1.carrier=F2.carrier
          GROUP BY F2.carrier
          HAVING AVG(F1.arr_delay)>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 (SELECT COUNT(*)
        FROM   reg_dlys t
        WHERE  t.avg_delay IS NOT NULL) /
                CAST((SELECT *
                      FROM   carriers_sum) AS FLOAT)
             AS 'Part of continuous delays';
```

Part of continuous delays
0.333333333333333

Jak opóźnienia wylotów wpływają na opóźnienia przylotów?

```r
data <- DBI::dbGetQuery(con, "SELECT dep_delay,
       arr_delay
FROM   Flight_delays
WHERE  dep_delay IS NOT NULL AND arr_delay IS NOT NULL;")

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

x
0.9640129

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

```sql
WITH A1_23avg
AS
(
SELECT    carrier,
          AVG(avg1) AS 'avg'
FROM      (SELECT    carrier,
                     AVG(arr_delay) AS 'avg1'
           FROM      Flight_delays
           WHERE     month = 7
           GROUP BY carrier, day_of_month
           HAVING    day_of_month BETWEEN 1 AND 23) AS T1
GROUP BY  carrier
),
A24_31avg
AS
(
SELECT    carrier,
          AVG(avg2) AS 'avg'
FROM      (SELECT    carrier,
                     AVG(arr_delay) AS 'avg2'
           FROM      Flight_delays
           WHERE     month = 7
           GROUP BY carrier, day_of_month
           HAVING    day_of_month BETWEEN 24 AND 31) AS T2
GROUP BY  carrier
)
SELECT    T1.carrier AS 'Carrier',
          T1.avg-T2.avg AS 'Delay growth'
FROM      A1_23avg T1 INNER JOIN A24_31avg T2
```

```
                     ON T1.carrier = T2.carrier
ORDER BY T1.avg-T2.avg DESC;
```

Carrier	Delay growth
EV	10.7838577
UA	8.4554257
NK	7.5085901
F9	6.3302250
B6	5.5323884
AA	5.4171770
DL	2.0241505
OO	0.1918574
AS	-0.1368039
HA	-0.5680958
VX	-0.9593595
WN	-1.3205764

Wartości ujemne oznaczają spadek opóźnień

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

```
SELECT DISTINCT carrier AS 'Carrier'
FROM   Flight_delays
WHERE  origin LIKE 'SFO'
       AND
       (dest IN ('PDX', 'EUG'))
```

Carrier
AS
OO
UA
VX
WN

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) 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) DESC;
```

Origin	Destination	Delay
MDW	SFO	15.114286
ORD	SFO	14.215517
MDW	SJC	10.600000
ORD	SJC	7.311111
MDW	OAK	4.758064