

Projekt 1: linie lotnicze

Katarzyna Pierzyńska

2023-05-14

Projekt 1 - Język SQL w analizie danych

Projekt przedstawiający analizę danych dotyczących opóźnień połączeń lotniczych w USA w lipcu 2017 r.

Zadanie 1

Jakie było średnie opóźnienie przylotu?

```
SELECT CAST (avg(arr_delay_new) AS varchar(16)) AS "avg_delay"
FROM "Flight_delays";
```

Table 1: 1 records

avg_delay
15.9115212681785

Zadanie 2

Jakie było maksymalne opóźnienie przylotu?

```
SELECT max(arr_delay) AS "max_delay"
FROM "Flight_delays";
```

Table 2: 1 records

max_delay
1895

Zadanie 3

Który lot miał największe opóźnienie przylotu?

```
SELECT carrier,
       origin_city_name,
       dest_city_name,
       fl_date,
       arr_delay_new
FROM "Flight_delays"
WHERE arr_delay_new IN (SELECT max(arr_delay_new)
                        FROM "Flight_delays");
```

Table 3: 1 records

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?

```
SELECT W.weekday_name,
       CAST (avg(F.arr_delay_new) AS varchar(16)) 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(F.arr_delay_new) DESC;
```

Table 4: 7 records

weekday_name	avg_delay
Friday	20.8074719468028
Monday	18.0480053678174
Wednesday	16.1051434386363
Thursday	15.6469561616281
Saturday	15.2187564809822
Tuesday	12.8805640843657
Sunday	12.7760570000742

Zadanie 5

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

```
SELECT A.airline_name,
       avg(D.arr_delay_new) AS "avg_delay"
FROM   "Airlines" A
       INNER JOIN "Flight_delays" F
              ON A.airline_id = F.airline_id
              AND origin = 'SFO'
       INNER JOIN "Flight_delays" D
              ON A.airline_id = D.airline_id
GROUP BY A.airline_name
ORDER BY avg(D.arr_delay_new) DESC;
```

Table 5: Displaying records 1 - 10

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

airline_name	avg_delay
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?

```
SELECT CAST(count(DISTINCT carrier) AS float) / cast((SELECT count(DISTINCT carrier)
FROM "Flight_delays") AS float) AS "late_proportion"
FROM "Flight_delays"
WHERE carrier IN (SELECT carrier
FROM "Flight_delays"
GROUP BY carrier
HAVING avg(dep_delay_new) >= 10);
```

Table 6: 1 records

late_proportion
0.8333333

Zadanie 7

Jak opóźnienia wylotów wpływają na opóźnienia przylotów?

```
SELECT (avg(dep_delay_new * arr_delay_new) - avg(dep_delay_new) * avg(arr_delay_new))
/ (STDDEV(dep_delay_new) * STDDEV(arr_delay_new)) AS "Pearsons r"
FROM "Flight_delays";
```

Table 7: 1 records

Pearsons r
0.9737081

Zadanie 8

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

```
SELECT R.airline_name,
avg(B.arr_delay_new) - avg(A.arr_delay_new) AS "delay_increase"
FROM "Flight_delays" A
INNER JOIN "Airlines" R
ON A.airline_id = R.airline_id
AND A.day_of_month <= 23
INNER JOIN "Flight_delays" B
on A.airline_id = B.airline_id
AND B.day_of_month > 23
GROUP BY R.airline_name
ORDER BY "delay_increase" DESC
LIMIT 1;
```

Table 8: 1 records

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)?

```
SELECT DISTINCT airline_name
FROM   "Airlines" R
      INNER JOIN "Flight_delays" F
        ON R.airline_id = F.airline_id
        AND F.origin = 'SFO'
        AND F.dest = 'PDX'
      INNER JOIN "Flight_delays" F2
        ON F.airline_id = F2.airline_id
        AND F2.origin = 'SFO'
        AND F2.dest = 'EUG';
```

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