

# Linie lotnicze Projekt 2023

Katarzyna Kuryło

2023-05-14

Projekt nr 1 - Język SQL w analizie danych

Projekt przedstawiający polecenia wraz z wygenerowanymi tabelami z analizy bazy danych dotyczących opóźnień połączeń lotniczych w USA w lipcu 2017 r.

Jakie było średnie opóźnienie przylotu?

```
SELECT avg(arr_delay_new) AS avg_delay FROM public."Flight_delays"
```

Table 1: 1 records

avg_delay
15.91152

Jakie było maksymalne opóźnienie przylotu?

```
SELECT max(arr_delay_new) AS max_delay FROM public."Flight_delays"
```

Table 2: 1 records

max_delay
1895

Który lot miał największe opóźnienie przylotu?

```
SELECT carrier, origin_city_name, dest_city_name, fl_date, arr_delay_new  
FROM public."Flight_delays"  
WHERE arr_delay_new =(SELECT max(arr_delay_new)  
FROM public."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

Które dni tygodnia są najgorsze do podróżowania?

```
SELECT weekday_name, avg_delay
FROM public."Weekdays" AS W
JOIN (SELECT day_of_week, avg(arr_delay_new) AS avg_delay
FROM public."Flight_delays"
GROUP BY day_of_week) AS F ON
W.weekday_id = F.day_of_week
ORDER BY avg_delay DESC
```

Table 4: 7 records

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

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

```
SELECT a.airline_name, b.avg_delay
FROM public."Airlines" a
JOIN (SELECT airline_id, avg_delay
FROM (SELECT airline_id, avg(arr_delay_new) AS "avg_delay"
FROM public."Flight_delays"
GROUP BY airline_id) AS f
WHERE airline_id IN (SELECT DISTINCT airline_id
FROM public."Flight_delays"
WHERE origin = 'SFO')) AS b ON
a.airline_id = b.airline_id
ORDER BY b.avg_delay 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
Delta Air Lines Inc.: DL	12.258788
Alaska Airlines Inc.: AS	7.453927
Hawaiian Airlines Inc.: HA	4.202719

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

```
SELECT CAST((SELECT count(*)
FROM(SELECT avg(arr_delay_new) AS avg_delay
FROM public."Flight_delays"
GROUP BY airline_id) AS E
WHERE avg_delay>10) AS float) / CAST((SELECT count(*)
FROM(SELECT avg(arr_delay_new) AS avg_delay
FROM public."Flight_delays"
GROUP BY airline_id) AS F)AS float) AS "late_proportion"
```

Table 6: 1 records

late_proportion
0.8333333

Jak opóźnienia wylotów wpływają na opóźnienia przylotów?

```
SELECT CORR(dep_delay_new,arr_delay_new) FROM public."Flight_delays"
```

Table 7: 1 records

corr
0.9763465

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

```
SELECT a.airline_name, w.delay_increase
FROM public."Airlines" a
JOIN (SELECT second_avg-first_avg AS "delay_increase", A.airline_id
FROM(SELECT avg(arr_delay_new) AS "second_avg", airline_id
FROM public."Flight_delays"
WHERE day_of_month IN(24, 25, 26, 27, 28, 29, 30, 31)
GROUP BY airline_id) AS A
JOIN
(SELECT avg(arr_delay_new) "first_avg", airline_id
FROM public."Flight_delays"
WHERE day_of_month NOT IN(24, 25, 26, 27, 28, 29, 30, 31)
GROUP BY airline_id
LIMIT 1) AS B ON
A.airline_id = B.airline_id
ORDER BY delay_increase DESC) AS w ON
a.airline_id = w.airline_id
```

Table 8: 1 records

airline_name	delay_increase
Southwest Airlines Co.: WN	0.584763

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

```
SELECT airline_name
FROM public."Airlines" AS A
JOIN (SELECT airline_id
FROM(SELECT DISTINCT airline_id, origin, dest
FROM public."Flight_delays"
WHERE(origin = 'SFO'
      AND dest IN ('PDX', 'EUG')))) AS X
GROUP BY airline_id
HAVING count(airline_id)>1
) AS F ON
A.airline_id = F.airline_id
```

Table 9: 2 records

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

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 public."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_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